
Integrated “omics” approaches to study non-orthodox seed germination: the case of holm oak (Quercus ilex subsp. ballota [Desf.] Samp)

Aproximaciones -ómicas al estudio de la germinación de semillas de especies recalcitrantes: el caso de la encina (Quercus ilex subsp. ballota)



María Cristina Romero Rodríguez

Doctoral Thesis

2015



TITULO: *Integrated "omics" approaches to study non-orthodox seed germination: the case of holm oak (Quercus ilex subsp. ballota [Desf.] Samp)*

AUTOR: *María Cristina Romero Rodríguez*

© Edita: Servicio de Publicaciones de la Universidad de Córdoba. 2015
Campus de Rabanales
Ctra. Nacional IV, Km. 396 A
14071 Córdoba

www.uco.es/publicaciones
publicaciones@uco.es



Departamento de Bioquímica y Biología Molecular

**Integrated “omics” approaches to study non-orthodox
seed germination: the case of holm oak (*Quercus ilex*
subsp. *ballota* [Desf.] Samp)**

Aproximaciones -ómicas al estudio de la germinación de semillas de especies
recalcitrantes: el caso de la encina (*Quercus ilex* subsp. *ballota*)

María Cristina Romero Rodríguez

Córdoba

2015

Imagen de portada tomada de : <http://m.inmagine.com/image-ptg00440920-Acorn-germinating.html>

Imagen de contraportada tomada de: <http://destilerias.blogspot.com.es/2013/03/de-bellotas-sabor-iberico.html>



Departamento de Bioquímica y Biología Molecular

Integrated “omics” approaches to study non-orthodox seed germination: the case of holm oak (*Quercus ilex* subsp. *ballota* [Desf.] Samp)

Aproximaciones -ómicas al estudio de la germinación de semillas de especies recalcitrantes: el caso de la encina (*Quercus ilex* subsp. *ballota*)

Trabajo realizado por María Cristina Romero Rodríguez, Licenciada en Bioquímica para optar al grado de Doctora

Los directores

Nieves Abril y Jesús Jorrín Novo

Córdoba 2015

Nieves Abril, Profesora titular del Departamento de Bioquímica y Biología Molecular de la Universidad de Córdoba

Jesús V. Jorrín Novo, Catedrático del Departamento de Bioquímica y Biología Molecular de la Universidad de Córdoba

INFORMAN

Que el trabajo titulado "**Aproximaciones -ómicas al estudio de la germinación de semillas de especies recalcitrantes: el caso de la encina (*Quercus ilex subsp. ballota*)**" realizado bajo nuestra dirección por **María Cristina Romero Rodríguez**, reúne los requisitos exigidos para la obtención del grado de Doctora, a los que se refieren las normas reguladoras de los estudios de doctorado de la Universidad de Córdoba.

Y para que así conste y surta los efectos oportunos, expedimos el presente informe

Córdoba, 3 de marzo de 2015



Nieves Abril



Jesús Jorrín Novo

Departamento de Bioquímica y Biología Molecular
Universidad de Córdoba

D. Conrado Moreno Vivián, Director del Departamento de Bioquímica y Biología Molecular de la Universidad de Córdoba

INFORMA:

Que el trabajo titulado **“Aproximaciones -ómicas al estudio de la germinación de semillas de especies recalcitrantes: el caso de la encina (*Quercus ilex* subsp. *ballota*)”** realizado por **María Cristina Romero Rodríguez** en el Departamento que dirige, puede ser presentado para su exposición y defensa como Tesis Doctoral en la Universidad de Córdoba.

Y para que así conste, expido el presente informe

Córdoba, 3 de marzo de 2015



Conrado Moreno Vivián

Departamento de Bioquímica y Biología Molecular
Universidad de Córdoba



TÍTULO DE LA TESIS:

Aproximaciones -ómicas al estudio de la germinación de semillas de especies recalcitrantes: el caso de la encina (*Quercus ilex subsp. ballota*)

DOCTORANDA:

María Cristina Romero Rodríguez

INFORME RAZONADO DEL/DE LOS DIRECTOR/ES DE LA TESIS

La Tesis Doctoral que se presenta tiene como objetivo el estudio a nivel molecular de la germinación de una semilla recalcitrante, como es la semilla de *Quercus ilex*. Para ello, se han realizado trabajos de investigación centrados en el análisis transcripcional, proteínico y de metabolitos durante el proceso de germinación y crecimiento temprano de la plántula. Los resultados del análisis de proteínas, mRNAs, hormonas y azúcares nos ha permitido establecer las similitudes y diferencias entre semillas ortodoxas y recalcitrantes, aportando nuevos conocimientos a proceso de germinación de las últimas. El presente trabajo constituye una importante aportación al conocimiento de la biología de la semilla de encina. Además, la metodología puesta a punto es de utilidad traslacional en el campo forestal, en lo que hace referencia a la caracterización de biodiversidad, conservación de semillas, y en definitiva de manejo sostenible de recursos forestales. Los cuatro capítulos en los que se han organizado los resultados de la tesis están siendo preparados para enviarlos a revistas científicas del campo para su publicación. No obstante la autora del presente trabajo consta con varias publicaciones que se anexan a la tesis doctoral, de la cuales una es derivada directamente de la presente tesis.

Desde el punto de vista académico la presente tesis ha tenido un grado formativo importante en cuanto al uso y manejo de un buen número de técnicas, incluyendo aproximaciones clásicas y las más recientes "ómicas", así como la integración de datos y el uso de bioinformática en la dirección de la denominada biología de sistemas.

Los Directores informan que la presente tesis cumple con creces los objetivos originalmente propuestos, con resultados más que satisfactorios, ajustándose sin lugar a dudas a los estándares científicos de calidad.

Por todo ello, autorizan la presentación de la tesis doctoral.

Córdoba, 04 de marzo de 2015

Firma del/de los director/es

Fdo.: 
Prof. Nieves Abril, PhD

Fdo.: 
Prof. Jesús Jorrín Novo, PhD

Este trabajo ha sido realizado en el Departamento de Bioquímica y Biología Molecular de la Universidad de Córdoba y financiado por fondos de Ministerio de Ciencia e Innovación (AGL-2009-12243-C02-02), Junta de Andalucía y programa propio de la Universidad de Córdoba. **María Cristina Romero Rodríguez** ha sido beneficiaria de una beca del Programa de Becas de Itaipú Binacional - Paraguay.

Agradecimientos

A muy poco de cumplir con un sueño, a donde no hubiera llegado sin el apoyo y la intervención de mucha gente, a quienes les doy mi más sincero agradecimiento, mil gracias y de antemano pido disculpas si omito a alguien en las siguientes líneas.

Agradezco infinitamente a mis padres quienes me dieron una infancia feliz y desde muy pequeña me inculcaron valores que no se encuentran afuera, gracias por darme cuanto podían para que pueda llegar a cumplir las metas a las que he llegado, gracias por dejarme volar, gracias por comprender y apoyar mi inquieta curiosidad y por dejarme elegir mi camino, aunque eso significó mi ausencia en tantas ocasiones en que debería haber estado con ustedes. En estas líneas también quiero agradecer a mis abuelos, quienes con sus historias me han enseñado muchas cosas de la vida, a los que ya no están los llevo siempre en el corazón.

A mis hermanos, Gustavo y Carmen quienes a temprana edad tuvieron una hija que quizá les rompió todos sus esquemas y con mis padres, me han mimado y reñido cuando era necesario, gracias “mama chica” por hacerme rehacer las feas tareas en vacaciones. A mis cuñados Aurelia y Lilo, que han visto crecer a esta que escribe y han sabido tolerar cuando necesitaba de mis hermanos, lo que hacía de pequeña no cuenta, eso era defensa personal, eran dos intrusos que se estaban llevando mis hermanos, tenía que evitarlo de alguna manera, aunque ninguna de mis técnicas ha funcionado, luego me he dado cuenta que eran dos intruso con un gran corazón. Sin desmerecer a los demás permítanme agradecerle a Aurelia gracias por quererme como una hija más y tolerarme; y a mi hermano por haber confiado ciegamente en mí y apoyarme para llegar a las metas que he logrado, gracias “manano”.

Ahora les toca a mis pequeños, mis sobrinitos, ellos son el escape de todos los problemas, quienes con sus travesuras y sus inocentes planteamientos, enseñan, y te hacen olvidar de todo los malos ratos, ser tía es lo mejor que hay; aunque disfruté jugando con alguno de ellos, me siento culpable de haberme perdido la infancia de otros, aunque algunos ya están creciditos e incluso me pasan en la estatura, todos seguirán siendo mis sobrinitos, les quiero mucho pequeños, gracias por orden de aparición Valdo, Mónica, José, Gustavito, Elías, Ceci e Isaías. Uno de ellos merece una mención especial, hoy es el ángel de la familia y a quien Dios tuvo que llevar muy pronto porque necesitaba un “super ángel”, gracias José por estar siempre conmigo.

A mis directores de tesis, infinitamente gracias por todo. Nieves, gracias por tu confianza, paciencia y dedicación; por estar conmigo codo a codo en la mesa del laboratorio si era necesario, por animarme a seguir ante cualquier problema técnico que surgía, por no dejarme rendir ante un experimento con malos resultados, y hacerme ver que eso forma parte del aprendizaje y de la ciencia. Jesús, gracias por darme la oportunidad de formar parte de tu grupo y convencerme que vuelva para el doctorado, gracias por las lecciones magistrales de proteómica sin importar que ya estés alejado del laboratorio, eres sin duda un gran maestro, también gracias por la confianza que siempre me has tenido, por hacerme participe de tus actividades las cuales me hicieron crecer en varios aspectos de mi vida, gracias por la “presión” que a veces es necesaria, ahora ya lo entiendo. A ambos gracias por animarme en los duros momentos por los que pase, y por compartir mis alegrías. Con ustedes aprendí a parte de ciencia, grandes lecciones de vida. Sobre todo gracias por la calidez humana que tuvieron conmigo, lo cual hizo que esto de estar lejos de mi tierra sea más llevadero.

A Ana Maldonado, quien también formó parte de esta tesis y quien me enseñó a trabajar en el laboratorio la molécula más delicada con la que he trabajado nunca y que al principio me asustaba, el ARN, pero gracias a tus enseñanzas Ana mis ARNs deslumbraban integridad.

Gracias al Grupo de Evaluación y Restauración de Sistemas Agrícolas y Forestales por estar siempre dispuestos a ayudarnos con lo que hacía falta.

Gracias a los profesores Antonio Gill, Miguel Ángel Carvajal, y Carolina de la Torre por ayudarnos algunas determinaciones de metabolitos.

A Luis Valledor, una persona de la que se puede aprender con un solo minuto de charla, gracias por enseñarme y responder las dudas proteómicas desde alguna parte del mundo.

A la gente del laboratorio, a Mari Carmen por estar siempre pendiente de que no nos falte nada y de que estemos bien, a Manolo por su disposición en ayudar siempre, a los “piltrafas” como yo, Rosita, Irene, Sekvan y Luis Rodríguez gracias por las charlas, risas y los buenos momentos compartidos. Muhittin gracias por las correcciones del inglés. A los que ya se fueron y de los que he aprendido la proteómica del laboratorio y de escritorio Bisma, Jose, Raquel, Inma y Lyudmila, de corazón gracias. A Antonio Archidona gracias por las largas tardes de limpieza de bellotas, amenizadas con risas y Nutella. A los de la segunda planta a Noe, gracias por echarme una mano en lo que necesitaba, a Ricardo por estar dispuesto a aclarar dudas técnicas proteómicas. Todos ustedes fueron mi familia en el día a día.

A Cristina Huertas y Bruce George Sutherland por las correcciones del inglés del manuscrito, muchas gracias.

A la gente del servicio de Proteómica de SCAI de la Universidad de Córdoba, gracias por responder siempre a cualquier duda.

A mi familia en Córdoba, Marce y Diego gracias por estar y aguantarme, creo que haber vivido juntos nos mantuvo cerca de nuestro Paraguay, gracias por las aventuras, charlas, risas y muchas cosas más que hemos compartido, y por las que vendrán, compartir un hogar con ustedes fue una excelente experiencia.

A la Universidad Nacional de Asunción, que a través de sus representantes me han apoyado y me otorgaron el permiso correspondiente para poder hacer usufructo de la beca para la realización de la presente tesis doctoral, gracias. A mi querido Paraguay que ha permitido mi formación profesional, espero contribuir desde la Universidad con mi pequeño aporte en las mejoras que tanta falta nos hace.

Por supuesto, a mis amigos de aquí y allá; quienes estando cerca o lejos estuvieron en los buenos y malos momentos, y han hecho más placentera esta aventura, a todos ustedes gracias.

A la encantadora Córdoba, al “cordobé” que desde que crucé la primera palabra con uno de ellos (aunque no les haya entendido nada, y eso que hablamos el mismo idioma) sentí la calidez humana, y luego de todos estos años aquí me siento de este lugar y el cual llevaré conmigo siempre.

A mi familia política, que durante todo este tiempo siempre estuvieron pendiente que todo vaya bien, a mis suegros Teresita y Yoshimori, y mis cuñados Akira e Hitoshi, gracias.

Y no por ser el último sea el menos importante, a Julio mi compañero de la vida, mi cómplice, mi compañero de aventuras, quien conoce esto casi tanto como yo, quien siempre tiene una palabra de aliento, de ánimo, gracias por tu cariño, confianza e incondicional apoyo.

Permítanme dirigirles a todos unas palabras en mi dulce idioma guaraní “Aguije che py’aiteguive che pytyvõ haguere”, que en español significa “Muchas gracias por ayudarme”.

Córdoba, 2015

*A mis padres y hermanos, por que dieron todo por mí.
A José, nuestro ángel.
A Julio, por su cariño e incondicional apoyo.*

DOCTORAL THESIS

**INTEGRATED “OMICS” APPROACHES TO STUDY NON-
ORTHODOX SEED GERMINATION: THE CASE OF
HOLM OAK (*QUERCUS ILEX* SUBSP. *BALLOTA*)**

Index

Abbreviations	1
Summary	3
Resumen	5
Chapter 1: General introduction	7
1.1. Research with forest trees: the case of <i>Q. ilex</i>	10
1.2. Seeds, seed germination and seedling growth	12
1.3. Functional genomics. “OMICS” approaches in plant biology research	18
1.3.1. Transcriptomics approaches	19
1.3.2. Proteomics approaches	20
1.3.3. Metabolomics approaches	30
1.4. References	31
Chapter 2: Objectives	39
Chapter 3: Plant material	43
3.1. Plant material collection, study site and storage	45
3.2. Germination/seedling experiment and morphological aspect	46
3.3. References	48
Chapter 4: Transcriptional, hormonal and sugar content changes during germination and early development of <i>Quercus ilex</i> seeds	49
Abstract	51
4.1 Introduction	52
4.2. Materials & Methods	53
4.3. Results and discussion	62
4.4. Concluding remarks	81
4.5. References	82
Chapter 5: Subtractive libraries for prospecting differentially expressed genes between <i>Q. ilex</i> germinated seeds and seedlings	89
Abstract	91
5.1 Introduction	92
5.2. Materials & Methods	93
5.3. Results and discussion	99
5.4. Concluding remarks	111
5.5. Appendixes	112
5.6. References	120
Chapter 6: Changes in the protein profiles of <i>Q. ilex</i> seeds during germination and seedling	123
Abstract	125
6.1 Introduction	126
6.2. Materials & Methods	127
6.3. Results and discussion	132
6.4. Concluding remarks	150
6.5. Appendixes	152
6.6. References	176

Chapter 7: Phosphoproteomic analysis in <i>Q. ilex</i> germinated seeds and seedlings	181
Abstract	183
7.1 Introduction	184
7.2. Materials & Methods	185
7.3. Results and discussion	189
7.4. Concluding remarks	198
7.5. Appendixes	199
7.6. References	205
Chapter 8: General discussion	209
Chapter 9: Conclusions	219
Chapter 10: Appendixes	223
10.1. Appendix 1: Improving the quality of protein identification in non-model species. Characterization of <i>Quercus ilex</i> seed and <i>Pinus radiata</i> needle proteomes by using SEQUEST and custom databases	225
10.2. Appendix 2: Contributions	233

Abbreviations

ABA	Abscisic acid
ACO	1-aminocyclopropane-1-carboxylic acid oxidase
AGP	Glucose-1-phosphate adenyltransferase
BL	Brassinolide
BLAST	Basic local alignment search tool
BR	Brassinosteroid
BSA	Bovine serum albumin
CBB	Coomassie Brilliant blue
CDS	Coding sequence
CHAPS	3-[(3-cholamidopropyl)dimethylammonio] propanesulfonate
CID	Collision induced dissociation
Ck	Cytokinins
CLPB	Casein lytic proteinase B
CS	Castasterone
DHAPS	Phospho-2-dehydro-3-deoxyheptonate aldolase 1
DHAR	Dehydroascorbate reductase
DHN	Dehydrin
DTT	Dithiothreitol
DW	Dry weight
DZ	Dihydrozeatin
DZR	Dihydrozeatin riboside
EDTA	Ethylenediaminetetraacetic acid
EF	Elongation factor
ENO	Enolase
ESI	Electrospray ionization
EST	Expressed sequences tags
FDH	Formate dehydrogenase
FTMS	Fourier transform mass spectrometer
FW	Fresh weight
GA	Gibberellins
GAPDH	Glyceraldehyde 3-phosphato dehydrogenase
GASA	Giberellin-stimulated arabidopsis
GC-MS/MS	Gas Chromatography coupled to Mass Spectrometry
GOLS	Galactinol synthase
GSH	Glutathione
GST	Glutathione S-transferases
HSP	Heat shock protein
IAA	Indole acetic acid
iP	Isopentenyladenine
IPG	Immobilised pH gradients
iPR	Isopentenyladenine riboside
KEGG	Kyoto Encyclopedia of Genes and Genomes
LAP	Leucine aminopeptidase
LC	Liquid chromatography
LEA	Late embryogenesis abundant
LTQ	Linear trap quadrupole
MALDI-TOF	Matrix-assisted laser desorption/ionization – time of flight
MAP	Month after polinization
NBT	Nitro blue tetrazolium
NDH6	NADH dehydrogenase
NIRs	Near-infrared spectroscopy
NLP	Omega-amidase
NSA	Normalised spot abundance
NSAF	Normalized spectral abundance factor
OCP	Overexpressor of cationic peroxidase
OEE	Oxygen evolving enhancer protein
P3DB	Plant protein phosphorylation data base

PC	Principal component
PCA	Principal components analysis
PCR	Polymerase chain reaction
PetE	Plastocyanin a
PGK	Phosphoglycerate kinase
PMSF	Phenylmethylsulphonylfluoride
PP2C	Protein phosphatase 2 C
PPi-PFK	Pyrophosphate-dependent phosphofructokinase
PR	Pathogenesis related
Pro-Q DPS	Pro-Q Diamond phosphorylation staining
PVDF	Polyvinylidene difluoride
PVPP	Polyvinylpyrrolidone
qRT-PCR	Quantitative reverse transcription-PCR
RBCL	Ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit
RBCS	Ribulose 1,5-bisphosphate carboxylase/oxygenase small subunit
RFO	Raffinose
RIN	RNA integrity number
ROS	Reactive oxygen species
RWC	Relative water content
SA	Spot abundance
SAM	S-adenosylmethionine
SCF	<i>Skp1</i> -Cullin1-F-box
SDIR	Salt and Drought Induced RING Finger
SDS-PAGE	Sodium dodecyl sulfate- Polyacrylamide gel electrophoresis
SE	Somatic embryogenesis
Ser-CPs	Serine carboxypeptidase-like
SKP	S-phase kinase-associated protein
SOD	Superoxide dismutase
SPC	Spectra matching peptides
SSH	Subtractive suppressive hybridization
SUT	Sucrose transporter
TCA	Trichloroacetic acid
TEMED	Tetramethylethylenediamine
Trx	Thioredoxin
tZ	Trans-zeatin
tZR	Trans-zeatin riboside
UHPLC-MS	Ultra-high performance liquid chromatography coupled to Mass Spectrometry
WC	Water content
XTH	Xyloglucan endotransglycosylase

SUMMARY

This work is focused on holm oak (*Quercus ilex* L. subsp. *ballota* [Desf.] Samp.) a dominant tree species in natural forest ecosystems over large areas of the Western Mediterranean Basin. *Q. ilex* forest maintenance and sustainability are facing important problems and challenges related to seed viability/conservation, and plant mortality in both adult trees and young one-year-old plants after field transplantation resulting from adverse environmental conditions such as drought and the so-called decline syndrome.

In the present Doctoral Thesis, the germination and seedling growth of *Q. ilex* was studied applying a multidisciplinary “omics” approach combined with classical biochemical approaches, according to the actual trends in biosciences research. Within this general aim, specific objectives pursued in this thesis were:

1. Determine the expression profile of a selected group of protein coding genes involved in desiccation tolerance, regulation of ABA-signalling, metabolism and antioxidative defence.
2. Identify differentially expressed genes between germinated seeds and seedlings of *Q. ilex* by using a suppression subtractive hybridization (SSH).
3. Analyse changes in the proteomic profiles during germination and seedling growth through gel based and gel free approaches.
4. Analyse the dynamic protein phosphorylation changes during seed germination and seedling development.

The transcriptional analysis revealed that mature *Q. ilex* seeds show some of the intracellular physiological characteristics of orthodox seed that included (i) accumulation of non-reducing carbohydrates (sucrose) and protective proteins (DHN3) and (ii) accumulation of transcripts involved in the synthesis of certain osmoregulator of raffinose series oligosaccharides (*GolS*), the anti-oxidative defence (*Sod1*, *Gst*). But the holm oak mature acorns, like other recalcitrant seeds, have the ability to maintain a partially active metabolism. The transcriptional analysis results were verified and complemented with the determination of (i) plant hormones levels (ABA; the gibberellins GA3 y GA4; the auxin IAA and the cytokinins iP and iPR,) (ii) sugars accumulation (Suc, Glc, Fru) and (iii) proteins amounts determined by immunoblotting (DHN3, GAPDH, RBCL) and/or enzymatic activity (SOD) and good correlation was found with transcript levels.

The analysis of differentially expressed genes by SSH allowed the identification for the first time a large number of putative differentially expressed ESTs from the embryo axis in germinated seeds and from shoot seedlings of *Q. ilex* during the postgermination and seedling establishment. Thirty-one over-expressed genes were identified in germinated seeds at the germination stage. Proteins encoded by these genes are representative of eight functional categories: stress responses, transport, oxidation-reduction, cell wall modification, cell division cycle, protein metabolism, cellular component organization and translation. On the other hand, 39 non-redundant transcripts over-represented in *Q. ilex* shoots seedlings were grouped in seven functional categories: photosynthesis, secondary metabolism, transport, signalling, stress response, gene expression and cellular component organization. These data constitute an important genomics resource that should clearly benefit further germination and other biological process research on *Q. ilex*.

The gene expression analysis was complemented with proteomics analysis using different approaches. By a 2-DE approach, 103 variable spots for protein identification were selected and around 90 differentially proteins in all stages analysed were identified using 2-DE MALDI-TOF/TOF. The gel-based approach revealed important metabolic changes that occurred in the holm oak seed after the germination. Few proteins were differentially accumulated during the germination (from S0 to S3). A gel free approach was, hence, used to analyse the mature unimbibed and the germinated seed, trying to improve the coverage of proteome analysed. Through *n*LC- LTQ-Orbitrap- analysis of total extract, 153 differentially accumulated proteins were identified.

At posttranslational levels, phosphoproteomics changes were analysed. A total of 55 phosphoprotein differentially accumulated spots were detected, of which 20 putative phosphoproteins were identified by MALDI TOF-TOF analysis. They were representative of six functional categories: carbohydrate and amino acid metabolisms, protein folding, oxidation-reduction processes and stress response, and the metabolism of RNA.

As conclusion, the proteomics profile changes observed in *Q. ilex* seed germination and seedling establishment affected mainly the proteins related to carbohydrate metabolism (*i.e.* degradation sugars, activation of glycolysis), amino acid metabolism (*i.e.*, use of methionine for the synthesis of secondary metabolites, including ethylene) and oxidative stress response, which also found at the transcriptional level. However, the few changes detected at translational levels (protein) during germination (from S0 to S3), suggested that the mature non-orthodox seeds of *Q. ilex* have the mechanisms necessary to ensuring the rapid resume of the metabolic activities required to start and finish the germination process and to *de novo* synthesise the biomolecules required for growth, this could be explain the precocious germinations that present this seeds during storage even without imbibition. In addition, the phosphoproteomics analysis revealed that changes in phosphorylation status of protein, activated pathways necessary to finish germination. However, we cover only a minimal fraction of the phosphoproteome, therefore a large scale analysis of phosphorylation protein could be necessary to disclose more metabolic pathway involved in germination.

The results presented here increase the knowledge of the physiological and molecular changes that take place during *Q. ilex* seed germination and seedling establishment, and are the basis of knowledge that could contribute to long-term in management and storage of this recalcitrant seeds for the afforestation projects and restoration programmes under the impending climate change in Mediterranean regions.

RESUMEN

La presente Tesis Doctoral está centrada en el estudio de la encina (*Quercus ilex* L. subsp. *ballota* [Desf.] Samp.), una especie dominante en los sistemas silvopastorales mediterráneos. El mantenimiento y la sostenibilidad de los bosques de encina se enfrenta actualmente con problemas y retos mayormente relacionados con la conservación de semillas viables, ya que se trata de una semilla recalcitrante, y con la alta tasa de mortalidad de los individuos adultos y plántones luego de ser trasplantados a campo, resultado de las adversas condiciones ambientales.

En el presente trabajo se llevó a cabo el estudio de la germinación de semillas de encina, así como de plántulas en estadio temprano mediante una aproximación multidisciplinar que combina las técnicas clásicas de biología molecular con las “ómicas”. Con este objetivo general fueron propuestos los siguientes objetivos específicos

1. Realizar un análisis transcripcional de un grupo de genes candidatos implicados en tolerancia a la de desecación, regulación de la señalización de ABA, metabolismo y defensa contra el estrés oxidativo, a lo largo del proceso de germinación.
2. Identificar genes de expresión diferencial en semilla germinada y de plántulas muy jóvenes de *Q. ilex* mediante el uso de hibridación sustractiva por supresión (SSH).
3. Identificar proteínas cuya abundancia se ve alterada durante la germinación y desarrollo de plántulas en estadios tempranos, determinando los perfiles proteómicos mediante aproximaciones basadas y libres de gel.
4. Identificar el phosphoproteoma y los cambios que ocurren durante la germinación de semillas

Los resultados del análisis transcripcional del grupo de genes candidatos revelaron que las semillas maduras de *Q. ilex* presentan algunas características fisiológicas intracelulares propias de las semillas ortodoxas que incluyen (i) acumulación de azúcares (sacarosa) y proteínas protectoras (DHN3) y (ii) acumulación de transcritos implicados en la síntesis de oligosacáridos osmorreguladores de rafinosa (*GolS*) y en la defensa contra el estrés oxidativo (*Sod1*, *Gst*). Sin embargo, las semillas maduras de *Q. ilex*, como otras semillas recalcitrantes se dispersan con un alto contenido de agua y mantienen el metabolismo parcialmente activo. Los resultados de este análisis transcripcional fueron verificados y complementados con la determinación de (i) niveles de fitohormonas (ABA, GA3, GA4, IAA, iP and iPR), (ii) acumulación de azúcares (Sac, Glc, Fru) y (iii) determinación de proteínas mediante inmunodetección (DHN3, GAPDH, RBCL) y actividad enzimática (SOD); estos resultados mostraron una buena correlación con los transcritos analizados.

El análisis de expresión diferencial de genes mediante SSH permitió identificar por primera vez en *Q. ilex* un grupo de ESTs correspondiente a genes diferencialmente expresados en semillas germinadas y plántulas en desarrollo temprano. En las semillas germinadas se identificaron 31 genes sobre-expresados, entre los que se incluyen genes que codifican para proteínas implicadas en la respuesta a estreses, transporte, defensa contra estrés oxidativo, modificación de la pared celular, ciclo celular, metabolismo de proteínas y transducción. Por otro lado, 39 transcritos no redundantes se encontraron sobre-expresados en estadio temprano de plántulas. Las proteínas codificadas por estos transcritos fueron agrupadas en siete categorías funcionales que incluyen fotosíntesis, metabolismo secundario, transporte, señalización, respuesta a diversos estreses, expresión génica y organización de componentes

celulares. Estos datos constituyen un importante recurso genómico, hasta la fecha inexistente para *Q. ilex*, que podría ser utilizado para ampliar el análisis molecular de la germinación y de otros procesos biológicos o fisiológicos en la investigación de *Q. ilex* y otras especies filogenéticamente relacionadas.

Los análisis de expresión génica se complementaron con análisis proteómicos utilizando diferentes aproximaciones. Mediante análisis de geles 2-DE se determinaron 103 spots de proteínas cuya intensidad variable era estadísticamente significativa. De ellos, mediante MALDI TOF/TOF, se identificaron 90 proteínas cuya expresión cambia en todos los estadios analizados. Esta aproximación desveló los importantes cambios metabólicos que ocurren en la semilla de encina después de la germinación, pero detectó solo un pequeño número de cambios en la abundancia de proteínas durante la germinación propiamente dicha (desde S0 hasta S3). Para aumentar la cobertura del proteoma, y detectar los cambios a nivel proteico que ocurren durante la germinación, se utilizó una aproximación “*gel free -label free*” para analizar la semilla madura y germinada. El análisis del extracto proteico total mediante *n*LC LTQ-Orbitrap permitió identificar cambios de abundancia de 153 proteínas.

En el análisis proteómico, también fue incluido el estudio del fosfoproteoma y los cambios que ocurren a nivel de esta modificación postraduccional. En este análisis se detectó un total de 55 spots de fosfoproteínas cuyas intensidades mostraban diferencias estadísticamente significativas, de las cuales 20 fosfoproteínas fueron identificadas por MALDI-TOF/TOF. Estas proteínas fueron representativas de las siguientes categorías funcionales: metabolismo de carbohidratos y de aminoácidos, plegamiento de proteínas, defensa contra estrés oxidativo y estrés en general, y metabolismo de RNA.

Como conclusión, los cambios en los perfiles proteómicos observados durante la germinación y el desarrollo de plántulas en estadio temprano en *Q. ilex* afectó mayormente a proteínas relacionados con el metabolismo de carbohidratos (i.e., degradación de azúcares de reserva, potenciación de la glucólisis), metabolismo de amino ácidos (i.e., utilización de metionina para la síntesis de metabolitos secundarios, incluyendo etileno) y respuesta a estrés, en completa concordancia con los resultados obtenidos a nivel transcripcional. Sin embargo, los pocos cambios a nivel transduccional que se detectaron durante la germinación (S0 a S3) sugieren que la semilla madura de *Q. ilex* tiene la maquinaria bioquímica necesaria para asegurar la rápida reactivación metabólica y poder empezar y terminar el proceso de germinación, así como las biomoléculas necesarias para la síntesis “*de novo*” requerida para el crecimiento en la post-germinación. Todo esto explicaría la facilidad de germinación de las semillas durante el almacenamiento, incluso en ausencia de imbibición. El análisis fosfoproteómico reveló cambios en la fosforilación de diversas proteínas implicadas en activación de rutas metabólicas necesarias para la germinación. Sin embargo, la metodología empleada permitió cubrir sólo una mínima parte del fosfoproteoma, así que para revelar completamente la regulación de las rutas metabólicas implicadas en la germinación se necesitaría análisis a mayor escala.

Los resultados obtenidos en este trabajo contribuyen sin duda al conocimiento de los cambios a nivel fisiológico y molecular que se producen durante la germinación y desarrollo temprano de la plántula de *Quercus ilex*, y son la base del conocimiento que a largo plazo podrían contribuir al manejo y mantenimiento de esta semilla recalcitrante para los proyectos de reforestación y programas de restauración en el marco del inminente cambio climático en las regiones mediterráneas.

CHAPTER 1:
GENERAL INTRODUCTION

General introduction

Forests, including natural or artificial forests, cover approximately four billion hectares, about one-third of the total land area in the world. Forest trees constitute around 82% of the continental biomass and harbor more than 50% of the terrestrial biodiversity, with all that it implies for gene pools, pharmaceuticals, and other unique and valuable goods and services. Natural forests have evolved and reproduced themselves naturally without suffering significant anthropogenic modifications; they include closed forests (of broad-leaved and coniferous species) and open forests (or savannahs, also known as thin forests). They are a source of raw material for many of the essential needs of humans, including building material, paper products, firewood for heat and cooking, energy and many tree-crop foods¹.

Forests contain large amounts of sequestered carbon, and their destruction or degradation (especially by burning) is estimated to contribute between 10–30% of all CO₂ emissions into the atmosphere². Industrial and agricultural revolutions, urbanization, and other natural disasters, such as fire, pest and pathogens³ have caused an important deforestation.

Although, it is well recognized that study and research can have a decisive impact on sustainable and productive forest management, and despite of its economic and ecological relevance; biological, physiological and molecular studies on trees are almost insignificant if compared with model plant species, and especially with other living organisms such as *Saccharomyces*, mice or humans. Thus, an “ISI Web of Knowledge” search covering the period 1990–2015 (February), using as a search string “forest trees” generated 113,300 hits (55,400 for *Pinus*, 25,300 for *Quercus*, 18,700 for *Populus*), while the values for *Arabidopsis thaliana* and rice are 119,200 and more than 200,000, respectively.

This Doctoral Thesis is focused on holm oak (*Quercus ilex* L. subsp. *ballota* [Desf.] Samp.) a dominant tree species in natural forest ecosystems over large areas of the Western Mediterranean Basin⁴. *Q. ilex* forest maintenance and sustainability are facing important problems and challenges related to seed viability/conservation, and plant mortality in both adult trees and young one-year-old plants after field transplantation resulting from adverse environmental conditions like drought and the so-called decline syndrome⁵. Nowadays, forest restoration and reforestation are high priority objectives and *Q. ilex* has become a priority tree species for reforestation programs⁶. Thus, there is an increase in demand for holm oak seedlings and favouring their nursery production. In the present work transcriptomics, proteomics and metabolite analysis were combined to study at molecular level the germination and seedling growth of *Q. ilex*, to obtain a better understanding of molecular mechanisms

involved in these processes. This knowledge is essential for restoration and reforestation programs with this species.

In this section, the biological system (*Q. ilex*, an orphan tree species), the biological process (germination and seedling growth) as well as the experimental approaches used (functional genomics: transcriptomic, proteomic and metabolites analysis) will be briefly introduced.

1.1. Research with forest trees: the case of *Q. ilex*

The Fagaceae is a large angiosperm family with species spread throughout the northern hemisphere, from the tropical to the boreal regions. Oaks (*Quercus*), chestnuts (*Castanea*), and beeches (*Fagus*) are the only genera that are distributed in Asia, Europe, and North America, where they cover very large continuous forests and constitute important forest resources⁷. The genus *Quercus* (oaks) is one of the most widespread in the northern hemisphere, being the main components of the tree stratum in many forests and woodlands⁷. Holm oak (*Q. ilex*) is one of the main woody species in Mediterranean landscapes⁴. This species covers in Andalusia (southern Spain) around 750,000 Ha⁸ and, along with cork oak (*Q. suber* L.) constitute “dehesas” (Spanish) or “montados” (Portuguese), one of the most important European agroforest systems extending over almost four million hectares in the Iberian Peninsula⁹.

The “dehesa” ecosystem consists of oak woodland with an understory composed of a mosaic of croplands, grasslands and shrublands, where cattle, sheep, pigs and goats are extensively raised. Besides the great ecological importance, this species is of great economic interest because their seeds (acorns) are used for fattening pigs of the Iberian race, whose meat is the basis of a high-quality food industry⁹⁻¹¹. Most existing holm oak forests have originated through resprouting after disturbances such as fire, overgrazing, wood extraction or charcoal production¹². Besides all these factors, the survival of these oak species is threatened by, abandonment of agricultural practices, low natural regeneration and the oak decline and the current variable climatic conditions on Earth¹³⁻¹⁴.

Holm oak conservation and breeding programs are constrained by limited natural regeneration. Described as a recalcitrant species, maturation of holm oak acorns implies an incomplete desiccation process, which adapts the seed to a partial drought stress but does not completely inhibit the metabolism. Consequently, holm oak seeds suffer a rapid loss of viability on shedding¹⁵. Hence, acorns have a limited ability to persist in the field and generate a seedling bank able to play a major role in natural regeneration of holm oak stands at short or

medium temporal scales. Seed orchards are usually unable to meet the demand of improved seeds for plantation, making vegetative propagation highly desirable¹⁶. However, in oaks, the vegetative propagation by traditional techniques is impractical at the operational level. Somatic embryogenesis (SE) has been developed as a means of plant regeneration in several species of the *Quercus* genus¹⁷. SE is currently one of the main biotechnologies used for mass propagation and cryopreservation of genetic resources, which enables the implementation of multivarietal forestry, therefore, it has a great potential in genetic improvement programs^{10; 18}. However, clonal regeneration by SE or organogenesis is still difficult for many woody species and is often limited to the use of juvenile explants¹⁹. In a survey on forest biotechnology activities carried out by the FAO, 196 references, excluding genetic modifications, have been reported. Species surveyed belonged to 142 botanical genera, with 62% of the research carried out on less than 6 genera, including *Pinus* (20%), *Eucalyptus* (11%), *Picea* (9%), *Populus* (9%), *Quercus* (7%), and *Acacia* (6%)²⁰.

Tree breeding programs are designed to develop genetically improved varieties used for reforestation and afforestation, looking to increase the economic, environmental or social value of a forest, the latest being an integral component of most programs around the world. For such an objective, basic research and knowledge generated by the different physiological, biochemical and genetic disciplines, especially those related to molecular studies would be of great relevance^{1; 21-22}.

The emergence of structural, functional, and comparative genomics has ushered in a new era in plant biology. Until the 1990's, biological research was mainly focused on the *in vitro* studies of individual components in which genes and proteins were investigated one at a time. The development of high-throughput genomics technologies has made possible the ongoing and detailed characterization of the genetic foundations of life²³. This strategy later shifted to *in vivo* and molecular large-scale research. The advent of the genome era and the availability of a growing number of fully-sequenced genomes, from model species *Arabidopsis*²⁴, rice²⁵ to forest species as *Populus trichocarpa*²⁶ and *Eucalyptus grandis*²⁷, have changed the way in which research is performed, moving to the holistic, high-throughput analysis of the gene function by using the new "omics" technologies (transcriptomics, proteomics and metabolomics). In *Quercus*, a genera without a reference genome sequences, the expressed sequence tags (ESTs) generated by construction of suppression subtractive hybridization (SSH) libraries (*i.e.*, *Q. petraea*²⁸) or using RNAseq (*i.e.*, *Q. suber*²⁹) are important genomic resources to study related species such *Q. ilex* in which molecular analysis are scarce.

1.2. Seeds, seed germination and seedling growth

Plant seeds provide staple food for the world population. In addition, they are also important for the plant life cycle. Most plants generate their progenies through seeds, which can help them to avoid the adverse or even extreme environmental conditions. There are two types of seeds: orthodox seeds and recalcitrant. Orthodox seeds acquire desiccation tolerance during development, can dry to low water contents (generally less than 5%), and retain viability in the dry state for predictable periods. Recalcitrant seeds, on the other hand, are shed at high water contents, ranging from 0.4 to 4.0 g water per g dry matter (g/g) ($\approx 40\%$ of water content), are sensitive to desiccation, and are also metabolically active on shedding³⁰ (Fig. 1.1).

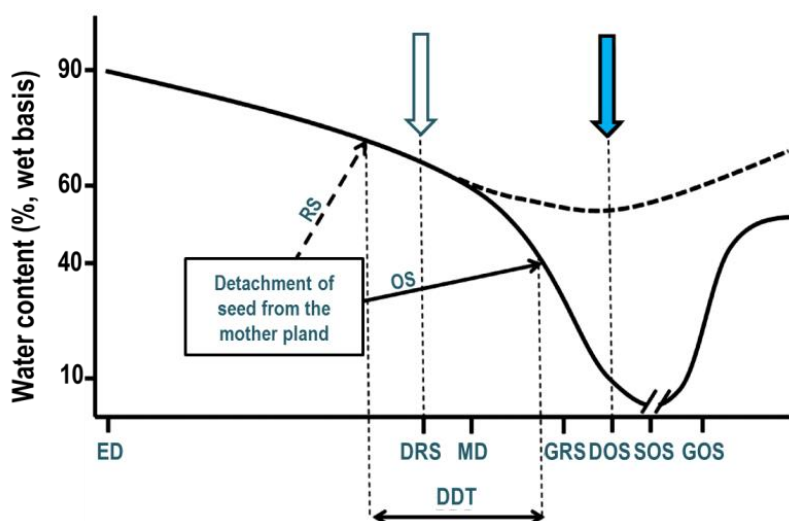


Figure 1.1: Schematic representation of the water content during maturation, dispersion, and germination of orthodox and recalcitrant seeds. Dotted line represents the behavior of recalcitrant seeds (RS) from the end of their maturation to the beginning of germination, and continuous line that of orthodox seeds (OS). ED: embryo starts development; DRS: dispersion of recalcitrant seeds (open arrow); DOS: dispersion of orthodox seeds (blue arrow); DDT: development of desiccation tolerance; MD: beginning of maturation drying; GRS: germination of recalcitrant seeds; SOS: storage of orthodox seeds; GOS: germination of orthodox seeds. SOS can vary from days to millennia. Taken from Barbedo *et al.*³¹

Quercus ilex seeds are considered recalcitrant, as they do not undergo maturation drying, and they are shed at relatively high water content, compared to orthodox seeds (Fig. 1.1). This makes them susceptible to desiccation injury, and to lose viability when stored for long periods, as reported in other *Quercus* spp.¹⁵. Problems associated with the loss of viability during collection and processing, and the short storage life of these seeds can have serious impacts on the afforestation projects and restoration programmes. For these reasons, it is important to understand why *Q. ilex* seeds do not achieve dehydration and remain metabolically active. Different aspects of recalcitrant seed metabolism, linked to germination, have been extensively examined, the main being the acquisition of desiccation sensitivity. Since metabolism in recalcitrant seeds is active also after shedding, dehydration would lead to its imbalance, causing significant cell damage and death of the embryo/seed³². However, to date,

the various deficiencies underlying desiccation sensitivity of recalcitrant seeds are generally conjectural as they are variably developed or expressed in the non-orthodox condition. A number of studies focused on germination, storage, desiccation sensitivity, and viability after storage of *Q. ilex* seeds have been conducted over the last few decades^{15; 33-42}. However the mechanisms involved in the recalcitrance of *Q. ilex* seeds remain elusive.

In orthodox seeds, the acquisition of desiccation tolerance, the last period previous to seed maturity, is associated with multiple cellular processes, including synthesis of storage compounds and heat-shock proteins, osmoprotective proteins and carbohydrates (e.g. sucrose, raffinose, galactinol and trehalose), and activation of antioxidative defenses (Angelovici *et al.* and Caccere *et al.*⁴³⁻⁴⁴ and refs. therein). Maturing seeds follow a metabolic switch that may indicate a mitochondrial transition between respiratory and anaerobe activity. The tricarboxylic acid cycle components and the respiratory rate decrease, whereas a strong accumulation of fumarate and succinate, linked to glyoxysomal fatty acid degradation by beta-oxidation has been described⁴⁵. Genes coding for glycolysis, tricarboxylic acid cycle, cell wall metabolism, DNA synthesis and transport of amino acids and nucleotides are downregulated during seed desiccation. But many other genes related to the translation machinery, the proteasome machinery, the energy metabolism or the DNA repair, among others, are either upregulated during seed desiccation and/or highly abundant in the stored mRNA population of dry seeds. After all these changes, the orthodox seeds can tolerate desiccation and are storable in a dry state for a long period of time (after-ripening). During after-ripening, the seed maintains low level of metabolic activity and maintain the viability for long periods at the dormant stage⁴³.

When dormancy is released, seeds can germinate under favorable conditions, specific to each species. The germination process, that begins with seed imbibition and finishes, proper sense, with the protrusion of the radicle, is divided into three distinct phases of water uptake (Fig. 1.2). Phase I starts with a fast water uptake and the activation of respiratory metabolism and transcriptional and translational activities. During this phase, known as 'physical' imbibition, a step-by-step activation of metabolic pathways results from the gradual increase in hydration, indicated with arrows in Fig. 1.2. When the level of hydration exceeds 60%, the rate of water acquisition slows (phase II) and new physiological mechanisms prepare cell expansion in the embryonic axes, culminating in the start of cell elongation. Osmotically active substances (solutes, such as sugars, amino acids, and potassium ions) are accumulated and acidification of the cell wall leads to a loosening of the bonds between cell-wall polymers. Later, in the phase III, water uptake resumes and testa rupture allows radicle protrusion. Here the *sensu stricto*, ends and starts the post-germination phase with high water uptake, mobilization of the major part of

reserves and first cell divisions, until the complete seedling development⁴⁶⁻⁴⁸. Seeds imbibition causes changes in the level of different metabolites. Storage nutrients (lipids, proteins or starch) accumulated in the embryo's cotyledon and/or endosperm start to be mobilized before completion of germination and are used in the post-germination steps to sustain the young plant in its early growth stages, before it becomes autotrophic. If the cell cycle resumes during germination, the first cell division (mitosis) occurs in the post-germinative phase⁴⁹.

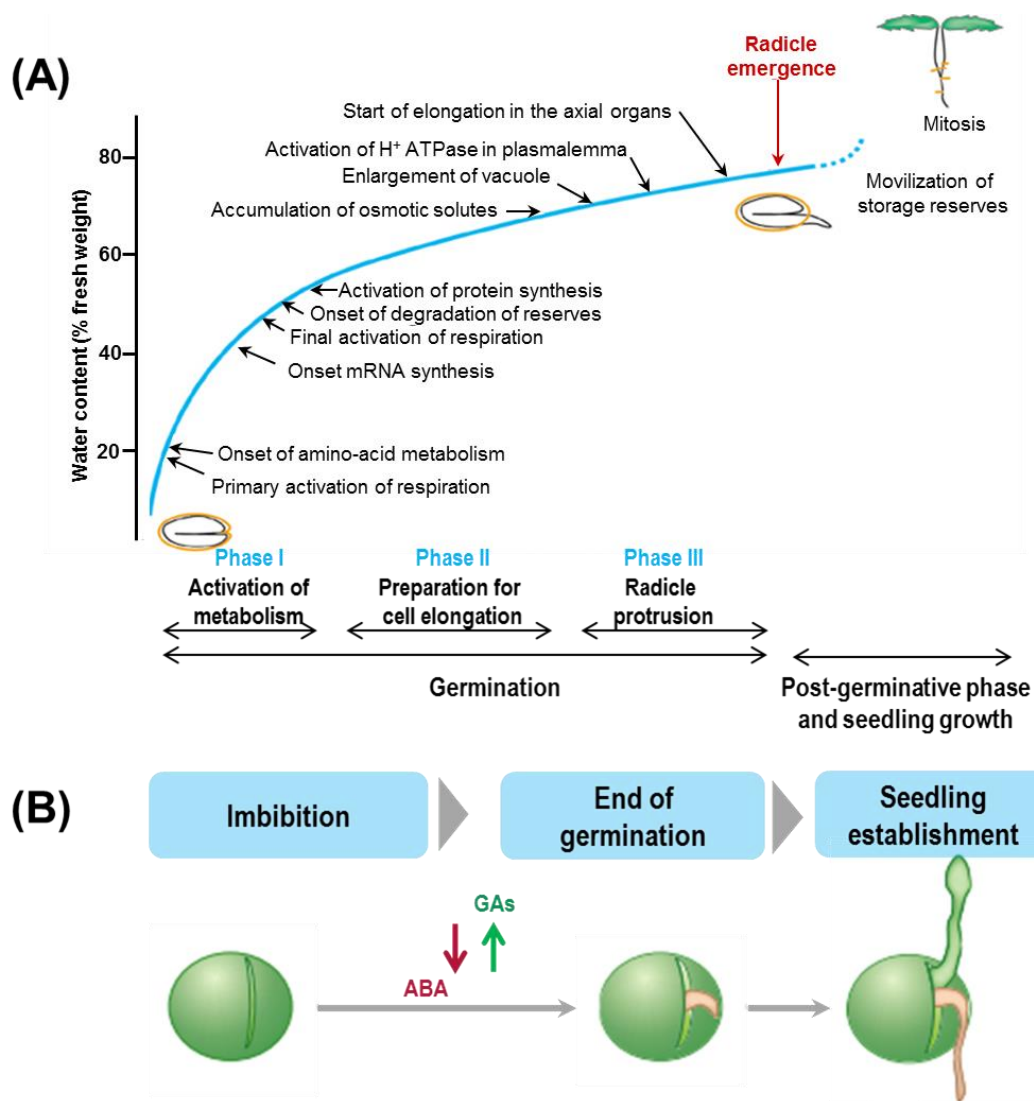


Figure 1.2: Phases of seed germination. (A). A rapid imbibition phase (phase I) launches the resumption of basic metabolism. When the level of hydration exceeds 60%, the rate of hydration slows (phase II) and new physiological mechanisms prepare cell expansion in the embryonic axes, culminating in the start of cell elongation. These events coincide with the activation of the H⁺ ATPase in the plasmalemma, which results in a further increase in water uptake (phase III) that may coincide with weakening of the surrounding tissues as the embryonic axes elongate and germination is completed. The arrows indicate the particular hydration levels that are known to correlate with individual metabolic events. **(B).** Morphology of seed germination and balance of ABA and GA necessary for germination. Adapted from Bove *et al.*⁴⁹ and Rajjou *et al.*⁵⁰.

The phytohormone abscisic acid (ABA), a sesquiterpene compound resulting from the cleavage of carotenoids, controls storage reserve accumulation and desiccation tolerance of

orthodox seeds. During seed development, ABA inhibits precocious germination and induces primary dormancy⁵¹. ABA regulates the expression of various sets of stress-responsive genes including those involved in the accumulation of compatible osmolytes and in the synthesis of LEA proteins, dehydrins and other stress-induced proteins. These protective proteins help in maintaining cellular water status and protect other proteins/enzymes and cellular organelles from collapsing under water stress (Sreenivasulu *et al.*⁵² and refs. therein). At its basal level, ABA also regulates various physiological processes of plant development such as seed development, embryo morphogenesis and maturation, dormancy and synthesis of storage proteins and lipids. A rise in the ABA level is observed in the middle stage of seed development (early maturation). The majority of ABA accumulated during this stage is originated and transported from the mother plant (Kanno *et al.*⁵³ and refs. therein) and triggers accumulation of many seed storage compounds. This “inherited” ABA is soon inactivated (catabolized through 8'-hydroxylation) and new ABA is synthesized in the zygotic tissues, giving rise to a second small peak of ABA accumulation in the late developmental stage of seed maturation, leading to acquisition of desiccation tolerance and seed dormancy (Kanno *et al.*⁵³, Nambara *et al.*⁵¹ and refs. therein).

Recent advances in *Arabidopsis* molecular genetics have revealed the core ABA signaling pathway. Central to ABA signaling in seeds are three core components: Pyrabactin Resistance/Pyrabactin-Like/Regulatory components of ABA receptors (PYR/PYL/RCAR), protein phosphatase 2Cs (PP2Cs) and sucrose non-fermenting 1 (SNF1)-related protein kinase 2s (SnRK2s)⁵¹ (Fig. 1.3). The binding of ABA to its receptor PYR/PYL/RCAR forms a complex, able to interact with and to inhibit the activity of PP2Cs, which in turn negatively regulates ABA signaling through repression of SnRK2s, the positive regulators of downstream targets. Inhibition of PP2Cs leads to de-repression of SnRK2s, which phosphorylate and activate downstream transcription factors including the basic leucine zipper (bZIP)-type transcription factors ABRE (ABA-responsive element)-binding factors (ABFs) and ABSCISIC ACID INSENSITIVE5 (ABI5), the APETALLA 2 (AP2)-type transcription factor ABI4, and the B3-type protein ABI3. All these transcription factors are essential to regulate the expression of ABA responsive genes in seeds. In the absence of ABA, PP2Cs dephosphorylates and deactivates SnRK2s and the complete pathway is switched off. The ABA signaling pathway constituted by these molecular components appears to be conserved in the seeds of both dicot and monocot species⁵⁴.

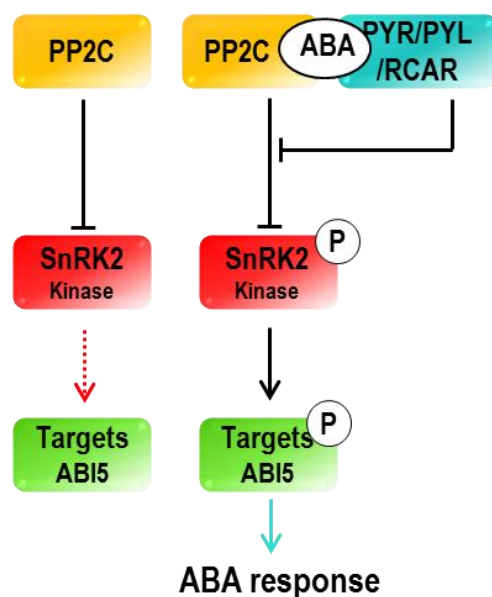


Figure 1.3. Model for ABA signaling pathway in plants. Left, ABA absence: SnRK2 kinase is inactivated by protein phosphatase 2C (PP2C), and thus the downstream targets, including ABI5, are inactive. Right, ABA-binding alters the conformation of PYR/PYL/RCAR and inhibits the activity of PP2C. De-repressed SnRK2 kinase activates the downstream targets. PYR/PYL/RCAR, pyrabactin resistance/PYR-like/regulatory components of ABA receptors; PP2C, protein phosphatase 2C; SnRK2, SNF1-related protein kinase2; ABI5, ABA insensitive 5.

Plant hormones function mostly in combination. Reduction in dormancy and increase in germination potential is accomplished by a decrease in sensitivity to germination-inhibiting signals (ABA) and a concomitant increase in the sensitivity of seeds to germination-stimulating signals, such as gibberellins (GA) (Fig 1.2B) ⁵¹. GAs negatively regulate those proteins that act as repressors of germination, and nowadays is accepted that the ABA/GA ratio is what regulates the metabolic transition required for germination ⁵⁰

Other hormones such as ethylene ⁵⁵, brassinosteroids ⁵⁶, salicylic acid ⁵⁷, cytokinins, auxins ^{56;58}, jasmonic acid ⁵⁹, and oxylipins ⁶⁰ also influence germination. Germination related phytohormones form an interlocked signaling network where all interact with one another to finely control germination, particularly in response to environmental constraints. Though the interplay of ABA with other endogenous signals is less documented, it has been reported that ethylene collaborates with GA to counteract the germination-inhibiting effect of ABA, and that ABA inhibition by cytokinin is required before GA can be effective in lettuce. In addition, the content of jasmonic acid decreases while that of indole acetic acid increase have been reported as necessary during the early germination phase ^{42; 51; 59}.

In contrast to the orthodox seeds, recalcitrant seeds contain embryos that do not tolerate desiccation. The reason(s) for the susceptibility to dehydration-induced damage in recalcitrant seeds remains unclear. It has been proposed that desiccation sensitivity and the precocious germination were caused by a low level of ABA in the embryo which would be

deprived of the ABA-driven accumulation of protective metabolites. Actually, the germination of some recalcitrant seeds (*i.e.*, *Theobroma cacao* L. and *Avicennia marina*) has been reported as similar to that of ABA anabolic mutations, which block ABA synthesis in the embryo axis, and of ABA-insensitive mutants displaying a lower sensitivity to the inhibiting effects of ABA ⁶¹. Moreover, it has been reported that the ABA content of *Q. robur* embryos decreased significantly (80%) at the end of maturation stage, what has been linked to the high precocious germination response of this plant ⁶². However, the levels of ABA were high (peaked at 44 ng g⁻¹ DW) by the time of the medium-maturation phase ⁶² and hence the recalcitrance of *Q. robur* or other species like *Araucaria angustifolia* or *Castanea sativa* do not seem due to an inability to accumulate storage or dehydrin proteins (ABA-regulated) or soluble sugars (ABA-regulated) during the maturation process ^{36;63}. Some other reports indicate that recalcitrant embryos lack the ability for the metabolic switch-off that occur during the maturation of orthodox seeds, and this has also been proposed as one of the possible basic reasons for recalcitrant seeds being desiccation sensitive and lacking of quiescence period ³².

The metabolism during the seed development of *Inga vera*, an important leguminous species whose seeds are among the most recalcitrant ones described up to date, where increased amounts of citric, glutamic, pyroglutamic, and aspartic acids are detected, corroborate the hypothesis of high metabolism in maturing seed, with a shift from fermentative to aerobic respiration at seed maturity ⁴⁴. The elevated rate of aerobic respiration would be cause of a high level of reactive oxygen species (ROS) in mature recalcitrant seed. Accordingly, in the recalcitrant seeds of *Q. robur*, *Acer saccharinum*, *Avicennia marina*, *Shorea robusta* etc., has been reported an excessive production of ROS in concurrence with a lowering in antioxidant enzyme protection, which may lead to cellular damage and desiccation sensitivity ⁶⁴ and ref. herein. But in contrast, in *A. angustifolia* seed an increase in proteins related to energy metabolism has been reported during germination process ⁶³, what suggests that the *A. angustifolia* non-germinating seed metabolism is not completely functional.

Studies of the mechanisms of carbohydrate regulation in plants have shown that the sugars themselves are often the signal molecules. A crosstalk between sugars and hormones has been reported as a regulation process of seed germinations and seedling growth ⁶⁵⁻⁶⁶. In addition, among the sugars, raffinose-family oligosaccharides (RFO) have been traditionally associated with desiccation tolerance and seed longevity ⁶⁷. RFO are suggested to protect cellular integrity during desiccation by stabilizing membranes during dehydration and provide substrates for energy generation during germination, and are also part of the scavenging machinery of hydroxyl radicals ⁴³.

1.3. Functional genomics. “OMICS” approaches in plant biology research

Large-scale genome projects have changed the biological experimentation paradigm study, as was exposed previously. The sequencing of the first plant genome by the end of 2000, that of *Arabidopsis thaliana* ²⁴, ushered the research in plant biology into a new era. Genomics can be defined as the structural and functional analysis of the complete genome of an organism, including mapping, sequencing and gene expression profiling; it has often been referred to as a new field that has led to a paradigm shift in the way science is performed.

During the late 1980s, structural genomics stated the generation and analysis of information about genes and genomes, thus long as that information could be produced systematically, by DNA sequencing projects. Currently the mass of genome data is being converted into gene-function data. The genome of *Populus trichocarpa*, the first tree species completely sequenced, is roughly 1/6 of the size of the human genome (≈ 3300 Mb) but this is an exception. The genome of most pines is more than 8-fold the human genome size ⁶⁸ and contain about 30,000 genes as estimated from known plant genomes ²¹. In the *Fagaceae* family considering *Quercus*, *Castanea* and *Fagus*, the species of these three genera have the same number of chromosomes ($2n = 24$) ⁶⁹, although their genome sizes vary almost two fold from 544 Mb/1C in *Fagus sylvatica* to 941 Mb/1C and 980 Mb/1C in *Castanea sativa* and *Q. ilex*, respectively ⁷⁰. Therefore, the size of genomes is one of the limitations in tree genomes sequencing, however, progress in technology has allowed the sequencing of many species that can be found in <http://www.plantgdb.org/>. During the middle 1990s, a new term, “functional genomics”, came to refer to generation and analysis of the information about what genes do. Since then, a number of other terms have been introduced: proteomics, transcriptomics, metabolomics. In fact, it was suggested that all such work be termed “omics” ⁷¹ (Fig. 1.4).

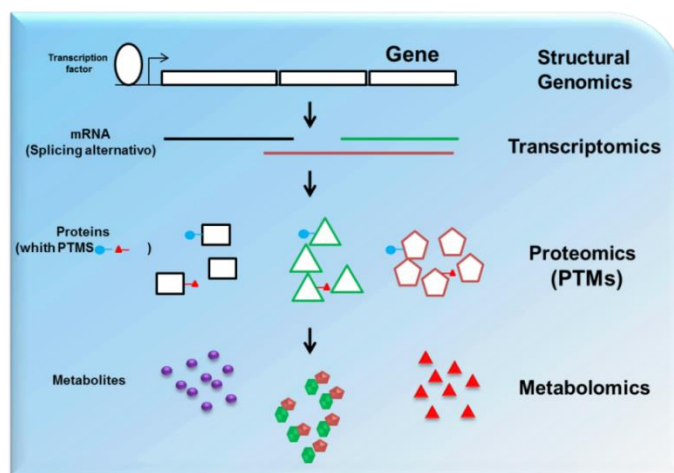


Figure 1.4: The “omics” cascade. “Omics” data sets that describe virtually all biomolecules in the cell are starting to become available. From the top, DNA (studied by genomics) is first transcribed to mRNA (transcriptomics) and translated into protein (proteomics), which can catalyze reactions that act on and give rise to metabolites (metabolomics), glycoproteins and oligosaccharides, and various lipids. Adapted from Joyce and Palsson ⁷².

The large-scale experiments, which include structural genomics (DNA), transcriptomics (RNA), proteomics (protein), and metabolomics (metabolites), together constitute the methodological bases of the Modern Systems Biology ⁷³, and have been recently incorporated to forest tree research ⁷⁴⁻⁷⁵. The most conclusive information about changes in gene expression levels can be gained from analysis of the varying qualitative and quantitative changes of messenger RNAs, proteins and metabolites. Classical biochemical assays, has been complemented with these new high throughput techniques. “Omics” researches should be viewed as synergistic with the more traditional studies of single molecules ⁷⁶⁻⁷⁷.

Plant functional genomics has benefited largely from genomic research performed on model organisms such as *Arabidopsis thaliana* ²⁴, the premier model system in plant science, *Oryza sativa* ²⁵, and *Populus trichocarpa* ²⁶. Model systems have provided excellent bases for the understanding of a wide range of biological processes. For *Arabidopsis*, large sets of genetic resources and analytical tools are now available to assist studies on the functions of plant genes, available in The Arabidopsis Information Resource (TAIR) (<http://www.arabidopsis.org/>). However, transferring information from *Arabidopsis* to other plant species still remains a considerable challenge ⁷⁸.

The following sections briefly introduce “omics” technologies that are used by researchers to investigate a biological process.

1.3.1. Transcriptomics approaches

The term ‘transcriptome’ encompasses all the genomic counterparts which are expressed as RNA transcripts, including coding (mRNA) and non-coding (e.g. tRNA or miRNA) RNAs at a given time in a cell or population of cells under a given set of environmental conditions ⁷⁶. A transcriptome is a snapshot of the gene expression provided by capturing the total RNA, which proporcionate a view into the gene action within a cell or tissue at a particular moment in time, as it permit to quantify transcripts. This quantification revelates the expression of active genes but also the combination of all isoform sequences (produced through alternative splicing and variant alleles) within the cells. ⁷⁹. The field of transcriptomics provides information about both the presence and the relative abundance of RNA transcripts, thereby indicating the active components within the cell ⁷². Transcriptomics has been used to study both the whole plant as well as comparative biology. This helps to comprehend the key genes responsible for general development or response to an external stimuli ⁸⁰.

Among the transcripts quantification techniques, northern blotting and quantitative RT-PCR (reverse transcription –RT, semi-quantitative polymerase chain reaction -PCR and qRT-PCR) ⁸¹ can be used for transcript analysis when a few genes are the interest. Over the

years, tremendous advancement has been made in techniques for transcriptome analysis (Fig. 1.5). Many high-throughput analytical methods for gene expression profiling were developed, including microarray technology⁸², differential display⁸³, serial analysis of gene expression (SAGE)⁸⁴, massively parallel signature sequencing (MPSS)⁸⁵, suppression subtractive hybridization (SSH)⁸⁶, expressed sequence tag (EST) sequencing⁸⁷ and most lately, the ultra-high-throughput sequencing (UHTS) technologies, often referred to as next-generation sequencing (NGS)⁸⁸, that has revolutionized the way and pace in which the whole transcriptome is sequenced, which is called RNAseq.

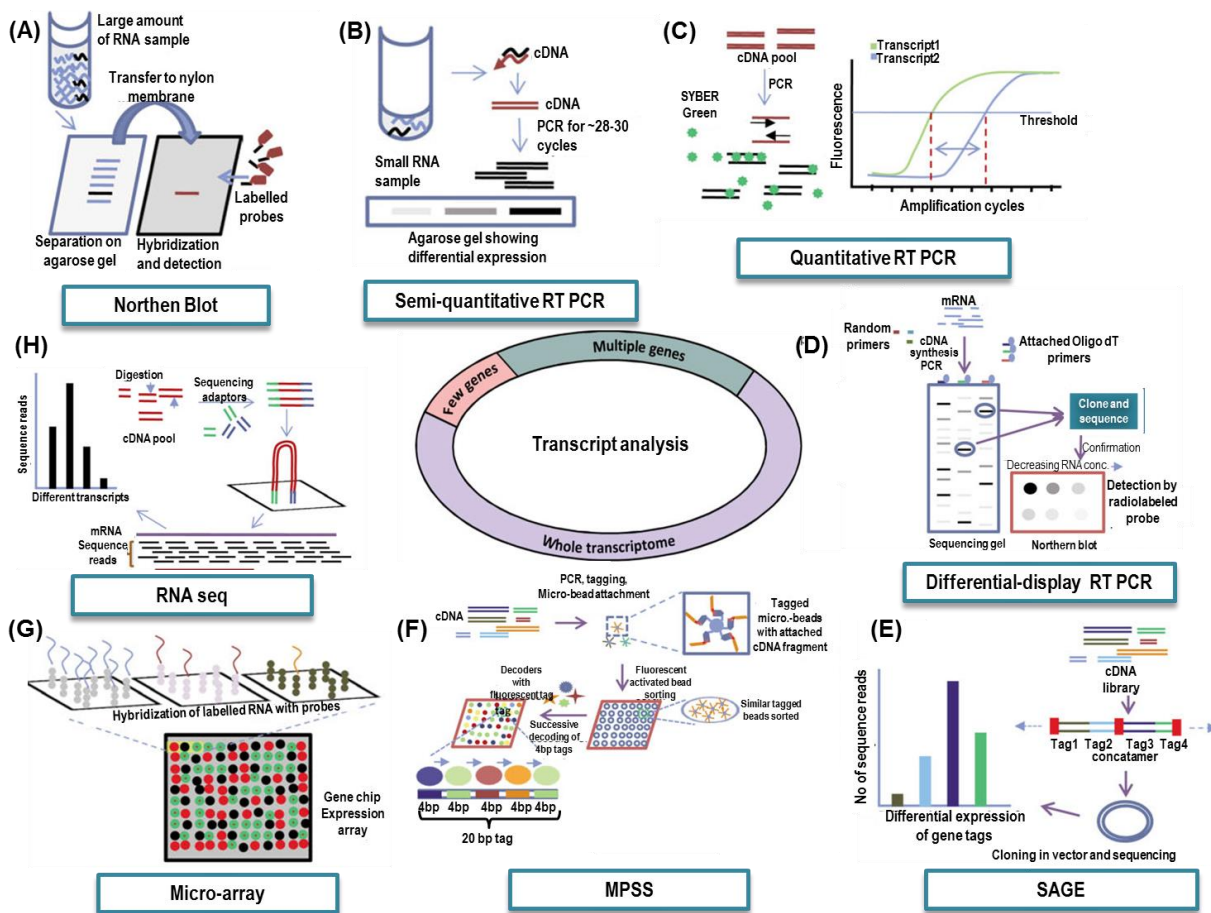


Figure 1.5. Progress in methods for expression analysis of genes from a few genes to multiple ones and whole transcriptome analysis. (A). Northern blotting is used for the detection of a single/few genes by hybridization of labeled probes to localized RNA samples run on an agarose gel and transferred to a membrane. (B). Semi-quantitative PCR is a visual estimation of the difference in expression level of one or few genes, by PCR, in the cDNA samples reverse transcribed from RNA. (C). Quantitative PCR is a real time estimation of the levels of transcript of one/few genes, often calculated as fold changes with respect to a control gene. (D). Differential display RT-PCR has been used to determine the differentially expressed genes between a control and test samples. (E, F). Serial analysis of gene expression and massively parallel signature sequencing determine the transcript number of known and novel genes in the tissues tested in terms of short sequences called as tags. (G). In microarray, RNA hybridized to probes spotted on a chip gives a signal value, which can be used to detect the differentially expressed as well as specific known genes. (H). RNA sequencing is transcript sequencing by NGS and shows the transcript number of known as well as novel genes along with alternative splicing, Adapted from Agarwal *et al.*⁸⁰.

The suppression subtractive hybridization (SSH) methods for the construction of cDNA libraries were developed for the detection of differentially expressed genes in two

distinct situations ⁸⁶. SSH permits a massive identification of genes differentially expressed at the mRNA level, resulting in a gene candidate list that, as it happens in the case of microarrays, have to be further validated by more accurate technique, such as real-time RT-PCR.

The methodology behind this technique is based on DNA hybridization. In contrast to other platforms for gene expression studies, SSH methods are non-target approaches, characterized by the independence of the previous knowledge of the gene sequences, becoming a key technology in the study of orphan species, such as *Q. ilex*. Furthermore, no specialized equipment is needed, since it makes use of traditional molecular biology methods. A critical issue in the application of SSH techniques is the difficulties to detect low abundance transcripts. To solve this problem, a protocol known as PCR-select cDNA subtraction is being nowadays routinely used in laboratories. It is based on the selective amplification of the differentially expressed sequences ⁸⁶. This technique has been used to study a wide variety of processes in many plant species such as: identification of genes involved in fruit ripening in oil palm (*Elaeis guineensis* Jacq) ⁸⁹, heat stress in leaves of *Pinellia ternate* ⁹⁰, pathogens infection ⁹¹, seed development in *Phaseolus vulgaris* ⁹² oil metabolism in maturing seeds of *Jatropha curcas* ⁹³ and seed germination ⁹⁴. Also some SSH studies have centered in forest tree species such as pine ⁹⁵ and oak ^{28;96}. As discussed in chapter 5 of this Thesis, two SSH libraries were constructed to investigate genes related to *Q. ilex* seeds germination and seedling growth.

Although it is not a “omics” methodology, real-time qRT-PCR is one of the best established technologies of the genomic era for detecting and quantifying mRNA in biological samples ⁹⁷. The adoption of this technology in individual research laboratories has resulted in its extensive application to functional genomics studies. Real-time RT-PCR has the particularity of the use of fluorochromes that allow monitoring the amplification of the different reactions in the same experiment and in each RT-PCR cycle. Over the past 10 years, the popularity of this method has grown exponentially, with the publication of well over 25,000 papers from diverse fields of science, including agricultural, environmental, industrial, and medical research, making reference to qRT-PCR data ⁹⁸. Several factors have contributed to the transformation of this technology into a mainstream research tool. In addition to its high sensitivity, its large dynamic range (>10⁷-fold) allows straight forward comparison between RNAs that differ widely in their abundance. It does not require post-PCR processing, in contrast with most transcriptomic techniques, and it has the potential for high throughput. Many of the key proteins, such as the transcription factors, are found in low abundance that real-time RT-PCR quantification constitutes the only technique with enough sensitivity to measure reliably their expression “*in vivo*” ^{97;99}.

The qRT-PCR can be used to provide both relative and absolute (number of molecules) quantification of specific mRNA molecules. The relative qRT-PCR method estimates fold change of expression difference between mRNAs from target and reference genes relative to a control condition through $2^{-\Delta\Delta C_t}$ calculation¹⁰⁰. Relative quantification using RT-qPCR has been used in diverse plant species¹⁰¹⁻¹⁰³. However, the problem with this relative quantification is its foundation on the expression of housekeeping genes, *i.e.*, genes that are expected to have a constant expression under the experimental conditions, for this reason a previous analysis to validate housekeeping gene in the experimental conditions are necessary¹⁰⁴⁻¹⁰⁵.

Absolute quantification of transcription allows the precise determination of copy number per cell, total RNA concentration, or unit mass of tissue. It requires the construction of an absolute standard curve, constructed by amplifying known amounts of the target under conditions identical to those of the sample^{97; 106}. This method has the advantage of giving information on the real biological relevance of a transcript abundance change, since a small relative variation in an abundant transcript may have more importance than a large relative change in rare or infrequent transcripts¹⁰⁷. Absolute quantification by real-time qRT-PCR has not been extensively applied in plants¹⁰⁸⁻¹⁰⁹. All the characteristics of qRT-PCR make it a very suitable technique for the validation of the huge amount of results obtained with other transcriptomic platforms, such as microarray or SSH. In this work, the absolute quantification of mRNAs from specific groups of genes was used to analyze genes implicated in germination (Ch. 4) and to verificate the results obtained by SSH (Ch. 5).

1.3.2. Proteomics approaches

The term “proteome” was coined by M. Wilkins and first used during the Siena meeting in 1994 (2-D Electrophoresis-From Protein Maps to Genomes, Siena, Italy, September 5-7, 1994) considered the year of birth for “proteomics”. In this meeting Prof. Wilkins presented, for the first time, to the audience the term “proteome”, in an attempt to circumvent the phrase “all proteins expressed by a genome, cell or tissue” while communicating his work¹¹⁰. The complexity of the proteome is much greater than that of the transcriptome due to the huge amount of possible post-translational modifications (PTMs) (phosphorylation, glycosylation, nitrosilation, etc), making the former highly dynamic. The main goal of proteomics is to study, know and understand “how”, “where”, “when”, and “what for” are the several hundred thousand of individual protein species/forms¹¹¹ produced in a living organism.

Different proteomics sub-areas can be defined according to the specific objectives:

i) Descriptive Proteomics, including subcellular proteomics (cataloguing as many as possible proteins species or forms from the experimental system, and hence answering to the where and when questions).

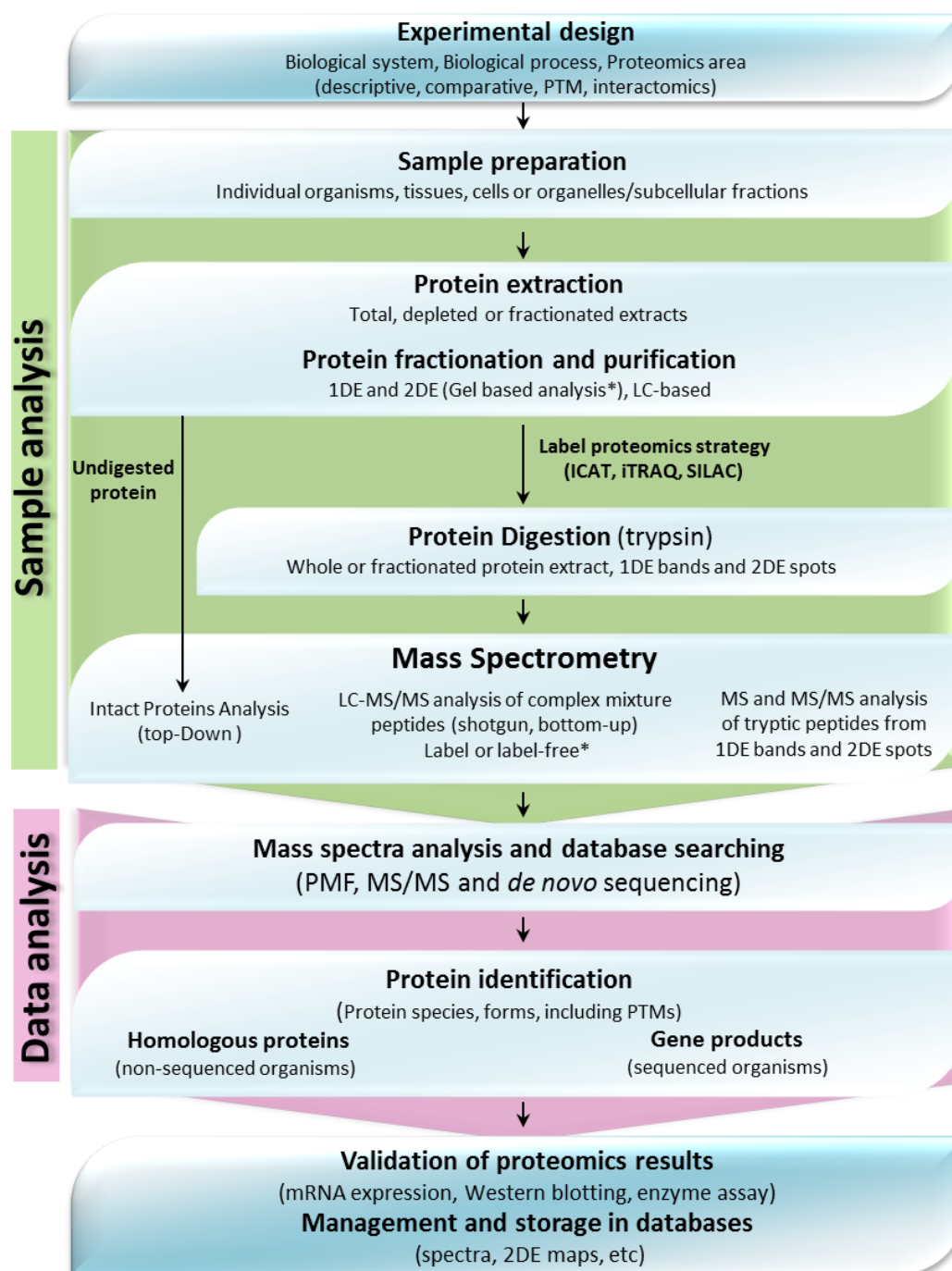
ii) Comparative Proteomics (as biological knowledge mainly comes from comparative studies, it is pretended to establish differences in the protein profile among genotypes, including transgenic and mutants, type of cells, tissues, and organs, developmental stages and external factors, and hence answering to the “what for”).

iii) Posttranslational Modifications (modifications to the first translation product and protein degradation, type of modification, modified residue and biological consequences, answering to how).

The workflow of a standard mass spectrometry (MS)-based proteomics experiment includes all or most of the following steps: experimental design, sampling, tissue/cell or organelle preparation, protein extraction/fractionation/purification, labelling/modification, separation, MS analysis, protein identification, statistical analysis of data, validation of identification, protein inference, quantification, and data analysis, management and storage. (Fig. 1.6).

Two MS-based proteomics approaches exist, namely, bottom-up and top-down proteomics, each with its own advantages and disadvantages for the identification of proteins and protein interacting partners, protein quantification, and analysis of PTMs. Peptide centric proteomics (bottom-up) are based on protein digestion, peptide identification, and later protein inference. In contrast, top-down proteomics, an emerging technology that studies proteins preserving the entire protein structure including PTMs, instead of measuring peptides produced from them by proteolysis ¹¹². This approach was initially limited to small proteins, however the use of new analytical procedures for protein ionization and dissociation has extended its use for proteins up to 200 kDa ¹¹³. The bottom-up has been used for first time in plants for characterizing the chloroplast proteins ¹¹⁴ or domain-domain interaction ¹¹⁵. However the contribution of this technique in plant proteomics is very scarce.

The different strategies for a proteomics experiment are the result of different protocol combinations in a specific sequence, and can be named and grouped into different categories. The following make reference to the evolution of the workflows and instrumentation in the most commonly employed bottom-up (peptide centric) approach.



* Statistical methods in gel based or gel free (label or label free) are required to select differentially expressed proteins.

Figure 1.6: Workflow of a proteomics experiment, from experimental design to data analysis and management. It includes alternative, complementary approaches or strategies, based on MS analysis of proteins (top-down) or tryptic peptides (bottom-up), either gel-based or gel-free and label-based (isotopic, isobaric) or label-free. PTM: posttranslational modifications; LC: liquid chromatography; iCAT: isotope-coded affinity tags; iTRAQ: isobaric tag for relative and absolute quantitation; SILAC: stable isotope labeling by amino acids in cell culture; MS: mass spectrometry; PMF: peptide mass fingerprinting. Adapted from Jorrin-Novo *et al*¹¹⁶.

As for other biological systems, *i.e.*, bacteria, fungi or animals, the dawn of plant proteomics can be defined by the use of 2-DE electrophoresis and N-terminal sequencing for identifying protein spots. Two dimensional electrophoresis reproducibility was moderate and

the rates of protein identification were low, mostly because of the absence of plant sequences in public databases. Mid-late 2000's brought the explosion of plant proteomics. During that dates the easiness and availability of high-throughput two dimensional electrophoresis systems (*i.e.*, DodecaCell from Biorad, or Ettan systems from Amersham) exponentially grew and so the number of plant species that were studied. Bigger gels and narrower isoelectric point (pI) ranges allowed the resolution of thousands of spots per gel and the introduction of fluorescent labels and stain incremented the sensibility of this technique also given the opportunity for gel multiplexing.

The development of sharper imaging systems and improved versions of proteomic software allowed faster and more accurate analyses ¹¹⁷. At the same time N-terminal identification of proteins was replaced by MALDI-TOF based peptide mass fingerprinting (PMF), despite this method was really useful only in fully sequenced species. New identification technologies, based on MALDI-TOF/TOF or LC-MS/MS, allowed an easy pipeline for *de novo* assembling. The algorithms for protein identification became tolerant to substitutions, making easier the cross-species protein identification ¹¹⁸⁻¹¹⁹. The appearance of these tools together with the release of new plant genomes and the improvement of available sequence data repositories ¹²⁰ increased the number of identified spots in a wider set of plant species.

The late 2000's were also defined for the introduction of monodimensional gel free LC-MS systems on plant proteomics, which were initially based on the isotopic labeling of peptides. In general, labeling technologies have not been popular in plant proteomics despite it can be considered as second generation of proteomics. Differential Gel Electrophoresis, DiGE, has been reported for first time in 1997 ¹²¹. However only a small number of plant proteomics work used this technology since 2006 (for a review see Jorrín et al. 2009). This methodology has two major disadvantages, in plant systems: it is difficult to get a reproducible labeling and it is pricey ¹²². Some methodologies that use labeling technologies are isotope-coded affinity tags (ICAT) ¹²³, stable isotope labeling by amino acids in cell culture (SILAC) ¹²⁴, isobaric tag for relative and absolute quantitation (ITRAQ) ¹²⁵, these approaches were quickly outdated by label free MS-based proteomics.

The development of gel-free label-free methodologies has provided a new environment for proteome research. The initial configurations for these studies were based on a nanoflow chromatography systems coupled to triple quadrupoles or ion traps, however the sensitivity, resolution, and speed limited the characterization and quantification of full proteomes. The introduction of a next generation mass spectrometers in proteomics, mainly the Orbitrap ¹²⁶⁻¹²⁷, allowed an unprecedented capability for a deep characterization of the peptides present in a sample with sub ppm accuracy ¹²⁸. Nowadays this family of instruments

can be considered the gold standard in proteomics and metabolomics as they provide higher resolution, speed, and dynamic range compared to classic quadrupole based instruments. The first applications of Orbitrap in proteomics dated from late 2005-2006¹²⁹. In early 2008 this instrumentation was first applied in plants, specifically for the characterization of the chloroplast proteome¹³⁰. The increased efficiency of chromatographs, search algorithms such as Mascot, SEQUEST and the improvement of genome annotations allowed the characterization and quantification of near complete proteomes, in *Saccharomyces* for example¹³¹. Current contributions in plant proteomics can describe thousands of proteins in model species. Close results have been reported in non-model species such as holm oak or pine when using custom databases that gathered all available information in public databases¹³².

One of the major drawbacks of the use of label-free mass spectrometry is the difficulty of obtaining precise quantitation, despite the recent advances in algorithms for processing peak intensities and area integrations. Though current precision can be considered good enough for relative abundance comparison in full proteomes, alternative procedures should be used for very accurate quantitation of targeted proteins¹³³.

An important area in proteomics is PTMs, the final turning for the proteins. PTMs are gaining increasing recognition as regulators of cellular processes, as they are known to affect properties such as protein activity, stability and localization. They are also involved in signaling, where not only signal amplification (*i.e.*, signal cascades) is important, but also the rapid removal of the signal as in the case for reversible phosphorylation and the kinase/phosphatase networks. To understand how PTMs affect biological processes, it is necessary to know which PTM sites change in a protein and how they change¹³⁴. PTMs, with the exception of partial proteolysis, involve the addition of different molecules or chemical groups as show in Fig. 1.7. Although hundreds of different PTMs have been reported, including phosphorylation, glycosylation, prenylation, redox proteome, ubiquitination¹³⁵, few of them have been studied. Recent and ongoing development of methodologies has made possible large-scale analysis of PTMs¹³⁶. Protein phosphorylation was one of the first to be studied as a major field within proteomic analysis because it is the common mechanism for altering the cellular activity of the protein. In plant science, phosphorylation is still the most prominently studied PTM (>75% of the available literature) followed by glycosylation and ubiquitination¹³⁷.

Phosphorylation of proteins is a key regulator of intra-cellular biological processes and is a reversible modification affecting both the folding (conformation) and function of proteins and regulating enzyme activities, substrate specificities and protein localization, complex formation and degradation¹³⁸. The relevance of phosphorylation is highlighted by the number of genes involved in phosphorylation processes. Regulation of this modification is

tightly controlled by two enzymes, protein kinases (attaching of phosphate groups to amino acid residues) and phosphatases (removing of phosphate groups) ¹³⁷.

The variety of functions in which phosphorylated proteins are involved depends on the number (multiple residues within a protein can be phosphorylated by different kinases) and the sites of phosphorylation. Multiple sites of phosphorylation allow a protein to adapt several different functions, depending on which phosphorylation site is modified. This is highlighted by phosphorylation at one particular amino acid residue which may lead to conformational change that in turn allows for the phosphorylation of different amino acid residues within the same protein, or on the other hand, prevent the phosphorylation of nearby amino acids through steric hindrance ¹³⁹. The four most common types of phosphorylation occurring on amino acids are: O-phosphates, N-phosphates, S-phosphates and acyl-phosphates. O-phosphates are the most common class and they are mostly commonly attached to serine (Ser)-, threonine (Thr)- and tyrosine (Tyr)-residues (Fig. 1.7). The occurrence of phosphorylation on Ser- and Thr-residues is more frequent than on Tyr-residues. N-, S- and acyl phosphorylations are far less common ¹³⁷.

There are several methods for detecting phosphoproteins. One of the most sensitive methods is the radiolabelling of the phosphate groups using ³²P or ³³P (either *in vivo* or *in vitro*) followed by radio-immunoblotting ¹⁴⁰⁻¹⁴¹. The major advantage of this radiolabelling technique is the ability to detect all different types of phosphorylations and the signals can be quantitated, but ³²P is toxic for many cells and will, over time, cause damage to the cell. In 1981, the first documented phospho-antibody was produced in rabbits immunized with benzoyl phosphonate conjugated to keyhole limpet hemocyanin (KLH) ¹⁴². This antibody broadly recognized proteins containing phosphotyrosine. After that, there has been a rapid development in production of the phospho-antibodies. Nowadays, a large amount of phospho specific antibodies targeted to different amino acids (Ser, Thr, Tyr) at distinct sites in proteins have been produced, and widely used in the basic and clinic research ¹⁴³.

Phosphoproteins can be separated according to their pI and molecular weight using 2-DE gel electrophoresis (2-DE PAGE) and lately visualized on the gel by using phosphospecific stains. Commercially available phospho-stains are less sensitive but their use more convenient than techniques such as radiolabelling. Some phosphospecific stains such the fluorescent Pro-Q DPS or Pro-Q Diamond ¹⁴⁴ bind directly to the phosphate moiety of phosphoproteins with high quantitative linearity. The stain is compatible with other staining methods (*i.e.*, SYPRO-Ruby for total protein staining) and subsequent MS analysis ¹³⁷. Protein phosphorylation events are detected by increases in amino-acid residue mass of +80 Da, which indicates the addition of

Low sensitivity is a frequent obstacle when analyzing phosphopeptides or phosphoproteins by MS. Sub-stoichiometric phosphorylation often occurs, reducing phosphoanalyte abundances compared to corresponding unphosphorylated forms. In addition, phosphopeptides may show inefficient ionization or may be lost preferentially during handling by adsorption to metal or plastics. Thus, a large repertoire of techniques has been developed to enrich phosphoanalytes and improve detection sensitivity, particularly for highly complex samples. Many of these make use of reactive chemistries for covalent coupling or affinity purification, the last one are the most used. Strategies for noncovalent enrichment of phosphopeptides and proteins have used affinity purification based on charge properties and antibody recognition. Widespread methods use immobilized metal affinity chromatography (IMAC), which adsorbs phosphopeptides to chelated metal ions (Fe^{3+} , Ga^{3+}) through metal-phosphate ion-pair interactions. Titanium dioxide (TiO_2) and zirconium dioxide (ZrO_2) are also used to adsorb phosphopeptides, through bidentate interactions. Ion exchange resins allow partial enrichment of phosphopeptides based on charge separation, and high selectivity has been reported with metal affinity resins in combination with ion exchange chromatography. In addition, highly specific antibodies to phosphotyrosine have enabled selective immunopurification of phosphotyrosine-containing phosphopeptides as well as phosphoproteins¹³⁵.

All affinity capture methods suffer from phosphopeptide losses resulting from poor binding or recovery, and any method may yield biased results owing to chemical selectivity. Recovery of peptides by any method is difficult to estimate because total numbers of phosphopeptides are usually unknown. However, analyses of simple mixtures suggest that 30–50% of peptide sequences are recovered by Fe^{3+} -IMAC. Therefore, comparative phosphoproteomics analyses by using these methods are not reproducible.

In recent years, large-scale studies on protein phosphorylation based on mass spectrometry have been conducted on different organisms. Some of them were carried out in plants¹⁴⁴⁻¹⁴⁹. As a result, a number of phosphorylation databases emerged, most of which focus on mammalian and prokaryotic systems. Phospho.ELM¹⁵⁰ contains verified eukaryotic phosphorylation sites, but most are from mammals. PHOSIDA contains large-scale phosphorylation data in *Homo sapiens*, *Bacillus subtilis* and *Escherichia coli*¹⁵¹. There is also phosphorylation databases concentrated on plants. PlantsP¹⁵² that contains phosphorylation data on a few different plants, and are focused on the annotation of plant protein kinases and protein phosphatases. PhosphAt¹⁵³ provides a database of phosphorylation sites collected from current literature solely for the model organism *Arabidopsis thaliana*. P3DB database provides a resource of protein phosphorylation sites from various plant sources and contains multiple

embedded search capacities for querying the database. Most of the datasets in the database comes from large-scale experiments (MS/MS), although several smaller datasets were also deposited ¹⁵⁴⁻¹⁵⁵.

1.3.3. Metabolomics approaches

The term “metabolome”, suggested by Oliver et al. (1998), includes the entire set of small molecule metabolites, produced by any organism. Metabolomics, is hence, the comprehensive analysis of all metabolites in an organism under a given set of conditions ⁷⁶. The metabolome represents the output that results from the cellular integration of the transcriptome and proteome, and therefore provides not only a list of metabolite components but also a functional readout of the cellular state ⁷². A brief introduction of metabolomics is presented here, though in this work, the metabolomics approach was not used. Notwithstanding, some specific metabolites were measured to verify the hypothesis arising from functional genomic studies.

An estimated 200,000 metabolites exist in plant, although only ~50,000 have been elucidated. Metabolic profiles provide a biochemical phenotypic assessment of the plants and hence are the most valuable in systems biology studies, so regarded as a cornerstone of systems biology ^{76; 156-159}.

The general workflow of a metabolomic analysis comprises of four main stages: preparation of the sample, data acquisition using analytical methods, data mining and compound identification plus quantification using the statistical and bioinformatics analyses. The final task is to draw meaningful biological interpretations from the analyzed data. Various analytical platforms may be used in metabolomics like, nuclear magnetic resonance (NMR), LC-NMR; MS approach such as: gas chromatography-MS (GC-MS), capillary electrophoresis-MS (CE-MS), liquid chromatography-MS (LC-MS), LC-electrochemistry-MS (LC-EC-MS), direct infusion MS (DIMS), fourier transform ion cyclotron MS (FTMS); infrared spectroscopy (IR), thin layer chromatography (TLC), high-performance liquid chromatography (HPLC) equipped with different kinds of detectors: UV or photodiode array (PDA), fluorescent, electrochemical, etc., Fourier transform infrared (FT-IR)- and Raman spectroscopies. Amongst these, NMR and MS are the chiefly applied ⁷⁶.

1.4. References

1. Neale DB, Kremer A (2011) Forest tree genomics: growing resources and applications. *Nature Reviews Genetics* 12: 111-122
2. FAO (2009) State of the World's Forests 2009. Electronic publishing policy and support branch, communication Division. FAO, Rome.
3. Brasier CM (2008) The biosecurity threat to the UK and global environment from international trade in plants. *Plant Pathology* 57: 792-808
4. Pulido FJ, Díaz M, Hidalgo de Trucios SJ (2001) Size structure and regeneration of Spanish holm oak *Quercus ilex* forests and dehesas: effects of agroforestry use on their long-term sustainability. *Forest Ecology and Management* 146: 1-13
5. Gallego FJ, de Algaba AP, Fernandez-Escobar R (1999) Etiology of oak decline in Spain. *European Journal of Forest Pathology* 29: 17-27
6. MAPA (2006) Forestación de tierras agrícolas. Ministerio de Agricultura y Pesca, Madrid, pp 1-373
7. Kremer A, Abbott A, Carlson J, Manos P, Plomion C, Sisco P, Staton M, Ueno S, Vendramin G (2012) Genomics of Fagaceae. *Tree Genetics & Genomes* 8: 583-610
8. Joffre R, Rambal S, Ratte JP (1999) The dehesa system of southern Spain and Portugal as a natural ecosystem mimic. *Agroforestry Systems* 45: 57-79
9. Cañellas I, Roig S, Poblaciones MJ, Gea-Izquierdo G, Olea L (2007) An approach to acorn production in Iberian dehesas. *Agroforestry Systems* 70: 3-9
10. Barra-Jiménez A, Blasco M, Ruiz-Galea M, Celestino C, Alegre J, Arrillaga I, Toribio M (2014) Cloning mature holm oak trees by somatic embryogenesis. *Trees - Structure and Function* 28: 657-667
11. Vicente Á, Alés R (2006) Long Term Persistence of Dehesas. Evidences from History. *Agroforestry Systems* 67: 19-28
12. Gracia M, Retana J, Picó FX (2001) Seedling bank dynamics in managed holm oak (*Quercus ilex*) forests. *Annals of Forest Science* 58: 843-852
13. Corcobado T, Cubera E, Moreno G, Solla A (2013) *Quercus ilex* forests are influenced by annual variations in water table, soil water deficit and fine root loss caused by *Phytophthora cinnamomi*. *Agricultural and Forest Meteorology* 169: 92-99
14. Plieninger T, Pulido FJ, Schaich H (2004) Effects of land-use and landscape structure on holm oak recruitment and regeneration at farm level in *Quercus ilex* L. dehesas. *Journal of Arid Environments* 57: 345-364
15. Pasquini S, Mizzau M, Petrusa E, Braidot E, Patui S, Gorian F, Lambardi M, Vianello A (2012) Seed storage in polyethylene bags of a recalcitrant species (*Quercus ilex*): analysis of some bio-energetic and oxidative parameters. *Acta Physiologiae Plantarum*: 1-12
16. Savill P, Kanowski P (1993) Tree improvement programs for European oaks: goals and strategies. *Annals of Forest Science* 50: 368s-383s
17. Vieitez A, Corredoira E, Martínez M, San-José M, Sánchez C, Valladares S, Vidal N, Ballester A (2012) Application of biotechnological tools to *Quercus* improvement. *European Journal of Forest Research* 131: 519-539
18. Lelu-Walter M-A, Thompson D, Harvengt L, Sanchez L, Toribio M, Pâques L (2013) Somatic embryogenesis in forestry with a focus on Europe: state-of-the-art, benefits, challenges and future direction. *Tree Genetics & Genomes* 9: 883-899
19. Bonga JM, Klimaszevska KK, von Aderkas P (2010) Recalcitrance in clonal propagation, in particular of conifers. *Plant Cell, Tissue and Organ Culture (PCTOC)* 100: 241-254
20. FAO (2004) Preliminary review of biotechnology in forestry, including genetic modification. forest genetic resources working paper FGR/59E.
21. White TL, Adams WT, Neale DB (2007) *Forest Genetics*. CABI Publishing, Wallingford, UK
22. Abril N, Gion J-M, Kerner R, Müller-Starck G, Cerrillo RMN, Plomion C, Renaut J, Valledor L, Jorrin-Novo JV (2011) Proteomics research on forest trees, the most recalcitrant and orphan plant species. *Phytochemistry* 72: 1219-1242
23. Wullschlegler SD, Tuskan GA, DiFazio SP (2002) Genomics and the tree physiologist. *Tree Physiology* 22: 1273-1276

24. TAG I (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796-815
25. Sequencing Project International Rice G (2005) The map-based sequence of the rice genome. *Nature* 436: 793-800
26. Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen G-L, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J, Degroeve S, Déjardin A, dePamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehlting J, Ellis B, Gendler K, Goodstein D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R, Huang W, Islam-Faridi N, Jones S, Jones-Rhoades M, Jorgensen R, Joshi C, Kangasjärvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A, Kalluri U, Larimer F, Leebens-Mack J, Leplé J-C, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson DR, Nelson C, Nieminen K, Nilsson O, Pereda V, Peter G, Philippe R, Pilate G, Poliakov A, Razumovskaya J, Richardson P, Rinaldi C, Ritland K, Rouzé P, Ryaboy D, Schmutz J, Schrader J, Segerman B, Shin H, Siddiqui A, Sterky F, Terry A, Tsai C-J, Uberbacher E, Unneberg P, Vahala J, Wall K, Wessler S, Yang G, Yin T, Douglas C, Marra M, Sandberg G, Van de Peer Y, Rokhsar D (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313: 1596-1604
27. Myburg AA, Grattapaglia D, Tuskan GA, Hellsten U, Hayes RD, Grimwood J, Jenkins J, Lindquist E, Tice H, Bauer D, Goodstein DM, Dubchak I, Poliakov A, Mizrahi E, Kullar ARK, Hussey SG, Pinard D, van der Merwe K, Singh P, van Jaarsveld I, Silva-Junior OB, Togawa RC, Pappas MR, Faria DA, Sansaloni CP, Petroli CD, Yang X, Ranjan P, Tschaplinski TJ, Ye C-Y, Li T, Sterck L, Vanneste K, Murat F, Soler M, Clemente HS, Saidi N, Cassan-Wang H, Dunand C, Hefer CA, Bornberg-Bauer E, Kersting AR, Vining K, Amarasinghe V, Ranik M, Naithani S, Elser J, Boyd AE, Liston A, Spatafora JW, Dharmwardhana P, Raja R, Sullivan C, Romanel E, Alves-Ferreira M, Kulheim C, Foley W, Carocha V, Paiva J, Kudrna D, Brommonschenkel SH, Pasquali G, Byrne M, Rigault P, Tibbits J, Spokevicius A, Jones RC, Steane DA, Vaillancourt RE, Potts BM, Joubert F, Barry K, Pappas GJ, Strauss SH, Jaiswal P, Grima-Pettenati J, Salse J, Van de Peer Y, Rokhsar DS, Schmutz J (2014) The genome of *Eucalyptus grandis*. *Nature* 510: 356-362
28. Derory J, Léger P, Garcia V, Schaeffer J, Hauser M-T, Salin F, Luschnig C, Plomion C, Glössl J, Kremer A (2006) Transcriptome analysis of bud burst in sessile oak (*Quercus petraea*). *New Phytologist* 170: 723-738
29. Pereira-Leal J, Abreu I, Alabaca C, Almeida M, Almeida P, Almeida T, Amorim M, Araujo S, Azevedo H, Badia A, Batista D, Bohn A, Capote T, Carrasquinho I, Chaves I, Coelho A, Costa M, Costa R, Cravador A, Egas C, Faro C, Fortes A, Fortunato A, Gaspar M, Goncalves S, Graca J, Horta M, Inacio V, Leitao J, Lino-Neto T (2014) A comprehensive assessment of the transcriptome of cork oak (*Quercus suber*) through EST sequencing. *BMC Genomics* 15: 371
30. Pammenter NW, Berjak P (2000) Aspects of recalcitrant seed physiology. *R. Bras. Fisiol. Veg.* 12 56-69
31. Barbedo CJ, Centeno DdC, Ribeiro RdCLF (2013) Do recalcitrant seeds really exist? *Hoehnea* 40: 583-593
32. Berjak P, Pammenter N (2013) Implications of the lack of desiccation tolerance in recalcitrant seeds. *Frontiers Plant Sciences* 4
33. Bonner FTV, John A. (1987) *Seed Biology and Technology of Quercus*. Gen. Tech. Rep. SO-66. New Orleans, LA: U.S. Dept of Agriculture, Forest Service, Southern Forest Experiment Station
34. Caliskan S (2014) Germination and seedling growth of holm oak (*Quercus ilex* L.): effects of provenance, temperature, and radicle pruning. *iForest - Biogeosciences and Forestry* 7: 103-109
35. Connor KF, Sowa S (2003) Effects of desiccation on the physiology and biochemistry of *Quercus alba* acorns. *Tree Physiology* 23: 1147-1152
36. Finch-Savage WE, Blake PS (1994) Indeterminate development in desiccation-sensitive seeds of *Quercus robur* L. *Seed Science Research* 4: 127-133
37. Finch-Savage WE, Blake PS, Clay HA (1996) Desiccation stress in recalcitrant *Quercus robur* L. seeds results in lipid peroxidation and increased synthesis of jasmonates and abscisic acid. *Journal of Experimental Botany* 47: 661-667
38. Goodman RC, Jacobs DF, Karrfalt RF (2005) Evaluating desiccation sensitivity of *Quercus rubra* acorns using X-ray image analysis. *Can. J. For. Res.* 35: 2823-2831

39. Liu Y, Liu G, Li Q, Liu Y, Hou L, Li G (2012) Influence of pericarp, cotyledon and inhibitory substances on sharp tooth oak (*Quercus aliena* var. *acuteserrata*) germination. PLoS ONE 7: e47682
40. Berjak P, Pammenter NW (2008) From Avicennia to Zizania: seed recalcitrance in perspective. Annals of Botany 101: 213-228
41. Ntuli TM, Finch-Savage WE, Berjak P, Pammenter NW (2011) Increased drying rate lowers the critical water content for survival in embryonic axes of english oak (*Quercus robur* L.) seeds. Journal of Integrative Plant Biology 53: 270-280
42. Pieruzzi FP, Dias LLC, Balbuena TS, Santa-Catarina C, Santos ALWd, Floh EIS (2011) Polyamines, IAA and ABA during germination in two recalcitrant seeds: *Araucaria angustifolia* (Gymnosperm) and *Ocotea odorifera* (Angiosperm). Annals of Botany 108: 337-345
43. Angelovici R, Galili G, Fernie AR, Fait A (2010) Seed desiccation: a bridge between maturation and germination. Trends Plant Sci 15: 211-218
44. Caccere R, Teixeira SP, Centeno DC, Figueiredo-Ribeiro RdCL, Braga MR (2013) Metabolic and structural changes during early maturation of *Inga vera* seeds are consistent with the lack of a desiccation phase. Journal of Plant Physiology 170: 791-800
45. Fait A, Angelovici R, Less H, Ohad I, Urbanczyk-Wochniak E, Fernie AR, Galili G (2006) Arabidopsis seed development and germination is associated with temporally distinct metabolic switches. Plant Physiology 142: 839-854
46. Bewley JD (1997) Seed Germination and Dormancy. The Plant Cell Online 9: 1055-1066
47. Nonogaki H, Bassel GW, Bewley JD (2010) Germination—Still a mystery. Plant Science 179: 574-581
48. Weitbrecht K, Müller K, Leubner-Metzger G (2011) First off the mark: early seed germination. Journal of Experimental Botany
49. Bove J, Jullien M, Grappin P (2001) Functional genomics in the study of seed germination. Genome Biology 3: reviews1002.1001 - reviews1002.1005
50. Rajjou L, Duval M, Gallardo K, Catusse J, Bally J, Job C, Job D (2012) Seed Germination and Vigor. Annual Review of Plant Biology 63: 507-533
51. Nambara E, Okamoto M, Tatematsu K, Yano R, Seo M, Kamiya Y (2010) Abscisic acid and the control of seed dormancy and germination. Seed Science Research 20: 55-67
52. Sreenivasulu N, Harshavardhan VT, Govind G, Seiler C, Kohli A (2012) Contrapuntal role of ABA: Does it mediate stress tolerance or plant growth retardation under long-term drought stress? Gene 506: 265-273
53. Kanno Y, Jikumaru Y, Hanada A, Nambara E, Abrams SR, Kamiya Y, Seo M (2010) Comprehensive hormone profiling in developing Arabidopsis seeds: examination of the site of ABA biosynthesis, ABA transport and hormone interactions. Plant and Cell Physiology 51: 1988-2001
54. Kim H, Hwang H, Hong J-W, Lee Y-N, Ahn IP, Yoon IS, Yoo S-D, Lee S, Lee SC, Kim B-G (2012) A rice orthologue of the ABA receptor, OsPYL/RCAR5, is a positive regulator of the ABA signal transduction pathway in seed germination and early seedling growth. Journal of Experimental Botany 63: 1013-1024
55. KeÇpczyński J, KeÇpczyńska E (1997) Ethylene in seed dormancy and germination. Physiologia Plantarum 101: 720-726
56. Kucera B, Cohn MA, Leubner-Metzger G (2005) Plant hormone interactions during seed dormancy release and germination. Seed Science Research 15: 281-307
57. Lee S, Kim S-G, Park C-M (2010) Salicylic acid promotes seed germination under high salinity by modulating antioxidant activity in Arabidopsis. New Phytologist 188: 626-637
58. Belin C, Megies C, Hauserova E, Lopez-Molina L (2009) Abscisic acid represses growth of the Arabidopsis embryonic axis after germination by enhancing auxin signaling. Plant Cell 21: 2253-2268
59. Preston J, Tatematsu K, Kanno Y, Hobo T, Kimura M, Jikumaru Y, Yano R, Kamiya Y, Nambara E (2009) Temporal expression patterns of hormone metabolism genes during imbibition of *Arabidopsis thaliana* seeds: A comparative study on dormant and Non-dormant accessions. Plant and Cell Physiology 50: 1786-1800
60. Dave A, Hernández ML, He Z, Andriotis VME, Vaistij FE, Larson TR, Graham IA (2011) 12-oxo-phytodienoic acid accumulation during seed development represses seed germination in Arabidopsis. The Plant Cell Online 23: 583-599

61. Farrant JM, Moore JP (2011) Programming desiccation-tolerance: from plants to seeds to resurrection plants. *Current Opinion in Plant Biology* 14: 340-345
62. Prewein C, Endemann M, Reinöhl V, Salaj J, Sunderlikova V, Wilhelm E (2006) Physiological and morphological characteristics during development of pedunculate oak (*Quercus robur* L.) zygotic embryos. *Trees - Structure and Function* 20: 53-60
63. Balbuena TS, Jo L, Pieruzzi FP, Dias LLC, Silveira V, Santa-Catarina C, Junqueira M, Thelen JJ, Shevchenko A, Floh EIS (2011) Differential proteome analysis of mature and germinated embryos of *Araucaria angustifolia*. *Phytochemistry* 72: 302-311
64. Parkhey S, Naithani SC, Keshavkant S (2014) Protein metabolism during natural ageing in desiccating recalcitrant seeds of *Shorea robusta*. *Acta Physiologiae Plantarum* 36: 1649-1659
65. Ramon M, Rolland F, Sheen J (2008) Sugar Sensing and Signaling. *The Arabidopsis Book*: e0117
66. Finkelstein RR, Lynch TJ (2000) Abscisic acid inhibition of radicle emergence but not seedling growth is suppressed by sugars. *Plant Physiology* 122: 1179-1186
67. Bailly C, Audigier C, Ladonne F, Wagner MH, Coste F, Corbineau F, Côme D (2001) Changes in oligosaccharide content and antioxidant enzyme activities in developing bean seeds as related to acquisition of drying tolerance and seed quality. *Journal of Experimental Botany* 52: 701-708
68. Neale D, Wegrzyn J, Stevens K, Zimin A, Puiu D, Crepeau M, Cardeno C, Koriabine M, Holtz-Morris A, Liechty J, Martinez-Garcia P, Vasquez-Gross H, Lin B, Zieve J, Dougherty W, Fuentes-Soriano S, Wu L-S, Gilbert D, Marcais G, Roberts M, Holt C, Yandell M, Davis J, Smith K, Dean J, Lorenz W, Whetten R, Sederoff R, Wheeler N, McGuire P, Main D, Loopstra C, Mockaitis K, deJong P, Yorke J, Salzberg S, Langley C (2014) Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. *Genome Biology* 15: R59
69. Ribeiro T, Loureiro J, Santos C, Morais-Cecílio L (2011) Evolution of rDNA FISH patterns in the Fagaceae. *Tree Genetics & Genomes* 7: 1113-1122
70. Alves S, Ribeiro T, Inácio V, Rocheta M, Morais-Cecílio L (2012) Genomic organization and dynamics of repetitive DNA sequences in representatives of three Fagaceae genera. *Genome* 55: 348-359
71. Brent R (2000) Genomic Biology. *Cell* 100: 169-183
72. Joyce AR, Palsson BO (2006) The model organism as a system: integrating 'omics' data sets. *Nature Reviews Molecular Cell Biology* 7: 198-210
73. Weckwerth W (2011) Green systems biology — From single genomes, proteomes and metabolomes to ecosystems research and biotechnology. *Journal of Proteomics* 75: 284-305
74. Myburg A, Bradfield J, Cowley E, Creux N, de Castro M, Hatherell TL, Mphahlele M, O'Neill M, Ranik M, Solomon L, Victor M, Zhou H, Galloway G, Horsley T, Jones N, Stanger T, Bayley A, Edwards N, Janse B (2008) Forest and fibre genomics: biotechnology tools for applied tree improvement. *Southern Forests: a Journal of Forest Science* 70: 59-68
75. Pijut PM, Woeste KE, Vengadesan G, Michler CH, Lakshmanan P (2007) Technological advances in temperate hardwood tree improvement including breeding and molecular marker applications. *In Vitro Cellular and Developmental Biology - Plant* 43: 283-303
76. Sheth B, Thaker V (2014) Plant systems biology: insights, advances and challenges. *Planta* 240: 33-54
77. Valledor L, Jorrín JsV, Rodríguez JL, Lenz C, Meijón Mn, Rodríguez R, Cañal MJs (2010) Combined proteomic and transcriptomic analysis identifies differentially expressed pathways associated to *Pinus radiata* needle maturation. *Journal of Proteome Research* 9: 3954-3979
78. Nakagami H, Sugiyama N, Mochida K, Daudi A, Yoshida Y, Toyoda T, Tomita M, Ishihama Y, Shirasu K (2010) Large-scale comparative phosphoproteomics identifies conserved phosphorylation sites in plants. *Plant Physiology*
79. Ward JA, Ponnala L, Weber CA (2012) Strategies for transcriptome analysis in nonmodel plants. *American Journal of Botany* 99: 267-276
80. Agarwal P, Parida SK, Mahto A, Das S, Mathew IE, Malik N, Tyagi AK (2014) Expanding frontiers in plant transcriptomics in aid of functional genomics and molecular breeding. *Biotechnology Journal* 12: 1480-1492
81. Heid CA, Stevens J, Livak KJ, Williams PM (1996) Real time quantitative PCR. *Genome Research* 6: 986-994
82. Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270: 467-470

83. Liang P, Pardee A (1992) Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* 257: 967-971
84. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW (1995) Serial analysis of gene expression. *Science* 270: 484-487
85. Brenner S, Johnson M, Bridgham J, Golda G, Lloyd DH, Johnson D, Luo S, McCurdy S, Foy M, Ewan M, Roth R, George D, Eletr S, Albrecht G, Vermaas E, Williams SR, Moon K, Burcham T, Pallas M, DuBridge RB, Kirchner J, Fearon K, Mao J-i, Corcoran K (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nature Biotechnology* 18: 630 - 634
86. Diatchenko L, Lau YF, Campbell AP, Chenchik A, Moqadam F, Huang B, Lukyanov S, Lukyanov K, Gurskaya N, Sverdlov ED, Siebert PD (1996) Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proceedings of the National Academy of Sciences* 93: 6025-6030
87. Parchman T, Geist K, Grahnen J, Benkman C, Buerkle C (2010) Transcriptome sequencing in an ecologically important tree species: assembly, annotation, and marker discovery. *BMC Genomics* 11: 180
88. Simon SA, Zhai J, Nandety RS, McCormick KP, Zeng J, Mejia D, Meyers BC (2009) Short-read sequencing technologies for transcriptional analyses. *Annual Review of Plant Biology* 60: 305-333
89. Al-Shanfari A, Abdullah S, Saud H, Omidvar V, Napis S (2012) Differential gene expression identified by suppression subtractive hybridization during late ripening of fruit in oil palm (*Elaeis guineensis* Jacq.). *Plant Molecular Biology Reporter* 30: 768-779
90. Lu H, Xue T, Zhang A, Sheng W, Zhu Y, Chang L, Song Y, Xue J (2013) Construction of an SSH library of *Pinellia ternata* under heat stress, and expression analysis of four transcripts. *Plant Molecular Biology Reporter* 31: 185-194
91. Xu L, Liu Z-Y, Zhang K, Lu Q, Liang J, Zhang X-Y (2013) Characterization of the *Pinus massoniana*. Transcriptional response to *Bursaphelenchus xylophilus* infection using suppression subtractive hybridization. *International Journal of Molecular Sciences* 14: 11356-11375
92. Abid G, Sassi K, Muhovski Y, Jacquemin J-M, Mingeot D, Tarchoun N, Baudoin J-P (2012) Identification and analysis of differentially expressed genes during seed development using suppression subtractive hybridization (SSH) in *Phaseolus vulgaris*. *Plant Molecular Biology Reporter* 30: 719-730
93. Chandran D, Sankararamasubramanian HM, Kumar MA, Parida A (2014) Differential expression analysis of transcripts related to oil metabolism in maturing seeds of *Jatropha curcas* L. *Physiology and Molecular Biology of Plants* 20: 181-190
94. Gimeno-Gilles C, Lelièvre E, Viau L, Malik-Ghulam M, Ricoult C, Niebel A, Leduc N, Limami AM (2009) ABA-mediated inhibition of germination is related to the inhibition of genes encoding cell-wall biosynthetic and architecture: modifying enzymes and structural proteins in *Medicago truncatula* embryo axis. *Molecular Plant* 2: 108-119
95. Canales J, Flores-Monterrosso A, Rueda-López M, Avila C, Cánovas F (2010) Identification of genes regulated by ammonium availability in the roots of maritime pine trees. *Amino Acids* 39: 991-1001
96. Soler M, Serra O, Molinas M, Huguet G, Fluch S, Figueras M (2007) A genomic approach to sSuberin biosynthesis and cork differentiation. *Plant Physiology* 144: 419-431
97. Bustin S (2000) Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *Journal of Molecular Endocrinology* 25: 169-193
98. Taylor S, Wakem M, Dijkman G, Alsarraj M, Nguyen M (2010) A practical approach to RT-qPCR—Publishing data that conform to the MIQE guidelines. *Methods* 50
99. Bustin SA, Benes V, Nolan T, Pfaffl MW (2005) Quantitative real-time RT-PCR – a perspective. *Journal of Molecular Endocrinology* 34: 597-601
100. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2- $\Delta\Delta$ CT method. *Methods* 25: 402-408
101. 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
102. Xu M, Zhang B, Su X, Zhang S, Huang M (2011) Reference gene selection for quantitative real-time polymerase chain reaction in *Populus*. *Analytical Biochemistry* 408: 337-339

103. Soler M, Serra O, Molinas M, García-Berthou E, Caritat A, Figueras M (2008) Seasonal variation in transcript abundance in cork tissue analyzed by real time RT-PCR. *Tree Physiology* 28: 743-751
104. Brunner A, Yakovlev I, Strauss S (2004) Validating internal controls for quantitative plant gene expression studies. *BMC Plant Biology* 4: 14
105. Gutierrez L, Mauriat M, Guenin S, Pelloux J, Lefebvre J, Louvet R, Rusterucci C, Moritz T, Guerneau F, Bellini C, Van Wuytswinkel O (2008) The lack of a systematic validation of reference genes: a serious pitfall undervalued in reverse transcription-polymerase chain reaction (RT-PCR) analysis in plants. *Plant Biotechnology Journal*
106. Ruiz-Ruiz S, Moreno P, Guerri J, Ambrós S (2007) A real-time RT-PCR assay for detection and absolute quantitation of *Citrus tristeza* virus in different plant tissues. *Journal of Virological Methods* 145: 96-105
107. Prieto-Álamo M-J, Abril N, Osuna-Jiménez I, Pueyo C (2009) *Solea senegalensis* genes responding to lipopolysaccharide and copper sulphate challenges: Large-scale identification by suppression subtractive hybridization and absolute quantification of transcriptional profiles by real-time RT-PCR. *Aquatic Toxicology* 91: 312-319
108. Gachon C, Mingam A, Charrier B (2004) Real-time PCR: what relevance to plant studies? *Journal of Experimental Botany* 55: 1445 - 1454
109. Remans T, Keunen E, Bex GJ, Smeets K, Vangronsveld J, Cuypers A (2014) Reliable gene expression analysis by reverse transcription-quantitative PCR: reporting and minimizing the uncertainty in data accuracy. *The Plant Cell Online* 26: 3829-3837
110. Wilkins MR, Pasquali C, Appel RD, Ou K, Golaz O, Sanchez J-C, Yan JX, Gooley AA, Hughes G, Humphery-Smith I, Williams KL, Hochstrasser DF (1996) From Proteins to Proteomes: Large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Nature Biotechnology* 14: 61-65
111. Jorrín-Novo JV, Maldonado AM, Echevarría-Zomeño S, Villedor L, Castillejo MA, Curto M, Valero J, Sghaier B, Donoso G, Redondo I (2009) Plant proteomics update (2007–2008): Second-generation proteomic techniques, an appropriate experimental design, and data analysis to fulfill MIAPE standards, increase plant proteome coverage and expand biological knowledge. *Journal of Proteomics* 72: 285-314
112. Siuti N, Kelleher NL (2007) Decoding protein modifications using top-down mass spectrometry. *Nature Methods* 4: 817-821
113. Han X, Jin M, Breuker K, McLafferty FW (2006) Extending top-down mass spectrometry to proteins with masses greater than 200 kilodaltons. *Science* 314: 109-112
114. Zabrouskov V, Giacomelli L, van Wijk KJ, McLafferty FW (2003) A new approach for plant proteomics: characterization of chloroplast proteins of *Arabidopsis thaliana* by top-down mass spectrometry. *Molecular & Cellular Proteomics* 2: 1253-1260
115. Guda C, King BR, Pal LR, Guda P (2009) A top-down approach to infer and compare domain-domain interactions across eight model organisms. *PLoS ONE* 4: e5096
116. Jorrín-Novo JV, Pascual J, Lucas RS, Romero-Rodríguez MC, Ortega MR, Lenz C, Villedor L (2015) Fourteen years of plant proteomics reflected in “Proteomics”: Moving from model species and 2-DE based approaches to orphan species and gel-free platforms. *PROTEOMICS*: n/a-n/a
117. Tsakanikas P, Manolakos ES (2011) Protein spot detection and quantification in 2-DE gel images using machine-learning methods. *PROTEOMICS* 11: 2038-2050
118. Grossmann J, Fischer B, Baerenfaller K, Owiti J, Buhmann JM, Gruissem W, Baginsky S (2007) A workflow to increase the detection rate of proteins from unsequenced organisms in high-throughput proteomics experiments. *PROTEOMICS* 7: 4245-4254
119. Vaudel M, Burkhardt JM, Sickmann A, Martens L, Zahedi RP (2011) Peptide identification quality control. *PROTEOMICS* 11: 2105-2114
120. Eisenacher M, Schnabel A, Stephan C (2011) Quality meets quantity – quality control, data standards and repositories. *PROTEOMICS* 11: 1031-1036
121. Ünlü M, Morgan ME, Minden JS (1997) Difference gel electrophoresis. A single gel method for detecting changes in protein extracts. *Electrophoresis* 18: 2071-2077

122. McNamara LE, Kantawong FA, Dalby MJ, Riehle MO, Burchmore R (2011) Preventing and troubleshooting artefacts in saturation labelled fluorescence 2-D difference gel electrophoresis (saturation DiGE). *PROTEOMICS* 11: 4610-4621
123. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnology* 17: 994-999
124. Ong S-E, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular & Cellular Proteomics* 1: 376-386
125. Wiese S, Reidegeld KA, Meyer HE, Warscheid B (2007) Protein labeling by iTRAQ: A new tool for quantitative mass spectrometry in proteome research. *PROTEOMICS* 7: 340-350
126. Makarov A (2000) Electrostatic axially harmonic orbital tTrapping: a high-performance technique of mass analysis. *Analytical Chemistry* 72: 1156-1162
127. Hardman M, Makarov AA (2003) Interfacing the orbitrap mass analyzer to an electrospray ion source. *Analytical Chemistry* 75: 1699-1705
128. Yates JR, Ruse CI, Nakorchevsky A (2009) Proteomics by mass spectrometry: approaches, advances, and applications. *Annual Review of Biomedical Engineering* 11: 49-79
129. de Souza G, de Godoy L, Mann M (2006) Identification of 491 proteins in the tear fluid proteome reveals a large number of proteases and protease inhibitors. *Genome Biology* 7: R72
130. Zybilov B, Rutschow H, Friso G, Rudella A, Emanuelsson O, Sun Q, van Wijk KJ (2008) Sorting signals, N-terminal modifications and abundance of the chloroplast proteome. *PLoS ONE* 3: e1994
131. Nagaraj N, Alexander Kulak N, Cox J, Neuhauser N, Mayr K, Hoerning O, Vorm O, Mann M (2012) System-wide perturbation analysis with nearly complete coverage of the yeast proteome by single-shot uHPLC runs on a bench top orbitrap. *Molecular & Cellular Proteomics* 11
132. Valledor L, Weckwerth W (2014) An improved detergent-compatible gel-fractionation LC-LTQ-Orbitrap-MS workflow for plant and microbial proteomics. In: Jorrin-Novo JV, Komatsu S, Weckwerth W, Wienkoop S (eds) *Plant Proteomics*. Humana Press, pp 347-358
133. Gallien S, Peterman S, Kiyonami R, Souady J, Duriez E, Schoen A, Domon B (2012) Highly multiplexed targeted proteomics using precise control of peptide retention time. *PROTEOMICS* 12: 1122-1133
134. Ytterberg AJ, Jensen ON (2010) Modification-specific proteomics in plant biology. *J Proteomics* 73: 2249-2266
135. Witze ES, Old WM, Resing KA, Ahn NG (2007) Mapping protein post-translational modifications with mass spectrometry. *Nature Methods* 4: 798-806
136. Nørregaard Jensen O (2004) Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry. *Current Opinion in Chemical Biology* 8: 33-41
137. Bond AE, Row PE, Dudley E (2011) Post-translation modification of proteins; methodologies and applications in plant sciences. *Phytochemistry* 72: 975-996
138. Thingholm TE, Jensen ON, Larsen MR (2009) Analytical strategies for phosphoproteomics. *PROTEOMICS* 9: 1451-1468
139. Paradela A, Albar JP (2008) Advances in the analysis of protein phosphorylation. *Journal of Proteome Research* 7: 1809-1818
140. Bendt AK, Burkovski A, Schaffer S, Bott M, Farwick M, Hermann T (2003) Towards a phosphoproteome map of *Corynebacterium glutamicum*. *PROTEOMICS* 3: 1637-1646
141. Immler D, Gremm D, Kirsch D, Spengler B, Presek P, Meyer HE (1998) Identification of phosphorylated proteins from thrombin-activated human platelets isolated by two-dimensional gel electrophoresis by electrospray ionization-tandem mass spectrometry (ESI-MS/MS) and liquid chromatography-electrospray ionization-mass spectrometry (LC-ESI-MS). *ELECTROPHORESIS* 19: 1015-1023
142. Ross AH, Baltimore D, Eisen HN (1981) Phosphotyrosine-containing proteins isolated by affinity chromatography with antibodies to a synthetic hapten. *Nature* 294: 654-656
143. Riedel J, Tischner R, Mack G (2001) The chloroplastic glutamine synthetase (GS-2) of tobacco is phosphorylated and associated with 14-3-3 proteins inside the chloroplast. *Planta* 213: 396-401
144. Agrawal GK, Thelen JJ (2006) Large scale identification and quantitative profiling of phosphoproteins expressed during seed filling in oilseed rape. *Molecular & Cellular Proteomics* 5: 2044-2059

145. Sugiyama N, Nakagami H, Mochida K, Daudi A, Tomita M, Shirasu K, Ishihama Y (2008) Large-scale phosphorylation mapping reveals the extent of tyrosine phosphorylation in Arabidopsis. *Molecular Systems Biology* 4
146. Meyer LJ, Gao J, Xu D, Thelen JJ (2012) Phosphoproteomic Analysis of Seed Maturation in Arabidopsis, Rapeseed, and Soybean. *Plant Physiology* 159: 517-528
147. Ino Y, Ishikawa A, Nomura A, Kajiwara H, Harada K, Hirano H (2014) Phosphoproteome analysis of Lotus japonicus seeds. *PROTEOMICS* 14: 116-120
148. Wang Y, Wang Y, Zhao Y, Chen D, Han Z, Zhang X (2014) Protein phosphorylation differs significantly among ontogenetic phases in Malus seedlings. *Proteome Science* 12: 31
149. Han C, Wang K, Yang P (2014) Gel-based comparative phosphoproteomic analysis on rice embryo during germination. *Plant and Cell Physiology*
150. Diella F, Gould CM, Chica C, Via A, Gibson TJ (2008) Phospho.ELM: a database of phosphorylation sites—update 2008. *Nucleic Acids Research* 36: D240-D244
151. Gnäd F, Ren S, Cox J, Olsen J, Macek B, Orosi M, Mann M (2007) PHOSIDA (phosphorylation site database): management, structural and evolutionary investigation, and prediction of phosphosites. *Genome Biology* 8: R250
152. Tchieu JH, Fana F, Fink JL, Harper J, Nair TM, Niedner RH, Smith DW, Steube K, Tam TM, Veretnik S, Wang D, Gribskov M (2003) The PlantsP and PlantsT Functional Genomics Databases. *Nucleic Acids Research* 31: 342-344
153. Durek P, Schmidt R, Heazlewood JL, Jones A, MacLean D, Nagel A, Kersten B, Schulze WX (2010) PhosPhAt: the *Arabidopsis thaliana* phosphorylation site database. An update. *Nucleic Acids Research* 38: D828-D834
154. Gao J, Agrawal GK, Thelen JJ, Xu D (2009) P3DB: a plant protein phosphorylation database. *Nucleic Acids Research* 37: D960-D962
155. Yao Q, Ge H, Wu S, Zhang N, Chen W, Xu C, Gao J, Thelen JJ, Xu D (2014) P3DB 3.0: From plant phosphorylation sites to protein networks. *Nucleic Acids Research* 42: D1206-D1213
156. Pichersky E, Gang DR (2000) Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective. *Trends in Plant Science* 5: 439-445
157. De Luca V, St Pierre B (2000) The cell and developmental biology of alkaloid biosynthesis. *Trends in plant science* 5: 168-173
158. Hall RD (2006) Plant metabolomics: from holistic hope, to hype, to hot topic. *New Phytologist* 169: 453-468
159. Saito K, Matsuda F (2010) Metabolomics for functional genomics, systems biology, and biotechnology. *Annual Review of Plant Biology* 61: 463-489

CHAPTER 2:

OBJECTIVES

Quercus ilex (Holm oak) is an important tree species of the Spanish Mediterranean forest, with great ecological value, as it represents, with *Q. suber*, the predominant tree species in the forestal ecosystems of the region. The Holm oak has also an important economic value, being widely used in reforestation and conservation forestry systems and silvicultural practices. The research group in which this work was developed has studied different aspects of *Q. ilex* biology (natural variation, physiology, stress-response) using a multidisciplinary approach that includes techniques of classical biochemistry and molecular biology combined with proteomics approaches. However, the knowledge at the molecular level in this species is still scarce.

Q. ilex forest maintenance and sustainability are facing important problems and challenges related to seed viability/conservation, and plant mortality in both adult trees and young one-year-old plants after field transplantation due to adverse environmental conditions. Thus, there is an increase in demand for holm oak seedlings and favouring their nursery production.

Therefore, the objectives of this PhD thesis were:

General objective

Study the germination and seedling growth of *Q. ilex* applying a multidisciplinary “omic” approach combined with classical biochemical approaches, according to the actual trends in biosciences research, named systems biology

Specific objectives

1. Determine the expression profile of twelve protein coding genes involved in desiccation tolerance, regulation of ABA-signalling, metabolism and antioxidative defence using a targeted transcriptional analysis of mature and germinating seeds through absolute quantification of transcripts based on qRT-PCR (Ch. 4).
2. Identify differentially expressed genes between germinated seeds and seedlings of *Q. ilex* by using a Suppression Subtractive Hybridization (SSH) approach (Ch. 5).
3. Analyse changes in the proteomic profiles during germination and seedling growth through gel based (SDS-PAGE, 2-DE) and gel free (LC-MS/MS) approaches (Ch. 6).
4. Analyse the dynamic protein phosphorylation changes during seed germination and seedling development using multiplex-staining of high-resolution 2-DE gels for total protein (SYPRO Ruby) and phosphoproteins (Pro-Q Diamond) (Ch. 7).

CHAPTER 3:

PLANT MATERIAL

Plant material and the study site

This section describes the plant material used in this work, the characteristics of the site where it was collected, the storage methods and germination experiments performed to obtain the embryo axes to be used in the different procedures that conform this Doctoral Thesis.

3.1. Plant material collection, study site and storage

Mature acorns were harvested during October-November from healthy holm oaks from Cerro Muriano-Córdoba (Córdoba, Spain 37°59'57.74"N, 4°46'57.93"W) in 2011 and 2012. The acorns provenance area measures approximately 80 km² and is situated in the north of the Andalusian province of Córdoba, at SW Spain (Fig. 3.1)



Figure 3.1. Location of the sampling area, Cerro Muriano, in the Córdoba Province (SW Spain; GPS coordinates 37°, 59', 57.74" N, 4°, 46', 57.93").

The most typical landscape in the area is the so-called Spanish dehesa, a savannah-like woodland that represents one of the better preserved ecosystems in southern Europe and that has been declared a biosphere reserve by UNESCO. Large areas of the dehesas are covered by Holm oak (*Q. ilex*) and Cork oak (*Q. suber*), surrounded by meadows and Mediterranean scrub. This peculiar agroforestry system is conditioned by a Mediterranean climate, albeit somewhat influenced by the Atlantic Ocean. The presence of soils with low fertility and poorly developed and the hilly topography of Sierra Morena makes arable farming unprofitable.

To reduce variability and growth differences due to seed size, only those acorns with a FW between 3.1 and 3.9 g, which represented the upper and lower quartiles of acorn FW, were used in the study. After collection and selection, the acorns were processed according to Bonner¹. Briefly, acorns were immersed in water in the laboratory and all of those still floating after 1 h were removed. Healthy acorns were swan in 2.5% (v/v) sodium hypochlorite solution for 10 min, and then washed with water. The acorns were air dried and stored at 4±1 °C in airtight polyethylene bag until use, as recommended by². Polyethylene bags reduce the acorns

metabolism and allow the accumulation of ethylene inside the storage atmosphere, inducing a dormant-like state and prolonging seed storage ².

3.2. Germination/seedling experiment and morphological aspect

In the first week after harvesting, undamaged, mature acorns were sterilized by immersion in 2.5 % sodium hypochlorite for 10 minutes, washed abundantly with water and finally dried with filter paper. In order to get a homogeneous germination ³⁻⁴ and to avoid delayed and uneven germination, acorns were peeled off eliminating the pericarp and cutting off parts of the distal ends of the acorns and put in plastic boxes containing one sheet of whatman N°3 filter paper over wet perlite (Fig. 3.2), a double layer of tissue paper was used to cover the seeds to insure uniform hydration. Previously, was described in other *Quercus* species, that the cotyledon reserves at the apex of acorns are more important than those at the base in supporting acorn viability and seedling establishment ⁵.

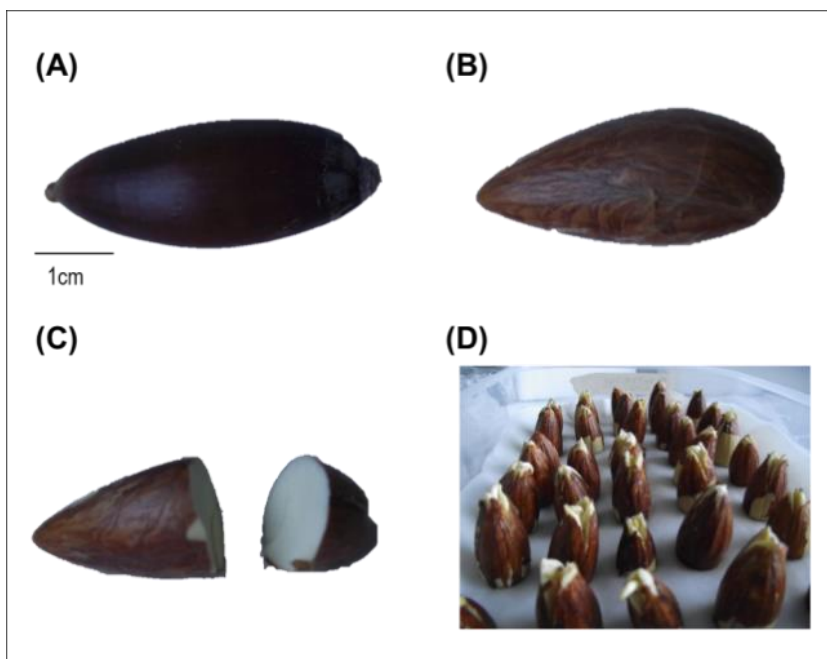


Figure 3.2: Process in acorn germination. (A). Acorn with pericarp, (B). Acorn dehulling of, (C). Cut acorn in the distal end and (D). Boxes with filter paper and perlite with germinated acorn.

Germination and seedling growth were carried out at 22 ± 1 °C for up to 216 h in darkness incubators. Radicle protrusion of 2 mm was used as the criterion of germination; this can be seen mainly at 24 hours post imbibition. Analyzed stages different sections are shown in Fig. 3.3.

To obtain shoot seedling material used in Ch. 5 and 6, seed that were maintained in darkness for 216 h (when the protruding radicle achieved the length of the longest dimension of the seed) were placed individually in 400 mL pots containing perlite, which was

continuously watered. These pots were transferred to a greenhouse under a natural photoperiod of approximately 12/12 h light/dark cycle, temperature of 30 ± 5 °C and 60 ± 10 % relative humidity (RH).

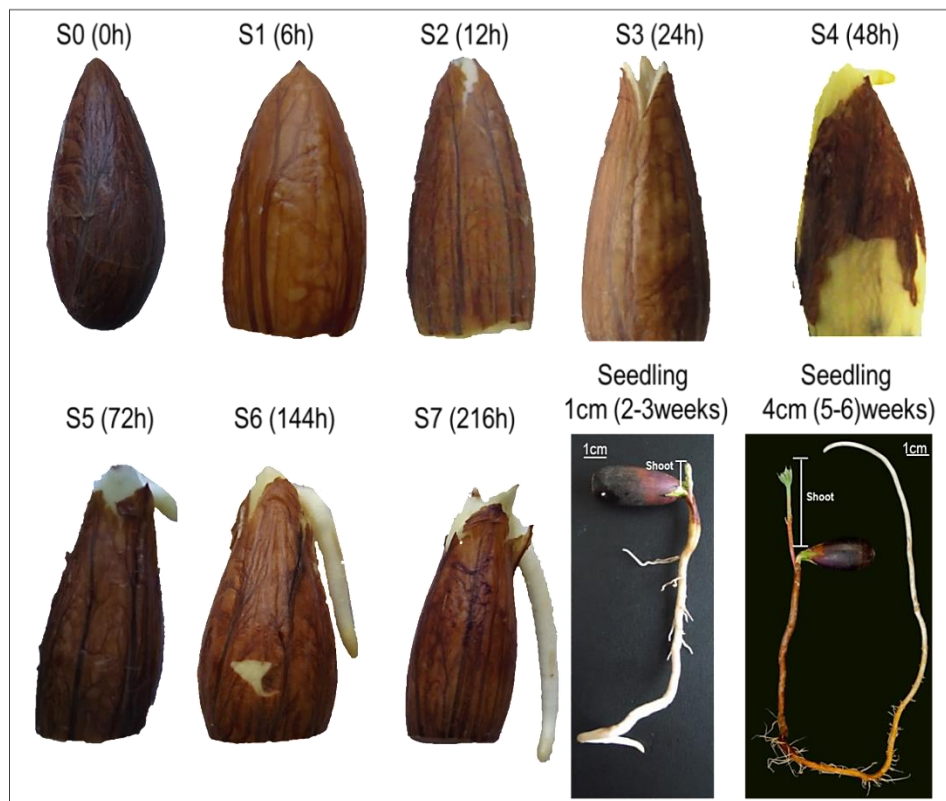


Figure 3.3: Morphological aspects of acorn and seed at different stages collected. In parenthesis are indicated approximate time in hour post imbibition to obtain the indicated stage. Approximate time for seedlings are shown in weeks.

All the experiments were performed with the embryo axis, including the radicle in S4 to S7 stages (Fig. 3.4), but not in the seedlings (Fig. 3.3). Embryo axes were rapidly removed, weighted and individually frozen in liquid N. Individuals were grouped in three mini-pools (20-30 individuals), that were considered as biological replicates.

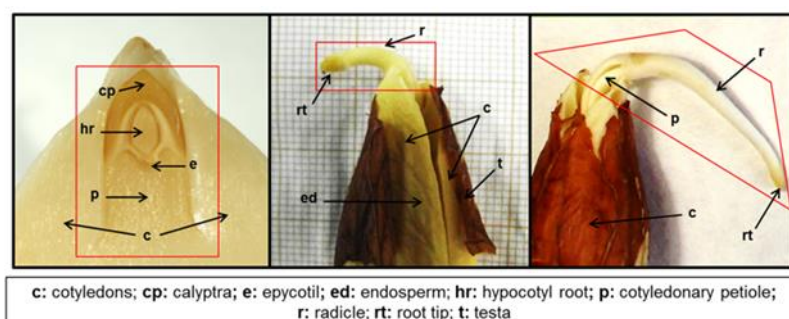


Figure 3.4. Material collected from the *Quercus ilex* seed for all experiments. Plant tissue collected from the seed and seedlings for all experiment are indicated with red line.

3.3. References

1. Bonner FT (2003) Collection and care of acorns. A practical guide for seed collectors and nursery managers version
2. Pasquini S, Mizzau M, Petrusa E, Braidot E, Patui S, Gorian F, Lambardi M, Vianello A (2012) Seed storage in polyethylene bags of a recalcitrant species (*Quercus ilex*): analysis of some bio-energetic and oxidative parameters. *Acta Physiologiae Plantarum*: 1-12
3. Liu Y, Liu G, Li Q, Liu Y, Hou L, Li G (2012) Influence of pericarp, cotyledon and inhibitory substances on sharp tooth Oak (*Quercus aliena* var. *acuteserrata*) germination. *PLoS One* 7: e47682
4. Giertych M, Suszka J (2011) Consequences of cutting off distal ends of cotyledons of *Quercus robur* acorns before sowing. *Annals of Forest Science* 68: 433-442
5. Hou X, Yi X, Yang Y, Liu W (2010) Acorn germination and seedling survival of *Q. variabilis*: effects of cotyledon excision. *Annals of Forest Science* 67: 711-711

CHAPTER 4:

TRANSCRIPTIONAL, HORMONAL AND SUGAR CONTENT CHANGES DURING GERMINATION AND EARLY DEVELOPMENT OF *QUERCUS ILEX* SEEDS

Abstract

We have used a targeted strategy based on reverse transcription of total RNA and real time-PCR amplification for absolute quantitation of the transcript levels of a group of twelve genes in germinating acorns. The transcriptional analysis results have been verified and complemented with the determination of (i) plant hormones levels (ABA; the gibberellins GA3 y GA4; the auxin IAA and the cytokinins iP and iPR,) (ii) sugars accumulation (Suc, Glc, Fru) and (iii) proteins amounts determined by immunoblotting (DHN3, GAPDH, RBCL) and/or enzymatic activity (SOD). We found that mature *Q. ilex* seeds show some of the intracellular physical characteristics of orthodox seed that included (i) accumulation of non-reducing carbohydrates (sucrose) and insoluble proteins (DHN3) that contribute to the intracellular vitrified state in seeds; and (ii) accumulation of transcripts involved in the synthesis of certain osmoregulator raffinose series oligosaccharides (*Gols*), the anti-oxidative defence (*Sod1*, *Gst*) and the preparation for the development of an adult plant (*RbcL*). But the holm oak mature acorns share with other recalcitrant seeds the ability to maintain a partially active metabolism, with high level of glycolytic (*GapdH*) and mitochondrial respiratory enzymes (*Ndh6*) and the absence of ABA (*Pp2c*, *Skp1*, *Sdir1*, *Ocp3*). However, imbibition increased the respiratory rate, paralleling the soluble carbohydrate sugar increase, and indicating that mitochondria resulted affected during acorn maturation. Formate, generated as a byproduct of photorespiration, from cell wall degradation during the germination process and from the excess of glycolytic intermediates, drives the increase of FDH synthesis, helped by the increased levels of ABA in the post-germination stage, to recover carbon. The results presented here will help to increase the knowledge of the physiological changes that take place during *Q. ilex* seed germination, illustrate the importance of considering the behaviour of seeds for the afforestation projects and restoration programmes under the impending climate change in Mediterranean regions.

4.1. Introduction

A number of studies focused on germination, storage, desiccation sensitivity, and viability after storage of holm oak seeds have been conducted over the last few decades ¹⁻⁸. However the mechanisms involved in the recalcitrance of holm oak seeds remain elusive. In this work was used a targeted strategy based on reverse transcription of total RNA and real time-PCR amplification for absolute quantitation of the transcript levels of a group of twelve genes in germinating acorns.

The germination process begins with seed imbibition and finishes with radicle protrusion, as was detailed in general introduction. In the orthodox seeds the major metabolic changes observed during this period are the significant reductions in the levels of the majority of different metabolites accumulated during the period of seed desiccation. In this new condition, the levels of sugar decrease, which may signify the entrance of sugars into the glycolytic pathway. The levels of some tricarboxylic acid cycle intermediates dramatically increase shortly after the imbibition, which suggests the initiation of metabolic and respiratory functions of the mitochondria ⁹. Significant changes in the seed transcript content are also described for this period, which generally signified major increases in mRNA levels of genes associated with biosynthetic processes and reductions in mRNA levels of genes associated with the degradation of amino acids or the biosynthesis of trehalose ⁹.

Abscisic acid (ABA) is considered an essential messenger in the adaptive response of plants against abiotic and biotic stresses. The accumulation of ABA in developing seeds contributes to seed maturation, acquisition of desiccation tolerance and seed dormancy. Hence, the decline of ABA shortly after imbibition but also the apparition other plant hormones (giberellins, cytokinin, indole acetic acid) are prerequisites for germination to be completed (Kanno *et al.*¹⁰ and ref. therein), the regulation of ABA signaling was detailed in general introduction.

In short, to date, the various deficiencies underlying desiccation sensitivity of recalcitrant seeds are generally conjectural as they are variably developed or expressed in the non-orthodox condition. Our research group are interested in determining the mechanisms that are involved in seed development and germination of holm oak seeds, a vital tree for soil and water conservation in the Mediterranean area, with seeds that form part of the diet of many animals, and that may constitute an important element in reforestation programs. Knowledge of the underlying biochemistry and metabolic status before and after the germination process could be important for the development and optimization of strategies for large scale propagation and germplasm conservation in this species. To this end, in the current study was

conducted a targeted transcriptional analysis of mature and germinating seeds by using an absolute quantitation approach based on reverse transcription and real-time PCR amplification. The transcriptional profiles of twelve genes coding proteins involved in desiccation tolerance (*Dnh3*, *Gols*), in the ABA-signaling regulatory proteins (*Ocp3*, *Skp1*, *Pp2c*, *Sdir1*), in the metabolism (*Fdh*, *GapdH*, *RbcL*, *Ndh6*) and in the antioxidative defence (superoxide dismutase *Sod1*, *Gst*), have been determined. The transcriptional analysis results have been verified and complemented with the determination of (i) plant hormones levels (ABA; the gibberellins GA3 y GA4; the auxin IAA and the cytokinins iP and iPR,) (ii) sugars accumulation (Suc, Glc, Fru) and (iii) proteins amounts determined by immunoblotting (DHN3, GAPDH, RBCL) and/or enzymatic activity (SOD). The results that we present here will help to increase the knowledge of the physiological changes that take place during *Q. ilex* seed germination.

4.2 Materials & Methods

4.2.1. Plant material

Samples collected at 0 (dry), 6h, 12h, 24h, 48h, 72h, 144 h and 216 h after imbibition start as described in Ch.3 were used (Fig. 3.3, p. 47). Three independent germination experiments were conducted. Individual embryonic axes from the different technical replicates collected at each sampling time were pooled and each of the three pools was considered as a biological replicate in subsequent experiments.

4.2.2. Relative water content (RWC) in *Q. ilex* seeds

Measurements of RWC were performed on 10-15 individual embryonic axes collected at each of the indicated sampling times along the three replicated germination experiments. Seeds/seedlings were weighed by using an analytical scale, with precision of 0.0001 g. The fresh weight (FW) obtained from each sample was between 3.1 and 3.9 g, which represented the upper and lower quartiles of acorn FW. Samples were placed in a pre-heated oven at 70 °C for 48 h, in order to obtain the dry weight (DW). The relative water content (RWC) in germinating seeds was expressed as percentage of lost weight $[(FW - DW) \times 100]$ relative to FW.

4.2.3. Desiccation resistance assay of germinating *Q. ilex* seeds

This assay was conducted as described by An and Lin ¹¹ with minor modifications. The seed/seedling of each given stage (Fig. 3.3, p. 47) were harvested and dried at room temperature (25±2°C) in plastic boxes with dry perlite for two weeks. The dehydrated

seeds/seedlings were re-germinated at the same condition as described in the germination experiment (plastic boxes with waterlogged filter paper on perlite at 22 °C in darkness). The desiccation resistance of the seeds/seedlings at each stage was determined as percentage of survival, defined as the percentage of seeds/seedling that elongated the radicle at least 2 mm within four or seven days of rehydration. Three independent replications were conducted for each time point. Sample size for each desiccation resistance assay ranged from 14 to 46 seeds.

4.2.4. Sugar extraction and GC-MS/MS separation-detection

Extraction procedure. The extraction of sugar was performed as described in ¹² with minor modification. Samples were prepared by mixing 20-50 embryonic axes collected of imbibed acorns at each studied stage and lyophilized (-40°C, 1 mBar, in a SubliMate® Bench Top Laboratory Freeze Dryers FDL-2S8) 100 mg of pooled lyophilized sample was placed in a test tube with 6.5 mL of extractant (2:1 dichloro- methane/methanol). The tube was immersed into a water bath at room temperature for a 10 min preset time under ultrasonic irradiation. The extracts were subsequently centrifuged at 10000g for 2 min and concentrated to dryness under vacuum at 30°C. The dried extracts were reconstituted in 1 mL of deionized water (18 MΩ3cm, Milli-Q Millipore, Bedford, MA) water and filtered through 0.45 µm Millipore nylon membrane. Sugar extracted (contained in dry 20 µL) were separated and identified by GC-MS of their per-*O*-trimethylsilylated methyl oxymes ¹³. Calibration curves from standards D-(+)-glucose, D-(-) fructose and D-(+)-Sucrose (5, 10, 20, 30, 40, and 50 µg) were prepared for comparison. In any case, 10 µg of xylitol were added as reference to the samples and freeze-dried. All standards and reference were purchased from Sigma-Aldrich. For derivatisation, 30 µL of a solution of MEOX (methoxyl-amine hydrochloride) in pyridine (20 mg/mL) were added and kept at 40 °C for 1 h; then, 80 µL of BSTFA (N,*O*-bis(trimethylsilyl)-trifluoroacetamide) were added for 1 h at 40 °C. All preparation was performed in triplicate.

GC-MS/MS Separation-Detection. Gas-liquid chromatography coupled with mass spectrometry (GC-MS) was performed on an Agilent Technologies GC system 7890A coupled to a mass spectrometer 5975C fitted with a column HP-5MS (30 m × 0.25 mm, Agilent). The temperature program was isothermal at 150 °C for 3 min, followed by a 5 °C/min gradient up to 210 °C, a 15 °C/min gradient up to 310 °C, and isothermal for 2 min. The ionization potential was 70 eV, and spectra were recorded in low-resolution mode. All measure was performed per duplicated.

4.2.5. Phytohormone determinations by using LC-MS/MS

Extraction procedure. Phytohormones were extracted from 150 mg of lyophilized tissue prepared from pooled embryos axis (20-50) of 0 (dry), 24 and 216 hours post imbibition (Fig. 3.4, p. 47). Samples were extracted with 1500 μ L of 2:1:0.002 2-propanol/water/Hydrochloric acid, in this phase internal standards had been added (2 ng 2H6-ABA, 2H5-AIA, 2H3-CS, 2H3DHJA, 2H3-DHZ, 2H2-GA3, 2H2-GA9, Olchemin Ltd., Czeck Republic). Each sample was placed in an orbital shaker for 30 min at 4°C, and then 1000 μ L of dichloromethane were added and stirred for 30 min more at 4 °C. The organic phase was carefully removed and the residual liquid was re-extracted with 1000 μ L of dichloromethane. Following a further 30 min of shaking at 4°C, the organic phase was pooled; this re-extraction step was repeated one time. Organic phase of three extractions were reunified into one fraction and reduced to approximately to half of the initial volume by a mild nitrogen flow. All stages were processed in triplicate.

LC-MS/MS Separation-Detection. Samples were re-suspended in 200 μ L of 100% methanol and filtered through a 0.2 μ m regenerated cellulose filter (Agilent Technologies). All the compounds were separated and quantified by an ultra-high performance liquid chromatography (UHPLC) in a 6460 Triple Quad LC/MS (Agilent Technologies) using the protocol described by Novak et al (2008) ¹⁴ and performed for the plant phytohormones analyzed. A chromatographic separation was performed using a reverse phase column (Zorbax SB-C18 2.1 x 50 mm column). The column was held at 40 °C and the mobile phase used in the chromatography consisted of (A) 99.9% methanol, 0.1% formic acid and (B) ammonium formate (15 mM, pH 4). A linear gradient of methanol from 10% to 50% and then reaching 100% in 7 and 2 minutes, respectively was used to analytes elution. Phytohormones were quantified by dynamic multireaction monitoring (MRM) of their [M+H]⁺ and the appropriate product ions, using optimized cone voltages and collision energies for diagnosis of each PGRs analyzed. Acquisition and quantification were made using Mass Hunter software (Agilent MassHunter Workstation, California, USA) and the obtained amounts of the different phytohormones based on the plant dry mass were used to be statistically analysed. All measure was performed per duplicated.

4.2.6. RNA isolation and cDNA synthesis

For RNA isolation, 20-50 germinating seeds/seedlings per sampling time were carefully selected based on the relative time point and morphology defined for each studied stage, to reduce the heterogeneity of embryo axis tissues. The embryo axis, including the radicle

in the seedlings (Fig. 3.4, p. 47), were rapidly removed, weighted and individually frozen in liquid nitrogen. Individuals were grouped in three mini-pools which were considered as biological replicates. Frozen plant material was hand grounded using a mortar and pestle in liquid nitrogen and stored at -80 °C till use. Total RNA was extracted from 20 mg per sample by using the InviTrap® Spin Plant RNA Mini Kit (Invitek), following the manufacturer indications with modifications¹⁵. Contaminating genomic DNA was removed by DNase I (Ambion) treatment. Total RNA quantitation was performed with the Qubit® RNA Assay Kit in the Qubit® 2.0 Fluorometer (Invitrogen) and the RNA quality was checked electrophoretically (Agilent 2100 Bioanalyzer). Only high-quality RNAs with RIN values > 8 and A260:A280 ratios of approximately 2.0 were used for subsequent experiments¹⁶⁻¹⁷. Lack of any remaining DNA contamination was confirmed by PCR amplification of the RNA samples without the previous step of cDNA synthesis.

The cDNA was generated from 1 µg of total RNA using the iScript cDNA synthesis kit (BioRad) according to the manufacturer's instructions. Reverse transcriptions were set up from RNA samples of identical concentrations in order to add the same volume to the RT reaction¹⁷. A "normalizer" RNA with a known amount of transcripts of the A170 gene was introduced in each experiment (i) to guarantee the quality of the retro-transcription and (ii) to set the threshold in the different qRT-PCR experiments.

4.2.7. Sequence analysis and primer design

Candidate genes involved in germination were identified by searching the literature. The sequences of *Dhn3*, *GapdH*, *Sod1* and *RbcL* was obtained from the GenBank database (<http://www.ncbi.nlm.nih.gov>). To obtain the *Q. ilex* genic sequences of *Gols*, *Ocp3*, *Skp1*, *Pp2c*, *Sdir1*, *Fdh*, *Ndh6* and *Gst*, the orthologous sequences from different phylogenetically related species (preferentially *Quercus* > *Fagaceas* > *Fagales* > other plants) were obtained from the GenBank database (<http://www.ncbi.nlm.nih.gov>), and aligned for each gene, using the ClustalW software (Megalign, DNASTAR Lasergene, v.6). From each alignment, the conserved sequences among species were defined, and the primer pair was designed over these conserved sequence zones. Possible primer pairs were obtained using the Primer-BLAST tool of NCBI (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/NCBI/>). As an example, Fig. 4.1 shows the alignment of *Fdh* sequences from different species related to *Q. ilex*, as well as the position of the primers used for amplification and sequencing of the corresponding *Q. ilex Fdh* fragment sequence.



Figure 4.1: Location of the primers used to amplify and sequencing of a fragment of *Fdh* coding sequence. Orthologous sequences were aligned and the primers designed over the close sequence. Primers were located in highly conserved sites of the 5' CDS.

Then, the proposed primer pairs were analysed with the OLIGO Primer Analysis Software v 7.58 (Molecular Biology Insights, Inc) and one pair for each gene, with high Tm and free of hairpins and duplex structures, was chosen. Orthologous primers (Table 4.1) were used to amplify synthesized cDNA from pooled *Q. ilex* RNAs.

Table 4.1. Primers used in this work for the determination of *Q. ilex* sequences. Primers based on orthologous sequences were designed to obtain partial coding sequences of *Q. ilex* genes.

Target ^a	Genebank Acc. number	Sequence 5'-3' ^b	Length ^c (bp)
<i>Fdh</i>	AJ577266	F: TCACTAGGCATCTTCATGCTTCTCCTGG R: CCCAGCACCAACAGTTCCAACCTGTC	545
<i>Gst</i>	DQ673318	F: GCGGGCAAGGTGGTTCTGTTGGATT R: GCCACAGCCTCCTTCTGCATGCAC	584
<i>GolS</i>	XM_002515187	F: GATTACATTAAGGCGTTGTTGGGTTGG R: CTCCATATTCTCTTTTCCAGTATATCTCC	729
<i>Ndh6</i>	FJ669574	F: TGATGGTTGCACGTGCTAAAAATCC R: TCCAACCTTCGTACCTTCCGGCATAAA	360
<i>Ppc2</i>	AJ277743	F: CCTTTACATTGCCAACCTTGGTGATTC R: CTGCCAACTCCATCTGCGTAGTATAA	705
<i>Ocp3</i>	XM_002324251	F: CCACTCTTTTCCGAGGCAGGTTTGTGAC R: CTCTTTTCCGAGGCAGGTTTGTGAC	261
<i>Sdir1</i>	NM_115410	F: AAGACATGAGCTTTGTTTCCGGGGAAG R: AGGATCCAACACCTTGTACTTGTGAACTGG	530
<i>Skp1</i>	XM_002510530	F: ACTAGCAGGAAAATCACTTTGAAGAGTTCTGAC R: AAGGCCCACTGATTTTCTCTCCGAAC	461

^aGene symbols are according to the NCBI Gene database.

^bSequence of forward (F) and reverse (R) primers specific for *Q. ilex* gene sequences

^cPCR product size in base pair (bp).

PCR amplification was attained by mixing 50 ng of *Q. ilex* cDNA with 0.75U of Platinum® Taq DNA Polymerase (Life Technologies), 0.3 μM each primer, 0.3 mM dNTPs, 1mM MgSO₄ and following the manufacturer's recommendations. The amplification was carried out on T100™ Thermal Cycler PCR System (BioRad). Cycling conditions consisted in one step at 95 °C for 3 min, and 40 two-step amplification cycles at 95 °C for 30 s for the Platinum® Taq activation and, depending on the amplified gene, at 58 to 70 °C for 30 s for annealing/extension. PCR products were electrophoretically separated and visualized on agarose gels (2%) containing GelRed™ (Biotium). DNA bands of the predicted size were excised from the gel, purified (Wizard® SV Gel and PCR Clean-Up System kit, Promega) and sequenced on an ABI PRISM™ 3130 XL sequencer (Applied Biosystems). Sequences were edited to remove primers and terminal ambiguities by using the software Chromas and ChromasPro v.1.5 (Technelysium). The identity of the trimmed sequences was confirmed using tBLASTx algorithm on the BLAST server at the NCBI databank and the *Q. ilex* sequences were deposited in the GenBank database (accession numbers in Table 4.2). These sequences were used to design primers that exactly complemented the *Q. ilex* genes for the absolute quantification of transcript levels by real-time RT-PCR (qRT-PCR). To obtain high specificity and a better performance, primers free of hairpin and duplex structures, were required to have high T_m (≥70°C), and optimal 3'-ΔG (≤ -3 kcal/mol) value to be used in two-step 94°C -70°C PCR reactions. All primer pairs produced amplicons of the predicted size (Table 4.2). All PCR products were further verified by nucleotide sequencing.

Table 4.2. Specific primers used for absolute quantification by real-time PCR of *Q. ilex* transcripts.

Target ^a	Genebank Acc. number	Sequence 5'-3' ^b	Length ^c (bp)	Amplification efficiency ^d (correlation coeficiente)
Dessiccation protective proteins				
<i>Dhn3</i>	FN548081	F: CAACCCAGGTTACTACGAGGGCCAACA R: AGTTGAAGTTGTGTGACCTTGCCCAACACC	107	1.0273 (99.85)
<i>GoIS</i>	KC595272	F: TAATCTATCCACATACCATGATCTCCTTG R: GGGAGATTGGCTTGTAAATATCATTGAAG	116	0.9968 (99.99)
ABA signaling pathway				
<i>Ocp3</i>	KC150868	F: ACCTGCACCAACCTTATCATTATCGGTGCC R: CATGAACAGGTGCTTGCAGCTTGGTCCC	118	1.0187 (99.88)
<i>Sdir1</i>	KC150870	F: GTCCTCTTGTGCAAAAAGCTATTGATGGTGGC R: ATGCAACTTGTCAACAACCTCAAGCTCAGG	123	0.999 (99.44)
<i>Skp1</i>	KC150871	F: AGATCGAAGAGTGTGGCCTGGTCAACCCTG R: AGCAGTTGAGTCCACCAGTGCCGATGAG	103	0.9989 (99.83)
<i>Pp2c</i>	KC150869	F: GCAAGTGTGTAATGAGGCAGGTTTATTCCACC R: CCATTTCCACGATGACATTACAGTGATTGTTG	132	0.9981 (99.83)
Metabolism				
<i>Ndh6</i>	KC150873	F: GAAGAGCATTTCACCAAAAAGATCAGTCC R: TCCATATTCAAATAGCGGAGATTCACGAAG	95	0.9934 (99.72)
<i>GapdH</i>	HM629510	F: GGGCCGTGGAGCTGCACAAAACATCA R: TGGGACACGGAAGGCCATTCCAGTA	116	0.9938 (99.35)
<i>Fdh</i>	KC150872	F: CGTGACCTCTGCGACTGTTAATCCAG R: GGCTGAAAGGATCCAAAAGGCCAAG	118	0.9915 (99.99)
<i>RbcL</i>	AB125020	F: CGCATAAATGGTTGGGAATTAACGTTCT R: GGGATTATCCGCTAAGAATTACGGTAGA	105	0.9929 (99.71)
Redox homeostasis/detoxification				
<i>Sod1</i>	KM262658	F: CGCAGATCCAGATGATCTTGGCGAGGG R: AGCACACAACAGAGTAGGGATTAGAAGACG	137	0.9941 (99.68)
<i>Gst</i>	KC150875	F: TAATCTTGGGGCACTCCGCCTCTCTG R: ACAAGCCTTACTTTGGGGGTGAAACATTTG	126	0.9954 (99.97)
Calibrator gene^e				
<i>A170</i>	U57413	F: GGAAGAGAAGCCGCTGACACCCACT R: CCCGTCAGTTTGTCTGACTTCCGAAG	113	1.0001 (99.9%)

^aGene symbols are according to the NCBI Gene database.

^bSequence of forward (F) and reverse (R) primers specific for *Q. ilex* gene sequences

^cPCR product size in base pair (bp).

^dThe real-time PCR efficiencies (E) were calculated from each standard curve according to the equation $E = 10^{(-1/\text{slope})} - 1$. E is in the range from 0 (minimum value) to 1 (maximum and optimum), i.e., E = 1 is equal to 100% efficiency.

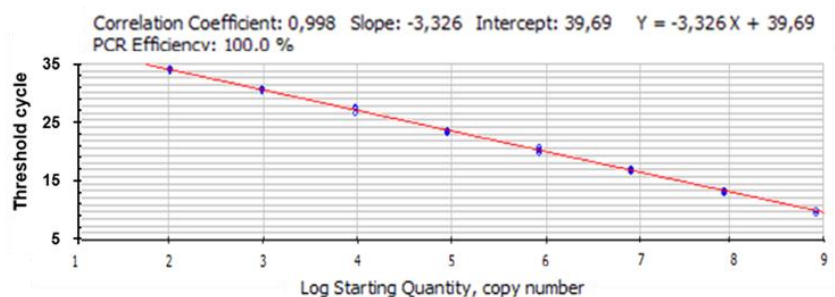
^ePrimers based on *M. musculus* sequences were designed for absolute quantification by real-time RT-PCR of A170 gene in liver mice tissue, used to guarantee the quality of the retro-transcription and to establish the threshold position.

4.2.8. qRT-PCR

Real-time PCR reactions were performed in quadruplicate with 50 ng/well of cDNA in an iCycler iQ thermocycler (BioRad) and the iQ SYBR Green SuperMix (BioRad), according to the manufacturer's indications. The amplification program consisted in one cycle at 95 °C for 3 min, and 40 two-step amplification cycles at 95 °C for 15 s and 68 °C for 30 s, respectively. After 1 min at 95°C, a melting curve was obtained by following the fluorescence intensity during a gradual cooling from 95 to 65 °C.

To carry out absolute qRT-PCR, a calibration curve was constructed with an in vitro synthesized RNA (Fig. 4.2A), as detailed previously¹⁸⁻¹⁹. The absolute calibration curve, obtained by plotting the log of the starting RNA molecules versus the threshold cycle (Ct), was linear ($r = 0.998$) over 7 orders of magnitude (Fig. 4.2A). The slope of the calibration curve indicates that the calibrator is amplified with 100% efficiency ($E = 10^{[-1/\text{slope}] - 1}$). This calibration curve ($\text{Ct} \times 50000 \text{ pg} = -3.326 \times \log N + 39.655$) was used to calculate the number of copies of each mRNA species in each total RNA sample, as described previously²⁰. The reliability of an absolute quantification depends on identical amplification efficiencies for both the target and the calibrator. Our primers were designed to amplify all amplicons with optimal ($\sim 100\%$) efficiencies and high linearity ($r > 0.99$) in the range of 20 to 2×10^5 pg of total RNA input as shown in Table 4.2 and exemplified for *Pp2c* (Fig. 4.2.B). Data from transcriptional measurements were analysed by *t*-test with the software InStat v2.05/00 (GraphPad).

(A) Calibration curve



(B) Efficiency curve of *Pp2c* transcripts

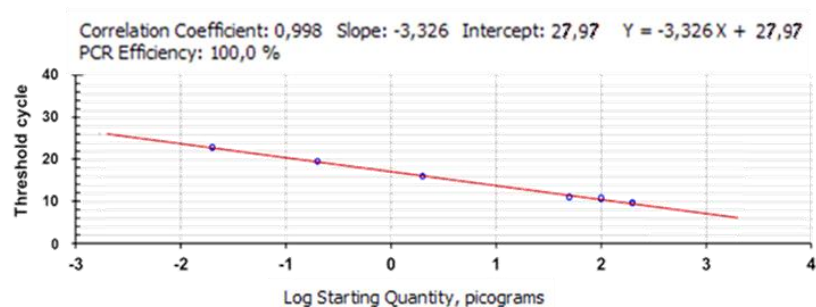


Figure 4.2. (A). Absolute standard curve used to calculate the number of copies of each experimental transcript per pg of total RNA and (B). the efficiency curve for the amplification of *Pp2c* transcripts.

4.2.9. Protein extraction and immunoblotting

Pooled embryonic protein extracts were obtained by crushing from cryohomogenates samples (200 mg, each sample in triplicate) by using trichloroacetic acid (TCA)–acetone–phenol protocols as described²¹. The final pellet was suspended for protein solubilization in 9 M urea, 4% CHAPS, 0.5% TritonX-100, and 100 mM DTT. Once the pellet was solubilized and the

insoluble material eliminated by centrifugation, the protein content was quantified by Bradford method ²², using bovine albumin as standard. Protein samples (25 µg) and markers were loaded in 12% polyacrylamide stain-free gels (Mini-PROTEAN® TGX Stain-Free™ Precast Gels, BioRad) and SDS-PAGE was performed ²³ by using the Mini Protean tetra-Cell (Bio-Rad). After electrophoresis, image of total protein were digitalized by ChemiDoc™ MP Imaging System (BioRad) and then protein were transferred onto polyvinylidene difluoride (PVDF) membranes by using the Trans-Blot® Turbo™ Transfer System (Bio-Rad). After transfer, the PVDF membranes were rinsed briefly in distilled water and the image of transfer protein were digitalized by ChemiDoc™ MP Imaging System (BioRad) to assess the quality of the transference. These images indicating the total amount of protein per lane were used to normalise the intensity of the luminescence observed after incubating the membrane with the corresponding primary and secondary antibodies ²⁴. PVDF membranes were blocked with 2% nonfat milk powder. Blots were incubated with rabbit antibodies raised against GAPDH (dilution:1:500), dehydrin (dilution:1:100) or RBCL (dilution:1:100), washed, and incubated with the secondary goat anti-rabbit IgG antibody (Sigma: A-9169, 1:2000), conjugated to horseradish peroxidase, for an additional 1 h. Blots were developed using the Clarity™ Western ECL Detection System (BioRad). Image captures and densitometric analyses were performed with the ChemiDoc MP Imaging system and ImageLab 4.1 software (BioRad), respectively.

4.2.10. SOD isoform identification and activity staining

Protein extracts were obtained from cryohomogenates samples, mixing 500 mg of each sample (in triplicate) with 1mL of 10 mM de Tris-HCl buffer (pH 7.4), 5 mM DTT, 2 mM EDTA; 0.5% (v/v) TritonX-100, 5 mM ascorbic acid, 100 mM PMSF and 10% (w/v) PVPP. The homogenates were then vortexed and sonicated 4 x 10 s (6 W). Supernatants were collected by centrifugation at 10000g for 10min at 4°C and used for enzyme assay. The protein content was quantified by Bradford method ²², using bovine albumin as standard. To preserve the enzymatic activity, 30% (w/v) saccharose were added. Total extract thus was stored at -80 °C till use. Twenty micrograms of protein was loaded in each well for activity staining assays using the native gels. SOD isoforms were separated on a 10% non-denaturing polyacrylamide (PA) gels at 4°C using a mini protean electrophoresis unit (Bio-Rad, USA). After electrophoresis, the gels were stained for SOD activity as described in ²⁵. Gels were soaked in 50 mM potassium phosphate buffer, pH 7.8, containing 1.25 mM nitro blue tetrazolium (NBT)-2HCl for 30 min in the dark at 25°C, followed by immersing in 50 mM potassium phosphate buffer, pH 7.8, 28 mM de TEMED and 30 µM riboflavin, which were then exposed under light source at room temperature. Isoforms of SOD were differentiated by activity staining of gels previously

incubated for 30 min at 25°C in 50 mM potassium phosphate buffer, pH 7.8, containing either 3 mM KCN (inhibitor of Cu/Zn SOD) or 5 mM H₂O₂ (inhibitor of Cu/Zn SOD and FeSOD)²⁶. SOD activities were quantified by converting the stained area and intensity into a relative unit by scanning the gel using the ChemiDoc MP Imaging system with the ImageLab 4.1 software (BioRad).

4.3. Results and discussion

4.3.1. Plant material

Mature acorns were harvested from healthy holm oaks from Cerro Muriano (Fig. 3.1, p. 45). Three independent germination experiments were conducted (Fig. 3.2, p. 46). After imbibition, embryo axes were collected at 0 (dry), 6 h, 12 h, 24 h, 48 h, and whole seedling collected at 72 h, 144 h and 216 h, as described in Ch. 3 (Fig. 3.3, p. 47).

4.3.2. Morphology, water uptake and desiccation sensitivity during *Q. ilex* seed germination progress

Seed germination is a complex process comprising events from seed imbibition to growth. Morphologically, initiation of growth corresponds to radicle emergence; subsequent growth is generally defined as seedling growth. To clearly define the distinct physiological germination stages of *Q. ilex* seeds, we carried out a germination experiment with healthy, regular sized acorns imbibed in water. We got an accelerated, homogeneous, synchronized germination for the 100 % of seeds by dehulling acorns prior imbibition. Morphology, water up-take and loss of seed desiccation resistance were examined over the course of seed germination, and seven developmental stages representing distinct physiology of *Q. ilex* over the course of seed germination were selected and referred as S0 to S7 stages. Figure 4.3 shows the morphology of germinating acorns at each developmental stage and the averaged time typically taken for acorns to reach the given stage. Though seed hydration already was evident in the S1 stage (6 h after imbibition, Fig. 4.3.A, B, C), changes in the acorns morphology were observed during the germination process only after reaching the S2 stage, about 12 h after imbibition, when the rupture of the testa was appreciable. Radicle emergence started to be visible 24 h (S3) after imbibition, reaching lengths of 5, 10, 20 and 55 mm at 50, 72, 144 and 216 h, respectively. Epicotil emergence from the embryonic axis and growth started to be visible 72 h (S5) after imbibition, and plumule emergence from cotyledonary petioles was observed 144 h (S6) after imbibition. After 216 h, the shoot emerged. These changes permitted to state that the holm oak acorn *early germination* period covers the first 12 h after imbibition,

ending with the *testa* rupture (stage S2) (Fig. 4.3.D). The *late germination* period covers the following 12 h, ending with the radicle emergence (stage S3) and characterized by a diminished rate of water uptake. At this point, the *post-germination* period (from S4 to S7 in our experiment) started (Fig. 4.3.D).

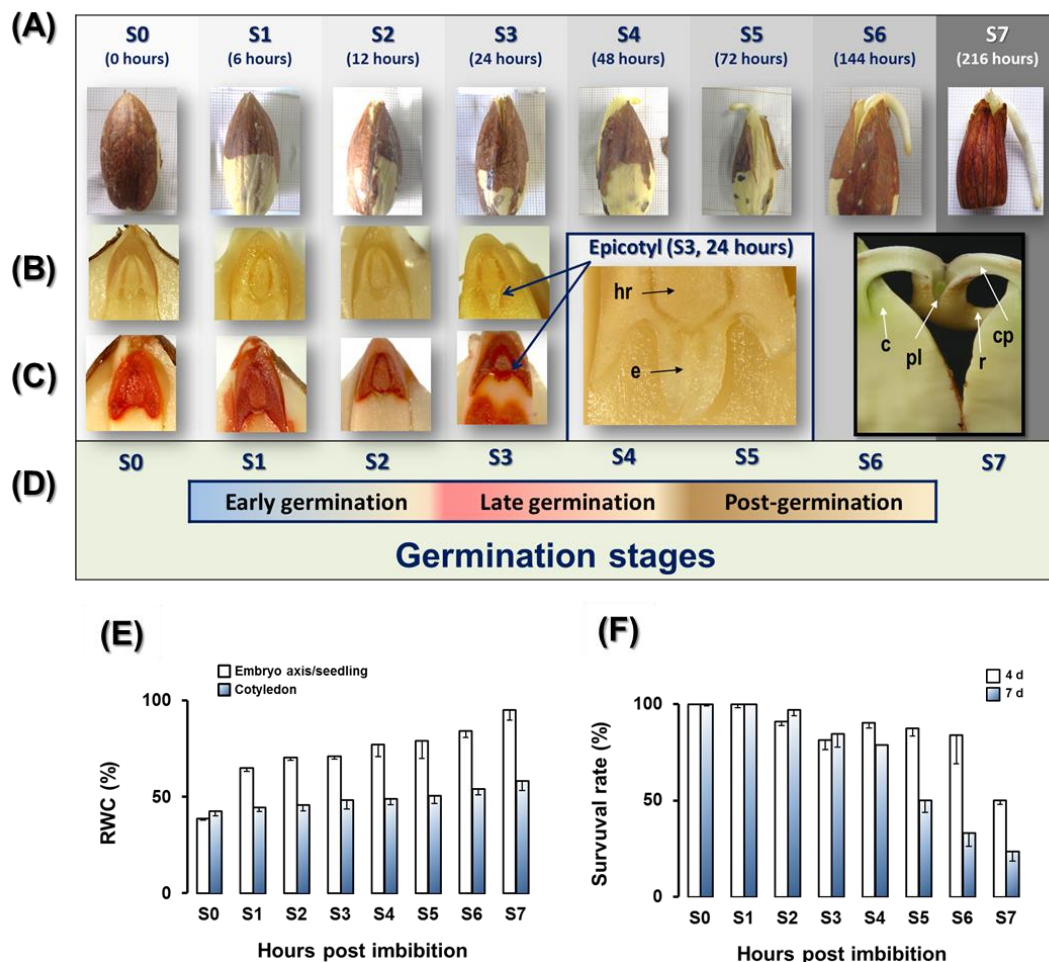


Figure 4.3. Morphology and physiology of germinating holm oak seeds at different developmental stages. (A) Morphology of germinating acorns and young seedlings at each developmental stage; the relative time taken for low hydrated seeds to reach each stage is also shown; germinating dehulled seeds shows important morphological changes after imbibition at S2: testa rupture at 12 hours of germination. S3: Radicle emergence at 24 h; S5: epycotil emergence from the embryonic axis; and S7: plumule emergence from differentiated cotyledonary petioles. (B) Seed sections showing the embryonic axes and the progressively differentiated parts of the embryo: hr: hypocotile radicle; e: epycotil; pl: plumule; c: cotyledon; cp: cotyledonary petiole; r: radicle; (C) TZ staining (1% solution of 2, 3, 5-triphenyl tetrazolium, 15 min, 35 °C, in darkness) were used to facilitate the visualization of the morphological changes produced in the embryonic axes during germination. (D) Correspondence among the seven developmental stages and the three phases of *Q. ilex* seeds germination: early (S0-S2), late (S2-S5) and post-germination (S5 and ahead) phases. (E) Percentage of relative water content (RWC %) of complete germinating holm oak seeds or their cotyledons at the different developmental stages. The fresh (FM) and dried (DM) weights (70 °C for 48 h) of 10-15 germinating acorns or seedlings were plotted per stage, with vertical bars representing \pm SE of the mean. All mass measurements were made using an analytical scale, with precision of 0.0001 g. The relative water content (RWC) in germinating seeds was expressed as percentage of lost weight $[(FM - DM) * 100]$ relative to fresh weight (FW). All weighting was done to the nearest mg. (F) Desiccation resistance of germinating *Q. ilex* seeds at each developmental stage. The seeds were collected at each developmental stage, dehydrated and then re-germinated for days. The percentage of the dehydrated germinating acorns that could revive to their growth at 4th and 7th days was defined as survival rate. Data are means \pm SD of three independent replicates. Sample size for each desiccation resistance assay ranged from 14 to 46 seeds.

The percentages of relative water content (RWC %), expressed as percentage of lost weight relative to FW of germinating holm oak seeds or their cotyledons at the different developmental stages are shown in Fig. 4.3.E. The relative water content in the mature non-germinating holm oak seeds was 38 %, according to previous reports for recalcitrant seeds (*i.e.*, Tilki *et al.* and í-Zomeño *et al.* ²⁷⁻²⁸). This relative water content increased rapidly at a rate of 2.34 % per hour during the first 12 h; then, a 36 h plateau was observed, with a rate of only 0.02 % per hour. Afterward, the water uptake increased at a rate of 0.20 % per hour until the end of the experiment (216 h), when seeds reached a RWC of 98.56 % (Fig. 4.3.E). The triphasic uptake of water by *Q. ilex* seed, with a rapid initial uptake (phase I, *i.e.* imbibition; S0-S2) followed by a plateau phase (phase II, S2-S5), and a further increase (phase III S5 and forward) as the embryo axis elongates and breaks through the covering layers to complete germination, is in agreement with previous reports in other *Quercus spp.* seeds ²⁹.

4.3.3. Quantification of sugars in embryo axis of *Q. ilex* seed and seedlings

GC-MS/MS method was applied to analyse and quantify the soluble carbohydrates sucrose, fructose and glucose contents in embryo axes and seedlings of *Q. ilex* at the pre-germination S0 stage and at the S3 and S7 post germination stages, in an effort to determine the contribution of these sugars to the metabolism of mature and germinated recalcitrant seeds. Soluble carbohydrates analysis in *Q. ilex* and other recalcitrant seeds are scarce. In orthodox seeds, maturation and acquisition of desiccation tolerance is reported to correlate well with the accumulation of sucrose and raffinose family oligosaccharides (RFOs: galactinol, raffinose, and stachyose) ³⁰. Accumulation of these non-reducing sugars offers protection by stabilization of membranes and proteins during the desiccation occurred in the last stage of orthodox seed maturation. Some authors ³¹ have proposed that sucrose must be present in relatively high quantities (10–20% of DW, such as in *Glycine max* or *Zea mays*) ³² to be effective as a cryoprotectants in mature seeds. In these conditions, the biosynthesis of raffinose and stachyose occurs via sequential transfers of galactosyl units to sucrose by raffinose synthase enzyme (Fig. 4.4.A). The galactosyl donor is galactinol, synthesized by the enzyme galactinol synthetase (GOLS) ³³. However, even in the orthodox seeds the amount of sucrose have been reported to differ from 33% in the African oil bean seeds, to only 0.3% to 3% in the seeds of some crops such as chick pea or black gram ³⁴. The level of sucrose ($83 \pm 0.4 \mu\text{g g}^{-1}$, <1% of DW, Fig. 4.4.B) accumulated in the embryonic axis of *Q. ilex* seeds (S0 stage) was in the lower range, as expected from a just partial desiccation process described for recalcitrant seeds ³⁵. The amount of reducing sugars like glucose or fructose make up only a very small amount (3.0 ± 0.7 and $6.7 \pm 0.7 \mu\text{g g}^{-1}$ DW, respectively) of *Q. ilex* seed in this S0 stage, as described for the

orthodox seed³⁴. The levels of both monosaccharides increased dramatically (10-20 folds) during the late germination stage (S7) accompanying to the decrease in the amount of sucrose observed at this stage. The germination and post germination processes are characterized by a high metabolic activity³⁶. The decrease of sucrose and the increase of monosaccharides observed in stages S3 to S7 might be associated with the increased hydrolysis of nutrients necessary for the development of seedling, because soluble sugars are usually utilized early in germination as an immediate energy source³⁷.

Data reported here suggest that during the maturation of *Q. ilex* seed a partial adaptation to dryness occurred, that implied (i) a medium-level accumulation of sucrose, the precursor of RFOs, what offers drought protection by stabilization of membranes and protein; and (ii) a reduction in the metabolic activity, by maintaining low levels of glucose and fructose, and that would avoid a precocious germination process.

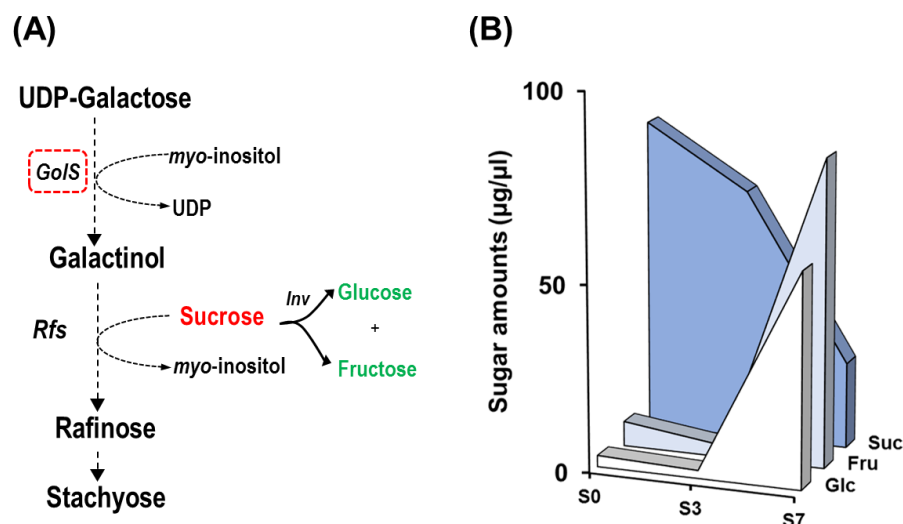


Figure 4.4. Sugar content in germinating *Q. ilex* seeds. (A) Biosynthetic pathway of raffinose family oligosaccharides (RFOs: galactinol, raffinose, and stachyose) (adapted from³³; and sucrose initial catabolism during active metabolism (*GoS*: galactinol synthase, *Rfs*: raffinose synthase, *Inv*: invertase). (B) Sucrose, glucose and fructose content in the embryo axis at the S0 (0h), S3 (24h) and S7 (216h) stages were determined by GC-MS. Data correspond to the mean \pm SE of measurements made on 20-50 seeds per sampling time, grouped in three pools. Changes between the S3 and S7 stages were statistically significantly different at the $p < 0.001$ value (ANOVA).

4.3.4. Phytohormone profiles in *Q. ilex* embryo axis and seedlings

Plant hormones are biochemical substances controlling many physiological and biochemical processes in the plant, including seed dormancy and germination³⁸. The complexity of any plant matrix and their small concentrations make phytohormone analyses particularly difficult. By applying the selectivity of reversed-phase chromatography coupled to tandem mass spectrometry, we have assessed the quantity of abscisic acid (ABA), gibberellins (GA3, GA4), indole-3-acetic acid (IAA), cytokinins (CKs: dihydrozeatin riboside, DHZR;

dihydrozeatin, DHZ; isopentenyladenine, iP; isopentenyladenine riboside, iPR); trans-zeatin, (tZ; trans-zeatin riboside, tZR) and the brassinolide castasterone (CS) in embryo axis and seedling in different germination and post germination stage of *Q. ilex* seeds. The results are summarized in Fig. 4.5.A.

Among the internal growth regulators of seed germination, abscisic acid (ABA) and gibberellin (GA) are the two phytohormones that have most pronounced effects, (reviewed by Locascio *et al.* 39). ABA establishes and maintains the seeds dormancy, whereas GA has the opposite effect of breaking the dormancy and inducing seed germination. In the mature non-germinating orthodox seed there are high levels of ABA and low levels of GA. Only when the situation changes, the increased levels of GA can counteract the germination-inhibiting effect of ABA.

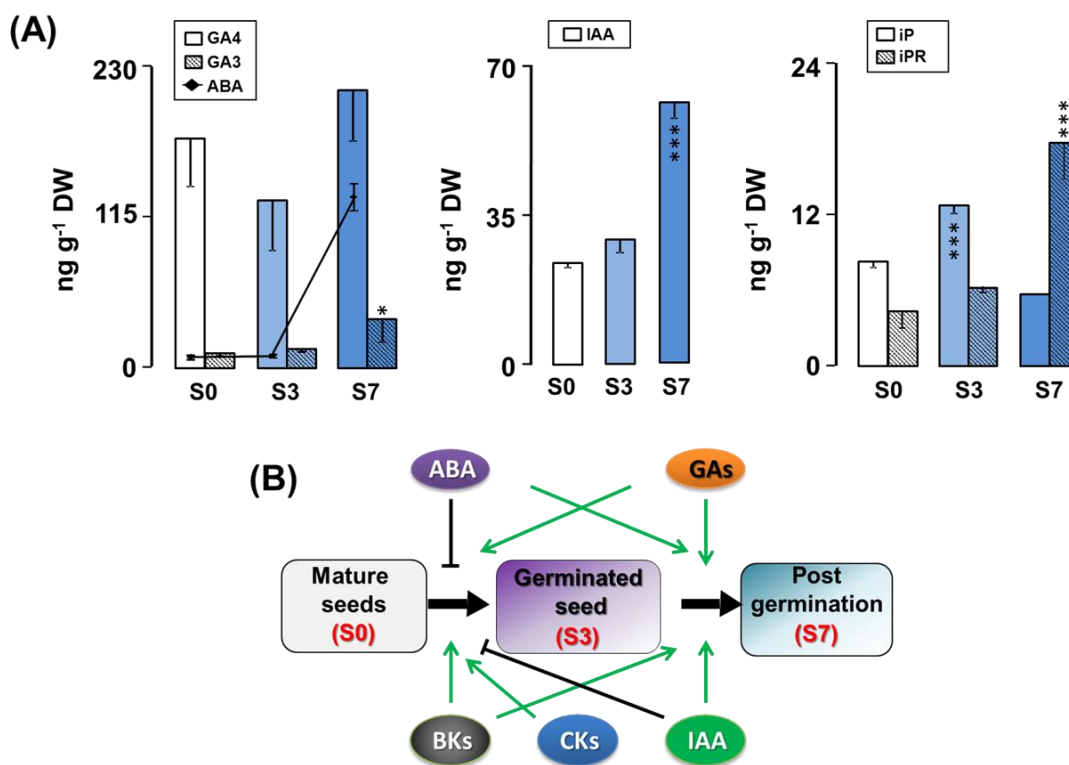


Figure 4.5. Changes in the endogenous concentrations of phytohormones in *Q. ilex* embryo axis and seedling during germination and post germination stages and cross-talk among them. (A) The levels of abscisic acid (ABA), gibberellins (GA3, GA4), auxin (IAA) and cytokinins (iP, iPR,) were determined by using GC-MS/MS. Data correspond to the mean \pm SE of measurements made on 20-50 seeds per sampling time, grouped in three pools. Statistical significances were determined in a one-way ANOVA Tukey test. The *P* values are given as * = significant at *P* < 0.05 level and *** = significant at *P* < 0.001 level. **(B).** Phytohormone interaction during germination and seedling development.

The ABA content in mature *Q. ilex* embryo axis (6.4 ± 1.0 ng g⁻¹ DW) reported here is similar to those found for the embryo axis of *Q. robur* 40 and other recalcitrant seeds 41. However, this value is considerably lower than those reported for orthodox seeds of *Arabidopsis thaliana* (140 ng g⁻¹ DW) 42 or *Solanum lycopersicum* (50 ng g⁻¹ DW) 43. The

desiccation sensitivity and absence of dormancy in *Q. ilex* seeds are probably related to their low ABA levels, emphasizing the role of this phytohormone in preventing premature germination and desiccation tolerance acquisition. No differences were observed during the germination process (S0 to S3) in the *Q. ilex* embryo axis but the ABA amount increased considerably (>20-folds) in the post-germination stage (from S3 to S7). A similar increase has been described for the recalcitrant species *Araucaria angustifolia* embryo,⁴¹ The high levels of ABA in the S7 stage might be related to the development of radicle. Some previous reports indicated a positive role for ABA in promoting root meristem maintenance and root growth in *Arabidopsis*, though the involved molecular mechanisms remain unknown⁴⁴⁻⁴⁵.

Gibberellins (GA) are required to generate sufficient embryo growth potential to rupture the endosperm and testa in the final stages of germination⁴⁶. Active GAs, GA3 and GA4 are present at very low levels in young orthodox embryonic axes and their levels increase to overcome seed dormancy (*i.e.*, Chen *et al.*, Ogawa *et al.*, Farrant *et al.* and Weitbrecht *et al.*⁴⁷⁻⁵⁰). Time-course changes in endogenous contents of GA3 and GA4 has been determined in the *Q. ilex* embryo axis using GC-MS (Fig. 5.6.A) that revealed levels of 10.6 ± 2.9 and 174.8 ± 37.1 ng/g DW, respectively. The GA3 levels are usually undetectable in orthodox non-germinating seed but its accumulation is a prerequisite for dormancy breaking. In the case of *Phellodendron amurense*, germination is accompanied by an increase of up to ~ 60 ng/g DW⁴⁷. In *Q. ilex* embryos the levels of GA3 increased from 10.6 ± 2.9 to 37.0 ± 18.1 ng/g DW (>3-fold increase) during the post germination stage (S3-S7). In other recalcitrant species such as *Avicennia marina* the GA3 levels were higher at the S0 stage (~ 35 ng/g DW) reaching levels of ~ 55 ng/g DW (<1.5-fold increase) at the radicle protrusion stage (Farrant, Berjak *et al.* 1993). GA4 has been proposed as the major endogenous active GA in germinating seeds and shoots⁴⁸. In *Arabidopsis* or *Phellodendron amurense* GA4 is present at about 3 ng/g DW in the mature, non-germinating seed and its levels increased 2- to 7-folds, respectively, at the S3 stage⁴¹. The GA4 concentration in *Q. ilex* embryos was considerably higher than that found in germinated orthodox seeds, what might explain its stability along the course of the experiment.

Other hormones are involved in seed developmental processes, and hormone interactions in relation to seed dormancy and germination have been discussed^{38; 51} (Fig. 4.5.B). The auxin indole-3-acetic acid (IAA) has been implicated in plant growth⁵² and roots development⁵³. The physiological role of IAA during seed maturation is not clear. Cross-talk between IAA and ABA during seed germination has been reported, as IAA enhances ABA-mediated inhibition of seed germination⁵⁴. In the present study, we found that the IAA levels in *Q. ilex* embryo axis and seedling were 23.9 ± 1.4 and 28.9 ± 2.8 ng g⁻¹ DW, at in S0 and S3 respectively, and these levels increased to 61.4 ± 3.8 ng g⁻¹ DW at S7 stage (Fig. 4.5.A). The

levels of IAA in mature seeds was lower than reported to *Avicenia marina* (500 ng g⁻¹ DW)⁵⁵, but IAA increases during post-germination similar to that reported here have been described for several recalcitrant seeds^{41;55}. The time-course change in *Q. ilex* seeds of endogenous IAA contents is matched by the changes in the levels of ABA, supporting the idea that of a crosstalk between IAA and ABA responses⁵⁴ and that both hormones are not involved in the processes occurring during *Q. ilex* seed germination.

We also analyzed the levels of some cytokinins and the brassinolide castasterone (CS) in embryo axis and seedling in different germination and post germination stage of *Q. ilex* seeds. Naturally occurring CKs (cytokinins) are modified forms of the nucleic acid base adenine with plant growth regulating activity. CKs play a role in many aspects of plant growth and development, including cell division, cell enlargement, senescence and differentiation and have also been implicated in favouring the germination and early post-germination events⁵⁶. The CKs iP, Z and DZ are predominant in higher plants. The free bases and their ribosides (iPR, ZR, DZR) are thought to be the biologically active compounds⁵⁷. Among the CKs we measured here (tZ, iP, DZ, iPR, tZR and DZR) only iP and iPR showed statistically significant increases along the studied stages (Fig. 4.5.A). The levels of tZ, tZR, DZ and DZR (data not showed) were about two orders of magnitude lower than described for other recalcitrant seeds⁵⁵ and their increases did not duplicate the basal S0 levels. The iP content in *Q. ilex* seeds/seedlings increased from 8.5±0.5 to 12.6 ±0.6 ng g⁻¹ DW during germination (S0 to S3), and then decreased at S7 stage (Fig. 4.5.A) reaching levels lower than in the S0 stage. Again, the *Q. ilex* seed content of this CK was lower than described in other recalcitrant species, but showed a similar time-course profile^{55;58}. In contrast to iP, iPR showed a continuous increase during germination and post germination stages (from 4.2 ± 1.3 ng g⁻¹ DW in S0 to 17.7 ± 2.9 ng g⁻¹ DW in S7). To the best of our knowledge, this is the first study to report of the time-course increase of iPR in this type of seeds and demonstrates that in the recalcitrant *Q. ilex* seeds the increase of anti-dormancy hormones accompanies the germination process.

Plant brassinosteroid (BR) hormones play a role similar to GA in promoting germination in that it helps to break ABA-induced dormancy and stimulate germination⁵⁹⁻⁶¹. We measured here the amount castasterone (CS), a precursor of brassinolide (BL), the active component of BR, in the *Q. ilex* seed germination and post germination events. In *Q. ilex* seeds CS during germination and post germination showed constant values between 49.7 ± 19.1 ng g⁻¹ DW and 53.5 ± 9.8 ng/g DW (data not shown). These findings indicate that de novo CS synthesis is not operating in the germinating *Q. ilex* seed, a result clearly opposite to the data reported for Arabidopsis, pea, and tomato⁶⁰. However, when the levels of CS in *Q. ilex* seeds at the S0, S3 and S7 stages (about 50 ng/g DW) are compared with the 0.5-2 ng/g DW reported in

Arabidopsis⁶² or pea seeds⁶⁰ it can be assumed that the levels of CS in *Q. ilex* seeds are high enough to finish the germination process.

Though several reports addressed the effects of phytohormones on growth and development, the studies on changes in the endogenous levels of phytohormones during seed germination in recalcitrant seeds, or centered in their crosstalk, are less documented. We report here data of the levels of some of these phytohormones in a recalcitrant seed. As a whole, the phytohormone data reported here support the idea of a non-totally incomplete level of maturation of *Q. ilex* seeds, where the low ABA levels are unable to prevent premature germination but where, also, the biosynthesis of GA, auxin and CK are needed for germination completion.

4.3.5. Isolation of *Q. ilex* gene fragment sequences for transcriptional analysis

We studied here a set of twelve genes involved in germination (Table 4.3). These genes were chosen after an extensive literature search reflecting multiple sources of experimental evidence (genetic, biochemical or high-throughput expression data) that definitely established their participation in regulating the plant response mechanisms to situations in which water availability is limited as is the case of a mature orthodox seed. The selection included members of the two major groups of drought-stress regulated genes⁶³, encoding for both regulatory or functional proteins. DHN3 and GOLS are ABA-regulated, drought-responsive proteins that protect plant proteins and membranes from the loss of water and help to maintain the cellular integrity during seed desiccation. *Skp1*, *Pp2c*, *Sdir1* and *Ocp3* genes code for regulatory elements involved in the ABA signalling pathway, the key responsible for the main defence responses to drought and salt stresses and the maintaining of seed dormancy (Cutler *et al.*⁶⁴ and ref. herein). Four genes coding for metabolic enzymes (FDH, GAPDH, RBLC and NDH6) were also included, since metabolism is expected to be widely affected during germination. The phase II metabolizing enzyme GST is considered a negative component of the ABA-mediated signal transduction pathway⁶⁵ in addition to its intrinsic defensive role against stresses during dormancy. Finally, superoxide dismutase (SOD) protect seed against ROS produced during seed desiccation, germination, and ageing, which may lead to cellular damage and seed deterioration.

One major limitation of transcripts quantification approaches in non-model organisms like *Quercus ilex* is the almost absolute lack of information regarding genic sequences in publicly available databases. Hence, as part of designing the qRT-PCR assay, we sequenced coding region segments of the selected gene loci from *Q. ilex*. When the sequence of the *Q. ilex* gene was not available in a public database, cross-species amplification experiments were

performed with gene-specific primer pairs designed over published orthologous sequences (Table 4.1). Partial coding sequences corresponding to eight *Q. ilex* genes (*Fdh*, *Gst*, *GolS*, *Ndh6*, *Ppc2*, *Ocp3*, *Sdir1* and *Skp1*) were isolated. After being trimmed and verified, the identity of each transcript was assigned on the basis of sequence similarity to proteins with known functions in GenBank using the NCBI-tBLASTx tool, with an e -value $\leq 10^{-20}$ and the CDS were deposited at the GenBank database (Table 4.2). To exactly complement the desired template location, primers for the absolute quantification of selected genes transcript levels by real-time RT-PCR (qRT-PCR) were designed (Table 4.2).

Table 4.3. Genes studied in this work and the functions of the proteins that they encode.

Functional category			
Gene name^a	Gene annotation (synonyms)	Protein Function	References
Desiccation protective proteins			
Dhn3	Dehydrin LEA isoform 3	Intrinsically disordered protein, potent chaperone, protects plant proteins and membranes from the loss of water; ABA-regulated.	Sunderlikova <i>et al.</i> ⁶⁶
GolS	Galactinol synthase (Galacturonosyl transferase-Like 4; <i>GATL</i>)	Key enzyme in the biosynthesis of raffinose, an osmolyte that protects the cellular integrity during desiccation; ABA-regulated.	Li <i>et al.</i> ⁶⁷
ABA signaling pathway			
Ocp3	Overexpressor of cationic peroxidase 3	Negative regulator of ABA signaling transduction.	Ramírez <i>et al.</i> ⁸⁸
Sdir1	Salt-And Drought-Induced Ring Finger 1	RING-type E3 ubiquitin ligase; modulates the ABA signaling during drought stress.	Li <i>et al.</i> ⁶⁸
Skp1	S-Phase Kinase-Associated Protein	Involved in ABA signalling by SCF-mediated protein degradation.	Li <i>et al.</i> ⁶⁹
Pp2c	Protein Phosphatase Type 2C	Negative regulator of ABA signaling by blocking downstream protein kinase SnRK2.	Sun <i>et al.</i> ⁷⁰
Metabolism			
Ndh6	NADH dehydrogenase subunit 6	Involved in dissipating energy and maintaining ATP supply under conditions of low CO ₂ in response to stress.	Abdeen <i>et al.</i> ⁷¹
GapdH	Glyceraldehyde 3-Phosphate Dehydrogenase (<i>GAPA</i> ; <i>GAPA 1</i>)	Catalyzes the oxidative phosphorylation of glyceraldehyde-3-phosphate.	Aranjuelo <i>et al.</i> ⁷²
Fdh	Formate dehydrogenase	Catalyzes the oxidation of formate into CO ₂ in the presence of NAD ⁺ .	Alekseeva <i>et al.</i> ⁷³
RbcL	Ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCo)	Enzyme fixing CO ₂ during photosynthesis.	Demirevska <i>et al.</i> ⁷⁴
Redox homeostasis/detoxification			
Sod1	Superoxide dismutase 1	Catalyzes the dismutation of superoxide (O ₂ ⁻) into oxygen and hydrogen peroxide	Roach <i>et al.</i> ⁹⁶
Gst	Glutathione S-Transferase isoform 2	Phase II metabolizing enzyme; negative component of stress-mediated signal transduction pathways (including ABA signaling) in adaptive responses to drought and salt stresses.	Chen <i>et al.</i> ⁶⁵

^aGene symbols are according to the NCBI Gene database.

4.3.6. Transcriptional profiles changes during *Q. ilex* seed germination and early seedling growth

Studies in orthodox seeds demonstrate that developmental phases during subsequent germination and early seedling are characterized by temporal patterns of gene expression. We determined by qRT-PCR the absolute transcript number for each gene in our panel at four different developmental stages of the *Q. ilex* seeds: S0 (mature acorn); S2 (early germination stage); S3 (late germination stage) and S7 (post-germination, early seedling period). We also quantified the transcript copy numbers of these 12 genes in mature fully expanded leaves of one-year old holm oak plants in order to clarify the metabolic state and drought adaptation of mature non-imbibed acorns (Fig. 4.6).

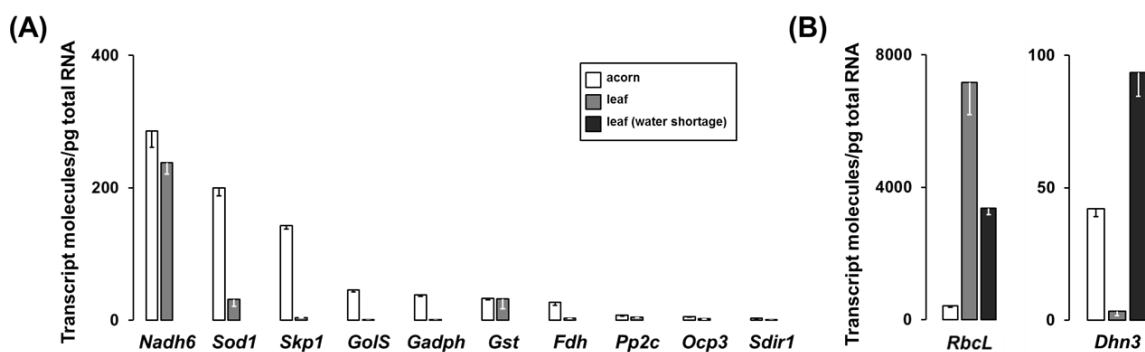


Figure 4.6. Absolute quantitation of transcript molecules in embryo axis tissue isolated from non-imbibed acorns (S0) and in mature, fully expanded leaves of one-year old *Q. ilex* plants. (A) Transcript data are the means \pm SD of transcript molecules/pg of total RNA from three biological replicates made in quadruplicate. Each biological replicate was a pool generated by mixing equal amounts of homogenized tissue from 10-30 embryo axis/leaves. **(B).** The same comparison, but including the *RbcL* and *Dhn3* transcript amounts measured in leaf tissues from water-stressed plants. One-year-old holm oak plants were water withheld for 28-days, which caused 70% decreases in soil moisture levels, an statistically significant ($P < 0.001$) reduction in the xylem water potential, which dropped from -0.72 ± 0.06 MPa to -2.2 ± 0.10 MPa after 28 days, in the stomatal conductance (66.2 ± 2.4 vs. 129.2 ± 4.8 mmol min⁻¹ s⁻¹) and diminished the photosynthetic efficiency by 88% at the end of the experiment. Statistical significances were determined with Student's t-test. Differences between acorns and well-watered leaves (a) and between well-watered and water-stressed leaves (b) transcript counts were all statistically significant at least at the $P < 0.05$ level.

With the exception of *Gst* and *RbcL*, all the studied genes displayed higher amounts of transcripts in the holm oak acorns than in the leaves of the one-year-old plants. Transcript accumulation has been described for low-hydrated mature seeds (the orthodox seeds), especially of mRNA species associated with proteins involved in the acquisition of desiccation tolerance (*i.e.* dehydrins) ⁷⁵ and other biological processes (*i.e.* ribosomal proteins and translation initiation factors; proteases and/or peptidases; proteins associated with energy metabolism, DNA repair and lipid degradation) ⁷⁶. Transcript accumulation protects the seed during desiccation, providing the mRNAs needed during the early stages of germination, where

no transcription is possible⁷⁵. Such mRNA accumulation would not be expected in a recalcitrant seed where the moisture grade is over 38 % in the mature stage and where there is not a complete cessation of metabolism⁷⁷. We show here, however, that the recalcitrant holm oak acorns also seemed to be adapted to cope with desiccation and to initiate the germination process by accumulating high levels of mRNAs. In these line, and despite of the high level of *RbcL* transcript counts in *Q. ilex* seeds, data in Fig. 4.6.B indicate that much more higher levels are needed for a tissue being functional in CO₂ fixation such a mature leaf. Data in Fig. 4.6.B also suggest a relative desiccation adaptation in the acorn, since a reduction in the *RbcL* and an increase in *Dhn3* mRNA molecules accompanied the water-stress response in the plant. These results corroborated previous findings regarding the accumulation of dehydrins⁷⁸ in *Quercus spp.*, and expanded the list of processes involved in the recalcitrant *Q. ilex* mature seed protection at the transcripts level.

Changes in the transcriptional profiles of genes coding for desiccation protective proteins

Figure 4.7 summarizes the transcripts level of *Dhn3* and *Gols* mRNA species corresponding to each studied developmental stage, and it shows different expression patterns along the experimental period. DHN3 and GOLS are ABA-regulated, drought-responsive proteins that protect plant proteins and membranes from the loss of water and help to maintain the cellular integrity during seed desiccation. The dehydrin DHN3 is a Group II Late Embryogenesis Abundant (LEA) family member, whose transcripts and proteins accumulate in vegetative tissues upon various stress factors that cause cell dehydration⁷⁹. These proteins show an intrinsic structural disorder that makes them resistant to crystallization and denaturalization, and that has been proposed to be the cause of dehydrins acting as potent chaperones under the conditions of dehydration stress⁶⁶. *Dhn3* transcripts and protein abundance dropped dramatically and constantly (>120 fold) from the S0 stage till the end of the experiment (S6), moving in parallel but opposite directions to the relative water content (RWC) in germinating seeds (Fig. 4.7.A, B).

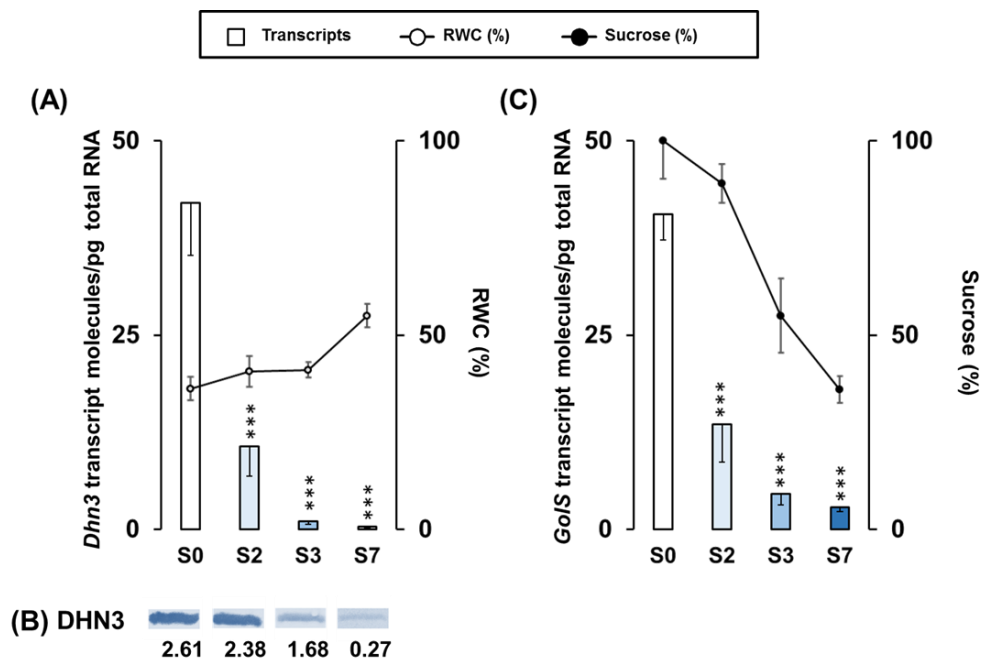


Figure 4.7. (A) Absolute quantitation of *Dhn3* transcript molecules in embryo axis tissue isolated from non-imbibed seed (S0) or after 10 h (S2), 24h (S3) and 216h (S7) of germination, compared with the relative water content (RWC) in germinating seeds. Transcript data are the means \pm SD of transcript molecules/pg of total RNA from three biological replicates made in quadruplicate. Each biological replicate was a pool generated by mixing equal amounts of homogenized tissue from 10-30 embryo axis of the same sampling time. Statistical significances were determined by a one-way ANOVA. Differences between S0 samples and each other were all statistically significant (***: $P < 0.001$). RWC data, expressed as percentage of lost weight relative to fresh weight, correspond to those in Fig. 4.4 and are included for correlation with *Dhn3* transcript amount. (B). Western blotting of DHN3 protein in *Q. ilex* seeds samples. Proteins were extracted from the same pools used in the transcriptional analysis. Numbers indicate the arbitrary Western blotting signal intensities normalized to the total protein contents, using Stain-free technology for total protein quantification. (C). Absolute quantitation of *GolS* transcript amounts in the same samples S0, S2, S3 and S7 and conditions describes in (A), compared with the sucrose content in germinating seeds. Sucrose data correspond to those in Fig. 4.5 and included for correlation with *GolS* transcript amount.

A similar profile was obtained for *GolS* mRNA (Fig. 4.7.C). During orthodox seeds germination, and coinciding with the loss of tissue desiccation tolerance, RFOs (raffinose family oligosaccharides, galactinol, raffinose, and stachyose) are quickly degraded and their synthesis stopped in the embryonic axis by directing accumulated sucrose to supply metabolic pathways. The disappearance of RFOs has been reported to be caused by the decrease in GOLS activity (cited in Lahuta *et al.*⁸⁰), the key unique enzyme in the RFOs synthesis. We show here that the *GolS* transcripts, accumulated at high levels (>40 transcript/pg total RNA) in the mature *Q. ilex* embryo axis (S0), dramatically dropped (>14-fold decrease) when imbibition started (S2 and ahead), moving in parallel with the relative sucrose content in germinating seeds (Fig. 4.7.C). Data suggest that *GolS* transcript accumulation in mature holm oak seeds was probably related to sucrose accumulation and both seems to indicate that in the *Q. ilex* recalcitrant seed there is also an, at least partial, adaptation to desiccation.

Changes in the transcriptional profiles of genes coding proteins regulating the ABA signaling response

Plants utilize hormones to integrate endogenous and exogenous signals. *Pp2c*, *Skp1*, *Sdir1* and *Ocp3* genes code for regulatory elements involved in the ABA signaling pathway, the key responsible for the main defence responses to drought and salt stresses and the maintaining of seed dormancy (Cutler *et al.* ⁶⁴ and ref. herein). Perception of ABA is mediated by a suite of receptors named pyrabactin resistance 1 (PYR1)/PYR1-like (PYL)/Regulatory component of ABA receptor (RCAR) ⁸¹. The binding of ABA to PYR/PYL/RCAR forms a complex that inhibits the activity of protein phosphatase PP2C, a negative regulator of ABA signaling through repression of SnRK2s, the positive regulators of downstream targets. The ABA activated protein-kinases SnRKs (Fig. 4.8.B) are then able to phosphorylate and activate downstream transcription factors including the bZIP-type transcription factors ABFs, ABI5 and ABI3 that are key to regulate the expression of ABA responsive genes in seeds ⁸².

The PP2C family is the largest P-protein phosphatase family in Arabidopsis ⁸³. *Pp2c* is a relatively abundant transcript (~ 8 molecules/pg total RNA) in the embryonic axes of unimbibed seeds, whereas the level of expression in leaves was almost undetectable (Fig. 4.6). The *Pp2c* mRNA was found to maintain constant counts in the S0-S3 stages, and significantly decreased after radicle emergence (S7 stage) (Fig. 4.8.A), when the ABA level start to raise (Fig. 4.5). As a negative inhibitor of ABA, downregulation of *Pp2c* gene and the likely subsequent diminished level of PP2C protein would facilitate the role of ABA at early stages of seedling development.

The ubiquitin–proteasome pathway has been implicated in ABA signaling as mutations in subunits of the 26S proteasome were shown to result in changes in hormone sensitivity. S-phase kinase-associated protein 1 (SKP1) is a component of a *Skp1*-Cullin1-F-box (SCF) complex E3 ligase and negatively regulates ABA signaling through promoting degradation of specific kinases that phosphorylate and activate ABI5 protein ⁶⁹. Mutants *skp1*-like of Arabidopsis exhibit reduced ABA sensitivity ⁶⁹, and overexpression of the SKP1 in tobacco has been reported to cause delayed seed germination ⁸⁴. We quantitated a high amount of *Skp1* transcript molecules (> 140 molecules/pg total RNA, Fig. 4.6.A) in the acorns of *Q. ilex*, which diminished when the seed enter in the germination process (from S0 to S3 stages) and started to rise (Fig. 4.8.A) with the rise in the phytohormone levels (from S3 to S7, Fig. 4.5). These data suggest that SKP1 is required to maintain low levels of ABI5 to ensure seedling establishment, as has been reported for other SCF-E3 ligases ^{69; 85}. The results for absolute transcripts counts displayed in Fig. 4.8 are in agreement with previous reports showing that *Skp1* and *Pp2c* genes are down-regulated in the absence of ABA.

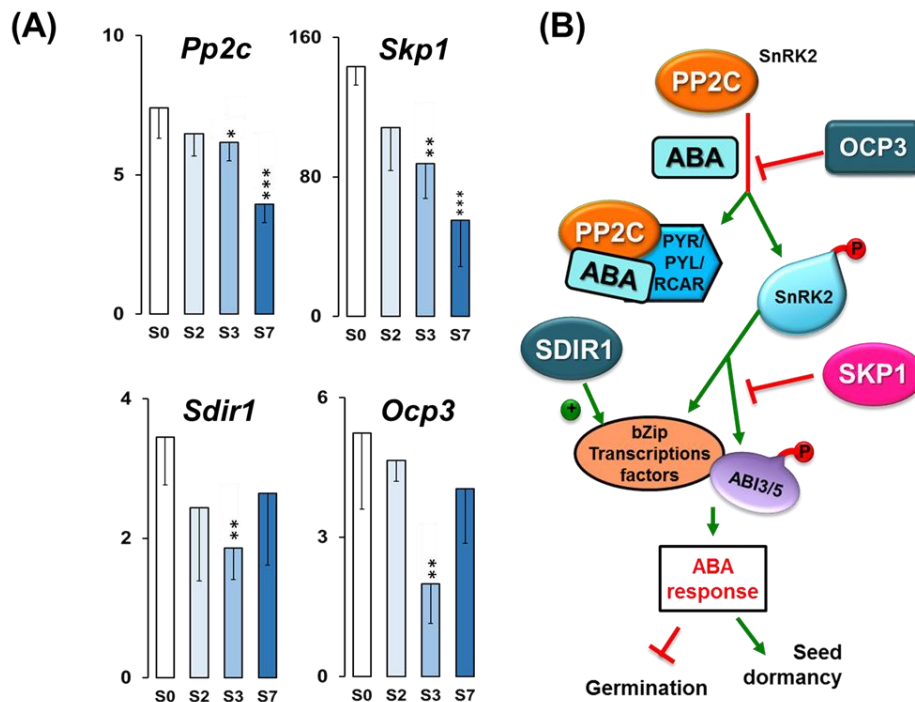


Figure 4.8. (A) Absolute quantitation of *Pp2c*, *Skp1*, *Sdir1* and *Ocp3* transcript molecules in embryo axis tissue isolated from non-imbibed seed (S0) or after 10 h (S2), 24h (S3) and 216h (S7) of germination. Transcript data are the means \pm SD of transcript molecules/ μ g of total RNA from three biological replicates made in quadruplicate. Each biological replicate was a pool generated by mixing equal amounts of homogenized tissue from 10-30 embryo axis of the same sampling time. Statistical significances were determined by a one-way ANOVA. Differences between S0 samples and each other were statistically significant at the level $P < 0.05$ (*); $P < 0.01$ (**); $P < 0.001$ (***). **(B)** A simplified working model of the ABA signalling pathway controlling the germination process indicating the participation of the *Pp2c*, *Skp1*, *Sdir1* and *Ocp3* gene products.

In addition to SKP1, many other E3 ligases have been found to be involved in ABA responses. The RING type E3s Salt and Drought Induced RING Finger 1 (SDIR1) acts as an active RING-type E3 ubiquitin ligase, upstream of ABA-responsive transcription factors, in a feedback mechanism to enhance the ABA-signal⁸⁶. *Sdir1* has been reported to be expressed in all tissues of *Arabidopsis* and is upregulated by drought and salt stress, but not by ABA⁸⁷. Data in Fig. 4.6 shown that the *Sdir1* transcripts are >8-fold more abundant in the acorn than in the leaves of *Q. ilex*. Acorn *Sdir1* transcript numbers decreased ~2-fold during the imbibition period (from 3.5 at S0 vs 1.8 at S3 molecules/ μ g total RNA, Fig. 4.8.A, increasing later in parallel to the ABA level increase (Fig. 4.5.A).

OCP3 is a member of the homeobox transcription factor family, and is considered a negative regulator of the early response of the plant to drought stress⁸⁸, as the *Ocp3* loss of function yields a hyper-susceptibility to the ABA hormone in *Arabidopsis*⁸⁸. The number of *Ocp3* mRNAs diminished 2-fold during the imbibition period (from 5.2 at S0 vs 2.3 at S3 molecules/ μ g total RNA, Fig. 4.8.A), and increased thereafter.

Changes in the transcriptional profiles of genes coding metabolic enzymes

In the orthodox seed, germination represents a switch from a period of quiescence to an energetically demanding state that necessitates ATP and hence, a rapid increase in respiration rate accompanies the early stages of germination after imbibition (Logan *et al.*⁸⁹ and ref. herein). By contrast, metabolism in recalcitrant seeds is active also after shedding, though their relative dehydration, the presence of the pericarp and testa, acting as barriers to gas exchange and the absence of fully functional plastids lead to a certain metabolic imbalance⁸. Four genes (*GapdH*, *Ndh6*, *RbcL* and *Fdh*) coding for metabolic enzymes were also included in this study to analyse putative changes in metabolism during germination of *Q. ilex* acorns. Figure 4.9 shows the variation of the transcripts amounts for these four genes along the germination and early post-germination stages.

GAPDH is a key enzyme for energy metabolism and the production of ATP and pyruvate through anaerobic glycolysis in the cytoplasm. Recent studies have shown that GAPDH has multiple functions independent of its role in energy metabolism and increased GAPDH gene expression and enzymatic function is associated with cell proliferation, transcriptional and posttranscriptional gene regulation, vesicular transport, receptor mediated cell signaling chromatin structure and the maintenance of DNA integrity⁹⁰⁻⁹¹. *GapdH* transcripts and protein remained almost constant during germination and early post-germination stages of *Q. ilex* seeds (Fig. 4.9.A, E). The levels of *GapdH* mRNA found in fully expanded leaves (Fig. 4.6) were lower than in the seed, which suggested a special necessity of GAPDH for the surviving and germination of *Q. ilex* seeds.

Upon imbibition, the embryo releases of a variety of hydrolytic enzymes for the degradation of storage compounds, what will supply energy and carbon sources to the developing embryo for seedling establishment⁹². We observed a decrease of sucrose amounts in stages S3 to S7 of *Q. ilex* seeds germination, associated to the increase of soluble sugars Glc and Fru (Fig. 4.4.B) that can be utilized early in germination as an immediate energy source³⁷ in respiration. The mitochondrial gene that codifies subunit 6 of NADH (*Ndh6* or *NdhF*) provides instructions for making a protein, NADH dehydrogenase 6, which is part of Complex I in the respiratory chain. Data in Fig. 4.9.C shows that the *Ndh6* transcript numbers peaked early during acorns germination, coinciding with the rupture of the testa, and dropped later to the levels found in the non-germinated seed as it reached the stage when both the coleoptile and radicle had clearly elongated. Similar ordered increase in the abundance of *Ndh6* and other transcripts encoding mitochondrial proteins has been reported for *Arabidopsis* during the maturation of the mitochondria that occurs after imbibition⁹³.

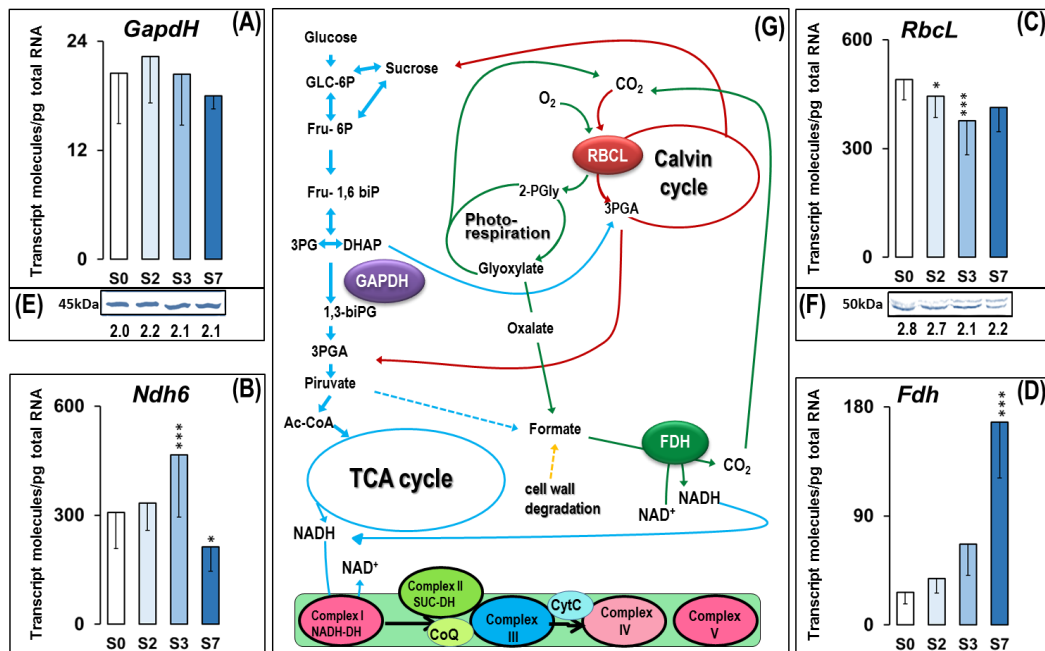


Figure 4.9. (A, B, C, D) Absolute quantitation of *GapdH*, *Ndh6*, *RbcL* and *Fdh* transcript molecules in embryo axis tissue isolated from non-imbibed seed (S0) or after 10 h (S2), 24h (S3) and 9d (S7) of germination. Transcript data are the means \pm SD of transcript molecules/pg of total RNA from three biological replicates made in quadruplicate. Each biological replicate was a pool generated by mixing equal amounts of homogenized tissue from 10-30 embryo axis of the same sampling time. Statistical significances were determined by a one-way ANOVA. Differences between S0 samples and each other were statistically significant at the level $P < 0.05$ (*); $P < 0.01$ (**); $P < 0.001$ (***). (E, F) Western blotting of GAPDH and RBCL proteins in *Q. ilex* seeds samples. Proteins were extracted from the same pools used in the transcriptional analysis. Numbers indicate the arbitrary Western blotting signal intensities normalized to the total protein contents, using Stain-free technology for total protein quantification. (G) A simplified working model of some metabolic pathways indicating the participation of the *GapdH*, *Ndh6*, *RbcL* and *Fdh* gene products.

Ribulose 1,5-bisphosphate carboxylase/oxygenase (RuBisCO) is a key enzyme in photosynthesis and the most abundant leaf protein. It catalyses two competing reactions, CO₂ fixation in photosynthesis (Calvin cycle) and the production of 2-phosphoglycolate in the photorespiratory pathway. In higher plants, RuBisCO is composed of eight small subunits, encoded by a nuclear multigene family (*RbcS*) and eight large subunits, encoded by a single gene (*RbcL*) in the chloroplast genome. Transcription of *RbcL* has been shown to be repressed by soluble sugars⁹⁴, what may explain the decrease in the *RbcL* transcript and protein amount during the *Q. ilex* germination (Fig. 4.9.C, F). In *Q. ilex* seeds *RbcL* is a transcript highly abundant (~ 400 molecules/ng total RNA). However, the *RbcL* transcript levels are ~ 20 -fold higher in leaf (> 7000 molecules/ng total RNA), suggesting the presence of a non-fully functional photosynthetic machinery in the acorn (Fig. 4.6).

Plants contain small, metabolically active pools of formate. Formate is generally considered to come from the photorespiratory pathway through a nonenzymatic, H₂O₂-dependent decarboxylation of glyoxylate. Formate may also arise in formate-producing fermentation pathways and from cell wall degradation (Fig. 4.9.G). Formate enters folate-mediated 1-C metabolism leading to serine synthesis via a reduction initiated by 10-formyl-

tetrahydrofolate synthetase or may be oxidized to CO₂, a reaction catalyzed by formate dehydrogenase (FDH) ⁹⁵. Li *et al.* ⁹⁵ reported that formate can be toxic to Arabidopsis seedlings, as it causes inhibition in the water oxidation reaction as well as in electron transfer in photosystem II. Although it is still not very clear how and why formate is formed in plants, it can be inferred that excess formate has to be broken down and maintained at a level below that at which it becomes toxic. FDH is a soluble mitochondrial enzyme capable of oxidizing formate into CO₂ in the presence of NAD⁺. The *Fdh* mRNA levels (Fig. 4.9.D) showed a constant increase along the different developmental stages studied here, becoming an abundant transcript, with more than 150 molecules/pg total RNA, at the end of the experiment. However, in the mature leaf the amount of *Fdh* transcripts was 8-fold lower (Fig. 4.6), suggesting that formate overproduction is caused by an stress situation in the early seedling stage. FDH is one of the enzymes whose gene is expressed under the action of ABA ⁷³. The increase of endogenous concentrations of this phytohormone in *Q. ilex* seedling during post germination stages (Fig. 4.5) might explain this dramatic increase in *Fdh* mRNA.

Changes in the transcriptional profiles of genes coding antioxidative enzymes

Reactivation of metabolism in germinating seeds provides an important source of reactive oxygen species (ROS). Excessive production of ROS causes oxidative damage to cell macromolecules. However, ROS play also normal physiological roles in cells, acting as signaling molecules in many processes, including seed germination. In orthodox seeds, elevated rates of ROS production upon seed imbibition have been suggested to be involved in cell wall loosening and in defending the emerging seedling against pathogens (Reviewed in Roach *et al.* ⁹⁶). Recalcitrant seeds are desiccation-sensitive and maintain relative high water contents and metabolic rates from seed maturation until germination. Desiccation disrupts their metabolism, leading to the accumulation of potentially harmful reactive oxygen species (ROS) ⁸. We investigated the contribution of two antioxidative enzymes, superoxide dismutase 1 (SOD1) and glutathione S-transferase (GST), to the protection of *Q. ilex* seeds and analysed the changes in their transcriptional profiles along the germination process (Fig. 4.10).

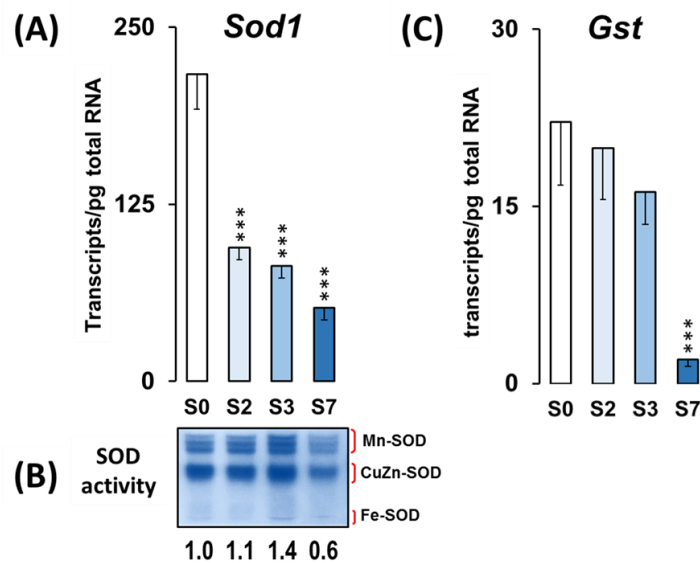


Figure 4. 10. (A, C). Absolute quantitation of *Sod1* and *Gst* transcript molecules in embryo axis tissue isolated from non-imbibed seed (S0) or after 10h (S2), 24h (S3) and 216h (S7) of germination. Transcript data are the means \pm SD of transcript molecules/pg of total RNA from three biological replicates made in quadruplicate. Each biological replicate was a pool generated by mixing equal amounts of homogenized tissue from 10-30 embryo axis of the same sampling time. Statistical significances were determined by a one-way ANOVA (***, $P < 0.001$). **(B)** SOD activity gel assay showing the SOD protein profiles on native PAGE of different *Q. ilex* seeds germination stages. Proteins were extracted from the same pools used in the transcriptional analysis. A 20 μ g aliquots of crude soluble proteins were loaded onto a 10% acrylamide gel. Numbers indicate the arbitrary signal intensities normalized to the total protein contents, using Stain-free technology for total protein quantification.

SODs are a family of metallo-enzymes whose presence has been demonstrated in both the cytosol and different cell organelles. Three types of SODs have been classified on the basis of the metal present at the catalytic site: Cu/Zn-SOD (SOD1, located in the cytosol and chloroplasts), Mn-SOD (in the mitochondrial matrix and peroxisomes), and Fe-SOD (in chloroplasts)⁹⁷. By using the classical gel method for assaying SOD activity in gel²⁵ and the SOD-inhibitors H_2O_2 and KCN, we identified three MnSOD polypeptides (resistant to H_2O_2 and KCN), a single Cu/ZnSOD polypeptide (sensitive to H_2O_2 and KCN) and a single FeSOD polypeptide (resistant to KCN and sensitive to H_2O_2) in *Q. ilex* seeds (Fig. 4.10.B). Bands of the five isoforms were present, though with different intensity, in all the studies stage. The SOD enzyme profiles of all the five isoforms increased slightly during imbibition to decrease later in the post-germination stage. The Cu/ZnSOD activity represented at least the 50% of total SOD activity in each sample, whereas the FeSOD was barely detectable. In agreement with the drop in SOD activity observed, the *Sod1* transcriptional profile in Fig. 4.10.A shows a significant decrease in the amount of *Sod1* mRNA molecules along the germination process. The decrease in the number of *Sod1* transcripts initiates at S2 stage and might be the cause of the drop in the SOD activity detected later (S7). These results suggest that antioxidative defense mechanisms have been activated during this recalcitrant seed maturation, leading to the accumulation of SOD to adapt the acorn to support a certain grade of desiccation. Indeed, the amounts of *Sod1*

mRNA molecules present in the un-stressed leaf was >6-fold lower than in the un-imbibed *Q. ilex* seeds (Fig. 4.6).

The plant glutathione S-transferases (GSTs) are a super-family of proteins of divergent sequence but conserved structure, which catalyze the conjugation of electrophilic xenobiotic substrates with the tripeptide glutathione (GSH) and which are selectively stress-inducible. Some GSTs can also act as glutathione peroxidases, protecting cells from oxygen toxicity⁹⁸. Our results revealed that the abundance of *Gst* transcripts decreases during germination in *Q. ilex* acorn (Fig. 4.10.C). Seed germination has been described to improve after over-expression of *Gst* in tobacco and it has been suggested that the higher amount of oxidized glutathione (GSSG) in dry embryos compared to germinating seeds could contribute to prevent the germination process, since in dormant wheat embryos GSSG (or the lack of GSH) seemed to block protein synthesis⁹⁹.

4.4. Concluding remarks

The aim of this study was to investigate the underlying biochemistry and metabolic status before and after the germination process in the *Quercus ilex* acorn. We found that mature *Q. ilex* seeds showed some of the intracellular physical characteristics that become modified with the acquisition of desiccation tolerance in orthodox seed that included (i) accumulation of non-reducing carbohydrates (sucrose) and insoluble proteins (DHN3) that contribute to the intracellular vitrified state in seeds; and (ii) accumulation of transcripts involved in the synthesis of certain osmoregulator raffinose series oligosaccharides (*GolS*), the anti-oxidative defence (*Sod1*, *Gst*) and the preparation for the development of an adult plant (*RbcL*). But the holm oak mature acorns share with other recalcitrant seeds the ability to maintain a partially active metabolism, with high level of glycolytic (*GapdH*) and mitochondrial respiratory enzymes (*Ndh6*) and the absence of ABA (*Pp2c*, *Skp1*, *Sdir1*, *Ocp3*). However, imbibition increased the respiratory rate, paralleling the soluble carbohydrate sugar increase, and indicating that mitochondria resulted affected during acorn maturation. As a byproduct of photorespiration, of cell wall degradation during the germination process and of the excess of glycolytic intermediates, formate is generated and FDH synthesis increased, regulated by increased levels of ABA, to detoxify formate into CO₂.

4.5. References

1. Bonner FTV, John A. (1987) Seed Biology and Technology of *Quercus*. In Gen. Tech. Rep. SO-66. New Orleans, LA: U.S. Dept of Agriculture, Forest Service, Southern Forest Experiment Station
2. Caliskan S (2014) Germination and seedling growth of holm oak (*Quercus ilex* L.): effects of provenance, temperature, and radicle pruning. *iForest - Biogeosciences and Forestry* 7: 103-109
3. Connor KF, Sowa S (2003) Effects of desiccation on the physiology and biochemistry of *Quercus alba* acorns. *Tree Physiology* 23: 1147-1152
4. Finch-Savage WE, Blake PS (1994) Indeterminate development in desiccation-sensitive seeds of *Quercus robur* L. *Seed Science Research* 4: 127-133
5. Finch-Savage WE, Blake PS, Clay HA (1996) Desiccation stress in recalcitrant *Quercus robur* L. seeds results in lipid peroxidation and increased synthesis of jasmonates and abscisic acid. *Journal of Experimental Botany* 47: 661-667
6. Goodman RC, Jacobs DF, Karrfalt RF (2005) Evaluating desiccation sensitivity of *Quercus rubra* acorns using X-ray image analysis. *Can. J. For. Res.* 35: 2823-2831
7. Liu Y, Liu G, Li Q, Liu Y, Hou L, Li G (2012) Influence of pericarp, cotyledon and inhibitory substances on sharp tooth oak (*Quercus aliena* var. *acuteserrata*) germination. *PLoS One* 7: e47682
8. Pasquini S, Mizzau M, Petrusa E, Braidot E, Patui S, Gorian F, Lambardi M, Vianello A (2012) Seed storage in polyethylene bags of a recalcitrant species (*Quercus ilex*): analysis of some bio-energetic and oxidative parameters. *Acta Physiologiae Plantarum*: 1-12
9. Fait A, Angelovici R, Less H, Ohad I, Urbanczyk-Wochniak E, Fernie AR, Galili G (2006) Arabidopsis seed development and germination is associated with temporally distinct metabolic switches. *Plant Physiology* 142: 839-854
10. Kanno Y, Jikumaru Y, Hanada A, Nambara E, Abrams SR, Kamiya Y, Seo M (2010) Comprehensive hormone profiling in developing Arabidopsis seeds: examination of the site of ABA biosynthesis, ABA transport and hormone interactions. *Plant and Cell Physiology* 51: 1988-2001
11. An Y-Q, Lin L (2011) Transcriptional regulatory programs underlying barley germination and regulatory functions of gibberellin and abscisic acid. *BMC Plant Biology* 11: 105
12. Gómez-González S, Ruiz-Jiménez J, Priego-Capote F, Luque de Castro MaD (2010) Qualitative and Quantitative Sugar Profiling in Olive Fruits, Leaves, and Stems by Gas Chromatography-Tandem Mass Spectrometry (GC-MS/MS) after Ultrasound-Assisted Leaching. *Journal of Agricultural and Food Chemistry* 58: 12292-12299
13. Cerdán-Calero M, Sendra JM, Sentandreu E (2012) Gas chromatography coupled to mass spectrometry analysis of volatiles, sugars, organic acids and aminoacids in Valencia late orange juice and reliability of the automated mass spectral deconvolution and identification system for their automatic identification and quantification. *Journal of Chromatography A* 1241: 84-95
14. Novák O, Hauserová E, Amakorová P, Doležal K, Strnad M (2008) Cytokinin profiling in plant tissues using ultra-performance liquid chromatography-electrospray tandem mass spectrometry. *Phytochemistry* 69: 2214-2224
15. Echevarría-Zomeño S, Abril N, Ruiz-Laguna J, Jorrín-Novo J, Maldonado-Alconada A (2012) Simple, rapid and reliable methods to obtain high quality RNA and genomic DNA from *Quercus ilex* L. leaves suitable for molecular biology studies. *Acta Physiologiae Plantarum* 34: 793-805
16. Fleige S, Pfaffl MW (2006) RNA integrity and the effect on the real-time qRT-PCR performance. *Molecular Aspects of Medicine* 27: 126-139
17. Taylor SC, Mrkusich EM (2014) The state of RT-quantitative PCR: firsthand observations of implementation of minimum information for the publication of quantitative real-time PCR experiments (MIQE). *J Mol Microbiol Biotechnol* 24: 46-52

18. Jurado J, Fuentes-Almagro CA, Prieto-Alamo MJ, Pueyo C (2007) Alternative splicing of c-fos pre-mRNA: contribution of the rates of synthesis and degradation to the copy number of each transcript isoform and detection of a truncated c-Fos immunoreactive species. *BMC Mol Biol* 8: 83-96
19. Prieto-Alamo MJ, Cabrera-Luque JM, Pueyo C (2003) Absolute quantitation of normal and ROS-induced patterns of gene expression: an in vivo real-time PCR study in mice. *Gene Expression Patterns* 11: 23-34
20. Prieto-Álamo M-J, Abril N, Osuna-Jiménez I, Pueyo C (2009) *Solea senegalensis* genes responding to lipopolysaccharide and copper sulphate challenges: Large-scale identification by suppression subtractive hybridization and absolute quantification of transcriptional profiles by real-time RT-PCR. *Aquatic Toxicology* 91: 312-319
21. Wang W, Vignani R, Scali M, Cresti M (2006) A universal and rapid protocol for protein extraction from recalcitrant plant tissues for proteomic analysis. *ELECTROPHORESIS* 27: 2782-2786
22. Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry* 72: 248-254
23. Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227: 680-685
24. Gürtler A, Kunz N, Gomolka M, Hornhardt S, Friedl AA, McDonald K, Kohn JE, Posch A (2013) Stain-Free technology as a normalization tool in Western blot analysis. *Analytical Biochemistry* 433: 105-111
25. Weydert CJ, Cullen JJ (2010) Measurement of superoxide dismutase, catalase and glutathione peroxidase in cultured cells and tissue. *Nature Protocols* 5: 51-66
26. Beauchamp C, Fridovich I (1971) Superoxide dismutase: Improved assays and an assay applicable to acrylamide gels. *Analytical Biochemistry* 44: 276-287
27. Tilki F (2010) Influence of acorn size and storage duration on moisture content, germination and survival of *Quercus petraea* (Mattuschka). *Journal of Environmental Biology* 31: 325-328
28. Echevarria-Zomeno S, Ariza D, Jorge I, Lenz C, Del Campo A, Jorrián JV, Navarro RM (2009) Changes in the protein profile of *Quercus ilex* leaves in response to drought stress and recovery. *Journal of Plant Physiology* 166: 233-245
29. Finch-Savage WE, Leubner-Metzger G (2006) Seed dormancy and the control of germination. *New Phytologist* 171: 501-523
30. Bailly C, Audigier C, Ladonne F, Wagner MH, Coste F, Corbineau F, Côme D (2001) Changes in oligosaccharide content and antioxidant enzyme activities in developing bean seeds as related to acquisition of drying tolerance and seed quality. *Journal of Experimental Botany* 52: 701-708
31. Hoekstra FA, Crowe LM, Crowe JH (1989) Differential desiccation sensitivity of corn and *Pennisetum* pollen linked to their sucrose content. *Plant, Cell and Environment* 15: 83-91
32. Koster KL, Leopold AC (1988) Sugars and desiccation tolerance in seeds. *Plant Physiology* 88: 829-832
33. Nishizawa A, Yabuta Y, Shigeoka S (2008) Galactinol and raffinose constitute a novel function to protect plants from oxidative damage. *Plant Physiology* 147: 1251-1263
34. Black M, Derek Bewley J (2000) *Seed Technology and Its Biological Basis*. Sheffield Academic Press Ltd, California, USA
35. Pammenter NW, Berjak P (2014) Physiology of desiccation-sensitive (recalcitrant) seeds and the implications for cryopreservation. *International Journal of Molecular Sciences* 175: 21-28
36. Rajjou L, Duval M, Gallardo K, Catusse J, Bally J, Job C, Job D (2012) Seed Germination and Vigor. *Annual Review of Plant Biology* 63: 507-533
37. Nkang Ani (2002) Carbohydrate composition during seed development and germination in two subtropical rainforest tree species (*Erythrina caffra* and *Guilfoylia monostylis*). *Journal of Plant Physiology* 159: 473-483

38. Miransari M, Smith DL (2014) Plant hormones and seed germination. *Environmental and Experimental Botany* 99: 110-121
39. Locascio A, Roig-villanova I, Bernardi J, Varotto S (2014) Current perspectives on the hormonal control of seed development in *Arabidopsis* and maize: a focus on auxin. *Frontiers Plant Sciences* 5
40. Prewein C, Endemann M, Reinöhl V, Salaj J, Sunderlikova V, Wilhelm E (2006) Physiological and morphological characteristics during development of pedunculate oak (*Quercus robur* L.) zygotic embryos. *Trees - Structure and Function* 20: 53-60
41. Pieruzzi FP, Dias LLC, Balbuena TS, Santa-Catarina C, Santos ALWd, Floh EIS (2011) Polyamines, IAA and ABA during germination in two recalcitrant seeds: *Araucaria angustifolia* (Gymnosperm) and *Ocotea odorifera* (Angiosperm). *Annals of Botany* 108: 337-345
42. Liu Y, Shi L, Ye N, Liu R, Jia W, Zhang J (2009) Nitric oxide-induced rapid decrease of abscisic acid concentration is required in breaking seed dormancy in *Arabidopsis*. *New Phytologist* 183: 1030-1042
43. Yang R, Yang T, Zhang H, Qi Y, Xing Y, Zhang N, Li R, Weeda S, Ren S, Ouyang B, Guo Y-D (2014) Hormone profiling and transcription analysis reveal a major role of ABA in tomato salt tolerance. *Plant Physiology and Biochemistry* 77: 23-34
44. Zhang H, Han W, De Smet I, Talboys P, Loya R, Hassan A, Rong H, Jürgens G, Paul Knox J, Wang M-H (2010) ABA promotes quiescence of the quiescent centre and suppresses stem cell differentiation in the *Arabidopsis* primary root meristem. *The Plant Journal* 64: 764-774
45. Finkelstein RR, Gampala SSL, Rock CD (2002) Abscisic acid signaling in seeds and seedlings. *The Plant Cell Online* 14: S15-S45
46. Goggin DE, Steadman KJ, Emery RJN, Farrow SC, Benech-Arnold RL, Powles SB (2009) ABA inhibits germination but not dormancy release in mature imbibed seeds of *Lolium rigidum* Gaud. *Journal of Experimental Botany* 60: 3387-3396
47. Chen S-Y, Chien C-T, Baskin JM, Baskin CC (2010) Storage behavior and changes in concentrations of abscisic acid and gibberellins during dormancy break and germination in seeds of *Phellodendron amurense* var. *wilsonii* (Rutaceae). *Tree Physiology* 30: 275-284
48. Ogawa M, Hanada A, Yamauchi Y, Kuwahara A, Kamiya Y, Yamaguchi S (2003) Gibberellin biosynthesis and response during *Arabidopsis* seed germination. *The Plant Cell Online* 15: 1591-1604
49. Farrant JM, Brandt W, Lindsey GG (2007) An overview of mechanisms of desiccation tolerance in selected angiosperm resurrection plants. *Plant Stress* 1: 13
50. Weitbrecht K, Müller K, Leubner-Metzger G (2011) First off the mark: early seed germination. *Journal of Experimental Botany*
51. Finkelstein R, Reeves W, Ariizumi T, Steber C (2008) Molecular aspects of seed dormancy. *Annual Review of Plant Biology* 59: 387-415
52. Santner A, Calderon-Villalobos LI, Estelle M (2009) Plant hormones are versatile chemical regulators of plant growth. *Nat Chem Biol* 5: 301-307
53. Tromas A, Perrot-Rechenmann C (2010) Recent progress in auxin biology. *Comptes Rendus Biologies* 333: 297-306
54. Liu PP, Montgomery TA, Kasschau KD, Nonogaki H, Carrington JC (2007) Repression of AUXIN RESPONSE FACTOR 10 by microRNA160 is critical for seed germination and post-germination stages. *The Plant Journal* 52: 133-146
55. Farrant JM, Berjak P, Cutting JGM, Pammenter NW (1993) The role of plant growth regulators in the development and germination of the desiccation-sensitive (recalcitrant) seeds of *Avicennia marina*. *Seed Science Research* 3: 55-63
56. Bromley JR, Warnes BJ, Newell CA, Thomson JC, James CM, Turnbull CG, Hanke DE (2014) A purine nucleoside phosphorylase in *Solanum tuberosum* L. (potato) with specificity for cytokinins contributes to the duration of tuber endodormancy. *Biochemical Journal* 458: 225-237

57. Werner T, Motyka V, Strnad M, Schmülling T (2001) Regulation of plant growth by cytokinin. *Proceedings of the National Academy of Sciences* 98: 10487-10492
58. Farrant JM, Moore JP (2011) Programming desiccation-tolerance: from plants to seeds to resurrection plants. *Current Opinion in Plant Biology* 14: 340-345
59. Steber CM, McCourt P (2001) A role for brassinosteroids in germination in *Arabidopsis*. *Plant Physiology* 125: 763-769
60. Nomura T, Ueno M, Yamada Y, Takatsuto S, Takeuchi Y, Yokota T (2007) Roles of brassinosteroids and related mRNAs in pea seed growth and germination. *Plant Physiology* 143: 1680-1688
61. Jiang W-B, Lin W-H (2013) Brassinosteroid functions in *Arabidopsis* seed development. *Plant Signal Behav* 8: e25928
62. Nomura T, Sato T, Bishop GJ, Kamiya Y, Takatsuto S, Yokota T (2001) Accumulation of 6-deoxocathasterone and 6-deoxocastasterone in *Arabidopsis*, pea and tomato is suggestive of common rate-limiting steps in brassinosteroid biosynthesis. *Phytochemistry* 57: 171-178
63. dos Reis SP, Lima AM, de Souza CRB (2012) Recent Molecular Advances on Downstream Plant Responses to Abiotic Stress. *International Journal of Molecular Sciences* 13: 8628-8647
64. Cutler SR, Rodriguez PL, Finkelstein RR, Abrams SR (2010) Abscisic acid: emergence of a core signaling network. *Annual Review of Plant Biology* 61: 651-679
65. Chen J-H, Jiang H-W, Hsieh E-J, Chen H-Y, Chien C-T, Hsieh H-L, Lin T-P (2012) Drought and Salt Stress Tolerance of an *Arabidopsis* Glutathione S-Transferase U17 Knockout Mutant Are Attributed to the Combined Effect of Glutathione and Abscisic Acid. *Plant Physiology* 158: 340-351
66. Eriksson SK, Harryson P (2011) *Dehydrins: molecular biology, structure and function*. Springer
67. Li X, Zhuo J, Jing Y, Liu X, Wang X (2011) Expression of a GALACTINOL SYNTHASE gene is positively associated with desiccation tolerance of *Brassica napus* seeds during development. *Journal of Plant Physiology* 168: 1761-1770
68. Li H, Jiang H, Bu Q, Zhao Q, Sun J, Xie Q, Li C (2011) The *Arabidopsis* RING Finger E3 ligase RHA2b acts additively with RHA2a in regulating abscisic acid signaling and drought response. *Plant Physiology* 156: 550-563
69. Li C, Liu Z, Zhang Q, Wang R, Xiao L, Ma H, Chong K, Xu Y (2012) SKP1 is involved in abscisic acid signalling to regulate seed germination, stomatal opening and root growth in *Arabidopsis thaliana*. *Plant, Cell and Environment* 35: 952-965
70. Sun H-L, Wang X-J, Ding W-H, Zhu S-Y, Zhao R, Zhang Y-X, Xin Q, Wang X-F, Zhang D-P (2011) Identification of an important site for function of the type 2C protein phosphatase ABI2 in abscisic acid signalling in *Arabidopsis*. *Journal of Experimental Botany* 62: 5713-5725
71. Abdeen A, Schnell J, Miki B (2010) Transcriptome analysis reveals absence of unintended effects in drought-tolerant transgenic plants overexpressing the transcription factor ABF3. *BMC Genomics* 11: 69
72. Aranjuelo I, Molero G, Erice G, Avice JC, Nogues S (2011) Plant physiology and proteomics reveals the leaf response to drought in alfalfa (*Medicago sativa* L.). *Journal of Experimental Botany* 62: 111-123
73. Alekseeva AA, Savin SS, Tishkov VI (2011) NAD (+) -dependent Formate Dehydrogenase from Plants. *Acta Naturae* 3: 38-54
74. Demirevska K, Zasheva D, Dimitrov R, Simova-Stoilova L, Stamenova M, Feller U (2009) Drought stress effects on Rubisco in wheat: changes in the Rubisco large subunit. *Acta Physiologiae Plantarum* 31: 1129-1138
75. Holdsworth MJ, Bentsink L, Soppe WJ (2008) Molecular networks regulating *Arabidopsis* seed maturation, after-ripening, dormancy and germination. *New Phytologist* 179: 33-54
76. Angelovici R, Galili G, Fernie AR, Fait A (2010) Seed desiccation: a bridge between maturation and germination. *Trends Plant Science* 15: 211-218

77. Berjak P, Pammenter NW (2008) From *Avicennia* to *Zizania*: seed recalcitrance in perspective. *Annals of Botany* 101: 213-228
78. Sunderlikova V, Salaj J, Kopecky D, Salaj T, Wilhem E, Matusikova I (2009) Dehydrin genes and their expression in recalcitrant oak (*Quercus robur*) embryos. *Plant Cell Reports* 28: 1011-1021
79. Hanin M, Brini F, Ebel C, Toda Y, Takeda S, Masmoudi K (2011) Plant dehydrins and stress tolerance: versatile proteins for complex mechanisms. *Plant Signal Behav* 6: 1503-1509
80. Lahuta LB, Pluskota WE, Stelmaszewska J, Szablińska J (2014) Dehydration induces expression of GALACTINOL SYNTHASE and RAFFINOSE SYNTHASE in seedlings of pea (*Pisum sativum* L.). *Journal of Plant Physiology* 171: 1306-1314
81. Park SY, Fung P, Nishimura N, Jensen DR, Fujii H, Zhao Y, Lumba S, Santiago J, Rodrigues A, Chow TF, Alfred SE, Bonetta D, Finkelstein R, Provart NJ, Desveaux D, Rodriguez PL, McCourt P, Zhu JK, Schroeder JI, Volkman BF, Cutler SR (2009) Abscisic acid inhibits type 2C protein phosphatases via the PYR/PYL family of START proteins. *Science* 324: 1068-1071
82. Gao F, Ayele BT (2014) Functional genomics of seed dormancy in wheat: advances and prospects. *Frontiers Plant Sciences* 5: 458
83. Schweighofer A, Hirt H, Meskiene I (2004) Plant PP2C phosphatases: emerging functions in stress signaling. *Trends in Plant Science* 9: 236-243
84. Hu D-L, Chen Q-Z, Zhang C-J, Wang Y, Zhang B-J, Tang C-M (2013) Identification of cotton SKP1-like gene GhSKP1 and its function in seed germination and taproot growth in tobacco. *Canadian Journal of Plant Science* 93: 817-825
85. Stone SL (2014) The role of ubiquitin and the 26S proteasome in plant abiotic stress signaling. *Frontiers Plant Sciences* 5: 135
86. Lyzenga WJ, Stone SL (2012) Abiotic stress tolerance mediated by protein ubiquitination. *Journal of Experimental Botany* 63: 599-616
87. Zhang Y, Yang C, Li Y, Zheng N, Chen H, Zhao Q, Gao T, Guo H, Xie Q (2007) SDIR1 is a RING Finger E3 ligase that positively regulates stress-responsive abscisic acid signaling in *Arabidopsis*. *The Plant Cell Online* 19: 1912-1929
88. Ramírez V, Coego A, López A, Agorio A, Flors V, Vera P (2009) Drought tolerance in *Arabidopsis* is controlled by the OCP3 disease resistance regulator. *The Plant Journal* 58: 578-591
89. Logan DC, Millar AH, Sweetlove LJ, Hill SA, Leaver CJ (2001) Mitochondrial biogenesis during germination in maize embryos. *Plant Physiology* 125: 662-672
90. Nicholls C, Li H, Liu JP (2012) GAPDH: a common enzyme with uncommon functions. *Clin Exp Pharmacol Physiol* 39: 674-679
91. Sirover MA (2011) On the functional diversity of glyceraldehyde-3-phosphate dehydrogenase: Biochemical mechanisms and regulatory control. *Biochimica et Biophysica Acta (BBA) - General Subjects* 1810: 741-751
92. He D, Han C, Yang P (2011) Gene expression profile changes in germinating rice. *Journal of Integrative Plant Biology* 53: 835-844
93. Howell KA, Millar AH, Whelan J (2006) Ordered assembly of mitochondria during rice germination begins with pro-mitochondrial structures rich in components of the protein import apparatus. *Plant Molecular Biology* 60: 201-223
94. Sheen J (1990) Metabolic repression of transcription in higher plants. *Plant Cell* 2: 1027-1038
95. Li R, Moore M, Bonham-Smith PC, King J (2002) Overexpression of formate dehydrogenase in *Arabidopsis thaliana* resulted in plants tolerant to high concentrations of formate. *Journal of Plant Physiology* 159: 1069-1076
96. Roach T, Beckett RP, Minibayeva FV, Colville L, Whitaker C, Chen H, Bailly C, Kranter I (2010) Extracellular superoxide production, viability and redox poise in response to desiccation in recalcitrant *Castanea sativa* seeds. *Plant Cell Environ* 33: 59-75

-
97. Fridovich I (1989) Superoxide dismutases. An adaptation to a paramagnetic gas. *J Biol Chem* 264: 7761-7764
 98. Dixon DP, Steel PG, Edwards R (2011) Roles for glutathione transferases in antioxidant recycling. *Plant Signal Behavior* 6: 1223-1227
 99. Tommasi F, Paciolla C, de Pinto MC, Gara LD (2001) A comparative study of glutathione and ascorbate metabolism during germination of *Pinus pinea* L. seeds. *Journal of Experimental Botany* 52: 1647-1654

CHAPTER 5:

SUBTRACTIVE LIBRARIES FOR PROSPECTING DIFFERENTIALLY EXPRESSED GENES BETWEEN *Q. ILEX* GERMINATED SEEDS AND SEEDLINGS

Abstract

Suppression subtractive hybridization (SSH) is an efficient technique for generating cDNAs enriched for differentially expressed genes. We used this SSH methodology to identify differentially expressed genes between *Q. ilex* germinated seeds (stage S3) and seedling shoots of only 4 cm length (SS-4). Two libraries were designed to obtain clones of genes with increased or decreased mRNA levels in tissues from germinated seeds (reverse library) and seedling (forward library). A total of 664 clones with insert (375 in the reverse library and 289 clones in the forward library, respectively) were randomly selected and analyzed by PCR using vector-based primers. Only PCR products with a size above 300 pb were selected for sequencing. After elimination of the sequences with poor quality, 173 qualified trimmed ESTs, with a size ranging from 87 to 627 bp, were obtained. Ninety-two ESTs corresponded to transcripts enriched in germinated seeds and 80 ESTs to transcripts enriched in shoot seedling of holm oak. A total of 71 unique sequences were found in the two subtractive libraries. Thirty nine transcripts were more abundant in young development seedling and 31 transcripts in germinated seeds. Only one gene, coding for the sugar transport ERD6, was identified in the two libraries, thereby indicating that subtraction by hybridization was quite effective and that the abundance of a particular cDNA in one library was a consequence of the differential expression of the gene in the 'tester' sample used. Thirty-one over-expressed genes were identified in germinated seeds. Proteins encoded by these genes are representative of eight functional categories: stress responses, transport, oxidation-reduction, cell wall modification, cell division cycle, protein metabolism, cellular component organization and translation. In the other hand, 39 non-redundant transcripts over-represented in *Q. ilex* shoots seedlings were grouped in seven functional categories: photosynthesis, secondary metabolism, transport, signaling, stress response, gene expression and cellular component organization. Here we have identified for the first time a large number of putative differentially expressed ESTs from the embryo axis in germinated seeds and from shoot seedlings of *Q. ilex* during the post-germination and seedling establishment. Our data constitute an important genomics resource that should clearly benefit further germination and other biological process research on *Q. ilex*, given that this economically important forest tree species remains largely unexplored at the genomic level

5.1. Introduction

We are interested in determining the mechanisms that are involved in seed development and germination of holm oak seeds. Knowledge of the underlying biochemistry and metabolic status before and after the germination process could be important for the development and optimization of strategies for large-scale propagation and germplasm conservation in this species. In Ch. 4 we evaluated the role of a group of selected genes of interest in the *Q. ilex* germination process by using a hypothesis-driven approach for candidate genes. This approach relied on first sequencing the gene of interest by using orthologue sequences. Although targeted gene sequencing approaches are straightforward and inexpensive, they are time consuming, biased and have low throughput. Furthermore, owing to the complex nature of germination, these candidate-based strategies fail to reveal the complete landscape of all the genomic changes that occur in in this process. The development of genome-wide experimental approaches, encompassed with new and potent bioinformatics tools, allows biologists to take on the study of adaptive genetic diversity and its association with phenotypic trait variation in non-model organisms such as forest trees. In the last 30 years, genomics has become an integral part of forest tree research, and forest trees, rarely viewed before as model systems in plant biology, have gained much attention for population, evolutionary and ecological genomic studies ¹.

DNA-based microarrays represent a powerful high-throughput analytic technology for examining multigene expression patterns. The use of microarrays manufactured from transcripts of one species to probe gene expression in another, related, species, eliminates the need to fabricate a new microarray platform for every new species of interest ²⁻³. These procedures are less applicable to experiments with non-model organisms, which frequently comprise non-sequenced genomes as is the case of *Q. ilex*, underrepresented in public sequence databases. Suppression subtractive hybridization (SSH) is an efficient technique for generating cDNAs enriched for differentially expressed genes ⁴. Because *Q. ilex* is underrepresented in public sequence databases, we performed the SSH methodology aimed to identify differentially expressed genes between *Q. ilex* seeds at S3 stage (24h after imbibition, Ch. 4) and seedling shoots of 4 cm length (SS-4). The S3 stage coincides with the emergence of the radicle, which is considered the completion of germination *sensu stricto*. At this stage, the embryo cells experiment many metabolic changes needed for seedling formation. Understanding how the cell cycle genes work at this particular phase of plant development might help to clarify the cellular and structural events that bring a quiescent embryo to a metabolically active plant. To identify changes in the expression of genes involved in all this process we chose a very young seedling shoots where leaves primordia are visible (SS-4) for comparing with the S3 acorns.

Sixty-five differentially expressed genes were detected, and most of them were identified as genes involved in a variety of biological functions. In the reverse library, representing to genes over-expressed in germinated seeds at the S3 stage, mechanisms involved particularly in seed protection against water shortage and reserves mobilization were observed. Genes in the forward library, over-expressed in seedlings, were mainly implicated in the transition switch from the heterotrophic to photoautotrophic metabolism and the biosynthesis of secondary metabolites. To verify the SSH results and assess inter-individual variability eight transcripts (*Eno*, *Lea-5*, *Sod1*, *NdhF (Ndh6)*, *PetE*, *RbcS*, *RbcL* and *Pp2c*) were selected for absolute quantification by real-time qRT-PCR. The qRT-PCR data revealed substantial differences in the absolute amounts of the eight transcripts and confirmed their up- or down-regulation observed in the SSH assay. Additionally, a good correlation between transcript variation along seed germination and protein/activity abundance was found for some of the studied genes. This data indicate that in our experimental system, transcript quantification provides good predictive value with respect to the extent of protein changes in abundance. Overall, this work provides novel insights into the molecular pathways that could mediate seed germination and seedling establishment in holm oak. An EST collection from *Q. ilex* after SSH was generated, which constitutes an important genomics resource to investigate germination and other biological process in this interesting forest tree.

5.2. Material & Methods

5.2.1. Plant material

We used mature acorns harvested from healthy holm oaks from Cerro Muriano (Córdoba, SW Spain) (Fig. 3.1, p. 45) described in Ch.3, including immature acorns collected in August, September and October (five, six and seven month after polinization (MAP), respectively) for RBCL western blots experiments. Germination and growth experiment was performed as described in Ch.3 (Fig. 3.2, p. 46).

Embryo axes from holm oak seeds were excised 24 h after imbibition (stage S3, Fig. 5.1.A). The leaf primordia, including the proximal tail, were excised from seedling when they reached a length of approximately 4 cm (Fig. 5.1.B). Also were collected embryo axis tissue at S0, S2, S3, S5 and S7 stage to perform the transcripts profile analysis seedlings (Fig. 3.3, p 47) and embryo axis at 5, 6 and 7 MAP stages (Fig. 5.1.C). Individual samples were immediately frozen in liquid N₂ and stored at -80°C. Three independent pools (biological replicates) were prepared for each sampling time (we referred to them as mini-pools). Each mini-pool was derived from 20-30 individuals.

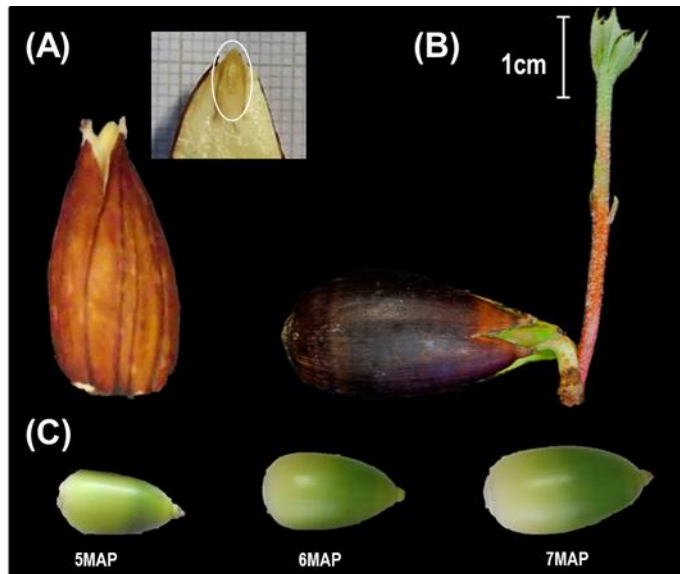


Figure 5.1: Tissues used for the construction of SSH libraries. (A). Embryo axes from holm oak seeds were excised 24 h after imbibition (stage S3). This point coincides with the rupture of the testa and the extrusion of the radicle, and mark the end of the germination *sensu stricto*. (B). Leaf primordia, including the proximal tail, excised from seedling shoots with a length of approximately 4 cm (stage SS-4). (C). Inmature acorns collected in August, September and October corresponding to 5, 6 and 7 month after polinization (MAP).

5.2.2. Construction of suppression subtractive hybridization (SSH) libraries

Forward (F) and reverse (R) libraries were designed to obtain clones of genes with increased or decreased mRNA levels in tissues from holm oak acorns collected at two different stages (S3 and SS-4). Total RNA was extracted from each mini-pool prepared per sampling time using the InviTrap® Spin Plant RNA Mini Kit (Invitex) with slight modifications as previously described⁵. Genomic contamination with DNA was eliminated by DNase I (Ambion) treatment. The RNA quantitation was performed with the Qubit® RNA Assay Kit in the Qubit® 2.0 Fluorometer (Invitrogen). The RNA quality was checked by using an Agilent 2100 Bioanalyzer (Agilent Tech.). Only RNAs with an integrity number (RIN) >8 and a ratio A260/A280 of ~2 were used. Equal amounts of each RNA sample with good quality were pooled to for poly-A mRNA purification; one pool was prepared for each sampling stage (S3 and SS-4). The mRNA were isolated from total RNA with Dynabeads® Oligo (dT)₂₅ (Dynabeads® mRNA Purification Kit, Life Technologies) according to the manufacturer's protocol. Absence of rRNA was checked by using the Agilent 2100 Bioanalyzer (Agilent Tech.). The SSH libraries were produced using the PCR-Select™ cDNA subtraction kit (Clontech) following the manufacturer's protocol, as previously described⁶. The cDNA of SS-4 tissue was used as a 'tester' and the cDNA of S3 seeds was used as a 'driver' to construct the forward subtracted library. The reverse subtraction was performed by using the cDNA from S3 seeds as a 'tester' and the cDNA from SS-4 tissue as a 'driver'.

A workflow followed in SSH construction is shown in Fig. 5.2. Briefly, double stranded cDNA was digested with *RsaI* and two tester populations were generated with adapters 1 and 2R. No adapter was ligated to the driver sample. A first hybridization was performed by mixing an excess of driver to each tester population (adapter 1 or 2R). In this first hybridization, the concentration of high- and low-abundance sequences is equalized among the type *a* molecules because re-annealing is faster for the more abundant molecules due to the second-order kinetics of hybridization. At the same time, type *a* molecules are significantly enriched for differentially expressed sequences while cDNAs that are not differentially expressed form type *c* molecules with the driver. The two primary hybridization samples are mixed together without denaturing (second hybridization) in order to generate PCR templates from differentially expressed sequences. Only the remaining equalized and subtracted ss tester cDNAs (molecules *a*) can reassociate and form new type *e* hybrids. These new hybrids are ds tester molecules with different ends, which correspond to the sequences of Adaptors 1 and 2R. Fresh denatured driver cDNA is added (again, without denaturing the subtraction mix) to further enrich fraction *e* for differentially expressed sequences. After filling in the ends by DNA polymerase, only the type *e* molecules (the differentially expressed tester sequences) have different annealing sites for the nested primers on their 5' and 3' ends. In consequence, only type *e* molecules are exponentially amplified (type *b* molecules contain long inverted repeats at the end and form stable "panhandle-like" structure and type *a* and type *c* molecules have a linear amplification). A secondary PCR amplification is performed using nested primers to further reduce any background PCR products and enrich for differentially expressed sequences. An analysis of results in each phase was performed following the manufacturer indications.

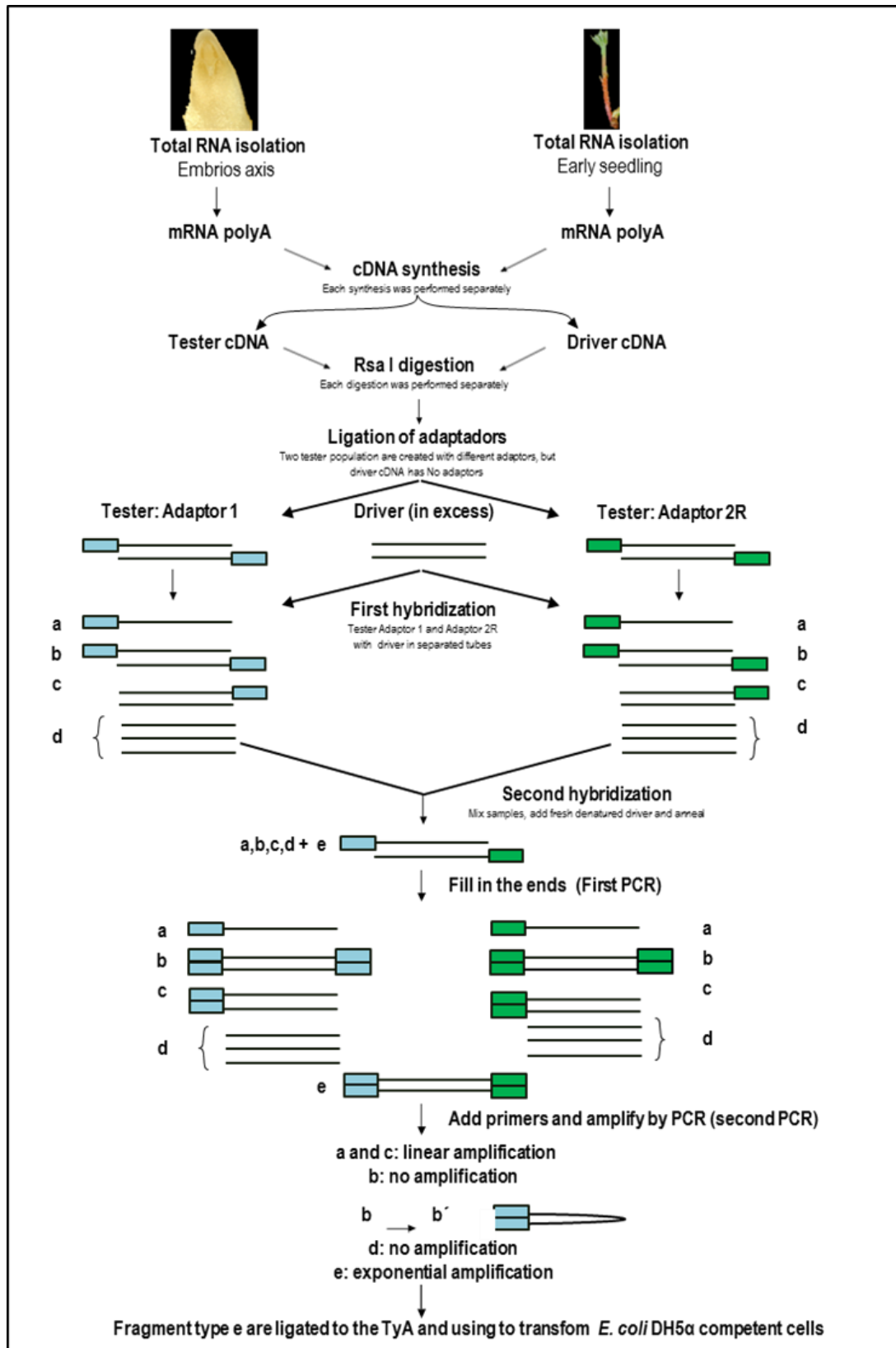


Figure 5.2: Construction of the forward library showing the SSH methodology workflow (adapted from Diatchenko et al, 1996). Type e molecules are formed only if the sequence is upregulated in the tester cDNA. Lines represent the Rsa I-digested tester or driver cDNA. Boxes represent the Adaptor 1 (blue) and 2R (green).

PCR products were ligated into the T&A Cloning Vector (Eastern Biotech Co. Ltd) and introduced in *E. coli* DH5 α (ECOS 101, Yeastern Biotech Co, Ltd) competent cells. Transformed bacteria were grown overnight at 37°C in Luria-Bertani medium supplemented with ampicillin (100 mg/mL), isopropyl-beta-D-thiogalactopyranoside (100mM) and 5-bromo-4-chloro-3-indolyl-beta-D-galacto-pyranoside (40 mg/ml). Transformant colonies containing plasmids with foreign DNA inserted into the plasmid LacZ gene were selected by blue-white screening ⁷.

5.2.3. Clones sequencing and sequence analysis

Clones were randomly selected and analyzed by PCR, using vector-based primers (M13 universal primers). PCR products were loaded onto 1% of agarose gel to verify the presence of one unique band and then purified by the NucleoSpin Gel and PCR Clean-up (Macherey-Nagel). PCR products were sequenced on ABI PRISM™ 3130 XL sequencer (Applied Biosystems) using standard M13 primers. Sequencing output was edited to remove vector sequences, PCR primers, and terminal ambiguities by using the software Chromas and ChromasPro v.1.5 (Technelysium). Trimmed sequences were aligned using the SeqMan module of the DNASTAR Lasergene software, to assemble the ESTs into contigs. The assembly conditions were 80% minimum homology and 25 base minimum overlap.

The BLAST server at NCBI (<http://www.ncbi.nlm.nih.gov/blast/Blast.cgi>) was used to search homologous nucleotide and protein sequences. BLAST searches were performed using the tBLASTx and BLASTn algorithm with default search conditions against the GenBank non redundant (nr) nucleotide collection, expressed sequence tags (EST_NCBI), whole-genome shotgun contigs (WGS_NCBI) and Quercus_DB (custome database)⁸. Classification of the annotated sequences was performed based on BLAST results and Gene Ontology annotation (<http://www.geneontology.org>). Sequences were deposited in the expressed sequence tag database (dbEST) of GenBank under the accession numbers from JZ794738 to JZ794825.

5.2.4. Primers design and qRT-PCR

Primers used to quantify the transcript molecules of *RbcL*, *Sod1*, *Pp2c* and *NdhF* genes were described in Ch. 4. Primers directed against other four selected genes (*RbcS*, *PetE*, *Lea-5* and *Eno*) were designed with OLIGO Primer Analysis Software v 7.58 (Molecular Biology Insights, Inc) as previously described (Pueyo *et al.*⁹ and Ch.4). Briefly, primers were required to have high T_m ($\geq 70^{\circ}\text{C}$), optimal 3'- ΔG (≥ -3 kcal/mol) values, and to be free of hairpins and duplex structures. All primer pairs produced specific PCR products of the predicted size and all PCR products were further verified by nucleotide sequencing.

Table 5.1: Primers used for absolute quantification by qRT-PCR of a selected group of differentially expressed transcripts in S3 and SS-4 stages seeds of *Q. ilex*. The primers used for cDNA calibration are included.

Target ^a	Genebank Acc. number	Sequence 5'-3' ^b	Length ^c (bp)	Amplification efficiency ^d (r%)
<i>PetE</i>	JZ794781	F: AGTAACTTTTCCCACCATCCCAGCTCCCT R: ATGCTCCAGGAGAGACATACGCTGTGACC	101	1.0004 (99.9%)
<i>Pp2c</i>	KC150869	F: GCAAGTGTGTAATGAGGCAGGTTTATTCCACC R: CCATTTCCACGATGACATTACAGTGATGTGTTG	132	0.9981 (99.8%)
<i>NdhF</i>	KC150873	F: GAAGAGCATTTCACCAAAAAGATCAGTCC R: TCCATATTCAAATAGCGGAGATTCACGAAG	95	0.9934 (99.7%)
<i>RbcL</i>	AB125020	F: CGCATAAATGGTTGGGAATTAACGTTCT R: GGGATTATCCGCTAAGAATTACGGTAGA	105	0.9929 (99.7%)
<i>Sod1</i>	KM262658	F: CGCAGATCCAGATGATCTTGGCCGAGGG R: AGCACACAACAGAGTAGGGATTAGAAGACG	137	0.9941 (99.7%)
<i>RbcS</i>	JZ794783	F: ACTGGGTGATGTGGAAGCTTCCCATGTT R: GCACTGCACTTGACGCTTGTTCGAATCC	140	0.9921 (99.9%)
<i>Eno</i>	JZ794741	F: CCCTTGTGCAACTCCCCGTAAACAAC R: GCTTGGCTCAGAAGCAGTCTATGCTGG	113	0.9980 (99.6%)
<i>Lea-5</i>	JZ794754	F: AAATGGTTCCACAATAAGTCCAAGGGCAG R: GCAGCGGAAATTGATGCAGTCGAGC	147	0.9966 (99.6%)
Calibrator gene^e				
<i>A170</i>	U57413	F: GGAAGAGAAGCCGCTGACACCCACT R: CCGTCAAGTTTGTGACTTCCGAAG	113	1.0001 (99.9%)

^aGene symbols are according to the NCBI Gene database

^bSequence of forward (F) and reverse (R) primers specific for *Q. ilex* gene sequence

^cPCR product size in base pair (bp)

^dThe real-time PCR efficiencies (E) were calculated from each standard curve according to the equation $E=10^{(-1/\text{slope})}-1$. E is the range from (minimum value) to 1(maximum and optimum), i.e., E=1 is equal to 100% efficiency.

^ePrimer based on *M. musculus* sequences were designed for absolute quantification by real-time PCR of A170 gene in liver mice tissue, used to guarantee the quality of retro-transcription and to establish the threshold position.

Real-time qRT-PCR was performed as previously described¹⁰. cDNA was generated from 2 µg of total RNA, and PCR was performed in quadruplicate using 50 ng/well of cDNA. The absolute quantification relates the PCR signal (Ct value in real-time PCR) to input copy number using a calibration curve, as described in Ch. 4 (p. 60). No primer dimers were detected. All primers designed for absolute real-time qRT-PCR amplified with the same optimal PCR efficiency (~100%) in the range of 20 to 2x10⁵ pg of total RNA input with high linearity (r > 0.99) (Table 5.1). An absolute calibration curve was constructed with an external standard in the range of 10² to 10⁹ RNA molecules, and the number of transcript molecules was calculated from the linear regression of the standard curve as previously described¹⁰.

5.2.5. Protein extraction and immunoblotting

Proteins were extracted from embryo axis of collected at 5, 6 and 7 month after pollinization (MAP), from the S0, S3, S5 and leaf primordia, including the proximal tail from SS-4. All process implicated in protein extraction and immunoblotting were performed as described in Ch.4 (p. 60). Blocked PVDF membranes were incubated with rabbit antibodies raised against or RBCL (dilution:1:500), washed, and incubated with the secondary (1:2000) goat anti-rabbit IgG (Sigma: A-9169) antibody conjugated to horseradish peroxidase, for an additional 1 h. Blots

were developed using the Clarity™ Western ECL Detection System (BioRad). Image captures and densitometric analyses were performed with the ChemiDoc™ MP Imaging System (BioRad) and ImageLab 4.1 software (BioRad).

5.2.6. Superoxide dismutase activity

Protein extracts from S0, S2, S3, S5, S7 and SS-4 were obtained as described in Ch. 4 (p. 61). The different isoenzymes of SOD were identified by staining parallel gels previously incubated at 25 °C for 20 min in either 50 mM potassium phosphate buffer pH 7.8 or in buffer containing 2 mM KCN or 5 mM H₂O₂ as detailed in Ch. 4. Gels were stained with Coomassie brilliant blue R-250 (BioRad). Image captures and densitometric analyses were performed with the ChemiDoc™ MP Imaging System (BioRad) and ImageLab 4.1 software (BioRad), respectively. Relative bands volumens of SOD are normalized by the protein amount in the line and expressed in arbitrary units.

5.3. Results and discussion

Transcriptomics, like other “omics” methodologies, is a formidable tool to investigate complex metabolic processes. This work focused on assessing the potential utility of transcript expression profiling for unrevealing the metabolic changes involved in seeds germination and seedling establishment of the recalcitrant species *Q. ilex*. Here, we used a SSH strategy to overcome the limited genomic sequence information available for this orphan species.

5.3.1. Plant material

We selected seeds at S3 and SS-4 stages for the construction of the SSH libraries. As described in Ch. 4, *Q. ilex* seeds needs 24h after imbibition to end the germination process *sensu stricto* in our experimental condition. Is in this S3 stage that the *testa* is broken and the radicle emergence occurs. In a previous work described in Ch.4 we observed that in this stage: (i) there is an important reduction in the level of transcripts accumulated during the seed maturation; (ii) the osmoprotective proteins (DHN3) and carbohydrates (sucrose, raffinose/GOLS) decreased; (iii) the amount of monosaccharides (Glc, Fru) are still low for a fully active energetic metabolism; and (iv) there is low levels of phytohormones (ABA, GA, IAA, iP, iPR.). In the SS-4 stage shoots, a group of primordial green leaves can already be observed, indicating that metabolism must be in this stage adapted to photosynthesis. Hence, the comparison of seeds at S3 and SS-4 stages, can let to identify the set of genes and metabolic pathways involved in holm oak seedlings establishment.

5.3.2. Subtractive libraries

SSH is a powerful strategy to identify differentially expressed transcripts without a previous knowledge of their sequences. Previous studies using a SSH approach have been centred in other *Quercus* species mainly aiming to discover genes involved in the response to stressing environmental conditions or the biosynthesis of economically interesting products (*i.e.*, cork, Kremer *et al.*¹¹ and refs. herein). Derory *et al.*¹² used this SSH methodology to identify differentially expressed genes among early bud development stages in sessile oak *Quercus petraea*. They identified genes involved in the cell rescue/defense-, metabolism-, protein synthesis-, cell cycle-, and transcription-related functional categories. Here, the SSH method was applied for the first time to identify genes that showed differential mRNA expression in association with the postgermination and early seedling establishment of *Q. ilex* seeds. To improve seed conservation and avoid abiotic stress damage in the context of global climate change, an improved understanding of the metabolic and signaling pathway for seed germination and sprouts survival is crucial for biomass production and forest management. By using SSH instead of a heterologous microarrays approach, we also obtained the sequences of the gene fragments isolated. This is an additional advantage of SSH methodology that permits to enrich the number of *Q. ilex* cDNA sequences accumulated in public databases. We used this sequences to design primers for verification by qRT-PCR of a selected group of the SSH identified genes in our studied samples. We also analysed by qRT-PCR the changes in their transcriptional profiles during six stages covering the germination, post-germination and early seedling stages.

mRNA quality

The quality of total RNA and mRNA is an important factor to consider in SSH library construction. We only used total RNA with A260/280 ratios ≥ 2 and RIN numbers > 8 as determined by using the Agilent 2100 Bioanalyzer (Fig. 5.3.A). In the case of mRNA, we also used this approach to assess the absence of rRNA. Figure 5.3.B shows the electropherogram of the S3 mRNA sample and indicates that the sample is enriched in mRNA and only contains a minor amount of contaminating ribosomal RNA.

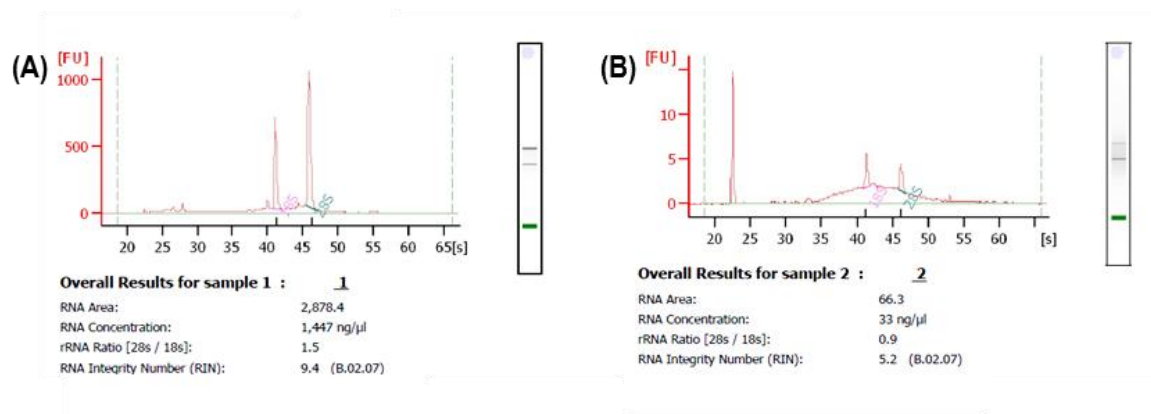


Figure. 5.3: Analysis of total RNA and mRNA quality. Total and mRNA preparations were analyzed with the Agilent 2100 bioanalyzer. As an example, the electropherograms of the S3 total (A) and mRNA (B) samples are shown alongside the gel-like images.

Controls in the process of construction of the SSH libraries

Two SSH libraries were constructed. For the reverse (R) library construction, selective amplification and enrichment of abundant cDNAs in germinated seeds (S3) samples (tester) was achieved by incubating and hybridizing S3 cDNA with an excess of seedling shoots (SS-4) cDNA (driver). The forward (F) subtraction, which identify transcripts more abundant in SS-4 than in S3, was performed by using the cDNA from SS-4 as a ‘tester’ and the cDNA from S3 as a ‘driver’.

The processes of subtraction were analysed at different levels along the experiment (Fig. 5.4). Quality of the *Rsa*I digestion of cDNA was analysed by comparing undigested and digested cDNAs on 1% agarose gel. Figure 5.4.A shows that cDNA size distribution increased considerably in the digested samples, indicating a successful digestion.

The ligation efficiency of each tester population was analysed by PCR amplification (Fig. 5.4.B). A primer pair (named *Pp2c-SSH*) was designed to specifically amplify a fragment corresponding to the *Pp2c* gene. The *Pp2c-SSH* primer 5’ was compatible with the PCR Primer 1 and the PCR Primer 2R, both included in the SSH kit. Tester 1-1 and tester 1-2R populations (corresponding to the reverse library, where S3 cDNA was ligated to one or another adapter) were amplified by using one gene specific primer (*Pp2c* primer 5’) and the PCR Primer 1 or PCR Primer 2R, respectively, both included in the SSH kit. The intensity of the obtained bands (300 pb) was compared with that obtained by using the two *Pp2c* primers 5’ and 3’ (200 pb). A similar procedure was used to analyse the ligation efficiency during the forward library construction. In all cases the intensities differed by less than 75%, which is recommended to ensure efficient subtraction, accordingly to the manufacturer's guidelines.

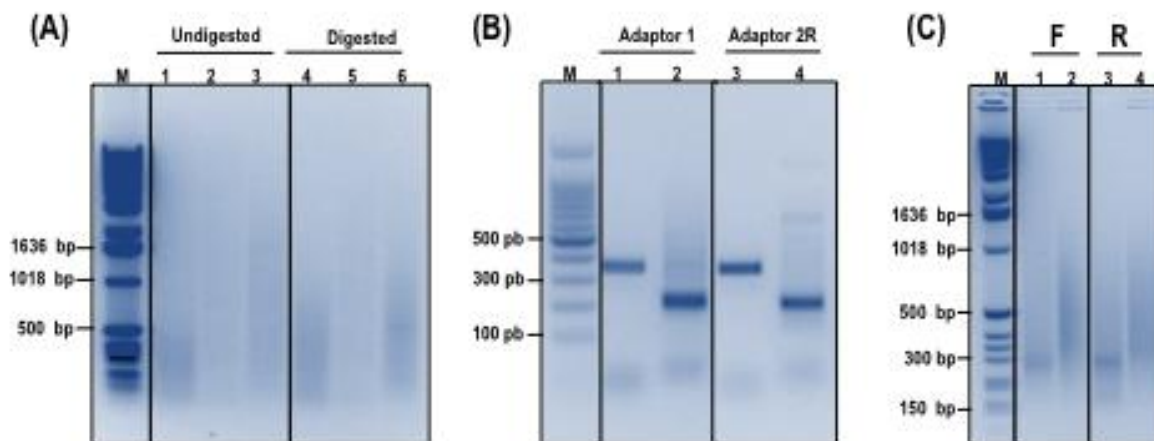


Figure 5.4: Analysis of results in the most important phase in SSH construction. (A) Analysis of double-strand cDNA digestion with *Rsa*I. Electrophoresis on a 1% agarose of 2.5 μ l of undigested, ds cDNA and 5 μ l of *Rsa* I-digested cDNA of embryo axis (lanes 1, 4), early seedling (lanes 2, 5) and Control Kit (lane 3, 6) samples, respectively. The average cDNA size is reduced after *Rsa* I digestion. M: Molecular weight marker X (0.07 to 1 5.2 Kbp DNA). **(B)** Analysis of the adaptors ligation efficiency in the R-library construction. After the ligation of adaptors 1 and 2R to the tester cDNA (S3 in this example), a 40 cycles PCR amplification was performed. Line 1: PCR products using Tester 1-1 (Adaptor 1-ligated) as the template, and the *Pp2c* 5' and PCR Primer 1 primers. Lane 2: PCR products using Tester 1-1 (Adaptor 1-ligated) as the template and the *Pp2c* primers 5' and 3'. Lane 3: PCR products using Tester 1-2R (Adaptor 2R-ligated) as template, and the *Pp2c* primers 5' and PCR Primer R2. Lane 4: PCR products using Tester 1-2 (Adaptor 2R-ligated) as the template, and the *Pp2c* primers 5' and 3'. Amplicons were loaded on 2% agarose/1x Gel Red. M: 100 bp DNA ladder molecular weight marker. *Pp2c* primers **F**: CGAAACGTTCTTCGTCGTTTC and **R**: GTGTAATGAGGCAGGTTTATTC **(C)** Control of subtraction experiments at the secondary PCR step. PCR band patterns for the unsubtracted (Line 1, 3) and subtracted (Lines 2, 4) cDNAs ligated to the adaptors, used to construct the forward and the reverse libraries. PCR products were loaded on 2% agarose/1x Gel Red. M: Molecular weight marker X (0.07 to 1 5.2 Kbp DNA).

We realized the analysis of subtraction efficiency by comparing the banding patterns obtained after PCR amplification after the secondary PCR (Fig. 5.4.C). Unsubtracted and subtracted cDNAs used for the construction of the forward and reverse libraries, were amplified by using adapter nested primers included in the SSH-kit. The banding patterns of unsubtracted cDNAs ligated with both adaptors were different from the banding patterns the experimental subtracted cDNAs samples, indicating a successful subtraction, accordingly to the manufacturer's guidelines.

Clone selection, sequencing and identification of genes enriched in the F and R libraries

A total of 664 clones with insert (375 and 289 clones from the R and F libraries, respectively) were randomly selected and analyzed by PCR using vector-based primers. Only PCR products with a size above 300 pb (Fig 5.5) were selected for sequencing. After elimination of the sequences with poor quality, 173 qualified trimmed ESTs, with a size ranging from 87 to 627 bp were obtained. Ninety-two ESTs corresponded to transcripts enriched in germinated seeds (R library) and 80 ESTs to transcripts enriched in shoot seedling (F library) of holm oak acorns (Table 5.3 and 5.4, see appendixes). Only 10% of ESTs represented redundant

sequences or various fragments of the same cDNA. There were two highly redundant transcripts. One, identified as *Lea-5*, accounted for 43 out of the 92 ESTs from the R library, what is in agreement with previous works reporting LEA-5 as an abundant protein in oak seeds¹³⁻¹⁴. The other is *RbcS*, that accounted for 17 out of 80 ESTs from the F library, something expected as the F library was constructed from a green tissue and *RbcS* forming part of the photosynthetic machinery. Redundant sequences were assembled into contigs with the SeqMan software (DNASTAR, Lasergene). Except for one *Lea-5* fragment, all redundant sequences for each gene formed just one contig.

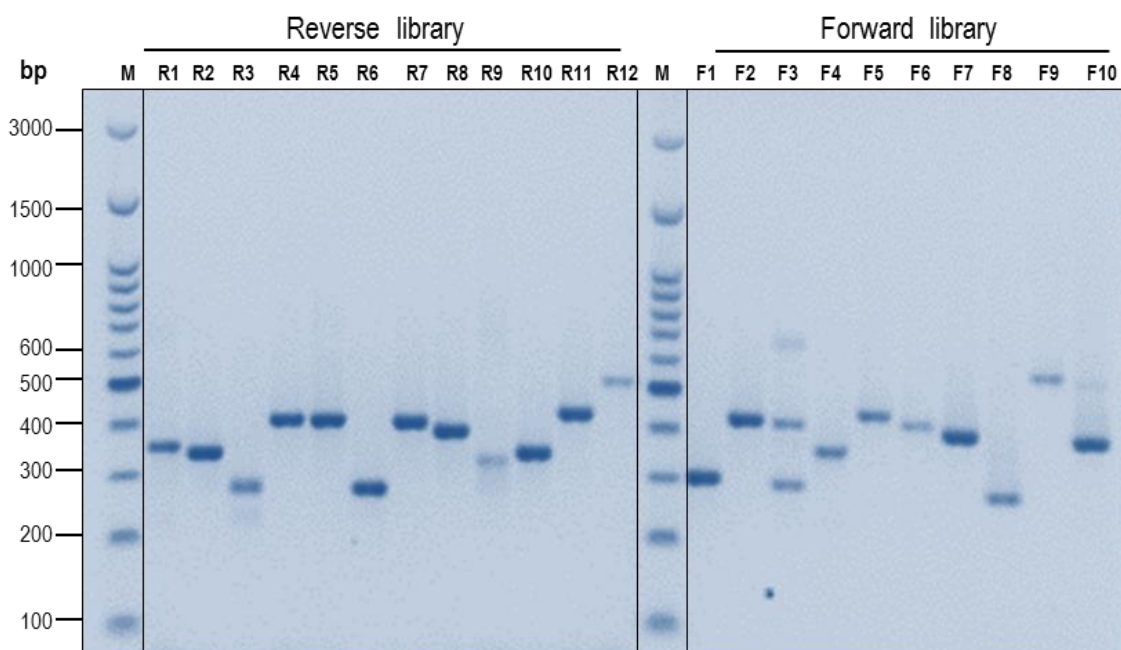


Figure 5.5: Analysis of transformant by PCR. Representative gels showing the size of the inserts included in the vector DNA in 10-12 clones from the R and the F libraries. Bacterial cells were lysed and the plasmids amplified by PCR using the vector M13 primers. Bands were separated on 2% agarose gels. M: Molecular weight marker X (0.1 to 3 Kbp DNA).

As discussed ahead, RuBisCo is composed by an equal amount of its small (RBCS) and large (RBCL) subunits. However, we isolated 17 clones with *RbcS* inserts vs. only one clone that matched with *RbcL*. This paradox can be explained if considering that RBCL is coded by a chloroplastidic gene, that chloroplast transcripts generally lack poly-A tails, and that the methodology used to purify mRNA was based on poly-A hybridization with oligo (dT) immobilized on beads.

A total of 71 unique sequences were found in the two subtractive libraries. Thirty nine transcripts were more abundant in young development seedling (F library) and 31 transcripts in germinated seeds (R library). Only one gene, coding for the sugar transport ERD6, was identified in the F and R libraries, thereby indicating that subtraction by hybridization was quite effective and that the abundance of a particular cDNA in one library was a consequence of

the differential expression of the gene in the 'tester' sample used. Around 80% of ESTs sequences showed high homology with plant genes of known and unknown functions registered in databases using tBLASTn or tBLASTx algorithms in the NCBI_nr database. Interestingly, all but one of the identified sequences were found or are of public access for the first time in *Q. ilex*. The remained 20% ESTs sequences (12 from R- and 19 from F-libraries) had no significant homology with any plant gene. The relatively high proportion of non-identified sequences is not surprising given that the only three tree genomes sequenced to date are from species distantly related (*Populus*, *Eucalyptus* and *Citrus*) to *Q. ilex*. These 31 novel identified sequences were blasted against EST_NCBI and WGS_NCBI databases. The nucleotide sequence hits with a percentage of identity higher than 80% was used to perform tBLASTx annotation by using the NCBI nr (with taxonomy restrictions to *Viridiplantae*) and Quercus_DB⁸ databases. By means of these approaches, a total of 10 ESTs for both libraries have been annotated (Table 5.5). As discussed elsewhere, the unidentified sequences may represent poorly conserved 3'-untranslated regions or unspliced introns^{4,6}.

Verification of SSH results and analysis of selected genes during germination and seedling establishment by real time qRT-PCR

Real-time RT-PCR was conducted to confirm that the presence of ESTs in the subtractive libraries truly represented their differential mRNA expression. Eight genes were chosen for this analysis. Selected genes were related to stress response (*Eno*, *Lea-5*), cell redox homeostasis (*Sod1*, *NdhF*), photosynthesis (*PetE*, *RbcS*, *RbcL*) and response to stimulus (*Pp2c*). All selected transcripts were quantified in an absolute manner to prevent the inaccuracy of most internal standards as quantitative references, and to express the transcript level changes in terms of molecule copy numbers, the only way to fully understand the biological significance of the observed transcriptional changes as discussed in Ch.1 and 4.

Quantifications were first made on the same pooled S3 and SS-4 samples used for the construction of R and F libraries. The results are shown in Table 5.2, where genes are listed according to their mRNA steady-state levels in embryo axes (*Lea-5*, *NdhF*, *Eno*, *Sod1*) or shoots (*RbcS*, *RbcL*, *PetE*, *Pp2c*). The qRT-PCR quantifications confirmed the expression patterns predicted by SSH, *i.e.* the up-regulation of genes involved in photosynthesis, CO₂ assimilation and cell signaling processes in the seedlings, and the downregulation in germinated seeds of genes involved in the protection of mature unimbibed acorns against the water loss.

Table 5.2: Comparison of the number of clones isolated in the R and F libraries with data of the absolute quantification by qRT-PCR of a selected group of differentially expressed transcripts in S3 and SS-4 stages seeds of *Q. ilex*.

<i>Gene Symbol</i>	<i>Reverse library</i>		<i>Forward library</i>	
	<i>No. of clones</i>	<i>mRNA molec/ pg total RNA</i>	<i>No. of clones</i>	<i>mRNA molec/ pg total RNA</i>
<i>Lea-5</i>	43	4884 ± 245		8 ± 1.4
<i>NdhF</i>	2	558 ± 41		102 ± 8
<i>Eno</i>	1	310 ± 23		94 ± 10
<i>Sod1</i>	1	87 ± 4		43 ± 1.8
<i>RbcS</i>		1 ± 0.2	17	1173 ± 25
<i>RbcL</i>		377 ± 32	1	2394 ± 87
<i>PetE</i>		0.3 ± 0.8	1	171 ± 30
<i>Pp2c</i>		6 ± 0.2	1	15 ± 0.5

As expected, substantial differences in abundance were found depending on the transcript and tissue examined. Thus, highly abundant (>500 molecules/pg of total RNA) and rare (≤ 1 molecules/pg) mRNA species were quantified in embryo axes and seedlings. *Lea-5* was the most abundant of all quantified transcripts and decreased more than 600-fold (4884 vs 8 mRNA molec/pg total RNA in S3 and SS-4 samples, respectively) during the post-germination. *RbcS* and *RbcL* were the most abundant transcripts in seedlings. However, the numbers for *RbcS* increased a thousand-fold during postgermination, whereas *RbcL* mRNA only showed a 6-fold increase, perhaps because it already had high transcript levels in seed. These data most likely explain the aforementioned redundancy of *Lea-5* and *RbcS* transcripts in the R and F libraries, respectively. Moreover, data in Table 5.2 highlight the relevance of the absolute measurements when comparing the increments in transcript molecules with the conventional fold variations. For instance, although a relative 6-fold increase in *RbcL* transcript levels might look modest compared to the 1000-fold increment in *RbcS* mRNA, the actual scenario is that *RbcL* (highly abundant mRNA in germinated acorns) exhibited a higher increase (from 377 to 2394) in copy number than *RbcS* (low abundant mRNA, that change from 1 to 1173 molecules).

Pooling of samples reduces biological variation by minimizing individual variation and increase statistical performance constituting an advantage in omic approaches that are arduous and time-consuming. The use of pooled samples to characterize populations can also yield more precise and less biased parameter estimations, compared to the use of individual samples¹⁵. A potential risk of using pooled samples is that a single aberrant sample can negatively impact the quality of a particular pool and hence confound the results and their interpretations. Another pivotal point is that biological replicates are crucial when estimating natural variability and thereby when defining the “biological noise” beyond which differential expression can be definitively established. To accurately validate the up- or down-regulation of the selected transcripts, quantifications were performed using biological replicates, each formed by mixing equal amounts of total RNA from 20-30 individuals. We also included in the study samples corresponding to the S0, S2, S5 and S7 stages, in order to analyze the variation of these

transcripts during germination (S0 to S3), postgermination (S3 to S7) and early seedling (S7 to SS-4) (Fig. 5.6).

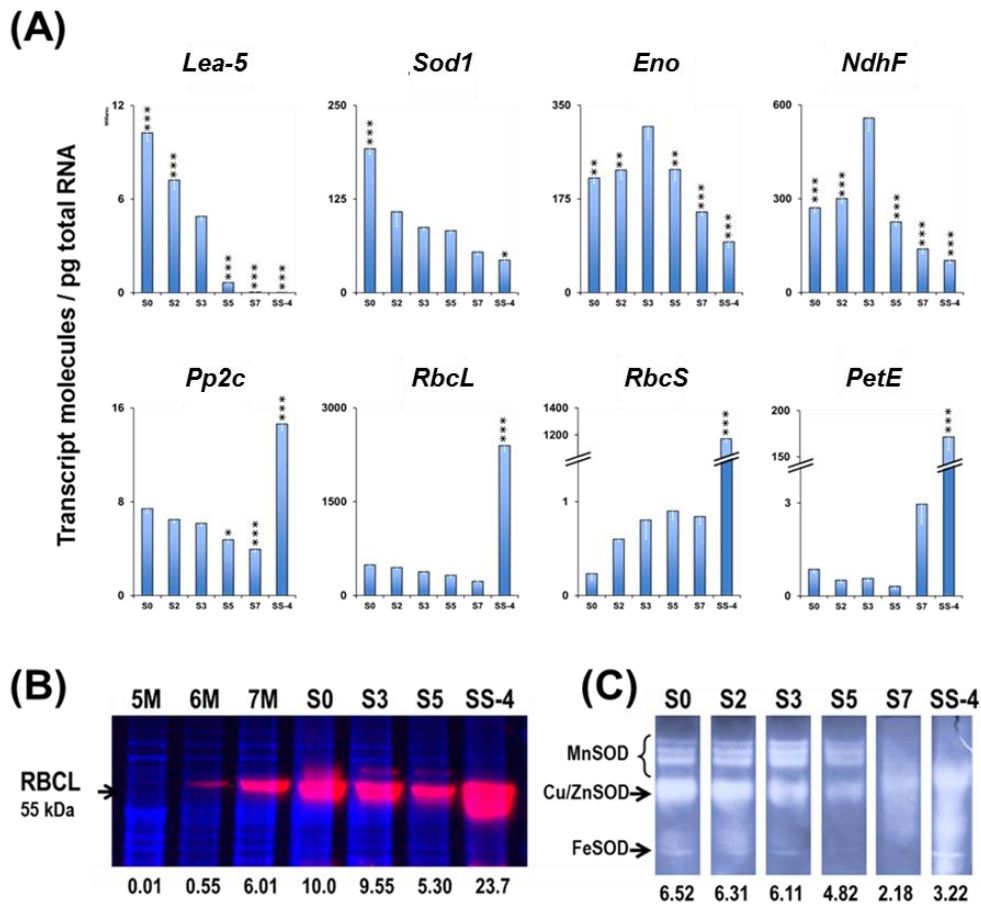


Figure 5.6: Verification of SSH results. (A) Developmental stage associated differences in transcript levels of selected genes. Absolute quantification of transcripts were performed by qRT-PCR. Data are means \pm SEM ($n=3$ biological replicates per condition) of mRNA molecules/pg of total RNA. Statistical significance vs. S3 embryo axes tissue is expressed as: *** $P < 0.001$, ** $P < 0.01$, and * $P < 0.05$. **(B)** Western blotting of RBCL protein in *Q. ilex* seeds samples. Proteins were extracted from the same pools used in the transcriptional analysis. Numbers indicate the arbitrary Western blotting signal intensities normalized to the total protein contents, using Stain-free technology for total protein quantification. M is referred to MAP. **(C)** SOD activity gel assay showing the SOD protein profiles on native PAGE of different *Q. ilex* seeds germination stages. Proteins were extracted from the same pools used in the transcriptional analysis. A 20 μ g of crude soluble proteins were loaded onto a 10% acrylamide gel. SOD activity was visualised by the NBT staining method. Numbers indicate the arbitrary signal intensities normalized to the total protein contents, using Stain-free technology for total protein quantification.

The variation inter-pools was low and similar in all the groups. Accordingly, we were able to verify at an individual level a substantial part of the changes in transcript copy number previously quantified on pooled samples. These results suggest that the quantifications in Table 5.2 should not be particularly open to misinterpretation due to inter-individual differences in transcript levels. Three patterns of variation were observed for the eight transcripts analyzed (Fig. 5.6.A). *Lea-5* and *Sod1* showed high levels of mRNA molecules in the mature acorn that decreased gradually as the development of the seed progressed. *Eno* and *NdhF* are abundant transcripts in mature un-imbibed seeds, but their levels increased during imbibition, picking at

the S3 stage and decreasing later during post germination and early seedling stages. *Pp2c* and *RbcL* showed a medium or high level of abundance along all the studied period and increased about one magnitude order at the SS-4 stage. Finally, the amounts of *Rbcs* and *PetE*, barely detectable during the germination and post-germination periods, picked dramatically in the seedling.

Verification of SSH results by immunoblotting and in-gel SOD activity staining

Studies at the mRNA levels are largely based on the supposition that RNA profiling might provide indirect support for protein expression levels in relation to gene functions and phenotypes. To assess the correlation between the differential expression of specific mRNAs and corresponding proteins in this study we performed Western blot experiments for RBCL protein. We included in the analysis embryo axes isolated from immature holm oak seeds collected at the fifth, sixth and seventh month after pollination. The immunoblotting results (Fig. 5.6.B) validate our SSH data, though the magnitude of the changes was higher at the mRNA level. There might be methodological and biological reasons for such discrepancies, including differences in the protein and mRNA degradation rates, protein translation efficiencies, and/or the steady-state mRNA and protein abundances¹⁶. Figure 5.6.B shows that the RBCL protein is accumulated during the seed maturation process, picking at the S0 stage. After imbibition, the amount of protein decreased, in agreement with the transcript profile, to pick up again in the seedling. The in-gel SOD activity staining in Fig. 5.6.C indicated a diminution in the levels of this antioxidant enzyme running in parallel with the decrease in the *Sod1* transcript amounts quantified along the seed development. Many authors have stated that correlation between mRNA and protein abundances in the cell is notoriously poor. Clearly, transcription and translation are far from having a linear and simple relationship and the transcript/protein discordance represents a critical layer of regulatory processes at the post-transcriptional level that is often neglected¹⁷⁻¹⁸. However, data presented here (Fig. 5.6) sustain that transcript levels, at least for SOD and RBCL, in our experimental system, provide good predictive value with respect to the extent of protein changes in abundance.

5.3.3. Biological functions of proteins encoded by differential expressed genes identified by SSH methodology

Functional classification of genes expressed in germinated seeds (reverse library)

Thirty-one over-expressed genes were identified in germinated seeds at the S3 stage. Proteins encoded by these genes are representative of eight functional categories: stress responses (3 genes, 50 clones), transport (4 genes, 7 clones), oxidation-reduction (3 genes, 4

clones), cell wall modification (2 genes, 3 clones) cell division cycle (1 gen, 1 clon), protein metabolism (2 genes, 2 clones), cellular component organization (1 gen, 1 clon), translation (2 genes, 5 clones). Another five genes were placed into a miscellanea category (Table 5.3).

The *stress response* group holds the *Lea-5* gene, the highest represented in the SSH reverse library, with 43 clones (Table 5.3). *Lea-5* (Late embryogenesis abundant) is an osmoprotective protein described to accumulate during seed maturation as part of the mechanism of seeds desiccation tolerance¹⁹. The DHN3 protein discussed in Ch.4 is also included in this family. In view of the amount of *Lea-5* transcripts accumulated in S3 (Fig. 5.6.A), it can be assumed that holm oak acorns are not fully recalcitrant seeds, but share part of the orthodox seeds characteristics, in agreement with the data obtained and discussed in Ch.4 and 6 of this Thesis. Enolase (ENO) is a ubiquitous enzyme that catalyzes the conversion of 2-phosphoglycerate to phosphoenolpyruvate, occurring at the end of glycolysis. Enolases are needed in the first stages of embryogenesis and seed formation and has been described as one of the most abundant proteins in many matured stages. As a glycolytic enzyme, is not easy to explain neither its abundance in quiescent seeds, nor its diminution in the stage SS-4, where the metabolism is fully active to supply the energy needed for the shoot growth. Several authors have reported that enolases lost a central sequence parts essential for enolase activity at the end of the seed maturation, becoming and behaving as storage proteins. These “small enolases” are highly abundant in storage tissues like zygotic embryos and endosperm, but are missing in the germinating embryos²⁰. In the *stress response* group was also included the gene *Chi11*, coding for chitinase. Chitinases are hydrolytic enzymes found in different organisms. These enzymes are present in various plant tissues and specific chitinases accumulate in seeds of many species as part of their normal developmental programme. Several authors suggest that chitinases are responsible for the protection of seeds against pathogens during the germination period, especially at the time that the rupture of the testa occurs and the radicle emerges²¹.

The proteins CCH and ERD6 are included, among other, in the *transport* category. The sugar transporter ERD6 is implicated in the high fluxes of sucrose that occur in seeds when storage products are accumulated or remobilized during development and germination, respectively²². As described for other plants, ERD6 could be implicated in the transport of sucrose from cotyledon to embryo axis in germinated embryo axis of *Q. ilex*. In plants, copper (Cu) is a cofactor for plastocyanin, Cu/Zn SOD, cytochrome-c oxidase, and the ethylene receptors for several apoplasmic oxidases. Copper (Cu) chaperones constitute a family of small Cu⁺-binding proteins required for Cu trafficking and the prevention of cytoplasmic exposure to copper ions in transit²³. One of the copper network components in *Arabidopsis thaliana* is the copper chaperone CCH. CCH has been associated with the maintaining of protein conformation

during the very active period of plant development that is seed dormancy breaking and germination²⁴.

The *oxidation-reduction* category includes the *Sod1* and the *NdhF* genes. Accumulation of *Sod1* transcripts (Fig. 5.6.A) in mature seeds can be seen also as a desiccation protective mechanism, as it produces high levels of oxygen reactivities species (ROS)²⁵. The decrease in *Sod1* transcripts and activity (Fig.5.6.B) observed along the developmental process can be explained because ROS have important roles in endosperm weakening, mobilization of seed reserves, and many other processes during seed germination and vigorous seedling²⁶⁻²⁸. Some Cu chaperones inserts copper into Cu/Zn-SOD. Interestingly, the transcripts of *Cch* and *Sod1* gene, coding for the *Q. ilex* Cu/Zn-SOD enzyme, were highly abundant in mature or germinating seeds, but scarce in seedlings (Table 5.3; Fig. 5.6.A). The *NdhF* transcript numbers peaked early during acorns germination, coinciding with the rupture of the testa and pericarp, and dropped later to the levels found in the ungerminated seed as it reached the stage when both the coleoptile and radicle had clearly elongated. The ordered increase in the abundance of *NdhF* and other mitochondrial proteins during germination and the posterior decrease in the early seedling, has been reported for *Arabidopsis* during the maturation of the mitochondria that occurs after imbibition²⁹.

Among the genes in the *cell wall modification* category, beta-glucosidase 29-like and endoglucanase 6-like are involved in the degradation of cellobiose, essential for cell elongation and radicle extension in the seed during germination³⁰⁻³². Two genes, carboxypeptidase-like (Ser-CPs) and proteasome subunit beta type 6 are linked to protein metabolism, implicated in the breakdown of the storage protein³³.

In addition, we identify in germinating seeds genes related to *cell division cycle* (histone), and to *traslation* and synthesis of proteins (ribosomal protein S12, S21 and arabinogalactan peptide 20-like). Some of the ribosomal protein coding genes mRNAs could in mature seeds of *Q. ilex* as it has been reported in the literature for other species³⁴.

Functional classification of genes expressed in shoot seedlings

The 39 non-redundant transcripts over-represented in *Q. ilex* shoots seedlings were grouped in seven functional categories: photosynthesis (10 genes, 30 clones), secondary metabolism (4 genes, 7 clones), transport (2 genes, 2 clones), signaling (4 genes, 7 clones), stress response (2 genes, 2 clones), gene expression (1 gen, 3 clones) and cellular component organization (1 gen, 1 clon). An additional category included genes with proteins of unknown function (Table 5.4).

Ten genes related with light energy conversion to chemical energy were over-represented in the forward library that correspond to shoots seedlings. Among these were included genes coding for several components of the photosynthetic apparatus (chlorophyll a-b binding proteins-CPBs, plastocyanins-Pet, plastoquinone oxidoreductase-Ndhk) and CO₂ assimilation (large and small subunits of RuBisCo-*RbcL*, *RbcS*). Successful seedling establishment after germination requires efficient utilization of endogenous storage reserves but also resources from the environment, adapting both the developmental and the metabolic programs³⁵. Up-regulation of this group of genes clearly showed the normal transition from the heterotrophic metabolism present in germination and post-germination stages to photoautotrophic growth in young seedling. The transcriptional profiles of *RbcL*, *RbcS* and *PetE*, are shown in Fig.5.6. For the three genes, a dramatic increase was observed in the transition from S7, when the plumule emerges from cotyledonary petioles, to SS-4, when a group of primordial green leaves is already clearly observed. Transcripts of genes associated with photoautotrophic growth, including RuBisCO subunits are expected to be present only at a very low level dry seeds, and be induced in postgermination. That is the case for *RbcS* and *PetE* (Fig. 5.6.A). However, *RbcL* protein is accumulated as the acorn matured, becoming abundant at both the level of transcripts and proteins in mature acorns (Fig. 5.6.A, B). The *RbcL* levels diminished after imbibition, to increase later to numbers that justify RBCL being the most abundant protein in mature fully expanded leaves. In higher plants, RuBisCo is composed of eight small subunits, encoded by a nuclear multigene family (*RbcS*) and eight large subunits, encoded by a single gene (*RbcL*) in the chloroplast genome. For synthesis of the Rubisco holoenzyme, both genes need to be expressed coordinately. However, it has been considered that expression of *RbcL* is modulated at the level of its translation for the coordinated expression between *RbcL* and *RbcS*³⁶.

In the group of *secondary metabolism* was *Cybr5*, the gene coding the NADH-cytochrome B5 reductase, the major electron-transfer enzyme involved in desaturation of fatty acids and sterol precursors³⁷. The enzyme ACO (1-aminocyclopropane-1-carboxylic acid oxidase) catalyzes the final step of ethylene biosynthesis³⁸. As discussed elsewhere, ethylene is a phytohormone implicated in a wide range of plant responses and developmental steps, promotion of seedling shoot growth, as reported for barley, oats and rice (Locke *et al.*³⁹ and refs. therein). The *Xth* gene codes for xyloglucan endotransglycosylase, an enzyme involved in the biosynthesis of the primary cell walls of dicotyledonous plants⁴⁰. Several studies have described the expression of *Xth* genes during cell growth, differentiation and cell expansion⁴¹⁻⁴².

A third category with interesting genes is *signaling*, where the *Pp2c* gene is included. As shown in Fig. 5.6. *Pp2c* mRNAs maintained constant counts during germination, and

significantly decreased after radicle emergence (S7 stage). In the previous Ch.4, we found that the ABA level start to raise from this stage on, to collaborate in growth progression. As a negative inhibitor of ABA, but under the regulation of ABA⁴³, downregulation of *Pp2c* gene at S7 would facilitate the role of ABA at early stages of seedling development, what in change promote the *Pp2c* overexpression from the SS-4 stage.

Finally, genes involved in *sugar transport (Mst2)*, *stress (MT1, Gst) biosynthesis of the nuclear envelope (Mlp1)*, the *regulation of gene expression (Lug)*, the *organization of cellular components (AdF)* were up-regulated during seedling growth.

As the sequences of all the genes described here are from now available at public databases, further studies could be implemented to uncover the role of these genes in the *Q. ilex* seedling establishment.

5.4. Concluding remarks

The objective of the present study were to identify and quantify the expression of genes of *Q. ilex* related to germination and seedling development by using suppression subtractive hybridization (SSH) and to analyze the transcriptional profile of some selected genes during this process by qRT-PCR. SSH is an efficient technique for generating cDNAs enriched for differentially expressed genes when the studied species is underrepresented in sequence databases.

Here we have identified for the first time a large number of putative differentially expressed ESTs from the embryo axis in germinated seeds and from shoot seedlings of *Quercus ilex* during the postgermination and seedling establishment. Additionally, we have provided the first absolute (molecule number) quantitative analysis by real-time RT-PCR of transcription profiles displayed by eight of the genes revealed by SSH. Our data constitute an important genomics resource that should clearly benefit further germination and other biological process research on *Q. ilex*, given that this economically important forest tree species remains largely unexplored at the genomic level.

5.5. Appendixes

Table 5.3: Classification based in Biological Process of Gene Ontology (Reverse library)

Gene Symbol	Putative Identity	GeneBank	Size (bp)	E-value	Frame	Closest species	Accession No.	No. of clones
Stress responses								
<i>Eno</i>	2-phospho-D-glycerate hydrolase	JZ794741	236	1.00E-07	+2/+3	<i>Prunus armeniaca</i>	AY958170	1
<i>Chi1</i>	Chitinase	JZ794739	228	6.00E-09	+3/+3	<i>Castanea sativa</i>	CSU48687	5
<i>Lea-5</i>	Group 5 late embryogenesis abundant protein	JZ794754	228	2.00E-04	+3/+2	<i>Citrus unshiu</i>	DQ424891	43
<i>Lea-5</i>	Group 5 late embryogenesis abundant protein	JZ794761	179	8.00E-06	+2/+2	<i>Fragaria vesca subsp. Vesca</i>	XP_004296856	1
Transport								
<i>Sec61G</i>	Protein transport Sec61 subunit gamma-like (LOC100827752)	JZ794747	208	6.00E-28	Plus/Plus	<i>Brachypodium distachyon</i>	XM_003563327	2
<i>ArF</i>	ADP-ribosylation factor, transcript variant 3, 3'UTR	JZ794760	186	5.00E-04	Plus/Plus	<i>Glycine max</i>	XM_003553886	1
<i>Cch</i>	Copper chaperone, 5'UTR	JZ794748	411	2.00E-40	+1/+2	<i>Populus alba x Populus tremula var. glandulosa</i>	AY603358	1
<i>ERD6</i>	Sugar transporter ERD6-like 16-like (LOC100816850)	JZ794749	410	3.00E-09	+2/+3	<i>Cicer arietinum</i>	XM_004508267	2
	Sugar transporter ERD6-like 16-like	JZ794750	346	3.00E-09	Plus/Plus	<i>Cicer arietinum</i>	XM_004508267	1
Oxidation-reduction								
<i>Sod1</i>	Partial <i>Sod1</i> mRNA for superoxide dismutase, Cu/Zn	KM262658	225	3.00E-22	+2/+1	<i>Fagus sylvatica</i>	AJ586519	1
<i>NdhF</i>	NADH dehydrogenase subunit F	JZ794738	312	8.00E-44	+2/+2	<i>Castanea seguinii</i>	AY586354	2
<i>RCOM_0593170</i>	NADH dehydrogenase FAD NAD binding oxidoreductases, putative, mRNA	JZ794759	104	3.00E-06	+3/+2	<i>Ricinus communis</i>	XM_002526853	1
Cell wall modification								
<i>Bglu</i>	Beta-glucosidase 29-like	JZ794742	301	8.00E-03	+3/+3	<i>Fragaria vesca subsp. vesca</i>	XM_004305875	1
<i>Cel8</i>	endoglucanase 6-like, 3'UTR	JZ794740	263	8.00E-53	Plus/Plus	<i>Vitis vinifera</i>	XM_002270844	2
Cell division cycle								

<i>Gene Symbol</i>	<i>Putative Identity</i>	<i>GeneBank</i>	<i>Size (bp)</i>	<i>E-value</i>	<i>Frame</i>	<i>Closest species</i>	<i>Accession No.</i>	<i>No. of clones</i>
<i>Hta909</i>	Histone 2	JZ794743	223	9.00E-40	+2/+3	<i>Populus trichocarpa</i>	XM_002328499	1
Protein metabolism								
<i>ScpL</i>	Serine carboxypeptidase-like 18-like (LOC101310608)	JZ794744	272	3.00E-12	+3/+2	<i>Fragaria vesca subsp. Vesca</i>	XM_004299733	1
<i>PsmB</i>	Proteasome subunit beta type 6	JZ794745	257	4.00E-19	+1/+1	<i>Malus x domestica</i>	XM_002527949	1
Cellular component organization								
<i>Pme</i>	Pectinesterase-like	JZ794746	222	7.00E-21	+2/+2	<i>Nicotiana tomentosiformis</i>	XM_009627181	1
Translation								
<i>RpS12</i>	Ribosomal protein S12, 3'UTR	JZ794751	259	1.00E-127	Plus/Plus	<i>Quercus nigra</i>	HQ664601	3
<i>RpS21</i>	40S ribosomal protein S21-2-like, transcript variant 2	JZ794752	227	1.00E-10	+2/+2	<i>Cucumis sativus</i>	XM_004171568	2
Miscellaneous								
<i>AgP20</i>	Arabinogalactan peptide 20-like	JZ794753	288	4.00E-09	+2/+1	<i>Fragaria vesca subsp. Vesca</i>	XM_004287458	1
<i>Fgfr10P</i>	FGFR1 oncogene partner(RCOM_1071630)	JZ794762	95	1.00E-09	+3/+2	<i>Ricinus communis</i>	XM_002523887	2
<i>RCOM_0710110</i>	21 kDa protein precursor	JZ794763	159	2.00E-05	+1/+1	<i>Malus x domestica</i>	XM_008393501	1
<i>Sp1L1</i>	Protein SPIRAL1-like 1-like, transcript variant 2	JZ794764	183	9.00E-12	+1/+1	<i>Fragaria vesca subsp. vesca</i>	XM_004297177	1
<i>Ag13</i>	mRNA for Ag 13 protein	JZ794765	196	3.00E-19	Plus/Plus	<i>Alnus glutinosa</i>	Y08435	1
Unknown function								
	Hypothetical protein	JZ794755	336	1.00E-10	+1/+1	<i>Populus trichocarpa</i>	XM_002309824	1
	Hypothetical protein (PRUPE_ppa012024mg) mRNA, complete cds	JZ794766	199	4.00E-07	Plus/Plus	<i>Prunus persica</i>	XM_007209589	1
	Hypothetical protein (PRUPE_ppa011596mg) mRNA, complete cds	JZ794756	242	7.00E-09	Plus/Plus	<i>Prunus persica</i>	XM_007202528	1
No sequence hits nr NCBI database								
	EST expressed in germinated acorn of <i>Quercus ilex</i> (Clone R2_74)	JZ794757	224					1
	EST expressed in germinated acorn of	JZ794758	219					1

<i>Gene Symbol</i>	<i>Putative Identity</i>	<i>GeneBank</i>	<i>Size (bp)</i>	<i>E-value</i>	<i>Frame</i>	<i>Closest species</i>	<i>Accession No.</i>	<i>No. of clones</i>
	<i>Quercus ilex</i> (Clone R2_211)							
	EST expressed in germinated acorn of <i>Quercus ilex</i> (Clone R2_10)	JZ794767	162					1
	EST expressed in germinated acorn of <i>Quercus ilex</i> (Clone R2_399)	JZ794768	157					1
	EST expressed in germinated acorn of <i>Quercus ilex</i> (Clone R2_403)	JZ794769	163					1
	EST expressed in germinated acorn of <i>Quercus ilex</i> (Clone R1T76)	JZ794770	105					1
	EST expressed in germinated acorn of <i>Quercus ilex</i> (Clone R1T31)	JZ794771	132					1
	EST expressed in germinated acorn of <i>Quercus ilex</i> (Clone R1T37)	JZ794772	94					1
	EST expressed in germinated acorn of <i>Quercus ilex</i> (Clone R2_32)	JZ794773	122					1
	EST expressed in germinated acorn of <i>Quercus ilex</i> (Clone R2_374)	JZ794774	176					1

Table 5.4: Classification based in Biological Process of Gene Ontology (Forward library)

<i>Gene Symbol</i>	<i>Putative Identity</i>	<i>GeneBank</i>	<i>Size (bp)</i>	<i>E-value</i>	<i>Frame</i>	<i>Close species</i>	<i>Accession No.</i>	<i>No. of clones</i>
Photosynthesis								
<i>ABC</i>	Chlorophyll a/b binding protein	JZ794776	454	6E-772	+1/+1	<i>Carya cathayensis</i>	DQ471302	1
<i>Cpb</i>	Chlorophyll a-b binding protein	JZ794777	260	2E-52	0.5	<i>Citrus sinensis</i>	XM_006467448	3
<i>Cpb26</i>	Chlorophyll a-b binding protein CP26	JZ794778	294	2E-50	+2/+2	<i>Glycine max</i>	XM_003536141	2
<i>Cpb</i>	Chlorophyll a-b binding protein 151, chloroplastic-like (LOC101212963)	JZ794779	360	2E-41	+2/+1	<i>Cucumis sativus</i>	XM_004144495	1
<i>PsA</i>	Photosystem I reaction center subunit II, chloroplastic-like	JZ794780	259	4E-35	+3/+3	<i>Vitis vinifera</i>	XM_003631941	1
<i>PetE</i>	Plastocyanin a	JZ794781	257	1E-20	+2/+2	<i>Populus nigra</i>	Z50185	1
<i>PetA</i>	<i>Quercus rubra</i> plastid	JZ794812	165	7E-78	Plus/Plus	<i>Quercus rubra</i>	JX970937	1
<i>NdhK</i>	NADH-plastoquinone oxidoreductase sub. K	JZ794775	210	1E-40	+1/+1	<i>Quercus nigra</i>	GQ998672	2
<i>RbcS</i>	Ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit	JZ794783	220	2E-33	+3/+2	<i>Eucalyptus globulus</i>	AB537499	17
<i>RbcL</i>	Rubisco large subunit pseudogene	JZ794813	156	0.005	+2/+3	<i>Orobanche cumana</i>	AF090350	1
Secondary metabolism								
<i>Cyb5R</i>	NADH-cytochrome B5 reductase	JZ794782	266	6E-06	Plus/Plus	<i>Populus trichocarpa</i>	XM_002328054	1
<i>Aco1</i>	1-aminocyclopropane-1-carboxylic acid oxidase	JZ794786	298	2E-15	+2/+1	<i>Betula luminifera</i>	HM357150	3
<i>Cpx</i>	Coproporphyrinogen-III oxidase	JZ794787	322	1E-27	+1/+3	<i>Brassica rapa</i>	XM_009121288	1
<i>Xth</i>	Xyloglucan endotransglycosylase/hydrolase precursor	JZ794784	223	7E-35	+2/+1	<i>Populus tremula</i> <i>x Populus tremuloides</i>	EF194050	2
Transport								
<i>Mst2</i>	Monosaccharide transporter	JZ794789	332	4E-07	Plus/plus	<i>Theobroma cacao</i>	XM_007025749	1
<i>ERD6</i>	Sugar transporter ERD6-like	JZ794790	308	5E-10	Plus/plus	<i>Citrus sinensis</i>	XM_006471810	1

<i>Gene Symbol</i>	<i>Putative Identity</i>	<i>GeneBank</i>	<i>Size (bp)</i>	<i>E-value</i>	<i>Frame</i>	<i>Close species</i>	<i>Accession No.</i>	<i>No. of clones</i>
Signaling								
<i>Lrr-rlKs</i>	Probable inactive leucine-rich repeat receptor-like protein kinase At3g03770-like	JZ794791	340	1E-29	+2/+2	<i>Cicer arietinum</i>	XM_004493152	1
<i>Pp2c</i>	Protein phosphatase-2c, 3'UTR	JZ794816	131	5E-19	Plus/Plus	<i>Populus trichocarpa</i>	XM_006389316	1
<i>Pp7L</i>	Serine/threonine-protein phosphatase 7 long form homolog mRNA, complete cds	JZ794792	358	7E-15	+1/+2	<i>Glycine max</i>	XM_006598404	1
<i>AcpL1</i>	Acid phosphatase 1-like (LOC100853464), mRNA	JZ794817	145	4E-12	+2/+1	<i>Eucalyptus grandis</i>	XM_010035523	4
Stress response								
-	Wound-responsive protein 15.46 mRNA, 3'-UTR	JZ794788	278	3E-98	Plus/Plus	<i>Castanea sativa</i>	AY055745	1
<i>Mt1</i>	Partial mRNA for metallothionein-like protein	JZ794815	191	1E-29	+2/+2	<i>Quercus robur</i>	AJ577263	1
Gene expression								
<i>Lug</i>	Transcriptional corepressor LEUNIG-like	JZ794785	367	7E-32	+2/+3	<i>Prunus mume</i>	XM_008221647	3
Cellular component organization								
<i>AdF</i>	Actin-depolymerizing factor 1-like, 3' UTR	JZ794814	183	5E-09	Plus/plus	<i>Vitis vinifera</i>	XM_002273922	1
Unknown function								
	Contig VV78X026477.7, whole genome shotgun sequence	JZ794793	213	4E-10	Plus/Plus	<i>Vitis vinifera</i>	AM431847	1
	Clone QsuP5.2 sequence	JZ794818	120	8.00E-23	Plus/Plus	<i>Quercus suber</i>	AF281042	1
	Contig VV78X240868.24, whole genome shotgun sequence/ <i>Vitis vinifera</i> contig VV78X240868.24	JZ794794	204	6E-34	Plus/Plus	<i>Vitis vinifera</i>	AM480217	1
	Contig VV78X122321.6, whole genome shotgun sequence, 3'UTR	JZ794795	250	4E-99	Plus/Plus	<i>Vitis vinifera</i>	AM460086	1
<i>LOC101210359</i>	Uncharacterized LOC101210359	JZ794796	322	5E-45	+3/+1	<i>Cucumis sativus</i>	XM_004142551	1
	Whole genome shotgun sequence, contig VV78X269505.8, clone ENTAV 115	JZ794797	627	2E-43	+1/-1	<i>Vitis vinifera</i>	AM486441	1
	Clone SS0ACG1YJ14	JZ794819	184	9E-12	Plus/plus	<i>Vitis vinifera</i>	FQ381038	1

<i>Gene Symbol</i>	<i>Putative Identity</i>	<i>GeneBank</i>	<i>Size (bp)</i>	<i>E-value</i>	<i>Frame</i>	<i>Close species</i>	<i>Accession No.</i>	<i>No. of clones</i>
	Hypothetical protein (PRUPE_ppa007296mg) mRNA	JZ794799	255	4E-12	Plus/plus	<i>Prunus persica</i>	XM_007223248	1
<i>No sequence hits nr NCBI database</i>								
	EST expressed <i>Q. ilex</i> shoot (Clone F2_240)	JZ794798	340					1
	EST expressed <i>Q. ilex</i> shoot (Clone F2_59_129)	JZ794800	231					2
	EST expressed <i>Q. ilex</i> shoot (Clone F2_105)	JZ794801	202					1
	EST expressed <i>Q. ilex</i> shoot (Clone F2_259)	JZ794802	264					1
	EST expressed <i>Q. ilex</i> shoot (Clone F2_4)	JZ794820	168					1
	EST expressed <i>Q. ilex</i> shoot (Clone F2_38)	JZ794803	364					1
	EST expressed <i>Q. ilex</i> shoot (Clone F2_40)	JZ794804	309					1
	EST expressed <i>Q. ilex</i> shoot (Clone F2_46)	JZ794805	209					1
	EST expressed <i>Q. ilex</i> shoot (Clone F2_52)	JZ794821	173					1
	EST expressed <i>Q. ilex</i> shoot (Clone F2_57)	JZ794806	461					1
	EST expressed <i>Q. ilex</i> shoot (Clone F2_74)	JZ794807	248					1
	EST expressed <i>Q. ilex</i> shoot (Clone F2_84)	JZ794808	237					1
	EST expressed <i>Q. ilex</i> shoot (Clone F2_102)	JZ794822	124					1
	EST expressed <i>Q. ilex</i> shoot (Clone F2_108)	JZ794823	146					1
	EST expressed <i>Q. ilex</i> shoot (Clone F2_110)	JZ794824	139					1
	EST expressed <i>Q. ilex</i> shoot (Clone F2_122)	JZ794809	272					1
	EST expressed <i>Q. ilex</i> shoot (Clone F2_148)	JZ794810	260					1
	EST expressed <i>Q. ilex</i> shoot (Clone F1T49)	JZ794825	124					1
	EST expressed <i>Q. ilex</i> shoot (Clone F2_13)	JZ794811	238					1

Table 5.5: Sequences without hits in the NCBI-nt database

GeneBank	OBS	BLASTn				BLASTx Annotation				
		Database	E-value	% Identity	Accession No.	Closest species	Putative identity	E-value	Closest species	Accession No.
Reverse Library										
JZ794757		EST_NCBI	1.00E-70	89	FN700828	<i>Quercus robur</i>	Sugar transporter ERD6-like 16	2.00E-20	<i>Prunus mume</i>	XP_008223080
JZ794758		EST_NCBI	1.00E-90	95	FN699270	<i>Quercus robur</i>	Sugar transporter ERD6-like 16	2.00E-20	<i>Prunus mume</i>	XP_008223080
JZ794767		EST_NCBI	7E-39	91	FN748440	<i>Quercus petraea</i>		No hits		
JZ794768						No hits		No hits		
JZ794769		WGS_NCBI	1E-72	98	GAOS01046715	<i>Notholithocarpus densiflorus</i>	Gibberellin-regulated protein 6	4.00E-49	<i>Vitis vinifera</i>	XP_002284937
JZ794770	3'UTR	WGS_NCBI	9.00E-40	97	GAOS01005093	<i>Notholithocarpus densiflorus</i>	Proteasome subunit alpha type-1-A-like	1.00E-164	<i>Malus domestica</i>	XP_008388039
JZ794771		WGS_NCBI	5.00E-26	99	GAOS01014527	<i>Notholithocarpus densiflorus</i>	Basic blue copper family protein	1.00E-65	<i>Populus trichocarpa</i>	XP_002298184
JZ794772		WGS_NCBI	3.00E-38	100	GAOS01004068	<i>Notholithocarpus densiflorus</i>	40S ribosomal protein S4-like	2.00E-118	<i>Vitis vinifera</i>	XP_002276986
JZ794773		WGS_NCBI	5.00E-25	86	JRKL01012822	<i>Castanea mollissima</i>		No hits		
JZ794774		WGS_NCBI	7.00E-70	94	GAOS01034165	<i>Notholithocarpus densiflorus</i>	Monoglyceride lipase-like	0	<i>Fragaria vesca subsp. Vesca</i>	XP_004291688
Forward Library										
JZ794798		WGS_NCBI	6.00E-49	84	JRKL01181633	<i>Castanea mollissima</i>		No hits		
JZ794800						No hits		No hits		
JZ794801		EST_NCBI	2E-28	80	FP025185	<i>Quercus robur</i>		No hits		
JZ794802		EST_NCBI	3E-34	79	FP025185	<i>Quercus robur</i>		No hits		
JZ794820		WGS_NCBI	1.00E-72	98	JRKL01000436	<i>Castanea mollissima</i>	Hypothetical protein PRUPE_ppa001266mg	1E-85	<i>Prunus persica</i>	XP_007207147

GeneBank	OBS	BLASn				BLASx Annotation				
		Database	E-value	% Identity	Accession No.	Closest species	Putative identity	E-value	Closest species	Accession No.
JZ794803		EST_NCBI	2E-100	96	EC993414	<i>Vitis vinifera</i>		No hits		
JZ794804		EST_NCBI	53-95	86	FP055967	<i>Quercus robur</i>		No hits		
JZ794805		WGS_NCBI	2E-97	99	GAOS01023571	<i>Notholithocarpus densiflorus</i>	Signal recognition particle protein	4E-63	<i>Medicago truncatula</i>	KEH18962
JZ794821		WGS_NCBI	3E-80	99	GAOS01002366	<i>Notholithocarpus densiflorus</i>	Uncharacterized protein LOC103336026, partial	0	<i>Prunus mume</i>	XP_008237287
JZ794806		EST_NCBI	0	96	FN760742	<i>Quercus petraea</i>		No hits		
JZ794807	3'UTR	WGS_NCBI	4E-113	97	GAOS01023741	<i>Notholithocarpus densiflorus</i>	Protein MLP1	0	<i>Prunus mume</i>	XP_008238878
JZ794808		EST_NCBI	3E-64	82	FR663529	<i>Quercus suber</i>		No hits		
JZ794822	3'UTR	EST_NCBI	4E-53	99	FP047535	<i>Quercus petraea</i>	Ubiquitin conjugating enzyme, partial	4.00E-45	<i>Nicotiana tomentosiformis</i>	XP_009604386
JZ794823	3'UTR	WGS_NCBI	2E-64	99	GAOS01007206	<i>Notholithocarpus densiflorus</i>	5'-nucleotidase surE	3E-153	<i>Morus notabilis</i>	EXB41281.1
JZ794824				No hits				No hits		
JZ794809				No hits				No hits		
JZ794810	5'UTR	WGS_NCBI	1E-113	96	GAOS01002875	<i>Notholithocarpus densiflorus</i>	Glutathione S-transferase-like	6E-110	<i>Prunus mume</i>	XP_008235033.1
JZ794825				No hits				No hits		
JZ794811				No hits				No hits		

5.6. Referencias

1. Abril N, Gion JM, Kerner R, Muller-Starck G, Cerrillo RM, Plomion C, Renaut J, Valledor L, Jorrin-Novo JV (2011) Proteomics research on forest trees, the most recalcitrant and orphan plant species. *Phytochemistry* 72: 1219-1242
2. Buckley BA (2007) Comparative environmental genomics in non-model species: using heterologous hybridization to DNA-based microarrays. *Journal of Experimental Botany* 210: 1602-1606
3. Abril N, Ruiz-Laguna J, García-Sevillano MA, Mata AM, Gómez-Ariza JL, Pueyo C (2014) Heterologous microarray analysis of transcriptome alterations in *Mus spretus* mice living in an industrial settlement. *Environ Sci Technol* 48: 2183-2192
4. Diatchenko L, Lau YF, Campbell AP, Chenchik A, Moqadam F, Huang B, Lukyanov S, Lukyanov K, Gurskaya N, Sverdlov ED, Siebert PD (1996) Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc Natl Acad Sci U S A* 93: 6025-6030
5. Echevarría-Zomeño S, Abril N, Ruiz-Laguna J, Jorrín-Novo J, Maldonado-Alconada A (2012) Simple, rapid and reliable methods to obtain high quality RNA and genomic DNA from *Quercus ilex* L. leaves suitable for molecular biology studies. *Acta Physiologiae Plantarum* 34: 793-805
6. Prieto-Alamo MJ, Abril N, Osuna-Jimenez I, Pueyo C (2009) Solea senegalensis genes responding to lipopolysaccharide and copper sulphate challenges: large-scale identification by suppression subtractive hybridization and absolute quantification of transcriptional profiles by real-time RT-PCR. *Aquat Toxicol* 91: 312-319
7. Sherwood AL (2003) Virtual elimination of false positives in blue-white colony screening. *Biotechniques* 34: 644-647
8. Romero-Rodríguez MC, Pascual J, Valledor L, Jorrin-Novo J (2014) Improving the quality of protein identification in non-model species. Characterization of *Quercus ilex* seed and *Pinus radiata* needle proteomes by using SEQUEST and custom databases. *Journal of Proteomics* 105: 85-91
9. Pueyo C, Jurado J, Prieto-Alamo MJ, Monje-Casas F, Lopez-Barea J (2002) Multiplex reverse transcription-polymerase chain reaction for determining transcriptional regulation of thioredoxin and glutaredoxin pathways. *Methods Enzymol* 347: 441-451
10. Jurado J, Prieto-Alamo MJ, Madrid-Risquez J, Pueyo C (2003) Absolute gene expression patterns of thioredoxin and glutaredoxin redox systems in mouse. *J Biol Chem* 278: 45546-45554
11. Kremer A, Abbott A, Carlson J, Manos P, Plomion C, Sisco P, Staton M, Ueno S, Vendramin G (2012) Genomics of Fagaceae. *Tree Genetics & Genomes* 8: 583-610
12. Derory J, Leger P, Garcia V, Schaeffer J, Hauser MT, Salin F, Luschnig C, Plomion C, Glossl J, Kremer A (2006) Transcriptome analysis of bud burst in sessile oak (*Quercus petraea*). *New Phytologist* 170: 723-738
13. Šunderlíková V, Salaj J, Kopecký D, Salaj T, Wilhem E, Matušíková I (2009) Dehydrin genes and their expression in recalcitrant oak (*Quercus robur*) embryos. *Plant Cell Reports* 28: 1011-1021
14. Vornam B, Gailing O, Derory J, Plomion C, Kremer A, Finkeldey R (2011) Characterisation and natural variation of a dehydrin gene in *Quercus petraea* (Matt.) Liebl. *Plant Biol (Stuttg)* 13: 881-887
15. Caudill SP (2010) Characterizing populations of individuals using pooled samples. *J Expo Sci Environ Epidemiol* 20: 29-37
16. Guo Y, Xiao P, Lei S, Deng F, Xiao GG, Liu Y, Chen X, Li L, Wu S, Chen Y, Jiang H, Tan L, Xie J, Zhu X, Liang S, Deng H (2008) How is mRNA expression predictive for protein expression? A correlation study on human circulating monocytes. *Acta Biochim Biophys Sin (Shanghai)* 40: 426-436
17. Maier T, Güell M, Serrano L (2009) Correlation of mRNA and protein in complex biological samples. *FEBS Letters* 583: 3966-3973
18. Vélez-Bermúdez IC, Schmidt W (2014) The conundrum of discordant protein and mRNA expression. Are plants special? *Frontiers Plant Sciences* 5
19. Liu Y, Wang L, Jiang S, Pan J, Cai G, Li D (2014) Group 5 LEA protein, ZmLEA5C, enhance tolerance to osmotic and low temperature stresses in transgenic tobacco and yeast. *Plant Physiology and Biochemistry* 84: 22-31

20. Rode C, Gallien S, Heintz D, Dorsselaer A, Braun H-P, Winkelmann T (2011) Enolases: storage compounds in seeds? Evidence from a proteomic comparison of zygotic and somatic embryos of *Cyclamen persicum* Mill. *Plant Molecular Biology* 75: 305-319
21. Santos IS, Oliveira AEA, Da Cunha M, Machado OLT, Neves-Ferreira AGC, Fernandes KVS, Carvalho AO, Perales Js, Gomes VM (2007) Expression of chitinase in *Adenantha pavonina* seedlings. *Physiologia Plantarum* 131: 80-88
22. Aoki N, Scofield GN, Wang X-D, Offler CE, Patrick JW, Furbank RT (2006) Pathway of sugar transport in germinating wheat seeds. *Plant Physiology* 141: 1255-1263
23. Pilon M, Abdel-Ghany SE, Cohu CM, Gogolin KA, Ye H (2006) Copper cofactor delivery in plant cells. *Current Opinion in Plant Biology* 9: 256-263
24. Pawlowski T (2009) Proteome analysis of Norway maple (*Acer platanoides* L.) seeds dormancy breaking and germination: influence of abscisic and gibberellic acids. *BMC Plant Biology* 9: 48
25. Rajjou L, Duval M, Gallardo K, Catusse J, Bally J, Job C, Job D (2012) Seed Germination and Vigor. *Annual Review of Plant Biology* 63: 507-533
26. Du D, Zhang Q, Cheng T, Pan H, Yang W, Sun L (2013) Genome-wide identification and analysis of late embryogenesis abundant (LEA) genes in *Prunus mume*. *Molecular Biology Reports* 40: 1937-1946
27. Morohashi Y (2002) Peroxidase activity develops in the micropylar endosperm of tomato seeds prior to radicle protrusion. *Journal of Experimental Botany* 53: 1643-1650
28. Dogra V, Ahuja PS, Sreenivasulu Y (2013) Change in protein content during seed germination of a high altitude plant *Podophyllum hexandrum* Royle. *Journal of Proteomics* 78: 26-38
29. Howell KA, Millar AH, Whelan J (2006) Ordered assembly of mitochondria during rice germination begins with pro-mitochondrial structures rich in components of the protein import apparatus. *Plant Mol Biol* 60: 201-223
30. Ketudat Cairns J, Esen A (2010) β -Glucosidases. *Cellular and Molecular Life Sciences* 67: 3389-3405
31. Morant AV, Jørgensen K, Jørgensen C, Paquette SM, Sánchez-Pérez R, Møller BL, Bak S (2008) β -Glucosidases as detonators of plant chemical defense. *Phytochemistry* 69: 1795-1813
32. Gallardo K, Job C, Groot SPC, Puype M, Demol H, Vandekerckhove J, Job D (2002) Proteomics of Arabidopsis seed germination. A comparative study of wild-type and gibberellin-deficient seeds. *Plant Physiology* 129: 823-837
33. Dal Degan F, Rocher A, Cameron-Mills V, von Wettstein D (1994) The expression of serine carboxypeptidases during maturation and germination of the barley grain. *Proceedings of the National Academy of Sciences* 91: 8209-8213
34. Koornneef M, Bentsink L, Hilhorst H (2002) Seed dormancy and germination. *Current Opinion in Plant Biology* 5: 33-36
35. Martin T, Oswald O, Graham IA (2002) Arabidopsis seedling growth, storage lipid mobilization, and photosynthetic gene expression are regulated by carbon: nitrogen availability. *Plant Physiology* 128: 472-481
36. Suzuki Y, Makino A (2013) Translational downregulation of RBCL is operative in the coordinated expression of Rubisco genes in senescent leaves in rice. *Journal of Experimental Botany* 64: 1145-1152
37. Fukuchi-Mizutani M, Mizutani M, Tanaka Y, Kusumi T, Ohta D (1999) Microsomal electron transfer in higher plants: cloning and heterologous expression of NADH-Cytochrome b 5 reductase from Arabidopsis. *Plant Physiology* 119: 353-362
38. Wang KL-C, Li H, Ecker JR (2002) Ethylene biosynthesis and signaling networks. *The Plant Cell Online* 14: S131-S151
39. Locke JM, Bryce JH, Morris PC (2000) Contrasting effects of ethylene perception and biosynthesis inhibitors on germination and seedling growth of barley (*Hordeum vulgare* L.). *Journal of Experimental Botany* 51: 1843-1849
40. Stratilová E, Ait-Mohand F, Řehulka P, Garajová S, Flodrová D, Řehulková H, Farkaš V (2010) Xyloglucan endotransglycosylases (XETs) from germinating nasturtium (*Tropaeolum majus*) seeds: Isolation and characterization of the major form. *Plant Physiology and Biochemistry* 48: 207-215
41. Imoto K, Yokoyama R, Nishitani K (2005) Comprehensive approach to genes involved in cell wall modifications in *Arabidopsis thaliana*. *Plant Molecular Biology* 58: 177-192

42. Rose JKC, Bennett AB (1999) Cooperative disassembly of the cellulose-xyloglucan network of plant cell walls: parallels between cell expansion and fruit ripening. *Trends in Plant Science* 4: 176-183
43. Hirayama T, Umezawa T (2010) The PP2C-SnRK2 complex: the central regulator of an abscisic acid signaling pathway. *Plant Signal Behav* 5: 160-163

CHAPTER 6:

CHANGES IN THE PROTEIN PROFILES OF *Q. ILEX* SEEDS DURING GERMINATION AND SEEDLING ESTABLISHMENT

Abstract

The changes that occur in the protein profiles during the germination and establishment of *Q. ilex* seedlings were investigated by means of a proteomic analysis of embryo axis, radicle and shoots tissues, excised at different stages along the growth process. PCA analysis of the normalised bands intensities obtained by SDS-PAGE clearly separated germination, postgermination and early seedling tissue proteomes. However, the band pattern was not greatly different between unimbibed and germinating seeds. This initial 1-DE proteomic approach was used for the selection of the samples (S0, S3, SS-1 and SS-4) to be analysed through a more resolutive technique, as it is 2-DE. A total of 732 spots were resolved with 2-DE, of which 103 variable spots were selected for protein identification. Some 90 differentially accumulated proteins were identified using 2-DE MALDI-TOF/TOF. The gel-based approach disclosed important metabolic changes that occurred in the holm oak seed after the germination (from S3 onward). Again, few proteins resulted altered in their abundance during the germination period (from S0 to S3). A gel free approach was, hence, used to analyse the mature unimbibed (S0) and the germinated seed (S3), trying to improve the coverage of proteome analysed. Through *n*LC Orbitrap LTQ analysis of total extract, 153 differentially accumulated proteins were identified. These two complementary approaches increased more than 2-fold the coverage the amount of analyzed proteome. Data suggested that the mature non-orthodox seeds of *Q. ilex* have the mechanisms necessary to ensuring the rapid resume of the metabolic activities requires to start the germination process and to *de novo* synthesise the biomolecules required for growth. Proteins related to energy metabolism and photosynthesis were up-accumulated during seedling establishment. Our results also indicated that the use of genus specific database combined with public database improve the quality and quantity of protein identification in orphan species. The diverse proteomic approaches we used gave similar clues about the metabolic state of the mature *Q. ilex* seed before the germination starts, and the metabolic switched experimented by the imbibed accord till the seedling is established. Data are in fully agreement with those obtained at the transcriptional level (Ch. 4 and Ch. 5), thereby strengthening each other.

6.1 Introduction

Germination is an important process in which a seed embryo develops into a seedling. It involves the activation of the metabolic pathways from physiologically quiescent status, and so leads to the initiation of growth by the emergence of the radicle and plumule or shoot. At the molecular level, germination implies a series of complex signal transduction pathways which cause changes in the gene expression profile ¹. Regulation of seed germination is one of the critical adaptive traits in plants in the long history of evolution. Optimal seed germination is a prerequisite for successful seedling establishment and plant vigour. The emergence of the radicle marks the end of the germination phase and beginning of the seedling establishment, the next phase in plant growth that ends when the seedling has exhausted the food reserves stored in the seed ². Further details of seed, germination and seedling establishment biology and physiology were presented in the general introduction of this Doctoral Thesis.

Proteomics is a global protein analysis strategy that provides information on a multitude of events in complex processes such as germination ³. Proteomic analyses have been conducted on germinating seeds of a number of species including the model species *Arabidopsis thaliana* ³⁻⁵ and *Medicago truncatula* ⁶, as well as on agronomically important species such as *Hordeum vulgare* ⁷⁻⁸, *Solanum lycopersicum* ⁹, *Oryza sativa* ¹⁰⁻¹¹, *Zea mays* ¹², *Triticum aestivum* ¹³, or *Glycine max* ¹⁴. Most of the plant species studied produce orthodox seeds. However the study of recalcitrant seeds is very limited, even in species of great interest such as forest trees. Attempts to understand the changes which occur in the proteome during the germination of woody plants have been carried out in *Havea brasiliensis* ¹⁵, *Fagus sylvatica* ¹⁶, *Prunus campanulata* ¹⁷, *Jatropha curcas* ¹⁸, *Acer platanoides* ¹⁹, *Phoenix dactylifera* ²⁰ and *Araucaria angustifolia* ²¹. These studies on the recalcitrant seeds germination demonstrated alterations in the abundance of specific functional classes of proteins associated to after-ripening, dormancy and germination. These include storage proteins, carbohydrate catabolism and biosynthesis, and stress-related proteins ²².

Our group has focused for years on the study of the genetic and physiological variability of plants, acorns and pollen ²³⁻²⁵ of *Quercus ilex*, a species with recalcitrant seeds, and its responses to various biotic and abiotic stresses ²⁶⁻³⁰ by using proteomic approaches. In the previous Ch. 4 and Chp. 5 we addressed the holm oak acorns germination and seedling establishment by transcriptional analysis (SSH and qRT-PCR). Given that proteins are direct executors of life processes and protein levels are major biomonitoring end points, we next investigated whether the changes in mRNA expression were reflected at the protein level. To this end, we initiated a new proteomic approach to unravel the mechanisms involved in seed development and germination of holm oak seeds to better understand the molecular and

physiological basis of embryo differentiation, development and germination. Proteomics can result of great utility in analyzing the dynamic of this complex biological process.

Proteomic studies utilizing intact seeds face the problem of the presence in the seed of highly abundant storage proteins (*i.e.*, legumin) that represent the primary limiting factor in the detection of other interesting but low-abundance proteins in proteomics analyses³¹⁻³³. The use of embryos dissected from germinating seeds facilitates the identification of many of these low-abundance key proteins involved in germination³⁴. In the present work, we used tissue samples excised from mature and germinating seeds (embryo axes), from acorns in the post-germinated stage (radicle) and from established seedlings (shoot, leaf primordia including the proximal tail tissues) to determine the changes originated by imbibition in the proteomic profiles by SDS-PAGE, 2-DE and LC-MS/MS. This approach has provided novel information about germination and seedling establishment of *Q. ilex* acorns. About 90 differentially accumulated proteins were identified using 2-DE MALDI-TOF/TOF and 153 proteins using *n*LC Orbitrap LTQ. These results provide a reference to analyze the regulation of seed germination and seedling establishment of *Q. ilex* in further studies.

6.2 Materials & Methods

6.2.1 Experimental design and plant material

Mature acorns were collected from a *Quercus ilex* population, Cerro Muriano-Cordoba (Fig. 3.1, p. 45). Germination and growth experiments were performed as described in the Ch. 3 (Fig 3.2, p. 46). The embryonic axis was removed from seeds at S0 to S3 stages and the whole seedling at S5 to S7 stages; seedling shoots of 1 (SS-1) and 4 cm length (SS-4) were also included (Fig. 3.3, p. 47). Samples from each time, washed, blot dried and frozen in liquid nitrogen, were pooled in three lots (1-2 g fresh weight) of embryo axes, radicles or seedling shoots per pool, and the three pools used as biological replicates. Samples were frozen in liquid nitrogen and stored at -80 °C until protein extraction.

6.2.2 Protein extraction

Samples were grounded to a fine powder using a mortar and a pestle. Three independent protein extractions were performed per experimental condition (one per biological replicate) by trichloroacetic acid (TCA)/acetone and phenol methods using 200 mg of the tissue powder as previously described^{25; 35-37}. Briefly, the powder was suspended in 2 mL of 10% (w/v) TCA /acetone solution and the mixture mixed by vortexing and sonicated 4 x 10 s (6 W). After centrifugation (14000g for 10 min) the pellet was recovered, washed with 2 mL of

cold (-20 °C) 80% (v/v) methanol containing 0.1 M ammonium acetate and mixed by vortexing. Centrifugation (14000g for 10 min) and washing were repeated with 2 mL of 80% (v/v) acetone. The pellet was air-dried and re-suspended in 0.6 mL of phenol (Tris-buffered, pH 8.0; Sigma St. Louis, MO, USA) and 0.6 mL of SDS buffer (30% [w/v] saccharose, 2% [w/v] SDS, 0.1 M Tris-HCl, pH 8.0, 5% [w/v] 2-mercaptoethanol). The mixture was vortexed thoroughly for 30 s, incubated on ice for 5 min and the phenol phase separated by centrifugation at 14000g for 10 min. The upper phenol phase was transferred to fresh 1.5 mL tubes. At least 5 volumes of cold 100% methanol containing 0.1 M ammonium acetate were added to the phenol phase and the mixture was stored at -20 °C overnight. Precipitated proteins were recovered at 10000g for 5 min, and then washed with cold methanol and later with cold 80% (v/v) acetone. The final pellet was suspended in 7 M urea, 2 M thiourea, 4% (w/v) CHAPS, 0.5% (w/v) Triton X-100 and 100 mM DTT. Insoluble material was removed by centrifugation at 14000g for 15 min. The protein concentration in the extract was determined by using the Bradford method³⁸⁻³⁹ with bovine serum albumin (BSA) as standard. Samples were frozen in liquid nitrogen and stored at -80 °C until analysis.

6.2.3 Gel electrophoresis, staining, image capture and analysis

For 1-DE analysis, 80 µg of protein BSA equivalent of the three biological replicates from each nine S0, S1, S2, S3, S5, S6, S7, SS-1 and SS-4 samples were separated by SDS-PAGE electrophoresis⁴⁰ on 12% polyacrylamide gels by using a Protean XL-II (20 x 20 cm) system (Bio-Rad). Gels were run at 80 V until the dye reached the bottom of the gel.

In the 2-DE experiment, the three biological replicates from stages S0, S3, SS-1 and SS-4 were analyzed. Previous work performed in the laboratory²⁴ was considered to select the working pH range to be used. Immobilised pH gradients (IPG) strips were employed (17 cm, 5-8 pH gradient; Bio-Rad). IPG strips were passively rehydrated for 16 h with 400 µg protein BSA equivalent in 400 µL of IEF solubilization buffer (7 M urea, 2 M thiourea, 4% [w/v] CHAPS, 0.2% [v/v] IPG buffer 5-8, 100 mM DTT and 0.01% [w/v] bromphenol blue). The strips were loaded onto a Bio-Rad Protean IEF Cell system and proteins were initially electrofocussed at 20°C first using a gradually increasing voltage (250 V-10000 V) and then reaching 55000 Vh. Strips were immediately equilibrated according to⁴¹. Second dimension SDS-PAGE was performed on 12% polyacrylamide gels using Protean Dodeca Cell systems (Bio-Rad). Gels were run at 80 V until the dye reached the bottom of the gel.

Gels were stained with colloidal Coomassie Brilliant blue (CBB) as previously described⁴². Briefly, the gels were rinsed in deionised water (2 x 1 min) and stained for 24 h in a solution containing 10% (w/v) ammonium sulfate, 2% (v/v) phosphoric acid, 0.1% (w/v)

CBB G250 (Bio-Rad) and 20% (v/v) methanol. After staining, the gels were washed 3 min in 0.1 M Tris-phosphoric acid (pH 6.5), followed by a 1 min wash in 25% (v/v) methanol and finally a 24 h wash in 20% (w/v) ammonium sulfate. CBB images, 1-DE and 2-DE, were acquired with a GS-800 Calibrated Densitometer (Bio-Rad).

Image analysis in 1-DE experiments was performed with the QuantityOne software (Bio-Rad). Bands were detected automatically and then confirmed by visual validation. 2-DE images analysis (three gels, one per biological replicate, was run for each stage) was performed with PDQuest v.8.01 software (Bio-Rad), using tenfold over background as a minimum criterion for presence/absence for the guided protein spot detection method. Visual validation of automated analysis was done thereafter to increase the reliability of the matching⁴³⁻⁴⁴.

6.2.4 Mass spectrometry analysis

For the MALDI-TOF/TOF analysis, spots with statistically significant differences were automatically excised from the gel employing Investigator ProPic robotic workstation (Genomic Solutions), transferred to multiwell 96 plates and digested with modified porcine trypsin (sequencing grade; Promega) by using a ProGest (Genomics Solution) digestion station^{35;45}. Briefly, gels pieces were destained by two washes at 37 °C for 30 min with 200 mM ammonium bicarbonate in 40% (v/v) acetonitrile (ACN). Slices were then washed twice, first with 25 mM ammonium bicarbonate for 5 min and later with 25 mM ammonium bicarbonate in 50 % (v/v) ACN for 15 min. respectively, dehydrated with 100 % ACN and finally dried at room temperature for 10 min. Trypsin (20 µL) solution was added to the dry gel pieces at a final concentration of 12.5 ng/µL in 25 mM ammonium bicarbonate, and the digestion proceeded at 37 °C overnight (approximately 16 h). Digestion was stopped by adding 10 µL of 0.5% trifluoroacetic acid (TFA) in water; peptides were desalted by using a resin C18 micro column (ZipTip, Millipore). Peptides were eluted directly with a matrix solution (α -cyanohydroxycinnamic acid at a concentration of 5 mg/mL in 70% ACN/0.1% TFA) onto MALDI plate using the dry droplet method. The MS analyzed in a 4700 Proteomics Analyzer MALDI-TOF/TOF Mass Spectrometer (Applied Biosystems), in the m/z range from 800 to 4000, with an accelerating voltage of 20 kV in reflectron mode and with a delayed extraction set to “on” and an elapsed time 120 ns. Spectra were internally calibrated using peptides from trypsin autolysis (M+H+=842.509, M+H+=2211.104) with an m/z precision of \pm 20 ppm. Most abundant peptide ions were subjected to MS/MS analysis, providing information that can be used to define the peptide sequence³⁵. A combined search (MS plus MS/MS) was performed against UniProtKB/TrEMBL, UniProtKB/SwissProt, NCBI non redundant (nr) protein database and Quercus_db (custom database of *Quercus* genera, detailed in appendix 10.1) using Mascot

searching engine (Matrix Science Ltd., London; <http://www.matrixscience.com>) with the following parameters: taxonomy restrictions to *Viridiplantae* for public database, one missed cleavage, 100 ppm mass tolerance in MS and 0.5 Da for collision-induced dissociation (CID) data, cysteine carbamidomethylation as a fixed modification and methionine oxidation as a variable modification. The confidence in the peptide mass fingerprinting matches ($p < 0.05$) was based on the MOWSE score (higher than 70), further confirmed by the accurate overlapping of the matched peptides with the major peaks of the mass spectrum. Proteins with unknown proteins hits were subjected to BLASTp search against NCBI nr protein database, considering hits with e -value $< 10^{-10}$ and identity $\geq 75\%$.

The *n*LC LTQ Orbitrap analysis was performed as previously described ⁴⁶. A first standard SDS-PAGE step was introduced for protein pre-fractionation and washing (removal of detergents and other contaminants). Briefly, 100 μ g of Laemmli-dissociated proteins per sample were subjected to SDS-PAGE electrophoresis in 12.5 % acrylamide gel (maximum thickness 1 mm) overlaid with a stacking gel of similar length, by using mini Protean (7 x 7 cm) system (Bio-Rad). The gel was run at 80 V until the bromophenol blue reached the resolving gel. The gel was then stained with CBB and the unique resulting bands per line excised using a clean scalpel and transferred to individual 1.5 mL tubes. The digestion was performed as described above, but the peptides elution step from C18 resin was performed by adding 200 μ L of solution containing 70% (v/v) ACN and 0.1% of TFA, which was later completely evaporated in speed-vac. The peptides were resuspended in 50 μ L of 4 % (v/v) ACN plus 0.25% (v/v) formic acid. This resuspension volume is equivalent to half the quantity of micrograms of protein loaded in a gel. The peptides were transferred into a microvial for autosampler injection. The LC settings were as follows: $\approx 2 \mu$ g of digested peptides in 5 μ L were loaded per injection onto a one-dimensional (1D) nano-flow LC-MS/MS system (Thermo Scientific). Peptides were eluted using a monolithic C18 column Acclaim PepMap (Thermo Scientific) of 15 cm length and 0.075 mm internal diameter during a 180 min gradient from 5 to 35% B with a controlled flow rate of 400 nL per minute (mobile phase A: 0.1% formic acid; mobile phase B: 90% ACN and 0.1% formic acid). LC was coupled to MS using an ESI source. MS analysis was performed on an LTQ Orbitrap XL mass spectrometer (Thermo Scientific). Specific tune settings for the MS were set as follows: spray voltage of 1.3 kV using a 10 μ m inner diameter needle (PicoTip Emitter; NewObjective, USA) and temperature of the heated transfer capillary was set to 180 °C. Fourier Transform Mass Spectrometer (FTMS) was operated as follows: full scan mode, centroid, resolution of 30,000, covering the range 400–1,500 m/z . Each full MS scan was followed by ten dependent MS/MS scans, in which the ten most abundant peptide molecular ions were dynamically selected during 150s. Ions with unassigned charge, charge+1 and higher than +3 were excluded for fragmentation.

For protein identification, raw-files were processed with the Proteome Discoverer software v.1.4 (Thermo Scientific) using the Quercus_DB protein database ⁴⁷ combined with UniProtKB/TrEMBL, UniProtKB/SwissPrto and NCBI nr databases, with taxonomy restrictions to *Viridiplantae*, using the SEQUEST algorithm (implemented in Proteome Discoverer v.1.4, Thermo Scientific). Precursor mass tolerance was set to 10 ppm and fragment ion mass tolerance fixed to 0.8 Da. Only charge states +2 or greater were used. Identification confidence was set to a 5% FDR and acetylation of N terminus, oxidation of methionine and carbamidomethyl cysteine formation were set as variable modifications. No fixed modifications were set. Trypsin was set as proteolytic enzyme and a maximum of two miss cleavages were set for all searches. The threshold for protein identification was one unique peptide (peptide that only appears once in the entire database) with a *X*-Correlation value 0.25 greater than the charge state (*i.e.*, 2.25 for peptides with charge +2) ⁴⁸. Proteins with unknown proteins hits were subjected to BLASTp search against NCBI nr protein database, with an *e*-value cutoff of 10^{-10} and identity $\geq 75\%$.

6.2.5 Functional classification of proteins

Identified proteins by *n*LC LTQ Orbitrap analysis were subjected to gene ontology (GO) annotation using Blas2GO ⁴⁹ based on BLASTp results against NCBI nr protein database (*e*-value $< 10^{-3}$). Differentially accumulated protein identified by MALDI-TOF/TOF and *n*LC LTQ Orbitrap analysis were extracted and classified based on their putative function according to Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway, also using Blas2GO ⁴⁹ or according to the annotation in UniProtKB protein database.

6.2.6 Statistical analysis

The band and spot abundances obtained in gel image analysis were normalised as previously described ⁵⁰⁻⁵¹. Briefly, the normalised spot abundance (NSA) from spot *k* of gel *i* was obtained after the division of the spot abundance (SA) by the sum of SA of all *N* spots present in gel *i*.

$$(NSA)_{k} = \frac{(SA)_{k}}{\sum_{i=1}^N (SA)}$$

In 1-DE analysis the normalisation of individual bands was performed by division of the band abundance (BA) by the sum of BA of all *N* bands present in the lane *i*.

Only consistent bands and spots that were present in the three biological replicates were considered for statistical calculations of protein expression levels ^{29; 50}. Experimental pI

was determined using a 5-8 linear scale over the total length of the IPG strip. *Mr* values were calculated by mobility comparisons with protein standards markers (SDS Molecular weight standards, Broad range, Bio-Rad) run in the same gel in 1-DE and in a separate lane in the gel in 2-DE.

For statistical and cluster analysis of protein bands or SA values in gel based approach, the web-based software NIA array analysis tool was utilised ⁵² available online at <http://lgsun.grc.nia.nih.gov/anova/index.html>, as reported in ²⁹. Through this software, statistically valid protein spots were selected based on analysis of variance (ANOVA). After uploading the data table containing normalized and transformed consistent spots data and indication of biological replications, the data were statistically analyzed using the following settings: error model max (average, actual), 0.01 proportions of the highest variance values to be removed before variance averaging, 10 degrees of freedom for the Bayesian error model, 0.05 FDR thresholds, and zero permutations. The entire data set was analyzed by principal component analysis (PCA) according to the following settings: covariance matrix type, four principal components, 2-fold change threshold for clusters, and 0.7 correlation threshold for clusters. PCA results were represented as a biplot, with consistent proteins in those experimental situations located in the same area of the graph.

Identified proteins by *n*LC LTQ Orbitrap were quantified by a peptide count measurement using NSAF (Normalized spectral abundance factor) approach, as previously described ^{46; 53}. For each protein *k* identified, we calculated a spectral abundance factor normalized against the whole protein complex (NSAF) as follows:

$$(NSAF)_k = \frac{(Spc/L)_k}{\sum_{i=1}^N (Spc/L)_i}$$

in which the total number of tandem MS spectra matching peptides from protein *k* (*Spc*) was first divided by the protein length (*L*), then divided by the sum of *Spc/L* for all *N* proteins in each stage analyzed. This measurement is limited to peptides that have been assigned to proteins, but not to those absent in the database or those with posttranslational modifications (PTMs) not defined in the SEQUEST search step ⁴⁶. To determine quantitative difference, a ratio between stage S3 and S0 (*S3/S0*) ≥ 2 was considered germination enriched (S3), and a ratio ≤ 0.5 were considered mature seed enriched (S0). In addition, only proteins with 3 or more peptide assigned were considered as differentially accumulated.


6.3 Results and discussion

To study holm oak seed germination and seedling establishment, two complementary proteomic approaches, gel based and gel free (shotgun), were used in order to obtain the

largest possible proteome coverage and hence, the maximal information about the changes in protein accumulation along these complex processes.

Proteins were extracted from cryo-homogenated embryo axes (stages S0-S3), whole seedling (S4 to S7) and shoots (SS-1 and SS-4) by using the TCA/acetone and phenol precipitation method. Statistically significant variations were observed in protein yield, and depending on the stage, ranged from 1.92 ± 0.38 mg/g FW (S7) to 15.21 ± 1.64 mg/g FW (S0) (Table 6.1), decreasing as the water content in the tissue increased. This decrease in protein content observed during germination and seedling establishment has been previously reported in other similar studies^{9; 20} and is mainly related to storage mobilisation.

Table 6.1: Protein yield in embryo axes from different stages of germination and seedling establishment.

Stages of germination and development	S0	S1	S2	S3	S5	S6	S7	SS-1	SS-4
									
Protein yield (mg g ⁻¹ of fresh weigh) ¹	15.21 ^a	13.08 ^a	12.32 ^b	13.66 ^a	6.58 ^{c,d}	3.17 ^{d,e}	1.92 ^a	8.67 ^{b,c}	5.02 ^{c,d,e}
SD	1.64	3.45	3.01	0.63	0.40	0.43	0.38	0.81	0.25

¹Means of protein yield with the same letter are not significantly different according to Tukey's HSD test ($\alpha=0.05$)

The values shown in Table 6.1 were in the range of those reported for seeds, radicle and shoots in different genera of *Fagaceae*, including *Quercus* spp. and other species^{24; 54-56}. The percentage of protein recovery with the used methodology was approximately 30% in seeds considering protein contents estimated by NIRS reported for *Q. ilex* seeds by²³. No data are available for the protein content in other holm oak tissues, but our results agree with those found in other species⁵⁷⁻⁵⁸.

6.3.1 One-dimensional SDS-PAGE protein profile

One-dimensional SDS-PAGE is a very useful tool to separate protein molecules by size, less expensive and time consuming than other proteomics approaches. We used this methodology to obtain a first vision of the proteomic changes occurring in *Q. ilex* seed at different times after imbibition that let us to choose only informative stages for posterior 2-DE and nLC LTQ Orbitrap approaches.

For each assayed time/stage, >100 individuals were distributed in three groups and the individuals in each group pooled; the three pools per stage, were used as biological replicates. Pooling of samples in *omics* experiments is a valid and potentially valuable procedure when many individuals are analysed. Usually the protein expression in a pool

matches the mean expression of the individuals making up the pool for the majority of proteins and the biological variance between pools is reduced compared with that between individuals⁵⁹. As we pooled individuals of each stage in three different pools, we still could assess the biological variability. The importance of biological over technical replication has been emphasised both in *omics* work, because treatment effects should be tested against biological rather than technical error⁵⁹.

Protein extracts (80 µg) from embryo axes at the nine development stages (three biological replicates per stage) were loaded per line, separated by SDS-PAGE stained and images analysed with the QuantityOne software (Bio-Rad) (Fig. 6.1). Data were subjected to multivariate statistical analysis and clustering in order to establish different groups. A total of 60 bands were resolved in the 5–221 kDa Mr range, most of them located in the 31–116 kDa Mr range (Fig. 6.1). Mean of normalized and transformed relative values for each band, as well as standard deviation (SD), are provided in Table 6.2. ANOVA (with FDR < 0.05) detected 43 bands with statistically significant qualitative (*i.e.*, absence or presence) or quantitative differences (*i.e.*, different intensities). Only 18 of these 43 bands, including quantitative and qualitative difference, contributed to distinguish the different stages (Fig. 6.2).

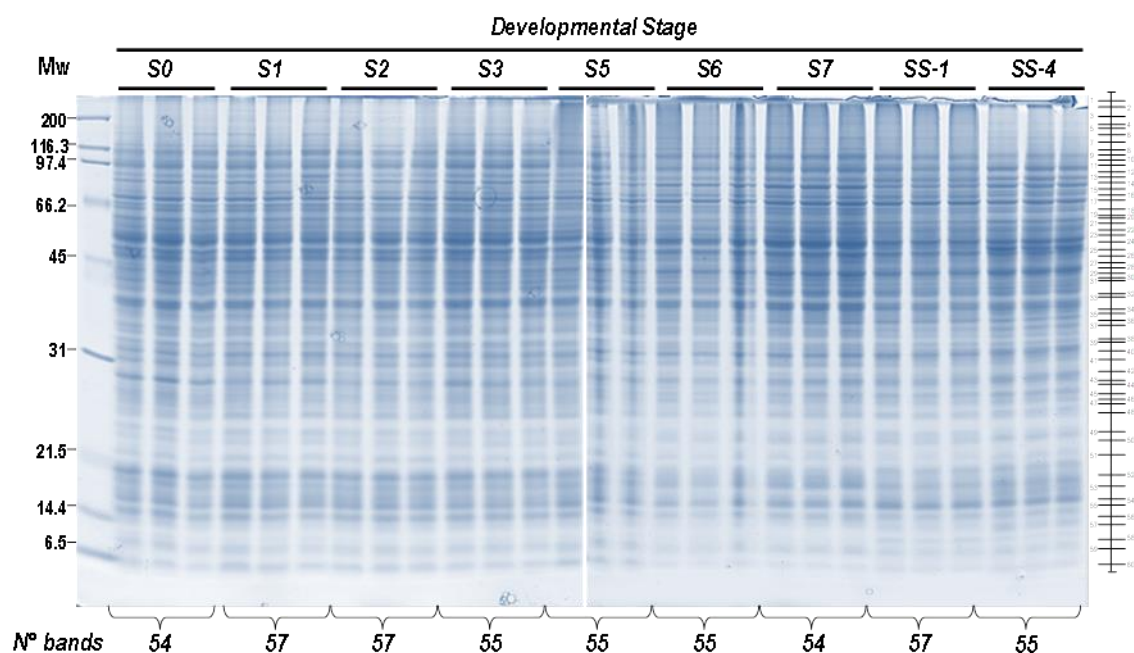


Figure 6.1: 1-DE analysis of changes in the protein profile during germination (S0 to S3), post germination (S5 to S7) and seedling establishment (SS-1 and SS-3) of *Q. ilex* seeds. Protein extracts (80 µg) from embryo axes at the nine development stages (three biological replicates per stage) were loaded per line, separated by SDS-PAGE stained and images analysed with the QuantityOne software (Bio-Rad). Each line correspond to one sampling time and biological replicate (three biological replicates per sampling time). On the right is shown a diagram representing the 60 resolved bands; only consistent bands were considered. Asterisks and crosses indicate the bands with quantitative and qualitative differences, respectively. Data were subjected to multivariate statistical analysis and clustering in order to establish different groups (Table 6.2 and Fig. 6.2).

Table 6.2: Normalised and transformed relative volumes of the 1-DE resolved bands. Normalised band abundance by dividing the band abundance (BA) by the sum of BA of all bands in each line as detailed in M&M.

Band Number	Mr (Kda)	S0 (0h)	S1 (4h)	S2 (10h)	S3 (24h)	S5 (72h)	S6 (144h)	S7 (216h)	SS-1 (Seedling 1 cm)	SS-4 (Seedling 4 cm)	Statistical analysis
											FDR
1	221.47	0.03 ± 0.01	0.07 ± 0.02	0.04 ± 0.01	0.04 ± 0.01	0.76 ± 0.60	1.10 ± 1.01	0.69 ± 0.27	1.14 ± 0.17	0.01 ± 0.00	0.00000
2	195.83	0.07 ± 0.10	0.14 ± 0.08	0.03 ± 0.01	0.13 ± 0.11	0.13 ± 0.39	1.11 ± 0.86	0.53 ± 0.22	0.67 ± 0.10	0.05 ± 0.01	0.00000
3	165.06	0.04 ± 0.02	0.05 ± 0.00	0.08 ± 0.04	0.19 ± 0.05	0.19 ± 0.20	0.71 ± 0.95	0.08 ± 0.01	0.15 ± 0.02	0.05 ± 0.03	0.00007
4	153.04	0.44 ± 0.26	0.37 ± 0.26	0.18 ± 0.11	0.71 ± 0.03	0.71 ± 0.22	1.13 ± 0.79	0.65 ± 0.07	0.40 ± 0.13	0.16 ± 0.08	0.00081
5	131.86	0.18 ± 0.07	0.22 ± 0.12	0.12 ± 0.10	0.40 ± 0.08	0.40 ± 0.06	0.85 ± 0.50	0.30 ± 0.08	0.18 ± 0.11	0.11 ± 0.05	0.00460
6	123.78	0.95 ± 0.10	0.74 ± 0.05	0.79 ± 0.33	0.98 ± 0.06	0.98 ± 0.06	1.11 ± 0.59	0.53 ± 0.08	0.17 ± 0.05	0.08 ± 0.01	0.00000
7	111.56	3.70 ± 0.85	3.84 ± 0.66	3.45 ± 0.92	3.85 ± 0.19	3.85 ± 0.03	4.29 ± 0.67	3.90 ± 0.21	2.35 ± 0.10	1.68 ± 0.17	0.01000
8	101.19	1.90 ± 0.33	1.80 ± 0.12	1.97 ± 0.24	1.65 ± 0.22	1.65 ± 0.11	1.07 ± 0.04	0.94 ± 0.19	0.82 ± 0.12	0.67 ± 0.12	0.00002
9	97.11	4.33 ± 0.75	3.86 ± 0.47	4.04 ± 0.34	4.79 ± 0.19	4.79 ± 0.55	3.79 ± 0.73	4.12 ± 0.11	2.76 ± 0.13	3.88 ± 0.21	0.62913
10	90.43	nd.	nd.	nd.	nd.	nd.	nd.	nd.	0.58 ± 0.09	nd.	0.00000
11	89.17	1.41 ± 0.33	1.83 ± 0.78	1.27 ± 0.09	1.64 ± 0.08	1.64 ± 0.24	1.86 ± 0.37	2.00 ± 0.15	3.04 ± 0.29	3.04 ± 0.48	0.00228
12	84.19	2.11 ± 0.21	1.86 ± 0.15	1.99 ± 0.20	2.23 ± 0.17	2.23 ± 0.33	3.95 ± 0.40	4.52 ± 0.10	2.79 ± 0.41	4.03 ± 0.02	0.00067
13	80.21	0.48 ± 0.25	0.88 ± 0.80	0.35 ± 0.03	0.62 ± 0.11	0.62 ± 0.16	0.57 ± 0.02	0.70 ± 0.04	0.83 ± 0.19	1.20 ± 0.13	0.00981
14	76.34	1.22 ± 0.16	1.94 ± 0.57	1.80 ± 0.10	1.62 ± 0.13	1.62 ± 0.47	2.06 ± 0.24	1.73 ± 0.13	1.98 ± 0.24	1.57 ± 0.31	0.69021
15	72.01	4.88 ± 0.55	4.31 ± 0.71	4.35 ± 0.13	4.33 ± 0.26	4.33 ± 0.58	4.94 ± 0.82	4.79 ± 0.19	5.23 ± 0.48	4.56 ± 0.25	0.99801
16	67.94	2.18 ± 0.11	2.60 ± 0.57	2.68 ± 0.05	2.23 ± 0.09	2.23 ± 0.28	1.99 ± 0.15	1.86 ± 0.18	1.09 ± 0.15	0.91 ± 0.13	0.00007
17	65.26	2.26 ± 0.14	2.50 ± 0.27	2.44 ± 0.30	2.24 ± 0.19	2.24 ± 0.35	1.70 ± 0.25	1.27 ± 0.04	0.99 ± 0.21	1.51 ± 0.21	0.00324
18	61.12	2.73 ± 0.29	2.62 ± 0.40	2.45 ± 0.16	2.70 ± 0.14	2.70 ± 0.21	1.38 ± 0.08	1.52 ± 0.16	1.84 ± 0.15	2.66 ± 0.18	0.05471
19	59.00	1.15 ± 0.10	1.18 ± 0.42	1.04 ± 0.03	0.97 ± 0.10	0.97 ± 0.13	1.34 ± 0.24	1.32 ± 0.13	1.57 ± 0.16	1.88 ± 0.47	0.33700
20	58.52	2.30 ± 0.39	1.96 ± 0.26	1.46 ± 0.06	1.91 ± 0.21	1.91 ± 0.22	2.33 ± 0.27	3.00 ± 0.39	2.64 ± 0.27	2.81 ± 0.43	0.21817
21	55.95	5.27 ± 1.56	4.44 ± 0.49	4.28 ± 0.30	4.39 ± 0.08	4.39 ± 0.29	2.81 ± 0.37	2.19 ± 0.30	2.82 ± 0.11	nd.	0.00000
22	54.58	6.17 ± 4.83	8.93 ± 1.76	9.71 ± 0.77	9.55 ± 0.43	9.55 ± 0.58	11.79 ± 0.57	10.72 ± 0.64	7.09 ± 1.44	7.11 ± 0.72	0.22854
23	50.58	nd.	nd.	nd.	nd.	nd.	nd.	nd.	3.33 ± 0.55	7.64 ± 0.15	0.00000
24	49.77	4.98 ± 0.53	5.55 ± 1.10	4.87 ± 0.47	5.05 ± 0.61	5.05 ± 0.65	5.93 ± 0.89	5.04 ± 0.17	5.65 ± 0.71	5.08 ± 1.08	0.99801
25	48.82	5.09 ± 1.27	5.16 ± 0.71	5.42 ± 0.46	3.40 ± 0.57	3.40 ± 0.47	1.73 ± 0.39	1.21 ± 0.02	3.00 ± 0.03	3.51 ± 0.16	0.00000
26	46.67	0.52 ± 0.11	0.99 ± 0.57	1.20 ± 0.05	0.53 ± 0.14	0.53 ± 0.18	0.68 ± 0.16	0.58 ± 0.54	2.53 ± 0.33	nd.	0.00000
27	44.54	2.51 ± 0.50	2.62 ± 0.64	2.93 ± 0.42	1.94 ± 0.28	1.94 ± 0.62	5.01 ± 0.75	4.88 ± 0.54	5.62 ± 0.50	8.38 ± 0.52	0.00000
28	43.83	nd.	1.62 ± 0.49	1.59 ± 0.17	1.30 ± 0.19	1.30 ± 0.15	1.86 ± 0.24	1.89 ± 0.27	1.59 ± 0.10	2.18 ± 0.46	0.00000
29	42.81	2.49 ± 0.09	2.10 ± 0.41	2.40 ± 0.04	2.47 ± 0.12	2.47 ± 0.32	1.79 ± 0.04	1.76 ± 0.28	1.67 ± 0.12	2.13 ± 0.49	0.76743
30	41.10	nd.	nd.	nd.	nd.	nd.	nd.	nd.	1.87 ± 0.27	2.53 ± 0.18	0.00000
31	41.80	1.84 ± 0.19	2.36 ± 0.13	2.92 ± 0.46	2.68 ± 0.19	2.68 ± 0.43	4.41 ± 0.17	4.10 ± 0.68	3.10 ± 0.20	2.87 ± 0.15	0.04442
32	39.62	4.22 ± 0.40	3.42 ± 0.35	3.26 ± 0.47	3.57 ± 0.48	3.57 ± 0.67	4.14 ± 0.45	4.05 ± 0.65	2.60 ± 0.46	2.08 ± 0.21	0.09404
33	39.24	4.20 ± 0.43	2.91 ± 0.09	3.28 ± 0.28	3.89 ± 0.24	3.89 ± 0.69	3.81 ± 0.69	4.04 ± 0.24	3.34 ± 0.34	3.24 ± 0.33	0.91919
34	37.90	1.19 ± 0.15	1.14 ± 0.30	1.32 ± 0.33	1.41 ± 0.30	1.41 ± 0.10	1.03 ± 0.10	1.23 ± 0.22	0.92 ± 0.20	0.91 ± 0.36	0.76448
35	37.68	0.58 ± 0.21	0.74 ± 0.39	0.37 ± 0.25	0.00 ± 0.00	0.00 ± 0.04	1.18 ± 0.12	2.05 ± 0.23	2.58 ± 0.09	1.57 ± 0.21	0.00000
36	36.17	1.23 ± 0.03	0.96 ± 0.28	0.77 ± 0.20	0.55 ± 0.26	0.55 ± 0.13	0.81 ± 0.16	1.28 ± 0.15	1.11 ± 0.18	1.23 ± 0.05	0.00939
37	35.66	1.01 ± 0.18	0.95 ± 0.30	1.07 ± 0.31	1.34 ± 0.35	1.34 ± 0.52	0.75 ± 0.30	1.11 ± 0.21	0.62 ± 0.10	0.56 ± 0.02	0.02218
38	34.12	1.24 ± 0.28	1.21 ± 0.51	0.81 ± 0.11	1.62 ± 0.21	1.62 ± 0.43	0.68 ± 0.29	0.82 ± 0.04	0.89 ± 0.10	0.65 ± 0.11	0.00927
39	33.70	nd.	0.48 ± 0.08	0.44 ± 0.12	0.67 ± 0.31	0.67 ± 0.14	0.77 ± 0.30	1.18 ± 0.06	0.44 ± 0.77	nd.	0.00000
40	32.72	2.56 ± 0.19	3.01 ± 0.11	2.90 ± 0.48	3.17 ± 0.15	3.17 ± 0.32	3.03 ± 0.21	3.52 ± 0.30	4.25 ± 0.21	3.23 ± 0.16	0.82597
41	31.35	2.00 ± 0.37	1.17 ± 0.30	1.24 ± 0.21	1.66 ± 0.19	1.66 ± 0.22	0.84 ± 0.44	1.47 ± 0.45	0.72 ± 0.36	0.61 ± 0.12	0.00000
42	29.83	nd.	0.40 ± 0.21	0.32 ± 0.05	0.40 ± 0.04	0.40 ± 0.03	0.23 ± 0.17	0.33 ± 0.15	0.35 ± 0.20	0.06 ± 0.05	0.00000
43	29.13	3.82 ± 0.35	2.39 ± 0.30	1.92 ± 0.09	3.26 ± 0.93	3.26 ± 0.19	1.58 ± 0.84	2.64 ± 0.44	2.18 ± 0.52	2.48 ± 0.23	0.02656
44	28.62	0.35 ± 0.09	0.30 ± 0.26	0.42 ± 0.15	0.00 ± 0.00	0.00 ± 0.31	0.06 ± 0.11	nd.	nd.	nd.	0.00000
45	27.94	1.27 ± 0.45	0.48 ± 0.25	0.58 ± 0.15	1.00 ± 0.21	1.00 ± 0.20	0.65 ± 0.53	1.19 ± 0.38	0.98 ± 0.27	0.95 ± 0.20	0.01452
46	27.67	0.88 ± 0.15	0.36 ± 0.06	0.33 ± 0.07	0.89 ± 0.21	0.89 ± 0.32	0.61 ± 0.36	0.60 ± 0.12	0.39 ± 0.24	0.40 ± 0.15	0.00003
47	27.25	0.33 ± 0.09	0.14 ± 0.04	0.18 ± 0.05	0.31 ± 0.05	0.31 ± 0.17	nd.	nd.	nd.	0.16 ± 0.03	0.00000
48	26.69	0.43 ± 0.04	0.44 ± 0.14	0.51 ± 0.03	0.57 ± 0.11	0.57 ± 0.05	0.36 ± 0.18	0.33 ± 0.13	0.26 ± 0.05	0.12 ± 0.03	0.00000
49	24.98	0.21 ± 0.03	0.27 ± 0.01	0.26 ± 0.03	0.20 ± 0.03	0.20 ± 0.03	0.15 ± 0.02	0.15 ± 0.03	0.24 ± 0.02	0.13 ± 0.05	0.05704
50	24.17	0.16 ± 0.02	0.18 ± 0.00	0.20 ± 0.03	0.19 ± 0.01	0.19 ± 0.03	0.20 ± 0.06	0.23 ± 0.02	0.26 ± 0.03	0.20 ± 0.05	0.82330
51	22.57	0.16 ± 0.03	0.23 ± 0.04	0.31 ± 0.02	0.16 ± 0.01	0.16 ± 0.05	0.14 ± 0.02	0.13 ± 0.00	0.06 ± 0.01	0.14 ± 0.02	0.00000
52	19.97	4.05 ± 0.82	3.62 ± 0.82	3.84 ± 0.30	3.31 ± 0.13	3.31 ± 0.61	0.89 ± 0.72	1.25 ± 0.30	0.53 ± 0.07	0.70 ± 0.21	0.00000
53	18.79	1.93 ± 0.48	1.91 ± 1.01	2.32 ± 0.28	1.84 ± 0.49	1.84 ± 0.43	1.44 ± 1.22	2.16 ± 0.39	1.67 ± 0.14	2.38 ± 0.27	0.09115
54	16.95	1.17 ± 0.33	2.13 ± 0.69	1.72 ± 0.18	1.09 ± 0.16	1.09 ± 0.00	nd.	nd.	nd.	1.68 ± 0.20	0.00000
55	16.19	2.42 ± 0.84	2.71 ± 1.14	3.10 ± 0.07	2.15 ± 0.30	2.15 ± 0.18	2.02 ± 1.35	2.41 ± 0.27	4.76 ± 0.48	3.06 ± 0.37	0.00907
56	14.63	1.75 ± 0.87	1.37 ± 0.97	1.85 ± 0.28	1.52 ± 0.15	1.52 ± 0.27	0.53 ± 0.33	0.47 ± 0.13	0.54 ± 0.05	0.40 ± 0.05	0.00000
57	12.83	0.30 ± 0.05	0.19 ± 0.09	0.28 ± 0.04	0.23 ± 0.03	0.23 ± 0.08	0.16 ± 0.08	0.12 ± 0.05	0.22 ± 0.01	0.34 ± 0.10	0.00033
58	8.94	0.10 ± 0.04	0.05 ± 0.02	0.05 ± 0.02	0.04 ± 0.02	0.04 ± 0.02	0.08 ± 0.03	0.05 ± 0.02	0.15 ± 0.00	0.08 ± 0.06	0.01870
59	8.30	0.33 ± 0.06	0.27 ± 0.03	0.34 ± 0.02	0.27 ± 0.02	0.27 ± 0.13	0.39 ± 0.04	0.30 ± 0.01	0.26 ± 0.03	0.20 ± 0.07	0.19294
60	5.78	0.24 ± 0.21	0.10 ± 0.03	0.12 ± 0.03	0.14 ± 0.04	0.14 ± 0.17	0.19 ± 0.05	0.10 ± 0.02	0.17 ± 0.01	0.15 ± 0.04	0.00037

Data are mean ± SD of three biological replicates. nd.: not detected spots; they were considered qualitative differences

A hierarchical clustering of the analysed biological replicates indicated that 1-DE band patterns were quite similar between S1 and S2 stages and among S5, S6 and S7 (Fig. 6.2A). PCA of developmental stages (Fig. 6.2B) and protein band intensities (Fig. 6.2C) were also conducted. PC1 and PC2 accounted for 39.22% and 32.11% of the total variability, respectively. PCA analysis separate two groups in the developmental stages, one corresponding to S0, S1, S2 and S3 and other to the seedling development stages from S5 to S7. SS-1 and SS-4 were excluded of both groups. This result indicated that PC1 separated tissue types: embryo axis and seedling on the one hand and shoots by another. It also indicate that there are at least four differential protein band groups in these tissues analyzed. However, the band pattern was not greatly different between unimbibed and germinating seeds. This preliminary experiment was used for the selection of the samples (S0, S3, SS-1 and SS-4) to be analysed through a more resolutive technique, as it is 2-DE.

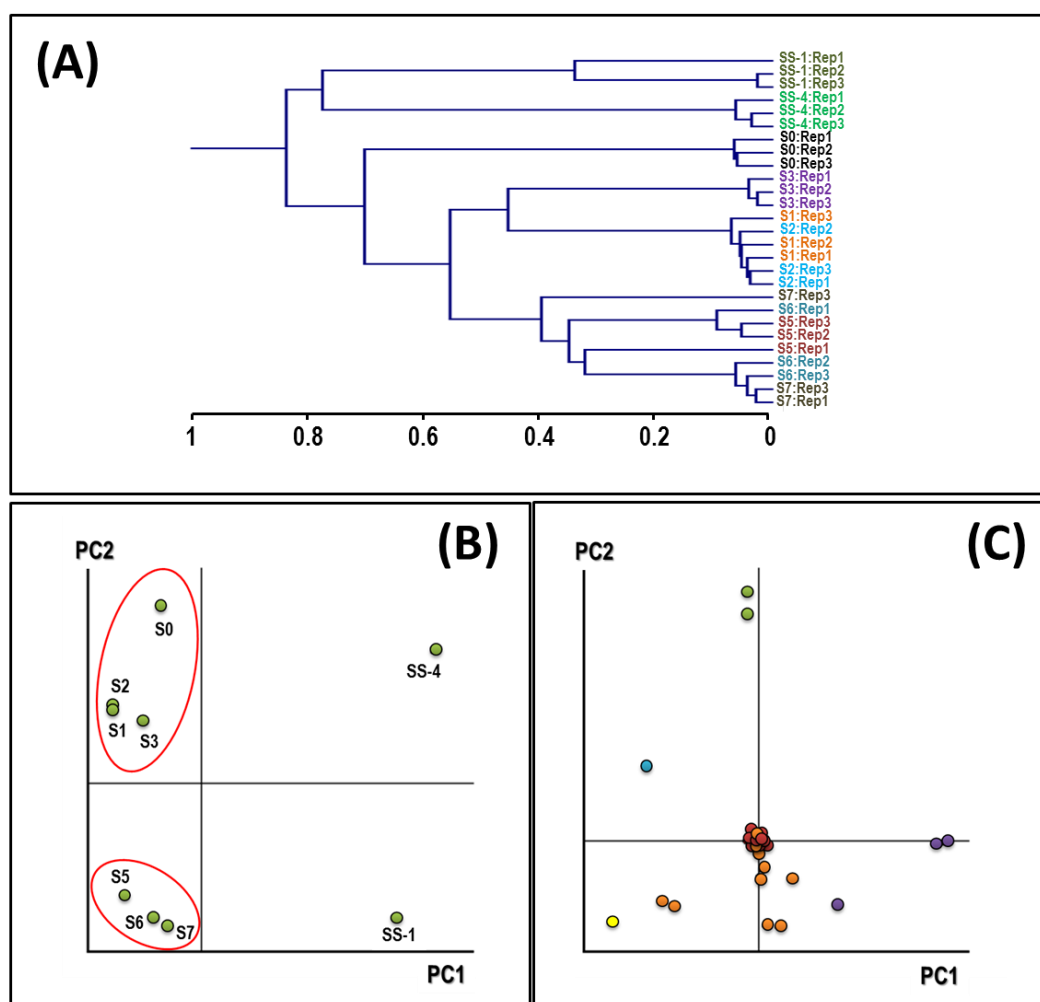


Figure 6.2: Clustering of biological replicates and samples based on main principal components found after PCA of data (Table 6.2) obtained after 1-DE analysis of changes in the protein profile during germination (S0 to S3), post germination (S5 to S7) and seedling establishment (SS-1 and SS-4) of *Q. ilex* seeds. (A). Hierarchical clustering of the three biological replicates of samples analysed. Two-dimensional biplots show associations between experimental samples (B), and protein band (C), generated by PCA. Samples and protein bands were plotted in the first and second component spaces.

6.3.2 Two-dimensional gel electrophoresis and differential abundance spots identification by using MALDI-TOF/TOF

Two-dimensional polyacrylamide gel electrophoresis (2-DE) is considered a powerful tool used for separation and fractionation of complex protein mixtures as it allows separation of thousands of proteins in one gel. Although 2-DE gels contain less proteins than 1-DE, their resolution is much higher as the use of two dimensions decreases the possibility of co-migration of proteins^{24; 26; 60}. Additionally, 2-DE provides direct visual confirmation of changes in protein abundance and in the abundance of post-translational modifications (PTMs), something that cannot be predicted from the genomic sequence. From data in 1-DE experiment, four stages were chosen to perform 2-DE analysis. Stages S0 (mature unimbibed seed) and S3 (24 h after imbibition, when the rupture of the testa and radicle protrusion occurs, indicating the end of the germination phase) were included to identify germination specific proteins. The inclusion of SS-1 (shoots seedling of 1cm) and SS-4 (shoots seedling of 4cm) helped to identify proteins involved in the early holm oak seedling establishment.

Previous 2-DE experiments performed in our laboratory with small 3-10 IPG range strip²⁴ showed that most of the proteins in the *Q. ilex* acorn were concentrated in the 5-8 pH range^{25-26; 28-29; 35; 44; 61}. Considering these results and in order to increase spot resolution, 2-DE was performed in the 5-8 pH range using 17 cm IPG strips. Three biological replicate per stage were analysed, being each replicate one pool of 30-50 individuals not included in any other pools. A virtual master gel representing all consistent protein spots in all analyzed samples is shown in Fig. 6.3, where protein spots with statistically significant differences in intensity in the various studied samples compared to the S3 stage, are remarked.

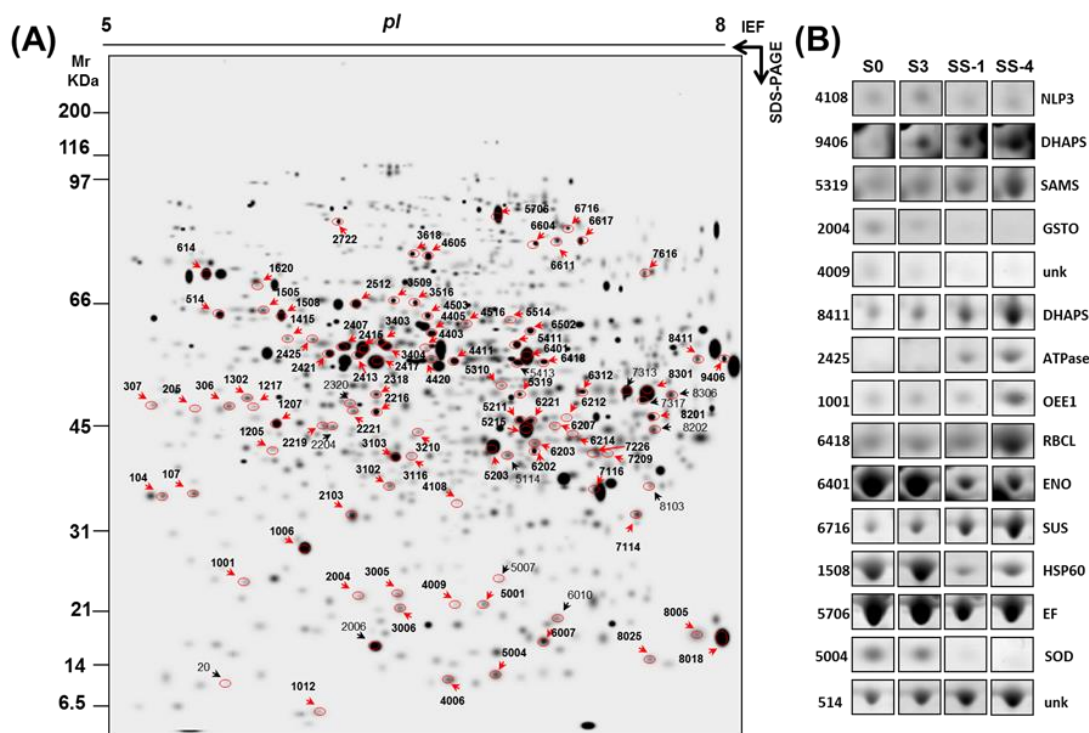


Figure 6.3: 2-DE analysis of changes in the protein profiles during germination (S0 to S3), post germination (S5 to S7) and seedling establishment (SS-1 and SS-3) of *Q. ilex* seeds. **(A).** Master gel combining spots images of S0, S3, SS1 and SS4. The relative *Mr* is given on the left and the *pI* is given at the top of the figure. Protein spots with a significant change in abundance between each studied stage and S3 are indicated with circles. Dark arrows indicate the spots that could not be identified, while red arrows point to spots that represent an identified protein. **(B).** Representative spots showing changes in protein abundance along the germination process. The spot number is indicated on the left of the image and the protein name on the right. Uncharacterized proteins were indicated as “unk”.

After normalization of protein spot images using PD-Quest software and posterior manual verification to increase reliability⁵⁰, 540 to 675 resolved spots were detected, which were distributed in the range of a mass weight between 16 and 131 KDa. The main cluster of protein spots was observed at the molecular mass of 30 to 70 kDa from pH range 6–7.5 (Fig. 6.3). SS-4 had the lowest number of protein spots detected (540 ± 1) and embryo axis of 24 h post imbibition (S3) had the highest number of protein spots detected (675 ± 4.58) (Table 6.3). The number of resolved spot in *Q. ilex* samples was similar to that described for other species^{17-18; 20; 62}. The proteome coverage depends on several factors, starting from protein extraction methods, fractionated strategies and finally the protein identification. The percentage of *Q. ilex* proteome that was analysed in these experimental conditions was estimated in about 2%. This estimations were based on an average number of 35,000 putative protein-coding genes⁶³⁻⁶⁵, without having in consideration that each gene can encode more than one protein and that not all genes are expressed in a tissue and in a given condition. We also did not consider possible post-translational modifications and assumed that each spot in gels is composed by a unique protein (which in reality is not the case, as one spot may contain more than one protein).

Table 6.3: Number of total protein spots resolved in 2-DE gels for each studied stage and number of spots showing qualitative or quantitative differences in abundance referred to S3 stage, based on ANOVA test ($p < 0.5$).

Stage	Number of resolved spots in 2-DE				
	Total	Up-accumulated		Down-accumulated	
		Qualitative	Quantitative	Qualitative	Quantitative
S0	646	1	4	0	0
S3	653	-	-	-	-
SS-1	528	9	23	16	17
SS-4	519	10	33	16	25

*Differentially expressed protein spots were selected after ANOVA analysis, with FDR < 0.05 and fold change ≥ 2 . To ensure identifications only spot with intensity higher than 500 in PD-Quest software analysis was selected.

A total of 732 consistent spots were analyzed by PCA (Fig. 6.4). The first four PCs accounted for 99.99% of the biological variability. PC1 and PC2 explained 81.79% and 15.10% of total variability, respectively. PC1 separated the two tissues (embryo axis and shoot seedling) and PC2 separated shoot seedling of 1 from 4 cm. However, these two first PCs could not separate the mature seeds at S0 and S3 stages. Data were further analyzed by ANOVA to generate a list of protein spots that differentiated between S3 group and any other, and 481 out of the 732 resolved consistent spots, with FDR < 0.05 , were selected. From these, a new list of 103 differentially abundant proteins was generated with an arbitrary fold change cutoff of ≥ 2 -fold (Table 6.4). Of this total, 38 protein spots showed qualitative variation (presence/absence) and other 65 spots showed significant quantitative variations relative to the S3 stage, used as reference. These 103 variable protein spots were subjected to in-gel tryptic digestion followed by MALDI-TOF/TOF analysis for their identification.

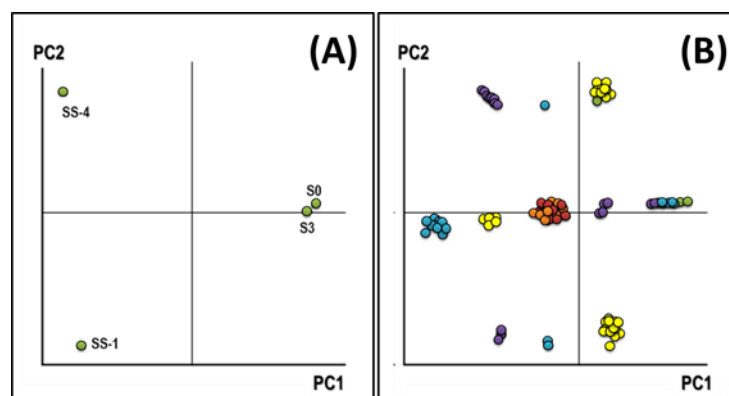


Figure 6.4. Principal components analysis of consistent spots in 2-DE protein profile during germination (S0 and S3) and seedling establishment (SS-1 and 4). Two-dimensional biplots showing associations between experimental samples (A) and protein band (B) generated by principal component analysis (PCA). Samples and protein spots were plotted in the first and the second component spaces.

Table 6.4: Normalised and transformed relative volumes of the 2-DE resolved protein spots. Differentially expressed protein spots were selected after ANOVA analysis, with FDR <0.05 and fold change ≥ 2 .

Spot Number	Mr (kDa)	pI	S0	S3	SS-1	SS-4	FDR
20	19.16	5.51	0.12 ± 0.04	0.02 ± 0.02	nd.	nd.	0.00000
104	37.34	5.19	0.35 ± 0.05	0.27 ± 0.08	0.67 ± 0.08	1.11 ± 0.63	0.00001
107	37.62	5.34	0.34 ± 0.04	0.33 ± 0.06	nd.	nd.	0.00000
205	50.90	5.35	0.10 ± 0.02	0.13 ± 0.04	0.25 ± 0.04	0.38 ± 0.06	0.00003
306	51.43	5.51	nd.	nd.	0.33 ± 0.06	0.70 ± 0.19	0.00000
307	51.56	5.14	nd.	nd.	nd.	0.38 ± 0.15	0.00000
514	71.20	5.47	0.59 ± 0.15	1.03 ± 0.04	2.02 ± 0.67	2.63 ± 0.31	0.00000
614	82.42	5.40	2.73 ± 0.42	3.75 ± 0.31	3.59 ± 1.00	1.02 ± 0.41	0.00001
1001	27.46	5.60	0.20 ± 0.03	0.17 ± 0.01	0.34 ± 0.12	0.93 ± 0.25	0.00000
1006	31.02	5.89	2.29 ± 0.49	3.08 ± 0.32	0.85 ± 0.29	0.59 ± 0.03	0.00000
1012	17.26	5.96	0.50 ± 0.07	0.39 ± 0.16	nd.	nd.	0.00000
1205	44.10	5.73	0.13 ± 0.02	0.13 ± 0.00	0.22 ± 0.04	0.42 ± 0.06	0.00031
1207	48.33	5.75	1.06 ± 0.41	1.25 ± 0.42	3.18 ± 0.80	5.26 ± 0.41	0.00000
1217	51.28	5.64	nd.	nd.	0.14 ± 0.05	0.35 ± 0.25	0.00000
1302	52.93	5.61	0.36 ± 0.06	0.29 ± 0.02	0.06 ± 0.04	0.08 ± 0.01	0.00000
1415	65.48	5.80	nd.	nd.	0.20 ± 0.03	0.45 ± 0.04	0.00000
1505	72.31	5.68	0.37 ± 0.13	0.28 ± 0.05	nd.	nd.	0.00000
1508	71.06	5.77	1.54 ± 0.66	1.97 ± 0.46	0.38 ± 0.06	0.48 ± 0.18	0.00000
1620	78.74	5.66	0.11 ± 0.02	0.07 ± 0.01	0.17 ± 0.05	0.33 ± 0.03	0.00000
2004	26.08	6.17	0.50 ± 0.13	0.17 ± 0.09	nd.	nd.	0.00000
2006	21.84	6.24	nd.	nd.	3.81 ± 0.36	0.85 ± 0.28	0.00000
2103	34.77	6.13	1.02 ± 0.17	0.85 ± 0.04	nd.	nd.	0.00000
2204	47.74	6.03	0.23 ± 0.04	0.21 ± 0.02	0.59 ± 0.09	0.72 ± 0.35	0.00001
2216	50.38	6.25	0.59 ± 0.26	0.70 ± 0.20	6.77 ± 1.66	10.19 ± 0.90	0.00000
2219	48.03	5.99	nd.	nd.	0.29 ± 0.02	0.54 ± 0.30	0.00000
2221	50.42	6.14	0.27 ± 0.05	0.29 ± 0.06	0.45 ± 0.06	0.79 ± 0.25	0.00085
2318	53.66	6.25	0.43 ± 0.06	0.31 ± 0.02	nd.	nd.	0.00000
2320	51.78	6.12	nd.	nd.	0.61 ± 0.10	0.54 ± 0.31	0.00000
2407	63.66	6.09	2.03 ± 0.31	1.73 ± 0.73	nd.	nd.	0.00000
2413	61.80	6.16	1.07 ± 0.25	1.18 ± 0.19	0.52 ± 0.02	0.48 ± 0.08	0.00188
2415	63.30	6.18	3.69 ± 0.48	4.09 ± 0.95	nd.	nd.	0.00000
2417	60.18	6.25	6.02 ± 1.26	6.68 ± 0.44	1.19 ± 0.39	0.98 ± 0.41	0.00000
2421	62.02	6.02	0.15 ± 0.06	0.13 ± 0.06	0.48 ± 0.16	1.51 ± 0.34	0.00000
2425	65.20	5.93	nd.	nd.	0.30 ± 0.02	0.44 ± 0.07	0.00000
2512	74.04	6.15	1.81 ± 0.04	1.89 ± 0.42	0.37 ± 0.02	0.37 ± 0.10	0.00000
2722	99.32	6.06	0.19 ± 0.16	0.22 ± 0.11	0.42 ± 0.09	0.80 ± 0.36	0.00579
3005	26.25	6.36	0.38 ± 0.08	0.33 ± 0.04	0.21 ± 0.03	0.16 ± 0.03	0.02180
3006	25.01	6.37	0.57 ± 0.01	0.36 ± 0.08	nd.	nd.	0.00000
3102	38.74	6.32	0.34 ± 0.05	0.28 ± 0.02	0.09 ± 0.02	0.04 ± 0.03	0.00000
3103	42.90	6.34	1.72 ± 0.34	2.01 ± 0.46	6.88 ± 0.11	8.30 ± 0.98	0.00000
3116	43.08	6.42	0.23 ± 0.05	0.18 ± 0.03	0.98 ± 0.34	0.59 ± 0.23	0.00000
3210	46.10	6.43	0.23 ± 0.06	0.17 ± 0.05	0.51 ± 0.06	0.62 ± 0.18	0.00001
3403	64.47	6.27	0.70 ± 0.14	0.80 ± 0.25	0.26 ± 0.01	0.39 ± 0.05	0.00041
3404	63.75	6.30	1.41 ± 0.11	1.75 ± 0.16	nd.	nd.	0.00000
3509	74.93	6.33	0.52 ± 0.12	0.41 ± 0.21	nd.	0.04 ± 0.01	0.00000
3516	74.36	6.44	0.38 ± 0.10	0.50 ± 0.12	0.03 ± 0.00	0.05 ± 0.01	0.00000
3618	88.48	6.43	0.31 ± 0.05	0.23 ± 0.05	0.51 ± 0.12	0.80 ± 0.25	0.00019
4006	19.40	6.61	0.79 ± 0.16	0.62 ± 0.23	nd.	nd.	0.00000
4009	25.35	6.65	0.29 ± 0.11	0.12 ± 0.07	nd.	nd.	0.00000
4108	36.60	6.65	0.23 ± 0.11	0.09 ± 0.03	0.11 ± 0.03	0.12 ± 0.03	0.01794
4403	63.06	6.50	0.20 ± 0.10	0.14 ± 0.01	0.53 ± 0.06	0.38 ± 0.07	0.00001
4405	66.61	6.52	1.76 ± 0.89	1.69 ± 0.75	1.07 ± 0.33	0.74 ± 0.14	0.03840
4411	60.39	6.64	1.98 ± 1.14	1.32 ± 0.89	0.67 ± 0.22	0.48 ± 0.10	0.00745
4420	60.78	6.52	nd.	nd.	0.52 ± 0.28	0.35 ± 0.18	0.00000
4503	71.02	6.50	0.55 ± 0.27	0.63 ± 0.16	0.11 ± 0.03	0.11 ± 0.01	0.00000
4516	69.04	6.69	0.35 ± 0.06	0.25 ± 0.04	0.44 ± 0.17	0.72 ± 0.43	0.01925
4605	87.74	6.51	1.02 ± 0.54	1.28 ± 0.42	2.88 ± 0.58	3.22 ± 0.40	0.00005
5001	25.38	6.79	0.27 ± 0.06	0.29 ± 0.08	0.19 ± 0.05	0.14 ± 0.03	0.06144
5004	19.74	6.84	0.84 ± 0.17	0.62 ± 0.10	nd.	nd.	0.00000
5007	27.71	6.87	0.10 ± 0.04	0.09 ± 0.02	0.33 ± 0.08	0.36 ± 0.05	0.00000
5114	43.19	6.90	nd.	nd.	0.57 ± 0.10	0.45 ± 0.01	0.00000
5203	44.50	6.83	8.34 ± 2.24	7.64 ± 2.53	3.03 ± 0.57	2.78 ± 0.53	0.00005
5211	49.04	6.96	0.51 ± 0.22	0.57 ± 0.37	0.60 ± 0.08	1.24 ± 0.31	0.07915
5215	47.38	6.99	8.59 ± 3.38	8.88 ± 2.21	0.47 ± 0.03	0.22 ± 0.05	0.00000
5310	55.32	6.87	0.19 ± 0.02	0.14 ± 0.03	0.27 ± 0.05	0.36 ± 0.03	0.01332
5319	53.64	6.96	0.36 ± 0.10	0.65 ± 0.09	1.76 ± 0.57	3.83 ± 0.94	0.00000
5411	64.07	6.95	1.36 ± 0.71	1.47 ± 0.43	0.59 ± 0.03	0.36 ± 0.04	0.00000
5413	60.15	6.95	1.30 ± 0.73	1.39 ± 0.45	7.66 ± 1.67	20.59 ±	0.00000
5514	69.95	6.91	0.15 ± 0.02	0.16 ± 0.05	0.37 ± 0.12	0.45 ± 0.09	0.00013
5706	101.95	6.85	6.53 ± 1.75	7.06 ± 3.19	3.14 ± 0.56	2.78 ± 1.71	0.02285
6007	22.26	7.08	0.69 ± 0.27	0.62 ± 0.13	0.43 ± 0.08	0.30 ± 0.13	0.02140
6010	24.10	7.14	0.52 ± 0.05	0.32 ± 0.07	nd.	nd.	0.00000
6202	43.82	7.03	0.30 ± 0.05	0.45 ± 0.13	nd.	nd.	0.00000

Spot Number	Mr (kDa)	pI	S0	S3	SS-1	SS-4	FDR
							<i>Cont.</i>
6203	45.04	7.04	0.48 ± 0.06	0.58 ± 0.12	0.25 ± 0.07	0.16 ± 0.04	0.00003
6207	48.03	7.13	0.23 ± 0.07	0.27 ± 0.08	0.50 ± 0.06	0.83 ± 0.20	0.00003
6212	49.41	7.20	0.09 ± 0.02	0.07 ± 0.03	1.99 ± 0.22	2.07 ± 0.28	0.00000
6214	46.76	7.22	0.10 ± 0.03	0.11 ± 0.05	0.53 ± 0.07	0.54 ± 0.01	0.00000
6221	48.94	7.02	nd.	nd.	0.76 ± 0.04	0.68 ± 0.09	0.00000
6312	54.17	7.27	0.42 ± 0.21	0.28 ± 0.10	0.20 ± 0.05	nd	0.00000
6401	61.45	7.00	10.32 ± 2.74	10.00 ± 2.95	1.56 ± 0.28	1.59 ± 0.66	0.00000
6418	60.22	7.08	0.30 ± 0.07	0.25 ± 0.07	0.48 ± 0.02	1.20 ± 0.25	0.00000
6502	67.34	7.01	0.48 ± 0.18	0.72 ± 0.30	1.52 ± 0.14	2.43 ± 0.32	0.00000
6604	91.88	7.04	0.32 ± 0.02	0.37 ± 0.19	1.92 ± 0.70	2.82 ± 0.60	0.00000
6611	92.81	7.15	0.20 ± 0.05	0.26 ± 0.01	0.43 ± 0.18	0.76 ± 0.47	0.00310
6617	92.57	7.26	0.40 ± 0.28	0.75 ± 0.21	2.37 ± 0.61	3.54 ± 1.00	0.00000
6716	97.03	7.20	0.11 ± 0.04	0.13 ± 0.03	0.65 ± 0.16	1.23 ± 0.27	0.00000
7114	34.95	7.55	0.54 ± 0.13	0.42 ± 0.12	1.26 ± 0.58	3.09 ± 0.48	0.00000
7116	38.14	7.33	nd.	nd.	nd.	0.84 ± 0.53	0.00000
7209	43.37	7.40	0.11 ± 0.02	0.13 ± 0.06	0.48 ± 0.07	0.57 ± 0.08	0.00000
7226	43.45	7.35	nd.	nd.	0.66 ± 0.03	0.53 ± 0.02	0.00000
7313	54.26	7.49	6.46 ± 1.85	3.59 ± 0.46	nd.	nd	0.00000
7317	52.56	7.57	1.33 ± 0.45	0.45 ± 0.18	0.12 ± 0.02	0.19 ± 0.04	0.00000
7616	82.41	7.58	0.50 ± 0.11	0.29 ± 0.08	0.49 ± 0.08	0.82 ± 0.43	0.01500
8005	22.76	7.84	0.70 ± 0.19	0.70 ± 0.17	nd.	nd	0.00000
8018	22.54	7.97	9.54 ± 5.83	8.20 ± 1.57	1.30 ± 0.57	1.41 ± 0.67	0.00000
8025	20.86	7.61	nd.	nd.	0.45 ± 0.07	0.86 ± 0.23	0.00000
8103	38.81	7.61	0.27 ± 0.06	0.19 ± 0.08	0.32 ± 0.14	0.37 ± 0.10	0.18859
8201	49.42	7.62	0.86 ± 0.56	0.75 ± 0.29	0.66 ± 0.12	0.24 ± 0.03	0.00059
8202	47.20	7.62	0.30 ± 0.20	0.50 ± 0.13	nd.	nd	0.00000
8301	53.85	7.59	13.65 ± 3.60	9.03 ± 0.60	nd.	nd	0.00000
8306	53.54	7.71	0.53 ± 0.01	0.40 ± 0.10	nd.	nd	0.00000
8411	60.87	7.84	0.27 ± 0.06	0.28 ± 0.04	0.61 ± 0.19	3.81 ± 1.74	0.00000
9406	60.86	7.98	nd.	0.40 ± 0.05	0.81 ± 0.07	2.16 ± 1.65	0.00000

Data are mean ± SD of three biological replicates. nd.: not detected spots; they were considered qualitative differences. Numbers in bold correspond to spots which changed in intensity between not imbibed (S0) and S3

Seed germination is a complex physiological period in which many biochemical processes are initiated or resumed ⁶⁶. Nevertheless, in this study only few changes were detected during germination *sensu stricto* (S0 vs. S3) by 2-DE approach compared with previous reports. Only five protein spots changed in intensity between S0 (not imbibed) and S3 stages in *Q. ilex* seeds (in bold in Table 6.4 and indicated with a “g” in Table 6.5). Changes in the abundance (up- and down-regulation) of higher number of proteins has been described during germination *sensu stricto* in *Arabidopsis* ³ and other orthodox seeds ^{10; 67}. However, our results are similar to that obtained in *Hevea brasiliensis* germination, also a recalcitrant seeds, in which only eight differential proteins were identified ¹⁵. In previous Ch. 4 and Ch. 5 we demonstrated at the transcriptional level that the mature holm oak seed maintains a relatively high level of metabolic activity, as described for other recalcitrant seeds. In consequence, minor differences between not imbibed (S0) and germinated embryo axis (S3) should be expected.

Qualitative as well as quantitative variable spots were observed when 2-DE protein profiles of the shoot seedling tissue (SS-1, SS-4 stages) and germinated seeds (S3) were compared. A total of 87 spots were differentially accumulated, being the differences qualitative (mainly absence in seedling samples) for 26 of them, and showing >2-fold changes for other 61 protein spots (quantitative changes, mainly up-accumulation in seedling samples) (Table 6.3). Differentially accumulated protein spots in mature, germinated embryo and shoot seedlings were identified combining results of a stringent sequence similarity search against the

Quercus_DB protein database ⁴⁷ combined with UniProtKB/TrEMBL, UniProtKB/SwissProt and NCBI nr databases. Ninety out of 103 analyzed spots were identified and showed MOWSE scores greater than 70. For most of them (89 spots), only one significant protein match was found. Only in one spot (spot 514) two different proteins were identified. Sixteen identified proteins were found in more than one spot, which resulted in a total of 72 non-redundant proteins (Table 6.5).

Identified proteins listed in Table 6.5 were classified in categories based on their putative function according to Kyoto Encyclopedia of Genes and Genomes (KEGG) (Fig. 6.5). The principal categories: carbohydrate metabolism (29 proteins), amino acids metabolism (14 proteins), energy metabolism (12 proteins), protein metabolism (7 proteins), biosynthesis of secondary metabolites (5 proteins) and oxidation-reduction process (5 proteins)

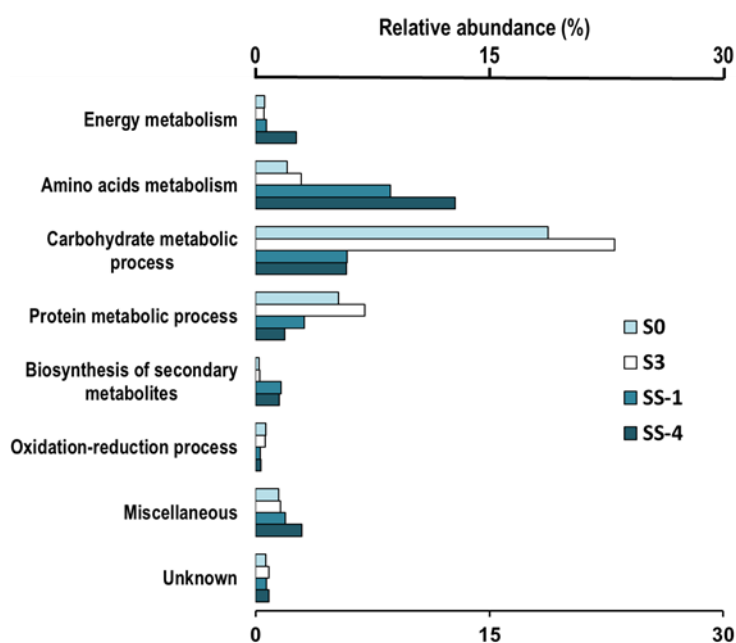


Figure 6.5: Distribution of identified proteins in functional categories and the percentage of proteins within each functional category as a function of identified proteins. Values represent the mean contribution of each category expressed as percentage of total spots volume of the identified proteins respect to the total spot volume in gel per analysed stages. Identified proteins in S0, S3, SS1 and SS4 represented 27.6%, 31.0%, 28.2 and 31.4% of total spot volumes respectively.

In mature (S0) and germinated seeds (S3) the most represented category corresponded to *carbohydrate metabolism* with 29 identified proteins, though little differences in identified spots abundance were found between the two stages (Fig. 6.5). The importance of this category diminished in seedling growth (Fig. 6.5). Most proteins included in this category was involved in glycolysis and implicated in storage mobilization (Table 6.5). Their accumulation in mature and germinated seeds (e.g. spot 6401: enolase, Fig. 6.3) indicates an active glycolysis that provides the energy necessary to complete the germination process and to

initiate seedling development¹⁰. Some enzymes in this category (*i.e.*, spot 6313: citrate synthase; spot 6716: sucrose synthase, Fig. 6.3) up-accumulated in S0 and S3, are also implicated in the mobilization of fatty acids and other in sucrose and starch mobilization and degradation of polysaccharide resources to provide substrate to the glycolytic pathway during germination^{68, 23, 69}. These processes are not necessary in photosynthetically active seedlings and this fact may explain, at least partially, the apparent decrease in the proteins related to carbohydrate metabolism in seedling.

The seven proteins corresponding to the *protein metabolism* functional category exhibited a reduction in expression as growth progressed. In this category are included proteins as heat shock protein (HSP60, spot 1508) and elongation factor (EF, spot 5706), described as developmentally regulated, being abundant in dry mature seeds, and disappearing during germination in both *Arabidopsis*³ and the recalcitrant *A. angustifolia*²¹ seeds, which is agreement with our data (Fig. 6.3 and Table 6.5). We did not find any indication of differential abundance in any of the studied developmental stages of proteins involved in proteolysis in germinated holm oak seeds. This was unexpected, given that proteases participate in all aspects of the plant life cycle, including the mobilisation of storage proteins during seed germination⁷⁰⁻⁷¹. This may be caused by the breaking down during germination of small amounts of storage proteins mediated by stored proteases⁷², which are not differentially accumulated during germination. However, it can not be excluded that proteases, which generally have acid isoelectric point, are lost during the isoelectric focusing step⁷³⁻⁷⁴.

Four proteins related to *oxidation-reduction processes*, DHAR (spot 6007), SOD (spot 5004), and two GSTs (spots 2004 and 4006), were highly abundant in mature and germinated seeds (Fig. 6.3 and Table 6.5) and scarce or absent in seedlings. Our results suggest that antioxidative defense mechanisms have been activated during holm oak seed maturation, leading to the accumulation of antioxidative enzymes to adapt the acorn to support a certain grade of desiccation (Ch.4). Rehydration of seeds has been associated with high levels of ROS production leading high ROS production during germination⁷⁵. The decrease in antioxidative defences observed along the developmental process can be, however, explained as ROS have important roles in endosperm weakening, mobilization of seed reserves, and many other process during seed germination and vigorous seedling⁷⁶⁻⁷⁷. In addition, this data corroborated the hypothesis that GST accumulated in mature, ungerminated seed, contributes to prevent the germination process as the protein synthesis results blocked because of GSH deficiency⁷⁸⁻⁷⁹. In germinating seeds, down-regulation of GST would restore the GSH levels. Data of SOD and GST protein abundance agree with those referred in previous Ch. 4 at the enzymatic activity (SOD) and the transcriptional (SOD and GST) levels. Transcription and translation are far from having

a linear and simple relationship and many reports indicated that the correlation between mRNA and protein abundances in the cell is notoriously poor. However, our data sustain that transcript levels, at least for SOD, GST and RBCL (see below) and in our experimental system, provide good predictive value with respect to the extent of protein changes in abundance.

Proteins included in all other functional categories were less abundant in mature and germinated seeds and increased their expression during seedling growth (Fig. 6.3; Table 6.5). The category of proteins related to *energy metabolism* included ATPase (spot 2425); OEE1 (spot 1001), a protein related to photosystem II; and RBCL (spot 6418). The increase in the abundance of these proteins along the germination process is in accordance with the transition of a heterotrophic phase to a photosynthetic and autotrophic phase. The results suggest that the plastid translational apparatus is established early during plant development, presumably to allow the development of the photosynthetic system of which several components are encoded by plastid genes, as was previously described in Arabidopsis³.

The amino acids produced from the breakdown of storage proteins by proteases are transported to growing seedling and for this purpose, where participate in the de novo biosynthesis of amino acids and proteins and the synthesis of many other metabolites. In the category of *amino-acid metabolism*, nine different proteins, corresponding to 14 spots, were included. Six of these differentially accumulated protein spots identified enzymes involved directly or indirectly in the metabolism of methionine (Met). Met is a fundamental metabolite because it functions not only as a building block for protein synthesis but also as the precursor of S-adenosylmethionine (SAM), the universal methyl-group donor, and as the precursor of polyamines, the plant-ripening hormone ethylene, and the vitamin biotin. The de novo biosynthesis of Met is catalyzed by 5-methyltetrahydropteroyltriglutamate-homocysteine S-methyltransferase (MET6), the cobalamin-independent MS activity present in plants. Three different protein spots (6617, 6604 and 6611) corresponding to MET6 were identified as differentially expressed in seedling (Table 6.5). This protein was present at low level in dry mature seeds, and its level was increased strongly after imbibition, as described for Arabidopsis seeds⁸⁰. SAM synthase (spots 9406 and 8411) catalyses the conversion of Met in SAM, the precursor of the ethylene and the spermidine/spermine biosynthesis pathways. SAM synthetase is considered a key regulator of metabolism in the transition from a quiescent to a highly active state during Arabidopsis seed germination¹⁹. Its accumulation was observed at the radicle emergence step (S3) and continued during seedling (Fig. 6.3.B; Table 6.5).

Three protein spots with high intensity in seedlings were identified as phospho-2-dehydro-3-deoxyheptonate aldolase (DHAPS, spots 9406 and 8411) and shikimate dehydrogenase (SD, spot 6502), involved in the shikimate pathway. This biosynthetic sequence

is employed by plants to synthesize aromatic amino acids and many secondary compounds including hormones (IAA) and pigments (anthocyanins). This pathway is important in the production of constituents for active cellular metabolism in germinating tissues. Data in Fig. 3.B and Table 6.5 agree with previous reports showing the increase in DAHP synthase specific activity during seed germination and seedling development in several plant species ⁸¹.

Finally, five spots identified three different proteins related to the biosynthesis of secondary metabolites, which showed higher abundance in the seedling growth phase (Fig. 6.5). Flavanone 3-hydroxylase (spot 3116) and chalcone synthase (spots 6221, 6207 and 6212) are implicated in the synthesis of flavonoids. The synthesis of these compounds has been observed to increase (at the transcriptional level) at the early stage of seed development in Arabidopsis ⁸².

In summary, by using a 2-DE proteomic approach, we separated the complex *Q. Ilex* protein mixtures isolated from embryo axes and early seedling, and identified the metabolic switches that experiment this recalcitrant seed after imbibition. Seventy two unique proteins changed in abundance during the postgermination period. Proteins involved in the metabolism of carbohydrate and proteins and in the oxidation/reduction processes were highly accumulated in germinated and mature seeds and decreased during postgermination. Data indicate that holm oak acorns are protected against oxidative stress caused by the partial desiccation they experimented during maturation, which only partially affects the seed metabolism. These results agreed with the transcriptional data presented in Ch.4 and Ch.5.

The end of the germination period defines the metabolic changes that permit seedling establishment and growth. An increase in the abundance of proteins included in the energy metabolism category determines the transition of a heterotrophic phase to a photosynthetic and autotrophic phase. An important increase in the biosynthesis of amino acids and several derived metabolites (ethylene, spermidine/spermine, hormones, pigments, flavonoids) helps the development and establishment of the seedling.

6.3.3 Identification and quantification of differentially accumulated proteins through *n*LC-MS Orbitrap LTQ analysis

The gel-based approach described in the previous section disclosed important metabolites changes that occurred in the holm oak seed after the germination (from S3 onward). However, few changes were detected during the germination (S0 vs. S3), affecting only to protein of the Met metabolism (SAMS, NLP3), the shikimate pathway (DHAPS) and the antioxidative defence. A gel free approach was hence used to analyse the mature unimbibed (S0) and the germinated seed (S3), trying to improve the coverage of proteome.

Protein identification was performed with combined database searches including Quercus_DB⁴⁷, UniprotKB/TrEMBL, UniprotKB/SwissProt, and NCBI nr protein databases, with taxonomy restrictions to *Viridiplantae* and using the SEQUEST algorithm implemented in the Proteome Discoverer v.1.4 software. This approach allowed the identification of 1650 protein species, 1155 in embryo axes at S0 stage, and 1198 in the S3 stage. With the same considerations made in the previous section, the percentage of *Q. ilex* proteome that was analysed with the gel-free approach was estimated in about 5%, more than twice the percentage of proteome covered with the gel based approach. Proteins were subjected to GO annotation using Blas2GO⁴⁹, and about 1250 were grouped into 18 categories (other 400 proteins species -25% were not annotated). The categories more represented were those related to *protein metabolism* (15%), *response to abiotic stimulus* (10%), *carbohydrate metabolic processes* (10%), and *metabolism of aromatic and nitrogen compounds* (10% each) (Fig. 6.6).

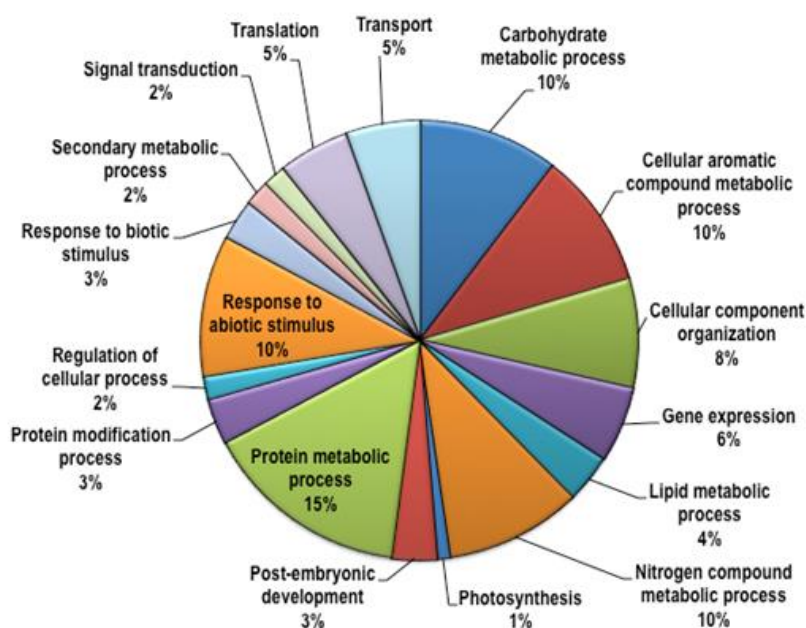


Figure 6.6: Functional categorisation of proteins identified in *Q. ilex* unimbibed (S0 stage) and germinated (S3 stage) seeds by nLC-MS Orbitrap LTQ analysis. Data was searched against the the Quercus_DB protein database combined with UniProtKB/TrEMBL, UniProtKB/SwissPrto and the NCBI nr databases, with taxonomy restrictions to Viridiplantae, using the SEQUEST algorithm. A total of 1250 proteins were annotated using gene ontology (Blas2GO) and categorized according to biological process domain.

Identified proteins were quantified by a standard peptide count measurement using a NSAF (Normalized spectral abundance factor) approach^{46;53}. A new list with 153 proteins with differential accumulation was generated, which included proteins enriched in non-imbibed seeds (80 proteins, being 64 specific of S0 seeds) and proteins enriched in germinated embryos

(73 proteins, of which 65 were specific of S3 seeds). Both groups were classified according to the KEGG pathway or the UniProt protein database (Fig. 6.7 and Table 6.6). As *Q. ilex* genome has not been sequenced, only $\approx 40\%$ spots were successfully identified.

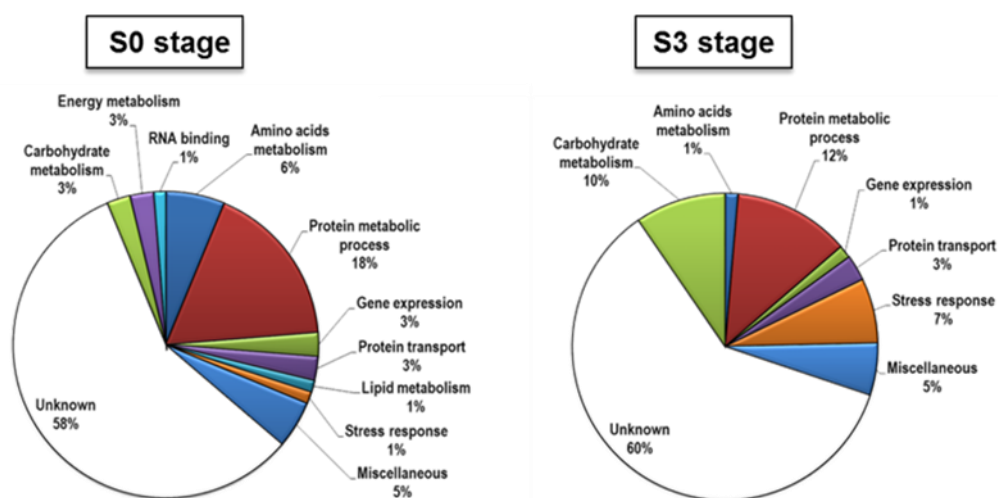


Figure 6.7: Categorization of differentially identified proteins in *Q. ilex* unimbibed (S0 stage) and germinated (S3 stage) seeds by nLC-MS Orbitrap LTQ analysis. A total of 80 proteins enriched in S0 seeds and 73 proteins enriched in S3 seeds were annotated using gene ontology (Blas2GO) and categorized according to biological process domain.

Protein metabolism and *amino acids metabolism* were the two predominant categories with a 24% of the 64 annotated proteins enriched in non-imbibed seeds. Most of proteins included in the first category were related to protein folding. The HSP60 and other heat shock proteins were also identified by 2-DE as abundant in *Q. ilex* seed before becoming a young seedling, but no differences were detected between S0 and S3 seeds. Other important groups of proteins included in this category were related to protein degradation, something that it was not found by the 2-DE proteomic approach. The casein lytic proteinase B (CLPB) proteins are chaperones that act to remodel/disassemble protein complexes and/or aggregates. These enzymes have been associated to protein degradation or processing during early germination and plastid proteolysis and to degradation of precursor proteins in plant mitochondria. Leucine aminopeptidase (LAP) is an exopeptidase that play important role in the mobilization of storage proteins at the cotyledon during seed germination. It has been reported that proteinase inhibitors accumulate as storage proteins during the development of seeds and tubers. High activities of all these proteins in the imbibed seeds with subsequent decline during germination of holm oak seeds (Table 6.6) resemble the sequence of events in other plant genera^{83 84}. Decarboxylation, deamination and transamination reactions of amino acids by the proteins included in the amino acids metabolism category, are catabolic reactions that drive the formation of ketoacids, which are important for fueling the respiration pathways during

germination. Several of these proteins accumulated only in embryo axes has been proposed to function in the signaling pathway that leads to radicle protrusion, growth, or preparation for biotic interactions⁸⁵.

Of note in the identification of dehydrin as a stress protein accumulated in S0 acorns that decline during germination, as these results are in agreement with transcriptional data presented in Ch.4, supporting the idea that changes in mRNA expression are reflected at the protein level.

In germinated S3 seeds, the protein metabolism category were also the most represented, with 12% of total annotated accumulated proteins (Fig. 6.7 and Table 6.6). However, in contrast to S0, the proteins included in this category at the S3 stages are involved in the resumption of protein synthesis (*i.e.*, several eukaryotic initiation factors 4A) and folding (*i.e.*, chaperonin CPN60) needed to sustain postgerminative growth.

Some proteins include in the *carbohydrate metabolism* group were also identified in the 2-DE analysis, but without differences between S0 and S3 seeds. Proteins related to the metabolism of starch and sucrose, glycolysis and the TCA cycle were accumulated in germinated seed, providing substrates for the obtention of energy and the biosynthetic pathways needed for seedling establishment.

A third category is important in S3 seeds, with a 7% of total proteins. Proteins related to *stress response* and *oxidation-reduction processes* were accumulated in germinated seeds. Annexins are a multigene family in most plant species and are suggested to play a role in a wide variety of essential cellular processes, including germination and involved in ABA and abiotic stress responses. Our results indicate that annexin is present in germinating seeds and that their abundance increase at S3 acorns, coinciding with data reported for other plant species. Annexins may be implicated in cell expansion due to their involvement in the Golgi-mediated secretion of the components of the plasma membrane and cell wall, leading to growth restoration after germination⁸⁶. Germination is accompanied by extensive change in the redox state of seed proteins. Proteins present in oxidized form in dry seeds are converted to the reduced state following imbibition, becoming functional. The increase in abundance of thioredoxin (Trx) appears to play a role in this transition⁸⁵. The continuous increase of these defence proteins during early germination strongly implies that they play a role in protecting the germinating seeds from possible stresses, including ROS generated after seed imbibition. This feature has also been found in orthodox seeds such as *Arabidopsis* and tomato^{9;87}. The differentially expressed proteins related to oxidation-reduction process between S0 and S3 were not found using the 2-DE MALDI-TOF/TOF approach.

The results obtained by *nLC-MS Orbitrap LTQ analysis* were complementary to those obtained in the 2-DE experiment, as both techniques cover slightly different parts of the proteome. Compared to the gel based approach, the shotgun analysis using label-free peptides has several advantages such as greater coverage of the proteome, less sample manipulation, and a dye- or tag-binding independent quantification ³¹.

The contribution of this experimental approach in an orphan species such as *Q. ilex* is important. However, several authors uphold that it is difficult to draw clear conclusions about gel free/label free proteomic quantitative data and the results need to be verified ⁸⁸. Nevertheless, the diverse transcriptional and proteomic approaches reported in this Thesis give similar clues about the metabolic state of the mature *Q. ilex* seed before the germination starts, and the metabolic switched experimented by the imbibed accord till the seedling is established.

6.3.4 Combined database search improves protein identification

The construction of a customised database for a particular species of genera improves the protein identification in two aspects: the number of protein identified and the confidence of their identification (see apPendix 10.1, p. 225) ⁴⁷.

In the MALDI-TOF/TOF analysis, the combination of a database specific to *Quercus*, the Quercus_DB protein database ⁴⁷, with UniProtKB/TrEMBL, UniProtKB/SwissProt and NCBI nr databases in Mascot search engine improved the coverage of annotated proteome. Most of the proteins were commonly identified with two or three different databases. In addition, six proteins were identified with Uniprot_KB and six with Quercus_DB, as shown in the Venn diagram (Fig. 6.8.A). The density plots of the scores of the identified proteins by 2-DE followed MALDI-TOF/TOF analysis (Fig. 6.8.C) showed that Uniprot-SwissProt, a manually curated protein database, had the highest score in a small number of proteins species. Coverage of the proteome was improved in the gel free approach by the use of the SEQUEST algorithm in combined database search. A number of proteins were identified in only one database, which illustrates the importance of using of the combined database search in protein identification (Fig. 6.8.B).

The Quercus_DB showed the lowest density in the peak of protein scores between 2.25 and 10 being the curve shifted to the right, a pattern which is more evident in *nLC LTQ Orbitrap*. It is clear that the scores obtained when employing these databases are, in general, higher than those obtained employing NCBI_Viridiplantae, UniProtKB/TrEMBL (Viridiplantae) and UniProtKB/SwissProt (Viridiplantae) (Fig. 6.8.D). These results show the importance of the database used in proteomic analysis.

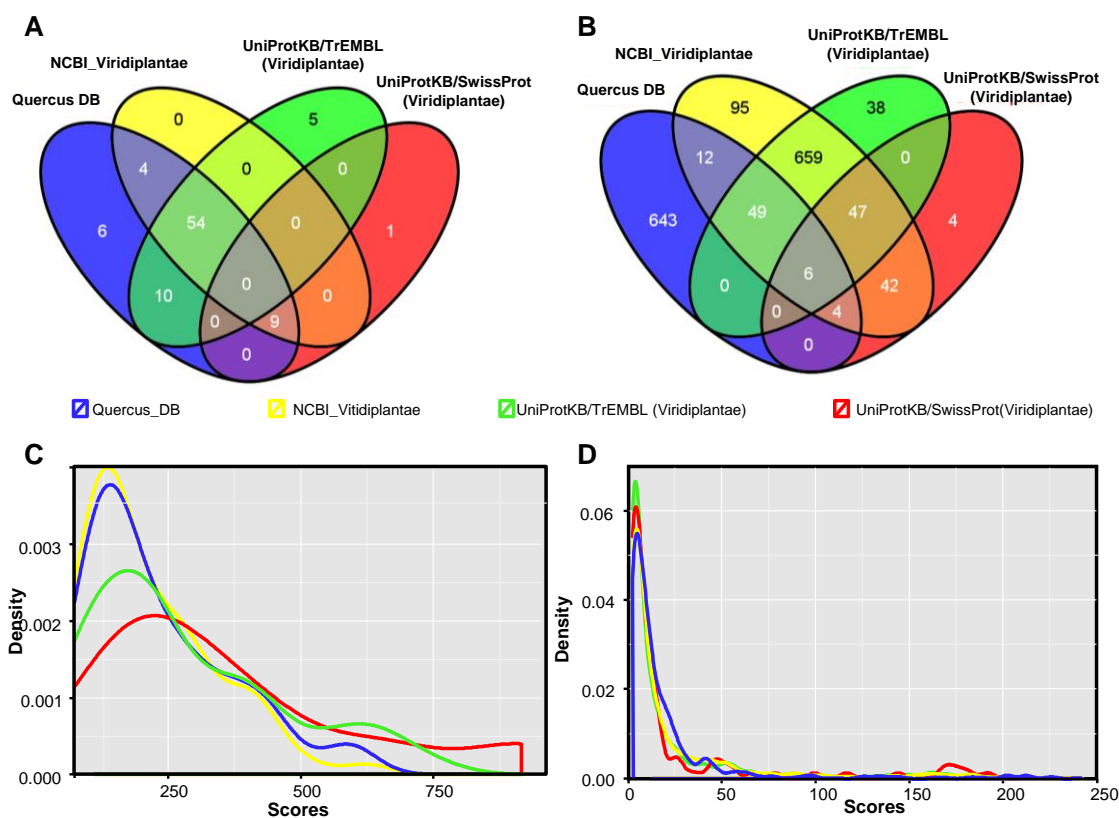


Figure 6.8: Differentially abundant identified proteins by MALDI-TOF/TOF and *n*LC LTQ Orbitrap using Mascot SEQUEST algorithms respectively, by using combined protein databases search including a custom genus specific database. Venn diagrams representing the number of identified proteins by MALDI-TOF/TOF (**A**) and *n*LC-LTQ Orbitrap (**B**); diagrams were plot using Venny (<http://bioinfogp.cnb.csic.es/tools/venny/>). Density plots of the scores of the proteins identified when MALDI-TOF/TOF and Mascot were employed (**C**). Density plots of the scores when *n*LC LTQ Orbitrap and the SEQUEST algorithm were used (**D**). The four indicated databases were used in each case. Blue lines indicate the developed databases, corresponding to Quercus_DB. Areas under each curve are equal and normalised to arbitrary value 1.

6.4 Concluding remarks

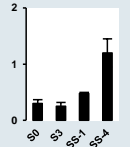
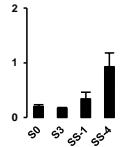
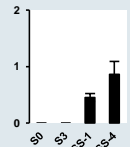
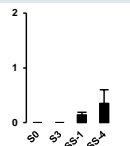
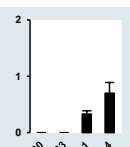
The metabolic process implicated in germination and seedling establishment of non-orthodox seed of *Q. ilex* was similar to that described for orthodox seeds. The proteomic changes observed in *Q. ilex* seeds affected mainly to proteins related to carbohydrate metabolism, amino acid metabolism and oxidative stress response. The up-accumulation of proteins related to glycolysis in mature and germinated seeds suggested that this process is essential for the energy supply and to provide molecules that serve as intermediaries in other metabolic pathways in *Q. ilex* seeds. Therefore, it can be hypothesized that the mature non-orthodox seeds of *Q. ilex* have the machinery necessary to resume rapidly metabolic activities and start the germination process. This is in contrast with the situation described for orthodox seeds, in which all metabolic activity ceases in mature dry seeds. Proteins related to energy metabolism and photosynthesis were up-accumulated during seedling establishment, thus ensuring the machinery necessary to synthesise *de novo* the biomolecules required for growth.

The results also indicated that the use of genus specific database combined with public database improve the quality and quantity of protein identification in orphan species.

The contribution of this experimental approach in an orphan species such as *Q. ilex* is important. The diverse proteomic approaches gave similar clues about the metabolic state of the mature *Q. ilex* seed before the germination starts, and the metabolic switched experimented by the imbibed accord until the seedling is established. Data are in fully agreement with those obtained at the transcriptional level (Ch. 4 and Ch. 5), thereby strengthening each other. Further studies should investigate post-translational modifications that occur during germination, such as phosphorylation, which are important in the regulation of germination and other processes of plant life cycle.

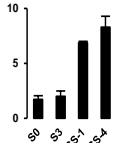
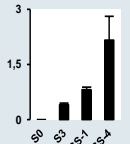
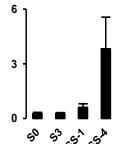
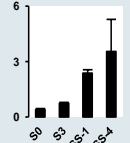
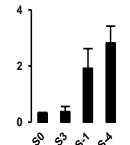
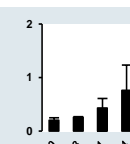
6.5 Appendixes

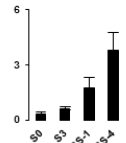
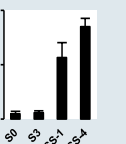
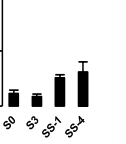
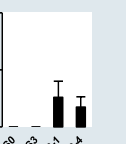
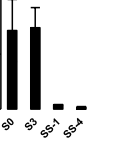
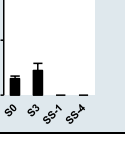
Table 6.5: List of proteins differentially abundant and identified by MALDI-TOF/TOF MS in *Q. ilex* seeds. Only protein spots that changed in abundance at least 2-fold in all three replicates for a given development stage are included. The positions of the spots are shown in Fig. 3. Spot normalized intensity values (mean of three biological replicates and standar deviation) are shown. These values were subjected to ANOVA statical test (FDR<0.05). Functional classification was based on Kyoto Encyclopedia of Genes and Genomes (KEGG) or biological process described in UniProt.

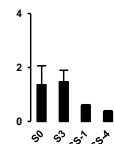
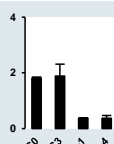
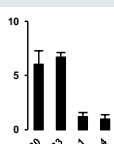
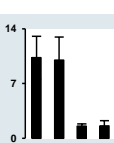
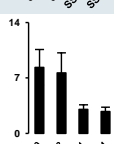
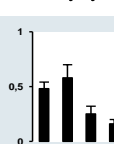
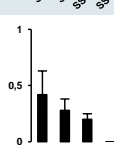
Spots Numbers ^a	Protein name	Accession numbers ^b	Mr (pI)		MOWSE score ^e	Peptides match	Sequence coverage (%)	Fragmented Ion (Ion Score)	Normalized spots volume intensities ^f
			Theor. ^c	Exp. ^d					
Energy metabolism									
6418	Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (RBCL), <i>Quercus multinervis</i>	Q597K9	51.4 (6.13)	60.2 (7.1)	668	35	51	LTYTTPDYQTK (55) DTDILAAFR (76) ALRLEDLR (32) TFQGPPHGIQVER (102) YGRPLLGCTIKPK (34) GGLDFTKDDENVNSQPFR (26) DNGLLLHIR (86)	
1001	Oxygen evolving enhancer protein 1 (OEE1), <i>Bruguiera gymnorhiza</i>	Q9LRC4	35.3 (6.48)	27.5 (5.6)	190	15	32	NAPPEFQNTK (29) LTYTLDEIEGPFVSPDGTVK (38) DGIDYAAVTQLPGER (19) GGSTGYBNAVALPAGGR (23)	
8025	Oxygen-evolving enhancer protein 2, chloroplast precursor_AT1G06680.1, <i>Quercus</i> spp.	FN729552_4	12.5 (6.51)	20.9 (7.6)	108	5	50	SITDYGSPEEFLSK (20) TNTDFLPYNGEGFK (49) YEDNFDNSNSVVIINSTDKK (18)	
1217	Ribulose bisphosphate carboxylase/oxygenase activase, chloroplastic, <i>Malus domestica</i>	Q40281	48.2 (8.2)	53.1 (5.6)	244	18	27	GLAFDTSDDQDITR (65) SFQCELVFAK (45) YREAADIIR (29) LVDTFPGQSIDFFGALR (35)	
306	Ribulose bisphosphate carboxylase/oxygenase activase, chloroplastic, <i>Malus domestica</i>	Q40281	48.2 (8.2)	51.4 (5.5)	389	20	28	GLAFDTSDDQDITR (64) SFQCELVFAK (47) VPIIVTGNDFSTLYAPLIR (86) IGVCIGIFR (38) LVDTFPGQSIDFFGALR (65)	

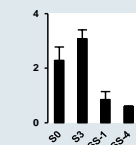
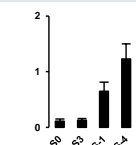
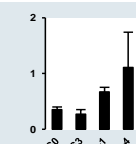
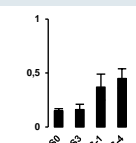
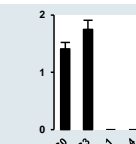
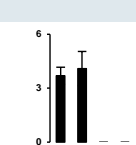
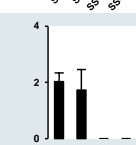
Spots Numbers ^a	Protein name	Accession numbers ^b	Mr (pI)		MOWSE score ^e	Peptides match	Sequence coverage (%)	Fragmented Ion (Ion Score)	Normalized spots volume intensities ^f
			Theor. ^c	Exp. ^d					
307	Putative uncharacterized protein (Ribulose biphosphate carboxylase/oxygenase activase 1_B9T427) ^g , <i>Vitis vinifera</i>	D7THJ7	52.3 (5.7)	51.6 (5.1)	226	23	37	GLAFDESDDQDITR (48) SFQCELVFAK (8) MCCLFINDLDAGAGR (8) VPIHVTGDNDFSTLYAPLIR (3) IGVCTGIFR (24) LVDTFPQSIDFFGALR (28)	
1205	Phosphoribulokinase, <i>Populus trichocarpa</i>	B9GZT5	45.3 (5.9)	44.1 (5.7)	171	12	26	ANNFDLMYEQVK (47) KPDFDAYIDPQK (56) FYGEVTQMLK (24)	
7616	10-formyltetrahydrofolate synthetase_AT1G50480.1, <i>Quercus</i> spp.	TC24939_32	43.4 (6.3)	82.4 (7.6)	98	8	45	GAPTGFTLPIR (24) GAPTGFTLPIRDVR (16) YSLGTPQCAVIVATIR (8) NAALAAGAYDAVICTHHAHGGK (31)	
3005	Transcription factor APFI family protein(Uncharacterized protein_U5GRZ5, Gamma carbonic anhydrase 1) ^g , <i>Populus trichocarpa</i>	A9PFJ3	29.5 (6.2)	26.2 (6.4)	169	9	35	LQGNYYFQEQLSR (95) GSSIWYGCVLR (29)	
1415	ATP synthase subunit alpha, chloroplastic, <i>Castanea mollissima</i>	E3W0R3	55.4 (5.4)	65.5 (5.8)	660	27	38	VVNTGTVLQVGDGIAR (93) IAQIPVSEAYLGR (93) VINALAKPIDGR (39) LIESPAPGHSR (79) EAYPGDVLYLHSR (106) TNKPQFQEIHSSTK (86)	
2425	ATP synthase subunit alpha (ATPase), chloroplastic, <i>Anredera baselloides</i>	H2BBA4	55.3 (5.2)	62.5 (5.9)	601	22	32	VVNTGTVLQVGDGIAR (102) IAQIPVSEAYLGR (83) LIESPAPGHSR (59) HTLIYDDLK (83) EAYPGDVLYLHSR (112) TNKPQFQEIHSSTK (67)	

Spots Numbers ^a	Protein name	Accession numbers ^b	Mr (pI)		MOWSE score ^e	Peptides match	Sequence coverage (%)	Fragmented Ion (Ion Score)	Normalized spots volume intensities ^f
			Theor. ^c	Exp. ^d					
2421	ATP synthase subunit beta, chloroplastic, <i>Castanea sativa</i>	Q6QBP2	53.9 (5.4)	62.0 (6.0)	916	40	59	IAQHGPVLDVTFPPR (99) DTAGQQINVTCEVQQLGNR (97) GMEVIDTGAPLSVPVGGATLGR (107) IFNVLGEVDNLGPDVTR (103) AHGGVSVFPGVGER (122) VALVYGMNEPPGAR (24) FVQAGSEVSALLGR (81) GIYPAVDPLDSTSTMLQPR (30)	
Amino acid metabolism									
6502	Shikimate dehydrogenase, <i>Ricinus communis</i>	B9REY0	57.6 (6.1)	67.3 (7.0)	383	16	25	IATTALDITDCAR (89) FGGYLTYGALEAGAISAPGQPTAK (69) DLLDLYNFR (83) SPLLFNAAFK (69) ESGAVIVYGTEMLIR (13)	
6214	Glutamate dehydrogenase, <i>Vitis vinifera</i>	Q1HDV6	44.8 (6.4)	46.8 (7.2)	211	21	39	SLLIPFR (21) SLLIPFREIK (22) DDGTLASFVGR (26) FHGYSPAVVTGKPTDLGGSLGR (17) MGAFTLGVNR (14)	
4108*	Omega-amidase (NLP3)_Q8RUF8, <i>Quercus</i> spp.	TC22204_23	36.6 (6.8)	36.6 (6.7)	237	8	28	GDLYQLVDVK (13) FQELAMIYAAR (10) NLNITIVGGSIPER (58) AVDNQLYVATCSPAR (71) IFNTCCVFGTDGNLK (36) TLTAGETPTIVDTEVGR (55)	
3403	Glutamate decarboxylase_AT2G02010.1, <i>Quercus</i> spp.	TC19169_41	58.0 (5.9)	64.5 (6.3)	208	12	45	VVIREDFSR (30) ETPEEIATYWR (53) GSSQIIAQYYQFVR (69) NYVDMDEYPTTELQNR (46)	
2219	Glutamine synthetase, <i>Avicennia marina</i>	A5A7P7	48.0 (6.4)	48.0 (6.0)	175	6	8	TIPKPVEHPSELPK (44) AILNLSLR (31) HKEHISAYGEGNER (92)	

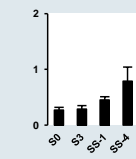
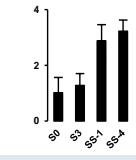
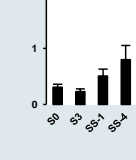
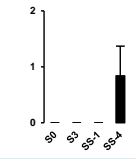
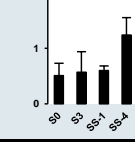
Spots Numbers ^a	Protein name	Accession numbers ^b	Mr (pI)		MOWSE score ^e	Peptides match	Sequence coverage (%)	Fragmented Ion (Ion Score)	Normalized spots volume intensities ^f
			Theor. ^c	Exp. ^d					
3103	Glutamine synthetase (Fragment), <i>Glycine max</i>	Q43759	17.5 (8.8)	42.9 (6.3)	230	8	40	HKEHIAAYGEGNER (127) HETADMNTFVWGVANR (52)	
9406 ^g	Phospho-2-dehydro-3-deoxyheptonate aldolase 1 (DHAPS), <i>Solanum tuberosum</i>	P21357	60.0 (8.9)	60.9 (7.8)	279	15	28	TIDEFPPIVFAGEAR (76) AYCQSAATLNLLR (69) QLDGAHVEFLR (78)	
8411	Phospho-2-dehydro-3-deoxyheptonate aldolase 1 (DHAPS), <i>Solanum tuberosum</i>	P21357	60.0 (8.9)	60.9 (7.8)	444	21	33	TIDEFPPIVFAGEAR (99) AYCQSAATLNLLR (82) AFATGGYAAMQR (86) AFATGGYAAMQR (19) QLDGAHVEFLR (84)	
6617	5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase, putative, <i>Ricinus communis</i>	B9SI90	84.9 (6.1)	92.6 (7.3)	593	25	25	GKYL FAGVVDGR (105) YLFAGVVDGR (74) AGINVIQIDEAALR (72) EGVKYGAGIGPGVYDIHSPR (88) YGAGIGPGVYDIHSPR (121)	
6604	5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase, <i>Ricinus communis</i>	B9SI90	84.9 (6.1)	91.2 (7.0)	633	27	31	GKYL FAGVVDGR (80) YLFAGVVDGR (86) AGINVIQIDEAALR (88) EGVKYGAGIGPGVYDIHSPR (77) YGAGIGPGVYDIHSPR (149)	
6611	5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase, <i>Ricinus communis</i>	B9SI90	84.9 (6.1)	92.8 (7.1)	400	15	12	GKYL FAGVVDGR (74) YLFAGVVDGR (95) AGINVIQIDEAALR (59) YGAGIGPGVYDIHSPR (129)	

Spots Numbers ^a	Protein name	Accession numbers ^b	Mr (pI)		MOWSE score ^e	Peptides match	Sequence coverage (%)	Fragmented Ion (Ion Score)	Normalized spots volume intensities ^f
			Theor. ^c	Exp. ^d					
5319 ^g	S-adenosylmethionine synthase (SAMS), <i>Glycine max</i>	I1JPQ2	43.2 (6.2)	53.6 (7.0)	695	22	42	TCPWLRPDGK (39) YLDDKTFHLNPSGR (111) TIFHLNPSGR (83) FVIGGPHGDAGLTGR (134) FVIGGPHGDAGLTGRK (79) TAAYGHFGR (74) TAAYGHFGRDPDFWETVK (53)	
2216	S-adenosylmethionine synthase 2, <i>Elaeagnus umbellata</i>	Q9AT55	43.6 (5.5)	50.4 (6.2)	650	22	44	TIGFVSDDVGLDADNCK (83) VLVNIQQSPDIAQGVGHFTK (97) TQVTVEYYNDKGMVPPV (17) TIFHLNPSGR (61) FVIGGPHGDAGLTGR (110) FVIGGPHGDAGLTGRK (89) TAAYGHFGR (79)	
3210	Fumarylacetoacetase, putative_AT1G12050.1, <i>Quercus</i> spp	TC23862_21	30.7 (5.5)	46.1 (6.4)	131	4	20	GQGHFAGNSPPYFGPSLK (78) KFLEDGDEVIFSGYSK (50)	
Carbohydrate metabolic process									
4420	UDP-glucose 6-dehydrogenase, <i>Vitis vinifera</i>	A5AVX9	53.5 (6.4)	60.8 (6.5)	274	16	30	AADLTYWESAAR (60) LAANAFLAQR (53) LSIYDPQVTEDQIQR (67) IYDNMQKPAFVFDGR (19)	
5215	Alcohol dehydrogenase, <i>Alnus glutinosa</i>	Q4A1D2	41.7 (6.28)	47.4 (7.0)	463	21	35	AAVAWEAGKPLVIEEVAPPQANEVR (108) GQTPLFPR (45) FGVTEFVNPK (65) TLKGTFFGNYKPR (62) GTFFGNYKPR (81)	
6202	Alcohol-dehydrogenase family protein, <i>Populus trichocarpa</i>	B9HEP6	41.8 (6.1)	43.8 (7.0)	369	24	44	GQTPLFPR (45) KGSSVAIFGLGAVGLAAAEGAR (43) FGVTEFVNPK (62) TLKGTFFGNYKPR (15) GTFFGNYKPR (61)	

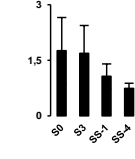
Spots Numbers ^a	Protein name	Accession numbers ^b	Mr (pI)		MOWSE score ^e	Peptides match	Sequence coverage (%)	Fragmented Ion (Ion Score)	Normalized spots volume intensities ^f
			Theor. ^c	Exp. ^d					
5411	Mitochondrial lipoamide dehydrogenase 1_AT1G48030.2, <i>Quercus</i> spp	TC26056_36	58.1 (8.31)	64.1 (6.9)	156	12	38	IVSSTGALALTEIPK (28) GALGGTCLNVGCIPSK (48) FSSVEIDLPAAMMAQK (13) ALLHSSHYHEAQHSFANHGVK (25)	
2512	2,3-biphosphoglycerate-independent phosphoglycerate mutase_AT1G09780.1, <i>Quercus petraea</i>	QP1063_77	61.0 (6.0)	74.0 (6.1)	476	10	25	DAILSGKFDQVR (46) FGHVTFWNGNR (77) AFEYEDFDKDFDR (67) LPSHYLVSPPEIDR (78) GTLHLIGLLSDGGVHSR (98) IQLTSHTCQPVPPIAIGGGLAPGCR (88) AHGSAVGLPTEDDMGNSEVGHNALGAGR (31)	
2417	Enolase 1, <i>Hevea brasiliensis</i>	Q9LEJ0	48.0 (5.6)	60.2 (6.2)	622	27	41	AAVPSGASTGIYEALR (128) LAMQEFMILPVGASSFK (41) VQIVGDDLLVTNPK (87) VNQIGSVTESIEAVK (74) IEEELGSEAVYAGANFR (147)	
6401	Enolase (ENO), <i>Annona cherimola</i>	COL7E2	48.1 (5.7)	61.4 (7.0)	757	29	50	KIPLYQHIANLAGNK (94) IPLYQHIANLAGNK (90) TLVLPVPAFNVINGGSHAGNK (113) VNQIGSVTESIEAVK (85) HAGWGVMAHR (35) SGETEDTFIADLSVGLATGQIK (106) FRAPVQPY (48)	
5203	Phosphoglycerate kinase_AT1G79550.1, <i>Quercus petraea</i>	QP6550_24	42.5 (6.15)	44.5 (6.8)	321	11	43	YSLKPIVPR (37) VILSTHLGRP (48) ELDYLVGAVSNPK (90) FLKPAVAGFLMQK (54) VDLNVPLDDNFNITDDTR (82)	
6203	Phosphoglycerate kinase_AT1G79550.1, <i>Quercus rubra</i>	QRU405_58	43.5 (6.7)	45.0 (7.0)	231	11	51	YSLKPIVPR (31) VILSTHLGRP (30) FLKPAVAGFLMQK (21) LVAEIPGGVLLLENVR (26) LASLADLYVNDAFGTAHR (49)	
6312	Citrate synthase, <i>Populus trichocarpa</i>	B9N619	53.8 (7.7)	54.2 (7.3)	196	11	17	VVPGFGHVLR (89) HLPDDPLFQLVSK (70)	

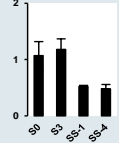
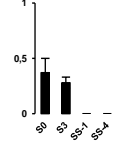
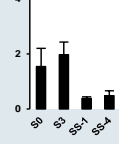
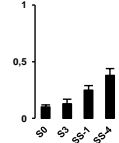
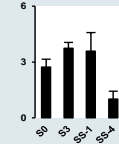
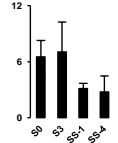
Spots Numbers ^a	Protein name	Accession numbers ^b	Mr (pI)		MOWSE score ^e	Peptides match	Sequence coverage (%)	Fragmented Ion (Ion Score)	Normalized spots volume intensities ^f
			Theor. ^c	Exp. ^d					
1006	Putative uncharacterized protein (Lactoylglutathione lyase_D2D330) ^a , <i>Vitis vinifera</i>	F6H7L5)	46.9 (5.7)	31.0 (5.9)	241	20	26	FLHVVYR (34) GPTPEPLCQVMLR (25) GPTPEPLCQVMLR (6) SAEVNVLVTKELGGK (35) ITSFLDPDGWK (72)	
6716	Sucrose synthase (SUS), <i>Citrus unshiu</i>	Q9SLS2	92.7 (5.9)	97.0 (7.2)	220	20	18	KHLTEGAFGEVLR (7) SIGNGVEFLNR (51) FQEIGLER (27) LLPDAVGTTCGQR (56)	
104	Fructokinase-2-like_AT1G06020.1, <i>Quercus</i> spp.	TC24048_19	37.5 (5.4)	37.3 (5.2)	421	8	28	TALAFVTLR (51) NFHGEVEAFR (54) APGGAPANVAIAVSR (106) IVDDQSILEDEPR (56) LGDDEFGHMLAGILR (15) ENGVVDEGINFDKGR (107) IVDDQSILEDEPRLR (34)	
5514	Pyrophosphate-dependent phosphofructokinase beta subunit, <i>Citrus sinensis</i> x <i>Citrus trifoliata</i>	A9YVC9	62.0 (6.3)	69.9 (6.9)	283	17	22	DKIETPEQFK (59) STGKYHFVVR (45) YYHFVR (37) QSHFFGYEGR (83)	
3404	Beta glucosidase 17_AT2G44480.1, <i>Quercus</i> spp.	QRO15180_40	47.8 (5.2)	63.8 (6.3)	119	10	48	GAYDFIGVNYTISR (103)	
2415	Beta glucosidase 17_AT2G44480.1, <i>Quercus</i> spp.	FN703157_40	27.8 (4.8)	63.3 (6.2)	86	5	38	GAYDFIGVNYTISR (72)	
2407	Beta glucosidase 17_AT2G44480.1, <i>Quercus</i> spp.	FN703157_40	27.8 (4.8)	63.7 (6.1)	100	4	35	GAYDFIGVNYTISR (95)	

Spots Numbers ^a	Protein name	Accession numbers ^b	Mr (pI)		MOWSE score ^e	Peptides match	Sequence coverage (%)	Fragmented Ion (Ion Score)	Normalized spots volume intensities ^f
			Theor. ^c	Exp. ^d					
8201	Cytosolic NADP+-dependent isocitrate dehydrogenase_AT1G65930.1, <i>Quercus rubra</i>	QRU821_31	48.3 (7)	49.4 (7.6)	153	11	33	HAFGDQYR (39) YFDLGLPYR (47) SEGGYWACK (49) LVPGWTKPICIGR (18)	
3509	Pyruvate decarboxylase, <i>Prunus armeniaca</i>	B0ZS79	66.2 (5.7)	74.9 (6.3)	235	12	23	ILHHTIGLPDFSQELR (124) EPVPFSLSPR (69)	
3516	Pyruvate decarboxylase, <i>Ricinus communis</i>	B9SWY1	66.6 (5.7)	74.4 (6.4)	243	14	19	ILHHTIGLPDFSQELR (126) EPVPFSLSPR (56) NWNVTGLVDIAIHNGEGK (9) VSAANSRPPNPQ (11)	
4503	Pyruvate decarboxylase family protein, <i>Populus trichocarpa</i>	B911N8	64.6 (5.8)	71.0 (6.5)	172	10	16	ILHHTIGLPDFSQELR (141)	
4411	UDP-glucose pyrophosphorylase, <i>Annona cherimola</i>	COL7E5	51.5 (5.8)	60.4 (6.6)	246	13	30	DGWYPPGHGDVFPSLR (96) VQLLEIAQVPDEHVNEFK (94)	
2318	Glucose-1-phosphate adenyltransferase, <i>Vitis vinifera</i>	D7TDB6	56.2 (6.5)	53.7 (6.2)	350	26	38	VDTTILGLDDER (59) KVPDFSFYDR (71) SSPIYTQPR (41) IINSDNVQEAAAR (40)	
7226	Putative uncharacterized protein (UDP-glucuronic acid decarboxylase 1_W9R8F1) ^g , <i>Vitis vinifera</i>	A5AXR4	39.1 (8.3)	43.5 (7.3)	389	28	45	FFQSNMR (11) VAETLMFDYHR (21) IFNTYGPR (53) VVSNFIAQALR (69) GEPLTVQAPGTQTR (46) SFCYVSDMVDGLIR (20)	

Spots Numbers ^a	Protein name	Accession numbers ^b	Mr (pI)		MOWSE score ^e	Peptides match	Sequence coverage (%)	Fragmented Ion (Ion Score)	Normalized spots volume intensities ^f
			Theor. ^c	Exp. ^d					
2221	Putative uncharacterized protein(UDP-D-apiose/UPD-D-xylose synthetase_D2D333) ^g , <i>Medicago truncatula</i>	B7F157	44.2 (5.7)	50.4 (6.1)	257	12	24	LIHFSTCEVYVK (37) MDFIPGIDGPSEGVPV (49) VLACFSNLLR (53) ANGHIFNVGNPNNEVTVR (80)	
4605	Transketolase, putative, <i>Ricinus communis</i>	B9RDA1	81.6 (6.5)	87.7 (6.5)	403	24	26	NPYWFNR (59) NPYWFNRDR (38) NGNTGYDEIR (61) ALPTYTPESPADATR (105) MFGDFQKDTPEER (39)	
3618	Putative uncharacterized protein (transketolase family protein_B9GPE7) ^g , <i>Populus trichocarpa</i>	A9PHE2	81.1 (5.9)	88.5 (6.4)	319	22	19	NPYWFNR (39) NGNTGYDEIR (52) ALPTYTPESPADATR (96) VLPGLLGGADLASSNMTLLK (21) MFGDFQKDTPEER (25) RPSILALSR (30)	
7116	Fructose-bisphosphate aldolase_AT4G38970.1, <i>Quercus</i> spp	TC20955_14	32.4 (5.9)	38.1 (7.3)	162	7	31	AAQEALLR (44) IVDVLVEQK (43) TAAYYQQGAR (25) NTPQIADYTLK (24) TVVSIPIGPTLAVK (28)	
5211	GDP-D-mannose-3',5'-epimerase, <i>Malpighia glabra</i>	A0EJL8	42.9 (5.8)	49.0 (7.0)	429	20	48	ISITGAGGFIASHIAR (50) SFTFIDECVEGVLR (94) KLPIHHIPGPEGV (75) VVGTPAPVQLGSLR (99)	

Protein metabolic process

4405	Mitochondrial processing peptidase beta subunit, <i>Cucumis melo</i>	Q9AXQ2	58.9 (6.6)	66.6 (6.5)	495	21	25	ALDILADILQNSK (44) TITKDHLSYIQTHTYAPR (148) SLLHLHIDGTSPPVAEDIGR (109) RIPFAELFAR (50) IPFAELFAR (43) TYWNRV (23)	
------	--	--------	------------	------------	-----	----	----	--	---

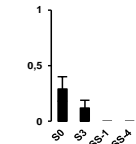
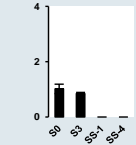
Spots Numbers ^a	Protein name	Accession numbers ^b	Mr (pI)		MOWSE score ^e	Peptides match	Sequence coverage (%)	Fragmented Ion (Ion Score)	Normalized spots volume intensities ^f
			Theor. ^c	Exp. ^d					
2413	Leucyl aminopeptidase 1 _ AT2G24200.1, <i>Quercus alba</i>	QA490_35	50.5 (5.8)	61.8 (6.2)	330	10	40	GLTFDSGGYNIK (68) AGQSTVLRPLPGLGSK (48) VGLLGLGQSASTPAAFR (83) RVGLLGLGQSASTPAAFR (68) LNSATAIASGTVLGLYEDNRYK (62)	
1505	Chaperonin-60kD, ch60, putative, <i>Ricinus communis</i>	B9RWQ2	61.6 (5.8)	72.3 (5.7)	98	15	23	GYISPYFITNQK (34)	
1508	Heat shock protein 60 (HSP60)_AT3G23990.1, <i>Quercus spp</i>	TC33448_39	63.9 (5.6)	71.1 (5.8)	436	11	23	AGHDPVKVIR (51) IGVQIIQNALK (91) NVVIEQSWGAPK (82) GYISPYFITNQK (73) SDEIAQVGTISANGER (72) AAVEEGIVPGGGVALLYASK (75)	
205	30S ribosomal protein S1, chloroplastic, <i>Spinacia oleracea</i>	P29344	45.0 (5.4)	50.9 (5.3)	116	12	20	GFVPFSQISSK (44) EIPLKFVEVDDEQSR (21) VSDIATVLQPGDTLK (14)	
614	Uncharacterized protein (heat shock cognate 70 kDa protein 4-like isoform 1_B9SR13) ^g , <i>Oryza brachyantha</i>	J3M7V9	71.0 (5.1)	82.0 (5.4)	579	40	39	TTPSYVAFTDTER (90) NAVVTVPAYFNDSQR (88) HNEPTAAAIAYGLDKK (91) STAGDTHLGGEDFDNR (9) ARFEELNMDLFR (3) VQQLLQDFFNKG (49) EQVFSTYSDNQPGVLIQVYEGER (64)	
5706	Elongation factor (EF), <i>Ziziphus jujuba</i>	G9JJS4	94.9 (5.9)	101.0 (6.9)	520	26	20	LWGENFFDPATKK (67) GFVQFCYEPIK (66) YRVENLYEGPLDDAYANAIR (85) VENLYEGPLDDAYANAIR (118) VIYASQLTAKPR (57) GHVFEEMQRPGTPLYNIK (33)	

Spots Numbers ^a	Protein name	Accession numbers ^b	Mr (pI)		MOWSE score ^e	Peptides match	Sequence coverage (%)	Fragmented Ion (Ion Score)	Normalized spots volume intensities ^f
			Theor. ^c	Exp. ^d					
Biosynthesis of secondary metabolites									
3116	Flavanone 3-hydroxylase (Fragment), <i>Acer palmatum</i>	I7DG89	34.7 (5.5)	43.1 (6.4)	92	13	26	HTDPGTITLLQDQVGLQATR (6) LSIATFQNPAPAEATVYPLK (14) SVLEEPITFAEMYR (13) SVLEEPITFAEMYRR (4)	
6221	Chalcone synthase 1, <i>Camellia sinensis</i>	P48386	42.9 (5.9)	48.9 (7.0)	168	16	37	AEGPATVMAIGTATPPNCVDQSTYPDYYFR (7) VLVVCSEITAVTFR (73)	
6207	Chalcone synthase (Fragment), <i>Cardamine monteluccii</i>	Q1G6T3	42.6 (6.1)	48.0 (7.1)	165	18	32	LLGLRPSVK (36) LLGLRPSVKR (15) LMMYQQGCFAGGTVLR (16)	
6212	Chalcone synthase, <i>Gossypium hirsutum</i>	G3G7Y8	42.9 (6.0)	40.9 (7.2)	202	24	50	LMMYQQGCFAGGTVLR (5) VLVVCSEITAVTFR (15) EVGLTFHLLKDVPLISK (19)	
7209	Cinnamyl alcohol dehydrogenase_AT4G39330.1, <i>Quercus</i> spp.	TC22105_49	40.1 (6.9)	43.4 (7.4)	91	9	43	DSSGLLSPFHFSR (30) HLGVAGLGLGHVAVK (8) ILYCGVCHSDLHNVK (12) TYGGYSDIVVVDEHYVLR (17)	
Oxidation-reduction process									
2004*	Glutathione S-transferase omega (GSTO)_D6BR66, <i>Quercus</i> spp	TC18312_19	28.2 (6.6)	26.0 (6.0)	75	8	36	LYISLSCPYAQR (24) EAGPAFDHLENALSK (14) WIEEVNKIDAYKPTK (11) YIDSNFEGPSLLPNDHAK (24)	

Spots Numbers ^a	Protein name	Accession numbers ^b	Mr (pI)		MOWSE score ^e	Peptides match	Sequence coverage (%)	Fragmented Ion (Ion Score)	Normalized spots volume intensities ^f
			Theor. ^c	Exp. ^d					
4006	Putative uncharacterized protein (Glutathione-s-transferase theta_B9T0U8) ^g , <i>Vitis vinifera</i>	D7TP00	24.9 (6.2)	19.4 (6.6)	107	6	22	NPFQIPVLEDGDLTFESR (38) AWWEDISSRPAFK (46)	
6007	Dehydroascorbate reductase (DHAR)_AT1G75270.1, <i>Quercus</i> spp	TC18689_13	24.2 (6.6)	22.3 (7.1)	268	8	44	AHGPIYAGEK (58) YKELLSR (35) EHVIAGWEPK (60) LYHLDVALGHFKK (60) ALDEHLKAHGPIYAGEK (34) HPEPSLTPPPEFASVGSK (12)	
4516	Methylenetetrahydrofolate reductase 2_AT2G44160.1, <i>Quercus robur</i>	QR09558_22	38.8 (7.2)	69.0 (6.7)	121	7	38	RPTNVFR (24) AYVEFFCSR (63) TIGWDQYPHGR (32)	
5004	Manganese superoxide dismutase (SOD) 1_AT3G10920.1, <i>Quercus</i> spp	TC29211_11	19.3 (7.9)	19.7 (6.8)	126	5	55	HHQAYITNKN (73) FNGGGHINHSIFWK (42) KLVVDTTANQDPLVTK (2)	
Stress response									
8005	Pathogenesis-related thaumatin superfamily protein_AT2G28790.1, <i>Quercus petraea</i>	QP7437_20	13.3 (7.1)	22.8 (7.8)	183	7	84	GQCPVVGCR (14) ASTFSEFFK (46) SPPGHGPPVACK (40) ANLLATCPDKLQLR (30) SGCEAFGTDELCCR (44)	
8018	Pathogenesis-related thaumatin superfamily protein_AT2G28790.1, <i>Quercus</i> spp	TC25355_10	28.5 (7.9)	22.5 (8.0)	312	10	32	ASTFSEFFK (46) ANLLATCPDKLQLR (80) SGCEAFGTDELCCR (83) GGFALHLLTHHSFPAPTQHWSSR (45) HACPATFTYAHDSPLMHECSSR (62)	

Spots Numbers ^a	Protein name	Accession numbers ^b	Mr (pI)		MOWSE score ^e	Peptides match	Sequence coverage (%)	Fragmented Ion (Ion Score)	Normalized spots volume intensities ^f
			Theor. ^c	Exp. ^d					
Miscellaneous									
2722	ATP-dependent clp protease, <i>Cucumis melo</i> subsp. <i>melo</i>	E5GBL8	103.4 (6.8)	99.3 (6.1)	390	40	31	GSGFVAVEIPFTPR (44) TKNNPCLIGEPGVGK (39) YRGEFER (22) GELQCIGATTLDEYRK (45) NPNRPIASFIFSGPTGVGK (13) LIGSPPGYVGYTEGGQLTEAVR (27)	
5310	Predicted protein (Sulfoquinovosyldiacylglycerol_AT4G3303 0.1) ^g , <i>Populus trichocarpa</i>	B9HDL9	54.3 (8.2)	55.3 (6.9)	123	12	22	ATDLNQGVVYGVV (58) FCVQAAVGHPLTVYVK (10)	
3102	Glutelin type-B 5_W9SME0, <i>Quercus</i> spp	TC21080_27	39.3 (8.6)	38.7 (6.3)	112	10	44	AGNLFIVPR (39) LALDKNGFALPR (62)	
1012	Putative uncharacterized protein_D7SY44, <i>Quercus</i> spp	TC33675_14	21.3 (5.9)	17.3 (6.0)	207	6	55	RFPDLR (20) AAGIGAIQAVSR (61) SLDELSSFVK (65) GGFFDLGHPLLNR (52)	
1302	DEAD box RNA helicase, <i>Pisum sativum</i>	Q8H1A5	47.1 (5.4)	52.9 (5.6)	258	24	46	GIYAYGFEPKPSAIQQR (60) ILSSGVHVVTGTPGR (28) VFDMLRR (16) MFVLDEADEMLSR (11) VLITTDLLAR (21)	
1620	Cell division protein ftsH, putative, <i>Ricinus communis</i>	B9S304	75.5 (6.4)	78.7 (5.7)	224	21	26	FLEYLDKDR (48) VRVQLPGLSQELLQK (3) SSGGMGGPGPGFPLAFGQSK (53) ADILDSALLRPGR (17)	
4403	Anthranilate N-benzoyltransferase protein, putative, <i>Ricinus communis</i>	B9RIV2	50.4 (5.4)	63.1 (6.5)	184	8	10	ELKPEDYTVFTVFADCR (79) DAGVNCVAVGSSPR (84)	

Spots Numbers ^a	Protein name	Accession numbers ^b	Mr (pI)		MOWSE score ^e	Peptides match	Sequence coverage (%)	Fragmented Ion (Ion Score)	Normalized spots volume intensities ^f
			Theor. ^c	Exp. ^d					
7114	AT4G39230.1_NmrA-like negative transcriptional regulator family protein (encodes a protein whose sequence is similar to phenylcoumaran benzylic ether reductase (PCBER)Phenylcoumaran benzylic ether reductase) ^g , <i>Quercus robur</i>	QRO2324_17	36.5 (6.8)	35.0 (7.5)	367	11	35	AGHPTFALVR (67) VLIIGGTGYIGK (68) FYPSEFGNDVDR (67) AIFNKEDDIGTYTIK (17) NLGVTLVHGDLYDHGSLVK (103) FYPSEFGNDVDRVHAVDPAK (17) GDHTNFEIEPSFGVEASQLYDPVK (33)	
1207	Actin, <i>Brassica napus</i>	Q9ZSD7	41.9 (5.3)	43.8 (5.8)	601	26	58	AVFPSIVGRPR (81) VAPEEHPVLLTEAPLNPK (107) GYMFTTAAER (11) SSSSVEKNYELPDGQVITIGAER (60) NYELPDGQVITIGAER (83) GEYDESGPSIVHR (86)	
107	Alpha/beta-Hydrolases superfamily protein_ AT4G02340.1, <i>Quercus alba</i>	QA3147_23	35.4 (5.2)	37.6 (5.3)	88	9	27	FALQVPYR (24) KFPGIEDYIR (42)	
5001	Aluminium induced protein with YGL and LRDR motifs_AT3G22850.1, <i>Quercus</i> spp.	TC18137_21	27.8 (7.0)	25.4 (6.8)	136	6	33	GCFFTSSGGLR (39) FAMILYDSSSK (25) DRGPYPADQVVR (16) SYEHLNVEVKPVPR (44) SPEALQSPQSGSVSTLK (15)	
3006	Aluminium induced protein with YGL and LRDR motifs_AT3G22850.1, <i>Quercus</i> spp.	TC18137_21	27.8 (7.0)	25.0 (6.4)	141	5	27	GCFFTSSGGLR (31) FAMILYDSSSK (49) SYEHLNVEVKPVPR (58) SPEALQSPQSGSVSTLK (4)	
Unknown									
514	Putative uncharacterized protein, <i>Vitis vinifera</i>	D7SLM9	64.9 (5.8)	71.2 (5.5)	419	20	30	VVAAGANPVLTR (61) GYISPYFVTDSEK (74) DLINVLEDAIR (76) AAVEEGIVVGGGCTLLR (126)	
	RuBisCO large subunit-binding protein subunit beta, chloroplastic, <i>Pisum sativum</i> Alternative name: 60 kDa chaperonin subunit beta	P08927	63.3 (5.8)	71.2 (5.5)	414	21	34	VVAAGANPVLTR (61) SAENSLYVVEGMQFDR (54) GYISPYFVTDSEK (74) AAVEEGIVVGGGCTLLR (126)	

Spots Numbers ^a	Protein name	Accession numbers ^b	Mr (pI)		MOWSE score ^e	Peptides match	Sequence coverage (%)	Fragmented Ion (Ion Score)	Normalized spots volume intensities ^f
			Theor. ^c	Exp. ^d					
4009 ^g	Uncharacterized protein, <i>Oryza brachyantha</i>	J3MD07	101.7 (5.1)	25.0 (6.6)	81	19	26	-	
2103	Putative cyclase family protein, <i>Arachis hypogaea</i>	C0L2U1	31.5 (5.04)	38.4 (6.1)	79	8	28	IFDISHR (36)	

^a Spot number as given on the 2-DE gel images in Fig. 3.

^b Uniprot and Quercus_DB accessions numbers. The accession whose first letters were FN, TC, QP, QRU, QRO and QA correspond to Quercus_DB accessions, the rest correspond to Uniprot_KB.

^c Molecular weight (kDa) and isoelectric point of protein calculated for each database.

^d Molecular weight (kDa) and isoelectric point of protein calculated by using molecular weight standards and the PD-Quest Advance (8.01) software.

^e Mascot score ($S = -10 \times \log(P)$): where P is the probability that the observed match is a random event, peptide matched in MS analysis, percentage of sequence coverage (into the brackets), and ions sequence matched (ion score into the brackets) from MS/MS analysis.

^f The figures represent the normalized spots volume intensities vs analyzed stages.

^g Proteins differentially accumulated between S0 and S3.

Table 6.6: List of identified proteins from non-imbibed seeds (S0) and germinated (S3) *Q. ilex* embryos by using nLC MS Orbitrap LTQ.

Protein description (close species)	Accession				Specific function	Mr(pI)	Scores	Cov (%)	Peptides Mached	Ratio S3/S0
	UniprotKB/ SwissProt	UniprotKB /TrEml	NCBI	Quercus_DB						
Up-accumulated in S0 (Qualitative differences)										
Amino acids metabolism										
AT4G33010.1_glycine decarboxylase P-protein 1_HC, <i>Quercus petraea</i>				QP8418_19	Glycine, serine and threonine metabolism	38.1 (7.61)	10.2	16.6	3	S0 specific
Mitochondrial glycine decarboxylase complex P-protein, <i>Populus tremuloides</i>				TC19232_29	Glycine, serine and threonine metabolism	50.7 (6.25)	9.8	10.7	3	S0 specific
2-oxoglutarate dehydrogenase, E1 component isoform 3; <i>Theobroma cacao</i>			590563613		Tryptophan metabolism	80.1 (7.34)	29.9	16.9	8	S0 specific
Uncharacterized protein(Methylmalonate-semialdehyde dehydrogenase [acylating], mitochondrial) ^a , <i>Musa acuminata</i> subsp. <i>malaccensis</i>		M0SR49			Valine, leucine and isoleucine degradation	57.7 (7.66)	7.9	6.3	3	S0 specific
Protein metabolic process										
Casein lytic proteinase B4, <i>Theobroma cacao</i>			590571986		Protein folding	108.7 (7.01)	27.8	8.9	7	S0 specific
AT5G15450.1_casein lytic proteinase B3_HC, <i>Quercus robur</i>				QR023829_16	Protein folding	24.3 (7.34)	22.1	28.2	5	S0 specific
AT3G11830.1_TCP-1/cpn60 chaperonin family protein_HC, <i>Quercus petraea</i>				QP1036_36	Protein folding	64.0 (7.88)	41.3	20.9	9	S0 specific
AT3G18190.1_TCP-1/cpn60 chaperonin family protein_LC, <i>Quercus petraea</i>				QP4157_33	Protein folding	59.7 (7.94)	32.3	17.8	8	S0 specific
Uncharacterized protein (T-complex protein 1 subunit epsilon-like) ^a , <i>Solanum lycopersicum</i>		K4BYC0	460386958		Protein folding	59.1 (5.82)	14.8	14.6	5	S0 specific
AT5G57870.1_MIF4G domain-containing protein / MA3 domain-containing protein_HC, <i>Quercus rubra</i>				QRU1014_115	Protein biosynthesis	79.6 (7.53)	14.7	8.0	5	S0 specific
Leucine aminopeptidase family protein; <i>Populus trichocarpa</i>		B9IMD7	566215336			60.9 (6.93)	22.5	12.0	5	S0 specific
Trypsin inhibitor A, <i>Glycine max</i>	P01070	P01070	125020			24.0 (5.11)	8.8	14.4	3	S0 specific
Gene expression										
Short-chain dehydrogenase/reductase, <i>Cucumis melo</i> subsp. <i>melo</i>		E5GBL0	307135996			108.4 (6.89)	22.6	6.2	6	S0 specific
AT2G27040.1_Argonaute family protein (Overexpressor of cationic peroxidase 11_HC, <i>Quercus rubra</i>				QRU2927_19		25.7 (9.64)	10.7	16.0	4	S0 specific
Protein transport										

Protein description (close species)	Accession				Specific function	Mr(pI)	Scores	Cov (%)	Peptides Mached	Ratio S3/S0
	UniprotKB/SwissProt	UniprotKB/TrEmI	NCBI	Quercus_DB						
AT5G60790.1_ABC transporter family protein_HC, <i>Quercus alba</i>				QA725_59	Protein transport	50.0 (7.78)	12.6	13.3	4	S0 specific
ADP,ATP carrier protein 3, <i>Morus notabilis</i>		W9QNM9	587854593		Protein transport	40.4 (9.69)	7.5	9.7	3	S0 specific
Lipid metabolism										
Uncharacterized protein(3-ketoacyl-CoA thiolase 1) ^a , <i>Brassica rapa</i> subsp. <i>pekinensis</i>		M4EP66			Fatty acid degradation and elongation; alpha-Linolenic acid metabolism,	47.3 (8.53)	11.1	13.3	3	S0 specific
Stress response										
Dehydrin, <i>Quercus spp.</i>				TC23570_14	Water stress	27.8 (6.16)	8.6	12.9	3	S0 specific
Miscellaneous										
Annexin-like protein RJ4, <i>Quercus spp.</i>				TC18250_18		36.1 (6.76)	25.3	25.7	7	S0 specific
AT5G54500.2_flavodoxin-like quinone reductase 1_LC, <i>Quercus alba</i>				QA4445_25		15.4 (5.82)	8.8	34.7	3	S0 specific
AT3G17810.1_Dihydropyrimidine dehydrogenase activity_HC, <i>Quercus robur</i>				QR02393_29		46.6 (6.34)	17.2	14.8	5	S0 specific
Dynamamin-2A, putative, <i>Ricinus communis</i>		B9SBU7	223537294			90.1 (8.87)	12.4	4.6	3	S0 specific
Unknown										
Predicted protein, <i>Quercus petraea</i>				QP8315_20		54.1 (5.34)	8.2	6.7	3	S0 specific
AT3G08030.1_Protein of unknown function, DUF642_LC, <i>Quercus petraea</i>				QP8179_17		28.2 (8.18)	13.0	21.6	4	S0 specific
Predicted protein, <i>Quercus robur</i>				QR012610_7		15.3 (10.45)	11.5	34.1	3	S0 specific
Predicted protein, <i>Quercus rubra</i>				QRU49717_11		8.6 (5.10)	16.6	44.4	3	S0 specific
Hypothetical protein EUGRSUZ_F01462, <i>Eucalyptus grandis</i>			629102255			92.5 (6.11)	74.8	25.0	15	S0 specific
Hypothetical protein EUGRSUZ_F04463, <i>Eucalyptus grandis</i>			629105924			36.0 (5.29)	48.6	27.3	8	S0 specific
Putative uncharacterized protein, <i>Arabidopsis lyrata</i> subsp. <i>lyrata</i>		D7MD24	297312917			39.6 (8.13)	15.6	10.9	3	S0 specific
Putative uncharacterized protein, <i>Glycine max</i>		C6TB24	574584672			23.0 (9.92)	10.3	19.2	3	S0 specific
Putative uncharacterized protein, <i>Medicago truncatula</i>		B7FM02	217074850			62.0 (5.20)	39.2	20.8	9	S0 specific

Protein description (close species)	Accession				Specific function	Mr(pI)	Scores	Cov (%)	Peptides Mchet	Ratio S3/S0
	UniprotKB/SwissProt	UniprotKB/TrEm1	NCBI	Quercus_DB						
Putative uncharacterized protein, <i>Selaginella moellendorffii</i>		D8RV35	300157077			70.6 (5.27)	15.9	7.3	4	S0 specific
Predicted protein, <i>Quercus</i> spp.				TC33465_22		32.4 (5.15)	9.0	15.2	3	S0 specific
Predicted protein, <i>Quercus</i> spp.				TC20900_15		21.8 (10.11)	16.4	40.0	5	S0 specific
Predicted protein, <i>Quercus</i> spp.				TC20465_44		68.4 (5.24)	11.6	8.3	4	S0 specific
Hypothetical protein ZEAMMB73_184776, <i>Zea mays</i>			414590819			43.7 (6.13)	15.1	13.5	4	S0 specific
Uncharacterized protein (Fragment), <i>Amborella trichopoda</i>		W1NSR4	586661534			93.3 (5.97)	18.3	6.1	5	S0 specific
Uncharacterized protein (Fragment), <i>Populus trichocarpa</i>		U7DX74	566255443			42.2 (5.36)	8.6	10.6	3	S0 specific
Uncharacterized protein, <i>Amborella trichopoda</i>		U5D3T2	586766811			77.2 (6.81)	19.5	6.7	4	S0 specific
Uncharacterized protein, <i>Amborella trichopoda</i>		U5DB51	586774182			110.2 (6.74)	23.0	6.8	6	S0 specific
Uncharacterized protein, <i>Capsella rubella</i>		R0G3M7	565479068			71.1 (5.07)	90.8	17.3	9	S0 specific
Uncharacterized protein, <i>Capsella rubella</i>		R0F0U4	565438714			77.0 (5.31)	13.2	6.7	4	S0 specific
Uncharacterized protein, <i>Citrus clementina</i>		V4UFF9	567919092			56.7 (8.29)	12.3	10.3	4	S0 specific
Uncharacterized protein, <i>Citrus clementina</i>		V4SJT5	567870081			71.1 (5.59)	104.1	18.2	10	S0 specific
Uncharacterized protein, <i>Citrus clementina</i>		V4UFD6	567864300			71.0 (5.21)	173.9	31.2	16	S0 specific
Uncharacterized protein, <i>Genlisea aurea</i>		S8E2N1	527198343			60.6 (6.28)	11.2	5.2	3	S0 specific
Uncharacterized protein, <i>Lotus japonicus</i>		I3SR66	388509384			10.8 (9.64)	12.5	37.0	3	S0 specific
Uncharacterized protein, <i>Medicago truncatula</i>		B7FL88	388501384			71.0 (5.19)	85.7	14.6	7	S0 specific
Uncharacterized protein, <i>Musa acuminata</i> subsp. <i>malaccensis</i>		M0RFS3				121.8 (5.29)	16.7	5.3	4	S0 specific
Uncharacterized protein, <i>Musa acuminata</i> subsp. <i>malaccensis</i>		M0SWY3				60.7 (6.42)	7.5	6.3	3	S0 specific
Uncharacterized protein, <i>Phaseolus vulgaris</i>		V7D0J0	593731242			17.8 (5.99)	7.5	18.7	3	S0 specific
Uncharacterized protein, <i>Populus trichocarpa</i>		B9GLH4	566147195			66.6 (7.87)	16.0	8.7	4	S0 specific
Uncharacterized protein, <i>Prunus persica</i>		M5WQL1	595852973			107.3 (9.17)	25.2	7.9	6	S0 specific

Protein description (close species)	Accession				Specific function	Mr(pI)	Scores	Cov (%)	Peptides Mached	Ratio S3/S0
	UniprotKB/SwissProt	UniprotKB/TrEml	NCBI	Quercus_DB						
Uncharacterized protein, <i>Prunus persica</i>		M5WY92	595962605			71.7 (5.38)	87.0	17.1	9	S0 specific
Uncharacterized protein, <i>Prunus persica</i>		M5VND4	595796780			71.3 (8.25)	17.9	12.1	5	S0 specific
Uncharacterized protein, <i>Prunus persica</i>		M5WR30	595857560			59.0 (7.14)	9.7	7.2	3	S0 specific
Uncharacterized protein, <i>Prunus persica</i>		M5XSL7	596299993			55.2 (7.46)	13.9	8.3	3	S0 specific
Uncharacterized protein, <i>Prunus persica</i>		M5WB08	595824521			47.5 (5.11)	18.4	18.9	6	S0 specific
Uncharacterized protein, <i>Prunus persica</i>		M5XQ59	596141602			37.0 (7.52)	171.6	40.4	9	S0 specific
Uncharacterized protein, <i>Prunus persica</i>		M5W463	595836837			37.2 (8.72)	11.4	8.3	3	S0 specific
Uncharacterized protein, <i>Prunus persica</i>		M5XKJ3	596054717			23.0 (6.68)	10.7	19.9	3	S0 specific
Uncharacterized protein, <i>Thellungiella salsuginea</i>		V4M1R7	567200069			71.2 (5.29)	116.3	19.7	10	S0 specific
Uncharacterized protein, <i>Zea mays</i>		B8A0W7	219886883			48.1 (5.26)	64.8	21.5	9	S0 specific
Uncharacterized protein, <i>Zea mays</i>		C4IZQ5	238006740			54.6 (7.14)	11.9	10.3	3	S0 specific
Up-accumulated in S0 (Quantitative differences)										
Protein metabolic process										
AT5G50920.1_ Chaperone protein ClpC1_HC, <i>Quercus petraea</i>				QP1770_109	Protein folding	83.3 (6.60)	45.1	23.1	11	0.47
Ubiquitin-activating enzyme E1, putative, <i>Ricinus communis</i>		B9SKU8	0			123.4 (5.45)	14.2	5.9	5	0.47
Uncharacterized protein(Ubiquitin-activating enzyme E1 2-like) ^a , <i>Glycine max</i>		K7KA83	571441365			120.5 (5.25)	13.7	5.4	4	0.47
AT3G02530.1_TCP-1/cpn60 chaperonin family protein_HC, <i>Quercus robur</i>				QR03715_43	Protein folding	60.8 (6.02)	26.0	18.8	8	0.43
Elongation factor 1-alpha, <i>Nicotiana paniculata</i>		Q9ZWH9	257327235		Protein biosynthesis	49.3 (9.13)	50.9	20.9	6	0.22
Chaperone protein ClpB4, mitochondrial, <i>Arabidopsis thaliana</i>	Q8VYJ7	Q8VYJ7	75161490		Protein folding	108.6 (6.98)	17.2	6.1	5	0.35
Carbohydrate metabolism										
Xylose isomerase, <i>Quercus</i> spp.				TC18422_30	Fructose and mannosa metabolism	54.2 (6.05)	14.1	11.9	5	0.47
AT4G35830.1_ aconitase 1_HC, <i>Quercus rubra</i>				QRU1997_91	TCA cycle	57.4 (7.49)	14.2	8.9	4	0.40
Energy metabolism										

Protein description (close species)	Accession				Specific function	Mr(pI)	Scores	Cov (%)	Peptides Mached	Ratio S3/S0
	UniprotKB/SwissProt	UniprotKB/TrEm1	NCBI	Quercus_DB						
NADH-ubiquinone oxidoreductase, putative, <i>Ricinus communis</i>		B9T118	223528410		Oxidative phosphorylation	80.7 (6.95)	13.9	6.9	4	0.47
AT4G11150.1_vacuolar ATP synthase subunit E1_HC, <i>Quercus</i> spp.				TC18506_12		20.6 (8.53)	9.0	14.2	3	0.51
Amino acids metabolism										
AT4G14880.1_O-acetylserine (thiol) lyase (OAS-TL) isoform A1_HC, <i>Quercus petraea</i>				QP1332_18	Cysteine and methionine metabolism	25.8 (7.39)	13.6	19.1	4	0.44
RNA binding										
AT3G15010.1_RNA-binding (RRM/RBD/RNP motifs) family protein_HC, <i>Quercus robur</i>				QR03293_21		45.9 (8.40)	16.9	13.2	4	0.40
Unknown										
Predicted protein, <i>Quercus</i> spp.				TC22919_31		53.4 (7.02)	30.8	16.5	6	0.27
Predicted protein (TUDOR-SN protein 1 isoform 3) ^a , <i>Quercus</i> spp.				TC31886_28		47.9 (8.28)	18.9	15.2	6	0.31
Uncharacterized protein, <i>Oryza brachyantha</i>		J3LDA2	573919270			94.0 (6.16)	37.6	12.2	9	0.45
Uncharacterized protein, <i>Oryza glaberrima</i>		I1NT64				70.9 (5.21)	69.0	20.4	8	0.47
Up-accumulated in S3 (Qualitative differences)										
Protein metabolic process										
Putative TCP-1/cpn60 chaperonin family protein isoform 1, <i>Zea mays</i>		C0P530	413953492		Protein folding	61.7 (5.49)	38.6	16.6	7	S3 specific
Chaperonin CPN60-like protein, <i>Medicago truncatula</i>		G7LHF7	355521119		Protein folding	61.6 (7.39)	19.2	9.0	4	S3 specific
Eukaryotic initiation factor 4A-8-like, <i>Cucumis sativus</i>			449476633		Protein biosynthesis	46.8 (5.57)	34.0	26.2	7	S3 specific
Translational initiation factor 4A-1, <i>Arabidopsis thaliana</i>		A8MRZ7	145332383		Protein biosynthesis	45.7 (6.28)	43.3	38.6	10	S3 specific
Eukaryotic initiation factor 4A-1, <i>Oryza sativa</i> subsp. <i>japonica</i>	P35683	P35683	97536398		Protein biosynthesis	47.1 (5.57)	54.0	37.7	11	S3 specific
Eukaryotic initiation factor 4A-3, <i>Arabidopsis thaliana</i>	Q9CAI7	Q9CAI7	75333652		Protein biosynthesis	46.7 (5.33)	47.8	33.1	10	S3 specific
AT1G54270.1_Eukaryotic initiation factor 4A-2_HC, <i>Quercus alba</i>				QA214_22	Protein biosynthesis	44.5 (6.15)	42.7	37.3	10	S3 specific
AT1G54270.1_Eukaryotic initiation factor 4A-2_HC, <i>Quercus petraea</i>				QP1475_18	Protein biosynthesis	36.7 (7.09)	30.7	31.2	7	S3 specific
Low molecular weight heat-shock protein, <i>Quercus</i> spp.				TC20725_11	Protein biosynthesis	18.2 (6.60)	7.8	28.5	3	S3 specific
Carbohydrate metabolism										

Protein description (close species)	Accession				Specific function	Mr(pI)	Scores	Cov (%)	Peptides Mached	Ratio S3/S0
	UniprotKB/SwissProt	UniprotKB/TrEml	NCBI	Quercus_DB						
Alcohol dehydrogenase class-3, <i>Quercus</i> spp.				TC21890_16	Glycolysis / Gluconeogenesis	31.1 (6.90)	23.0	12.0	3	S3 specific
Cinnamyl alcohol dehydrogenase, <i>Quercus</i> spp.				TC18401_20	Glycolysis / Gluconeogenesis	36.3 (8.35)	13.8	11.9	3	S3 specific
2-phospho-D-glycerate hydrolase, <i>Poncirus trifoliata</i>		D7NHW9	568856679		Glycolysis / Gluconeogenesis	47.8 (5.78)	147.9	27.8	8	S3 specific
Succinyl-CoA ligase [GDP-forming] beta-chain, mitochondrial precursor, <i>Oryza sativa</i> subsp. <i>japonica</i>				TC17986_22	TCA cycle	21.8 (5.01)	15.8	32.2	4	S3 specific
Alpha-1,4 glucan phosphorylase L isozyme, chloroplastic/amyloplastic-like, <i>Cucumis sativus</i>			449438839		Starch and sucrose metabolism	108.9 (5.77)	18.3	6.3	5	S3 specific
AT4G27270.1_Quinone reductase family protein_HC, <i>Quercus suber</i>				QS1415_39		22.1 (6.55)	17.6	31.7	4	S3 specific
Amino acids metabolism										
Aspartate aminotransferase P2, mitochondrial (Fragment), <i>Lupinus angustifolius</i>	P26563	P26563	112979		Cysteine, methionine, phenylalanine, tyrosine and tryptophan biosynthesis	49.9 (7.85)	8.4	7.3	3	S3 specific
Stress response										
AT5G12380.1_annexin 8_HC, <i>Quercus robur</i>				QR01593_18		36.1 (6.57)	12.0	13.0	4	S3 specific
AT2G22170.1_Lipase/lipoxygenase, PLAT/LH2 family protein_HC, <i>Quercus rubra</i>				QRU637_29		20.2 (4.98)	8.6	17.0	3	S3 specific
Acyl-coenzyme A oxidase, <i>Gossypium hirsutum</i>		E9L0E1	321438027		Fatty acid degradation, biosynthesis of unsaturated fatty acids	74.8 (7.42)	11.7	5.9	3	S3 specific
Protein transport										
AT1G09630.1_RAB GTPase 11C_HC, <i>Quercus petraea</i>				QP9969_18	Protein transport	24.4 (7.84)	9.9	20.0	4	S3 specific
Predicted protein(ATP-dependent transporter, putative) ^a , <i>Quercus</i> spp.				TC24103_33		58.4 (5.74)	10.4	8.7	3	S3 specific
Miscellaneous										
Putative uncharacterized protein(Transcription factor BTF3) ^a , <i>Vitis vinifera</i>		D7TXR6	296088008			17.2 (8.35)	17.2	25.6	3	S3 specific
Predicted protein (Esterase d, s-formylglutathione hydrolase) ^a , <i>Quercus</i> spp.				TC21959_22	Methane metabolism	33.9 (7.14)	9.1	14.6	3	S3 specific
Predicted protein(Endo-1,3;1,4-beta-D-glucanase-like) ^a , <i>Quercus</i> spp.				TC29292_21	Glycosidase	27.0 (7.15)	17.3	22.3	4	S3 specific
Unknown										

Protein description (close species)	Accession				Specific function	Mr(pI)	Scores	Cov (%)	Peptides Mached	Ratio S3/S0
	UniprotKB/SwissProt	UniprotKB/TrEmI	NCBI	Quercus_DB						
Endoplasmin homolog, <i>Cucumis sativus</i>			449444490			89.5 (5.66)	18.0	11.6	6	S3 specific
Predicted protein, <i>Quercus alba</i>				QA63945_8		9.3 (4.53)	10.1	43.9	3	S3 specific
Predicted protein, <i>Quercus alba</i>				QA11472_4		9.5 (5.06)	12.3	69.5	4	S3 specific
Predicted protein, <i>Physcomitrella patens</i> subsp. <i>patens</i>		A9TLR9	162669047			51.8 (5.78)	9.9	8.9	3	S3 specific
Predicted protein, <i>Physcomitrella patens</i> subsp. <i>Patens</i>		A9RKQ8	162694344			17.6 (10.98)	13.3	27.0	3	S3 specific
Predicted protein, <i>Physcomitrella patens</i> subsp. <i>patens</i>		A9S0A3	162689004			71.0 (5.33)	32.0	9.4	5	S3 specific
Predicted protein, <i>Physcomitrella patens</i> subsp. <i>patens</i>		A9TWR9	162665432			71.0 (5.25)	43.9	12.6	6	S3 specific
Predicted protein, <i>Physcomitrella patens</i> subsp. <i>patens</i>		A9U4N3	162662524			70.8 (5.21)	34.1	8.0	4	S3 specific
Predicted protein, <i>Physcomitrella patens</i> subsp. <i>patens</i>		A9TRK2	162667294			71.0 (5.36)	31.4	7.3	4	S3 specific
Putative uncharacterized protein, <i>Picea sitchensis</i>		C0PQU1	224286262			49.6 (5.10)	9.5	11.1	3	S3 specific
Putative uncharacterized protein, <i>Populus trichocarpa</i> x <i>Populus deltoides</i>		A9PIL8	566198751			21.9 (6.95)	10.5	23.4	3	S3 specific
Putative uncharacterized protein, <i>Selaginella moellendorffii</i>		D8R691	300165670			47.0 (5.81)	37.3	24.6	8	S3 specific
Putative uncharacterized protein, <i>Selaginella moellendorffii</i>		D8RBE2	300164169			71.5 (5.29)	66.7	16.2	8	S3 specific
Putative uncharacterized protein, <i>Selaginella moellendorffii</i>		D8RYR3	300155849			71.4 (5.33)	17.6	8.9	4	S3 specific
Putative uncharacterized protein Sb08g009580, <i>Sorghum bicolor</i>		C5YU58	241942772			74.4 (5.26)	17.2	9.1	5	S3 specific
Predicted protein, <i>Quercus</i> spp.				TC22417_25		41.7 (8.28)	12.0	13.0	4	S3 specific
Predicted protein, <i>Quercus</i> spp.				TC25340_15		27.4 (5.14)	18.1	24.6	4	S3 specific
Predicted protein, <i>Quercus</i> spp.				TC31608_14		20.0 (8.56)	9.3	24.1	3	S3 specific
Uncharacterized protein, <i>Amborella trichopoda</i>		W1P6P5	586685400			61.3 (5.87)	47.3	16.4	7	S3 specific
Uncharacterized protein, <i>Amborella trichopoda</i>		W1PYC8	586750760			48.0 (5.92)	170.0	33.5	11	S3 specific
Uncharacterized protein, <i>Capsella rubella</i>		R0HL55	565467720			20.5 (9.91)	15.9	24.2	3	S3 specific
Uncharacterized protein, <i>Citrus clementina</i>		V4TAP7	568836083			72.8 (5.35)	40.6	11.3	6	S3 specific

Protein description (close species)	Accession				Specific function	Mr(pI)	Scores	Cov (%)	Peptides Mached	Ratio S3/S0
	UniprotKB/SwissProt	UniprotKB/TrEml	NCBI	Quercus_DB						
Uncharacterized protein, <i>Citrus clementina</i>		V4RFD2	567852369			51.0 (6.43)	8.8	7.7	3	S3 specific
Uncharacterized protein, <i>Citrus clementina</i>		V4WJB1	567923216			60.8 (7.31)	29.3	15.6	7	S3 specific
Uncharacterized protein, <i>Citrus clementina</i>		V4SV61	568855345			61.0 (6.01)	27.9	9.0	4	S3 specific
Uncharacterized protein, <i>Citrus clementina</i>		V4S4J1	568854377			41.7 (5.49)	34.9	29.0	7	S3 specific
Uncharacterized protein, <i>Citrus clementina</i>		V4W3Z0	568830945			28.2 (10.86)	13.5	13.5	3	S3 specific
Uncharacterized protein, <i>Citrus clementina</i>		V4SA43	568866546			39.2 (8.29)	11.6	17.6	4	S3 specific
Uncharacterized protein, <i>Citrus clementina</i>		V4TQM7	567892811			118.3 (5.36)	8.0	3.4	3	S3 specific
Uncharacterized protein, <i>Citrus clementina</i>		V4TCG4	567886976			70.9 (5.21)	153.5	36.7	17	S3 specific
Uncharacterized protein, <i>Citrus clementina</i>		V4SNC1	567887042			11.7 (10.13)	9.3	30.1	3	S3 specific
Uncharacterized protein, <i>Lotus japonicus</i>		I3SYP8	388514657			61.9 (5.53)	34.7	17.4	8	S3 specific
Uncharacterized protein, <i>Lotus japonicus</i>		I3T2V1	388517563			36.4 (6.79)	10.5	13.8	3	S3 specific
Uncharacterized protein, <i>Phaseolus vulgaris</i>		V7CZ55	593799232			71.3 (5.22)	93.3	22.7	10	S3 specific
Uncharacterized protein, <i>Phaseolus vulgaris</i>		V7CGW4	593788678			17.6 (9.80)	16.1	27.1	4	S3 specific
Uncharacterized protein, <i>Phaseolus vulgaris</i>		V7B3U0	593331412			80.9 (7.12)	17.1	10.4	6	S3 specific
Uncharacterized protein, <i>Phaseolus vulgaris</i>		V7B0A5	593331628			71.1 (5.25)	99.2	23.7	12	S3 specific
Uncharacterized protein, <i>Prunus persica</i>		M5W737	595841933			64.4 (5.97)	66.3	29.3	12	S3 specific
Uncharacterized protein, <i>Prunus persica</i>		M5WXC0	595967112			59.6 (6.15)	18.6	12.4	6	S3 specific
Uncharacterized protein, <i>Prunus persica</i>		M5XEK3	596147928			46.6 (5.80)	50.6	36.8	11	S3 specific
Uncharacterized protein, <i>Prunus persica</i>		M5VPZ0	595804508			46.7 (5.58)	56.3	36.0	11	S3 specific
Up-accumulated in S3 (Quantitative differences)										
Oxidative stress										
AT1G65980.1_ thioredoxin-dependent peroxidase 1 _HC, <i>Quercus rubra</i>				QRU798_35	Oxidoreductase	16.6 (7.97)	34.6	33.3	4	2.34

Protein description (close species)	Accession				Specific function	Mr(pI)	Scores	Cov (%)	Peptides Mached	Ratio S3/S0
	UniprotKB/ SwissProt	UniprotKB /TrEml	NCBI	Quercus_DB						
Miscellaneous										
AT1G09620.1_ ATP binding;leucine-tRNA ligases;aminoacyl-tRNA ligases, <i>Quercus robur</i>				QR06584_49	Translation	75.1 (6.19)	14.5	21.4	5	2.26
AT3G14290.1_ 20S proteasome alpha subunit E2 _HC, <i>Quercus petraea</i>				QP547_45	Proteolysis	28.2 (4.88)	18.0	22.9	4	2.50
AT3G55440.1_ triosephosphate isomerase _HC, <i>Quercus rubra</i>				QRU489_20	Carbohydrate metabolism	27.3 (6.55)	39.1	32.6	6	2.81
AT1G11860.1_ Glycine cleavage T-protein family _HC, <i>Quercus petraea</i>				QP518_18	Amino acid metabolism	25.3 (9.11)	18.4	18.8	4	2.08
Unknown										
Uncharacterized protein, <i>Thellungiella salsuginea</i>		V4M1C4	567199488			46.7 (5.80)	27.7	25.0	6	4.21
Uncharacterized protein LOC101299909 isoform 2, <i>Fragaria vesca</i> subsp. <i>Vesca</i>			470106648			15.5 (9.20)	27.6	21.7	3	2.03
Putative uncharacterized protein, <i>Medicago truncatula</i>		B7FHH0	217071678			17.7 (6.20)	17.7	25.0	3	2.18

^a Blastp result homology of 75% and 10⁻¹⁰ evalue

6.6 References

1. He D, Yang P (2013) Proteomics of rice seed germination. *Frontiers Plant Sciences* 4
2. Derek Bewley J, Black M, Halmer P (2006) The encyclopedia of seeds: science, technology and uses Cabi Series. CABI, Wallingford, UK
3. Gallardo K, Job C, Groot SPC, Puype M, Demol H, Vandekerckhove J, Job D (2001) Proteomic analysis of *Arabidopsis* seed germination and priming. *Plant Physiology* 126: 835-848
4. Gallardo K, Job C, Groot SPC, Puype M, Demol H, Vandekerckhove J, Job D (2002) Proteomics of *Arabidopsis* seed germination. A comparative study of wild-type and gibberellin-deficient seeds. *Plant Physiology* 129: 823-837
5. Fu Q, Wang B-C, Jin X, Li H-B, Han P, Wei K-h, Zhan X-m, Zhu Y-X (2005) Proteomic analysis and extensive protein identification from dry, germinating *Arabidopsis* seeds and young seedlings. *Journal of Biochemistry and Molecular Biology* 38: 650-660
6. Boudet J, Buitink J, Hoekstra FA, Rogniaux H, Larré C, Sator P, Leprince O (2006) Comparative analysis of the heat stable proteome of radicles of *Medicago truncatula* seeds during germination identifies Late Embryogenesis Abundant proteins associated with desiccation tolerance. *Plant Physiology* 140: 1418-1436
7. Østergaard O, Finnie C, Laugesen S, Roepstorff P, Svensson B (2004) Proteome analysis of barley seeds: Identification of major proteins from two-dimensional gels (pI 4–7). *PROTEOMICS* 4: 2437-2447
8. Bønsager BC, Finnie C, Roepstorff P, Svensson B (2007) Spatio-temporal changes in germination and radical elongation of barley seeds tracked by proteome analysis of dissected embryo, aleurone layer, and endosperm tissues. *PROTEOMICS* 7: 4528-4540
9. Sheoran IS, Olson DJH, Ross ARS, Sawhney VK (2005) Proteome analysis of embryo and endosperm from germinating tomato seeds. *PROTEOMICS* 5: 3752-3764
10. Yang P, Li X, Wang X, Chen H, Chen F, Shen S (2007) Proteomic analysis of rice (*Oryza sativa*) seeds during germination. *PROTEOMICS* 7: 3358-3368
11. Sano N, Permana H, Kumada R, Shinozaki Y, Tanabata T, Yamada T, Hirasawa T, Kanekatsu M (2012) Proteomic analysis of embryonic proteins synthesized from long-lived mRNAs during germination of rice seeds. *Plant Cell Physiol* 53: 687-698
12. Huang H, Møller IM, Song S-Q (2012) Proteomics of desiccation tolerance during development and germination of maize embryos. *Journal of Proteomics* 75: 1247-1262
13. Mak Y, Willows RD, Roberts TH, Wrigley CW, Sharp PJ, Copeland L (2009) Germination of wheat: a functional proteomics analysis of the embryo. *Cereal Chemistry Journal* 86: 281-289
14. Han C, Yin X, He D, Yang P (2013) Analysis of proteome profile in germinating soybean seed, and its comparison with rice showing the styles of reserves mobilization in different crops. *PLoS One* 8: e56947
15. Wong P-F, Abubakar S (2005) Post-germination changes in *Hevea brasiliensis* seeds proteome. *Plant Science* 169: 303-311
16. Pawłowski TA (2007) Proteomics of European beech (*Fagus sylvatica* L.) seed dormancy breaking: Influence of abscisic and gibberellic acids. *PROTEOMICS* 7: 2246-2257
17. Lee C-S, Chien C-T, Lin C-H, Chiu Y-Y, Yang Y-S (2006) Protein changes between dormant and dormancy-broken seeds of *Prunus campanulata* Maxim. *PROTEOMICS* 6: 4147-4154
18. Yang M-F, Liu Y-J, Liu Y, Chen H, Chen F, Shen S-H (2009) Proteomic analysis of oil mobilization in seed germination and postgermination development of *Jatropha curcas*. *Journal of Proteome Research* 8: 1441-1451
19. Pawłowski T (2009) Proteome analysis of Norway maple (*Acer platanoides* L.) seeds dormancy breaking and germination: influence of abscisic and gibberellic acids. *BMC Plant Biology* 9: 48
20. Sghaier-Hammami B, Valledor L, Drira N, Jorriin-Novo JV (2009) Proteomic analysis of the development and germination of date palm (*Phoenix dactylifera* L.) zygotic embryos. *PROTEOMICS* 9: 2543-2554

21. Balbuena TS, Jo L, Pieruzzi FP, Dias LLC, Silveira V, Santa-Catarina C, Junqueira M, Thelen JJ, Shevchenko A, Floh EIS (2011) Differential proteome analysis of mature and germinated embryos of *Araucaria angustifolia*. *Phytochemistry* 72: 302-311
22. Holdsworth MJ, Bentsink L, Soppe WJ (2008) Molecular networks regulating Arabidopsis seed maturation, after-ripening, dormancy and germination. *New Phytol* 179: 33-54
23. Valero Galván J, Jorrín Novo J, Cabrera A, Ariza D, García-Olmo J, Cerrillo R (2012) Population variability based on the morphometry and chemical composition of the acorn in Holm oak (*Quercus ilex* subsp. *ballota* [Desf.] Samp.). *European Journal of Forest Research* 131: 893-904
24. Valero Galván J, Valledor L, Cerrillo RMN, Pelegrín EG, Jorrín-Novo JV (2011) Studies of variability in Holm oak (*Quercus ilex* subsp. *ballota* [Desf.] Samp.) through acorn protein profile analysis. *Journal of Proteomics* 74: 1244-1255
25. Valero Galván J, Valledor L, González Fernandez R, Navarro Cerrillo RM, Jorrín-Novo JV (2012) Proteomic analysis of Holm oak (*Quercus ilex* subsp. *ballota* [Desf.] Samp.) pollen. *Journal of Proteomics* 75: 2736-2744
26. Echevarría-Zomeño S, Ariza D, Jorge I, Lenz C, Del Campo A, Jorrín JV, Navarro RM (2009) Changes in the protein profile of *Quercus ilex* leaves in response to drought stress and recovery. *Journal of Plant Physiology* 166: 233-245
27. Jorge I, Navarro RM, Lenz C, Ariza D, Jorrín J (2006) Variation in the holm oak leaf proteome at different plant developmental stages, between provenances and in response to drought stress. *PROTEOMICS* 6: S207-S214
28. Jorge I, Navarro RM, Lenz C, Ariza D, Porras C, Jorrín J (2005) The holm oak leaf proteome: Analytical and biological variability in the protein expression level assessed by 2-DE and protein identification tandem mass spectrometry de novo sequencing and sequence similarity searching. *PROTEOMICS* 5: 222-234
29. Sghaier-Hammami B, Valero-Galván J, Romero-Rodríguez MC, Navarro-Cerrillo RM, Abdelly C, Jorrín-Novo J (2013) Physiological and proteomics analyses of Holm oak (*Quercus ilex* subsp. *ballota* [Desf.] Samp.) responses to *Phytophthora cinnamomi*. *Plant Physiology and Biochemistry* 71: 191-202
30. Valero-Galván J, González-Fernández R, Navarro-Cerrillo RM, Gil-Pelegrín E, Jorrín-Novo JV (2013) Physiological and proteomic analyses of drought stress response in holm oak provenances. *Journal of Proteome Research* 12: 5110-5123
31. Jorrin-Novo J (2014) Plant proteomics methods and protocols. In: Jorrin-Novo JV, Komatsu S, Weckwerth W, Wienkoop S (eds) *Plant Proteomics*. Humana Press, pp 3-13
32. Miernyk J (2014) Seed proteomics. In: Jorrin-Novo JV, Komatsu S, Weckwerth W, Wienkoop S (eds) *Plant Proteomics*. Humana Press, pp 361-377
33. Boschetti E, Righetti P (2014) Plant Proteomics methods to reach low-abundance proteins. In: Jorrin-Novo JV, Komatsu S, Weckwerth W, Wienkoop S (eds) *Plant Proteomics*. Humana Press, pp 111-129
34. Kim ST, Wang Y, Kang SY, Kim SG, Rakwal R, Kim YC, Kang KY (2009) Developing rice embryo proteomics reveals essential role for embryonic proteins in regulation of seed germination. *Journal of Proteome Research* 8: 3598-3605
35. Maldonado AM, Echevarría-Zomeño S, Jean-Baptiste S, Hernández M, Jorrín-Novo JV (2008) Evaluation of three different protocols of protein extraction for *Arabidopsis thaliana* leaf proteome analysis by two-dimensional electrophoresis. *Journal of Proteomics* 71: 461-472
36. Sghaier-Hammami B, Drira N, Jorrín-Novo JV (2009) Comparative 2-DE proteomic analysis of date palm (*Phoenix dactylifera* L.) somatic and zygotic embryos. *Journal of Proteomics* 73: 161-177
37. Wang W, Vignani R, Scali M, Cresti M (2006) A universal and rapid protocol for protein extraction from recalcitrant plant tissues for proteomic analysis. *ELECTROPHORESIS* 27: 2782-2786
38. Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry* 72: 248-254
39. Ramagli LS, Rodriguez LV (1985) Quantitation of microgram amounts of protein in two-dimensional polyacrylamide gel electrophoresis sample buffer. *ELECTROPHORESIS* 6: 559-563
40. Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227: 680-685
41. Görg A, Weiss W, Dunn MJ (2004) Current two-dimensional electrophoresis technology for proteomics. *PROTEOMICS* 4: 3665-3685

42. Mathesius U, Keijzers G, Natera SHA, Weinman JJ, Djordjevic MA, Rolfe BG (2001) Establishment of a root proteome reference map for the model legume *Medicago truncatula* using the expressed sequence tag database for peptide mass fingerprinting. *PROTEOMICS* 1: 1424-1440
43. Chich J-F, David O, Villers F, Schaeffer B, Lutomski D, Huet S (2007) Statistics for proteomics: Experimental design and 2-DE differential analysis. *Journal of Chromatography B* 849: 261-272
44. Valledor L, Castillejo MA, Lenz C, Rodríguez R, Cañal MJ, Jorrín Js (2008) Proteomic analysis of *Pinus radiata* needles: 2-DE Map and protein identification by LC/MS/MS and substitution-tolerant database searching. *Journal of Proteome Research* 7: 2616-2631
45. Shevchenko A, Wilm M, Vorm O, Mann M (1996) Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. *Analytical Chemistry* 68: 850-858
46. Valledor L, Weckwerth W (2014) An improved detergent-compatible gel-fractionation LC-LTQ-Orbitrap-MS workflow for plant and microbial proteomics. In: Jorrin-Novo JV, Komatsu S, Weckwerth W, Wienkoop S (eds) *Plant Proteomics*. Humana Press, pp 347-358
47. Romero-Rodríguez MC, Pascual J, Valledor L, Jorrin-Novo J (2014) Improving the quality of protein identification in non-model species. Characterization of *Quercus ilex* seed and *Pinus radiata* needle proteomes by using SEQUEST and custom databases. *Journal of Proteomics* 105: 85-91
48. Valledor L, Recuenco-Munoz L, Egelhofer V, Wienkoop S, Weckwerth W (2012) The different proteomes of *Chlamydomonas reinhardtii*. *Journal of Proteomics* 75: 5883-5887
49. 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
50. Valledor L, Jorrín J (2011) Back to the basics: Maximizing the information obtained by quantitative two dimensional gel electrophoresis analyses by an appropriate experimental design and statistical analyses. *Journal of Proteomics* 74: 1-18
51. Valledor L, Romero-Rodríguez MC, Jorrin-Novo J (2014) Standardization of data processing and statistical analysis in comparative plant proteomics experiment. In: Jorrin-Novo JV, Komatsu S, Weckwerth W, Wienkoop S (eds) *Plant Proteomics*. Humana Press, pp 51-60
52. Sharov AA, Dudekula DB, Ko MSH (2005) A web-based tool for principal component and significance analysis of microarray data. *Bioinformatics* 21: 2548-2549
53. Paoletti AC, Parmely TJ, Tomomori-Sato C, Sato S, Zhu D, Conaway RC, Conaway JW, Florens L, Washburn MP (2006) Quantitative proteomic analysis of distinct mammalian mediator complexes using normalized spectral abundance factors. *Proceedings of the National Academy of Sciences* 103: 18928-18933
54. Bonner FTV, John A. (1987) *Seed Biology and Technology of Quercus*. Gen. Tech. Rep. SO-66. New Orleans, LA: U.S. Dept of Agriculture, Forest Service, Southern Forest Experiment Station
55. Barclay AS, Earle FR (1974) Chemical analyses of seeds iii oil and protein content of 1253 species. *Economic Botany* 28: 178-236
56. Prinsi B, Negri A, Pesaresi P, Cocucci M, Espen L (2009) Evaluation of protein pattern changes in roots and leaves of *Zea mays* plants in response to nitrate availability by two-dimensional gel electrophoresis analysis. *BMC Plant Biology* 9: 113
57. Yeoh H-H, Wee Y-C (1994) Leaf protein contents and nitrogen-to-protein conversion factors for 90 plant species. *Food Chemistry* 49: 245-250
58. Dickie LB, Collins W, Young C (1984) Root protein quantity and quality in a seedling population of sweet potatoes. *HortScience* 19: 689-692
59. Diz AP, Truebano M, Skibinski DO (2009) The consequences of sample pooling in proteomics: an empirical study. *ELECTROPHORESIS* 30: 2967-2975
60. Plomion C, Lalanne C, Claverol S, Meddour H, Kohler A, Bogeat-Triboulot M-B, Barre A, Le Provost G, Dumazet H, Jacob D, Bastien C, Dreyer E, de Daruvar A, Guehl J-M, Schmitter J-M, Martin F, Bonneau M (2006) Mapping the proteome of poplar and application to the discovery of drought-stress responsive proteins. *PROTEOMICS* 6: 6509-6527
61. Valledor L, Jorrín JsV, Rodríguez JL, Lenz C, Meijón Mn, Rodríguez R, Cañal MJs (2010) Combined proteomic and transcriptomic analysis identifies differentially expressed pathways associated to *Pinus radiata* needle maturation. *Journal of Proteome Research* 9: 3954-3979

62. Wang W-Q, Møller IM, Song S-Q (2012) Proteomic analysis of embryonic axis of *Pisum sativum* seeds during germination and identification of proteins associated with loss of desiccation tolerance. *Journal of Proteomics* 77: 68-86
63. Argout X, Salse J, Aury J-M, Guiltinan MJ, Droc G, Gouzy J, Allegre M, Chaparro C, Legavre T, Maximova SN, Abrouk M, Murat F, Fouet O, Poulain J, Ruiz M, Roguet Y, Rodier-Goud M, Barbosa-Neto JF, Sabot F, Kudrna D, Ammiraju JSS, Schuster SC, Carlson JE, Sallet E, Schiex T, Dievart A, Kramer M, Gelley L, Shi Z, Berard A, Viot C, Boccara M, Risterucci AM, Guignon V, Sabau X, Axtell MJ, Ma Z, Zhang Y, Brown S, Bourge M, Golser W, Song X, Clement D, Rivallan R, Tahiri M, Akaza JM, Pitollat B, Gramacho K, D'Hont A, Brunel D, Infante D, Kebe I, Costet P, Wing R, McCombie WR, Guiderdoni E, Quetier F, Panaud O, Wincker P, Bocs S, Lanaud C (2011) The genome of *Theobroma cacao*. *Nature Genetics* 43: 101-108
64. Myburg AA, Grattapaglia D, Tuskan GA, Hellsten U, Hayes RD, Grimwood J, Jenkins J, Lindquist E, Tice H, Bauer D, Goodstein DM, Dubchak I, Poliakov A, Mizrahi E, Kullam ARK, Hussey SG, Pinard D, van der Merwe K, Singh P, van Jaarsveld I, Silva-Junior OB, Togawa RC, Pappas MR, Faria DA, Sansaloni CP, Petroli CD, Yang X, Ranjan P, Tschaplinski TJ, Ye C-Y, Li T, Sterck L, Vanneste K, Murat F, Soler M, Clemente HS, Saidi N, Cassan-Wang H, Dunand C, Hefer CA, Bornberg-Bauer E, Kersting AR, Vining K, Amarasinghe V, Ranik M, Naithani S, Elser J, Boyd AE, Liston A, Spatafora JW, Dharmwardhana P, Raja R, Sullivan C, Romanel E, Alves-Ferreira M, Kulheim C, Foley W, Carocha V, Paiva J, Kudrna D, Brommonschenkel SH, Pasquali G, Byrne M, Rigault P, Tibbits J, Spokevicius A, Jones RC, Steane DA, Vaillancourt RE, Potts BM, Joubert F, Barry K, Pappas GJ, Strauss SH, Jaiswal P, Grima-Pettenati J, Salse J, Van de Peer Y, Rokhsar DS, Schmutz J (2014) The genome of *Eucalyptus grandis*. *Nature* 510: 356-362
65. Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhale Rao RR, Bhale Rao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen GL, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J, Degroove S, Dejardin A, Depamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehling J, Ellis B, Gendler K, Goodstein D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R, Huang W, Islam-Faridi N, Jones S, Jones-Rhoades M, Jorgensen R, Joshi C, Kangasjarvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A, Kalluri U, Larimer F, Leebens-Mack J, Leple JC, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson DR, Nelson C, Nieminen K, Nilsson O, Pereda V, Peter G, Philippe R, Pilate G, Poliakov A, Razumovskaya J, Richardson P, Rinaldi C, Ritland K, Rouze P, Ryaboy D, Schmutz J, Schrader J, Segerman B, Shin H, Siddiqui A, Sterky F, Terry A, Tsai CJ, Uberbacher E, Unneberg P, Vahala J, Wall K, Wessler S, Yang G, Yin T, Douglas C, Marra M, Sandberg G, Van de Peer Y, Rokhsar D (2006) The genome of black cottonwood, *Populus trichocarpa*. *Science* 313: 1596-1604
66. Bewley JD (1997) Seed Germination and Dormancy. *The Plant Cell Online* 9: 1055-1066
67. Swigonska S, Weidner S (2013) Proteomic analysis of response to long-term continuous stress in roots of germinating soybean seeds. *Journal of Plant Physiology* 170: 470-479
68. Eckardt NA (2005) Peroxisomal Citrate Synthase provides exit route from fatty acid metabolism in oilseeds. *The Plant Cell Online* 17: 1863-1865
69. Sturm A, Tang G-Q (1999) The sucrose-cleaving enzymes of plants are crucial for development, growth and carbon partitioning. *Trends in Plant Science* 4: 401-407
70. Schaller A (2004) A cut above the rest: the regulatory function of plant proteases. *Planta* 220: 183-197
71. Tan-Wilson AL, Wilson KA (2012) Mobilization of seed protein reserves. *Physiologia Plantarum* 145: 140-153
72. Müntz K, Belozersky MA, Dunaevsky YE, Schlereth A, Tiedemann J (2001) Stored proteinases and the initiation of storage protein mobilization in seeds during germination and seedling growth. *Journal of Experimental Botany* 52: 1741-1752
73. Larocca M, Rossano R, Riccio P (2010) Analysis of green kiwi fruit (*Actinidia deliciosa* cv. Hayward) proteinases by two-dimensional zymography and direct identification of zymographic spots by mass spectrometry. *Journal of the Science of Food and Agriculture* 90: 2411-2418
74. Chevalier F (2010) Highlights on the capacities of "Gel-based" proteomics. *Proteome Science* 8: 23
75. Tan L, Chen S, Wang T, Dai S (2013) Proteomic insights into seed germination in response to environmental factors. *PROTEOMICS* 13: 1850-1870

76. Dogra V, Ahuja PS, Sreenivasulu Y (2013) Change in protein content during seed germination of a high altitude plant *Podophyllum hexandrum* Royle. *Journal of Proteomics* 78: 26-38
77. Morohashi Y (2002) Peroxidase activity develops in the micropylar endosperm of tomato seeds prior to radicle protrusion. *Journal of Experimental Botany* 53: 1643-1650
78. Wang KL-C, Li H, Ecker JR (2002) Ethylene biosynthesis and signaling networks. *The Plant Cell Online* 14: S131-S151
79. Tommasi F, Paciolla C, de Pinto MC, Gara LD (2001) A comparative study of glutathione and ascorbate metabolism during germination of *Pinus pinea* L. seeds. *Journal of Experimental Botany* 52: 1647-1654
80. Gallardo K, Job C, Groot SPC, Puype M, Demol H, Vandekerckhove J, Job D (2002) Importance of methionine biosynthesis for Arabidopsis seed germination and seedling growth. *Physiologia Plantarum* 116: 238-247
81. Jones JD, Henstrand JM, Handa AK, Herrmann KM, Weller SC (1995) Impaired wound induction of 3-deoxy-D-arabino-heptulosonate-7-phosphate (DAHP) synthase and altered stem development in transgenic potato plants expressing a DAHP synthase antisense construct. *Plant Physiology* 108: 1413-1421
82. Kleindt CK, Stracke R, Mehrtens F, Weisshaar B (2010) Expression analysis of flavonoid biosynthesis genes during *Arabidopsis thaliana* silique and seed development with a primary focus on the proanthocyanidin biosynthetic pathway. *BMC Research Notes* 3: 255
83. Mishra RC, Grover A (2014) Intergenic sequence between Arabidopsis caseinolytic protease B-cytoplasmic/heat shock protein100 and choline kinase genes functions as a heat-inducible bidirectional promoter. *Plant Physiology* 166: 1646-1658
84. Vale E, Heringer A, Barroso T, Ferreira A, da Costa M, Perales JE, Santa-Catarina C, Silveira V (2014) Comparative proteomic analysis of somatic embryo maturation in *Carica papaya* L. *Proteome Science* 12: 37
85. Zhang Q, Marsolais F (2014) Identification and characterization of omega-amidase as an enzyme metabolically linked to asparagine transamination in Arabidopsis. *Phytochemistry* 99: 36-43
86. Badowiec A, Weidner S (2014) Proteomic changes in the roots of germinating *Phaseolus vulgaris* seeds in response to chilling stress and post-stress recovery. *Journal of Plant Physiology* 171: 389-398
87. Job C, Rajjou L, Lovigny Y, Belghazi M, Job D (2005) Patterns of protein oxidation in Arabidopsis seeds and during germination. *Plant Physiology* 138: 790-802
88. Gonzalez-Fernandez R, Aloria K, Arizmendi JM, Jorrin-Novo JV (2013) Application of label-free shotgun nUPLC-MSE and 2-DE approaches in the study of *Botrytis cinerea* Mycelium. *Journal of Proteome Research* 12: 3042-3056

CHAPTER 7:

PHOSPHOPROTEOMIC ANALYSIS IN *Q. ILEX* GERMINATED SEEDS AND SEEDLINGS

Abstract

The changes in the phosphoproteome profile during the germination and development of *Q. ilex* seedlings were analyzed by multiplex-staining of high-resolution 2-DE gels. SYPRO-Ruby staining was used to reveal total protein spots and Pro-Q DPS for identifying phosphoprotein spots. By this double staining of gels it is possible the quantification of changes in phosphoprotein abundance and the detection of changes in phosphorylation status that are independent of total protein abundance. A total of 222 consistent proteins and phosphorylated in gel resolved spots were subjected to quantitative analysis. Among them, 55 phosphoprotein spots were differentially accumulated and 20 putative phosphoproteins were identified by MALDI TOF-TOF analysis. They were representative of five functional categories: carbohydrate and amino acid metabolism, defense, protein folding and oxidation-reduction processes. Among 20 putative phosphoproteins identified, seven showed changes in phosphorylation status. The most important proteins in this group were proteins related to glycolysis and biosynthesis of secondary metabolites from methionine. Phosphoproteome analysis detected an increase in these pathways from mature seeds to germinated seeds and seedling. In contrast, the total proteome analysis did not detect changes in abundance of these proteins between mature seeds to germinated seeds.

4.1. Introduction

Protein phosphorylation is one of the first and best-studied post-translational modifications (PTMs). These modifications increase the functional diversity of the proteome by the covalent addition of functional groups or proteins (phosphorylation, glycosylation, ubiquitination, nitrosylation, methylation, acetylation, lipidation), proteolytic cleavage of regulatory subunits or degradation of entire proteins and influence almost all aspects of normal cell biology and pathogenesis ¹. In the Ch. 1 (General Introduction), the biological relevance and methodology used to study phosphoproteins were briefly described.

Phosphorylation is a ubiquitous and reversible PTM, and it is an essential key regulator of intra-cellular biological processes that influence both the folding (conformation) and function of proteins ²⁻⁴. Phosphorylation events modulate a wide range of biological pathways in plants and other organisms ⁵. In eukaryotes, protein phosphorylation occurs predominantly on serine (Ser) and threonine (Thr) residues, whereas phosphorylation on tyrosine (Tyr) residues is less abundant ⁶. The characterisation of the phosphoproteome includes the detection and identification of phosphoproteins and phosphopeptides, localisation of the exact phosphorylation sites and the quantitation of phosphorylation status, which can be performed by gel based and gel free approaches. There are several strategies for detecting phosphoproteins (p. 27-30). Radioactive labelling of proteins with ³²P isotope, the first method developed, immunochemical techniques by using phosphospecific antibodies to phosphotyrosine (P-Tyr), phosphothreonine (P-Thr) and phosphoserine (P-Ser) ⁷⁻⁸, and staining protocols to reveal phosphoproteins such as Pro-Q Diamond phosphoprotein gel staining (Pro-Q DPS) ⁹, are all well known strategies for the detection of phosphoproteins.

Mass spectrometry (MS) can be then used for the identification of phosphoprotein (top-down) or phosphopeptides (bottom-up). However, phosphopeptides are notoriously difficult to analyse by MS, especially in the presence of the non-phosphorylated counterpart. The main challenges in analysing protein phosphorylation by MS, compared with proteomic analysis, is the low stoichiometry of phosphorylated proteins arising from the fact that only a small fraction of the protein will exist in a particular phosphorylated form, the lability of P-Ser and P-Thr residues when subjected to collision-induced dissociation (CID) and the phosphorylation being a reversible modification ¹⁰⁻¹¹. MS can hence become biased toward high abundant sample components, if a previous enrichment of phosphopeptides is not incorporated to the methodology ¹¹⁻¹².

Lower ionisation efficiency of phosphopeptides resulting in lower signal intensities in the presence of non-phosphorylated peptide ions is also another obstacle difficulting the detection of phosphorylation sites ¹³. Although phosphopeptides can be identified by LC-MS/MS

analysis, the identified phosphorylation site is sometimes ambiguous. This constitutes a problem for phosphopeptide identification by MS because the O-phosphate bond in Ser- and Thr-phosphorylated peptides is labile during ionization, resulting in a neutral loss of phosphoric acid from the precursor ion. In brief, while some methods are more efficient for phosphopeptide identification, other are superior for localisation of phosphorylation sites¹⁴. It is important to mention that “*in vitro*” analysis is greatly contributing to the study of protein phosphorylation and phosphorylation sites. This approach permits to identify consensus sequences containing phosphorylation sites for a specific protein kinase¹⁵, as those listed in¹⁶

Phosphorylation is involved in the regulation seed germination and seedling development¹⁷. Several reports indicate that the effect of ABA in seed germination and other process is mediated by a signaling cascade transduction pathway in which phosphorylation of proteins has a principal role (Fig. 1.3 p. 16)¹⁸⁻²⁰. Quantitative and qualitative profiling of phosphoproteins expressed during seed germination and seedling development have been performed using different proteomic approaches (gel based and gel-free) in different plant species such as *Arabidopsis thaliana*²¹⁻²³, *Medicago truncatula*²³⁻²⁴, *Phaseolus vulgaris*²⁵, *Zea mays*²⁶ and *Oryza sativa*^{17;27}. It is important to highlight that to the best of our knowledge, all previously investigated species produced orthodox seed and no report on phosphoproteomic analysis of non-orthodox or recalcitrant seeds has been published.

The *Q. ilex* seed germination and seedling development have been analysed by transcriptomics (Ch. 4 and 5) and proteomics (Ch. 6); in addition the sugars and phytohormones contents were measured (Ch. 4). In an attempt to complement these results, the dynamic protein phosphorylation changes during seed germination and seedling development were analysed using multiplex-staining of high-resolution 2-DE gels for total protein (SYPRO-Ruby) and phosphoproteins (Pro-Q DPS), which allow to quantify changes in overall abundant protein phosphorylation status in analysed processes. Herein, a phosphoproteome map of embryo axes of *Q. ilex* during germination was established and putative phosphorylated proteins associated to the regulation of germination and seedling development were identified.

4.2. Materials & Methods

4.2.1 General considerations

Embryo axis and seedlings were analysed at the following stages: S0, S3 and S7 (Fig. 3.3, p. 47). The plant material, sampling, total protein extraction, protein quantification, fractionation by 2-DE electrophoresis and protein identification by using MALDI TOF/TOF

were performed as described in materials and methods of Ch. 6 (p. 127). Specific methodologies for this chapter are given below.

4.2.2. Detection of phosphoproteins and total proteins and image analysis of 2-DE gels

All staining and washing steps were performed with continuous and gentle agitation at room temperature on an orbital shaker at 50 rpm. The workflow is showed in Fig. 7.1. For phosphoprotein detection, 2-DE gels were stained as reported by ²⁸. Gels were fixed with a solution containing 50% methanol and 10% acetic acid (2 x 30 min), and were then washed twice for 10 minutes with deionised water. Gels were subsequently stained in the dark with Pro-Q DPS solution (Molecular Probes, Invitrogen) for 2 hours and then destained with a solution containing 50mM sodium acetate (pH=4) and 20% acetonitrile (4 x 30 min) to remove gel-bound nonspecific staining. Images were acquired with Molecular Imager FX (Bio-Rad Laboratories, Inc.) at excitation/emission wavelengths of 532/555 nm. Gels were then washed with deionised water (2 x 10 min) and stained with SYPRO-Ruby (Molecular Probes, Invitrogen) as described ²⁹. The gels were destained with a solution containing 10% methanol, 7% acetic acid (2 x 30min) and were finally washed with deionised water prior to image capture with the Molecular Imager FX (Bio-Rad Laboratories) at excitation/emission wavelengths of 478/640 nm.

Gel image (Pro-Q DPS and SYPRO-Ruby) analysis was performed with PDQuest 8.0.1 software (Bio-Rad) as described in Ch. 6 (p. 128). To eliminate false positive phosphoprotein spots, a normalisation of detected spot in Pro-Q DPS contrasted against positive and negative phosphoprotein markers was performed as described ³⁰(Fig 7.1). A molecular weight standards marker (Bio-Rad) containing one phosphorylated protein (ovalbumin, 45.0 kDa) and nine unphosphorylated proteins was used as a positive (PM) and nine negative phosphoprotein markers (NM), respectively. A comparison was performed between the ratio of consistent spot volumes detected by Pro-Q DPS and SYPRO-Ruby in the same gels with the ratio of positive and negative phosphoprotein markers.

$$(NMPQ/NMSR)_i < (SvPQ_x/SvSR_x)_i \geq (PMPQ/PMSR)_i$$

In this equation, SvPQ_x, NMPQ and PMPQ are the normalised spot x volume, and the volumes of negative and positive phosphoprotein markers, detected in gel *i* by Pro-Q DPS staining respectively; and SvSR_x, NMSR, and PMSR are the normalised spot x volume, and the volumes of negative and positive phosphoprotein markers detected in gel *i* by SYPRO-Ruby staining, respectively. Only spots with a ratio SvPQ/SvSR similar or higher than the ratio of

PMPQ/PMSR and higher than NMPQ/NMSR were considered as phosphoprotein spots in the statistical analysis.

Phosphoprotein and total protein normalised spot volumes (individual spot intensity/normalisation factor, detailed in Ch. 6 (p. 131), calculated for each gel based on total quantity in valid spots, were determined and used for statistical analysis of protein expression levels. The web-based software NIA array analysis tool was utilized ³¹, to analyse phosphoprotein and total protein abundance values ³². The software was available online at <http://lgsun.grc.nia.nih.gov/anova/index.html>.

4.2.3. MALDI-TOF/TOF analysis

The phosphoprotein spots which showed statistically significant intensity differences were included in the analysis performed in Ch. 6 (p. 129). The MS results obtained in that chapter were used to perform a homology based search using the Mascot algorithm, including the phosphorylation of Ser, Thr and Tyr residues as dynamic modification.

4.2.4. BLAST alignment of putative phosphorylated proteins

Identified phosphoprotein sequences downloaded from UniprotKB, NCBI nr or available in Quercus_DB ³³ were subjected to BLAST analysis by using the phosphoprotein BLAST tool in the Plant Protein Phosphorylation DataBase (P3DB) ³⁴ available at <http://www.p3db.org/>, to find orthologous proteins whose phosphorylation sites were described previously in other species.

4.2.5. Functional classification of proteins

Proteins identified by MALDI TOF/TOF analysis were extracted and classified based on their putative function according to Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway, using Blas2GO ³⁵ based on BLASTp results against NCBI nr protein database (e-value<10⁻³), or according to annotations in UniProtKB protein database.

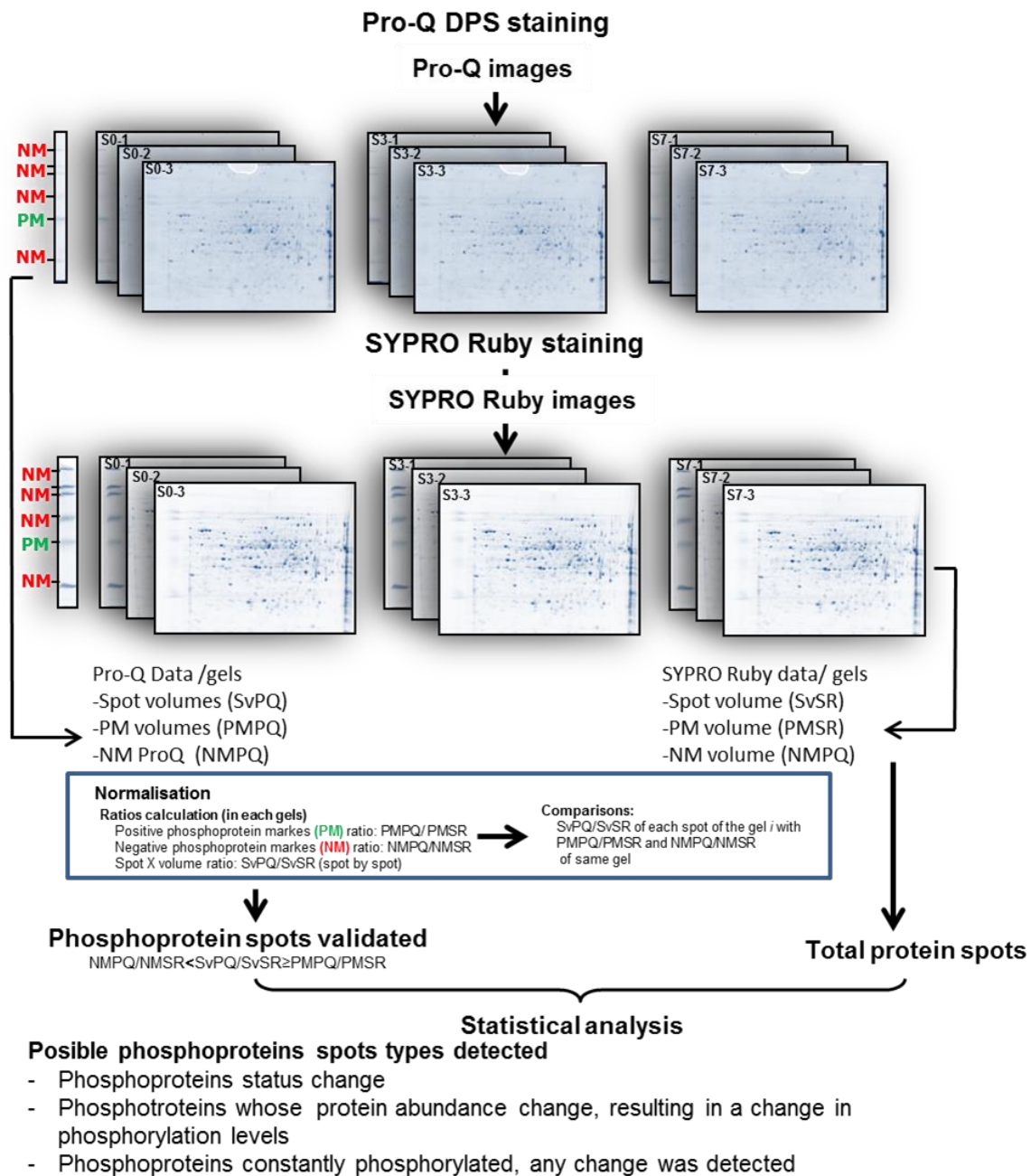


Figure 7.1: Phosphoproteome analysis workflow by multiplex-staining of high-resolution 2-DE gels. Pro-Q DPS was used to reveal phosphoproteins and SYPRO-Ruby for total proteins.

4.3. Results and discussion

The phosphoproteome changes which occur during holm oak seeds germination and seedling growth were analysed in embryo axes of mature (S0, unimbibed seeds) and germinated seeds (S3, radicle emergence) and in the whole seedling (S7). Similar analyses have been described in some model species including *Arabidopsis*, rice, soybean, rapeseed and maize [17; 26; 36]. By using the workflow described in (Fig. 7.1), the phosphoproteins were revealed in 2-DE gels by using Pro-Q DPS. Total proteins were detected in the same gel by using SYPRO-Ruby staining. The 2-DE gel images were analysed with PDQuest 8.0 software. Only consistent spots were considered for further statistical analysis.

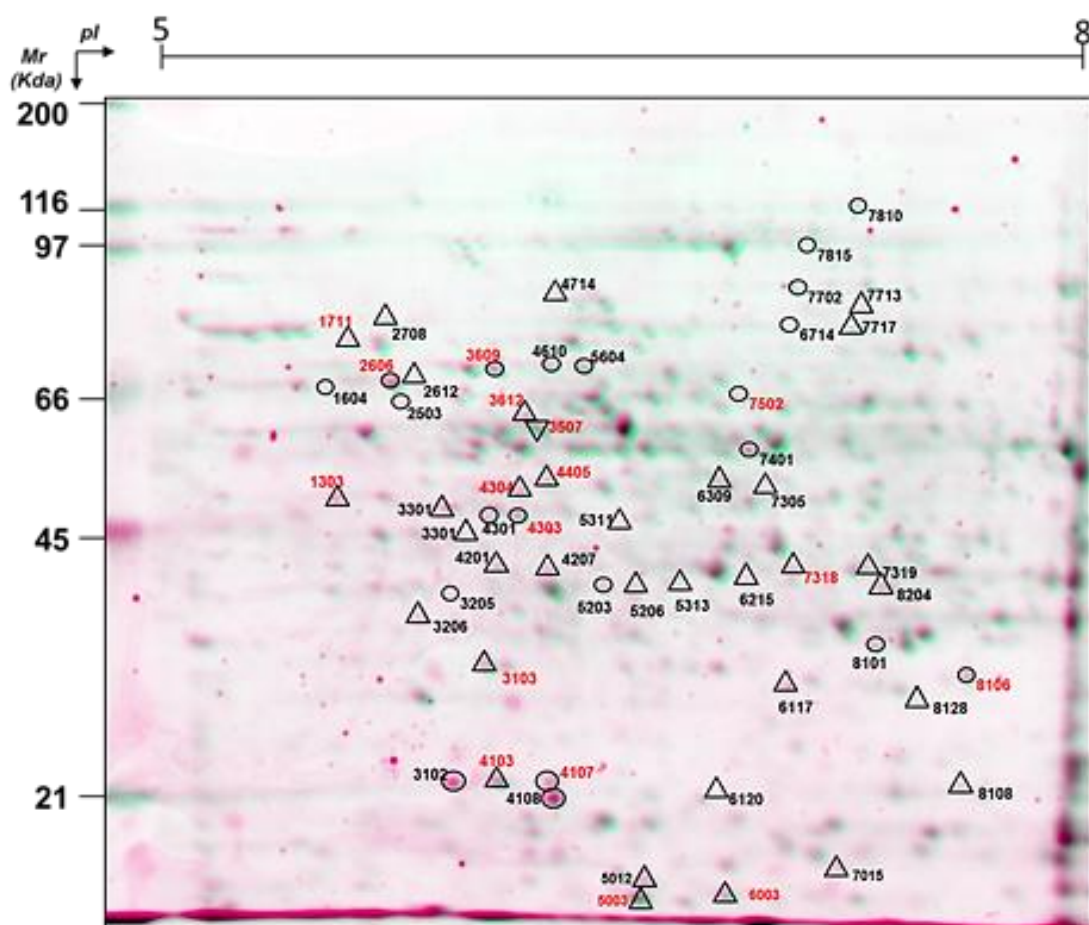


Figure 7.2: A virtual 2-DE gel showing the protein profile of *Q. ilex* mature seed embryo axis (S0 stage) obtained by successive Pro-Q DPS and SYPRO-Ruby staining. Proteins stained with SYPRO-Ruby appear in green, while Pro-Q DPS stained proteins appear in red. The statistically significant differential phosphoprotein spots were indicated with circles for quantitative differences and with triangles for qualitative (absence/presence) differences. Numbers in red indicate the protein spots that were identified by MALDI TOF/TOF.

Protein and phosphoprotein spots detected in protein extracts of dry mature seeds (S0), germinated seeds (S3) and seedlings (S7) are shown in Table 7.1. A total of 482 protein

spots were resolved by SYPRO-Ruby staining, with a distribution along the different stages similar to that described in Ch. 6. As the phosphorylation staining is unspecific, possible false positives were eliminated by normalizing the phosphoprotein spot volumes against the positive phosphoprotein marker (PM) (Fig. 7.1). The normalisation approach allowed the detection of 222 phosphoprotein spots by Pro-Q DPS. The phosphorylated protein spots were distributed evenly within the isoelectric point (pI) range 5–8 and Mr range 6–116 kDa. The percentage of phosphorylated proteins resolved in our experimental conditions was 46%, similar to that reported in other species. In example, around 500 spots were detected in germinating rice seeds¹⁷ and around 300 spots in three-week-old chickpea seedlings³⁷.

Table 7.1: Electrophoretic analysis of changes in the protein and phosphoprotein profile during germination and seedling growth. Number of spots detected by SYPRO-Ruby and Pro-Q DPS in different analyzed stages and number of differential spots in total protein and phosphoprotein are shown.

Stages	Spots detected by SYPRO-Ruby stain	Spots detected by Pro-Q DPS	Total number spot analyzed		Spots with change in total protein profile (SYPRO-Ruby)		Spots with change in phosphoprotein profile			
			SYPRO Ruby	Pro-Q DPS	Up	Down	With changes phosphorylation status*		Without changes in phosphorylation status*	
							Up	Down	Up	Down
S0	402	205			6	5	4	5	6	1
S3	412	211	482	222	-	-	-	-	-	-
S7	329	174			45	26	4	12	15	20

Up and down accumulated proteins were calculated respect to the S3 stage

*Phosphorylation status was considered changed when no difference was observed in SYPRO-Ruby staining but was statistically different in Pro-Q DPS. In contrast, it was considered unchanged when a difference was observed in both staining methods.

Consistent phosphoprotein spots were subjected to quantitative analysis, and 55 protein spots with statistically significant differences in Pro-Q staining intensity during germination and seedling development were selected (Table 7.2).

Table 7.2: Normalised relative volumes of phosphoproteins (Pro-Q DPS) and total protein (SYPRO-Ruby) spots differentially expressed in 2-DE gels. Mean \pm SD of three biological replicates. nd.: not detected spot (qualitative differences).

Spot Number	Pro-Q DPS				SYPRO-Ruby			
	S0	S3	S7	FDR	S0	S3	S7	FDR
Phosphoproteins with changes in phosphorylation status								
Quantitative differences								
2606	1.09 \pm 0.24	1.13 \pm 0.16	0.34 \pm 0.04	0.010	0.63 \pm 0.19	0.62 \pm 0.23	0.35 \pm 0.05	0.144
7502	0.10 \pm 0.06	0.24 \pm 0.02	0.35 \pm 0.14	0.010	0.17 \pm 0.04	0.19 \pm 0.02	0.30 \pm 0.09	0.254
7815	0.08 \pm 0.04	0.18 \pm 0.13	0.36 \pm 0.10	0.024	0.11 \pm 0.00	0.17 \pm 0.24	0.37 \pm 0.10	0.271
4107	0.75 \pm 0.14	0.63 \pm 0.16	0.25 \pm 0.10	0.027	0.32 \pm 0.03	0.30 \pm 0.09	0.18 \pm 0.01	0.288
1604	nd.	0.13 \pm 0.05	0.02 \pm 0.03	0.028	0.07 \pm 0.03	0.12 \pm 0.05	0.05 \pm 0.01	0.085
2503	0.10 \pm 0.03	0.12 \pm 0.09	0.05 \pm 0.01	0.028	0.06 \pm 0.03	0.09 \pm 0.04	0.06 \pm 0.01	0.525
6714	0.09 \pm 0.03	0.08 \pm 0.06	0.27 \pm 0.04	0.028	0.06 \pm 0.06	0.07 \pm 0.04	0.11 \pm 0.03	0.524
8101	0.21 \pm 0.08	0.15 \pm 0.01	0.46 \pm 0.02	0.028	0.24 \pm 0.07	0.13 \pm 0.07	0.18 \pm 0.10	0.358
Qualitative differences								
5311	0.09 \pm 0.02	nd.	nd.	-	nd.	nd.	nd.	-
6309	0.12 \pm 0.06	nd.	nd.	-	0.07 \pm 0.01	0.04 \pm 0.02	nd.	-
3612	0.39 \pm 0.13	nd.	0.16 \pm 0.05	-	0.35 \pm 0.06	0.37 \pm 0.22	0.23 \pm 0.11	-
7318	0.29 \pm 0.07	nd.	0.54 \pm 0.10	-	0.21 \pm 0.06	0.34 \pm 0.26	nd.	-
1711	0.14 \pm 0.12	0.28 \pm 0.20	nd.	-	0.21 \pm 0.01	0.26 \pm 0.06	0.15 \pm 0.01	-
4201	0.20 \pm 0.06	0.14 \pm 0.09	nd.	-	0.10 \pm 0.00	0.15 \pm 0.13	0.16 \pm 0.07	-
4304	0.27 \pm 0.03	0.40 \pm 0.18	nd.	-	0.20 \pm 0.06	0.27 \pm 0.04	nd.	-
4714	0.15 \pm 0.05	0.20 \pm 0.06	nd.	-	0.14 \pm 0.05	0.13 \pm 0.02	0.16 \pm 0.06	-
5012	0.31 \pm 0.06	0.22 \pm 0.05	nd.	-	0.23 \pm 0.08	0.24 \pm 0.07	0.16 \pm 0.04	-
6117	0.11 \pm 0.06	0.09 \pm 0.03	nd.	-	0.13 \pm 0.03	0.12 \pm 0.04	0.08 \pm 0.02	-
7713	0.08 \pm 0.04	0.11 \pm 0.03	nd.	-	0.09 \pm 0.01	0.09 \pm 0.01	0.04 \pm 0.01	-
7717	0.09 \pm 0.08	0.07 \pm 0.05	nd.	-	0.13 \pm 0.04	0.05 \pm 0.03	0.02 \pm 0.02	-
Phosphoproteins without changes in phosphorylation status								
Quantitative differences								
3102	1.55 \pm 0.55	0.76 \pm 0.09	0.13 \pm 0.04	0.000	0.22 \pm 0.03	0.28 \pm 0.10	0.07 \pm 0.02	0.000
3205	0.06 \pm 0.01	0.00 \pm 0.00	0.21 \pm 0.08	0.000	0.03 \pm 0.00	0.05 \pm 0.02	0.12 \pm 0.03	0.003
7702	0.17 \pm 0.05	0.23 \pm 0.03	0.86 \pm 0.10	0.000	0.20 \pm 0.04	0.18 \pm 0.03	0.85 \pm 0.12	0.000
4610	0.41 \pm 0.10	0.49 \pm 0.14	0.12 \pm 0.08	0.000	0.36 \pm 0.08	0.31 \pm 0.05	0.07 \pm 0.01	0.000
4301	0.14 \pm 0.05	0.20 \pm 0.02	0.66 \pm 0.17	0.001	0.15 \pm 0.01	0.18 \pm 0.02	0.56 \pm 0.09	0.000
8106	0.43 \pm 0.02	0.37 \pm 0.06	1.57 \pm 0.18	0.001	0.32 \pm 0.01	0.26 \pm 0.06	1.00 \pm 0.19	0.000
4108	1.24 \pm 0.12	0.50 \pm 0.11	0.38 \pm 0.29	0.010	0.38 \pm 0.04	0.25 \pm 0.05	0.13 \pm 0.05	0.004
5203	0.22 \pm 0.11	0.21 \pm 0.03	0.71 \pm 0.38	0.010	0.19 \pm 0.03	0.20 \pm 0.05	0.54 \pm 0.04	0.000
4303	0.44 \pm 0.01	0.51 \pm 0.17	1.38 \pm 0.23	0.024	0.32 \pm 0.07	0.32 \pm 0.01	1.05 \pm 0.05	0.000
7810	0.21 \pm 0.08	0.00 \pm 0.00	0.53 \pm 0.11	0.024	0.22 \pm 0.04	0.22 \pm 0.04	0.49 \pm 0.18	0.036
3609	0.73 \pm 0.28	0.77 \pm 0.16	0.27 \pm 0.05	0.028	0.59 \pm 0.16	0.59 \pm 0.06	0.26 \pm 0.07	0.011
5604	0.34 \pm 0.04	0.51 \pm 0.20	0.16 \pm 0.03	0.028	0.34 \pm 0.06	0.29 \pm 0.06	0.11 \pm 0.02	0.001
7401	0.45 \pm 0.15	0.42 \pm 0.10	1.20 \pm 0.26	0.028	0.22 \pm 0.03	0.26 \pm 0.07	0.52 \pm 0.07	0.025
Qualitative differences								
3206	nd.	nd.	0.52 \pm 0.03	-	nd.	0.03 \pm 0.06	0.35 \pm 0.08	-
5206	nd.	nd.	0.49 \pm 0.06	-	0.08 \pm 0.07	0.10 \pm 0.03	0.49 \pm 0.09	0.000
5313	nd.	nd.	0.25 \pm 0.08	-	0.03 \pm 0.05	0.01 \pm 0.02	0.22 \pm 0.08	0.000
6120	nd.	nd.	0.35 \pm 0.17	-	nd.	0.04 \pm 0.05	0.18 \pm 0.03	0.015
7305	nd.	nd.	0.70 \pm 0.12	-	0.23 \pm 0.05	0.22 \pm 0.05	0.53 \pm 0.06	0.007
1303	0.58 \pm 0.24	0.63 \pm 0.05	nd.	-	0.13 \pm 0.06	0.17 \pm 0.03	0.07 \pm 0.01	0.040
3103	0.62 \pm 0.07	0.44 \pm 0.30	nd.	-	0.57 \pm 0.04	0.40 \pm 0.27	0.01 \pm 0.02	0.000
3507	0.42 \pm 0.13	0.65 \pm 0.28	nd.	-	0.58 \pm 0.06	0.55 \pm 0.10	0.17 \pm 0.03	0.000
4103	0.76 \pm 0.04	0.38 \pm 0.10	nd.	-	0.47 \pm 0.04	0.31 \pm 0.08	0.03 \pm 0.03	0.000
4405	0.18 \pm 0.08	0.12 \pm 0.08	nd.	-	0.18 \pm 0.03	0.16 \pm 0.03	0.05 \pm 0.03	0.000
5003	0.87 \pm 0.65	0.58 \pm 0.22	nd.	-	0.73 \pm 0.15	0.59 \pm 0.13	0.14 \pm 0.03	0.000
6003	0.47 \pm 0.18	0.35 \pm 0.02	nd.	-	0.42 \pm 0.23	0.48 \pm 0.09	0.02 \pm 0.02	0.000
8204	0.36 \pm 0.12	0.33 \pm 0.19	nd.	-	0.11 \pm 0.02	0.20 \pm 0.05	0.00 \pm 0.00	0.021
6215	nd.	nd.	0.45 \pm 0.08	-	nd.	nd.	0.38 \pm 0.11	-
4207	0.31 \pm 0.04	0.18 \pm 0.35	nd.	-	0.18 \pm 0.02	0.06 \pm 0.07	nd.	-
2612	0.12 \pm 0.03	0.20 \pm 0.17	nd.	-	0.07 \pm 0.01	0.10 \pm 0.04	nd.	-
2708	0.20 \pm 0.09	0.42 \pm 0.49	nd.	-	0.16 \pm 0.04	0.15 \pm 0.04	nd.	-
3301	0.67 \pm 0.24	0.43 \pm 0.11	nd.	-	0.17 \pm 0.06	0.17 \pm 0.08	nd.	-
7015	0.33 \pm 0.19	0.10 \pm 0.07	nd.	-	0.15 \pm 0.06	0.10 \pm 0.03	nd.	-
7319	0.27 \pm 0.08	0.21 \pm 0.08	nd.	-	0.17 \pm 0.04	0.16 \pm 0.05	nd.	-
8108	0.15 \pm 0.11	0.14 \pm 0.04	nd.	-	0.18 \pm 0.02	0.13 \pm 0.04	nd.	-
8128	0.07 \pm 0.05	0.08 \pm 0.03	nd.	-	0.09 \pm 0.02	0.10 \pm 0.04	nd.	-

Two-dimensional biplots indicating associations between experimental samples and protein spots were generated (Fig. 7.3) by principal component analysis (PCA) in NIA array analysis tools. PCA results showed that the first (PC1) and the second (PC2) components explained 88.57% and 11.42% of the total variability, respectively. Samples from different

stages (Fig. 7.3A) and phosphoprotein spots (Fig. 7.3B) were plotted in PC1 and PC2 spaces. The consistent phosphorylated spots were sufficiently different to establish groups among the analysed tissues and even between mature and germinated seeds. The three analysed stages were separated among them; the first component separated stages S0 and S3 from S7, and the second component separated the three stages. After this rigorous statistical analysis to eliminate false positives, the 55 phosphoprotein spots were selected for MALDI-TOF/TOF analysis.

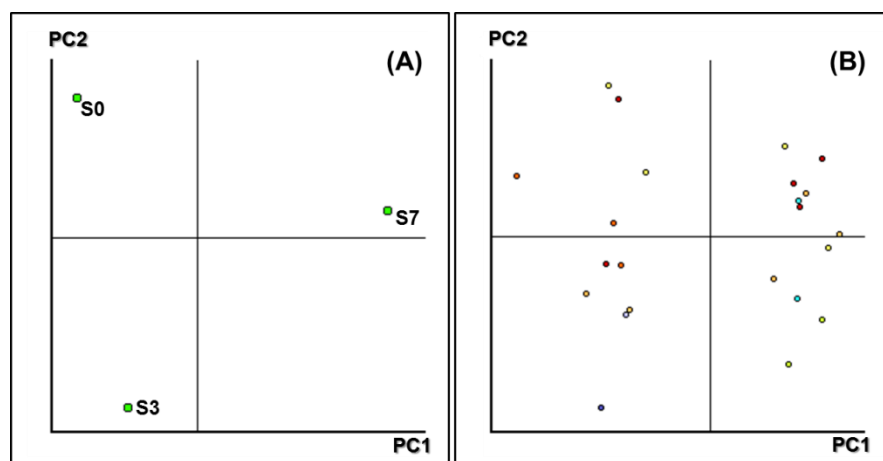


Figure. 7.3. Principal component analysis plots. (A) Representation of the samples based on the main principal components found after PCA. **(B)** Plot component PC1 vs PC2 of differentially expressed spots among three analysed stages.

To avoid considering as differentially phosphorylated those protein spots with differential accumulation of total protein, an ANOVA analysis was performed on the normalised volume of consistent spots detected with SYPRO-Ruby staining gels and the results were compared with differential spots obtained by Pro-Q DPS. Based on this ANOVA results comparison, we identify 20 out of 55 putative differentially phosphorylated proteins with no differences in total protein profile (Fig. 7.4, Table 7.2). Data indicated that these 20 protein spots have quantitative (8 spots) or qualitative (12 spots) alterations in phosphorylation levels during germination. In contrast, the other 35 spots, whether showing quantitative and qualitative differences, differed in both Pro-Q DPS and SYPRO-Ruby staining intensities and were denoted as phosphoproteins without changes in phosphorylation status (Table 7.2).

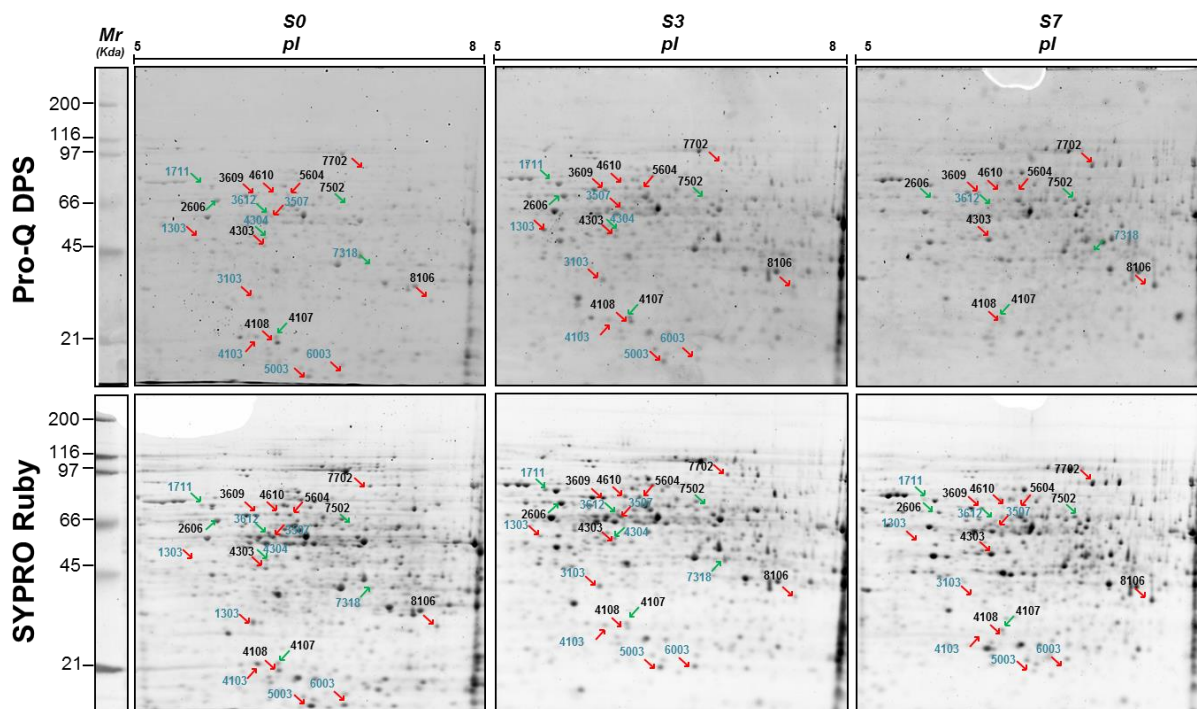


Figure 7.4. Representative 2-DE phosphoproteome profiles of *Q. ilex* embryos during germination and seedling. The gels were successively stained with Pro-Q DPS and SYPRO-Ruby as indicated in the figure and the images captured and analysed. Arrows indicate the identified proteins and the colors indicate the changes in phosphorylation status (green) or in protein levels (red).

4.3.1 Phosphoprotein identification

After MALDI-TOF/TOF analysis, a total of 20 proteins were identified, corresponding seven protein spots to proteins with changes in their phosphorylation status. The other 13 spots were considered as phosphoproteins with changes in abundance but not in their phosphorylation status. These identified proteins are listed in Tables 7.3 and 7.4 (Appendixes of this chapter), where appeared grouped in the following functional categories, based in the KEGG pathways database.

To investigate whether these 20 phosphoproteins had been previously identified in other species or biological process, as a way to ensure that they are phosphorylated proteins, we used the BLAST tool in the Plant Protein Phosphorylation DataBase (P3DB; <http://www.p3db.org/>)²³. With the single exception of the phosphoprotein spot number 3103, a putative cyclase family protein, the other *Q. ilex* proteins had at least one orthologous phosphoprotein in *A. thaliana*, *M. truncatula*, *Nicotiana. tabacum* and *Glycine max*. Table 7.5 (Appendixes of this chapter) list the orthologous proteins and their host species^{5; 21-22; 38-42}.

We try to identify by MALDI TOF/TOF analysis the phosphorylation sites in the peptides used to identify the phosphorylated proteins. The phosphorylation sites were identified in two phosphopeptides sequences based on Mascot probability score of 95%. One of








the phosphorylation sites was found in Ser³⁷⁹ present in the full length sequence of the protein pyruvate decarboxylase assigned to the spot 5604. This phosphorylation site corresponded to Thr³⁸⁰ pyruvate decarboxylase of *Arabidopsis*. The other phosphorylation site identified was Ser³⁷⁸ in the full length amino acidic sequence of 5-methyl-tetrahydropteroyl-triglutamate-homocysteine methyltransferase like, identified in the spots 7702. Two phosphorylation sites (Tyr⁶⁹⁸ and Ser⁷⁰²) has been described for this protein in *Arabidopsis*.

The identification of only these two phosphopeptides in the current study may be attributed to the methods used in MS analysis. The fragmentation of P-Ser- and P-Thr-phosphopeptides by CID in a MS frequently results in the loss of phosphoric acid from the full-length peptide, which complicates the confident assignment of the phosphorylation site to a particular residue in the peptide ^{6; 14}. As Ser is one of the preferred phosphorylation sites in identified phosphopeptides in plants ^{5; 21; 26}, probably we identified these two phosphosites because of their abundance.

4.3.2. Functional relevance of differentially phosphorylated proteins

Most of the proteins with altered phosphorylation status along the germination process (Table 7.3) were included in the *carbohydrate metabolism* category, and they showed their lowest phosphorylation state in unimbibed seeds. Similar results were reported in phosphoproteome analysis of rice germination ¹⁷. In contrast, phosphoproteins related to amino acids metabolism and protein folding showed their highest values in the seedling samples.

Table 7.3. List of identified phosphoprotein spots in 2-DE with changes in phosphoprotein status.

Spot number ^a	Protein name	Accession numbers ^b	Mr (pI)		MOWSE score ^e	Peptide matches	Seq cov (%)	Fragmented ion (Ion Score)	Normalised spot volume intensities ^f
			Theor. ^c	Exp. ^d					
Carbohydrate metabolism									
7502	Pyrophosphate-dependent phosphofructokinase beta subunit. <i>Citrus sinensis</i> x <i>Citrus trifoliata</i>	A9YVC9	62.0 (6.3)	53.8 (6.8)	283	17	22	DKIETPEQFK (59) STGKYHFVFR (45) YHFVFR (37) GQSHFFGYEGR (83)	
7318	Phosphoglycerate kinase_AT1G79550.1, <i>Quercus rubra</i>	QRU405_58	43.5 (6.7)	40.0 (7.1)	231	11	51	YSLKPIVPR (31) VILSTHLGRPK(30) FLKPAVAGFLMQR(21) LVAEIPGGVLLLENVR(26) LASLADLYVNDAFGTAHR (49)	
4304	Glucose-1-phosphate adenylyltransferase, <i>Vitis vinifera</i>	D7TDB6	56.2 (6.5)	37.4 (6.1)	350	26	38	VDTTILGLDDER (59) KPVPDFSFYDR (71) SSPIYTPQR (41) IINSDNVQEAAR (40)	
Amino acid metabolism									
3612	Glutamate decarboxylase_AT2G02010.1, <i>Quercus</i> spp.	TC19169_4 1	58.0 (5.9)	50.2 (6.1)	208	12	45	VVIREDFSR (30) ETPEEIATYWR (53) GSSQIIAQYYQFVR (69) NYVDMDEYPTTELQNR (46)	
Protein folding									
2606	Heat shock protein 60_AT3G23990.1, <i>Quercus</i> spp	TC33448_3 9	63.9 (5.6)	68.1 (5.8)	436	11	23	AGIHPVKVIR (51) IGVQIIQNALK (91) NNVIEQSWGAPK (82) GYISPYFITNQK (73) SDEIAQVGTISANGER (72) AAVEEGIVPGGGVALLYASK (75)	
UnKnown									
4107	Unknown protein	A9PFJ3	29.5 (6.2)	20.2 (6.3)	169	9	35	LQGNYYFQEQLSR (95) GSSIWYGCVLR (29)	
1711	Cell division protein ftsH, putative, <i>Ricinus communis</i>	B9S304	75.5 (6.4)	78.7 (5.7)	224	21	26	FLEYLDKDR (48) VRVQLPGLSQELLQK (3) SSGGMGPGGPGFPLAFGQSK (53) ADILDSALLRPR (17)	

^a Spot number as given on the 2-D gel images in Fig. 7.2

^b Uniprot and Quercus_DB accession numbers. The accession whose first letters were TC and QRU correspond to Quercus_DB

^c Molecular weight (kDa) and isoelectric point of each database

^d Molecular weight (kDa) and isoelectric point calculated by using molecular weight standards and the PD-Quest Advance (8.01) software.

^e Mascot score ($S = -10 \times \log(P)$): where P is the probability that the observed match is a random event, peptide matched in MS analysis, percentage of sequence coverage and ions sequence matched from MS/MS analysis.

^f The bar charts represent the normalised spot volume intensities vs analysed stages.

Phosphorylation and dephosphorylation of metabolic enzyme may regulate carbohydrates metabolism and mobilization during seed germination, and constitute an important pathway to provide energy during early seedling development. The proteins with changes in their phosphorylation status were mainly involved in glycolysis: pyrophosphate-dependent phosphofructokinase (P_{Pi}-PFK, spot 7502), phosphoglycerate kinase (PGK, spot 7318) and glucose-1-phosphate adenylyltransferase (AGP, spot 4304) (Tabla 7.3 and Fig. 7.5).

P_{Pi}-PFK is an exclusively cytosolic enzyme that catalyzes the phosphorylation of fructose-6-phosphate to fructose-1,6-bisphosphate in the glycolytic direction, using inorganic pyrophosphate as the phosphoryl donor. This process makes fructose flows into glycolysis to provide energy. PGK catalyzes the conversion of 1,3-diphosphoglycerate to 3-phosphoglycerate, the first substrate-level phosphorylation reaction in the glycolytic pathway for production of ATP. AGP catalyses the synthesis of ADP-glucose, which is the active glucoside for starch synthesis (Fig. 7.5).

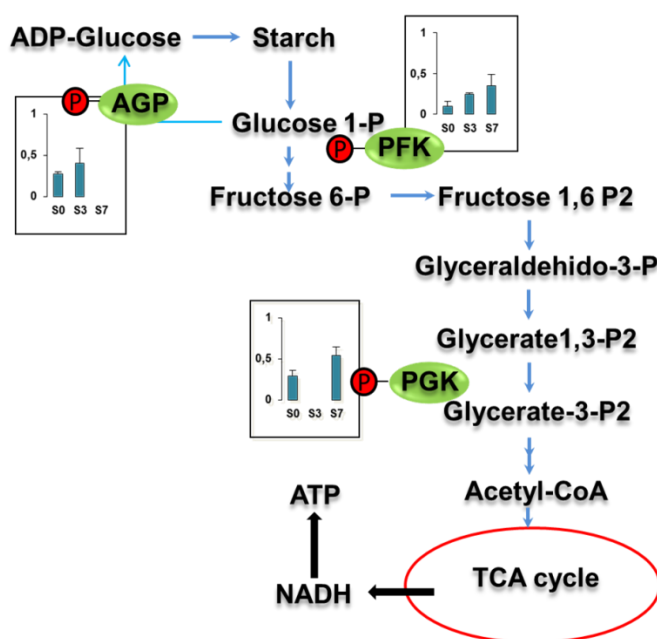


Figure 7.5: Carbohydrate metabolic pathway showing proteins with a change in phosphorylation (indicated by green circle). The bar charts represent the normalised spot volume intensities vs analysed stages of each protein.

Overall, the phosphorylation states of these three enzymes increased along the germination process. Phosphorylation modification of many glycolytic enzymes has been to cause a significant increase in enzyme activity⁴³, increasing the glycolysis rate and the generation of energy to afford the needs of the developing seedlings. These results are in agreement with previous studies on rice germination and seedling^{5; 17; 27}. In addition, they are supported by data of sugars content measurements of Ch.4.

Other glycolysis enzyme, pyruvate decarboxylase (Table 7.4), showed decreased abundance as germination and seedling establishment progressed, but no changes in their phosphorylation status were detected in this work. Pyruvate decarboxylase (PDC) catalyzes the decarboxylation of pyruvate to yield CO₂ and acetaldehyde, which is then reduced to ethanol by alcohol dehydrogenase. Phosphorylation of PDC increases its activity and redirect pyruvate to alcoholic fermentation. We not detected changes in its phosphorylation levels, in contrast with previous reports on rice germination, where the phosphorylation status of the two enzymes increased¹⁷. However, the results presented here would better explain the use of pyruvate in glycolysis, not in fermentation, something that probably should be avoid in a developmental organisms like a germinating seed.

Glutamate decarboxylase (GDC) (Table 7.3) is an enzyme that catalyzes the decarboxylation of glutamate to GABA and CO₂. GABA, as a non-protein amino acid, is involved in various stress tolerances in plants and has been described to accumulate in germinating seeds of rice and tomato⁴⁴⁻⁴⁵. Some isoforms of this enzyme result inhibited by phosphorylation. If applicable to GDC, the reduction in its phosphorylation status observed here might imply an increase in the activity of this enzyme, to eliminate the excess of glutamate and glutamine originated by the high rates of stored proteins degradation occurring during germination.

Proteins like 5-methyltetrahydropteroyltriglutamate homocysteine methyltransferase like and S-adenosylmethionine synthase 2, increased their abundance and their phosphorylation status along the germination process (Table 7.4). These enzymes are involved in the utilization of amino acids produced from the breakdown of storage proteins by proteases for the *de novo* biosynthesis of amino acids and proteins and the synthesis of many other metabolites. The *de novo* biosynthesis of Met is catalysed by 5-methyltetrahydropteroyltriglutamate-homocysteine S-methyltransferase. Met is a fundamental metabolite because it functions not only as a building block for protein synthesis but also as the precursor of S-adenosylmethionine (SAM), the universal methyl-group donor. SAM synthase catalyses the conversion of Met in SAM, the precursor of the ethylene and the spermidine/spermine biosynthesis pathways. SAM synthetase is considered a key regulator of metabolism in the transition from a quiescent to a highly active state during Arabidopsis seed germination⁴⁶.

The phosphorylation status of heat shock proteins (HSPs), involved in protein folding, has been described to decrease during rice germination¹⁷. In agreement with that, HSP60 showed high levels of phosphorylation in non-imbibed (S0) and germinated seeds (S3), and then decreased levels in seedling development (S7).

Germination is a rehydration process in which the levels of oxidative stress increase and result in the production and accumulation of reactive oxygen species (ROS) ⁴⁷. In rice, oxidation-reduction enzyme glutathione S-transferase (GST) was found to be phosphorylated during seedling development, and retaining high enzyme activity even though its protein abundance was significantly decreased ¹⁷. We found here the decrease in GST transcripts and proteins abundance described in previous chapters of this work is accompanied by a parallel diminution in the level of its phosphorylation.

4.4. Concluding remarks

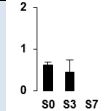
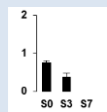
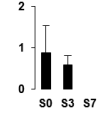
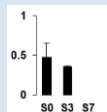
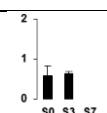
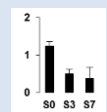
We have used here a combination of Pro-Q DPS/SYPRO-Ruby differential staining methodology to obtain from the same 2-DE gel the phosphoproteins and total proteins profiles of *Q. ilex* germinating seeds.

Over 200 putative phosphoproteins spots were detected by this method. Among them, 20 proteins exhibited significant changes in their phosphorylation status, of which seven were identified. A further 35 phosphoproteins exhibited statistically significant changes in the total protein levels, of which 13 were identified. The identified 20 phosphoproteins could be classified into the following five functional categories: carbohydrate and amino acid metabolism, defence, protein folding and reduction-oxidation. Alterations in phosphorylation status of proteins related to glycolysis and secondary metabolites biosynthesis from Met, indicated that these pathways increased from mature seeds to germinated seeds and seedling. Although only a small part of the phosphoproteome was covered and analysed, integration of data obtained from total proteome analysis and phosphoproteomics analysis suggested that metabolic machinery present in the recalcitrant seeds receives an activation signal to activate and resume the most important metabolic pathways (the glycolysis) in *Q. ilex* to start the germination and the establishment of the seedlings.

4.5. Appendixes

Table 7.4. List of identified phosphoprotein spots in 2-DE without changes in phosphorylations status

Numbers ^a	Protein name	Accession numbers	Mr (pI)		MOWSE score ^e	Peptide matches	Seq cov (%)	Fragmented ion (Ion Score)	Normalized spot volumen intensities ^f
			Theor. ^c	Exp. ^d					
Carbohydrate metabolism									
4610	Pyruvate decarboxylase. <i>Prunus armeniaca</i>	B0ZS79	66.2 (5.7)	61.9 (6.2)	235	12	23	ILHHTIGLPDFSQELR(124) EPVVPFSLSPR(69)	
3609	Phosphoglycerate mutase_AT1G09780.1. <i>Quercus petraea</i>	QP1063_77	61.0 (6.0)	60.5 (6.0)	476	10	25	DAILSGKFDQVR (46) FGHVTFWNGNR (77) AFEYEDFDKFDK (67) LPSHYLVSPPEIDR (78) GTLHLIGLSDGGVHSR (98) IQILTSHTCQVPVPIAIGGPGLAPGCR (88) AHGSAVGLPTEDDMGNSEVGHNALGAGR (31)	
5604	Pyruvate decarboxylase. putative. <i>Ricinus communis</i>	255563082	64.2 (5.9)	61.1 (6.3)	112	2	4	ILHHTIGLPDFSQELR (97) IFVPSGVPLK (22) Phosphorilated S379 #	
3507	Beta glucosidase 17_AT2G444480.1, <i>Quercus</i> spp.	QRO15180_40	47.8 (5.2)	60.3 (6.2)	119	10	48	GAYDFIGVNYTTSR (103)	
Amino acid metabolism									
7702	5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase-like. <i>Solanum lycopersicum</i>	460407874	85.0 (6.0)	82.5 (7.0)	304	4	6	YLFAGVVDGR(76) ALSGAKDEAFPSANAAAQASR(41) Phosphorilated S378# EGVKYAGIGPGVYDIHSPR (64) YGAGIGPGVYDIHSPR(122)	
4303	S-adenosylmethionine synthase 2, <i>Elaeagnus umbellata</i>	Q9AT55	43.6 (5.5)	47.0 (6.2)	650	22	44	TIGFVSDVGLDADNCK (83) VLVNIQQSPDIAQGVHGHFTK(97) TQVTVEYYNDKGMVPVR(17) TIFHLNPSGR(61) FVIGGPHGDAGLTGR(110) FVIGGPHGDAGLTGRK (89) TAAYGHFGR (79)	

Numbers ^a	Protein name	Accession numbers	Mr (pI)		MOWSE score ^e	Peptide matches	Seq cov (%)	Fragmented ion (Ion Score)	Normalized spot volumen intensities ^f
			Theor. ^c	Exp. ^d					
3103	Putative cyclase family protein, <i>Arachis hypogaea</i>	COL2U1	31.5 (5.04)	30.7 (6.5)	79	8	28	IFDISHR (36)	
Oxidation-reduction process									
4103	Glutathione S-transferase omega_D6BR66, <i>Quercus</i> spp	TC18312_19	28.2 (6.6)	26.0 (6.0)	75	8	36	LYISLSCPYAQR (24) EAGPAFDHLENALSK (14) WIEEVNKIDAYKPTK (11) YIDSNFEGPSLLPNDHAK (24)	
5003	Putative uncharacterized protein (Glutathione-s-transferase theta_B9T0U8), <i>Vitis vinifera</i>	D7TP00	24.9 (6.2)	15.3(6.6)	107	6	22	NPFQIQPVLEGGDLTLFESR (38) AWWEDISSRPAFK (46)	
6003	Manganese superoxide dismutase 1_AT3G10920.1, <i>Quercus</i> spp	TC29211_11	19.3 (7.9)	15.4 (6.8)	126	5	55	HHQAYITNYNK (73) FNGGGHINHSIFWK (42) KLVVDTTANQDPLVTK(2)	
RNA metabolism									
1303	DEAD box RNA helicase, <i>Pisum sativum</i>	Q8H1A5	47.1 (5.4)	49.6 (5.6)	258	24	46	GIYAYGFEKPSAIQQR (60) ILSSGVHVVVGTTPGR (28) VFDMLRR (16) MFVLDEADEMLSR (11) VLITTDLLAR (21)	
Stress response									
4108	Aluminium induced protein with YGL and LRDR motifs_AT3G22850.1. <i>Quercus</i> spp.	TC18137_21	27.8 (7.0)	20.7 (6.3)	141	5	27	GCFFTSSGGLR(31) FAMILYDSSSK (49) SYEHPLNEVKPVPR (58) SPEALQSPQSGSVSTLK (4)	
Unknown									

Numbers ^a	Protein name	Accession numbers	Mr (pI)		MOWSE score ^e	Peptide matches	Seq cov (%)	Fragmented ion (Ion Score)	Normalized spot volumen intensities ^f								
			Theor. ^c	Exp. ^d													
8106	AT4G39230.1_NmrA-like negative transcriptional regulator family protein). <i>Quercus robur</i>	QRO2324_17	36.5 (6.8)	26.2 (7.6)	367	11	35	AGHPTFALVR(67) VLIIGGTGYIGK(68) FYPSEFGNDVDR(67) AIFNKEDDIGTYTIK(17) NLGVTLVHGDLYDHGSLVK(103) FYPSEFGNDVDRVHAVDPAK(17) GDHTNFEIEPSFGVEASQLYPDVK(33)	<table border="1"> <caption>Normalized spot volume intensities</caption> <thead> <tr> <th>Spot</th> <th>Intensity</th> </tr> </thead> <tbody> <tr> <td>S0</td> <td>~0.5</td> </tr> <tr> <td>S3</td> <td>~0.5</td> </tr> <tr> <td>S7</td> <td>~1.5</td> </tr> </tbody> </table>	Spot	Intensity	S0	~0.5	S3	~0.5	S7	~1.5
Spot	Intensity																
S0	~0.5																
S3	~0.5																
S7	~1.5																

^a Spot number as given on the 2-DE gel images in Fig 7.2.

^b Uniprot, NCBI nr and Quercus_DB accessions numbers. The accession whose first letters were TC, QRU, QRO and QP correspond to Quercus_DB. The accession numbers without letters correspond to NCBI nr

^c Molecular weight (kDa) and isoelectric point of each database

^d Molecular weight (kDa) and isoelectric point calculated by using molecular weight standards and the PD-Quest Advance (8.01) software.

^e Mascot score ($S = -10 \times \log(P)$): where P is the probability that the observed match is a random event, peptide matched in MS analysis, percentage of sequence coverage, and ions sequence matched from MS-MS analysis.

^f The bar charts represent the normalized spot volume intensities vs analysed stages.

^g The phosphorylation sites is referred to the position of amino acids in the proteins

Table 7 5. Phosphorylation sites and close species in which the phosphoprotein identified in *Q. ilex* germinated seeds and seedling development were identified.

<i>Spot Number</i>	<i>Protein name</i>	<i>Protein ID</i>	<i>Close species</i>	<i>Phosphorylation sites</i>	<i>Process in which were described</i>	<i>References</i>
2606	Heat shock protein 60 AT3G23990.1, <i>Quercus</i> spp	TC33448_39	<i>Arabidopsis thaliana</i>	474 S		Reiland et al. 2009
7318	Phosphoglycerate kinase_AT1G79550.1, <i>Quercus rubra</i>	QRU405_58	<i>Arabidopsis thaliana</i>	81S 86T 87S	Phosphoproteome characterization of <i>Arabidopsis</i> seedlings shoots and rosette leaves using IMAC and TiO ₂ phosphopeptide enrichment strategies	Reiland et al. 2009
1711	Cell division protein ftsH, putative, <i>Ricinus communis</i>	B9S304	<i>Arabidopsis thaliana</i>	86T		Reiland et al. 2009
1303	DEAD box RNA helicase, <i>Pisum sativum</i>	Q8H1A5	<i>Arabidopsis thaliana</i>	76S 86S 85Y 105S 716S 723S		Reiland et al. 2009
4107	Unknown protein	A9PFJ3	<i>Oryza sativa</i>	235S 261S		Nakagami, et al. 2010
3612	Glutamate decarboxylase_AT2G02010.1, <i>Quercus</i> spp.	TC19169_41	<i>Arabidopsis thaliana</i>	8S 10S 13S	Large-scale analysis of rice phosphorylation sites from non-stimulated suspension-cultured rice cells	Nakagami, et al. 2010
4304	Glucose-1-phosphate adenylyltransferase, <i>Vitis vinifera</i>	D7TDB6	<i>Arabidopsis thaliana</i>	77S		Nakagami, et al. 2010
3609	Phosphoglycerate mutase_AT1G09780.1, <i>Quercus petraea</i>	QP1063_77	<i>Arabidopsis thaliana</i>	82 S	Phosphoproteome role in tobacco pollen activated in vitro and large-scale phosphoproteomics Large-scale phosphoproteomics analysis in roots	Nakagami, et al. 2010 and ⁴¹
3507	Beta glucosidase 17_AT2G44480.1, <i>Quercus</i> spp.	QRO15180_40	<i>Medicago truncatula</i>	82 T		Grimsrud et al 2010
4610	Pyruvate decarboxylase.	B0ZS79	<i>Nicotiana tabacum</i>	380T	Phosphoproteome role in	Fila et al. 2012

Spot Number	Protein name	Protein ID	Close species	Phosphorylation sites	Process in which were described	References
	<i>Prunus armeniaca</i>				tobacco pollen activated in vitro	
5604	Pyruvate decarboxylase, Putative, <i>Ricinus communis</i>	gi55563082	<i>Nicotiana tabacum</i>	380 T		
8106	AT4G39230.1_NmrA-like negative transcriptional regulator family protein, <i>Quercus robur</i>	QR02324_17	<i>Medicago truncatula</i>	171 T	Integrated large-scale approach to investigate changes in the phosphoproteome, proteome, and transcriptome that occur one hour after Nod factors treatment in <i>Medicago truncatula</i>	Rose et al. 2012
4303	S-adenosylmethionine synthase 2, <i>Elaeagnus umbellata</i>	Q9AT55	<i>Medicago truncatula</i>	131 S 266 S		
4103	Glutathione S-transferase omega_D6BR66, <i>Quercus</i> spp	TC18312_19	<i>Medicago truncatula</i>	10 T		
4108	Aluminium induced protein with YGL and LRDR motifs_AT3G22850.1, <i>Quercus</i> spp.	TC18137_21	<i>Arabidopsis thaliana</i>	215S 216S 240S	Whole cell suspension line, seedlings and seed maturation of rapessed, Arabidopsis and soybean phosphoproteome	Sugiyama et al 2008, Meyer et al. 2012
6003	Manganese superoxide dismutase 1_AT3G10920.1, <i>Quercus</i> spp	TC29211_11	<i>Glycine max</i>	173S	Analysis of seed maturation in arabidopsis, rapeseed, and Soybean	Meyer et al. 2012
7502	Pyrophosphate-dependent phosphofructokinase beta subunit. <i>Citrus sinensis</i> x <i>Citrus trifoliata</i>	A9YVC9	<i>Arabidopsis thaliana</i>	12T 16S	Nuclear phosphoproteins analysis of Arabidopsis.	Jones et al. 2009, Reiland et al. 2009
5003	Putative uncharacterized protein (Glutathione-s-transferase theta_B9T0U8), <i>Vitis vinifera</i>	D7TP00	<i>Arabidopsis thaliana</i>	12 S	Large-scale phosphoproteome analysis of Arabidopsis cell suspension line,	Sugiyama et al. 2008

<i>Spot Number</i>	<i>Protein name</i>	<i>Protein ID</i>	<i>Close species</i>	<i>Phosphorylation sites</i>	<i>Process in which were described</i>	<i>References</i>
7702	5-methyltetrahydropteroyltriglutamate homocysteine methyltransferase-like, <i>Solanum lycopersicum</i>	460407874	<i>Arabidopsis thaliana</i>	698 Y 702 S-		
3103	Putative cyclase family protein, <i>Arachis hypogaea</i>	COL2U1	No hits			

4.6. References

1. Ytterberg AJ, Jensen ON (2010) Modification-specific proteomics in plant biology. *J Proteomics* 73: 2249-2266
2. Bond AE, Row PE, Dudley E (2011) Post-translation modification of proteins; methodologies and applications in plant sciences. *Phytochemistry* 72: 975-996
3. Kersten B, Agrawal GK, Iwahashi H, Rakwal R (2006) Plant phosphoproteomics: A long road ahead. *PROTEOMICS* 6: 5517-5528
4. Hunt L, Holdsworth MJ, Gray JE (2007) Nicotinamidase activity is important for germination. *The Plant Journal* 51: 341-351
5. Nakagami H, Sugiyama N, Mochida K, Daudi A, Yoshida Y, Toyoda T, Tomita M, Ishihama Y, Shirasu K (2010) Large-scale comparative phosphoproteomics identifies conserved phosphorylation sites in plants. *Plant Physiology*
6. de la Fuente van Bentem S, Hirt H (2010) Protein tyrosine phosphorylation in plants: more abundant than expected? *Trends in Plant Science* 14: 1360-1385
7. Kaufmann H, Bailey JE, Fussenegger M (2001) Use of antibodies for detection of phosphorylated proteins separated by two-dimensional gel electrophoresis. *PROTEOMICS* 1: 194-199
8. Woods Ignatoski K (2001) Immunoprecipitation and western blotting of phosphotyrosine-containing proteins. In: Reith A (ed) *Protein Kinase Protocols*. Humana Press, pp 39-48
9. Agrawal GK, Thelen JJ (2005) Development of a simplified, economical polyacrylamide gel staining protocol for phosphoproteins. *PROTEOMICS* 5: 4684-4688
10. Wu R, Haas W, Dephoure N, Huttlin EL, Zhai B, Sowa ME, Gygi SP (2011) A large-scale method to measure absolute protein phosphorylation stoichiometries. *Nature Methods* 8: 677-683
11. Rigbolt KT, Blagoev B (2012) Quantitative phosphoproteomics to characterize signaling networks. *Seminars in Cell & Developmental Biology* 23: 863-871
12. Cox J, Mann M (2011) Quantitative, high-resolution proteomics for data-driven systems biology. *Annual Review of Biochemistry* 80: 273-299
13. Thingholm TE, Jensen ON, Larsen MR (2009) Analytical strategies for phosphoproteomics. *PROTEOMICS* 9: 1451-1468
14. Kosako H, Nagano K (2011) Quantitative phosphoproteomics strategies for understanding protein kinase-mediated signal transduction pathways. *Expert Review of Proteomics* 8: 81-94
15. Ivanov KI, Puustinen P, Gabrenaite R, Vihinen H, Rönstrand L, Valmu L, Kalkkinen N, Mäkinen K (2003) Phosphorylation of the potyvirus capsid protein by protein kinase CK2 and its relevance for virus infection. *The Plant Cell Online* 15: 2124-2139
16. Ubersax JA, Ferrell Jr JE (2007) Mechanisms of specificity in protein phosphorylation. *Nature Reviews Molecular Cell Biology* 8: 530-541
17. Han C, Wang K, Yang P (2014) Gel-based comparative phosphoproteomic analysis on rice embryo during germination. *Plant and Cell Physiology*
18. Fujii H, Chinnusamy V, Rodrigues A, Rubio S, Antoni R, Park S-Y, Cutler SR, Sheen J, Rodriguez PL, Zhu J-K (2009) In vitro reconstitution of an abscisic acid signalling pathway. *Nature* 462: 660-664
19. Cutler SR, Rodriguez PL, Finkelstein RR, Abrams SR (2010) Abscisic acid: emergence of a core signaling network. *Annual Review of Plant Biology* 61: 651-679
20. Umezawa T, Sugiyama N, Takahashi F, Anderson JC, Ishihama Y, Peck SC, Shinozaki K (2013) Genetics and phosphoproteomics reveal a protein phosphorylation network in the abscisic acid signaling pathway in *Arabidopsis thaliana*
21. Sugiyama N, Nakagami H, Mochida K, Daudi A, Tomita M, Shirasu K, Ishihama Y (2008) Large-scale phosphorylation mapping reveals the extent of tyrosine phosphorylation in *Arabidopsis*. *Molecular Systems Biology* 4
22. Reiland S, Messerli G, Baerenfaller K, Gerrits B, Endler A, Grossmann J, Gruissem W, Baginsky S (2009) Large-scale *Arabidopsis* phosphoproteome profiling reveals novel chloroplast kinase substrates and phosphorylation networks. *Plant Physiology* 150: 889-903
23. Kersten B, Agrawal GK, Durek P, Neigenfind J, Schulze W, Walther D, Rakwal R (2009) Plant phosphoproteomics: An update. *PROTEOMICS* 9: 964-988

24. Rose CM, Venkateshwaran M, Grimsrud PA, Westphall MS, Sussman MR, Coon JJ, Ané J-M (2012) Medicago PhosphoProtein Database: a repository for *Medicago truncatula* phosphoprotein data. *Frontiers Plant Sciences* 3
25. Alonso J, Zapata C (2014) Evidence for phosphorylation of the major seed storage protein of the common bean and its phosphorylation-dependent degradation during germination. *Plant Molecular Biology* 84
26. Lu T-C, Meng L-B, Yang C-P, Liu G-F, Liu G-J, Ma W, Wang B-C (2008) A shotgun phosphoproteomics analysis of embryos in germinated maize seeds. *Planta* 228: 1029-1041
27. Chen F, Jiang L, Zheng J, Huang R, Wang H, Hong Z, Huang Y (2014) Identification of differentially expressed proteins and phosphorylated proteins in rice seedlings in response to strigolactone treatment. *PLoS One* 9: e93947
28. Schulenberg B, Goodman TN, Aggeler R, Capaldi RA, Patton WF (2004) Characterization of dynamic and steady-state protein phosphorylation using a fluorescent phosphoprotein gel stain and mass spectrometry. *ELECTROPHORESIS* 25: 2526-2532
29. Berggren K, Chernokalskaya E, Steinberg TH, Kemper C, Lopez MF, Diwu Z, Haugland RP, Patton WF (2000) Background-free, high sensitivity staining of proteins in one- and two-dimensional sodium dodecyl sulfate-polyacrylamide gels using a luminescent ruthenium complex. *ELECTROPHORESIS* 21: 2509-2521
30. Agrawal GK, Thelen JJ (2006) Large scale identification and quantitative profiling of phosphoproteins expressed during seed filling in oilseed rape. *Molecular & Cellular Proteomics* 5: 2044-2059
31. Sharov AA, Dudekula DB, Ko MSH (2005) A web-based tool for principal component and significance analysis of microarray data. *Bioinformatics* 21: 2548-2549
32. Sghaier-Hammami B, Valero-Galván J, Romero-Rodríguez MC, Navarro-Cerrillo RM, Abdelly C, Jorrín-Novo J (2013) Physiological and proteomics analyses of Holm oak (*Quercus ilex* subsp. *ballota* [Desf.] Samp.) responses to *Phytophthora cinnamomi*. *Plant Physiology and Biochemistry* 71: 191-202
33. Romero-Rodríguez MC, Pascual J, Valledor L, Jorrin-Novo J (2014) Improving the quality of protein identification in non-model species. Characterization of *Quercus ilex* seed and *Pinus radiata* needle proteomes by using SEQUEST and custom databases. *Journal of Proteomics* 105: 85-91
34. Gao J, Agrawal GK, Thelen JJ, Xu D (2009) P3DB: a plant protein phosphorylation database. *Nucleic Acids Research* 37: D960-D962
35. 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
36. Meyer LJ, Gao J, Xu D, Thelen JJ (2012) Phosphoproteomic analysis of seed maturation in Arabidopsis, rapeseed, and soybean. *Plant Physiology* 159: 517-528
37. Subba P, Barua P, Kumar R, Datta A, Soni KK, Chakraborty S, Chakraborty N (2013) Phosphoproteomic dynamics of chickpea (*Cicer arietinum* L.) reveals shared and distinct components of dehydration response. *Journal of Proteome Research* 12: 5025-5047
38. Jones AME, MacLean D, Studholme DJ, Serna-Sanz A, Andreasson E, Rathjen JP, Peck SC (2009) Phosphoproteomic analysis of nuclei-enriched fractions from *Arabidopsis thaliana*. *Journal of Proteomics* 72: 439-451
39. Li H, Wong WS, Zhu L, Guo HW, Ecker J, Li N (2009) Phosphoproteomic analysis of ethylene-regulated protein phosphorylation in etiolated seedlings of Arabidopsis mutant ein2 using two-dimensional separations coupled with a hybrid quadrupole time-of-flight mass spectrometer. *PROTEOMICS* 9: 1646-1661
40. Grimsrud PA, den Os D, Wenger CD, Swaney DL, Schwartz D, Sussman MR, Ané J-M, Coon JJ (2010) Large-scale phosphoprotein analysis in *Medicago truncatula* roots provides insight into in vivo kinase activity in legumes. *Plant Physiology* 152: 19-28
41. Fíla J, Matros A, Radau S, Zahedi RP, Čapková V, Mock H-P, Honys D (2012) Revealing phosphoproteins playing role in tobacco pollen activated in vitro. *PROTEOMICS* 12: 3229-3250
42. Rose CM, Venkateshwaran M, Volkening JD, Grimsrud PA, Maeda J, Bailey DJ, Park K, Howes-Podoll M, den Os D, Yeun LH, Westphall MS, Sussman MR, Ané J-M, Coon JJ (2012) Rapid phosphoproteomic and transcriptomic changes in the rhizobia-legume symbiosis. *Molecular & Cellular Proteomics* 11: 724-744

43. Li X, Zhuo J, Jing Y, Liu X, Wang X (2011) Expression of a GALACTINOL SYNTHASE gene is positively associated with desiccation tolerance of *Brassica napus* seeds during development. *Journal of Plant Physiology* 168: 1761-1770
44. Taji T, Ohsumi C, Iuchi S, Seki M, Kasuga M, Kobayashi M, Yamaguchi-Shinozaki K, Shinozaki K (2002) Important roles of drought- and cold-inducible genes for galactinol synthase in stress tolerance in *Arabidopsis thaliana*. *Plant J* 29: 417-426
45. Li H, Jiang H, Bu Q, Zhao Q, Sun J, Xie Q, Li C (2011) The Arabidopsis RING Finger E3 ligase RHA2b acts additively with RHA2a in regulating abscisic acid signaling and drought response. *Plant Physiology* 156: 550-563
46. Pawlowski T (2009) Proteome analysis of Norway maple (*Acer platanoides* L.) seeds dormancy breaking and germination: influence of abscisic and gibberellic acids. *BMC Plant Biology* 9: 48
47. Tan L, Chen S, Wang T, Dai S (2013) Proteomic insights into seed germination in response to environmental factors. *PROTEOMICS* 13: 1850-1870

CHAPTER 8:

GENERAL DISCUSSION

This Doctoral Thesis is focused on Holm oak (*Quercus ilex* L. subsp. *ballota* [Desf.] Samp.) a dominant tree species in natural forest ecosystems over large areas of the Western Mediterranean Basin ¹. *Q. ilex* forest maintenance and sustainability are facing important problems and challenges related to seed viability/conservation, and plant mortality in both adult trees and young one-year-old plants after field transplantation resulting from adverse environmental conditions like drought and the so-called decline syndrome ².

It is now evident that a cellular function is determined by many gene products and constitutes a complex network. These networks are regulated at transcriptional and post-transcriptional level, being necessary a multidisciplinary approach integrating transcriptomics, proteomics, metabolomics, bioinformatics and statistical analysis. Thus, “omics” techniques are useful for study complex biological processes, hardly achieved with classical molecular biology. However, the classical techniques are not displaced by “omics” approaches, they are generally used to validate data generated by holistic methods.

In the present work a multidisciplinary “omics” and metabolite analysis were combined to study at molecular level the germination and seedling growth of *Q. ilex*, to obtain a better understanding of molecular mechanisms involved in these processes. This knowledge is essential for restoration and reforestation programs with this orphan, recalcitrant forest tree species of great ecological importance and great economic interest. This Discussion attempts to integrate the changes observed at the transcriptional, translational and posttranslational levels, with the determinations of some metabolites (sugars, plant hormones) that arise or disappear in the cell as a consequence of those changes, to provide novel insights into the molecular pathways that could mediate seed germination and seedling establishment in Holm oak. Along the discussion, data will aid to compare the recalcitrant seeds biology with that of orthodox seeds, much more studied at molecular level.

A targeted strategy allowed to analyse a group of twelve protein coding genes involved in (i) desiccation tolerance (*Dhn3*, *GolS*), (ii) regulation of ABA-signalling (*Ocp3*, *Skp1*, *Pp2c*, *Sdir1*), (iii) metabolism (*Fdh*, *Gapdh*, *RbcL*, *Ndh6*) and (iv) antioxidative defence (*Sod1*, *Gst*) during germination and post germination of *Q. ilex* seeds. In holm oak mature acorns it was found that, like other recalcitrant seeds, maintain a partially active metabolism, with high level of glycolytic (*Gapdh*) and mitochondrial respiratory enzymes (*Ndh6*). In contrast, mature *Q. ilex* seeds show some of the intracellular physiological characteristics of orthodox seed that included accumulation of sucrose, DHN3 and

transcripts involved in the synthesis of certain osmoregulator raffinose series oligosaccharides (*GolS*) and in the anti-oxidative defence (*Sod1*, *Gst*).

Despite the high metabolic activity in mature seeds, during imbibition (at the S2), the transcript levels of *Ndh6* and *Fdh* (Fig. 8.1) start to increase. This coincides with phase II of seed germination, which is the most critical stage in orthodox seeds³, during which all necessary metabolic pathways and physiological processes are reactivated, and hence decision is made to initiate the germination or not. The transcriptional analysis results were complemented with the determination of sugars levels and plant hormones. In the sugar analysis, a decrease of sucrose and increase of reducing sugars like glucose or fructose during imbibition and germination were found. This might be associated with the mobilization of storage polysaccharides (starch) and increased catabolism of the glucose in order to satisfy the energetic and carbon request to support early seedling growth⁴⁻⁶.

The plant hormones analysis revealed a low level of ABA and high level of GA4, comparing with mature and dry orthodox seed⁷⁻¹⁰, however the content of ABA and GA was similar to described in other recalcitrant species including *Q. robur* and *Avicennia marina*¹¹⁻¹². Seed germination is antagonistically controlled by the phytohormones GA and ABA, GA promotes seed germination and ABA blocks this process. A dynamic balance between synthesis and catabolism of these hormones controls the equilibrium between dormancy and germination¹³. GAs are able to induce a range of genes which are necessary for modifying proteins of cell wall including α -amylase, proteases and β -glucanases, to promote the radicle protrusion¹⁴⁻¹⁷. On the other hand cytokinins such as iP and iPR, play a role in many aspects of plant growth and development, (including cell division, cell enlargement, senescence and differentiation): and have also been implicated in favouring the germination and early post-germination events¹⁸. The time-course analysis of iPR in the recalcitrant *Q. ilex* seeds showed an increase of this anti-dormancy hormones during the germination process.

The transcriptional analyses were extended with a prospecting of differentially expressed genes between *Q. ilex* germinated seeds and seedlings. It was found that in germinating embryo axis the most abundant genes were related to stress response, transport and oxidation-reduction process. These were in good agreement and validated the results obtained in targeted strategy focused in some genes. High transcript level of *Lea5* was found, which together with the *Dhn3*, are protective molecules¹⁹⁻²⁰. The up-accumulation of an endoglucanase in germinated seeds could be related to the high levels of GA4. In previous study was described that GAs induce genes related to cell wall

modification ²¹. Genes over-expressed in seedlings, were mainly implicated in the transition switch from the heterotrophic to photoautotrophic metabolism and the biosynthesis of secondary metabolites. In the seedlings an increase of *RbcL*(Fig. 8.1)the transcript and protein levels were found by RT-qPCR and immunoblotting analysis was found, along with the up-accumulation of transcripts of chlorophyll a/b binding protein, photosystem I reaction centre and coproporphyrinogen-III oxidase, all of them implicated in carbon fixation or photosynthesis ²².

Given that proteins are direct executors of the genetic information and they are major biomonitoring end points, it was investigated whether the changes in mRNA expression were reflected at the protein level. The comparative proteome profile analysis performed between mature and germinated seeds and seedlings revealed that enzymes of the carbohydrate and protein metabolic pathways, as well as those involved in redox homeostasis were up-accumulated in germinated and mature seeds; then decreasing along the germination and early seedling growth. The accumulation of proteins related to carbohydrate metabolism in mature acorns suggested that this pathway is essential for energy supply during *Q. ilex* seed germination as described in other species²³. Similarly, the accumulation of oxidative stress related enzymes was interpreted as a defence strategy against the accumulation of ROS under low water potential at the maturation stage. An increase in the abundance photosynthesis related proteins was also observed during seedling growth. These results agreed with that observed at the transcript level. However, the few changes detected between mature and germinated seeds such as proteins related to Met metabolism (SAMS, NLP3), the shikimate pathway (DHAPS) and the antioxidative defence (Fig. 8.1) were revealed only by proteomics approach. These results were in agreement with other analysed recalcitrant seeds ²⁴⁻²⁵.

At posttranslational level, changes in phosphoproteins were analysed for the first time in a recalcitrant species. Here the identified proteins in germination and seedling were related to carbohydrate metabolism function, amino acid metabolism, oxidation-reduction process, protein folding, RNA metabolism and stress response. Almost all proteins were previously described as phosphorylated proteins ²⁶. Among them, protein involved in glycolytic pathway showed changes in phosphorylation status, similar to that reported in rice germination ²⁷. This suggested that the phosphorylation/dephosphorylation could be the signal to activate and resume the most important metabolic pathway during germination, but large scale analysis is needed to have a global vision of the phosphoproteomics changes.

During the germination process, changes in expression profile at transcript, proteomics and metabolite levels occur. These changes are more significant in orthodox seeds, in contrast to *Q. ilex* seeds, in our experimental conditions, where the changes were minimal. Transcription and translation are far from having a linear and simple relationship and many reports indicated that the correlation between mRNA and protein abundances in the cell is notoriously poor²⁸. However, the data obtained in this study indicate that transcript level, at least in this experimental system, provide a good predictive value of protein abundance (Fig. 8.1).

The classical biochemical techniques combined with “omics” approaches (proteomics and transcriptomics) allowed deepen the understanding of key aspects of seed germination and seedling growth of *Q. ilex*, an orphan species which little is known at molecular levels. This new knowledge could be used in biotechnological project in this species.

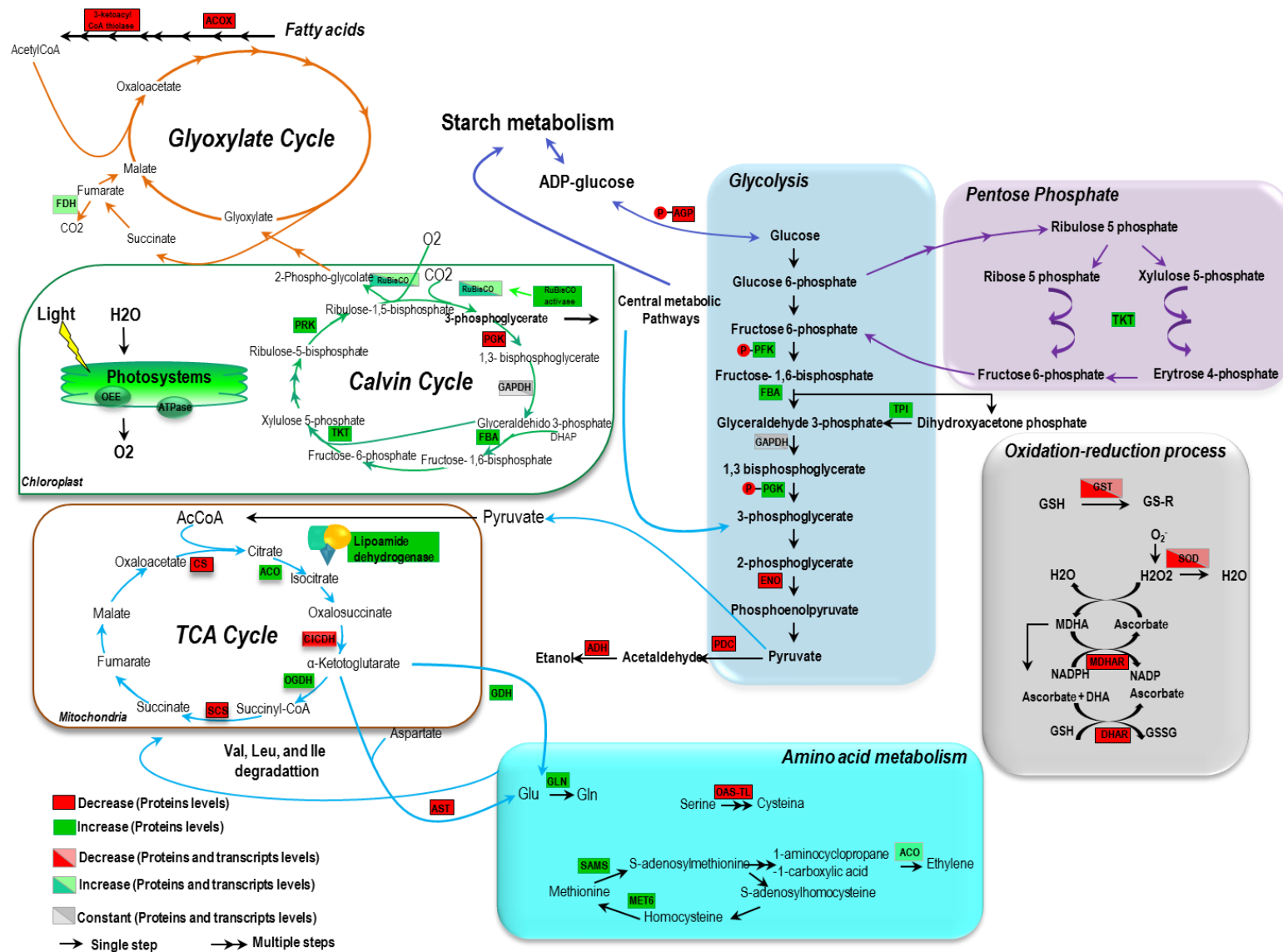


Figure 8.1. Changes in metabolic pathways during germination and seedling growth. The metabolic pathways were constructed based on the proteomic and gene expression data. The metabolic network was reconstructed based on the KEGG. Red indicated that protein or transcript decrease from S0 to S7 and SS-4; and green indicate an increase in the same direction. Box with two color indicated both transcript and proteins levels and box with one color indicate differences at protein or transcript level.

References

1. Pulido FJ, Díaz M, Hidalgo de Trucios SJ (2001) Size structure and regeneration of Spanish holm oak *Quercus ilex* forests and dehesas: effects of agroforestry use on their long-term sustainability. *Forest Ecology and Management* 146: 1-13
2. Gallego FJ, de Algaba AP, Fernandez-Escobar R (1999) Etiology of oak decline in Spain. *European Journal of Forest Pathology* 29: 17-27
3. He D, Han C, Yao J, Shen S, Yang P (2011) Constructing the metabolic and regulatory pathways in germinating rice seeds through proteomic approach. *PROTEOMICS* 11: 2693-2713
4. Nkang Ani (2002) Carbohydrate composition during seed development and germination in two sub-tropical rainforest tree species (*Erythrina caffra* and *Guilfoylia monostylis*). *Journal of Plant Physiology* 159: 473-483
5. Rajjou L, Duval M, Gallardo K, Catusse J, Bally J, Job C, Job D (2012) Seed Germination and Vigor. *Annual Review of Plant Biology* 63: 507-533
6. Kim ST, Wang Y, Kang SY, Kim SG, Rakwal R, Kim YC, Kang KY (2009) Developing rice embryo proteomics reveals essential role for embryonic proteins in regulation of seed germination. *Journal of Proteome Research* 8: 3598-3605
7. Liu Y, Shi L, Ye N, Liu R, Jia W, Zhang J (2009) Nitric oxide-induced rapid decrease of abscisic acid concentration is required in breaking seed dormancy in *Arabidopsis*. *New Phytologist* 183: 1030-1042
8. Yang R, Yang T, Zhang H, Qi Y, Xing Y, Zhang N, Li R, Weeda S, Ren S, Ouyang B, Guo Y-D (2014) Hormone profiling and transcription analysis reveal a major role of ABA in tomato salt tolerance. *Plant Physiology and Biochemistry* 77: 23-34
9. Chen S-Y, Chien C-T, Baskin JM, Baskin CC (2010) Storage behavior and changes in concentrations of abscisic acid and gibberellins during dormancy break and germination in seeds of *Phellodendron amurense* var. *wilsonii* (Rutaceae). *Tree Physiology* 30: 275-284
10. Weitbrecht K, Müller K, Leubner-Metzger G (2011) First off the mark: early seed germination. *Journal of Experimental Botany*
11. Prewein C, Endemann M, Reinöhl V, Salaj J, Sunderlikova V, Wilhelm E (2006) Physiological and morphological characteristics during development of pedunculate oak (*Quercus robur* L.) zygotic embryos. *Trees - Structure and Function* 20: 53-60
12. Farrant JM, Berjak P, Cutting JGM, Pammenter NW (1993) The role of plant growth regulators in the development and germination of the desiccation-sensitive (recalcitrant) seeds of *Avicennia marina*. *Seed Science Research* 3: 55-63
13. Nambara E, Okamoto M, Tatematsu K, Yano R, Seo M, Kamiya Y (2010) Abscisic acid and the control of seed dormancy and germination. *Seed Science Research* 20: 55-67
14. Voegelé A, Linkies A, Müller K, Leubner-Metzger G (2011) Members of the gibberellin receptor gene family *GID1* (*GIBBERELLIN INSENSITIVE DWARF1*) play distinct roles during *Lepidium sativum* and *Arabidopsis thaliana* seed germination. *Journal of Experimental Botany* 62: 5131-5147
15. Appleford NEJ, Lenton JR (1997) Hormonal regulation of α -amylase gene expression in germinating wheat (*Triticum aestivum*) grains. *Physiologia Plantarum* 100: 534-542
16. Yamaguchi S (2008) Gibberellin Metabolism and its Regulation. *Annu Rev Plant Biol* 59: 225-251
17. Miransari M, Smith DL (2014) Plant hormones and seed germination. *Environmental and Experimental Botany* 99: 110-121
18. Bromley JR, Warnes BJ, Newell CA, Thomson JC, James CM, Turnbull CG, Hanke DE (2014) A purine nucleoside phosphorylase in *Solanum tuberosum* L. (potato) with specificity for cytokinins contributes to the duration of tuber endodormancy. *Biochemical Journal* 458: 225-237
19. Liu Y, Wang L, Jiang S, Pan J, Cai G, Li D (2014) Group 5 LEA protein, ZmLEA5C, enhance tolerance to osmotic and low temperature stresses in transgenic tobacco and yeast. *Plant Physiology and Biochemistry* 84: 22-31

20. Kleinwächter M, Radwan A, Hara M, Selmar D (2014) Dehydrin expression in seeds: an issue of maturation drying. *Front Plant Sci* 5
21. Chen F, Nonogaki H, Bradford KJ (2002) A gibberellin-regulated xyloglucan endotransglycosylase gene is expressed in the endosperm cap during tomato seed germination. *Journal of Experimental Botany* 53: 215-223
22. Ishikawa A, Okamoto H, Iwasaki Y, Asahi T (2001) A deficiency of coproporphyrinogen III oxidase causes lesion formation in Arabidopsis. *The Plant Journal* 27: 89-99
23. Sturm A, Tang G-Q (1999) The sucrose-cleaving enzymes of plants are crucial for development, growth and carbon partitioning. *Trends in Plant Science* 4: 401-407
24. Wong P-F, Abubakar S (2005) Post-germination changes in *Hevea brasiliensis* seeds proteome. *Plant Science* 169: 303-311
25. Balbuena TS, Jo L, Pieruzzi FP, Dias LLC, Silveira V, Santa-Catarina C, Junqueira M, Thelen JJ, Shevchenko A, Floh EIS (2011) Differential proteome analysis of mature and germinated embryos of *Araucaria angustifolia*. *Phytochemistry* 72: 302-311
26. Gao J, Agrawal GK, Thelen JJ, Xu D (2009) P3DB: a plant protein phosphorylation database. *Nucleic Acids Research* 37: D960-D962
27. Han C, Wang K, Yang P (2014) Gel-based comparative phosphoproteomic analysis on rice embryo during germination. *Plant and Cell Physiology*
28. Maier T, Güell M, Serrano L (2009) Correlation of mRNA and protein in complex biological samples. *FEBS Letters* 583: 3966-3973

CHAPTER 9:

CONCLUSIONS

A multiple-level analysis transcriptional, translational and posttranslational combined with metabolites analysis to study germination and seedling growth in *Q. ilex* revealed some pathway that play an important factor in seed germination and seedling growth of this recalcitrant seeds.

The characterization of *Q. ilex* seeds revealed a triphasic uptake of water during germination, with a rapid initial uptake (phase I) followed by a plateau phase (phase II) and a further increase (phase III) as the embryo axis elongates and breaks the covering layers to complete germination.

The integrated transcriptomics, proteomics and metabolites analysis of *Q. ilex* seeds showed some of the intracellular physiological characteristics that are observed in orthodox seed which included accumulation of non-reducing carbohydrates (sucrose) and proteins that contribute to the intracellular vitrified state in seeds (transcripts and proteins of DHN3, and transcript of *Lea-5*); the anti-oxidative defence (*SOD*, *Gst*) and the preparation for the development of an adult plant (*RbcL*). But the Holm oak mature acorns showed an opposite profile found in mature orthodox seeds, low ABA content and high levels of GA in mature acorns. This could be one of the explanations of the precocious germination of *Q. ilex* seeds.

Holm oak mature acorns have the ability to maintain a partially active metabolism, with high level of glycolytic and mitochondrial respiratory enzymes, displayed at transcriptional, translational and posttranslational levels that could explain the recalcitrance of seeds. However, imbibition increased the respiratory rate, leading to an increase of the soluble carbohydrates. The proteomic changes observed in *Q. ilex* seed germination and seedling establishment affected mainly the proteins related to carbohydrate metabolism, amino acid metabolism and oxidative stress response, which also found at the transcriptional level. In addition, the phosphoproteomics analysis revealed that the glycolytic pathway was activated in germinated seeds comparing with mature seeds. The up-accumulation of proteins related to glycolysis in mature and the activation of this pathway in germinated seeds suggested that this process is essential for the energy supply and to provide molecules that serve as intermediate in other metabolic pathways during *Q. ilex* germination and seedling growth.

The results we presented here will help to increase the knowledge of the physiological changes that take place during *Q. ilex* seed germination, illustrate the importance of considering the behaviour of seeds for the afforestation projects and restoration programmes under the impending climate change in Mediterranean regions.

CHAPTER 10:

APPENDIXES

Available online at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/jprot

Technical note

Improving the quality of protein identification in non-model species. Characterization of *Quercus ilex* seed and *Pinus radiata* needle proteomes by using SEQUEST and custom databases[☆]



M. Cristina Romero-Rodríguez^{a,1,2}, Jesús Pascual^{b,2}, Luis Valledor^{c,d,*}, Jesús Jorrín-Novo^{a,**}

^aAgricultural and Plant Biochemistry and Proteomics Research Group, Dept. of Biochemistry and Molecular Biology, University of Córdoba, Spain

^bPlant Physiology, Faculty of Biology, Dept. of Organisms and Systems Biology, University of Oviedo, Spain

^cDept. of Biology & Centre for Environmental and Marine Studies, University of Aveiro, Aveiro, Portugal

^dGCRC, Adaption Biotechnologies, Academy of Sciences of the Czech Republic, Brno, Czech Republic

ARTICLE INFO

Available online 4 February 2014

Keywords:

Protein identification

Custom databases

Non-model species

SEQUEST

ESTs

NGS

ABSTRACT

Nowadays the most used pipeline for protein identification consists in the comparison of the MS/MS spectra to reference databases. Search algorithms compare obtained spectra to an *in silico* digestion of a sequence database to find exact matches. In this context, the database has a paramount importance and will determine in a great deal the number of identifications and its quality, being this especially relevant for non-model plant species. Using a single *Viridiplantae* database (NCBI, UniProt) and TAIR is not the best choice for non-model species since they are underrepresented in databases resulting in poor identification rates. We demonstrate how it is possible to improve the rate and quality of identifications in two orphan species, *Quercus ilex* and *Pinus radiata*, by using SEQUEST and a combination of public (*Viridiplantae* NCBI, UniProt) and a custom-built specific database which contained 593,294 and 455,096 peptide sequences (*Quercus* and *Pinus*, respectively). These databases were built after gathering and processing (trimming, contiging, 6-frame translation) publicly available RNA sequences, mostly ESTs and NGS reads. A total of 149 and 1533 proteins were identified from *Quercus* seeds and *Pinus* needles, representing a 3.1- or 1.5-fold increase in the number of protein identifications and scores compared to the use of a single database. Since this approach greatly improves the identification rate, and is not significantly more complicated or time consuming than other approaches, we recommend its routine use when working with non-model species.

[☆] This article is part of a Special Issue entitled: Proteomics of non-model organisms.

* Correspondence to: L. Valledor, GCRC, Adaption Biotechnologies, Academy of Sciences of the Czech Republic, Brno, Czech Republic.

** Corresponding author.

E-mail addresses: luis@valledor.info (L. Valledor), bf1jonoj@uco.es (J. Jorrín-Novo).

¹ Permanent address: Technological Multidisciplinary Research Centre, National University of Asunción, Paraguay.

² These authors contributed equally to this work.

Biological significance

In this work we demonstrate how the construction of a custom database (DB) gathering all available RNA sequences and its use in combination with *Viridiplantae* public DBs (NCBI, UniProt) significantly improve protein identification when working with non-model species. Protein identification rate and quality is higher to those obtained in routine procedures based on using only one database (commonly *Viridiplantae* from NCBI), as we demonstrated analyzing *Quercus* seeds and *Pine* needles. The proposed approach based on the building of a custom database is not difficult or time consuming, so we recommend its routine use when working with non-model species.

This article is part of a Special Issue entitled: Proteomics of non-model organisms.

© 2014 Elsevier B.V. All rights reserved.

Protein identification from tandem mass spectrometry (MS/MS) data is a central task in proteomics. The software tools that have been developed for peptide identification can be broadly divided into two categories: *de novo* sequencing (PEAKS [1], PepNovo [2], NovoHMM [3]) and database (DB) search by homology (the most popular being Mascot [4] and SEQUEST [5]).

Conceptually, and as it was coined and conceived, protein identification through a MS-based approach requires a sequenced and annotated genome of the species under investigation. In the case of plants, this is the exception better than the rule. Thus, the number of plant genomes sequenced and functionally annotated is of approximately 37 [6]; while the total number of cultivated plants amounts to 35,000 species, i.e. about 14% of higher plants [7]. Although not a crop, *Arabidopsis* (*Arabidopsis thaliana*) is the most popular model organism in classical genetics because it is small, shows a great natural variation, and has a short life cycle. The availability of its genome sequence [8] made *Arabidopsis* the plant paradigm for molecular biology and various “omics” approaches, including proteomics [9]. In ISI Web of Knowledge searching with the key words “proteom” and “*Arabidopsis thaliana*” returned 153 articles over the period 2000–2013, while 10 and 52 were obtained, respectively, with “proteome*” and “*Quercus**” or “*Pinus**”.

Cross-species identification is the only option for protein identification whenever a genome is poorly characterized. Search by homology, using a non-species-specific DB will tell us just which is the closest homologous gene product, and from that hypothesize or more commonly, speculate the protein function. Proteomic analyses of fully annotated species, for example *Arabidopsis* or rice, can achieve a high number of confident identifications with a standard homology search algorithm [6]. In contrast, using the same approach and databases for non-model species results in a lower number and confidence of identified proteins. This can be explained by the fact that either some species-specific proteins will not be present in DBs or homologous proteins will show small evolutionary differences in its sequences. In consequence the main target for homologous protein identification is the active site of the protein, which usually is highly conserved but not exact. In this scenario the use of tight algorithms like SEQUEST, in which the homology should be of 100%, is not the best approach; algorithms that tolerate amino acid substitution, such as Paragon [10], were developed to

allow non-perfect matched by evaluating punctual amino acid substitutions. This strategy increased the number of identified proteins in species like *Quercus* [11,12] and *Pine* [13,14]; however, this kind of identification is less confident than using a tight algorithm like SEQUEST, having strong implications for peptide to protein assignment and label-free quantitative purposes [15].

Immediately a question rises: How to proceed in the case of non-model plants? Nowadays the best choice would be an initial transcriptome analysis by using deep sequencing to generate our species-specific database to compare the proteins to. However this approach is complex, expensive, and requires skilled personnel, making it unreachable for most of the laboratories working on non-model species. A good alternative is to take advantage of the specific protein and RNA sequences, most of them in the form of ESTs that are available in public repositories [15]. This strategy has only been reported in a very limited number of cases [16,17]. Champagne and Boutry [6] make a revision of the recent literature in the plant field (January 2010 to August 2012) to assess which DBs were most commonly used for protein DB searches. They found that 94% of the research articles surveyed used the NCBI protein DB, either alone or combined with their DB, and only 15% used specific EST DB translated in silico.

We have previously published a number of papers [13,14,18,19], in which the proteome of *Q. ilex* and *P. radiata* was analyzed by gel-based (1- and 2-DE in combination with MALDI-TOF/TOF) and gel-free (nLC-LTQ Orbitrap MS) approaches. The identification of proteins from MS/MS spectra has been performed using different algorithms (Mascot, Paragon) against plant protein sequence DB: UniProtKB and/or NCBI *Viridiplantae*. For *Q. ilex* seeds, using gel based proteomics and Paragon algorithm only 60% of the 23 bands were identified and 28% of the 2-DE spots were identified [18]. The identification of leaf proteins under abiotic or biotic stresses resulted in a higher number of identifications, 57% [20] and 45% [21] respectively. On the other hand, the same approach resulted in the identification of the 67% of the excised spots when studying *P. radiata* needles [13]. The low percentage of identification can be related to different issues related to databases: the number of full-length proteins available in database is very limited (97 and 2568 in the case of *Q. ilex* and *P. radiata*, September 2013); full length cDNAs also follow the same trend; non-model species are sometimes

excluded from non-redundant databases; deep sequencing reads and ESTs are not processed to enrich protein/peptide databases.

The objective of this work was to evaluate the impact of using a custom genus-specific DB for protein identification. In this manuscript we demonstrate how it is possible and quality in two proteomic studies by the construction of an in-house protein DB from publicly available ESTs, and using a combined DB search.

In a first step two specific DBs were engineered, *Quercus* and Conifers DB. We included 669,011 ESTs from five *Quercus* species (*Q. alba*, *Q. petraea*, *Q. robur*, *Q. rubra*, and *Q. suber*) obtained from NCBI and the sequences available in the Oak Gene Index DB, composed of ESTs of several *Quercus* species and integrated research data from international *Quercus* EST sequencing; and 1099 protein sequences from all *Quercus* species available in Dendrome Project. Conifers DB was constructed using a total of 624,835 EST sequences from NCBI of three *Pinus* species (*P. radiata*, *P. pinaster*, and *P. taeda*) and EST collection from Dendrome Project and the Gene Index for *Pinus* and *Picea* and 85,747 protein sequences from NCBI for *Pinus* and *Picea* genera (Table 1).

DBs were constructed as reported in [17], with minor modification. The workflow for building the custom libraries is provided in Fig. 1. Briefly: i) Trimming: sequences were matched against the UniVec DB (<ftp://ftp.ncbi.nih.gov/pub/UniVec>) and regions contaminated with vectors or with a high number of indeterminations were removed. ii) Contig sequences were then independently assembled using SeqMan

[22]. All contigs and singletons obtained were compared sequentially using local BLASTx (NCBI) [23] to the following protein DBs: TAIR10_pep [*A. thaliana*, ftp://ftp.arabidopsis.org/home/tair/Proteins/TAIR10_protein_lists/, Release 10 (August 23, 2011)], UniProtKB/Swiss-Prot reviewed (<ftp://ftp.uniprot.org/pub/databases/>), Egrandis_201_protein [*Eucalyptus grandis*, <ftp://ftp.jgi-psf.org/pub/compngen/phytozome/v9.0>, Release 201 (December 12, 2012)], *Ptrichocarpa_210_protein* [*Populus trichocarpa*, <ftp://ftp.jgi-psf.org/pub/compngen/phytozome/v9.0>, Release 201 (December 12, 2012)] and all plant protein sequences in NCBI_nr. We considered a good homology those matches that showed e-values lower than 10⁻⁵⁰. Sequence annotation was performed based on hits obtained in all protein DBs in following priority order: TAIR10_pep > UniProtKB/Swiss-Prot > Egrandis_201_protein > *Ptrichocarpa_210_protein* > NCBI_nr annotations. The Gene Index data for two species were not subjected to assembly and BLASTx step because they were previously assembled and annotated by Computational Biology and Functional Genomics Laboratory. iii) Since almost all of the contigs were derived from transcriptomic analyses we assumed that they are translated in its full length. We used EMBOSS [24] to perform a six-frame translation and extract the longest *in silico* translation products of both strands. In this step the Gene Index data were included. iv) Then these translation products were BLASTp searched against the mentioned protein DBs, with the same parameters used in BLASTx. v) The BLASTp and BLASTx results obtained with each DB were compared, and translated product with the same result in BLASTx and BLASTp was used to annotate the sequence. vi) All possible polypeptide sequences

Table 1 – Sequences employed in this work. Number of ESTs and protein sequences, contigs, singletons, and in silico translation products. Database compiler and download source are also indicated.

Species	ESTs	Contigs	Singletons	Contig and singleton sequences	In silico translation products	Proteins	Compiler	Download URL
<i>Q. suber</i>	6660	1485	1058	2543	5086		NCBI (National Center for Biotechnology Information)	ftp://ncbi.nlm.nih.gov/blast/db/est
<i>Q. petraea</i>	58,230	8981	11,380	20,361	40,722			
<i>Q. robur</i>	81,617	11,046	16,209	27,255	54,510			
<i>Q. alba</i>	203,206	20,748	68,622	89,370	178,740			
<i>Q. rubra</i>	277,154	27,781	87,193	114,974	229,948			
<i>Quercus</i> spp.	42,144	19,674	22,470	42,144	84,288	1099	Fagaceae Genomics Web, The National Science Foundation Computational Biology and Functional Laboratory “The Gene Index Project” Dendrome Project	http://www.fagaceae.org/ http://compbio.dfci.harvard.edu/tgi/plant.html http://dendrome.ucdavis.edu/
Total	669,011	89,715	206,932	296,647	593,294	1099		
<i>P. radiata</i>	13,431	1964	7009	8973	17,946		Dendrome Project	http://dendrome.ucdavis.edu/
<i>P. pinaster</i>	48,839	7399	18,586	25,985	51,970		Dendrome Project	http://dendrome.ucdavis.edu/
<i>P. taeda</i>	405,830	46,468	43,617	90,085	180,170		Dendrome Project	http://dendrome.ucdavis.edu/
<i>Pinus</i> spp.	77,326	44,858	32,468	77,326	99,026	60,321	Computacional Biology and Functional Laboratory “The Gene Index Project” (ESTs) and NCBI (Proteins)	http://compbio.dfci.harvard.edu/tgi/ , ftp://ncbi.nlm.nih.gov/blast/db/est
<i>Picea</i> spp.	79,409	44,517	34,892	79,409	105,984	25,426	Computacional Biology and Functional Laboratory “The Gene Index Project” (ESTs) and NCBI (Proteins)	http://compbio.dfci.harvard.edu/tgi/ , ftp://ncbi.nlm.nih.gov/blast/db/est
Total	624,835	145,206	136,572	281,778	455,096	85,747		

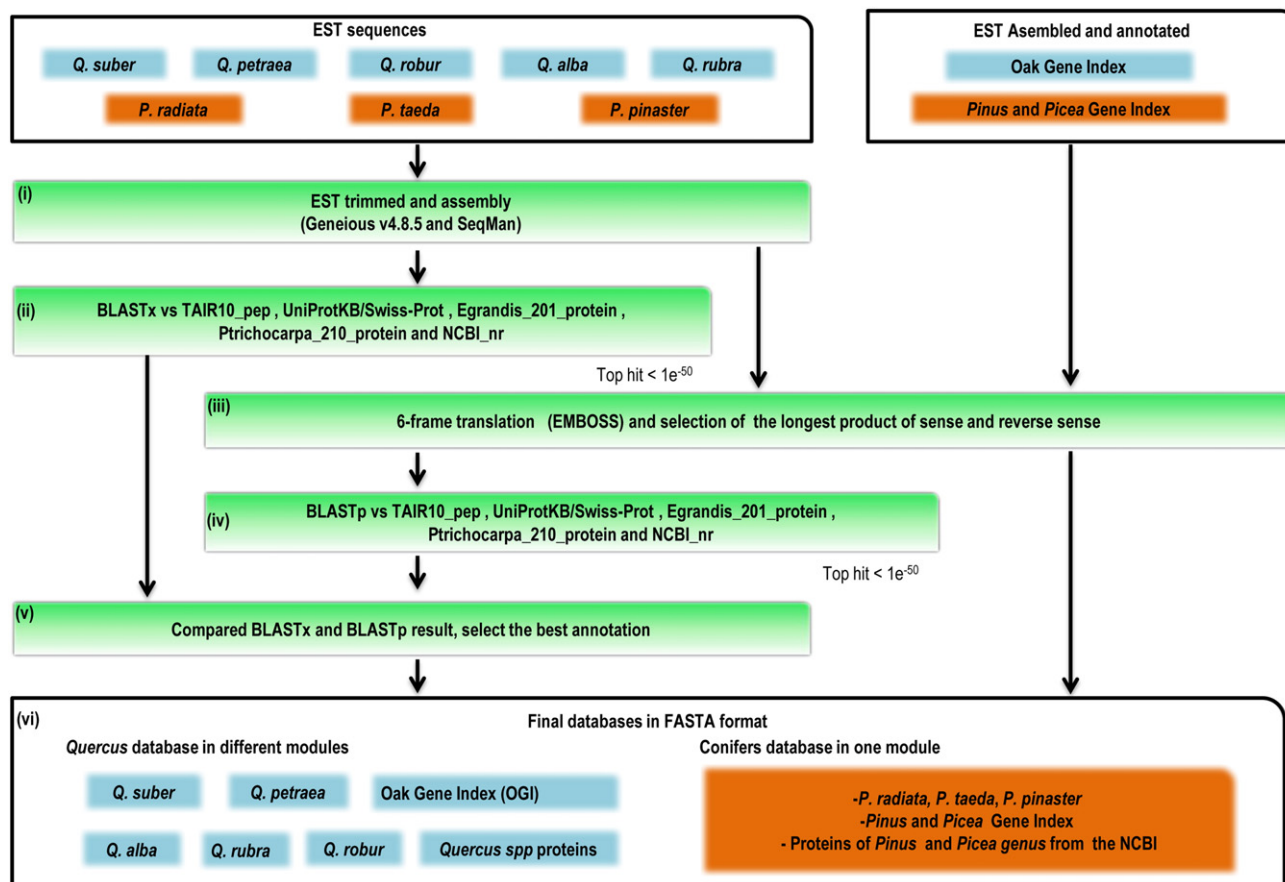


Fig. 1 – Flowchart of Quercus and Conifers DB construction. All the downloaded ESTs were trimmed using as reference UniVec vector DB to remove possible vector contaminations and using the software Geneious (Biomatters Inc.). After that, they were assembled into contigs using the software SeqMan, part of the DNASTar package (Lasergene). Contigs and singletons were annotated by BLASTx against TAIR10_pep, UniProtKB/Swiss-Prot, Egrandis_201_protein, Ptrichocarpa_210_protein, and NCBI_nr databases. At the same time, they were 6-frame translated using the EMBOSS tools (EBI) and the longest products (sense and antisense strands) were annotated employing the DBs described before. The better hits, based on e-values, were kept for naming the accessions of the DB. All protein sequences available in the Dendrome project and NCBI were merged to corresponding database.

obtained in the previous step, unannotated and annotated, were compiled in FASTA files, keeping each species separately for *Quercus*, considered as module; *Quercus* (1099) protein from Dendrome Project was included such as other modules; in Conifers DB all polypeptide sequences obtained were compiled in one FASTA file, including *Pinus* (60,321) and *Picea* (25,426) genus proteins from NCBI DB.

In the *Quercus* EST assembly 89,715 contigs and 206,932 singletons were obtained that were *in silico* translated into 593,294 peptides. In Conifers, a total of 145,206 contigs and 136,572 singletons were obtained which were *in silico* translated into 455,096 peptides (Table 1). Full length *Quercus* and *Pine* proteins were added to corresponding specific DB.

Custom DBs were tested against the most common DBs used in proteomic analysis of plant UniProtKB/Swiss-Prot, UniProtKB/TrEMBL and NCBI with *Viridiplantae* taxonomic restriction (downloaded in September 2013) by using SEQUEST algorithm and nLC-Orbitrap runs from *Q. ilex* seeds and *P. radiata* needles available in our laboratory. Proteins of *Q. ilex*

seed were extracted and processed as described by Romero-Rodríguez et al. [25] and that of *P. radiata* needles as described by Valledor and Weckwerth [26]. The LC-MS analysis of both protein extracts were performed according to Valledor and Weckwerth [26].

Spectra were processed using the SEQUEST algorithm available in Proteome Discoverer© 1.4 (Thermo-Scientific, USA) and the four described DBs (*Quercus*/Conifers and other three public DBs). The following settings were used: precursor mass tolerance was set to 10 ppm and fragment ion mass tolerance to 0.8 Da. Only charge states +2 or greater were used. Identification confidence was set to a 5% FDR and the variable modifications were set to: oxidation of methionine and to fixed modifications was set to carbamidomethyl cysteine formation. A maximum of two missed cleavages were set for all searches. The threshold used for considering confident protein identification was having at least one unique peptide of an X-correlation value greater than the charge state +0.25 [15].

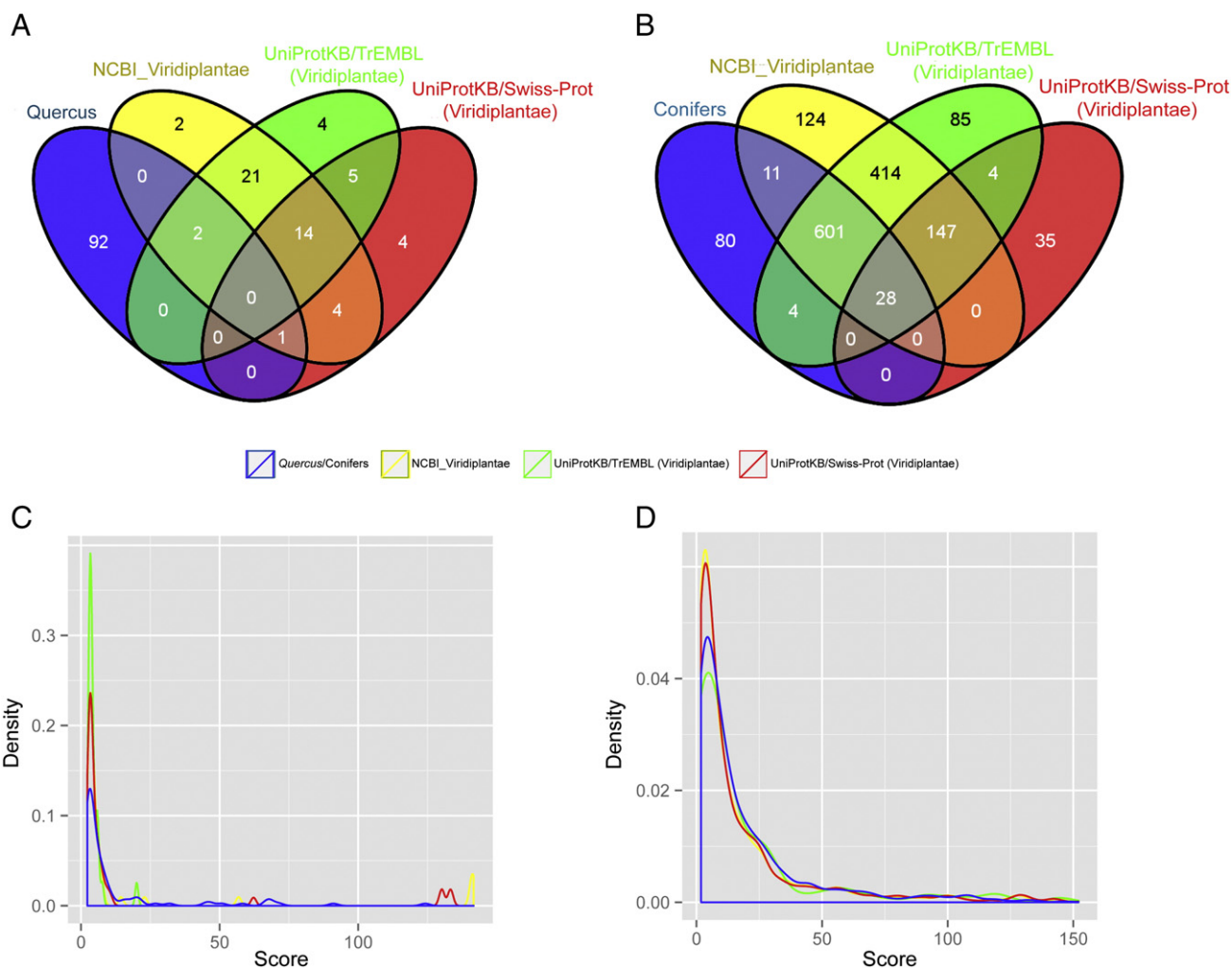


Fig. 2 – Venn diagrams representing the number of identified proteins in *Q. ilex* seeds (A) and *P. radiata* needles (B); diagrams were plot using Venny (<http://bioinfogp.cnb.csic.es/tools/venny/>). Density plots of the scores of the identified proteins in *Quercus ilex* seed (C), and *Pinus radiata* needles (D) by using SEQUEST algorithm and the four indicated databases. Blue lines indicate the developed databases, corresponding to *Quercus* or *Conifers* databases. These DBs showed a lower density in the peak of protein scores between 2.25 and 4, being the curve shifted to the right, indicating that the scores obtained employing these databases are, in general, higher than those obtained employing NCBI_Viridiplantae, UniProtKB/TrEMBL (Viridiplantae) and UniProtKB/Swiss-Prot (Viridiplantae). Areas under each curve are equal and normalized to arbitrary value 1.

The result files were merged and aligned into a single file, one per species, by using Proteome Discoverer® to perform a direct comparison between DBs. This analysis showed that most of the proteins were identified in all databases (Fig. 2A and B); however, each DB provided a unique set of identified proteins. The aim of this study was to demonstrate the importance of genus-specific DB to improve protein identification confidence, in achieving better results in orphan species researches. In *Q. ilex* seeds and *P. radiata* needles, we obtained a total of 149 and 1533 proteins, respectively, using a combined search against the four DBs; of which 63% and 47% of *Q. ilex* and *P. radiata* proteins, respectively, were identified in the genus-specific DB. Using only one Viridiplantae protein DB, for example UniProtKB/Swiss-Prot, the number of protein identified would be just 19% in *Quercus* and 14% for *Pinus*, which proves the usefulness of our approach in the

maximization of identification rates, since an important percentage of protein would not be identified if we have used only one of the DB instead of the combined approach using the four of them. On the other hand, in *Q. ilex* seeds we identified 95 proteins with *Quercus* DB (Fig. 2A), of which 92 were identified only with this genus-specific DB. In *P. radiata* needles, 724 proteins were identified using *Conifers* DB (Fig. 2B), of which 80 were specific for this DB.

Despite the fact that one of the main objectives of this work was to increase the number of identified proteins, we further investigated how the DB improved the quality of protein identification, in terms of SEQUEST score. Fig. 2C and D shows a density plot of the protein identification scores of each DB. The distribution of protein scores employing genus-specific and Swiss-Prot DB showed a lower density in the peaks of low scores, and curves are displaced toward

higher scores; in contrast to NCBI and TrEMBL, which show a high density at low score, both for *Q. ilex* seeds and *P. radiata* needles.

Despite increasing the number of identified proteins in both species and tissues, the result was different between *Q. ilex* and *P. radiata* datasets. Needles, a green photosynthetic organ, showed a great overlap of protein identification between DBs. This can be explained by the fact that photosynthetic proteins are mostly conserved, and in consequence *Viridiplantae* DB would perform adequately. However we cannot forget that the use of our strategy increased by 80 the number of identified proteins compared to *Viridiplantae* (NCBI, Swiss-Prot and TrEMBL). On the other hand, homologous proteins in different species are rarely identical; in *Q. ilex* transcriptomic and proteomic analyses of seed development are scarce, so the presence of related sequences in DB is not abundant, and also the phylogenetic distance between species increases significantly the number of sequence variations of storage proteins [27,28]. All these contribute to a minimal overlap between *Quercus* DB and other *Viridiplantae* DBs.

One of the main purposes of every mass-spectrometry proteomic analysis is to identify as many proteins as possible. However, the quality of the identifications is also an important factor, to ideally obtain as much high-quality identification as possible. In this way, we showed that with a genus-specific DB we maximized both factors. As we have already mentioned, by combining our selected DB, we identified more proteins and at the same time a higher number of high-quality identifications, as it can be seen in the density plot of score (Fig. 2C and D), where the density curves to *Quercus* and Conifers DB show a lower density in the peak of the score between 2.25 and 4, to NCBI and TrEMBL DB show a higher density in the peak of the low score, attributing to these databases poor quality in protein identification.

These results reinforce one of the protein identification basics that is usually forgotten: protein identification based on homology requires a perfect match between spectra and database to be successful, depending on the score on the length and composition of their amino acid composition. The employ of custom databases increased the confidence of identified proteins by “rescuing” more MS/MS spectra than would not have counterpart in nonspecific DBs, thus reducing the undermining of protein identification when using standard protein identification pipelines [29].

Our results showed the importance of specific DB in proteomic analysis by homology search in orphan species. Due to the lack of annotated reference genome in *Q. ilex* and *P. radiata*, a combination of customized and public DBs greatly improves the protein identification rate in terms of quantity and quality.

Acknowledgments

MCRR, JP, and LV were respectively supported by the predoctoral scholarship program Itaipú Binacional—Paraguay (Paraguay Government), FPU predoctoral program (Ministry of Education, Spain), and FCT postdoctoral fellowship (Portuguese Government).

REFERENCES

- [1] Ma B, Zhang K, Hendrie C, Liang C, Li M, Doherty-Kirby A, et al. PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Commun Mass Spectrom* 2003;17:2337–42.
- [2] Frank A, Pevzner P. PepNovo: de novo peptide sequencing via probabilistic network modeling. *Anal Chem* 2005;77:964–73.
- [3] Fischer B, Roth V, Roos F, Grossmann J, Baginsky S, Widmayer P, et al. NovoHMM: a hidden Markov model for de novo peptide sequencing. *Anal Chem* 2005;77:7265–73.
- [4] Perkins DN, Pappin DJC, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 1999;20:3551–67.
- [5] Eng J, McCormack A, Yates J. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* 1994;5:976–89.
- [6] Champagne A, Boutry M. Proteomics of nonmodel plant species. *Proteomics* 2013;13:663–73.
- [7] Khoshbakht K, Hammer K. How many plant species are cultivated? *Genet Resour Crop Evol* 2008;55:925–8.
- [8] The *Arabidopsis* genome initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 2000;408:796–815.
- [9] Koornneef M, Meinke D. The development of *Arabidopsis* as a model plant. *Plant J* 2010;61:909–21.
- [10] Shilov IV, Seymour SL, Patel AA, Loboda A, Tang WH, Keating SP, et al. The paragon algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra. *Mol Cell Proteomics* 2007;6:1638–55.
- [11] Jorge I, Navarro RM, Lenz C, Ariza D, Jorrín-Novo JV. Variation in the Holm oak leaf proteome at different plant developmental stages, between provenances and in response to drought stress. *Proteomics* 2006;6:S207–14.
- [12] Jorge I, Navarro RM, Lenz C, Ariza D, Porras C, Jorrín J. The Holm oak leaf proteome: analytical and biological variability in the protein expression level assessed by 2-DE and protein identification tandem mass spectrometry de novo sequencing and sequence similarity searching. *Proteomics* 2005;5:222–34.
- [13] Valledor L, Castillejo MA, Lenz C, Rodríguez R, Cañal MJ, Jorrín J. Proteomic analysis of *Pinus radiata* needles: 2-DE map and protein identification by LC/MS/MS and substitution-tolerant database searching. *J Proteome Res* 2008;7:2616–31.
- [14] Valledor L, Jorrín-Novo JV, Rodríguez JL, Lenz C, Meijón M, Cañal MJ. Combined proteomic and transcriptomic analysis identifies differentially expressed pathways associated to *Pinus radiata* needle maturation. *J Proteome Res* 2010;9:3954–79.
- [15] Valledor L, Recuenco-Munoz L, Egelhofer V, Wienkoop S, Weckwerth W. The different proteomes of *Chlamydomonas reinhardtii*. *J Proteomics* 2012;75:5883–7.
- [16] Lei Z, Dai X, Watson BS, Zhao PX, Sumner LW. A legume specific protein database (LegProt) improves the number of identified peptides, confidence scores and overall protein identification success rates for legume proteomics. *Phytochemistry* 2011;72:1020–7.
- [17] Dustin L, Mack Y, Jörg B. Spruce proteome DB: a resource for conifer proteomics research. *Tree Genet Genomes* 2009;5:723–7.
- [18] Valero Galván J, Valledor L, Cerrillo RMN, Pelegrín EG, Jorrín-Novo JV. Studies of variability in Holm oak (*Quercus ilex* subsp. *ballota* [Desf.] Samp.) through acorn protein profile analysis. *J Proteomics* 2011;74:1244–55.

- [19] Valero Galván J, Valledor L, González Fernandez R, Navarro Cerrillo RM, Jorrín-Novo JV. Proteomic analysis of Holm oak (*Quercus ilex* subsp. *ballota* [Desf.] Samp.) pollen. *J Proteomics* 2012;75:2736–44.
- [20] Echevarría-Zomeño S, Ariza D, Jorge I, Lenz C, Del Campo A, Jorrín J, et al. Changes in the protein profile of *Quercus ilex* leaves in response to drought stress and recovery. *J Plant Physiol* 2009;166:233–45.
- [21] Sghaier-Hammami B, Valero-Galván J, Romero-Rodríguez MC, Navarro-Cerrillo RM, Abdelly C, Jorrín-Novo J. Physiological and proteomics analyses of Holm oak (*Quercus ilex* subsp. *ballota* [Desf.] Samp.) responses to *Phytophthora cinnamomi*. *Plant Physiol Biochem* 2013;71:191–202.
- [22] Swindell S, Plasterer T. SEQMAN. In: Swindell S, editor. Sequence data analysis guidebook. New York: Springer; 1997. p. 75–89.
- [23] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215:403–10.
- [24] Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* 2000;16:276–7.
- [25] Romero-Rodríguez MC, Maldonado-Alconada A, Valledor L, Jorrin-Novo J. Back to Osborne. Sequential protein extraction and LC–MS analysis for the characterization of the Holm oak seed proteome. In: Jorrín-Novo JV, et al, editors. *Plant proteomics*. New York: Humana Press; 2014. p. 379–89.
- [26] Valledor L, Weckwerth W. An improved detergent-compatible gel-fractionation LC-LTQ-Orbitrap-MS workflow for plant and microbial proteomics. In: Jorrin-Novo JV, editor. *Plant Proteomics*. New York: Humana Press; 2014. p. 347–58.
- [27] Shewry PR, Napier JA, Tatham AS. Seed storage proteins: structures and biosynthesis. *Plant Cell* 1995;7:945–56.
- [28] Shutov AD, Bäumlein H, Blattner FR, Müntz K. Storage and mobilization as antagonistic functional constraints on seed storage globulin evolution. *J Exp Bot* 2003;54:1645–54.
- [29] Liska AJ, Shevchenko A. Expanding the organismal scope of proteomics: cross-species protein identification by mass spectrometry and its implications. *Proteomics* 2003;3:19–28.

10.2. Appendix 2: Contributions

1. Publications

- a. Jorrín-Novo JV, Pascual J, Lucas RS, **Romero-Rodríguez MC**, Ortega MR, Lenz C, Valledor L (2015) Fourteen years of plant proteomics reflected in “Proteomics”: Moving from model species and 2-DE based approaches to orphan species and gel-free platforms. PROTEOMICS: n/a-n/a
- b. **Romero-Rodríguez MC**, Pascual J, Valledor L, Jorrin-Novo J (2014) Improving the quality of protein identification in non-model species. Characterization of *Quercus ilex* seed and *Pinus radiata* needle proteomes by using SEQUEST and custom databases. Journal of Proteomics 105: 85-91
- c. **Romero-Rodríguez MC**, Maldonado-Alconada A, Valledor L, Jorrin-Novo J (2014) Back to Osborne. Sequential Protein Extraction and LC-MS Analysis for the Characterization of the Holm Oak Seed Proteome. In: Jorrin-Novo JV, Komatsu S, Weckwerth W, Wienkoop S (eds) Plant Proteomics. Humana Press, pp 379-389
- d. Valledor L, **Romero-Rodríguez MC**, Jorrín-Novo J (2014) Standardization of data processing and statistical analysis in comparative plant proteomics experiment. In: Jorrin-Novo JV, Komatsu S, Weckwerth W, Wienkoop S (eds) Plant Proteomics. Humana Press, pp 51-60

2. Presentation in conferences

- a. **M Cristina Romero-Rodríguez**; Luis Valledor; Christof Lenz, Henning Urlaub and Jesús V. Jorrín-Novo. Proteomics workflows and protocols for the study of orphan species in the study of a recalcitrant and orphan forest plant species: *Quercus ilex*. 13th Human Proteome Organization World Congress. 5 al 8 de octubre de 2014. Madrid, España.
- b. Lyudmila Simova-Stoilova, **M Cristina Romero-Rodríguez**, Jose Valero-Galvan, Rosa Sanchez, Jesús Jorrín-Novo. The nutraceutical potential of *Quercus ilex* L. revealed by proteomic analysis of acorns and roots. impact of the drought stress. 13th Human Proteome Organization World Congress. 5 al 8 de octubre de 2014. Madrid, España.
- c. Lyudmila Petrova Simova-Stoilova, **M. Cristina Romero-Rodríguez**, Jesús Jorrín-Novo. 2-DE Proteomics response of *Quercus ilex* roots to drought stress at early developmental stage. 1st INPPO (International Plant Proteomics Organization) World Congress on Plant Proteomics: Methodology to Biology. 31 de agosto al 4 de setiembre del 2014. Hamburgo, Alemania.
- d. **Ma Cristina Romero Rodríguez**. Aproximaciones -ómicas al estudio de la germinación de semillas de especies recalcitrantes: el caso de la encina (*Quercus ilex* subsp. ballota). Plan de Investigación de tesis. I Congreso Científico de Investigadores en Formación en Agroalimentación. II Congreso Científico de Investigadores en Formación de la Universidad de Córdoba. Córdoba, España, 8 y 9 de mayo 2012 (Presentación oral).
- e. **María Cristina Romero Rodríguez**, Jesús Jorrín-Novo, Ana M. Maldonado-Alconada. “Descripción del proteoma de bellota de la encina (*Quercus ilex*) mediante extracción secuencial de proteínas y análisis por LC-MS”. III Jornadas Bienales de Jóvenes Investigadores en Proteómica. Santiago de Compostela. (Presentación de Póster).

