VARIABILIDAD POBLACIONAL EN ENCINA (*Quercus ilex* subsp. *ballota* (Desf.) Samp.): MORFOMETRÍA, ESPECTROSCOPÍA DE INFRARROJO CERCANO Y PROTEÓMICA

José Valero Galván

Tesis Doctoral

MAYO 2012

TÍTULO: *Variabilidad poblacional en encina (quercus ilex subsp. Ballota (Desf.) Samp.): morfometría, espectroscopía de infrarrojo cercano y proteómica*

AUTOR: *José Valero Galván*

Campus de Rabanales Ctra. Nacional IV, Km. 396 A
14071 Córdoba

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

ETSIAM

GRUPO DE INVESTIGACIÓN DE BIOQUÍMICA Y PROTEÓMICA VEGETAL Y AGROFORESTAL

DEPARTAMENTO DE BIOQUÍMICA Y BIOLOGÍA MOLECULAR

UNIVERSIDAD DE CÓRDOBA

VARIABILIDAD POBLACIONAL EN ENCINA (*Quercus ilex* subsp. *ballota* (Desf.) Samp.): MORFOMETRÍA, ESPECTROSCOPÍA DE INFRARROJO CERCANO Y PROTEÓMICA

JOSÉ VALERO GALVÁN

TESIS DOCTORAL

MAYO DE 2012

TÍTULO DE LA TESIS:

Variabilidad poblacional en encina (Quercus iles subsp. ballota (desf.) samp.): morfometría, espectroscopía de infrarrojo cercano y proteómica.

DOCTORANDO/A:

José Valero Galván

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

(se hará mención a la evolución y desarrollo de la tesis, así como a trabajos y publicaciones derivados de la misma).

El documento presentado por D. José Valero Galván, y que lleva por título "Variabilidad poblacional en encina (*Quercus ilex* subsp. *ballota* (desf.) samp.): morfometría, espectroscopía de infrarrojo cercano y proteómica" corresponde a su trabajo de Tesis Doctoral. Ha ido dirigido a la caracterización de procedencias andaluzas de encina utilizando técnicas novedosas como son la proteómica y la de espectroscopía NIRS. Supone una aportación importante al campo de las ciencias forestales y, más concretamente, al de especies de interés para Andalucía, como es el caso de la encina. Cumple con los requisitos científicos y académicos exigidos para su presentación y defensa. Prueba de ello son las tres publicaciones derivadas del trabajo en revistas de alto IF (European Journal of Forest Research, IF 1,942 y Journal of Proteomics, IF 5,074). Los tres capítulos en los que se ha organizado la tesis corresponden a cada una de las publicaciones. Se ha caracterizado la composición química de bellotas mediante la tecnología NIRS y, por primera vez se describe el proteoma de bellotas y polen. Se ha podido catalogar procedencias y establecer agrupaciones y distancias entre ellas. Desde un punto de vista práctico, se sientan las bases para la selección de individuos y procedencias desde una óptica científica (molecular) más que empírica (fenotípica).

Por todo ello, se autoriza la presentación de la tesis doctoral.

Córdoba, 01 de mayo de 2012

Firma del/de los director/es

Fdo.: Rafael M. Navarro Cerrillo Fdo.: Jesús V. Jorrín Novo

Financiación personal y de investigación

Esta tesis doctoral ha sido posible gracias a la concesión de una beca del Programa Alßan, de la Unión Europea para América Latina [I06D00010MX] y ha estado incluida en el siguiente proyecto:

1. "Variabilidad, catalogación, respuesta a estreses y propagación clonal de encina (*Q. ilex*.)" (AGL2009-12243-C02-02). Proyecto Financiado por Ministerio de Ciencia e Innovación.

Agradecimientos

Quiero dar mis agradecimientos a todas aquellas personas que han contribuido a la realización de ésta tesis doctoral:

- A mis directores de tesis, Jesús V. Jorrín Novo y Rafael Mª Navarro Cerrillo, por la paciencia y dedicación que me han mostrado en todo este tiempo de desarrollo de la tesis, y por haberme abierto las puertas de sus respectivos grupos de investigación.
- A toda mi familia, en especial, a mis padres y a mis hermanos, que con todo su apoyo han contribuido a formarme como persona.
- A Raquel González Fernández por todo el apoyo que me ha dado.
- A mis compañeros del Máster (Carlos, Sira, Jaouad, Tito, Tania) que ahora son mis amigos.
- A todas aquellas personas que me apoyaron en la realización de mi tesis en especial a David Ariza, Ana María Maldonado, María del Carmen Molina, Inmaculada Redondo, Luis Valledor, Besma Sghaier-Hammami, Juan del Olmo, Augusto Gómez Cabrera, Antonio López y Manolo.
- A la Universidad de Córdoba.

Summary

Holm oak (*Quercus ilex* subsp. *ballota* [Desf.] Samp.) is the dominant tree species in natural forest ecosystems over large areas of the Western Mediterranean Basin. In Spain, it is widely used for conservation and forestry practices, covering an approximate surface of 2,039,563 ha, with estimated values of over 120 million of Euros in fruit acorn production per year. The Holm oak acorns are a major component in the feeding systems of many Mediterranean wild and livestock species. Moreover, it is also the basic feed ingredient for domestically bred high quality meat pigs. Nowadays, forest restoration and reforestation are high priority objectives, and Holm oak is acquiring more interest for Mediterranean forestry. However, in the *"dehesas"* of Mediterranean areas of Western and Southwestern Spain, where grasslands reach between 70 and 90% cover, tree regeneration is very low because of an inappropriate livestock management, together with a high acorn predation by a variety of vertebrate herbivores (i.e., cattle, deer, wild boar, mice and rabbits), water limitations, and effect of the decline syndrome. Those factors are considered the major limitations for the seedling establishment, and subsequent growth of tree stands, and plant distribution Holm oak ecosystems. Criteria for tree selection for restoration programs are frequently the production of big and abundant acorns per individual tree. However, in order to create systems which are able to self-regenerate, it is important to take into account the seedling traits related to abiotic (drought) and biotic (pest and diseases) tolerance. Our group is carrying out a functional genomics approach – transcriptomics and proteomics- to characterize variability among Holm oak populations and its response to biotic and abiotic stresses.

Within this general project, the objectives pursued in this PhD thesis were:

- Study of Holm oak population variability based on the morphometry and chemical composition of the acorn.
- Study of variability in Holm oak through the acorn protein profile analysis.
- Study of Holm oak pollen by proteomic analysis.

We have studied in this PhD thesis the natural variability in Holm oak by analyzing acorn morphometry and chemical composition in 13 populations from the Andalusia region (southern Spain). The acorn morphometry showed statistic significant differences between the populations with a higher acorn weight, length, and diameter in the Northern provenances (TJA, GSE, CTH, CHU, APS, and PCO), while the Southern provenances (RMA, SAA, CCO, BCA, and SCA) showed the lower value that even decreased with the altitude. The acorn chemical composition was analyzed by near-infrared spectrometry (NIRS), and also showed statistic significant differences between populations. Morphometry and chemical composition data were subjected to cluster analysis, and two major clusters were detected. The cluster analysis showed that Northern populations (CTH, TJA, SCA, CHU, PCO, GSE, and APS) showed a tendency to present values higher in acorn weight (5.4 g vs. 3.2 g); acorn length (37.2 mm vs. 28.7 mm); acorn diameter (16.1 mm vs. 13.7 mm); ash (1.8% vs. 1.5%); protein (4.8% vs. 4.3%); fat (12.5% vs. 9.9%); and oleic acid (66.9 vs. 66.4%). On the contrary, Southern populations (SSA, VJA, CCO, BCA, RG, and RMA) showed a tendency to present values lowest in sugar (10.2% vs. 11.2%); energy (20,076.7 kJ/kg vs. 20,149.2 kJ/kg); palmitic (16.1% vs. 16.3%) and linoleic acids (13.0% vs. 13.1%).

We have also studied the variability in Holm oak, using electrophoresis-based proteomic analysis of acorns. Ten populations distributed throughout the Andalusia region have been surveyed. Acorns were sampled from individual trees and proteins extracted from seed flour by using the TCA–acetone precipitation protocol. Extracts were subjected to SDS-PAGE and 2-DE for protein separation, gel images captured, spot or bands quantified, and subjected to statistical analysis (ANOVA, SOM and clustering). Variable bands or spots among populations were subjected to nLC-MS/MS and MALDI-TOF/TOF respectively for identification. The protein yield of the used protocol varied among populations, and it was in the 2.9–5.9 mg/g dry weight range. A total of 23 bands was resolved by SDS-PAGE in the 3–35 kDa Mr range, with 8 and 12, out of the total, showing respectively qualitative and quantitative statistically significant differences among populations. Data allowed grouping populations, with groups being correlated according to

geographical location and climate conditions, to northern and southern, as well as the discrimination of both mesic and xeric groups. Acorn flour extracts from the most distant populations were analyzed by 2-DE, and 56 differential spots were proposed as markers of variability. Identified proteins were classified into two principal categories; storage and stress/defense protein.

Additionally, we studied the Holm oak pollen proteome, together with an evaluation of the potentiality that a proteomic approach may have in the provenance variability assessment. Proteins were extracted from pollen of four Holm oak provenances, and they were analyzed by gel-based (1- and 2-DE in combination with MALDI-TOF/TOF) and gel-free (nLC-LTQ Orbitrap MS) approaches. A comparison of 1- and 2-DE protein profiles of the four provenances revealed significant differences, both qualitative and quantitative, in abundance (18 bands and 16 spots, respectively). Multivariate statistical analysis carried out on bands and spots clearly showed distinct associations between provenances, which highlight their geographical origins. A total of 100 spots selected from the 402 spots observed on 2-DE gels were identified by MALDI-TOF/TOF. Moreover, a complementary gel-free shotgun approach was performed by nLC-LTQ Orbitrap MS. The identified proteins were classified according to biological processes, and most proteins in both approaches were related to metabolism and defense/stress processes. The nLC-LTQ Orbitrap MS analysis allowed us the identification of proteins belonging to the cell wall and division, transport and translation categories.

Resumen

La encina (*Quercus ilex* subsp. ballota [Desf.] Samp.) es una especie dominante en los sistemas silvopastorales mediterráneos, de entre los cuales forma parte la dehesa. Este sistema de explotación es el más característico y tradicional de los recursos naturales de la Península Ibérica donde los usos agrícola, ganadero y forestal se integran de una forma muy compleja. En España, las poblaciones de encinas cubren una superficie aproximada de 2 039 563 de ha. La importancia económica de esta especie se estima en más de 120 millones de euros en cuanto a la producción de bellota, ya que es un componente alimenticio muy importante de la dieta del cerdo ibérico. Actualmente la encina tiene un gran interés en los programas de conservación forestal, debido a que los encinares están sufriendo un continuo deterioro y desaparición como consecuencia de prácticas agrícolas desproporcionadas y por factores bióticos (enfermedades) y abióticos (estrés por sequía). Por tanto, el declive de los valores naturales de la dehesa puede ser una situación dramática en un futuro. Debido a estas razones, se está desarrollando una línea de investigación enfocada a estudiar, por un lado, la variabilidad genética intra e inter poblacional y, por otro lado, la identificación de rodales e individuos selectos con respuesta diferencial a estreses bióticos y abióticos en la encina, con el fin de identificar marcadores moleculares adecuados de calidad, productividad y tolerancias a estreses que afectan a dicha especie.

Siguiendo con esta línea de estudio, en el presente trabajo de investigación se abordaron los siguientes objetivos:

- El estudio de la variabilidad poblacional de la encina basada en la morfometría y composición química de la bellota.
- El estudio de la variabilidad poblacional de la encina a través del análisis del perfil proteico de las bellotas.
- El estudio del proteoma del polen de encina.

En esta tesis hemos realizado un estudio de la variabilidad poblacional de la encina basada en la morfometría y composición química de bellota, a partir de 13 poblaciones de la región de Andalucía. La morfometría de la bellota mostró diferencias significativas entre

las poblaciones analizadas, presentando los valores más elevados en peso, longitud y diámetro aquellas localizadas en el norte (TJA, GSE, CTH, CHU, APS, y PCO), mientras que las poblaciones del sur (RMA, SAA, CCO, BCA, y SCA) presentaron los valores más bajos. La composición química de la bellota se analizó mediante la técnica de espectroscopía de infrarrojo cercano (NIRS) y también reveló diferencias significativas entre las poblaciones. El análisis de agrupamientos teniendo en cuenta los datos morfométricos y de composición química mostró dos agrupamientos de poblaciones bien diferenciados correspondientes a las poblaciones localizadas en el norte y aquellas localizadas en el sur de la región. Así, las poblaciones del norte (CTH, TJA, SCA, CHU, PCO, GSE, y APS) tuvieron una tendencia a presentar valores mas altos en peso (5,4 g vs. 3,2 g); longitud (37,2 mm vs. 28,7 mm); diámetro (16,1 mm vs. 13,7 mm); cenizas (1,8% vs. 1,5%); proteínas (4,8% vs. 4,3%); grasas (12,5% vs. 9,9%); y ácido oleico (66,9 vs. 66,4%). Por el contrario, las poblaciones del sur (SSA, VJA, CCO, BCA, RG, and RMA) mostraron una tendencia a presentar valores más bajos en azúcares (10,2% vs. 11,2%); contenido calórico (20 076,7 kJ/kg vs. 20 149,2 kJ/kg); y en los ácidos grasos palmítico (16,1% vs. 16,3%) y linoleico (13,0% vs. 13,1%).

El siguiente objetivo fue el estudio de la variabilidad poblacional en la encina, mediante un análisis proteómico basado en electroforesis en gel (1-DE y 2-DE) de las proteínas de la bellota. Este estudio se llevo a cabo en diez de las trece poblaciones analizadas a través de datos morfométricos y bromatológicos. Las proteínas de bellota fueron extraídas usando el protocolo de precipitación basado en TCA-acetona. Las proteínas se separaron mediante SDS-PAGE y 2-DE, las bandas o manchas proteicas fueron cuantificadas, y sujetas a un análisis estadístico (ANOVA, SOM y de agrupamientos). Las bandas o manchas proteicas variables entre las poblaciones fueron identificadas a través de MALDI-TOF/TOF y nLC-MS/MS. El rendimiento de las proteínas varió entre las poblaciones de 2,9 a 5,9 mg/g de peso seco. Un total de 23 bandas fueron separadas por SDS-PAGE en un rango de M*r* de 3–35 kDa, 8 de las cuales mostraron diferencias cualitativas y 12 diferencias cuantitativas entre las diferentes poblaciones analizadas. El análisis de los datos permitió agrupar a las poblaciones de acuerdo a las condiciones geográficas (norte y sur) y climáticas (xéricas, mésicas e intermedias). Las cuatro poblaciones geográficamente más alejadas y con los perfiles de proteína 1-DE más

diferentes fueron analizadas mediante 2-DE, obteniéndose 56 manchas proteicas significativamente variables, las cuales fueron propuestas como marcadores moleculares de variabilidad entre las distintas poblaciones. Las proteínas identificadas fueron clasificadas en dos categorías principales: proteínas de almacenamiento y de respuesta a defensa/estrés.

Igualmente hemos estudiado el proteoma del polen de encina, con el objetivo de estudiar la variabilidad poblacional en la región de Andalucía. En este estudio se realizó el análisis proteómico del polen de cuatro poblaciones, mediante dos aproximaciones metodológicas: técnicas basadas en gel (1-DE y 2-DE en combinación de MALDI-TOF/TOF) y el análisis masivo de proteínas (nLC-LTQ Orbitrap MS). La comparación del análisis de los perfiles proteicos de las cuatro poblaciones reveló diferencias cualitativas y cuantitativas (18 bandas y 16 manchas proteicas), las cuales podrían ser también posibles marcadores moleculares de variabilidad en polen, junto con las proteínas variables en el análisis proteómico de la bellota. El análisis multivariante de los datos normalizados, ya sea de las bandas o de las manchas proteicas, claramente mostró diferentes agrupamientos entre las poblaciones estudiadas, y los resultados fueron relacionados con la localización de las poblaciones. Un total de 100 manchas proteicas, de las 402 observadas en geles 2-DE, fueron identificadas por MALDI-TOF/TOF. De forma complementaria, se realizó un análisis masivo de proteínas usando nLC-LTQ Orbitrap MS. La mayoría de las proteínas identificadas estuvieron relacionadas con el metabolismo y la respuesta a defensa/estrés. Dicho análisis permitió también la identificación de proteínas relacionadas con la división y pared celular, transporte y traducción complementando a las proteínas identificadas por MALDI-TOF/TOF.

1. Introducción

1.1. Contexto de la tesis doctoral

Esta tesis doctoral se centra en la especie más importantes y representativa del ecosistema forestal andaluz, la dehesa, y del bosque mediterráneo: la encina (*Quercus ilex* subsp. ballota [Desf.] Samp.) [1-3]. A pesar de su interés, tanto desde el punto de vista medioambiental como económico, es una especie poco estudiada a nivel molecular. Dicho interés se pone de manifiesto y queda reflejado en: i) la importancia de las especies forestales en el secuestro del carbono, ciclo del agua, protección del suelo y reserva de biodiversidad; ii) el incremento en los programas de reforestación; iii) la elaboración de una normativa específica sobre la dehesa; y iv) en el incremento del número de publicaciones científicas sobre el género *Quercus*, en general, y la encina, en particular. La conservación de la dehesa es, en este momento, un objetivo prioritario, entendido como forma de preservar, desarrollar y revalorizar su riqueza económica, biológica, ambiental, social y cultural, por lo que se hace necesario promover su gestión de una manera integral y sostenible; todo ello a través de su estudio y el desarrollo de nuevas estrategias y técnicas.

Los principales problemas que presenta la dehesa como espacio natural y ente socio-económico productivo son la continua pérdida de arbolado, la baja productividad que presenta el sistema y la compleja gestión que requiere [2], lo que conduce a la necesidad de realizar un doble esfuerzo, tanto en la vertiente científico-técnica como en la socio-económica. La presente tesis se ajusta al análisis anterior y su interés práctico se justifica teniendo en cuenta la pérdida de masa forestal ocurrida en los últimos 30 años, asociada a una mala política forestal, con una falta acusada de diversidad de edades en el arbolado [2], a incendios y al decaimiento o síndrome de la seca, y la que se puede preveer, a largo plazo, debido a las condiciones de cambio climático [4-6]. La mortalidad de árboles observada durante las últimas décadas en la Península Ibérica ha sido relacionada con episodios de variabilidad climática. El clima está creando unas condiciones muy favorables para que se produzca un incremento de las plagas y enfermedades [5]. La predicción del comportamiento de la vegetación a medio o largo plazo, en respuesta a cambios climáticos inducidos por el efecto invernadero, es uno de los problemas científicos actuales más importantes [7]. En España las especies del género *Quercus* parecen haber sufrido de forma

especial una perturbación a gran escala sobre todo el sur de la península y, dada su importancia ecológica y económica, el estudio de este problema parece de especial relevancia [4].

El manejo y mantenimiento sostenible del sistema agroforestal requerirá la identificación de rodales selectos y genotipos productivos y adaptados a condiciones ambientales adversas, para lo que habrá que explotar la variabilidad natural existente en encina (típica de especies alógamas poco intervenidas por el hombre). En este contexto, en los últimos años, los grupos de investigación de Bioquímica y Proteómica Vegetal y Agroforestal del Dpto. de Bioquímica y Biología Molecular y el de Restauración Forestal del Dpto. de Ingeniería Forestal de la Universidad de Córdoba, han desarrollado una línea de investigación dirigida al estudio de la caracterización de una de las especies forestales más representativa de España, la encina, dentro del marco del desarrollo de estrategias de gestión sostenible y de conservación. Esta colaboración ha propiciado la elaboración de una serie de trabajos orientados al conocimiento de la diversidad natural de la especie, específicamente en aquellas poblaciones de encina que se desarrollan en la región de Andalucía. Dentro de los primeros trabajos publicados en esta especie destacan aquellos relacionados con el estudio de la variabilidad y respuesta a sequía a través del uso de la proteómica [8-10]. Complementariamente, se están elaborando una serie de trabajos científicos donde se utilizan aproximaciones genéticas [11], con el fin de tener una caracterización general de la especie y ver su capacidad de respuesta frente al conjunto de factores que determinan el síndrome de la seca, a corto plazo, y en un escenario de cambio climático, a largo plazo.

La encina como especie forestal representativa de los ecosistemas forestales andaluces es de gran interés tanto desde el punto de vista medioambiental como económico. Se trata de una especie cuya biología ha sido poco estudiada, en especial a nivel molecular. En esta tesis, mediante una aproximación multidisciplinar (estudios de morfometría, y uso de nuevas técnicas como NIRs y la proteómica), se ha estudiado la variabilidad poblacional.

1.2. Antecedentes

1.2.1. Los bosques mediterráneos

Los ecosistemas mediterráneos presentan una elevada diversidad de flora y fauna silvestre, así como una enorme variedad de paisajes o modos de vida y de aspectos socio-económicos de las poblaciones humanas que viven en él. La vegetación mediterránea se distribuye desde el norte de África hasta el Oriente próximo, y por el sur de Europa [12]. El clima mediterráneo está caracterizado por una marcada estacionalidad con respecto a la temperatura y distribución de las precipitaciones, así como por una baja predecibilidad intra- e interanual de dichos parámetros. En general, los veranos son calurosos y secos, lo cual genera un notable estrés hídrico y térmico muy característico en las especies que componen los ecosistemas vegetales mediterráneos. Sin embargo, es frecuente que en las zonas próximas al mar Mediterráneo las temperaturas sean moderadas y las precipitaciones elevadas desde el otoño a la primavera, lo cual hace que muchos ecosistemas alcancen su máxima productividad en ese periodo. Este esquema general se modifica como consecuencia de la topografía y la distancia al mar, introduciendo un grado variable de continentalidad y rigor climático [13].

Las formaciones vegetales características de la Cuenca Mediterránea están constituidas por gran variedad de matorrales y bosques dominados por especies tolerantes a el estrés por sequía, que limitan por el norte con bosques templados de especies caducifolias durante el invierno, y por el sur y el este con estepas y formaciones sub-desérticas de escaso desarrollo, aunque de gran interés biológico y ecológico. Las especies más características son las encinas, quejigos y robles (*Quercus* spp.), frecuentemente mezcladas con pinos (*Pinus* spp.) y madroños (*Arbutus unedo*), mientras que en sitios más secos o continentales, crece junto con enebros y sabinas (*Juniperus* spp.). Existen claras diferenciaciones geográficas donde las formaciones con chaparros (*Quercus coccifera*) y pino carrasco (*Pinus halepensis*) de la zona oeste encuentran su equivalente en formaciones con robles de Palestina (*Q. calliprinos*) y el pino de Chipre (*P. brutia*) en el este. O los sabinares ibéricos con *Juniperus phoenicea* y *J. thurifera* encuentran su equivalente en Turquía en formaciones dominadas por *J. excelsa* y *J. foetidissima*. Pero, posiblemente, una de las especies más representativas y abundantes de los bosques mediterráneos es la encina

(*Q. ilex*). Esta especie tiende a formar bosques propios y es una de las especies más comunes en los sistemas de dehesas y montados. La encina coexiste con robles (*Q. robur*) y hayas (*Fagus sylvatica*) en el norte (zonas húmedas), con sabinares (*J. sabina*) (zonas continentales), y con gran variedad de matorrales termófilos y xerofíticos en el sur de la cuenca mediterránea y en numerosas zonas costeras [14-17].

1.2.2. Género Quercus

El género *Quercus* pertenece a la familia de plantas Fagáceas. Dentro de esta familia, este género es el que presenta una mayor distribución en todo el mundo. Se le puede localizar en casi todos los bosques templados del hemisferio norte, así como también en algunas regiones tropicales y subtropicales del mismo. Su distribución abarca desde noreste de África al sureste de Asia, en el continente Americano desde Canadá hasta Colombia, y en la región Mediterránea se puede localizar desde Portugal hasta Siria y desde Marruecos hasta Francia.

Se reconocen dos zonas de diversidad para el género. La primera se localiza en el sureste de Asia con alrededor de 125 especies [18] . El segundo se presenta en México [18] particularmente en las regiones montañosas, donde forma parte importante de los bosques templados. Sin embargo, es muy complicado especificar el número total de especies descritas para el género *Quercus*, aunque algunos autores estiman que oscila entre 300 hasta 600 especies [19, 20].

El género *Quercus* está formado por árboles de gran porte por lo general, aunque también se incluyen arbustos. Los árboles poseen hojas que nacen a ambos lados del tallo (hojas alternas) y pueden ser de follaje perennes, caducifolios y marcescentes. La forma de sus hojas es muy variable, incluso dentro del mismo árbol según el sitio donde estén las hojas o dependiendo de la edad. Este género se caracteriza por ser monoico, es decir, tienen flores masculinas y femeninas separadas (unisexuales) pero presentes en cada árbol. Las flores masculinas están agrupadas en amentos o inflorescencias complejas colgantes, donde habitualmente cada flor posee entre cuatro y diez estambres (siendo lo más a menudo seis) de largos filamentos. Las flores femeninas aparecen aisladas u organizadas en espigas o cabezuelas, presentan tres estigmas, así como óvulos anátropos, y están rodeadas por una

estructura de escamas empizarradas que al madurar se convertirá en la cúpula [19, 20]. Lo más característico de este género es la fruta, llamada bellota, con su típico gorro o capuchón semiesférico en la base llamado cúpula. La bellota es un fruto solitario y de origen axil (de brote), con cotiledones planos. La corteza suele ser lisa en los ejemplares jóvenes pero se va agrietando en los ejemplares maduros con la edad. Suelen medir de 1 a 5 cm de longitud. Estos árboles se reproducen a través de la semilla, pero son de crecimiento lento. En la **Figura 1** se pueden observar algunas de las características que presenta el género *Quercus*.

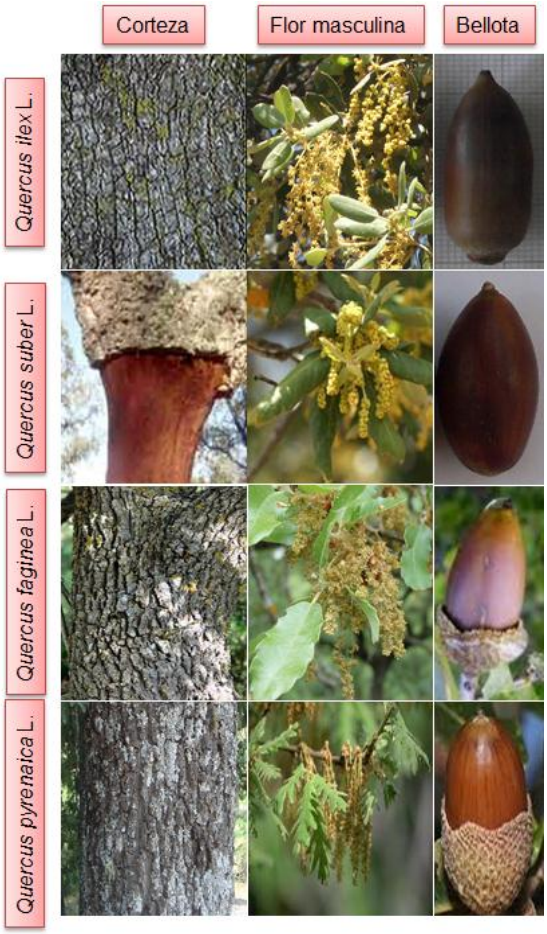


Figura 1. Corteza, flor masculina, y frutos de algunas especies forestales del género *Quercus* spp.

En la **Tabla 1** se presentan algunas de las especies más representativas de este género *Quercus* en España. Las más importantes de acuerdo a la superficie que ocupan son: *Q. ilex* subsp. *ballota*, *Q. pyrenaica*, *Q. faginea* y *Q. suber* (**Figura 2**).

Tabla 1. Topología de los bosques de *Quercus* ibéricos y superficie ocupada [21, 22] .

Tipo	Especie característica	Superficie (ha)
Robledales o carballeiras	*Q. robur*	286 987
Robledales albares	*Q. petraea*	45 802
Quejigares de roble pubescente	*Q. humilis*	111 162
Melojares	*Q. pyrenaica*	948 315
Quejigares	*Q. faginea*	636 530
Quejigares de *Q. canariensis*	*Q. canariensis*	9 476
Alsinares	*Q. ilex* subsp. *ilex*	158 038
Encinares carrascales interiores continentales	*Q. ilex* subsp. *ballota*	5 562 079
Alcornocales	*Q. suber*	655 098
Coscojares	*Q. coccifea*	1 627

1.2.3. La encina

La encina (*Q. ilex*) es un árbol monoico, alogamo y anemófilo incluido en el subgénero *sclerophyllodrys*, especie *ilex*, la cual presenta dos subspecies *ilex* y *ballota* [23]. La encina se encuentra ampliamente distribuida en la región basal del Mediterráneo [24], donde se distribuye desde Portugal hasta Siria y desde Marruecos hasta Francia. En España cubre aproximadamente 5,5 millones ha (Tabla 1). Su distribución altitudinal es amplia debido a que soporta bien condiciones extremas y puede extenderse desde el nivel del mar hasta los 2 000 m de altitud, pero su estado óptimo está entre 200 y 1 200 m. La encina es poco exigente en cuanto a humedad (tolera menos de 350 mm anuales) y resiste bien tanto temperaturas altas (28ºC) como bajas (-3ºC). No es exigente en cuanto al suelo, pudiéndose

encontrar tanto en sustratos calizos como silíceos, prefiriendo los suelos sueltos, ligeros y permeables, en los que adquiere el máximo desarrollo. Se adapta a suelos pedregosos aunque tolera mal los terrenos encharcados y los margosos o arcillosos excesivamente compactos [25].

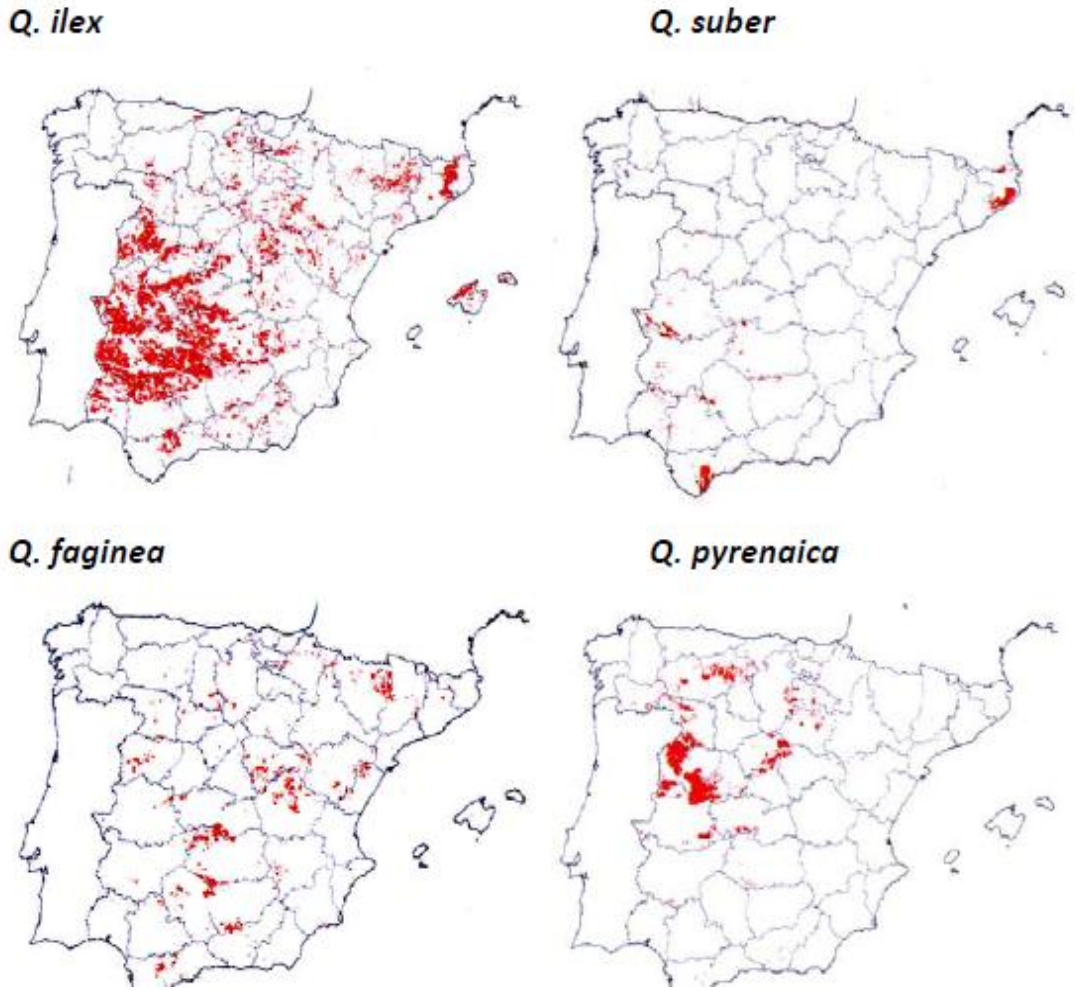


Figura 2. Mapa de distribución de las cuatro especies de género *Quercus* más representativas en el territorio español [26].

La encina es un árbol corpulento, de talla media, que puede desarrollarse hasta los 25 metros de altura. El tronco es recio y cilíndrico pudiendo ser derecho o algo torcido y ramificado por encima de los 2 metros (**Figura 3**). La corteza es de color ceniciento o parduzco y agrietada. Las ramas son abiertas, entre erguidas y horizontales, robustas y muy ramificadas; la copa suele ser amplia, densa y redondeada (**Figura 3**).

Figura 3. Árbol de encina desarrollándose en la población de Córdoba (PCO).

El sistema radical es muy penetrante, axonoforme, pivotante y de fuerte crecimiento inicial. Presenta hoja simple, alterna, persistente y puede exhibir vellosidades. La forma es variable, de manera que las que se sitúan cerca del suelo son más coriáceas y presentan espinas, mientras que las superiores suelen ser más redondeadas. El haz es verde glauco y liso y el envés grisáceo y tomentoso con un número variable de dientes enteros en el borde (**Figura 4**). Como se ha comentado anteriormente, es una especie monoica, teniendo flores masculinas y femeninas separadas. Las flores masculinas están agrupadas en amentos arracimados colgantes y las femeninas sobre las metidas del año, de forma solitaria o en parejas [27].

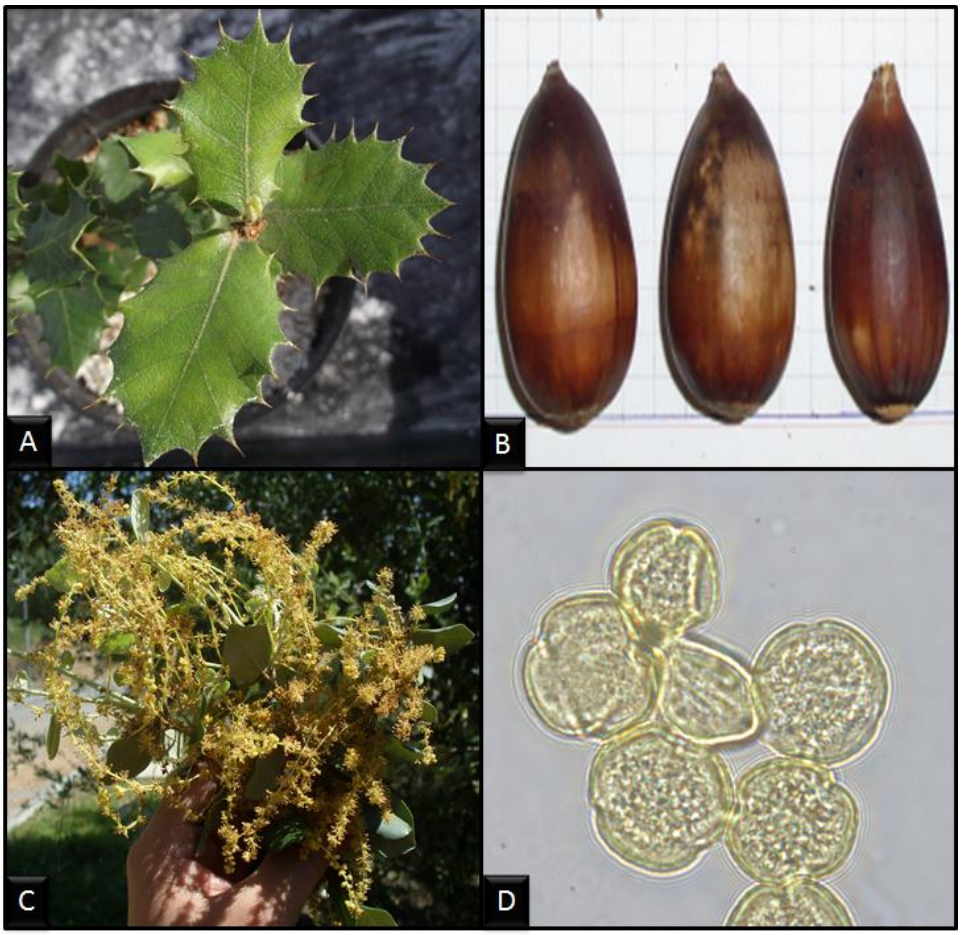


Figura 4. Características fenotípicas de la encina. A) Hojas; B) Fruto (bellota); C) Flores masculinas; D) Polen.

El fruto (bellota) es un aquenio de forma ovoide de hasta 5 cm, de color castaño negruzco en la madurez, con una cúpula hemisférica en la base de color grisáceo, con escamas planas muy apretadas y un pedúnculo (**Figura 4B**) [27]. La producción de bellota comienza a los 8-10 años y su cosecha suele comenzar a los 15-20 años, siendo regular y abundante en climas templados, con el máximo entre 50 y 100 años. En las zonas bajas de los montes adehesados, las encinas suelen producir fructificación anual abundante cada 2-3 años, a partir de los 10. Un resumen de la fenología del desarrollo vegetativo de la encina se indica en la **Tabla 2**.

Tabla 2. Fenología de la encina [28].

Desarrollo vegetativo	Fechas
Brotación	Primavera, rebrotación de raíz hasta la muerte y rebrotación de cepa hasta 200-300 años
Floración	Un mes después de la foliación; finales de abril a principios de mayo
Polinización	Embriogénesis y engrosamiento de la bellota durante el verano hasta el comienzo del otoño
Caída de la bellota	Octubre a diciembre
Edad de fructificación	8-10 años
Germinación	Otoño o letargo hasta final de primavera
Longevidad	700-800 años

1.2.3.1. Distribución geográfica en España y Andalucía

En España, la encina se distribuye por el Valle del Duero, Levante y Bajo Aragón, hasta el Valle del Ebro (**Figura 2**). Su existencia es rara en las vertientes al mar de las sierras y cordilleras, siendo propia de regiones de clima continental o subcontinental. En el norte se sitúa en localidades donde las precipitaciones son de hasta 800 mm al año, en el Pirineo y pre-Pirineo de 400 m, y en el Bajo Duero de hasta 300 mm, al igual que en el surco inter-ibérico; en la zona de Levante su limite está en zonas con 350 mm, al igual que en toda la mitad sur [27].

En la región de Andalucía ocupa una extensión territorial de aproximadamente 735 671 ha [29]. En la **Figura 5**, se muestra la distribución de los encinares en la región de Andalucía. Dentro de este territorio, los encinares se han clasificados en dos tipos: los luso-extremeños y los béticos. Los encinares luso-extremeños se localizan en la zona occidental y cuando están bien conservados son bosques bastante homogéneos y muy ricos en

especies, sobre todo en las umbrías donde se mantiene algo más la humedad durante el estío. Aparecen sobre suelos ácidos (pizarras, esquistos) en las provincias de Huelva, Sevilla y Córdoba. El estrato arbustivo lo forman madroños, coscoja, cornicabra (*Pistacia terebinthus*), cantueso (*Lavandula stoechas* subsp. *pedunculata*), cistáceas, ericáceas (especies del género *Erica* y la brecina, *Calluna vulgaris*), genistas e incluso acebuche (*Olea europaea* var. *sylvestris*). También hay otras especies destacables en estos encinares como piruétano (*Pyrus bourgaeana*), quejigo (*Q. faginea*), majuelo (*Crataegus monogyna*), fresno (*Fraxinus angustifolia*) y mirto o arrayán (*Myrtus communis*). En este tipo de encinares dominan los bosques con estructura de dehesa.

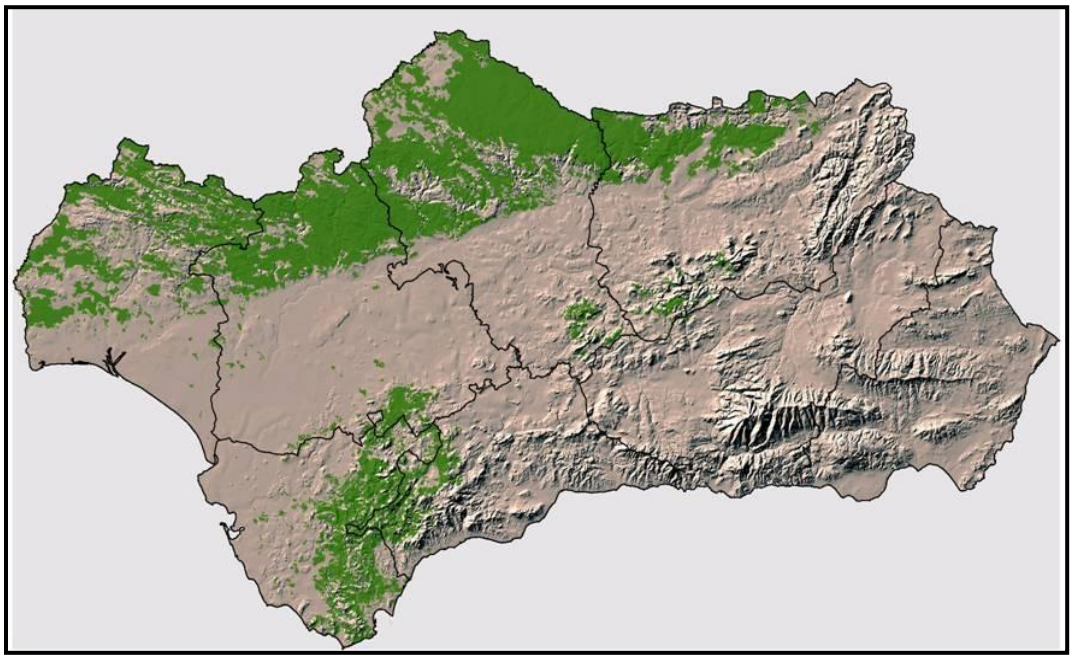


Figura 5. Mapa de distribución de los encinares en Andalucía (Consejería Medio Ambiente, 1996).

En el sector oriental, encontramos los encinares béticos, propios de las Sierras Béticas y Penibéticas. Se asientan sobre calizas, filitas y cuarcitas. Los bosques tienen porte arbustivo y alta densidad en las zonas bajas, donde la encina es la especie dominante junto al enebro (*J. oxycedrus*), el torvisco (*Daphne gnidium*), el tomillo (*Thymus* spp.), la salvia (*Salvia* spp.), el romero (*Rosmarinus officinalis*), el lentisco (*Pistacia lentiscus*) o el acebuche (*O. europaea* var. *sylvestris*). A medida que se asciende en altitud, aparecen

arbustos de montaña como el agracejo (*Berberis vulgaris* subsp. *australis*) y en cotas superiores a 1 400 m aparecen otras especies arbustivas como el majuelo (*Crataegus monogyna*), el endrino (*Prunus spinosa*), e incluso algunos caducifolios como el arce (*Acer monspessulanum*) y alguna trepadora como la hiedra (*Hedera helix*). Las zonas más degradadas están dominadas por los coscojares (*Q. coccifera*), las genistas (*Genista cinerea* subsp. *speciosa*) o retamares (*Retama sphaerocarpa*), en contacto frecuente con pinares de pino carrasco (*P. halepensis*). En áreas de montaña, las encinas se mezclan con quejigos y pino salgareño (*P. nigra*), y en zonas de la Serranía de Ronda y la Sierra de Grazalema con pinsapos (*Abies pinsapo*).

1.2.3.2. Importancia económica, problemas y retos de la encina

La encina es el componente principal de las dehesas, que se caracterizan por ser formaciones arboladas en las que la distancia entre los troncos permite el desarrollo y utilización de los pastizales subyacentes por la ganadería. La encina cumple funciones esenciales para la rentabilidad del sistema, proporcionando tanto rentabilidad ecológica como económica. Ecológicamente la encina protege al suelo de la erosión, ya que evita los efectos desecantes del viento, el exceso de radiación solar y facilita la condensación de las nieblas. Además, la circulación del ganado con sus aportes orgánicos suministra abono al suelo favoreciendo la implantación de especies. Mientras que a nivel económico, la encina aporta la extracción de madera o leña, y la bellota alimenta el ganado porcino en régimen de montanera.

Los problemas que afrontan los encinares de la región de Andalucía están relacionados con el manejo tradicional al que han sido sometidos desde épocas remotas, y a amenazas derivadas de plagas y enfermedades [30]. La transformación del monte original en dehesas supuso problemas de regeneración poblacional que conducen a unas masas envejecidas y debilitadas [31]. La propiedad privada de los montes ha dificultado el desarrollo de planes y estrategias de mejora o de conservación. Otro de los problemas que actualmente presentan las poblaciones de encina es el fenómeno conocido como "la seca", atribuido al ataques de hongos (como *Hypoxilon mediterraneum* o *Phytophthora cinnamomi*) [6, 32, 33], aunque el factor desencadenante puede deberse a la sucesión de varios años de sequía y temperaturas altas. Así, el estrés ambiental prolongado causaría un

debilitamiento en los individuos, siendo altamente susceptibles al ataque de plagas y enfermedades. Por último, la reforestación en tierras agrarias está suponiendo un movimiento de semillas que, a pesar de la normativa, no siempre es posible controlar, provocando una homogenización de las poblaciones. Por tanto, la base de los programas de mejora forestal deberían estar enfocados al estudio de la diversidad natural para la selección de genotipos superiores (árboles plus vs. selección poblacional) en poblaciones naturales, considerando que estos genotipos superiores podrían llegar a adaptarse mejor a unas condiciones adversas tales como estreses abióticos (sequía) o bióticos (ataques por hongos).

1.2.3.3. Investigaciones relacionadas con la encina

En la actualidad, las investigaciones realizadas en especies forestales se están enfocando en el estudio de marcadores moleculares que puedan ser relacionados con el crecimiento y la productividad. Sin embargo, la investigación forestal presenta algunos retos como sistema experimental, debido a que estas especies presentan genomas de mayor tamaño comparado con especies cultivadas, un ciclo de vida largo y son de crecimiento lento. Además, la preparación de las muestras para el análisis en el laboratorio es complicado, especialmente en aquellas especies que presentan semillas recalcitrantes o problemas de germinación [34]. Estos factores constituyen los principales retos en el desarrollo de los estudios moleculares y las técnicas clásicas utilizadas en los programas de mejora forestal de dichas especies [35]. Así, cabe destacar que especies como las representadas por el género *Quercus* están poco estudiadas, y que en el periodo de 1990-2011 tan sólo se han publicado 11 244 trabajos relacionados con este género, en contraste con los publicados en *Arabidopsis* (67 660) o el género *Oriza spp.* (75 115). Únicamente el 23% de los trabajos publicados en *ISI Web of Knowledge* estuvieron relacionadas a los *Quercus* mediterráneos, así tenemos que *Q. ilex* presentó sólo 1 613 publicaciones, mientras que *Q. suber* 704 y, finalmente, *Q. coccifera* 293. Del total de los trabajos publicados para *Q. ilex*, tan sólo el 9% estaban relacionados con estudios de variabilidad.

Para estudiar el polimorfismo y la variabilidad en el género *Quercus* se han utilizado medidas de las características anatómicas o químicas, de actividades enzimáticas, de marcadores de ADN, y estudios de introgresión o filogenéticos [36-42]. Sin embargo, en España son pocos los trabajos publicados referentes al estudio de la variabilidad en especies

forestales tales como la encina [38, 43-46]. Si acotamos más la busqueda a la región de Andalucía los estudios son muy escasos y los que se encuentran son estudios con pocas poblaciones .

a) Variabilidad morfométrica de la bellota

El género *Quercus* se ha caracterizado por presentar una alta diversidad genética, ya que presenta unas características que le hacen favorable para tener una amplia distribución geográfica, como son una vida larga, es alógama, anemófila y capaz de hibridar entre varias especies del mismo género. Numerosos estudios han establecido la gran variabilidad morfológica (en hoja y bellota) que presenta el género *Quercus* [47, 48]. En lo que se refiere a la encina, Vázquez et al. [49] clasificaron diferentes formas de la bellota de encina basándose en sus características morfológicas (**Tabla 3**), presentándose una elevada variabilidad morfométrica (longitud, diámetro y peso seco) en dicha semilla.

Tabla 3. Clasificación de las bellotas de *Q. ilex* de acuerdo a su morfología [49].

Forma	Longitud (mm)	Diámetro (mm)	Peso seco (g)
rotundifolia	30-40	12,0-17,0	1,5-5,0
macrocarpa	38-50	19,0-22,0	6,0-8,0
expanda	33-40	14,0-16,5	2,5-5,0
dolichocalyx	21-37	0,9-15,0	3,5-6,0
avellaeniformis	12-14	9,5-10,5	0,2-0,5
brevicupulata	28-30	16,0-18,0	2,5
crassicupulata	32-37	10,0-11,5	1,5-3,5

Algunos autores han relacionado estas variaciones morfométricas de la bellota a factores tales como las condiciones ambientales o el manejo del entorno en las que se haya desarrollado el fruto. Sin embargo, la especie responde a mecanismos de heredabilidad propios de la especie, lo que permiten presentar características de adaptación y resistencia, o de tolerancia a condiciones de estrés [50]. Además, se ha descrito una fuerte influencia

genética en las características de las bellotas de cada árbol, y las diferencias entre éstos explican el 62% de la variación de la biomasa de las bellotas [51].

b) Composición química de la bellota

La composición química de la bellota de encina adquiere importancia ya que se utiliza en la ceba del cerdo Ibérico, debido a que la bellota influye positivamente en la calidad de la carne y en su perfil de ácidos grasos [30, 52, 53]. La composición química de la bellota del género *Quercus* ha sido usada como herramienta para determinar variabilidad en dicho género [54-56], pero son pocos los estudios realizados en la bellota de la encina. En la **Tabla 4**, se muestran tres de los trabajos más sobresalientes de la composición química de la bellota. La bellota presenta altos contenidos de grasa (4-12%), además de altos contenidos en los ácidos oleico (> 63%), y palmítico y linoleico (12-20%). Sin embargo, estos porcentajes varían entre especies del género *Quercus*, y también lo hace en función de otros factores, como son la humedad ambiental, el periodo de la montanera, el estado de maduración y el ataque por plagas o fitopatógenos [30, 57, 58].

Actualmente no existen estudios en donde se haya comparado un número importante de parámetros en la composición química de harina de bellotas de varias procedencias, con el fin de llegar a catalogar árboles selectos para desarrollar programas forestales más eficientes. Para conocer la composición química de las bellotas, una técnica de gran utilidad es la espectroscopía de infrarrojo cercano (NIRS). Esta técnica analítica consiste esencialmente en la proyección de un haz de energía perteneciente a la región del espectro electromagnético (comprendido entre las longitudes de onda de 700 y 2 500 nm, lo que corresponde al infrarrojo cercano) sobre la muestra. Ésta, en función de su composición y de la naturaleza de los enlaces presentes en sus moléculas, realizará una absorción selectiva de fotones, reflejando o dejando pasar el resto que serán captados por unos receptores determinados. La absorción de energía por parte de la materia obedece a la ley de Lambert-Beer según la cual cuando una radiación electromagnética de longitud de onda definida incide sobre una sustancia, la fracción de la radiación absorbida es una función de la concentración de la sustancia en la trayectoria de la luz y del espesor de la muestra [59, 60]. Al desarrollar una calibración NIRS, se relaciona mediante un algoritmo la información espectral (óptica) con la información de la composición físico-química (método de

referencia) a través de la aplicación de modelos estadísticos como son la regresión múltiple, los componentes principales y los cuadrados mínimos parciales.

Tabla 4. Variación de la composición química de diferentes poblaciones de encina.

Composición química	Referencias		
	[30]	[61]	[62]
Pulpa DM (%)	56,72-60,57	68,70-66,00	56,00
Ceniza (%)[1]	1,99-1,89	1,50-1,50	1,80-2,00
Proteína Cruda (%)[1]	04,48-04,8	4,60-5,40	3,70-5,00
Fibra (%)[1]	2,71-2,92	5,70-5,90	0,81-1,06
Almidón (%)[1]	59,73-58,28	-	-
Azúcar (%)[1]	5,89-6,84	-	-
Grasa Cruda (%)[1]	10,05-10,76	5,60-7,00	9,50-10,40
C16 (%)[2]	13,97-14,28	-	16,30-20,60
C16:1 (%)[2]	0,12-0,11	-	-
C17 (%)[2]	0,10-0,10	0,10-0,10	-
C17:1 (%)[2]	0,09-0,1	-	-
C18 (%)[2]	3,15-3,28	3,60-3,40	3,30-4,20
C18:1 (%)[2]	64,99-65,45	56,70-62,90	53,64-62,50
C18:2 (%)[2]	15,21-15,31	17,70-18,50	12,20-18,20
C18:3 (%)[2]	0,73-0,81	1,10-1,70	-
C20 (%)[2]	0,44-0,46	0,30-0,30	-
C20:1 (%)[2]	0,56-0,57	-	-

[1] Como porcentaje de la pulpa DM; [2] Como porcentaje de grasa cruda.

A pesar de la cantidad de trabajos y publicaciones existentes sobre las aplicaciones de esta técnica a productos de origen vegetal y animal, su aplicación al análisis de bellota es aún muy elemental y no está muy documentado en comparación con otros productos. Sin embargo, una de las líneas de trabajo en el Dpto. de Producción Animal de la Universidad de Córdoba es el estudio de la calidad de la bellota mediante NIRS y los trabajos de Gómez et al. (datos no publicados) han desarrollado metodologías y aportado resultados de la aplicación de esta técnica en análisis cualitativos y cuantitativos de harina de bellotas de encina y alcornoque. El uso de técnicas de análisis no destructivas de la muestra será de suma importancia en un futuro de cara a la selección de semillas con las características adecuadas para producir plántulas selectas para forestación o en programas de mejora genética.

c) Proteómica vegetal: encina

La proteómica se encarga de estudiar el proteoma de los diferentes seres vivos, entendiendo éste como el conjunto de todas y cada una de las proteínas presentes en una unidad biológica (fracción subcelular, célula, tejido, órgano, organismo) en un tiempo (estadio de diferenciación y desarrollo) bajo condiciones ambientales determinadas [63]. La proteómica constituye, hoy en día, un área de investigación prioritaria en cualquier proyecto biológico, y la investigación vegetal no es una excepción. Los estudios de proteómica permiten identificar *a priori* el conjunto de todas y cada una de las proteínas de un sistema biológico y entre ellas aquellas directamente relacionadas con factores de productividad, valor nutricional, marcadores de genotipos, procedencias y respuestas a estreses bióticos y abióticos. Además, dichos estudios permiten la identificación de alérgenos plantas de especies herbáceas, como queda reflejado en las revisiones realizadas sobre el tema [64-66]. La mayoría de los estudios de proteómica de especies vegetales se han llevado a cabo en herbáceas [64-66]. Como técnica, la proteómica ha demostrado ser útil en la caracterización de genotipos, a través del estudio de las proteínas, además ha permitido establecer distancias genéticas y relaciones entre líneas, especies y géneros [67-73]. Junto con lo anterior, la proteomica se ha utilizado fundamentalmente en el estudio de la respuesta a estreses incluyendo la sequía [64-66]. En especies forestales destacan aquellos trabajos realizados en los géneros *Populus* [74, 75] y *Pinus* [67, 76, 77].

El uso de la proteómica en el estudio de la encina es prácticamente novedoso, debido a que los primeros trabajos que se realizaron fueron en el 2005. Desde un punto de vista metodológico, un experimento típico incluye las siguientes etapas: i) obtención de un extracto, generalmente mediante métodos de precipitación (TCA-acetona y TCA-fenol); ii) separación de proteínas por electroforesis mono- y bidimensional, tinción de geles (Coomassie coloidal), análisis de imágenes (PD-Quest), comparación del mapa proteico e identificación de bandas/spots diferenciales; iii) análisis por espectrometría de masas (por ejemplo, MALDI-TOF/TOF MS/MS) de fragmentos peptídicos obtenidos tras la digestión con tripsina; iv) identificación de proteínas por búsqueda en base de datos (por ejemplo, UniProt) a partir de datos de huella peptídica (PMF o espectros MS) o espectros de fragmentación (MS/MS) utilizando motores de búsqueda (por ejemplo, MASCOT) (**Figura 6**) [8-10].

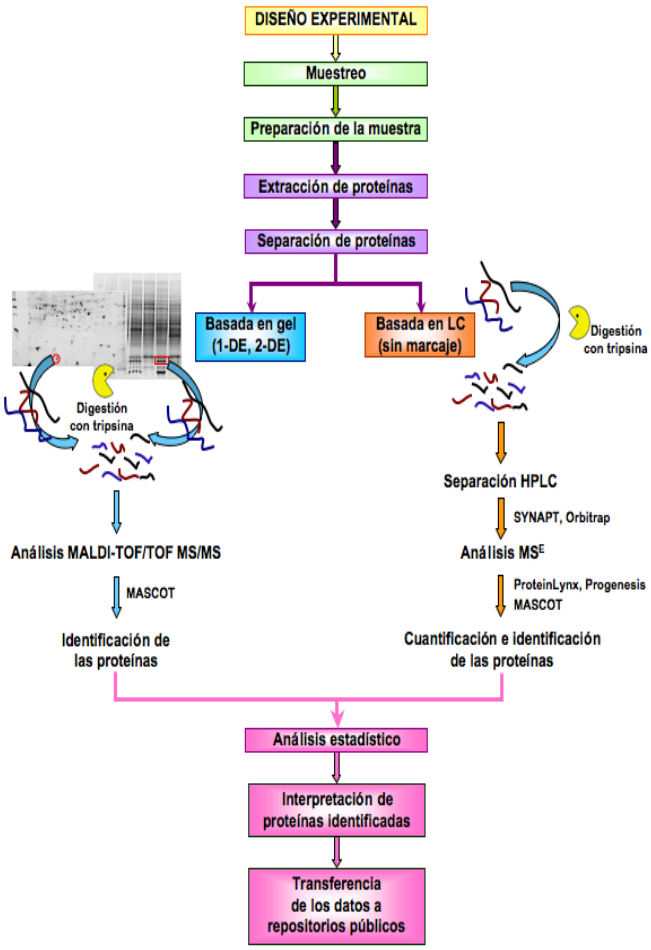


Figura 6. Secuencia metodológica típica en proteómica.

Jorge et al. [10] describieron el proteoma de hoja en la encina. Los autores estudiaron la varibilidad en hoja de encina donde la variabilidad biológica calculada para 100 manchas proteicas fue del 56%, y ésta dependió de la orientación de la hoja (norte, sur, este u oeste) y de la hora de muestreo. Estas variaciones demostraron, por una parte, el carácter dinámico del proteoma, y, por otra, la gran plasticidad de las plantas. Por ello, no se han encontrado marcadores claros asociados a la variabilidad poblacional. Sin embargo, y a pesar de las limitaciones, se realizó un estudio de variabilidad de la encina en tres regiones de Andalucia: Región Extremadurense y Sierra Morena Occidental (Qi11e), Sierra Nevada-Filabres (Qi16a), y Sierras Béticas Occidentales-Sierras Interiores (Qi14a). Al comparar los perfiles de proteínas de hojas entre las tres poblaciones, se pudieron localizar cuatro proteínas que fueron encontradas diferenciales entre poblaciones. La cadena alfa de la ATP sintasa y la fosfoglicerato mutasa independente de 2,3-bisfosfoglicerato se detectaron sólo en plántulas de Qi16a. Dos manchas proteicas no identificadas no se detectaron en los extractos proteicos de la población Qi11e y no se observó una diferencia significativa entre las poblaciones Qi14a y Qi16a. Martínez et al. [78] estudiaron la fracción de las glutelinas de extractos de bellotas de encina en ocho poblaciones localizadas en las provincias de Córdoba y Sevilla. Los resultados obtenidos demostraron que de las 36 bandas resueltas por SDS-PAGE, 32 fueron polimórficas entre las diferentes poblaciones. Dichas investigaciones demostraron una gran variabilidad, lo que recomendó que se abordase el estudio de variabilidad en la encina mediante el análisis proteómico de un material con un proteoma más estable, como es el caso de la bellota o el polen.

2. Objetivos

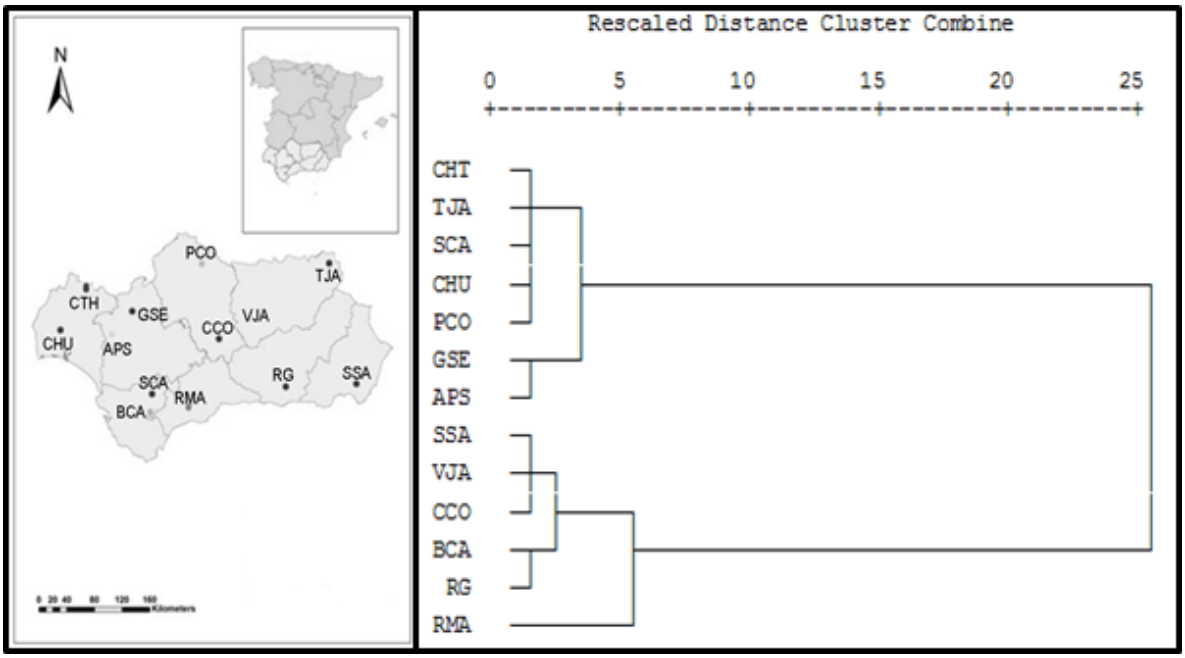
El objetivo general de esta tesis fue analizar la variabilidad en poblaciones de encina (*Q. ilex* subsp. ballota [Desf.] Samp.) localizadas en la región forestal de Andalucía. Para alcanzar este objetivo general se han planteado los siguientes objetivos específicos:

- El estudio de la variabilidad poblacional de la encina basada en la morfometría y composición química de la bellota.
- El estudio de la variabilidad poblacional de la encina a través del análisis del perfil proteico de las bellotas.
- El estudio del proteoma del polen de la encina.

3. CAPÍTULO I

El estudio de la variabilidad poblacional basada en la morfometría y composición química de la bellota en encina.

Population variability based on morphometry and chemical composition of the acorn in Holm oak (*Quercus ilex* L. subsp. *ballota* [Desf.] Samp.).



JOSÉ VALERO GAVÁN, JESÚS V. JORRÍN NOVO, AUGUSTO GÓMEZ CABRERA, DAVID ARIZA, JUAN GARCIA OLMO, RAFAEL Mª NAVARRO CERRILLO. 2011. Population variability and mother tree selection in holm oak (*Quercus ilex* L. subsp. *ballota* [Desf.] Samp) based on acorn morphometry and chemical composition. European Journal of Forest Research. DOI: 10.1007/s10342-011-0563-8.

Eur J Forest Res
DOI 10.1007/s10342-011-0563-8

ORIGINAL PAPER

Population variability based on the morphometry and chemical composition of the acorn in Holm oak (*Quercus ilex* subsp. *ballota* [Desf.] Samp.)

José Valero Galván · Jesús J. Jorrín Novo · Augusto Gómez Cabrera · David Ariza · Juan García-Olmo · Rafael M[a] Navarro Cerrillo

Received: 17 November 2010 / Revised: 24 June 2011 / Accepted: 11 August 2011

Abstract We have studied in this work the natural variability in Holm oak (*Quercus ilex* subsp. *ballota*) by analyzing acorn morphometry and chemical composition in 13 populations from the Andalusia region (southern Spain). Acorn morphometry showed statistic significant differences between the populations with a higher acorn weight, length, and diameter in the northern provenances (TJA, GSE, CTH, CHU, APS, and PCO), while the Southern provenances (RMA, SAA, CCO, BCA, and SCA) showed lower value that even decreased with the altitude. Acorn chemical composition also showed statistic significant differences between populations. Morphometry and chemical composition data were subjected to clustering analysis, and two major clusters were detected. This analysis showed that northern populations (CTH, TJA, SCA, CHU, PCO, GSE, and APS) showed a tendency to present values highest in acorn weight (5.42 g vs. 3.21 g); acorn length (37.21 mm vs. 28.73 mm); acorn diameter (16.12 mm vs. 13.79 mm); ash (1.88% vs. 1.55%); protein (4.82% vs. 4.38%); fat (12.55% vs. 9.92%); and oleic acid (66.93 vs. 66.4%). On the contrary, Southern populations (SSA, VJA, CCO, BCA, RG, and RMA) showed a tendency to present values lowest in sugar (10.28% vs. 11.25%); energy (20,076.71 kJ/kg vs. 20,149.26 kJ/kg); palmitic (16.11% vs. 16.39%) and linoleic acids (13.07% vs. 13.11%). The populations studied here seemed to integrate acorn morphology and chemical composition related to environmental conditions of the original provenances. These data have provided biochemical markers that help establish phylogenic associations between populations and also reveal potentially account as an alternative source for dietary nutrition.

Keywords *Quercus ilex* · Intraspecific variation · Geographical gradients

Communicated by J. Müller.

J. Valero Galván · J. J. Jorrín Novo
Department of Biochemistry and Molecular Biology, Agricultural and Plant Biochemistry and Proteomics Research Group, University of Córdoba, Córdoba, Spain

A. G. Cabrera
Department of Animal Production, University of Córdoba, Córdoba, Spain

D. Ariza · R. M. N. Cerrillo (✉)
Department of Forestry Engineering, ETSIAM, University of Córdoba, Córdoba, Spain
e-mail: rmnavarro@uco.es

J. García-Olmo
NIR/MIR Service, SCAI, University of Córdoba, Córdoba, Spain

Introduction

Holm oak (*Quercus ilex* L.) is the most abundant and representative *Quercus* tree species in the Mediterranean forest. In Spain, its species is covering an approximate surface of 2,039,563 ha, with estimated values of over 120 million of euros in fruit acorn production per year (Gea-Izquierdo et al. 2006). The Holm oak acorns are a major component in the feeding systems of many Mediterranean wild and livestock species. Moreover, it is also the basic feed ingredient for domestically bred high-quality meat pigs (Gea-Izquierdo et al. 2006), which has a relevant economic and social impact on the Iberian Peninsula (Soto et al. 2008). Nowadays, forest restoration and reforestation are high-priority objectives, and Holm oak is acquiring more interest for Mediterranean forestry, together with other evergreen species, namely *Quercus suber* L.,

Published online: 28 September 2011

Ceratonia siliqua L., and *Olea europaea* subsp. *sylvestris* Brot. Spain Holm oak reforestation programs have heavily increased in the last 10 years, overtaking the *Pinus* species, which were widely used in the past (Pausas et al. 2004). Only in Andalusia (Spain), more than 75,000 ha were reforested from 1993 to 1999, with the introduction of millions of new seedlings.

As natural, non-domesticated plant species, with great plasticity and phenotypic variability, a key challenge prior to massive clonal propagation is the establishment of techniques for the cataloging and selection of genotypes among provenances with high survival percentage and productivity under specific environmental conditions. The existence of a high variability and polymorphism in *Quercus* spp. has been concluded from data obtained by using anatomical, chemical, and enzyme activity indicators, as well as DNA marker techniques (Finkeldey and Matyas 2003; Gandour et al. 2007; Gömöry et al. 2003; Jiménez et al. 1999; Lumaret et al. 2005; Lumaret and Jabbour-Zahab 2009; Martín et al. 2009; Michaud et al. 1995; Toumi and Lumaret 1998). In this context, Spain holm oak populations have been showed to present a prolonged isolation and complex geographic patterns of genetic variation, being native and distributed through diverse forest areas (Lumaret et al. 2002; Petit et al. 2002). However, several problems of typification particularly in the wide-ranging population are still unresolved, and modern techniques for the cataloging and selection of genotypes among provenances are needed to improve phylogenetic analysis. A proteomic research program is being carried out in order to study variability in Andalusia Holm oaks (Echevarría-Zomeño et al. 2009; Jorge et al. 2005, 2006; Martín et al. 2009; Valero et al. 2011). In our previous publications, we attempted to contribute toward understanding the diversity and phylogeny by analyzing leaf and seed proteome of several populations. The results generally indicated an appreciable difference between Andalusia Holm oak populations. In the present study, we continue our efforts to characterize the morphometry and chemical composition from the acorns of thirteen Andalusia Holm oak populations. Acorn chemical composition of oaks has shown a valuable tool for assessing variability (Dodd et al. 1993; Özcan 2007; Rafii et al. 1991, 1993), but few studies have been carried out on acorn chemical composition of Andalusia Holm oak acorns from a natural variability point of view (Rodríguez-Estévez et al. 2009; Valero et al. 2011). Chemosystematic differentiation based on differences in acorn fatty acid composition between Italian and Spanish populations of *Q. ilex* and *Q. rotundifolia* yielded partial separations of the individual populations (Rafii et al. 1991). Such separation using acorn fatty acids for native and hybrid populations of *Q. agrifolia* and *Q. wislizenii* was also achieved (Dodd et al. 1993). French Mediterranean evergreen oak populations were reported to be intermediate and heterogeneous for the fatty acid profiles between Spanish and Italian oak, suggesting a zone of hybridization (Rafii et al. 1993).

On the other hand, near-infrared reflectance spectroscopy (NIRS) is a highly valuable method that could be adapted for rapid measurements of acorn chemical composition traits. It is a non-destructive, simple, reliable, and accurate technique, allowing a complete chemical analysis in a single experiment, including fiber, starch, protein, and fatty acids (Jiang et al. 2007; Kim et al. 2007; Niewitetzki et al. 2010; Stuth et al. 2003; Takahashi et al. 2010). NIR spectroscopy is a rapid and cost-effective method for forestry-breeding programs and can be successfully applied to the selection of 'high yield' trees (Schimleck et al. 2000), differentiation of species (Adedipe et al. 2008), and oil content (Sousa-Correia et al. 2007).

The present work was undertaken to explore Holm oak morphometric and chemical composition variability. To accomplish this aim, morphometric and chemical composition in the mature acorns of 13 Andalusia Holm oak was performed. Data were subjected to clustering analysis and were correlated with geographical and climatic conditions. These data have provided biochemical markers that help establish phylogenic associations between populations and also reveal potentially account as an alternative source for dietary nutrition.

Materials and methods

Plant material

The present study was performed with thirteen populations distributed throughout the Andalusia (southern Spain) region (Fig. 1). Geographical coordinates, altitude, mean annual precipitation, and mean maximum and minimum temperatures corresponding to each surveyed area are presented in Table 1. The populations sampled in this study are distributed geographically in three major Andalusia regions: south (SAA, RG, RMA, SCA, and BCA), northeast (PCO, CCO, VJA, and TJA), and northwest (GSE, APS, CTH and CHU). The climate is diverse and characterized by a mean annual precipitation of 277.9 mm (SAA) to 1,263.6 mm (BCA), a mean winter precipitation of 78.9 mm (SAA) to 466.6 mm (SCA), a mean autumn precipitation of 103.4 mm (SAA) to 431.3 mm (RMA), a mean monthly maximum temperature of 23.1°C (TJA) to 26.8°C (PCO), and a mean monthly minimum temperature of 4.4–11.5°C (RG). Acorns were harvested during November 2007 from 15 trees per population. Trees 100 m apart from each other were randomly selected.

Fig. 1 Spanish (*inside square*) and Andalusia maps with *points* indicating the locations where the 13 populations were surveyed. For more data see Table 1

Acorn morphometry

Healthy acorns (20 units per tree) were used for size and weight determinations. Length and maximum diameter were measured in acorn digitized images by using appropriate software (Win Seed V3.10b, Regent Instruments Inc., Canada 1996). Fresh weight was determined in an analytical balance (Mettler AJ150).

Acorn flour preparation

Acorns (20 units per tree) were scarified with a knife by making transversal and longitudinal cuts, thus permitting the pericarp to be rapidly removed. Once peeled, cotyledons (including embryos) were triturated in a blade mill (Moulinex AD56 42) until a powder was obtained. Powder was weighed and dried in a forced-air drier at 45°C for 48 h, homogenized in a Waring Blender (LB20E, Waring Products, New Hartford, CT, USA), and sieved (1 mm), to obtain a fine homogeneous flour. Samples were stored at 4°C in a desiccator until NIRS analysis. To determine the pulp dry weight of acorn, we weighted 2 g of powder and then we dried at 103°C in a ventilation oven for 24 h. Finally, we weighted them in order to obtain the dried weight of the powder (dry matter).

NIRS-derived chemical data

Ash, total protein, fat, starch, sugars, palmitic acid (C16:0), stearic acid (C18:0), oleic acid (C18:1), linoleic acid (C18:2), digestibility, and energy were analyzed at the University of Cordoba NIRS Service (http://www.uco.es/servicios/scai/nir.html). The instrument employed for spectra collection was a Foss-NIRSystems 6500 System II spectrophotometer (Foss-NIRSystems Inc., Silver Spring, MD, USA) equipped with a transport module and autogain detectors; one from 400 to 1,100 nm (known as the VIS region) and another from 1,100 to 2,500 nm (known as the NIR region). Samples were scanned by reflectance using a 1/4 rectangular cup, and spectra were collected every 2 nm using WinISI software 1.50 (Infrasoft International, Port Matilda, PA, USA). Before recording the spectra, the samples were thermostated at 24°C. The reflectance (log 1/R) spectra were collected in duplicate, and the mean spectrum of each sample was used for data analysis to determine relationship to reference characteristics. The scatter correction of standard normal variant and detrend (SNV-D) was applied to spectra, along with several different maths treatments for derivative order number, gap, and first smoothing. The second smoothing was set at 1 to indicate no second smoothing. Principal component analysis was used to identify and remove spectral outliers. Samples having spectra with Mahalanobis distance (H) values greater than 3.0 were considered to be outliers and were removed from the file. NIRS spectra were matched with reference data and analyzed using modified partial least squares (MPLS) regression, maths treatment 2, 5, 5, 1, and WinISI IIv. 1.04 (Infrasoft International, Port Matilda, PA) software. The standard error of calibration (SEC), the coefficient of determination in calibration (RSQ), the standard error of cross validation (SECV), coefficient of determination of the cross validation (1-VR), and RPD [standard deviation of laboratory reference data divided by the standard error of prediction (SEP)] were determined and used to find calibration equations that were acceptable for quantitative prediction (Deaville et al. 2009; Foss North America 2008; Shenk and Westerhaus 1990; Williams and Sobering 1996).

NIRS analysis

The descriptive statistics including mean, standard deviation (SD), and range, for chemical composition of acorn flour samples used in the calibration and validation sets, are shown in Table 2. Mean values of the dry matter were

Table 1 Geographical and climatic dates of the 13 populations of *Quercus ilex* subsp. *ballota* included in this study

Population	Code	Latitude–Longitude	Altitude (m)	Annual precipitation (mm)	Mean winter precipitation (mm)	Mean autumn precipitation (mm)	Mean monthly maximum temperature (°C)	Mean monthly minimum temperature (°C)
Sierra de Alhamilla (Almería)	SAA	36°59′N 6°05′W	1,241	277.9	78.9	103.4	25.2	8.9
Benamahoma (Cádiz)	BCA	36°45′N 5°27′W	649	1,263.6	375.1	408.9	24.9	9.8
Puerto Serrano (Cádiz)	SCA	36°54′N 5°31′W	373	1,000.5	466.6	565,3	25.5	9.5
Cabra (Córdoba)	CCO	37°28′N 4°25′W	891	593.3	209.3	225.1	26.0	8.0
Pozoblanco (Córdoba)	PCO	38°22′N 4°54′W	618	612.6	211.7	229.9	26.8	8.1
Arenas del Rey (Granada)	RG	36°57′N 3°54′W	963	489.3	171.6	204.5	24.7	11.5
Calaña (Huelva)	CHU	37°38′N 6°51′W	184	635.7	235.8	261.9	26.5	10.5
Corteconcepción (Huelva)	CTH	37°54′N 6°30′W	364	845.6	310.2	348.4	26.3	9.5
Valdepeñas (Jaén)	VJA	37°30′N 3°56′W	1,229	556.3	189.2	216.4	24.8	5.9
Torres de Albanchez (Jaén)	TJA	38°17′N 2°36′W	643	795.4	277.4	256.8	23.1	4.4
Ronda (Málaga)	RMA	36°44′N 5°11′W	1,063	952.2	355.6	431.3	24.3	9.6
El Garrobo (Sevilla)	GSE	37°37′N 6°10′W	289	948.6	356.3	384.0	25.6	9.5
Almaden de la Plata (Sevilla)	APS	37°52′N 6°05′W	482	722.1	248.5	307.6	26.4	9.5

Dates were obtained from the Spanish agency of meteorology (AEMET) (http://www.aemet.es/es/portada)

Table 2 NIRS calibration statistics of acorn chemical composition used for predictions

		Content (%) of variables in the sample				NIRS performance data					
						Calibration		Cross validation		Calibration precision	
Variable	*N*	Mean	SD	Min	Max	SEC	RSQ	SECV	1-VR	RPD	CV
Ash	369	02.13	0.24	1.50	03.39	0.10	0.83	0.11	0.79	2.18	5.16
Total protein	356	05.80	1.00	2.85	08.98	0.20	0.96	0.23	0.95	4.35	3.90
Crude fat	366	09.52	2.00	4.12	15.05	0.23	0.99	0.26	0.98	7.77	2.80
Starch	227	61.51	2.27	56.23	67.88	1.32	0.66	1.38	0.63	1.65	2.24
Sugars	217	06.09	2.74	2.55	14.02	0.34	0.98	0.43	0.97	6.33	7.06
Palmitic acid	253	15.45	1.47	10.8	21.05	0.48	0.89	0.71	0.77	2.08	4.59
Stearic acid	253	02.97	0.58	1.60	04.47	0.25	0.81	0.33	0.69	1.78	11.11
Oleic acid	246	64.20	2.26	57.32	69.70	0.77	0.88	1.10	0.77	2.06	1.71
Linoleic acid	246	15.66	1.98	10.80	23.43	0.49	0.94	0.71	0.88	2.78	4.53
Digestibility	150	72.68	5.24	60.76	84.98	2.66	0.74	3.26	0.62	1.61	4.48
Energy	141	4,706	80.6	4,478	4,959	33.04	0.83	35.38	0.81	2.28	0.75

N no. of samples, *SD* standard deviation, *Min* minimum, *Max* maximum, *SEC* typical error in the calibration, *RSQ* coefficient of determination of the calibration, *SECV* typical error of cross validation, *1-VR* coefficient of determination of the cross validation, *RPD* standard deviation of laboratory reference data divided by the standard error of prediction (SEP), *CV* SECV/Mean × 100

Table 3 Acorn size and weight

Population code	Weight (g)	Length (mm)	Maximum diameter (mm)
SAA	2.84 ± 1.00 (35)^{a}	28.31 ± 3.59 (12)^{bc}	13.28 ± 1.22 (16)^{bc}
BCA	3.53 ± 1.23 (34)^{abc}	31.34 ± 3.00 (9)^{cde}	14.55 ± 1.78 (17)^{abcdef}
SCA	4.80 ± 1.43 (29)^{bcdef}	36.25 ± 2.52 (6)^{fgh}	15.15 ± 1.97 (17)^{bcde}
CCO	3.46 ± 1.31 (37)^{ab}	27.56 ± 2.36 (8)^{b}	14.21 ± 1.62 (16)^{abcde}
PCO	6.12 ± 1.86 (30)^{f}	37.21 ± 3.42 (9)^{ghi}	17.06 ± 2.44 (23)^{f}
RG	4.32 ± 1.48 (34)^{bcde}	32.03 ± 3.74 (11)^{de}	15.48 ± 1.95 (19)^{cdef}
CHU	5.48 ± 1.88 (34)^{ef}	36.65 ± 6.25 (17)^{fgh}	16.43 ± 4.85 (35)^{ef}
CTH	5.57 ± 1.26 (22)^{ef}	35.79 ± 4.00 (11)^{fgh}	16.39 ± 1.24 (18)^{ef}
VJA	2.72 ± 0.96 (35)^{a}	30.38 ± 3.03 (9)^{bcd}	12.91 ± 1.65 (15)^{ab}
TJA	4.96 ± 1.47 (29)^{cdef}	34.67 ± 2.63 (7)^{efg}	15.80 ± 1.78 (19)^{def}
RMA	2.41 ± 0.61 (25)^{a}	22.81 ± 0.81 (3)^{a}	12.35 ± 1.15 (13)^{a}
GSE	5.06 ± 1.22 (24)^{def}	40.35 ± 3.38 (8)^{i}	15.64 ± 1.66 (18)^{cdef}
APS	5.98 ± 1.31 (22)^{f}	39.57 ± 3.72 (9)^{hi}	16.39 ± 1.67 (20)^{ef}

The descriptive statistics are presented in terms of the Mean ± SD (CV). Mean values with the same letters indicate homogeneous subsets for $P \leq 0.05$ according to Duncan test

2.13% of ash, 5.80% of total protein, 9.52% of crude fat, 61.51% of starch, 6.09% of sugar, 15.45 of palmitic acid, 2.97% of stearic acid, 64.20% of oleic acid, 15.66% of linoleic acid, 72.68% of digestibility, and 4,706 kcal of energy.

The standard error of calibration (SEC), the coefficient of determination in calibration (RSQ), the standard error of cross validation (SECV), coefficient of determination of the cross validation (1-VR), and RPD [standard deviation of laboratory reference data divided by the standard error of prediction (SEP)] were determined. The statistics of calibrations and cross-validations are shown in Table 2. For each trait, we selected the highest RSQ and 1-VR (closest to 1) and the lowest SEC and SECV. However, the RPD statistic is considered to be a useful indicator for evaluating the prediction capacity of an equation to predict the component of interest. The higher the RPD values, the greater the probability of the model to predict new samples accurately. If the RPD is >3, the calibration models are considered as being acceptable for analytical purposes (Williams and Sobering 1996; Williams 2001). According to Williams and Sobering (1996), calibration models with an RPD value of between 2 and 3 indicate approximate quantitative predictions. In this context, the equation models for total protein, crude, fat, sugar, and linoleic acid had high values of RSQ (0.96, 0.99, 0.98, and 0.94, respectively), 1-VR (0.95, 0.98, 0.97, and 0.88, respectively), and RPD (4.35, 7.77, 6.33, and 2.78, respectively), indicating a close relationship between reference values and NIRS estimated values. However, the models for ash, palmitic acid, oleic acid, and energy had relatively high values of RSQ (0.83, 0.89, 0.88, and 0.83, respectively), low 1-VR (0.79, 0.77, 0.77, and 0.81, respectively), and RPD (2.18, 2.08, 2.06 and 2.28, respectively), indicating that those were not strong models, but that they might be useful for screening purposes. Finally, the models for starch, stearic acid, and digestibility presented a relatively low RSQ (0.66, 0.81, and 0.74, respectively), relatively low 1-VR (0.79, 0.77, 0.77, and 0.81, respectively), and low RPD (1.65, 1.78, and 1.61, respectively), indicating that they were of a lesser quality, but that they might be useful for screening ends. Nevertheless, it has been considered to be convenient not to consider them capable of improving the quality of the corresponding models. Only the models for fat, sugar, total protein, linoleic acid, ash, energy, palmitic, and oleic acid were utilized for the characterization of Andalusia Holm oak populations.

Statistical analysis

Normality of frequency distributions was tested by the Kolmogorov–Smirnov test. A one-way ANOVA was performed to test for differences in acorn morphometry and chemical data. Pearson's correlation was carried out to test for correlation between acorn size, chemical composition, geographical, and climate data. Ward's clustering method using squared Euclidian distances was used with the aiming of classifying and establishing distances between acorn morphometry and chemical data. Statistical analysis was conducted using SPSS v.8.0 software (SPSS Inc. Chicago IL, USA).

Results

Acorn morphometry and weight

Table 3 shows morphometry variations between seed collection sites. Statistic significant ($P < 0.05$) differences

Table 4 Correlationship between acorn size, weight, chemical determinations, and geographical (altitude, latitude, and longitude) and climatic (annual precipitation (1), mean monthly maximum (2) and minimum (3) temperature) data

	Acorn size			Geographical data			Climate data		
	Weight	Length	Diameter	Altitude	Latitude	Longitude	1	2	3
Weight	1.00	0.96 (0.00)	0.90 (0.00)	−0.82 (0.00)	0.71 (0.01)	0.27 (0.37)	0.09 (0.77)	0.53 (0.06)	0.12 (0.68)
Length	0.96 (0.00)	1.00	0.85 (0.00)	−0.80 (0.00)	0.60 (0.03)	0.29 (0.34)	0.15 (0.63)	0.45 (0.12)	0.08 (0.80)
Diameter	0.90 (0.00)	0.85 (0.00)	1.00	−0.74 (0.00)	0.68 (0.01)	0.20 (0.50)	0.09 (0.76)	0.46 (0.12)	0.11 (0.72)
Pulp D.M.	0.32 (0.29)	0.54 (0.05)	0.17 (0.57)	−0.22 (0.46)	0.25 (0.40)	0.24 (0.42)	−0.30 (0.30)	0.03 (0.92)	−0.02 (0.95)
Ash	0.88 (0.00)	0.84 (0.00)	0.84 (0.00)	−0.66 (0.01)	0.60 (0.03)	0.14 (0.64)	0.15 (0.62)	0.29 (0.34)	0.03 (0.93)
Crude protein	0.41 (0.17)	0.19 (0.53)	0.50 (0.08)	−0.01 (0.97)	0.52 (0.07)	−0.01 (0.99)	−0.37 (0.21)	0.37 (0.21)	−0.06 (0.84)
Crude fat	0.76 (0.00)	0.87 (0.00)	0.68 (0.01)	−0.80 (0.00)	0.43 (0.14)	0.40 (0.18)	0.23 (0.46)	0.40 (0.17)	0.22 (0.48)
Sugar	−0.44 (0.13)	−0.29 (0.34)	−0.40 (0.18)	0.36 (0.23)	−0.28 (0.35)	0.04 (0.90)	−0.26 (0.40)	−0.36 (0.22)	0.00 (1.00)
Palmitic acid	−0.23 (0.46)	−0.34 (0.26)	−0.12 (0.69)	0.35 (0.25)	0.15 (0.62)	−0.18 (0.55)	−0.48 (0.10)	0.02 (0.96)	−0.25 (0.40)
Oleic acid	0.30 (0.33)	0.53 (0.06)	0.33 (0.27)	−0.22 (0.48)	0.04 (0.89)	0.22 (0.46)	−0.13 (0.67)	−0.08 (0.79)	0.09 (0.77)
Linoleic acid	−0.10 (0.75)	−0.07 (0.82)	−0.22 (0.46)	−0.02 (0.96)	−0.11 (0.73)	0.10 (0.74)	0.01 (0.96)	0.05 (0.87)	0.14 (0.65)
Energy	−0.10 (0.76)	−0.44 (0.13)	−0.03 (0.92)	0.21 (0.50)	0.09 (0.77)	−0.37 (0.21)	0.18 (0.56)	−0.01 (0.97)	−0.16 (0.61)

Pearson's correlation coefficient is indicated with level of significance ($P \leq 0.05$)

were observed in acorn weight, length, and maximum diameter between thirteen seed lots. Acorn weight ranged from 2.41 g (RMA population) to 6.12 g (PCO population). Acorn length ranged from 40.35 mm (GSE population) to 22.81 mm (RMA population), while acorn maximum diameter ranged from 17.06 mm (PCO population) to 12.35 mm (RMA population). Coefficient of variance (CV) showed that all three acorn sizes were not equally variable between the populations. Acorn weight showed a CV of 30%, and, depending on the population, it ranged from 22% (APS population) to 37% (CCO population), while maximum diameter showed a CV of 18%, and, depending on the population, it ranged from 35% (CHU population) to 13% (RMA population). However, length showed a CV of 9%, and, depending on the population, it ranged from 17% (CHU population) to 3% (RMA population). Acorn weight, length and diameter showed statistic significant ($P < 0.01$) inverse correlation with altitude, and positively with latitude range of seed source (Table 4; Fig. 2). When the acorn size was correlated with latitude, two geographical groups were obtained, with the southern populations (RMA, SAA, CCO, BCA, and SCA) showing lower values than the northern ones (TJA, GSE, CTH, CHU, APS, and PCO) (Fig. 2a–c). Similar results were observed when the acorn size was correlated with altitude data of population localities (Fig. 2d–f). In this direction, populations located geographically in the northern areas would be expected to have higher acorn morphometry and weight than southern populations. Moreover, it would decrease with the altitude location.

Acorn chemical data

Table 5 shows chemical composition variations between seed collection sites. Statistic significant ($P < 0.05$) differences were observed in pulp dry weight, ash, total protein, sugar energy, fat, oleic acid, linoleic and palmitic acids between thirteen seed lots. Pulp dry weight ranged from 63.90% (RMA) to 83.64% (APS); ash ranged from 1.34% (CCO) to 2.02% (PCO); total protein ranged from 3.90% (RMA) to 5.94% (PCO); sugar ranged from 8.95% (PCO) to 12.47% (RG); energy ranged from 19,934.48 kJ/kg (SSA) to 20,398.66 kJ/kg (PCO); fat ranged from 9.14% (CCO) to 14.95% (GSE); oleic acid ranged from 64.98% (RMA) to 67.81% (GSE); while linoleic acid ranged from 12.15% (CCO) to 14.65% (CHU) and palmitic acid ranged from 15.62% (SCA) to 17.14% (CCO). Coefficient of variance (CV) showed that acorn composition did not vary equally between the populations. Crude protein presented a CV of 17%, and, depending on the population, it ranged from 10% (RMA) to 23% (GSE); sugar showed a CV of 16% and ranged from 7% (CCO) to 35% (PCO); fat showed a CV of 14% and ranged from 9% (GSE) to 23% (RMA). However, pulp dry weight, energy, palmitic and oleic acid were mostly homogenous between the populations (CVs < 10%). Table 4 shows the correlation between chemical composition and acorn size, climate conditions and geographical distribution. Pulp dry weight was positively correlated ($P \leq 0.05$) with length ($r = 0.54$). Ash content was positively correlated ($P < 0.05$) with weight ($r = 0.88$), length ($r = 0.84$), diameter ($r = 0.84$), latitude ($r = 0.60$) and negatively correlated with altitude

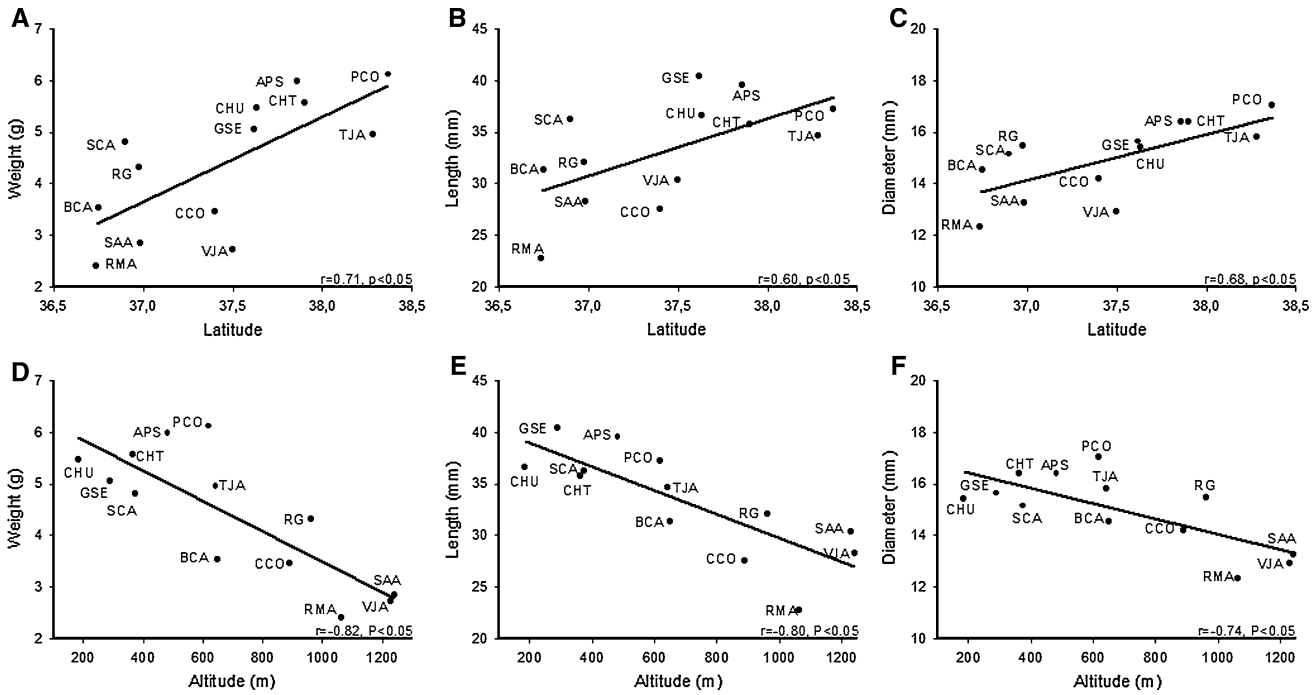


Fig. 2 Correlation between acorn morphometry and geographical data. **a** Weight and latitude, **b** length and latitude, **c** diameter and latitude, **d** weight and altitude, **e** length and altitude, and **f** diameter and altitude. Pearson's correlation coefficient is indicated with level of significance ($P \leq 0.05$). Population names are indicated in Table 1

($r = -0.66$) (Fig. 3a, b, c). Fat content was positively correlated ($P < 0.05$) with weight ($r = 0.76$), length ($r = 0.87$), diameter ($r = 0.68$) (Fig. 3d, e, f) and negatively correlated with altitude ($r = -0.80$).

Phylogenetic analysis

Morphometry and chemical composition data of thirteen seed lots were subjected to clustering analysis in order to establish groups of populations and distances among them. The tree groups obtained by Ward's clustering method using squared Euclidean distances can be observed in Fig. 4. Two clusters can be distinguished, with an average distance value (ADV) of 25. The first cluster, however, was then divided into two small clusters at ADV of 3.8. CTH, TJA, SCA, CHU, and PCO populations fell in one cluster with APS and GSE populations. The second principal was divided into two small clusters with an ADV of 5.3. SSA, VJA, and CCO fell in one cluster with BCA and RG (at ADV of 2.5), whereas RMA was the most distant inside this group (at ADV of 5.3). This analysis showed that northern populations (CTH, TJA, SCA, CHU, PCO, GSE, and APS) would be expected to have similar values in acorn morphometry and chemical composition. This group of population showed a tendency to present values highest in acorn weight (5.42 g vs. 3.21 g); acorn length (37.21 mm vs. 28.73 mm); acorn diameter (16.12 mm vs. 13.79 mm); ash (1.88% vs. 1.55%); protein (4.82% vs. 4.38%); fat (12.55% vs. 9.92%); and oleic acid (66.93% vs. 66.4%). These populations grow in intermediate climate with an annual precipitation ranging between 635.7 and 1,000.5 mm and an altitude ranging between 184 to 643 m. On the contrary, southern populations (SSA, VJA, CCO, BCA, RG, and RMA) showed a tendency to present values lowest in sugar (10.28% vs. 11.25%); energy (20,076.71 kJ/kg vs. 20,149.26 kJ/kg); palmitic (16.11% vs. 16.39%) and linoleic acids (13.07% vs. 13.11%). Of these populations, SSA, VJA, CCO, and RG populations grow in xeric climates with an annual precipitation ranging between 277.9 and 593.3 mm and an altitude ranging between 891 and 1,241 m.

Discussion

Morphometry characteristics of Holm oak acorn exhibited variation at the population level (Table 3). These variations in acorn size were influenced by altitudinal and latitudinal variation (Table 4; Fig. 2). Populations from the southern sites had the smallest average seed size, while northern ones had the longest and broadest acorns. Populations growing in lowest altitudes presented the longest and broadest acorns; on the contrary, populations growing highest altitudes presented the smallest and thinnest acorns.

Table 5 Chemical composition of the different population of *Quercus ilex* subsp. *ballota*

Population	Pulp DW (%)[a]	Ash (%)[b]	Crude proteins (%)[b]	Sugars (%)[b]	Energy (kJ/kg)
SAA	75.05 ± 4.88 (6)^{d}	1.48 ± 0.16 (10)^{ab}	4.43 ± 1.02 (23)^{abc}	11.47 ± 1.35 (12)^{cde}	19,934.48 ± 417.34 (2)^{ab}
BCA	67.54 ± 3.14 (6)^{abc}	1.60 ± 0.24 (14)^{bc}	4.48 ± 0.77 (17)^{abcd}	9.60 ± 1.72 (18)^{b}	20,250.59 ± 438.04 (2)^{bcd}
SCA	72.22 ± 3.52 (18)^{cd}	1.91 ± 0.24 (12)^{de}	4.13 ± 0.87 (21)^{ad}	9.22 ± 2.28 (25)^{b}	20,000.45 ± 465.56 (2)^{abc}
CCO	68.48 ± 4.36 (6)^{abc}	1.34 ± 0.23 (16)^{a}	4.30 ± 0.61 (13)^{abcd}	11.48 ± 0.76 (7)^{cde}	20,265.72 ± 230.52 (1)^{bcd}
PCO	66.10 ± 4.18 (17)^{ab}	2.02 ± 0.20 (9)^{e}	5.94 ± 0.88 (14)^{e}	8.95 ± 2.22 (35)^{a}	20,445.47 ± 364.54 (2)^{d}
RG	71.06 ± 3.11 (8)^{bcd}	1.77 ± 0.20 (11)^{cd}	4.33 ± 0.86 (16)^{abc}	12.47 ± 1.08 (9)^{e}	20,031.27 ± 394.35 (2)^{abcd}
CHU	75.29 ± 6.55 (8)^{d}	1.80 ± 0.22 (12)^{cd}	4.77 ± 0.81 (16)^{bcd}	9.75 ± 1.62 (17)^{bc}	19,997.34 ± 398.60 (2)^{abc}
CTH	71.66 ± 2.91 (4)^{cd}	1.87 ± 0.22 (11)^{de}	5.36 ± 0.89 (16)^{de}	11.84 ± 2.18 (18)^{de}	20,166.56 ± 355.72 (2)^{abcd}
VJA	76.14 ± 6.34 (6)^{d}	1.62 ± 0.25 (15)^{bc}	4.82 ± 0.75 (15)^{bcd}	12.23 ± 0.90 (7)^{e}	20,015.05 ± 441.02 (2)^{abc}
TJA	72.32 ± 6.75 (9)^{cd}	1.86 ± 0.18 (10)^{de}	4.56 ± 0.76 (16)^{abcd}	10.64 ± 1.63 (15)^{bcde}	20,154.97 ± 370.46 (2)^{abcd}
RMA	63.90 ± 3.51 (10)^{a}	1.49 ± 0.07 (5)^{ab}	3.92 ± 0.41 (10)^{a}	10.30 ± 1.63 (16)^{bcd}	20,398.66 ± 449.32 (2)^{cd}
GSE	74.55 ± 3.45 (2)^{d}	1.79 ± 0.16 (9)^{cd}	3.90 ± 0.93 (23)^{a}	11.43 ± 1.21 (15)^{cde}	20,015.65 ± 395.05 (2)^{a}
APS	83.64 ± 3.11 (4)^{e}	1.94 ± 0.25 (12)^{de}	5.14 ± 0.89 (20)^{cde}	10.14 ± 2.08 (21)^{bcd}	19,757.11 ± 230.02 (2)^{abc}

Population	Fat (%)[a]	Palmitic acid (%)[c]	Oleic acid (%)[c]	Linoleic acid (%)[c]
SAA	10.21 ± 1.17 (11)^{abc}	16.54 ± 1.17 (7)^{ab}	67.46 ± 0.89 (1)^{defg}	12.48 ± 1.41 (11)^{abc}
BCA	10.23 ± 1.62 (16)^{abc}	16.08 ± 1.18 (7)^{ab}	66.19 ± 1.71 (2)^{abcd}	13.12 ± 2.09 (15)^{bcd}
SCA	12.52 ± 1.43 (11)^{e}	15.62 ± 1.44 (9)^{a}	67.42 ± 1.02 (1)^{defg}	13.09 ± 1.45 (11)^{bcd}
CCO	9.14 ± 1.73 (19)^{a}	17.14 ± 1.05 (6)^{b}	65.95 ± 1.74 (2)^{abc}	12.15 ± 2.40 (19)^{ab}
PCO	12.69 ± 1.35 (11)^{e}	15.94 ± 0.80 (5)^{ab}	65.49 ± 1.49 (2)^{ab}	13.70 ± 1.76 (12)^{bcd}
RG	10.69 ± 1.03 (10)^{bcd}	16.33 ± 1.14 (7)^{ab}	67.56 ± 0.99 (1)^{efg}	13.18 ± 1.26 (9)^{bcd}
CHU	12.70 ± 1.57 (12)^{e}	16.13 ± 1.40 (9)^{ab}	66.12 ± 0.93 (1)^{abcd}	14.65 ± 1.41 (9)^{d}
CTH	11.81 ± 1.20 (10)^{de}	16.71 ± 1.15 (7)^{ab}	66.73 ± 1.10 (1)^{bcdef}	12.35 ± 1.40 (11)^{abc}
VJA	9.64 ± 1.67 (17)^{ab}	16.37 ± 1.01 (6)^{ab}	66.35 ± 1.66 (2)^{bcde}	14.06 ± 1.95 (13)^{cd}
TJA	11.28 ± 1.89 (17)^{cde}	16.38 ± 1.31 (8)^{ab}	67.20 ± 1.90 (2)^{cdefg}	12.50 ± 2.13 (17)^{abc}
RMA	9.63 ± 2.26 (23)^{ab}	15.89 ± 0.54 (3)^{ab}	64.98 ± 1.72 (2)^{a}	13.84 ± 1.87 (13)^{bcd}
GSE	14.95 ± 1.33 (9)^{f}	15.91 ± 1.51 (9)^{ab}	67.81 ± 1.69 (2)^{fg}	13.00 ± 2.24 (17)^{bcd}
APS	11.95 ± 1.58 (13)^{de}	16.10 ± 1.22 (7)^{ab}	67.75 ± 0.96 (1)^{fg}	12.26 ± 1.73 (14)^{abc}

The descriptive statistics are presented in terms of the Mean ± SD (CV). Mean values with the same letters indicate homogeneous subsets for $\alpha = 0.05$ according to Duncan test

[a] As percentage of fresh weight

[b] As percentage of pulp dry weight

[c] As percent of the total fatty acids

Populations variability with respect to acorn morphometry have earlier been reported in some *Quercus* spp. Seed length and weight exhibited significant inverse correlation with altitude of seed origin of *Quercus glauca* (Singh et al. 2010). Furthermore, acorn morphometry also has showed a positive correlation with high latitudes of seed origin of *Quercus suber* (Ramírez-Valiente et al. 2009). Aizen and Woodcock (1992) also found a positive relationship between acorn size and latitude in American *Quercus* species. Díaz-Fernández et al. (2004) established that the percentage of biennial acorns produced and their ripening time differed with latitudes across *Q. suber* populations. Bigger acorns lead to increased seedling growth, higher root and shoot ratios, which can improve the seedling performance by developing and maintaining deep and extensive root system accessing water from deep within the soil profile (Ramírez-Valiente et al. 2009; Singh et al. 2010). Selection for big acorns as a consequence of selection for tall seedlings is likely to be stronger in water-limiting environments, indeed, a positive relationship among acorn size; seedling height and summer survival have been reported for *Q. ilex* (Gómez 2004).

The development of a rapid, precise, robust, non-destructive, and economical method to evaluate quality components is of major interest to growers, processors, and breeders. NIR reflectance spectroscopy (NIRS) is routinely used for the prediction of quality traits in many crops. In this study, we have used this approach for determining the acorn chemical quality using an ample number of samples and chemical parameters. The calibration equations

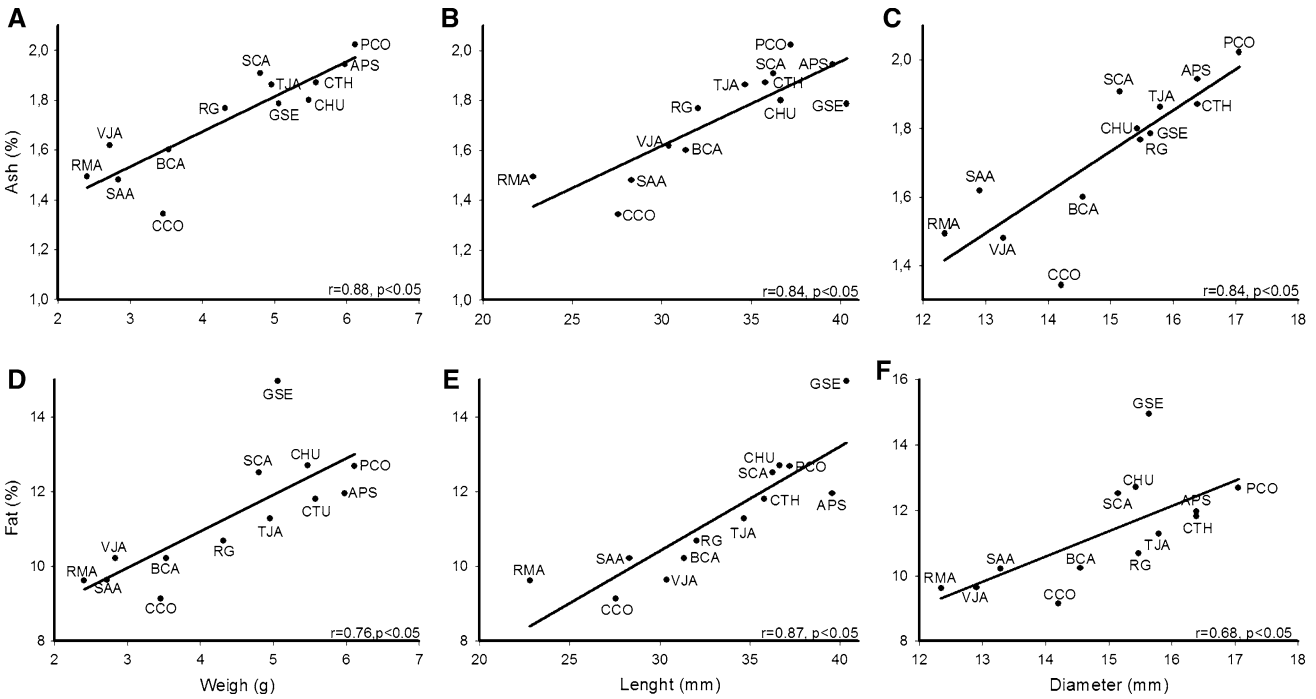


Fig. 3 Correlation between acorn morphometry and acorn chemical composition. **a** ash and weight, **b** ash and length, **c** ash and diameter, **d** crude fat and weight, **e** crude fat and length, **f** crude fat and diameter. Pearson's correlation coefficient is indicated with level of significance ($P \leq 0.05$). Population names are indicated in Table 1

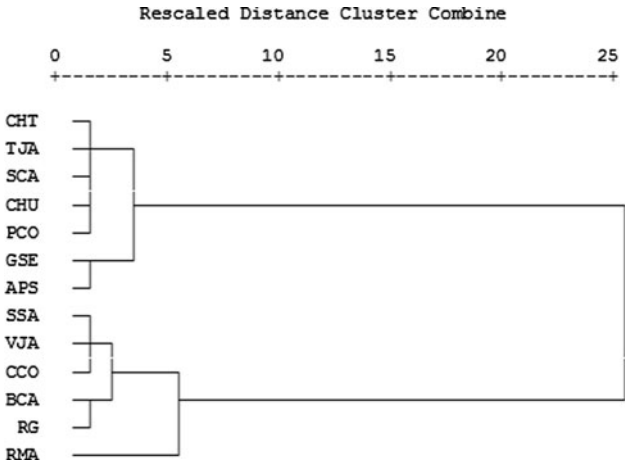


Fig. 4 Graphical representation of the distances of acorn morphometry and chemical compositions, established after Ward's clustering method analysis using squared Euclidian distance

employed for total protein, fat, sugar, and linoleic acid showed high values of RQS, low SEC and SECV, and high RPD, similar to those of other seeds and oilseed crops (Fontaine et al. 2002; Sousa-Correia et al. 2007; Baye and Becker 2004; Kim et al. 2007; Pazdernik et al. 1997), with higher RQS (>0.9) and lower SEC values and high RPD. However, the models for ash, palmitic acid, oleic acid, and energy had relatively high values indicating that these were not strong models, but that they might be useful for screening purposes. The models for starch, stearic acid, and digestibility did not have good statistics for RSQ, 1-VR, and RPD, indicating the threshold where an equation was not useful. However, CV values were similar or better than those presented in other crops using the Official Analytical Chemists Methods (AOAC 2000).

Acorn chemical composition of Holm oak has been little studied, and data in the literature are variable, especially when samples are small. In this direction, we have applied the NIRS technique to predict a complete chemical analysis of acorn flour in an attempt to compare, catalog, and characterize natural populations of Holm oak from the Andalusia region. The result indicated that there were statistic significant differences in acorn chemical composition between the different Holm oak populations (Table 5). Our results for ash, protein, and sugar content are in agreement with the findings on *Quercus ilex*. However, crude fat, oleic and palmitic acids contents showed relatively higher levels considering previous findings (Charef et al. 2008; León-Camacho et al. 2004; Rodríguez-Estévez et al. 2009; Vázquez 1998). Furthermore, linoleic acid percent showed relatively lower levels with respect to previous findings on *Quercus ilex* (Charef et al. 2008; León-Camacho et al. 2004; Rodríguez-Estévez et al. 2009; Tejerina et al. 2010). The energy content of the acorns ranged from

19,934.48 ± 417.34 to 20,398.66 ± 449.32 kJ/kg, while no other results are reported in the literature.

The relationship of geographical and climate data with acorn chemical composition of *Quercus ilex* has not been studied in much detail. However, in this study, the population located in the north (CHU, CTH, GSE, and APS) showed a tendency toward presenting high contents of ash, fat, oleic and linoleic acid, while the population located in the south (SAA, RG, BCA, and RMA) showed a tendency toward presenting low contents of sugar and palmitic acid. When the populations were classified according to climate data, those that grew in the driest site (VJA, SAA, CCO, and GR) gave high contents in total protein, sugar, and palmitic acid, the populations growing in mesic sites (RMA, BCA, and TJA) showed high contents in energy and linoleic acid, and, finally, the populations growing in an intermediate climate (CTH, CHU, APS, SCA, and GSE) presented a high content in ash, fat, and oleic acid. Similar geographical pattern differentiation has been observed when these populations were used to study the natural variability through acorn protein profiles (Valero et al. 2011). Variations in fatty acid levels were observed in different *Quercus* spp. taxas (Charef et al. 2008; Özcan 2007). Fatty acid profiles were affected by intra-annual and inter-annual time sampling (Tejerina et al. 2010). Differences in fatty acid were observed in different *Quercus* spp. populations (Dodd et al. 1993; Rafii et al. 1991, 1993). The results obtained did not allow one to decide whether the observed differences in total unsaturated acids are genetic or not; the data indicated that the differences were most likely related to the climatic or ecological variations in elevation (Rafii et al. 1991). Variations in the content of chemical composition in the acorns could be related to factors associated with (1) weather conditions, such as precipitation, temperature, and temperature oscillations and (2) the acorns physiological conditions such as ripening or germination and (3) sanitary conditions such as desiccation, dampening, rotting, and attacks from pests, such as *Curculio* spp., *Cydia* spp. or phytophages (Daza et al. 2007; Steele et al. 1996; Vázquez 1998). However, cross-pollination, that would enable gene flow, would prevent genetic divergence via drift and may account for the lack of spatial genetic structure of neutral markers (O'Brien et al. 2007; Soto et al. 2003). If selection is strong enough, local adaptation may evolve even in the presence of gene flow rates between populations situated at a geographical proximity, thus contributing to a stronger phenotypic similarity than what would be observed from the influence of environmental variations alone (Endler 1977), therefore representing a potential mechanism for explaining the structured distribution of phenotypic variation. The results of ongoing population genetic and phylogeographic studies of *Q. ilex* using nuclear and chloroplast DNA microsatellites may help to discern between these alternatives (Jiménez and Gil 2000).

Conclusions

In conclusion, based on averages, eastern populations of *Q. ilex* subsp. *ballota* are different from western populations in their acorn morphology and chemical composition, and the western populations display a more overall variation. Mother trees grown in native environments can be distinguished from each other using those variables. It is possible to restrict the material selection areas in accordance with the valuation parameter considered, although, in practically all the populations, trees with high–low values could be suitable for use in reforestation can be found. Because most populations are widely dispersed, the species is susceptible to a loss of genetic diversity, mainly among eastern populations. Additional studies on genetic diversity in *Q. ilex* at the molecular level would help to create efficient and effective conservation strategies for this species.

Acknowledgments Jose Valero was a recipient of an Alban Programme fellowship (I06D00010MX). This work is part of research projects financed by Spanish Ministry of Science and Innovation, being FEDER cofinanced: CGL2008-04503-C03-01/BOS, AGL2002-00530, and AGL2009-12243-C02-02.

References

Adedipe OE, Dawson-Andoh B, Slahor J, Osborn L (2008) Classification of red oak (*Quercus rubra*) and white oak (*Quercus alba*) wood using a near infrared spectrometer and soft independent modelling of class analogies. J Near Infrared Spectrosc 16:49–57

Aizen MA, Woodcock H (1992) Latitudinal trends in acorn size in eastern north-american species of *Quercus*. Can J Bot 70: 1218–1222

AOAC (2000) AOAC, Official methods of analysis, 17th edn. Association of Official Analytical Chemists, Gaithersburg

Baye T, Becker HC (2004) Analyzing seed weight, fatty acid composition, oil, and protein contents in *Vernonia galamensis* Germplasm by Near-Infrared Reflectance Spectroscopy. J Am Oil Chem Soc 81:641–645

Charef M, Yousfi M, Saidi M, Stocker P (2008) Determination of the fatty acid composition of acorn (*Quercus*), *Pistacia lentiscus* seeds growing in Algeria. J Am Oil Chem Soc 85:921–924

Daza A, López-Bote CJ, Barberan FAT, Espin JC, Carrasco CL, Olivares A, Rey AI (2007) Effect of Mediterranean forest parasite with *Curculio* sp on nutritional value of acorn for Iberian pig feeding and fat characteristics. Meat Sci 76:316–320

Deaville ER, Humphries DJ, Givens DI (2009) Whole crop cereals 2. Prediction of apparent digestibility and energy value from in vitro digestion techniques and near infrared reflectance spectroscopy and of chemical composition by near infrared reflectance spectroscopy. Anim Feed Sci Technol 149:114–124

Díaz-Fernández PM, Climent J, Gil L (2004) Biennial acorn maturation and its relationship with flowering phenology in

Iberian populations of *Quercus suber*. Trees-Struct Funct 18: 615–621

Dodd RS, Raffi ZA, Zavarin E (1993) Chemosystematic Variation in acorn fatty acids of Californian live Oaks *(Quercus agrifolia* and *Q. wislizenii)*. Biochem Syst Ecol 21:279–285

Echevarría-Zomeño S, Ariza D, Jorge I, Lenz C, Del Campo A, Jorrin JV, Navarro-Cerrillo RM (2009) Changes in the protein profile of *Quercus ilex* leaves in response to drought stress and recovery. Plant Physiol 166:233–245

Endler JA (1977) Geographic variation, speciation and clines. Princeton University Press, Princeton

Finkeldey R, Matyas G (2003) Genetic variation of oaks (*Quercus spp*.) in Switzerland. 3. Lack of impact of postglacial recolonization history on nuclear gene loci. Theor Appl Genet 106: 346–352

Fontaine J, Schirmer B, Horr J (2002) Near-infrared reflectance spectroscopy (NIRS) enables the fast and accurate prediction of essential amino acid contents. 2. Results for wheat, barley, corn, triticale, wheat bran/middlings, rice bran, and sorghum. J Agric Food Chem 50:3902–3911

Foss North America (2008) ISIscan and WinISI Software Training Class

Gandour M, Khouja ML, Toumi L, Triki S (2007) Morphological evaluation of cork oak (*Quercus suber*): Mediterranean provenance variability in Tunisia. Ann For Sci 64:549–555

Gea-Izquierdo G, Cañellas I, Montero G (2006) Acorn production in Spanish holm oak woodlands. Invest Agrar Sist Recur For 15:339–354

Gómez JM (2004) Bigger is not always better: conflicting selective pressures on seed size in *Quercus ilex*. Evolution 58:71–80

Gömöry D, Paule L, Shvadchak IM, Popescu F, Sulkowska M, Hynek V, Longauer R (2003) Spatial patterns of the genetic differentiation in European beech (*Fagus sylvatica* L.) at allozyme loci in the Carpathians and the adjacent regions. Silvae Genet 52:78–83

Jiang HY, Zhu YJ, Wei LM, Dai JR, Song TM, Yan YL, Chen SJ (2007) Analysis of protein, starch and oil content of single intact kernels by near infrared reflectance spectroscopy (NIRS) in maize (Zea mays L.). Plant Breed 126:492–497

Jiménez P, Gil L (2000) Conservación de recursos genéticos de los *Quercus* mediterráneos en España. Investigación agraria. Sistemas y recursos forestales. Fuera de Serie 2:103–115

Jiménez P, Agundez D, Alia R, Gil L (1999) Genetic variation in central and marginal populations of *Quercus suber* L. Silvae Genet 48:278–284

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

Jorge I, Navarro-Cerrillo 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(Suppl 1):S207–S214

Kim KS, Park SH, Choung MG (2007) Nondestructive determination of oil content and fatty acid composition in perilla seeds by near-infrared spectroscopy. J Agric Food Chem 55:1679–1685

León-Camacho M, Viera-Alcaide I, Vicario IM (2004) Acorn (*Quercus* spp.) fruits lipids: saponifiable and unsaponifiable fractions: a detailed study. J Am Oil Chem Soc 81:447–453

Lumaret R, Jabbour-Zahab R (2009) Ancient and current gene flow between two distantly related Mediterranean oak species, *Quercus suber* and *Q. ilex*. Ann Bot Lond 104:725–736

Lumaret R, Mir C, Michaud H, Raynal V (2002) Phylogeographical variation of chloroplast DNA in Holm oak (*Quercus ilex* L.). Mol Ecol 11:2327–2336

Lumaret R, Tryphon-Dionnet M, Michaud H, Sanuy A, Ipotesi E, Born C, Mir C (2005) Phylogeographical variation of chloroplast DNA in cork oak (*Quercus suber*). Ann Bot Lond 96: 853–861

Martín MA, Navarro-Cerrillo RM, Ortega P, Álvarez JB (2009) Cotyledon proteins as genetic markers in sweet holm oak. For Sci 55:526–531

Michaud H, Toumi L, Lumaret R, Li TX, Romane F, Digiusto F (1995) Effect of geographical discontinuity on genetic-variation *in Quercus ilex* (holm oak) evidence from enzyme polymorphism. Heredity 74:590–606

Niewitetzki O, Tillmann P, Becker HC, Mollers C (2010) A new Near-Infrared Reflectance Spectroscopy Method for high-throughput analysis of oleic acid and linoleic acid content of single seeds in oilseed rape (*Brassica napus* L.). J Agric Food Chem 58:94–100

O'Brien EK, Mazanec RA, Krauss SL (2007) Provenance variation of ecologically important traits of forest trees: implications for restoration. Appl Ecol 44:583–593

Özcan T (2007) Characterization of Turkish *Quercus L.* taxa based on fatty acid compositions of the acorns. J Am Oil Chem Soc 84:653–662

Pausas J, Bladél C, Valdecantos A, Seva J, Fuentes D, Alloza J, Vilagrosa A, Bautista S, Cortina J, Vallejo R (2004) Pines and oaks in the restoration of Mediterranean landscapes of Spain: New perspectives for an old practice—a review. Plant Ecol 171:209–220

Pazdernik DL, Killam AS, Orf JH (1997) Analysis of amino acid and fatty acid composition in soybean seed, using near-infrared reflectance spectroscopy. Agron J 89:679–685

Petit RJ, Brewer S, Bordács S, Burg K, Cheddadi R, Coart E, Cottrell J, Csaikl UM, van Dam B, Deans JD, Espinel S, Fineschi S, Finkeldey R, Glaz I, Goicoechea PG, Jensen JS, König AO, Lowe AJ, Madsen SF, Mátyás G, Munro RC, Popescu F, Slade D, Tabbener H, de Vries SGM, Ziegenhagen B, de Beaulieu J-L, Kremer A (2002) Identification of refugia and post-glacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence. For Ecol Manag 156:49–74

Rafii ZA, Zavarin E, Pelleau Y (1991) Chemosystematic differentiation of *Quercus ilex* and *Q. rotundifolia* based on acorn fatty-acids. Biochem Syst Ecol 19:163–166

Rafii ZA, Dodd RS, Pelleau Y (1993) Biochemical diversity and systematics of mediterranean evergreen oak from South East France. Biochem Syst Ecol 21:687–694

Ramírez-Valiente JA, Valladares F, Gil L, Aranda I (2009) Population differences in juvenile survival under increasing drought are mediated by seed size in cork oak (*Quercus suber* L.). For Ecol Manag 257:1676–1683

Rodríguez-Estévez V, García A, Gómez AG (2009) Characteristics of the acorns selected by free range Iberian pigs during the montanera season. Livest Sci 122:169–176

Schimleck LR, Raymond CA, Beadle CL, Downes GM, Kube PD, French J (2000) Applications of NIR spectroscopy to forest research. Appita J 53:458–464

Shenk JS, Westerhaus MO (1990) In: Biston R, Bartiaux-Thill N (eds) Proceedings of the 3rd international conference on near infrared spectroscopy. Agricultural Research Centre Publishing, Gembloux, Belgium, p 649

Singh B, Prakash SK, Bhatt BP (2010) Provenance variation in seed and seedlings attributes of *Quercus glauca* Thunb in Garhwal Himalaya. India Dendrobiol 63:59–63

Soto A, Lorenzo Z, Gil L (2003) Nuclear microsatellite markers for the identification of *Quercus ilex* L. and Q. suber L. hybrids. Silvae Genet 52:63–66

Soto E, Hoz L, Ordóñez JA, Hierro E, Herranz B, López-Bote C, Cambero MI (2008) Impact of feeding and rearing systems of

Iberian pigs on volatile profile and sensory characteristics of dry-cured loin. Meat Sci 79:666–676

Sousa-Correia C, Alves A, Rodrigues JC, Ferreira-Dias S, Abreu JM, Maxted N, Ford-Lloyd B, Schwanninger M (2007) Oil content estimation of individual kernels of *Quercus ilex* subsp. *rotundifolia* [(Lam) O. Schwarz] acorns by Fourier transform near infrared spectroscopy and partial least squares regression. J Near Infrared Spectrosc 15:247–260

Steele MA, Hadj-Chikh LZ, Hazeltine J (1996) Caching and feeding decisions by *Sciurus carolinensis*: responses to weevil-infested acorns. J Mammal 77:305–314

Stuth J, Jama A, Tolleson D (2003) Direct and indirect means of predicting forage quality through near infrared reflectance spectroscopy. Field Crop Res 84:45–56

Takahashi A, Ichihara Y, Isagi Y, Shimada T (2010) Effects of acorn tannin content on infection by the fungus *Ciboria batschiana*. For Pathol 40:96–99

Tejerina D, García-Torres S, Cabeza de Vaca M, Vázquez FM, Cava R (2010) Acorns (*Quercus rotundifolia* Lam.) and grass as natural sources of antioxidants and fatty acids in the "montanera" feeding of Iberian pig: Intra- and inter-annual variations. Food Chem 124:997–1004

Toumi L, Lumaret R (1998) Allozyme variation in cork oak (*Quercus suber* L.): the role of phylogeography and genetic introgression by other Mediterranean oak species and human activities. Theor Appl Genet 97:647–656

Valero GJ, Valledor L, Navarro RM, Gil E, Jorrín-Novo JV (2011) Studies of variability in Holm oak (*Quercus ilex* subsp. *ballota* [Desf.] Samp.) through acorn protein profile analysis. J Proteomics 74(2011):1244–1255

Vázquez FM (1998) Semillas del género *Quercus* L. (Biología, Ecología y Manejo). Servicio de Investigación y Desarrollo Tecnológico, Consejería de Agricultura y Comercio. Junta de Extremadura, Mérida (España), pp 234

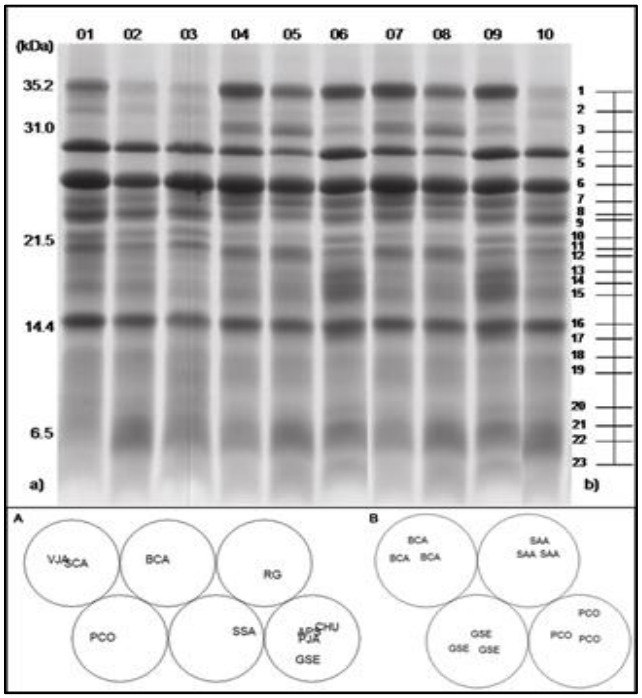
Williams P (2001) Implementation of near infrared technology. In: Williams PC, Norris KH (eds) Near-infrared technology in the agricultural and food industries. American Association of Cereal Chemists, St. Paul, pp 145–171

Williams P, Sobering DC (1996) How do we do it: a brief summary of the methods we use in developing near infrared calibrations. In: Davies AMC, Williams PC (eds) Near infrared spectroscopy. The future waves. NIR Publications, Montreal, pp 185–188

4. CAPÍTULO II

El estudio de la variabilidad poblacional de la encina a través del análisis del perfil proteico de las bellotas.

Studies of variability in Holm oak (*Quercus ilex* subsp. *ballota* [Desf.] Samp.) through acorn protein profile analysis.



JOSÉ VALERO GALVÁN, LUIS VALLEDOR, RAFAEL Mª NAVARRO CERRILLO, JESUS V. JORRÍN-NOVO. 2011. Studies of variability in Holm oak (*Quercus ilex* subsp. *ballota* (Desf.) Samp.) through proteomic analysis of acorns. Journal of Proteomics 74:1244-1255.

JOURNAL OF PROTEOMICS 74 (2011) 1244–1255

available at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/jprot

Studies of variability in Holm oak (*Quercus ilex* subsp. *ballota* [Desf.] Samp.) through acorn protein profile analysis

José Valero Galván[a,*], *Luis Valledor*[b], *Rafael M*[a]*. Navarro Cerrillo*[c], *Eustaquio Gil Pelegrín*[d], *Jesus V. Jorrín-Novo*[a]

[a]*Agricultural and Plant Biochemistry and Proteomics Research Group, Dept. of Biochemistry and Molecular Biology, University of Córdoba, Córdoba, Spain*
[b]*Dept. of Molecular Systems Biology (MOSYS), University of Vienna, Vienna, Austria*
[c]*Department of Forestry Engineering, ETSIAM, University of Córdoba, Córdoba, Spain*
[d]*Forestry Resources Department, Agrifood Research and Technology, Centre of Aragon, Zaragoza, Spain*

ARTICLE INFO

Article history:
Received 21 February 2011
Accepted 3 May 2011
Available online 12 May 2011

Keywords:
Acorn proteomics
Quercus ilex proteomics
Holm oak variability

ABSTRACT

Studies of variability in Holm oak (*Quercus ilex* subsp. *ballota* [Desf.] Samp.), the dominant tree species in the typical Mediterranean forest, have been carried out by using electrophoresis-based proteomic analysis of acorns. Ten populations distributed throughout the Andalusia region have been surveyed. Acorns were sampled from individual trees and proteins extracted from seed flour by using the TCA–acetone precipitation protocol. Extracts were subjected to SDS-PAGE and 2-DE for protein separation, gel images captured, spot or bands quantified, and subjected to statistical analysis (ANOVA, SOM and clustering). Variable bands or spots among populations were subjected to MALDI-TOF/TOF and LC-MS/MS for identification. The protein yield of the used protocol varied among populations, and it was in the 2.92–5.92 mg/g dry weight range. A total of 23 bands were resolved by SDS-PAGE in the 3–35 kDa Mr range, with 8 and 12, out of the total, showing respectively qualitative and quantitative statistically significant differences among populations. Data allowed grouping populations, with groups being correlated according to geographical location and climate conditions, to northern and southern, as well as the discrimination of both mesic and xeric groups. Acorn flour extracts from the most distant populations were analyzed by 2-DE, and 56 differential spots were proposed as markers of variability. Identified proteins were classified into two principal categories; storage and stress/defense protein. Besides providing the first reference map of mature acorn seeds, the use of SDS-PAGE and proteomics in characterizing natural biodiversity in forest trees will be discussed.

Abbreviations: DW, dry weight; SOM, Kohonen's self organizing maps; SSA, Sierra de Alhamilla (Almería); RG, Arenas del Rey (Granada); VJA, Valdepeñas (Jaén); PCO, Pozoblanco (Córdoba); CHU, Calaña (Huelva); APS, Almadén de la Plata (Sevilla); PJA, Pozoalcón (Jaén); GSE, El Garrobo (Sevilla); SCA, Puerto Serrano (Cádiz); BCA, Benamahoma (Cádiz).

* *Corresponding author at:* Departamento de Bioquímica y Biología Molecular, Universidad de Córdoba, Campus de Rabanales, Ed. Severo Ochoa, E-14071 Córdoba, Spain. Fax: +34 957 218439.
E-mail address: bb2vagaj@uco.es (J.V. Galván).
URLs: http://www.uco.es/investiga/grupos/probiveag/ (J.V. Galván), http://www.univie.ac.at/mosys/ (L. Valledor).

1874-3919/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jprot.2011.05.003

1. Introduction

Holm oak (*Quercus ilex* subsp. *ballota* [Desf.] Samp.) is the dominant tree species in natural forest ecosystems over large areas of the western Mediterranean Basin [1]. In Spain, it is widely used for conservation and forestry practices, covering an approximate surface of 2,039,563 ha, with estimated values of over 120 million of Euros in fruit acorn production per year [2]. The Holm oak acorns are a major component in the feeding systems of many Mediterranean wild and livestock species. Moreover, it is also the basic feed ingredient for domestically-bred high quality meat pigs [2].

Nowadays, forest restoration and reforestation are high priority objectives, and Holm oak is acquiring more interest for Mediterranean forestry, together with other evergreen species, namely *Quercus suber* L., *Ceratonia siliqua* L., and *Olea europaea* subsp. *sylvestris* Brot. [3–5]. Thus, this is increasing the demand for Holm oak seedlings and favoring their nursery production. As natural, non-domesticated plant species, with great plasticity and phenotypic variability, a key challenge prior to massive clonal propagation is the establishment of techniques for the cataloging and selection of genotypes among provenances with high survival percentage and productivity under specific environmental conditions. The existence of a high variability and polymorphism in *Quercus* spp. has been concluded from data obtained by using anatomical, chemical, and enzyme activity indicators, as well as DNA marker techniques, the latter also used for introgression and phylogeographical studies [6–12]. In Spain, Holm oak populations show a prolonged isolation and complex geographic patterns of genetic variation, being native to and distributed through diverse forest areas [13,14]. Some populations have survived under the extreme dry climate conditions prevailing in southern Spain, so these individuals are expected to present efficient mechanisms of adaptability to these conditions. A proteomic research program is being carried out with *Q. ilex* subsp. *ballota* in order to study variability of Holm oak populations in Andalusia [15–17].

The high variability in the leaf 2-DE protein profile, even within the same tree and depending on several factors (i.e. leaf orientation, crown position, sampling time, and develop stage), does not allow to clearly discriminate between trees or populations, with a low number of differential spots found [17]. According to these findings, we initiated a study directed target to Holm oak seed proteome, expected to be less variable than the leaf proteome. Seed protein polymorphism has been used in genetic studies in many species [18–20], but few studies have been carried out on seed protein analysis of *Quercus* species [21–23]. Genes, visualized as precise protein bands or spots, which reflect physiological status are good candidates for assessing variability, and establishing genetic distances and phylogenetic relationships between different species and individuals, being used in studies with herbaceous species [24–31]. In forest tree, mainly in *Pinus* spp., 2-DE analysis has shown to be a valuable tool providing informative markers for assessing variability and the establishment of linkage maps [24,32,33].

The present work was undertaken to explore Holm oak acorn protein profile, and to exploit it to study variability. To accomplish this aim, SDS-PAGE and 2-DE acorn protein analyses were performed. Data were subjected to multivariate statistical and clustering analysis to establish groups of populations and correlation with geographical and climatic conditions. Finally, MS analysis and protein identification allowed discussion of differences from a biological point of view.

2. Materials and methods

2.1. Acorn sampling and flour powder preparation

Holm oak mature acorns were collected from ten populations distributed throughout Andalusia region (southern Spain) (Supplementary Fig. 1). Geographical coordinates, altitude, mean annual precipitation, and mean maximum and minimum temperatures corresponding to each population are indicated in Supplementary Table 1. Undamaged, homogeneous mature acorns (Supplementary Fig. 2) were collected from ten individual trees (3 kg per tree) for each population (final N = 100), during November 2007. Once harvested, acorns were put in airtight polyethylene bag and stored at 4 ± 1 °C during no more of 12 h. Immediately after arriving to the Laboratory, and previously to protein extraction, a pool of twenty acorns per tree, were scarified with a knife by making transversal and longitudinal cuts, thus permitting the pericarp to be rapidly removed. Once peeled out, their embryos (including cotyledons) were triturated in a blade mill (Moulinex AD56 42) until a fine powder was obtained. Powder (Supplementary Fig. 2) was weighted and stored in a desiccator at 4 ± 1 °C, until protein extraction. For SDS-PAGE protein extraction, flour of ten independent samples per population were used; whereas for 2-DE protein extraction flour representing all studied accessions of each population was crushed together and its proteins were extracted.

2.2. Protein extraction

Proteins were extracted from 100 mg of flour powder by using the TCA–acetone protocol [34], as reported in [17]. The final pellet was suspended in 100 μL of 9 M urea, 4% CHAPS, 0.5% Triton X100, and 100 mM DTT. Insoluble material was eliminated by centrifugation, and the protein content in the supernatant was quantified by the method of Bradford [35], using bovine albumin as standard. Samples were stored at −70 °C until isoelectro focusing or 1-D electrophoresis.

2.3. Gel electrophoresis

Protein extracts, 70 μg BSA equivalent, were subjected to SDS-PAGE electrophoresis [36] on 15% polyacrylamide gels by using PROTEAN II (17 cm) cells (Bio-Rad, Hercules, USA). Gels were stained with colloidal Coomassie [37]. Images were digitized with a GS-800 Calibrated Densitometer (Bio-Rad, Hercules, USA) and analyzed with Quantity One software (Bio-Rad, Hercules, USA).

In preliminary 2-DE experiments with small (7 cm), 3–10 IPG strips, it was observed that most of the spots were

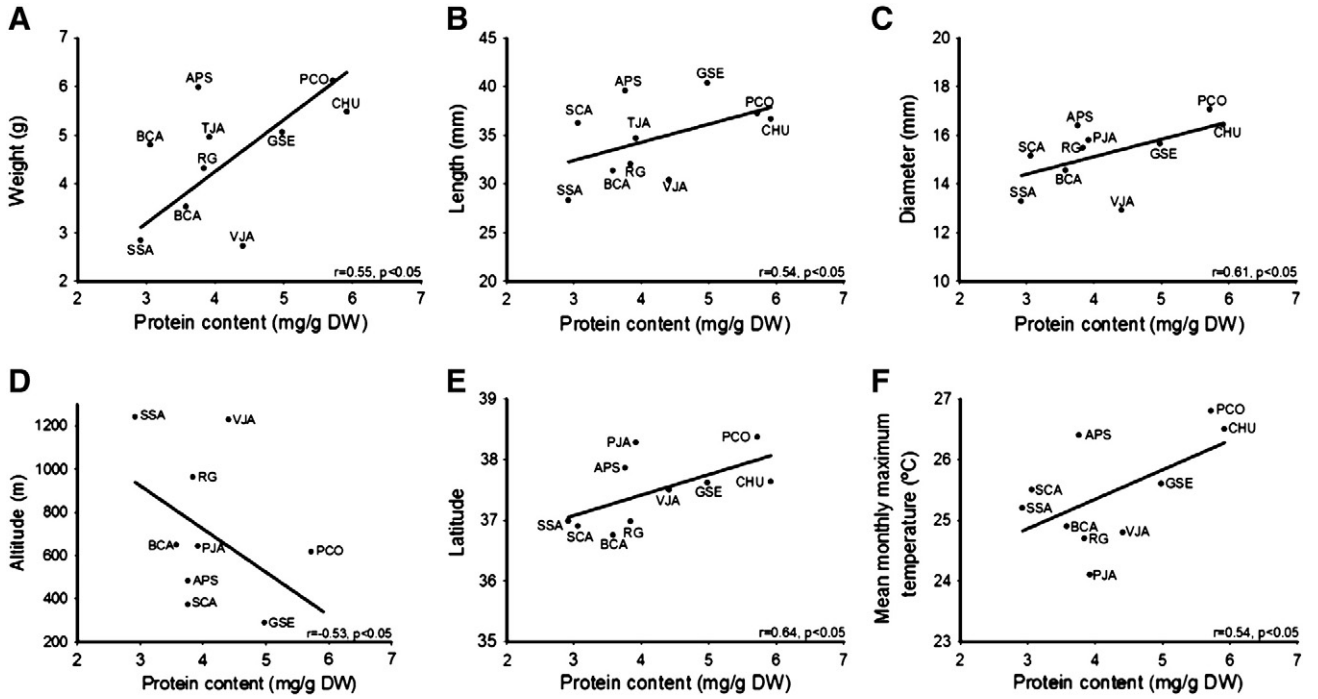


Fig. 1 – Correlation between protein content (X axes) and morphometric (acorn weight, length and diameter), or geographical (latitude and altitude) and climate (average monthly maximum temperature) parameters (Y axes).

concentrated in the pH 5–8 region (Supplementary Fig. 3). Considering this, and in order to increase the resolution by preventing protein co-migration as far as possible, IEF was performed using 5–8 pH linear range strips. IPG strips (17 cm, 5–8 pH linear gradient; Bio-Rad, Hercules, USA) were passively rehydrated for 12 h with 250 μL of rehydration buffer containing 300 μg protein [38]. The strips were loaded onto a Protean IEF Cell (Bio-Rad, Hercules, USA), and electrofocussed at 20 °C using a gradually increasing voltage: 0–250 V, 20 min, 250–10,000 V, 150 min and to 40,000 Vh with a maximum voltage of 10,000 V. After IEF, IPG strips were immediately equilibrated according to [38]. The second dimension was performed on 13% polyacrylamide gels using the Protean Dodeca Cell (Bio-Rad, Hercules, USA). Gels were run at 150 constant volts until the dye reached the bottom of the gel. Gels were stained with colloidal Coomassie [37]. Images were digitized with a GS-800 Calibrated Densitometer (Bio-Rad, Hercules, USA), and analyzed with PD-Quest software (Bio-Rad, Hercules, USA), using tenfold over background as a minimum criterion to assess the presence/absence. The analysis was re-evaluated by visual inspection, focusing on those spots which were present in all three biological replicates for each sample.

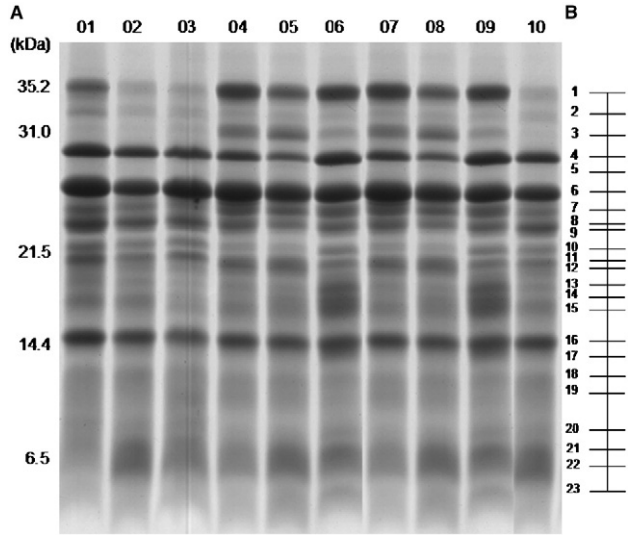


Fig. 2 – SDS-PAGE protein profile protein extracts from seed flour of *Quercus ilex* subsp. *ballota* (ten biological replicates corresponding to the CHU population). On the left, position of the molecular weight protein markers; on the right, diagram of the 23 resolved bands.

2.4. *Phylogenetic and statistical analyses*

Prior to statistical and phylogenetic analyses volumes of each band or spot were normalized, respectively, considering the total volume of bands detected in corresponding lane (SDS-PAGE), or the total volume of the valid spots in each gel (2-DE). Normalized values were cubic root transformed to reduce dependence between abundance and standard deviation. Differentially abundant bands or spots were defined applying ANOVA, followed by Duncan's test. Band and spots were clustered employing Ward's clustering method using Squared Euclidian distances by Unscrambler software (http://www.

camo.com/). Kohonen's self organizing maps (SOM), a type of artificial neural network [39], was applied employing R environment v2.11 software [40]. Furthermore, the correlation between inter-population distances found by multivariate statistical methods and its geographical coordinates or climate parameters was contrasted by employing a Mantel test.

2.5. *In-gel digestion and MS/MS analysis*

Bands were manually excised by using a scalpel; whereas spots were automatically excised by using the Investigator ProPic station (Genomics Solutions). In gel digestion and MS analysis were performed as reported in [41]. A combined search (PMF plus MS/MS) was carried out by using GPS Explorer™ software v 3.5 (Applied Biosystems) over non-redundant NCBI databases using the MASCOT search engine (Matrix Science Ltd., London; http://www.matrixscience.com). The following parameters were allowed: taxonomy restrictions to *Viridiplantae*, one missed cleavage, 100 ppm mass tolerance in MS and 0.5 Da for MS/MS 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, and confirmed by the accurate overlapping of the matched peptides with the major peaks of the mass spectrum. Those not identified proteins by MALDI-TOF/TOF were analyzed by LC-MS/MS. Samples were analyzed using a Finnigan Surveyor HPLC system coupled with an LTQ-Orbitrap XL (Thermo Fisher Scientific, USA). The peptide mixtures were loaded onto a C18 column (BioBasic C18, 75 μm × 10 cm, particle size 5 μm, Thermo Fisher Scientific, USA) to achieve an elution flow of 0.3 μL min^{-1} using an autosampler. Peptides were eluted during a 0–40% gradient (Buffer A, 0.1% formic acid, and 5% ACN; Buffer B, 0.1% formic acid and 95% ACN) over 60 min and detected in LTQ-Orbitrap equipped with a nanoelectrospray ion source (nESI). The general mass spectrometric conditions were as follows: spray voltage, 2 kV; no sheath and auxiliary gas flow; ion transfer tube temperature was of 170 °C; the capillary voltage was set to 39 V and tube lens value was 175. The Orbitrap was operated at resolution settings of 30,000 and full-scan spectra over the *m/z* range 400–1500 were acquired in the positive ion mode. The mass spectrometry method used was double play with CID activation, the normalized collision energy was 35, the activation time was 35 ms, the isolation width was 3.0 and the activation Q value was 0.25. Fragmented more intense ion was 5 and dynamic exclusion was set to on. Acquired data were analyzed with Proteome Discoverer 1.0 software (Thermo Fisher Scientific, USA). The database used was specific for Viridiplantae (02032010), and peptide search was carried out with a maximum of 2 missed cleavage sites. Cysteine carbamidomethylation and methionine oxidation were established as fixed and dynamic modification, respectively. Where appropriate, proteins identified with a false discover rate (FDR) <0.05 and further evaluated by the comparison with their calculated mass using the experimental values obtained from SDS-PAGE.

3. Results and discussion

The economic and ecological importance of Holm oak in the Andalusia region, and the practical consequences of large-scale surveys using molecular markers for future applications (such as tracing the geographic origin of seed lots, identifying the seed source of plantations, or conservation purposes), would be enough to justify the present work. Up to now, few studies of Holm oak variability based on proteomic approach have been displayed [16]. The aim of this work was to study natural variability in ten Holm oak Andalusia populations by analyzing the acorn protein profile through SDS-PAGE and 2-DE coupled to mass spectrometry. The populations sampled in this study are distributed geographically in three major Andalusia regions: South (SSA, RG, SCA and BCA), North-East (PCO, VJA and PJA), and North-West (GSE, APS, and CHU). Climate and physiographic variability are contrasted on these areas (Supplementary Fig. 1, Supplementary Table 1).

Populations were compared in terms of protein yield, number and intensity of bands (SDS-PAGE) or spots (2-DE). Proteins were extracted from acorn flour by using the TCA–acetone precipitation method. Statistically significant variations were observed in protein yield, and depending on the population, ranged from 2.92 (SSA population) to 5.92 (CHU population) mg/g DW (Table 1). These values were in the range of those reported for different *Quercus* spp. [42] and represented just a small fraction (5–10%) of the total protein content in acorns as determined by NIRS, both data being correlated ($r=0.877$, $P<0.05$) [43]. Such variability in the protein content could reflect environmental conditions as has been shown in other *Quercus* species [44]. The present study showed that the protein content was significantly correlated with acorn weight, length and diameter, latitude and altitude date, and average monthly maximum temperature (Fig. 1). According to our data, higher protein content would be expected in bigger than smaller acorns. In this direction, populations located

Table 1 – Protein content and number of bands resolved by SDS-PAGE. For details on electrophoresis see Fig. 2. Data with different superscript letters correspond to statistically significant differences according to Duncan's test (α = 0.5).

Population	Protein content (mg/g DW)	Number of bands		
		Total	Absent[1]	Variable[2]
Almería (SSA)	2.92 ± 1.25^{a}	21	2	15
Granada (RG)	3.84 ± 0.45^{abc}	22	1	17
Jaén 1 (VJA)	4.41 ± 0.88^{abcd}	21	2	15
Córdoba (PCO)	5.72 ± 0.93^{cd}	21	2	15
Huelva (CHU)	5.92 ± 2.20^{d}	20	3	13
Sevilla 1 (APS)	4.98 ± 0.37^{bcd}	21	2	15
Jaén 2 (PJA)	3.92 ± 1.09^{abc}	21	2	15
Sevilla 2 (GSE)	3.76 ± 0.00^{ab}	17	6	7
Cádiz 2 (SCA)	3.06 ± 0.40^{ab}	22	1	17
Cádiz 1 (BCA)	3.58 ± 2.05^{ab}	21	2	15

[1]Out of the 23 different bands resolved by SDS-PAGE.
[2]Present in all the populations but with statistically significant differences in intensity.

geographically in the northern areas would be expected to have higher protein content than southern populations. Moreover, it would decrease with the altitude location; and finally, the protein content would increase with the mean monthly maximum temperature where the populations are growing (Fig. 1).

A total of 23 bands were resolved by SDS-PAGE in the 3–35 kDa Mr range (Fig. 2; Table 1). Mean relative intensity values for each band, as well as SD, are provided in Supplementary Table 2. Qualitative (band intensity below detection limit in at least one population) and quantitative (band common to all populations but with different intensities) statistically significant differences (Duncan's test, $P<0.05$) were observed in 20 bands. Band numbers 5, 9, 12, 13, 14, 17, 20, and 21 (Supplementary Table 2; Fig. 2) showed qualitative differences; while, bands 1, 2, 3, 4, 7, 8, 10, 15, 18, 19, 22, and 23, reflected quantitative differences. When analyzing the glutelin fraction in Holm oak acorn extracts from eight populations located in two provinces, Córdoba and Sevilla, 32 polymorphic bands of the 36 resolved bands were detected [23].

Data were subjected to multivariate statistical and clustering analysis to establish groups of populations and distances among them. The tree groups obtained by Ward's clustering method using Squared Euclidean distances from the analysis of bands of the 100 protein profiles can be observed in Fig. 3. Two clusters can be distinguished, with an average distance value (ADV) of 10. The first clusters, however, were then divided into two small clusters with average distance value (ADV)=6.3. GSE fell in one cluster with APS (At ADV=5.49), whereas PJA fell in one cluster with CHU (At ADV=4), and the second one comprised the rest of the population (SCA, PCO, VJA, BCA, SSA and RG). SCA was the most distant among this group (ADV=7.3), PCO diverged of VJA at ADV=6.04, whereas, VJA diverged of BCA at ADV=5.9, and finally BCA diverged of SSA and RG at ADV=3.1. Neural Network-based Self Organizing Maps (SOM) is known to perform well in handling proteomics data [39], so this approach was explored aiming to classify the populations. SOM-based distance grouped the populations in several groups (Fig. 4A). The first group was formed by VJA and SCA populations. BCA, PCO, SSA, and RG populations remained as independent groups, while CHU, APS, PJA, and GSE were clustered together. Since the employed statistical methods tend to classify population together, we have used Mantel's test to study the possible correlation between these distances and the geographical and climate parameters indicated in Supplementary Table 1. SOM-based distances showed a high correlation with precipitations (P-value=0.046) and the combination of elevation and precipitations (P-value=0.052). Both analyses showed that Northern populations (GSE, APS, CHU and PJA) would be expected to have similar acorn protein profiles. These populations grow in intermediate climate with an annual precipitation among 635.7 to 948.6 mm and an altitude of 184 to 643 m. Southern populations (SSA, RG, SCA, and BCA) showed a tendency to clustering together with Ward's clustering method using, but these population remained as independent groups using SOM analysis. SSA, RG, PCO and VJA populations grow in xeric climates with an annual precipitation ranging between 277.9 and 612.6 mm and an altitude between 618 and 1241 m. SCA and BCA populations grow in mesic climates with an annual precipitation ranging between 1000.5 and 1263.6 mm and an altitude between 373 and 649 m. Similar geographical pattern was found using principal component analysis to acorn size and chemical composition of these populations. These analyses grouped the populations according to: (1) acorn size (big, intermediate, and small); (2) location (northern and southern); (3) climate conditions (xeric, mesic, and intermediate) [43]. Moreover, geographical and climate-dependent variability and introgression in *Quercus* spp. have been experimentally supported [45–48]. Andalusia Holm oak populations showed a high variability analyzing the glutelin acorn protein fraction [23], presenting Córdoba populations higher variability than Sevilla populations. Moreover, *Q. ilex* populations of different geographical regions (Greece, Italy, North Africa and the Iberian Peninsula) suggested that *Q. ilex* migration and genetic differentiation occurred from east to west [13] and this genetic differentiation may be very ancient and may characterize distinct geographical ice-age refuge (i.e. southern Balkan Peninsula, southern Italian Peninsula, North Africa and southern Iberian Peninsula) from which, after the

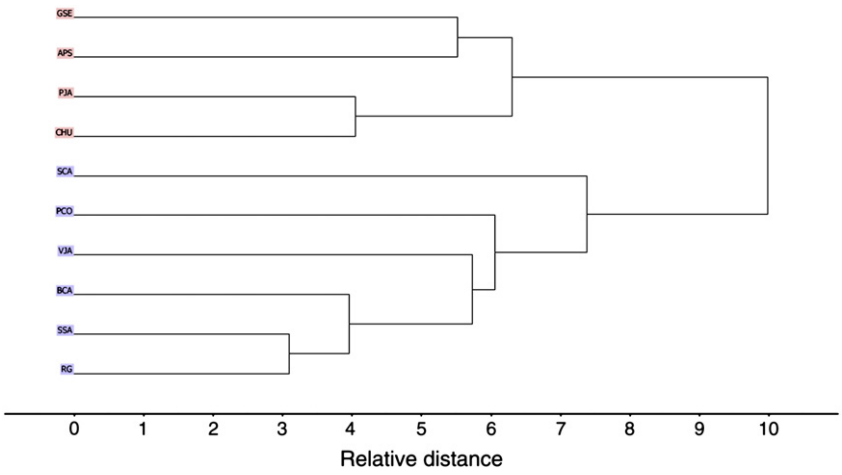


Fig. 3 – Graphical representation of the genetic distances of SDS-PAGE protein profiles, established after Ward's clustering method analysis using Squared Euclidian distance.

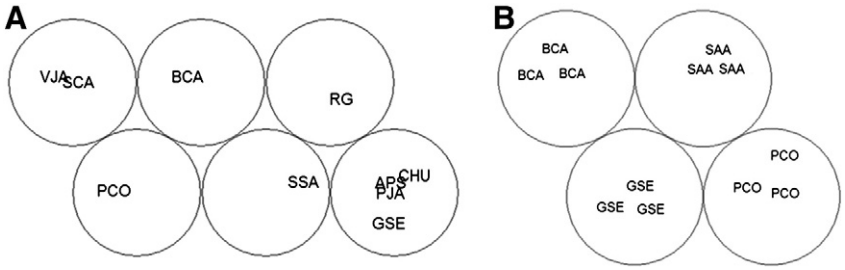


Fig. 4 – Graphical representation of the multivariate statistical analyses. (A) SOM represents the ten defined classes of SDS-PAGE protein profile of the studied populations. (B) SOM represents the defined classes of 2-DE protein profiles of the studied populations.

last glaciations, *Q. ilex* began to migrate to more northern areas. The Iberian Peninsula chlorotypes were related to those identified in North Africa. Nevertheless, a local chlorotype 14 was exclusively located in the south of Andalusia, and it may be considered as evolutionarily intermediate between the widely distributed Moroccan chlorotype 12 and the several other Iberian chlorotypes [13].

Four populations were chosen to be studied by 2-DE according to a combination of both analysis of SDS-PAGE data and geographical distance based criteria: GSE (North-West geographical location), PCO (North-East geographical location), SSA and BCA (South geographical location). Flour representing all studied accessions of each population was crushed together, and its proteins were extracted and electrophoresed as mentioned earlier. Three analytical replicates per population were analyzed. In order to increase spot resolution, 2-DE was performed in the 5–8 pH range (Supplementary Fig. 3), as most of the proteins concentrated in this range, also

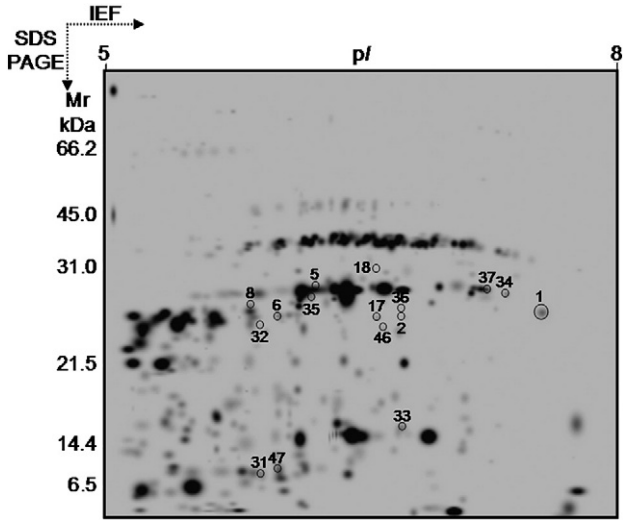


Fig. 5 – Master 2-DE gel of seed flour extracts from *Quercus ilex* subsp. *ballota*. Proteins were separated on 17 cm, pH 5–8 IPG strips (first dimension), and 13% SDS-PAGE (second dimension). On the left, position of the molecular weight marker proteins. Variable spots are numbered according to Supplementary Table 3. The real gel corresponded to the GSE population.

observed for other plants and organs [17,41,49,50]. A virtual master gel representing all protein spots present throughout all analyzed samples is shown in Fig. 5 and Supplementary Fig. 4. After normalization of protein spot images using PD-Quest and manual verification to increase reliability [39], we initially detected 200–286 well resolved spots. SSA population had the lowest number of protein spots detected (204 ± 10.58), followed by BCA (227 ± 3.21) and PCO (246 ± 0.57). However, GSE population had the highest number of protein spots detected (286 ± 5.05). Normalized and transformed spot relative volumes are presented in Supplementary Table 3. For statistical analysis, only consistent spots (present in each of the three analytical replicates per population), in a number of 161, were considered. Out of these 161 spots, 56 showed qualitative differences. Spots 1–16 were characterized by high expression in SSA, as well as to zero expression in GSE, PCO and BCA. Spots 17–30 were characterized by high expression in GSE, as well as to zero expression in SSA, PCO and BCA. Spots 31–45 were characterized by high expression in PCO, as well as to zero in SSA, GSE and BCA. Finally, spots 46–56 were characterized by high expression in BCA, as well as to zero expression in SSA, GSE and PCO (Supplementary Table 3). Ward's clustering method using Squared Euclidean distances from the analysis of 161 spots can be observed in Fig. 6. PCO was the most distant among the four population (at ADV = 10). SSA fell in one cluster with BCA and GSE (ADV = 8). BCA was closely related to GSE (ADV = 6.6) (Fig. 6). When the populations were analyzed using the SOM classification, no clusters were observed (Fig. 4B). 2-DE distances data showed a partial consistency with SDS-PAGE distances data. Only SSA and BCA showed a similar tendency to clustering like the SDS-PAGE distances. It is important to consider that the increased number of spots may lead to different classifications of populations, since proteins that constitute one SDS-PAGE bands may be split into different spots with different isoelectric points but the same molecular weight. This fact may also explain the discrepancies that can be seen comparing the total band and total spot number of populations in which those with for example the population with the higher spot number, GSE, is not the one with more bands. Moreover, is necessary to increase the number of populations using the 2-DE analysis, this approach would improve the grouping of the ten populations analyzed by SDS-PAGE. When the combination of SDS-PAGE and 2-DE data of the four populations was analyzed by Ward's clustering, the results were similar to 2-DE cluster (Supplementary Fig. 5). The

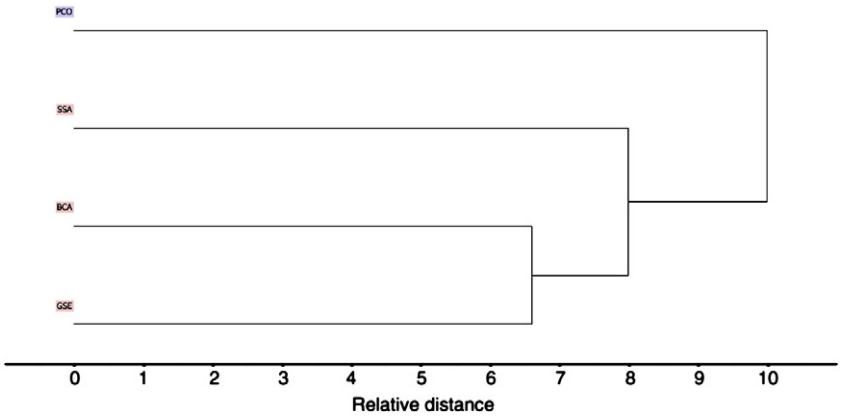


Fig. 6 – Graphical representation of the genetic distances between 2-DE protein profiles established after Ward's method analysis using Squared Euclidian distance.

higher resolution of the 2-DE, as compared with 1-DE, increased the number of proteins or group of proteins differentiating populations, favoring simultaneously the identification by MS and biological interpretation. 2-DE has been used for such a purpose, mainly with herbaceous, and to a lesser extent with woody plants, but never before with *Quercus* spp. [24–26,28,51,52].

All bands (23 in number) and variable spots (56 in number) were excised from the gels and were subjected to trypsin digestion for further analysis by MALDI-TOF/TOF and/or LC-MS/MS. In general, a low percentage of identification is characteristic of orphan tree species almost absent in databases. As reported in Table 2 (SDS-PAGE), 60.8% of the 23 bands were identified. However, 28.5% of the 2-DE spots were identified (Table 3 and Fig. 5). Proteins identified in this work were classified into two principal categories; storage and stress/defense protein. A total of fifteen proteins were identified as legumin precursor (band 6 and spots 1, 2, 5, 6, 8, 17, 18, 32, 34, 35, 36, 37, 46, and 47), ranging from Mr of 12.81 to 29.28 kDa and from pI of 6 to 8. These proteins corresponded to the same sequence accessions from *Quercus robur*. Two bands and two spots were identified as legumin (bands 16 and 20, and spots 31 and 33), ranging from Mr of 8.0 to 17.06 kDa and from pI of 6 to 7. These spots corresponded to four different sequenceaccessions. Three bands, corresponded to a unique uniprot accession (P15590) and were identified as globulin-1 S allele (bands 2, 10 and 13), ranging from Mr of 18 to 33 kDa. Storage protein heterogeneity has been also reported in other plant species and their presence has been attributed to: (i) expression of multigene families encoding for different primary sequences; (ii) differential proteolytic processing of expressed genes; and (iii) differential protein glycosylation [53–58]. Whether the identified proteins in this group can be explained by these three alternatives remains to be investigated. Obviously, the lack of information on Holm oak genome represented a strong limitation for an exhaustive elucidation of the complexity of these proteins. A genomic study in castor plant seed reveals that those 11S globulins are encoded by two diverged subfamilies comprising a total of nine genes and two putative pseudogenes. The predicted sizes for α subunit isoforms were 28.6–32.6 kDa and 20.2–21.6 kDa for β-subunit having a wide variation in pI of 5.41–10.21 [54]. Our data are within these ranges of Mr and pI, with the exception of band 20 and spots 31, 33 and 47, with lower Mr values indicating that these proteins could be degradation products. Based on measured Mr and pI values, we verified the occurrence of both polypeptides in SDS-PAGE and 2-DE protein profiles similarly to those already reported in other plant species [57,59]. Several proteins identified are involved in defense/stress responses. Bands 3, 12, and 14 were identified as heat shock proteins, which have been related to specifically increase in osmoprimed seeds and water stress [60]. We also identified a 2-Cys peroxiredoxin BAS1 (band 7) and Peroxiredoxin-2B (band 13) proteins reported to be involved in response to oxidative stress in plants. Chitinase; which was identified in bands 4 and 15, has been reported as playing an important role in plant defense against pathogens [61,62]. Glyceraldehyde-3-phosphate dehydrogenase, which was identified in bands 1, 2 and 3, has been reported as playing an important role in the seed dehydration process [63]. Other identified band proteins were involved in secondary metabolite metabolism (band 1), thiamine diphosphate biosynthesis (band 2) and translation (band 17).

In this work, we used a proteomic approach to identify specific protein markers of different populations. A comparison of these protein profiles reveals protein band as differentially expressed between populations. Statistical analysis carried out on variably expressed bands showed that bands 5, 9, 12, 13, 14, 17 and 20 were essential for population discrimination, thus proposing them as population markers. Also, these results suggest that Holm oak seed has developed a "machinery proteins" for protection of diverse biotic and abiotic stresses, and provide various nutritional proteins for the growing of seedling and enhance germination. Most of these proteins were also found in seeds of other species.

4. Concluding remarks

Nowadays, due to environmental concerns and reforestation programs, demand for tree seedlings and nursery production are increasing, as Holm oak seedlings (*Q. ilex* subsp. *ballota*

Table 2 – List of identified proteins in SDS-PAGE band.

Band number [a]	Mr exp. [b]	Mr theor. [c]	MS	Name	Species	Accession number	Protein score/ matched peptides (% coverage)/MS-MS ions (score) [d]
6	25.8	55.94	MALDI TOF/ TOF	Legumin precursor	*Quercus robur*	gi|1465711	354/14 (27)/ ADIYNPQAGR (14) ISTLNSHNLPVLR (23) WLQLSAEFGR (34) DAIYVPHWNR (56) AIPADVLANAFQLSR (128)
16	15.1	13.51	MALDI TOF/TOF	Legumin	*Quercus serrata*	gi|9971257	166/7 (61) AIPADVLANAFQLSR (100)
20	8	13.44	MALDI TOF/TOF	Legumin	*Quercus gilva*	gi|9971273	328/3 (38) TNDNAQISPLAGQTSVLR (183) AIPADVLANAFQLSR (123)
4	28.6	34.53	MALDI TOF/TOF	Chitinase	*Cucurbita*	gi|7435357	284/4(20)/ GPIQLTHNYNYGPAGR (92) VPGYGVITNIINGGLECGR (172)
15	17.3	35.22	MALDI TOF/TOF	Chitinase 2	*Triticum aestivum*	gi|18146827	105/2 (6) GPIQLTHNYNYGPAGR (97)
1	35.5	36.91	LC-MS/MS	Glyceraldehyde-3-phosphate dehydrogenase, cytosolic	*Arabidopsis thaliana*	P25858	21.9/6(8.2) VPTVDVSVVDLTVR LVSWYDNEWGYSSR TLLFGEKPVTVFGIR
1	35.5	32.55	LC-MS/MS	Lactoylglutathione lyase	*Oryza sativa* subsp. *japonica*	Q948T6	4.7/1(4.4) GPTPEPLcQVMLR
2	32.8	65.02	LC-MS/MS	Globulin-1 S allele	*Zea mays*	P15590	33.4/9(1.92) NPESFLSSFSK
2	32.8	36.91	LC-MS/MS	Glyceraldehyde-3-phosphate dehydrogenase. cytosolic	*Arabidopsis thaliana*	P25858	23.2/7(8.2) LVSWYDNEWGYSSR VPTVDVSVVDLTVR
2	32.8	36.66	LC-MS/MS	Thiazole biosynthetic enzyme. chloroplastic	*Arabidopsis thaliana*	Q38814	2.4/1(3.7) ALDMNTAEDAIVR
3	30.5	36.91	LC-MS/MS	Glyceraldehyde-3-phosphate dehydrogenase. cytosolic	*Arabidopsis thaliana*	P25858	5.3/1(4.1) LVSWYDNEWGYSSR
3	30.5	70.57	LC-MS/MS	Heat shock 70 kDa protein	*Zea mays*	P11143	10.1/1(3.15) IINEPTAAAIAYGLDKK
7	24.3	28.09	LC-MS/MS	2-Cys peroxiredoxin BAS1. chloroplastic	*Oryza sativa* subsp. *japonica*	Q6ER94	12/5(6.9) KSGGLGDLKYPLISDVTK SGGLGDLKYPLISDVTK
10	21.5	65.02	LC-MS/MS	Globulin-1 S allele	*Zea mays*	P15590	2/2(1.9) NPESFLSSFSK
12	19.7	17.37	LC-MS/MS	17.4 kDa class I heat shock protein	*Oryza sativa* subsp. *japonica*	P31673	8.2/8(27.2) ASMENGVLTVTVPK IDWKETPEAHVFK VEVEDGNVLQISGER
13	18.6	65.02	LC-MS/MS	Globulin-1 S allele	*Zea mays*	P15590	5.2/2(1.9) NPESFLSSFSK
13	18.6	17.42		Peroxiredoxin-2B	*Arabidopsis thaliana*	Q9XEX2	8.5/1(5.5) LDATANDIPSDTFDVK
14	17.5	16.46	LC-MS/MS	18.1 kDa class I heat shock protein (fragment)	*Medicago sativa*	P27879	4.6/4(18.8) AAMENGVLTVTVPK
17	14.3	13.69	LC-MS/MS	40S ribosomal protein S20-2	*Arabidopsis thaliana*	Q9STY6	5.6/2(9.8) VIDLFSSPDVVK

[a] Band numbers correspond to Fig. 2 and Supplementary Table 2.
[b] Molecular weight (kDa) calculated by using molecular weight standards.
[c] Molecular weight (kDa) annotated in the NCBI database.
[d] 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.

Table 3 – List of identified protein of 2-DE spots.

Number [a]	Mr/pI exp. [b]	Mr/pI theor. [c]	Name	Species	Accession number [d]	Protein score/ matched peptides (% coverage)/MS-MS ions (score) [e]
1	26.22/7.87	55.94/5.86	Legumin precursor	*Quercus robur*	gi|1465711	418/11 (24) ADIYNPQAGR (72) ISTLNSHNLPVLR (70) DAIYVPHWNR (51) TNDNAQISPLAGQTSVLR (161)
2	25.90/7.12	55.94/5.86	Legumin precursor	*Quercus robur*	gi|1465711	288/8 (18) ISTLNSHNLPVLR (50) ASSSEGFEWVAFK (32) TNDNAQISPLAGQTSVLR (136) AIPADVLANAFQLSR (31)
5	27.85/6.64	55.94/5.86	Legumin precursor	*Quercus robur*	gi|1465711	404/9 (20) ISTLNSHNLPVLR (52) DAIYVPHWNR (55) TNDNAQISPLAGQTSVLR (155) AIPADVLANAFQLSR (96)
6	25.25/6.60	55.94/5.86	Legumin precursor	*Quercus robur*	gi|1465711	315/9 (22) ADIYNPQAGR (78) DAIYVPHWNR (54) TNDNAQISPLAGQTSVLR (137)
8	26.74/6.35	55.94/5.86	Legumin precursor	*Quercus robur*	gi|1465711	254/11 (23) ISTLNSHNLPVLR (38) TNDNAQISPLAGQTSVLR (135) AIPADVLANAFQLSR (19)
17	25.88/6.97	55.94/5.86	Legumin precursor	*Quercus robur*	gi|1465711	267/7 (16) ADIYNPQAGR (48) ISTLNSHNLPVLR (39) TNDNAQISPLAGQTSVLR (149)
18	29.28/6.96	55.94/5.86	Legumin precursor	*Quercus robur*	gi|1465711	119/7 (17) ADIYNPQAGR (13) ISTLNSHNLPVLR (23) DAIYVPHWNR (5) TNDNAQISPLAGQTSVLR (47)
31	10.90/6.14	13.10/4.43	Legumin	*Castanea crenata*	gi|9971249	157/4 (38) ASSSEGFEWVAFK (22) TNDNAQISPLAGQTSVLR (103)
32	27.53/7.69	55.94/5.86	Legumin precursor	*Quercus robur*	gi|1465711	119/7 (17) ISTLNSHNLPVLR (13) TNDNAQISPLAGQTSVLR (76)
33	17.06/7.11	13.52/5.43	Legumin	*Quercus acutissima*	gi|9971255	232/5 (42) AVSSEGFEWVAFK (41) TNDNAQISPLAGQTSVLR (149)
34	27.53/7.69	55.94/5.86	Legumin precursor	*Quercus robur*	gi|1465711	311/7 (18) ADIYNPQAGR (65) ISTLNSHNLPVLR (71) TNDNAQISPLAGQTSVLR (144)
35	27.23/6.64	55.94/5.86	Legumin precursor	*Quercus robur*	gi|1465711	264/9 (20) ISTLNSHNLPVLR (51) TNDNAQISPLAGQTSVLR (167)
36	26.55/7.11	55.94/5.86	Legumin precursor	*Quercus robur*	gi|1465711	245/8 (20) ISTLNSHNLPVLR (24) DAIYVPHWNR (40) TNDNAQISPLAGQTSVLR (97) AIPADVLANAFQLSR (44)
37	27.83/7.59	55.94/5.86	Legumin precursor	*Quercus robur*	gi|1465711	437/10 (21) ADIYNPQAGR (59) ISTLNSHNLPVLR (105) DAIYVPHWNR (58) TNDNAQISPLAGQTSVLR (158)

Table 3 (*continued*)

Number [a]	Mr/pI exp. [b]	Mr/pI theor. [c]	Name	Species	Accession number [d]	Protein score/ matched peptides (% coverage)/MS-MS ions (score) [e]
46	25.19/7.01	55.94/5.86	Legumin precursor	*Quercus robur*	gi|1465711	271/8 (20) ADIYNPQAGR (70) DAIYVPHWNR (64) TNDNAQISPLAGQTSVLR (100)
47	12.81/6.60	55.94/5.86	Legumin precursor	*Quercus robur*	gi|1465711	174/12 (25) TNDNAQISPLAGQTSVLR (101)

[a] Numbers correspond to Fig. 4 and Supplementary Table 3.
[b] Molecular weight (kDa) and isoelectric point calculated by using molecular weight standards and the PD-Quest Advance (8.01) software.
[c] Molecular weight (kDa) and isoelectric point annotated in the NCBI database.
[d] NCBI database accession numbers.
[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.

[Desf.] Samp.) in southern Spain. This goal requires the establishment of techniques directed at characterizing, cataloging, and selecting genotypes, prior to sexual and vegetative propagation.

We have shown in this paper the utility of a basic proteomics approach, based on SDS-PAGE and 2-D electrophoresis analyses of protein extracts from mature seed flour, for variability studies in Holm oak. A total of ten different populations distributed throughout the Andalusia region have been used, together with a proper multivariate statistical analysis of the results. This methodology can be used in combination with other morphometric, physiological, biochemical, and genetic approaches, in order to characterize population variability, and establish distances among them. Results presented here prove the great inter-population variability of Holm oak in southern Spain, as corresponding to non-domesticated out-crossing species colonizing quite different habitats. This conclusion has also been reported in the literature for different *Quercus* spp. To what extent genetic or environmental factors do contribute to the great heterogeneity remains to be investigated. Even so, the simplest protein quantitation and the SDS-PAGE analysis allowed the separation and grouping of the populations according to its acorn morphometry, location (northern and southern), and climate conditions (xeric, mesic and intermediate), confirming previous results of our research group by using acorn morphometry, and NIRS chemical composition [43].

The 2-DE technique is complicated and difficult to use with large number of samples. On the contrary, it permits proteins identification, thus placing a blind code (band or spot) into a biological context, and allowing gen characterization, as well as proteins and mechanisms mediating both genotypic and phenotypic differences. In our 2-DE analysis around 240 spots could be resolved in the 5–8 pH range, with 161 consistent spots (do not reflect qualitative intra-population variability). Out of these 161 spots, 56 showed qualitative differences, and could be considered as markers of population variability.

Proteomics (1DE and 2DE) can be used to detect variability between different populations coming from different environments and there are some correlations to the environmental conditions. The limitation of the approach used is the low number of proteins identified, most of them corresponding to the reserve 11S globulin family of legumins which have a difficult interpretation from a biological point of view.

Supplementary materials related to this article can be found online at doi:10.1016/j.jprot.2011.05.003.

Acknowledgments

Jose Valero was a recipient of an Alban Program fellowship. This work is part of research projects funded by the Spanish Ministry of Science and Innovation, their being FEDER co-funded: CGL2008-04503-C03-01/BOS, AGL2002-00530, and AGL2009-12243-C02-02.

REFERENCES

[1] Pulido FJ, Díaz M, Hidalgo de Trucios SJ. Size structure and regeneration of Spanish Holm oak *Quercus ilex* forests and dehesas: effects of agroforestry use on their long-term sustainability. Forest Ecol Manage 2001;146:1–13.
[2] Gea-Izquierdo G, Cañellas I, Montero G. Acorn production in Spanish Holm oak woodlands. Invest Agrar Sist Recur For 2006;15:339–54.
[3] FAO. Evaluación de los recursos forestales mundiales. Informe nacional España; 2010 http://www.fao.org/forestry/20262-1-50.pdf.
[4] Pausas J, Bladé C, Valdecantos A, Seva J, Fuentes D, Alloza J, et al. Pines and oaks in the restoration of Mediterranean landscapes of Spain: new perspectives for an old practice a review. Plant Ecol 2004;171:209–20.
[5] Soto A, Lorenzo Z, Gil L. Differences in fine-scale genetic structure and dispersal in *Quercus ilex* L. and *Q. suber* L.: consequences for regeneration of Mediterranean open woods. Heredity 2007;99:601–7.
[6] Burgarella C, Lorenzo Z, Jabbour-Zahab R, Lumaret R, Guichoux E, Petit RJ, et al. Detection of hybrids in nature: application to oaks (*Quercus suber* and *Q. ilex*). Heredity 2009;102:442–52.
[7] Castro-Díez P, Villar-Salvador P, Pérez-Rontomé C, Maestro-Martínez M, Montserrat-Martí G. Leaf morphology and leaf chemical composition in three *Quercus* (Fagaceae)

species along a rainfall gradient in NE Spain. Trees-Structure and Function 1997;11:127–34.
[8] Jimenez P, de Heredia UL, Collada C, Lorenzo Z, Gil L. High variability of chloroplast DNA in three Mediterranean evergreen oaks indicates complex evolutionary history. Heredity 2004;93:510–5.
[9] Lumaret R, Jabbour-Zahab R. Ancient and current gene flow between two distantly related Mediterranean oak species, *Quercus suber* and *Q. ilex*. Ann Bot 2009;104:725–36.
[10] Lumaret R, Tryphon-Dionnet M, Michaud H, Sanuy A, Ipotesi E, Born C, et al. Phylogeographical variation of chloroplast DNA in cork oak (*Quercus suber*). Ann Bot 2005;96:853–61.
[11] Toumi L, Lumaret R. Allozyme characterisation of four Mediterranean evergreen oak species. Biochem Syst Ecol 2001;29:799–817.
[12] Vakkari P, Blom A, Rusanen M, Raisio J, Toivonen H. Genetic variability of fragmented stands of pedunculate oak (*Quercus robur*) in Finland. Genetica 2006;127:231–41.
[13] Lumaret R, Mir C, Michaud H, Raynal V. Phylogeographical variation of chloroplast DNA in Holm oak (*Quercus ilex* L.). Mol Ecol 2002;11:2327–36.
[14] Petit RJ, Brewer S, Bordács S, Burg K, Cheddadi R, Coart E, et al. Identification of refugia and post-glacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence. Forest Ecol Manage 2002;156:49–74.
[15] Echevarría-Zomeño S, Ariza D, Jorge I, Lenz C, Del Campo A, Jorrin JV, 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.
[16] Jorge I, Navarro Cerrillo RM, Lenz C, Ariza D, Jorrin J. Variation in the Holm oak leaf proteome at different plant developmental stages, between provenances and in response to drought stress. Proteomics 2006;6(Suppl 1):S207–14.
[17] Jorge I, Navarro Cerrillo RM, Lenz C, Ariza D, Porras C, Jorrin 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.
[18] Cherdouh A, Khelifi D, Carrillo JM, Nieto-Taladriz MT. The high and low molecular weight glutenin subunit polymorphism of Algerian durum wheat landraces and old cultivars. Plant Breed 2005;124:338–42.
[19] Karihaloo JL, Kaur M, Singh S. Seed protein diversity in *Solanum melongena* L. and its wild and weedy relatives. Genetic Resources Crop Evol 2002;49:533–9.
[20] Stoilova T, Cholakova N, Markova M. Variation in seed protein and isoenzyme patterns in *Cucurbita* cultivars. Biol Plantarum 2006;50:450–2.
[21] Fonseca PA, Ferreira RB, Teixeira AR. Seed proteins from *Quercus suber*. J Agric Food Chem 1997;45:3443–7.
[22] Collada C, Caballero RG, Casado R, Aragoncillo C. Seed storage proteins in Fagaceae: similarity between *Castanea globulins* and *Quercus glutelins*. Plant Sci 1991;75:145–54.
[23] Martín M, Navarro-Cerrillo R, Ortega P, Alvarez J. The use of cotyledon proteins to assess the genetic diversity in sweet Holm oak. J Forest Sci 2009;55:526–31.
[24] Bahrman N, Petit RJ. Genetic polymorphism in maritime pine (*Pinus pinaster* Ait.) assessed by two-dimensional gel electrophoresis of needle, bud, and pollen proteins. J Mol Evol 1995;41:231–7.
[25] Basha SM. Identification of cultivar differences in seed polypeptide composition of peanuts (*Arachis hypogaea* L.) by two-dimensional polyacrylamide gel electrophoresis. Plant Physiol 1979;63:301–6.
[26] Chevalier F, Martin O, Rofidal V, Devauchelle AD, Barteau S, Sommerer N, et al. Proteomic investigation of natural variation between *Arabidopsis* ecotypes. Proteomics 2004;4: 1372–81.
[27] Emre İ. Electrophoretic analysis of some *Lathyrus* L. species based on seed storage proteins. Genetic Resources Crop Evol 2009;56:31–8.
[28] Jacobsen S, Nesic L, Petersen M, Sondergaard I. Classification of wheat varieties: use of two-dimensional gel electrophoresis for varieties that can not be classified by matrix assisted laser desorpiton/ionization-time of flight-mass spectrometry and an artificial neural network. Electrophoresis 2001;22:1242–5.
[29] Ostergaard O, Melchior S, Roepstorff P, Svensson B. Initial proteome analysis of mature barley seeds and malt. Proteomics 2002;2:733–9.
[30] Singh N, Matta N. Variation studies on seed storage proteins and phylogenetics of the genus *Cucumis*. Plant Syst Evol 2008;275:209–18.
[31] Yüzbaşıoğlu E, Açik L, Özcan S. Seed protein diversity among lentil cultivars. Biol Plantarum 2008;52:126–8.
[32] Petit RJ, Bahrman N, Baradat PH. Comparison of genetic differentiation in maritime pine (*Pinus pinaster* Ait.) estimated using isozyme, total protein and terpenic loci. Heredity 1995;75:382–9.
[33] Vienne DD, Burstin J, Gerber S, Leonardi A, Guilloux ML, Murigneux A, et al. Two-dimensional electrophoresis of proteins as a source of monogenic and codominant markers for population genetics and mapping the expressed genome. Heredity 1996;76:166–77.
[34] Damerval C, De Vienne D, Zivy M, Thiellement H. Technical improvements in two-dimensional electrophoresis increase the level of genetic variation detected in wheat-seedling proteins. Electrophoresis 1986;7:52–4.
[35] Ramagli LS, Rodriguez LV. Quantitation of microgram amounts of protein in two-dimensional polyacrylamide gel electrophoresis sample buffer. Electrophoresis 1985;6: 559–63.
[36] Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 1970;227: 680–5.
[37] Mathesius U, Keijzers G, Natera SH, Weinman JJ, Djordjevic MA, Rolfe BG. Establishment of a root proteome reference map for the model legume *Medicago truncatula* using the expressed sequence tag database for peptide mass fingerprinting. Proteomics 2001;1:1424–40.
[38] Gorg A, Postel W, Baumer M, Weiss W. Two-dimensional polyacrylamide gel electrophoresis, with immobilized pH gradients in the first dimension, of barley seed proteins: discrimination of cultivars with different malting grades. Electrophoresis 1992;13:192–203.
[39] Valledor L, Jorrin J. Back to the basics: maximizing the information obtained by quantitative two dimensional gel electrophoresis analyses by an appropriate experimental design and statistical analyses. J Proteomics 2010;74:1–18.
[40] Team RDC. R: a language and environment for statistical computing. Vienna, Austria: Foundation for Statistical Computing; 2008.
[41] Maldonado AM, Echevarría-Zomeño S, Jean-Baptiste S, Hernández M, Jorrín-Novo JV. Evaluation of three different protocols of protein extraction for *Arabidopsis thaliana* leaf proteome analysis by two-dimensional electrophoresis. J Proteomics 2008;71:461–72.
[42] Bonner F, Vozzo J. 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; 1987. p. 21.
[43] Valero-Galvn J, Jorrín J, Gómez C, Ariza D, García-Olmo J, Navarro Cerrillo R. Population variability and mother tree selection in Holm oak (*Quercus ilex* L. subsp. *ballota* [Desf.] Samp.) based on acorn morphometry and chemical composition. Submitted to the European Journal of Forest Research. 2010.

[44] Özcan T. Total protein and amino acid compositions in the acorns of Turkish *Quercus* L. taxa. Genetic Resources Crop Evol 2006;53:419–29.
[45] Barbero M, Loisel R, Quézel P. Biogeography, ecology and history of Mediterranean *Quercus ilex* ecosystems. Plant Ecol 1992;99–100:19–34.
[46] Corine H, Serge R. An examination of the interaction between climate, soil and leaf area index in a *Quercus ilex* ecosystem. Annals Forest Sci 2003;60:153–61.
[47] Ducousso A, Guyon JP, Kremer A. Latitudinal and altitudinal variation of bud burst in western populations of sessile oak (*Quercus petraea* (Matt) Liebl). Annals Forest Sci 1996;53: 775–82.
[48] Morales R, Vicente JA, Galán de Mera A. cpDNA evidence of introgressions in *Quercus* L. (Fagaceae). The influence of the phytotopographic position. Flora 2005;200:222–8.
[49] Valledor L, Castillejo MA, Lenz C, Rodriguez R, Canal MJ, Jorrin 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.
[50] Valledor L, Jorrin JV, Rodriguez JL, Lenz C, Meijon M, Rodriguez R, et al. Combined proteomic and transcriptomic analysis identifies differentially expressed pathways associated to *Pinus radiata* needle maturation. J Proteome Res 2010;9: 3954–79.
[51] Consoli L, Damerval C. Quantification of individual zein isoforms resolved by two-dimensional electrophoresis: genetic variability in 45 maize inbred lines. Electrophoresis 2001;22:2983–9.
[52] Marques K, Sarazin B, Chane-Favre L, Zivy M, Thiellemen H. Comparative proteomics to establish genetic relationships in the Brassicaceae family. Proteomics 2001;1:1457–62.
[53] Casey R, Short MN. Variation in amino acid composition of legumin from *Pisum*. Phytochemistry 1981;20:21–3.
[54] Chileh T, Esteban-García B, Diego López A, García-Maroto F. Characterization of the 11S globulin gene family in the castor plant *Ricinus communis* L. J Agric Food Chem 2009;58:272–81.
[55] Derbyshire E, Wright DJ, Boulter D. Legumin and vicilin, storage proteins of legume seeds. Phytochemistry 1976;15: 3–24.
[56] Maplestone P, Allison J, Hussein EHA, Gamal El-Din AE-KY, Gatehouse JA, Boulter D. Variation of the legumin seed storage protein amongst *Vicia* species. Phytochemistry 1985;24:1717–23.
[57] Scippa GS, Rocco M, Ialicicco M, Trupiano D, Viscosi V, Di Michele M, et al. The proteome of lentil (*Lens culinaris* Medik.) seeds: discriminating between landraces. Electrophoresis 2010;31:497–506.
[58] Wobus U, Bäumlein H, Bassüner R, Heim U, Jung R, Müntz K, et al. Characteristics of two types of legumin genes in the field bean (*Vicia faba* L.var. minor) genome as revealed by cDNA analysis. FEBS Lett 1986;201:74–80.
[59] Bourgeois M, Jacquin F, Savois V, Sommerer N, Labas V, Henry C, et al. Dissecting the proteome of pea mature seeds reveals the phenotypic plasticity of seed protein composition. Proteomics 2009;9:254–71.
[60] Gallardo K, Le Signor C, Vandekerckhove J, Thompson RD, Burstin J. Proteomics of *Medicago truncatula* seed development establishes the time frame of diverse metabolic processes related to reserve accumulation. Plant Physiologists 2003;133: 664–82.
[61] Bestel-Corre G, Dumas-Gaudot E, Gianinazzi-Pearson V, Gianinazzi S. Mycorrhiza-related chitinase and chitosanase activity isoforms in *Medicago truncatula* Gaertn. Symbiosis 2002;32:173–94.
[62] Salzer P, Bonanomi A, Beyer K, Vögeli-Lange R, Aeschbacher RA, Lange J, et al. Differential expression of eight chitinase genes in *Medicago truncatula* roots during mycorrhiza formation, nodulation, and pathogen Infection. Mol Plant Microbe Interact 2000;13:763–77.
[63] Gallardo K, Job C, Groot SP, Puype M, Demol H, Vandekerckhove J, et al. Proteomic analysis of *Arabidopsis* seed germination and priming. Plant Physiologists 2001;126: 835–48.

Supplementary tables

Supplementary Table 1. Coordinates, altitude, mean annual precipitation, and mean maximum and minimum monthly temperatures corresponding to the geographical areas of the ten Holm oak Andalusia populations used in the present study (see Supplementary Figure 1). Populations are ordered according to increasing annual precipitation.

Province	Population Location Abbreviation	Latitude (North) Longitude (West)	Altitude (m)	Annual precipitation (mm)	Average monthly temperature (ºC)	
					Maximum	Minimum
Almería	Sierra de Alhamilla (SSA)	36º 59′ 06º 05′	1241	277.9	25.2	08.9
Granada	Arenas del Rey (RG)	36º 57′ 03º 54′	963	489.3	24.7	11.5
Jaén	Valdepeñas (VJA)	37º 30′ 03º 56′	1229	556.3	24.8	05.9
Córdoba	Pozoblanco (PCO)	38º 22′ 04º 54′	618	612.6	26.8	08.1
Huelva	Calaña (CHU)	37º 38′ 06º 51′	184	635.7	26.5	10.5
Sevilla	Almadén de la Plata (APS)	37º 52′ 06º 05′	482	722.1	26.4	09.5
Jaén	Pozoalcón (PJA)	38º 17′ 02º 36′	643	795.4	24.1	04.4
Sevilla	El Garrobo (GSE)	37º 37′ 06º 10′	289	948.6	25.6	09.5
Cádiz	Puerto Serrano (SCA)	36º 54′ 05º 31′	373	1000.5	25.5	09.5
Cádiz	Benamahoma (BCA)	36º 45′ 05º 27′	649	1263.6	24.9	09.8

Supplementary Table 2. Normalized and transformed relative volumes of the 1-DE resolved bands. Bands were numbered according to Figure 2. Mean ± SD of ten biological replicates. n.d.: no detected band (qualitative differences). Mean values with the same letters ($^{a.b.c.d.e}$) indicate homogeneous subsets for α=0.05 according to Duncan test.

Band	*Mr* (kDa)	POPULATION									
		SSA	RG	VJA	PCO	CHU	APS	PJA	GSE	SCA	BCA
1	35.5	2.60±2.25 ^{a}	3.08±1.17 ^{ab}	2.96±1.61 ^{ab}	4.91±0.86 ^{c}	3.97±0.43 ^{b}	3.51±0.66 ^{ab}	3.09±1.03 ^{ab}	3.40±0.41 ^{ab}	3.73±1.95 ^{b}	2.03±1.63 ^{a}
2	32.8	1.37±1.80 ^{ab}	1.89±1.23 ^{ab}	0.41±0.86 ^{a}	0.83±1.45 ^{a}	2.35±1.09 ^{b}	2.09±1.25 ^{b}	2.20±0.97 ^{b}	2.36±0.42 ^{b}	0.65±2.38 ^{a}	2.07±1.18 ^{b}
3	30.5	2.31±2.05 ^{a}	3.09±1.30 ^{ab}	3.63±0.53 ^{b}	3.26±1.60 ^{ab}	2.79±1.35 ^{ab}	3.32±0.99 ^{ab}	2.96±1.53 ^{ab}	2.60±0.55 ^{ab}	4.59±0.92 ^{c}	2.39±1.54 ^{a}
4	28.6	2.27±2.04 ^{a}	3.68±1.02 ^{b}	4.67±0.28 ^{c}	2.86±1.38 ^{a}	3.95±0.35 ^{b}	3.84±1.10 ^{bc}	4.07±0.50 ^{bc}	3.47±0.84 ^{b}	4.41±2.04 ^{c}	4.26±0.65 ^{bc}
5	27.8	0.55±1.17	0.90±1.60	0.64±1.35	n.d.	0.27±0.83	0.68±1.28	n.d.	n.d.	1.75±2.09	1.59±1.73
6	25.8	4.34±0.43 ^{a}	4.64±0.00 ^{a}	4.64±0.00 ^{a}	4.62±0.19 ^{a}	4.64±0.00 ^{a}	4.56±0.27 ^{a}	4.64±0.00 ^{a}	4.64±0.00 ^{a}	4.64±1.19 ^{a}	4.64±0.00 ^{a}
7	24.3	3.85±0.27 ^{cd}	3.15±0.86 ^{bc}	3.84±0.37 ^{cd}	4.25±0.55 ^{d}	3.26±0.32 ^{bc}	2.98±1.22 ^{ab}	3.19±0.44 ^{bc}	1.81±1.28 ^{a}	3.53±1.32 ^{bc}	2.62±1.59 ^{ab}
8	23.3	4.37±0.25 ^{c}	3.80±0.45 ^{ab}	4.51±0.22 ^{c}	3.63±1.54 ^{a}	3.69±0.30 ^{a}	3.81±0.45 ^{ab}	3.79±0.38 ^{ab}	3.81±0.45 ^{ab}	4.54±1.94 ^{c}	3.82±0.37 ^{ab}
9	22.8	n.d.	0.85±1.29	2.41±1.27	1.39±1.82	n.d.	n.d.	0.34±0.92	n.d.	1.62±1.65	2.09±1.28
10	21.5	3.77±0.36 ^{c}	3.13±0.51 ^{b}	3.87±0.27 ^{c}	3.49±1.02 ^{b}	3.07±0.28 ^{b}	3.20±0.39 ^{b}	3.18±0.48 ^{b}	2.71±0.29 ^{a}	3.89±1.24 ^{c}	3.50±0.40 ^{b}
11	20.3	3.80±0.25 ^{a}	3.29±0.33 ^{a}	3.65±0.07 ^{a}	3.55±1.04 ^{a}	3.35±0.85 ^{a}	3.40±1.03 ^{a}	3.25±0.94 ^{a}	3.62±0.25 ^{a}	3.86±1.52 ^{a}	3.69±0.57 ^{a}
12	19.7	2.13±1.85	2.12±1.52	2.63±1.51	2.39±1.78	0.86±1.38	1.27±1.61	1.76±1.66	n.d.	3.06±1.76	2.68±1.29
13	18.6	2.53±1.45	2.14±1.60	4.11±0.54	2.69±1.84	n.d.	2.34±1.19	n.d.	3.12±0.23	2.03±2.00	2.12±1.47
14	17.5	1.82±1.61	2.78±1.35	n.d.	2.86±1.95	2.96±0.80	n.d.	2.86±0.87	n.d.	4.30±1.79	2.62±1.57
15	17.3	3.04±0.72 ^{c}	2.59±1.50 ^{b}	3.21±1.84 ^{cd}	1.66±1.75 ^{a}	3.62±0.49 ^{d}	3.32±0.53 ^{d}	3.42±0.58 ^{d}	3.29±0.41 ^{cd}	2.09±1.94 ^{ab}	3.06±0.90 ^{c}
16	15.1	4.64±0.00 ^{a}	4.36±0.51 ^{a}	4.59±0.62 ^{a}	5.09±0.55 ^{a}	4.36±0.34 ^{a}	4.32±0.46 ^{a}	4.40±0.34 ^{a}	3.98±0.62 ^{a}	4.85±2.59 ^{a}	4.84±0.80 ^{a}
17	14.3	n.d.	n.d.	n.d.	1.62±1.71	n.d.	0.19±0.77	0.73±1.34	n.d.	n.d.	n.d.
18	11.6	3.86±0.49 ^{bc}	3.27±1.00 ^{b}	4.03±0.42 ^{c}	4.14±0.89 ^{c}	3.50±1.06 ^{bc}	2.54±1.51 ^{a}	2.79±1.46 ^{a}	3.22±0.79 ^{b}	4.35±1.49 ^{c}	3.99±1.45 ^{bc}
19	10.6	2.93±1.57 ^{b}	2.56±1.98 ^{b}	3.60±2.04 ^{c}	0.74±1.69 ^{a}	0.85±1.51 ^{a}	3.31±1.53 ^{c}	2.30±1.70 ^{b}	2.34±0.65 ^{b}	4.42±1.49 ^{ab}	3.11±1.70 ^{c}
20	8.0	0.72±1.51	1.08±1.73	1.49±1.93	n.d.	0.15±0.68	2.96±0.90	2.20±1.61	0.71±1.14	4.28±0.97	n.d.
21	7.4	4.04±1.07	2.99±1.14	3.89±0.28	2.79±1.67	2.41±1.70	3.22±1.68	3.56±0.45	n.d.	4.05±1.68	3.55±1.28
22	6.3	3.15±1.75 ^{b}	3.62±1.53 ^{b}	4.47±0.69 ^{c}	4.34±1.24 ^{c}	3.49±1.33 ^{b}	1.92±2.26 ^{a}	1.90±2.01 ^{a}	3.73±1.09 ^{b}	3.53±2.14 ^{b}	4.40±1.70 ^{c}
23	4.0	3.59±2.02 ^{c}	3.55±1.90 ^{c}	4.78±0.87 ^{d}	4.27±1.31 ^{d}	2.37±1.88 ^{b}	0.80±1.80 ^{a}	3.75±1.31 ^{c}	2.41±2.09 ^{b}	4.51±2.14 ^{ab}	4.57±2.42 ^{d}

Supplementary Table 3. Normalized and transformed relative volumes of the 2-DE resolved spots. Mean ± SD of three biological replicates. n.d.: non detected band (qualitative differences).

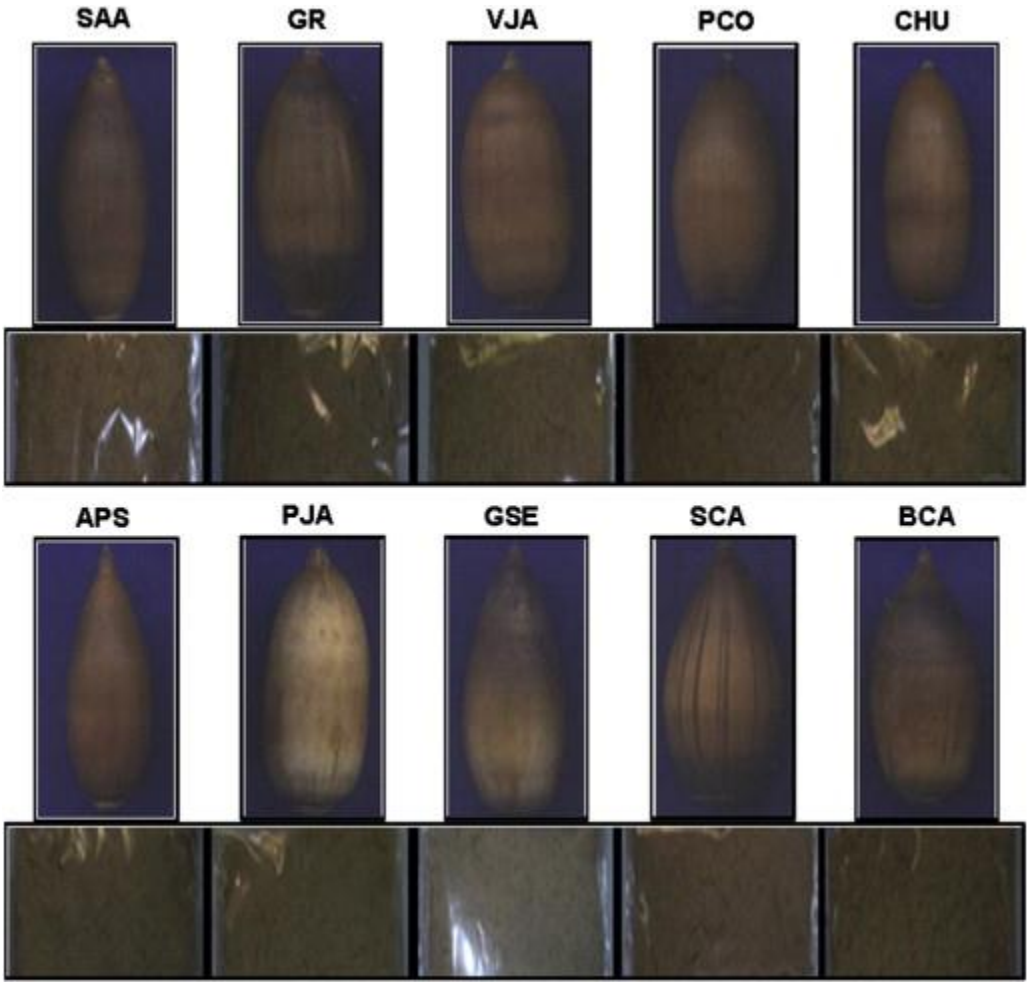
Spot number	POPULATION			
	SSA	GSE	PCO	BCA
1	6.80±1.00	n.d.	n.d.	n.d.
2	2.47±0.08	n.d.	n.d.	n.d.
3	3.18±0.16	n.d.	n.d.	n.d.
4	4.30±0.06	n.d.	n.d.	n.d.
5	7.59±0.89	n.d.	n.d.	n.d.
6	4.02±0.18	n.d.	n.d.	n.d.
7	6.66±0.29	n.d.	n.d.	n.d.
8	3.82±0.18	n.d.	n.d.	n.d.
9	5.76±0.46	n.d.	n.d.	n.d.
10	4.84±0.65	n.d.	n.d.	n.d.
11	4.42±0.22	n.d.	n.d.	n.d.
12	2.52±0.39	n.d.	n.d.	n.d.
13	3.59±0.23	n.d.	n.d.	n.d.
14	4.78±0.18	n.d.	n.d.	n.d.
15	3.74±0.40	n.d.	n.d.	n.d.
16	3.40±0.72	n.d.	n.d.	n.d.
17	n.d.	3.28±0.24	n.d.	n.d.
18	n.d.	2.58±0.08	n.d.	n.d.
19	n.d.	2.73±0.18	n.d.	n.d.
20	n.d.	3.37±0.54	n.d.	n.d.
21	n.d.	2.55±0.69	n.d.	n.d.
22	n.d.	3.45±0.16	n.d.	n.d.
23	n.d.	2.51±0.10	n.d.	n.d.
24	n.d.	2.84±0.13	n.d.	n.d.
25	n.d.	3.70±0.40	n.d.	n.d.
26	n.d.	2.80±0.24	n.d.	n.d.
27	n.d.	1.88±0.08	n.d.	n.d.
28	n.d.	2.57±0.15	n.d.	n.d.
29	n.d.	2.73±0.27	n.d.	n.d.
30	n.d.	2.56±0.19	n.d.	n.d.
31	n.d.	n.d.	4.66±0.18	n.d.
32	n.d.	n.d.	3.38±0.07	n.d.
33	n.d.	n.d.	4.56±0.12	n.d.
34	n.d.	n.d.	3.95±0.06	n.d.
35	n.d.	n.d.	4.01±0.29	n.d.
36	n.d.	n.d.	2.77±0.03	n.d.
37	n.d.	n.d.	6.52±0.08	n.d.
38	n.d.	n.d.	6.13±0.02	n.d.
39	n.d.	n.d.	3.73±0.73	n.d.
40	n.d.	n.d.	3.37±0.35	n.d.
41	n.d.	n.d.	4.04±0.19	n.d.

42	n.d.	n.d.	3.35±0.17	n.d.
43	n.d.	n.d.	4.37±0.20	n.d.
44	n.d.	n.d.	5.29±0.15	n.d.
45	n.d.	n.d.	4.89±0.15	n.d.
46	n.d.	n.d.	n.d.	3.17±0.31
47	n.d.	n.d.	n.d.	5.30±0.09
48	n.d.	n.d.	n.d.	4.59±0.54
49	n.d.	n.d.	n.d.	4.61±0.14
50	n.d.	n.d.	n.d.	4.53±0.35
51	n.d.	n.d.	n.d.	4.50±0.46
52	n.d.	n.d.	n.d.	1.90±0.19
53	n.d.	n.d.	n.d.	5.69±0.61
54	n.d.	n.d.	n.d.	3.39±0.94
55	n.d.	n.d.	n.d.	1.83±0.78
56	n.d.	n.d.	n.d.	4.71±0.27
57	4.93±1.40	6.19±0.92	2.39±0.80	7.24±2.25
58	7.93±3.69	8.26±0.28	7.08±0.29	9.43±0.74
59	9.68±0.34	9.63±0.26	8.10±0.46	10.07±0.35
60	8.72±0.93	8.83±0.61	6.52±0.51	8.58±0.51
61	10.76±1.06	10.78±0.49	7.38±0.80	10.44±0.80
62	6.36±0.36	4.92±0.39	3.93±0.28	7.62±0.54
63	5.31±0.34	3.92±0.58	3.74±0.54	5.28±0.33
64	5.79±0.58	5.73±0.57	5.31±0.04	5.78±0.12
65	5.91±0.18	6.00±0.12	4.78±0.09	7.02±0.12
66	4.30±0.51	3.84±0.23	4.62±0.01	4.14±0.60
67	4.97±0.14	4.24±0.39	4.96±0.13	5.99±0.48
68	5.17±1.89	4.10±0.10	3.81±0.22	4.58±1.11
69	4.91±1.02	4.79±0.06	4.42±0.13	4.66±2.90
70	5.91±0.16	5.39±0.15	7.33±0.16	5.23±0.62
71	10.42±0.83	10.26±1.41	7.23±1.61	11.90±0.91
72	10.92±1.26	11.39±0.25	11.88±0.41	9.61±0.42
73	3.70±0.35	4.77±0.28	2.43±1.23	4.17±0.56
74	4.50±0.13	4.24±0.52	4.78±0.00	4.44±0.37
75	11.88±2.80	11.76±1.66	15.66±0.51	11.65±0.52
76	7.08±0.37	8.60±0.30	5.57±0.09	7.40±0.32
77	6.39±0.80	4.95±0.41	4.22±1.17	6.68±0.79
78	3.90±0.96	3.99±0.18	2.57±0.65	5.36±0.23
79	5.84±0.76	6.56±0.10	5.22±0.33	8.21±0.19
80	10.30±2.68	6.40±0.13	8.48±0.36	8.56±3.61
81	5.63±0.73	5.60±0.29	4.97±0.61	5.19±0.17
82	5.96±0.59	5.62±0.14	5.68±0.21	6.35±0.32
83	9.88±1.27	10.45±0.29	8.92±2.22	7.74±0.72
84	6.12±0.35	4.73±0.74	4.17±0.98	5.18±0.43
85	4.42±0.27	3.62±0.20	3.61±0.74	3.57±0.40
86	5.02±1.10	5.42±0.31	4.05±0.46	6.42±0.18

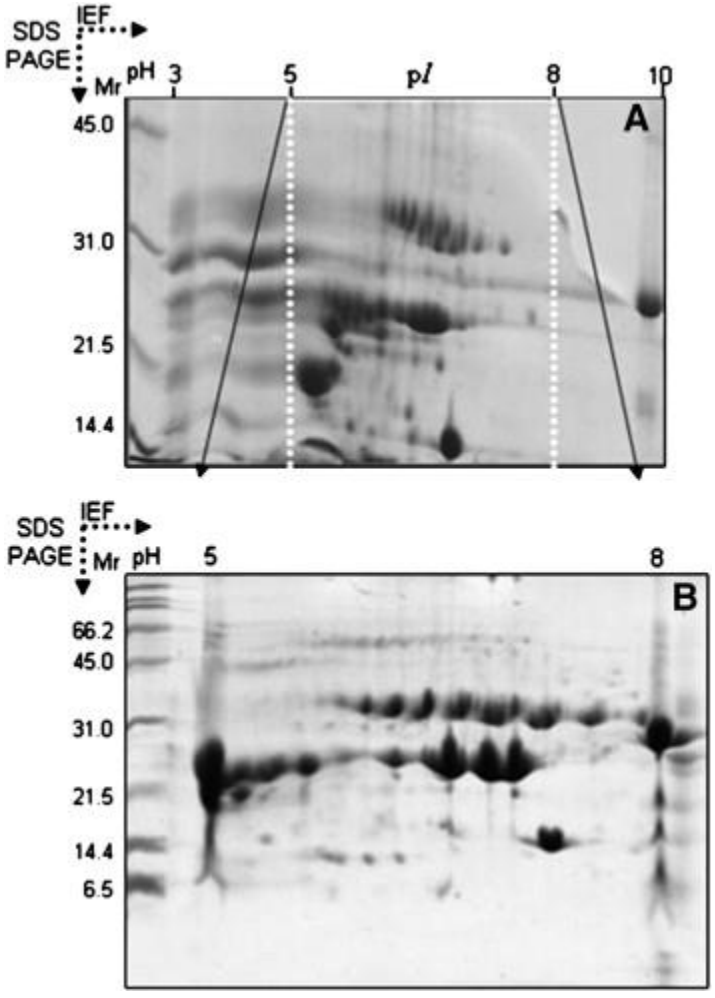
87	6.00±0.34	6.86±0.22	5.46±0.18	7.22±0.35
88	6.15±0.97	5.98±0.03	5.22±0.77	6.36±0.12
89	4.24±1.19	6.01±0.32	7.01±0.60	5.17±2.26
90	2.79±1.33	3.79±0.50	3.32±0.34	3.03±0.41
91	4.23±0.45	4.09±0.55	4.97±0.47	4.68±0.50
92	3.44±1.07	3.75±0.66	5.60±0.57	4.45±0.08
93	3.55±0.57	3.82±0.63	4.96±0.43	5.18±0.50
94	3.01±1.03	3.38±0.38	3.15±0.36	3.89±0.71
95	3.04±1.13	3.19±0.26	2.78±0.27	3.50±0.80
96	6.00±1.05	6.83±0.23	5.71±0.67	6.92±0.44
97	6.49±1.72	7.41±0.04	7.90±0.62	6.09±1.44
98	4.98±0.30	5.50±0.04	6.02±1.19	4.42±0.74
99	5.25±0.98	6.21±0.40	4.10±0.41	6.91±0.52
100	5.26±0.24	5.18±0.57	4.23±0.29	6.92±0.29
101	5.35±0.51	5.18±1.52	4.17±0.63	6.25±0.68
102	6.14±0.20	4.59±2.24	5.30±0.49	6.52±0.31
103	10.73±1.08	10.48±0.23	9.88±0.28	11.56±0.11
104	2.89±0.11	3.05±0.50	2.69±0.64	5.12±0.15
105	4.28±0.01	4.82±0.04	4.66±0.19	4.32±0.28
106	3.60±1.04	5.04±0.05	4.17±0.47	5.40±0.21
107	3.47±0.42	4.31±0.47	4.21±0.25	4.88±0.17
108	4.44±0.30	4.18±0.07	3.82±0.25	4.32±0.17
109	5.18±0.41	6.50±0.07	7.51±0.08	6.63±0.32
110	5.92±0.61	6.79±0.17	5.82±0.29	5.16±0.52
111	5.24±0.23	6.78±0.51	6.30±0.87	4.76±0.62
112	6.31±0.52	6.97±0.40	5.63±0.19	7.85±0.55
113	3.96±0.86	4.66±0.11	4.25±0.47	5.41±0.42
114	12.54±0.58	13.81±0.64	15.91±0.65	12.46±1.23
115	9.12±0.22	7.60±0.22	7.91±0.40	7.96±0.19
116	1.65±0.56	2.47±0.11	3.24±0.15	2.51±1.27
117	4.71±0.07	5.82±0.09	4.71±0.37	5.71±0.39
118	2.05±0.26	2.76±0.36	2.49±0.89	3.87±0.60
119	4.51±0.23	4.50±0.13	6.38±0.33	4.26±0.36
120	12.40±1.99	17.39±0.95	14.57±1.16	14.580.38
121	5.61±0.31	7.45±0.24	6.09±0.13	6.85±0.10
122	5.33±0.29	6.46±0.21	7.72±0.78	4.47±0.56
123	5.91±0.91	7.13±0.27	9.59±0.07	4.11±0.47
124	6.72±0.08	8.50±0.24	8.34±1.62	8.00±0.27
125	10.91±0.57	8.95±0.86	8.87±0.71	11.62±0.51
126	4.74±0.07	4.52±0.10	4.14±0.14	3.93±0.54
127	4.59±0.20	4.16±0.25	3.55±0.12	4.32±0.34
128	5.55±0.37	4.32± 0.0	4.28±0.33	4.86±0.23
129	4.70±0.24	7.51±0.06	5.96±0.08	6.73±0.72
130	4.00±0.06	3.37±0.45	3.31±0.14	3.89±0.61
131	10.65±0.20	12.49±0.37	13.66±1.02	10.13±0.66

132	3.37±0.30	2.42±0.99	3.23±0.32	1.56±0.80
133	5.97±1.57	7.97±0.30	8.13±2.03	5.61±0.40
134	5.81±0.28	6.77±1.46	5.37±0.98	5.68±0.76
135	9.48±0.99	8.03±0.40	9.80±2.06	8.14±0.57
136	7.40±0.99	7.03±0.56	7.41±0.62	6.41±1.37
137	6.05±0.55	7.94±0.37	9.57±1.12	4.74±0.36
138	7.14±1.43	7.63±0.35	8.68±0.50	6.28±0.21
139	7.77±0.44	8.40±1.72	11.26±0.26	5.97±0.14
140	5.74±0.11	6.82±0.36	8.22±0.67	4.57±0.47
141	4.46±0.45	2.86±0.20	3.30±1.02	4.29±0.55
142	11.11±0.06	12.56±0.19	15.48±0.42	11.69±0.53
143	4.05±0.39	4.41±0.19	4.21±0.30	3.41±1.05
144	4.30±0.42	3.02±1.13	4.02±0.28	4.73±0.36
145	3.78±0.55	3.76±0.30	3.21±0.37	4.54±0.29
146	5.21±0.18	4.25±0.05	4.37±0.25	5.03±0.17
147	7.19±1.69	6.60±0.22	10.52±0.18	6.39±1.31
148	6.15±0.14	5.99±0.14	3.88±0.34	5.07±0.22
149	4.47±0.29	5.30±0.30	10.36±0.37	3.61±0.83
150	5.72±1.46	4.74±0.58	6.68±2.04	5.70±0.60
151	5.02±0.87	3.97±0.26	4.72±3.50	4.57±0.35
152	6.03±1.04	6.30±0.21	7.48±0.13	5.73±0.20
153	6.74±0.63	7.48±0.32	9.15±0.27	5.25±0.73
154	5.44±0.09	6.94±0.37	7.29±0.43	4.62±0.46
155	2.50±0.78	3.05±0.12	3.14±0.33	1.18±0.41
156	13.43±0.60	13.43±3.34	16.36±0.11	14.09±0.49
157	5.46±0.46	7.22±0.26	7.00±0.05	5.06±0.63
158	4.22±0.49	3.54±0.11	6.15±1.06	3.23±1.11
159	1.83±1.17	2.07±0.47	3.22±0.08	2.27±1.10
160	2.71±0.49	2.40±0.27	4.37±0.71	2.52±1.25
161	4.50±0.25	5.23±0.26	5.39±0.20	4.59±0.30

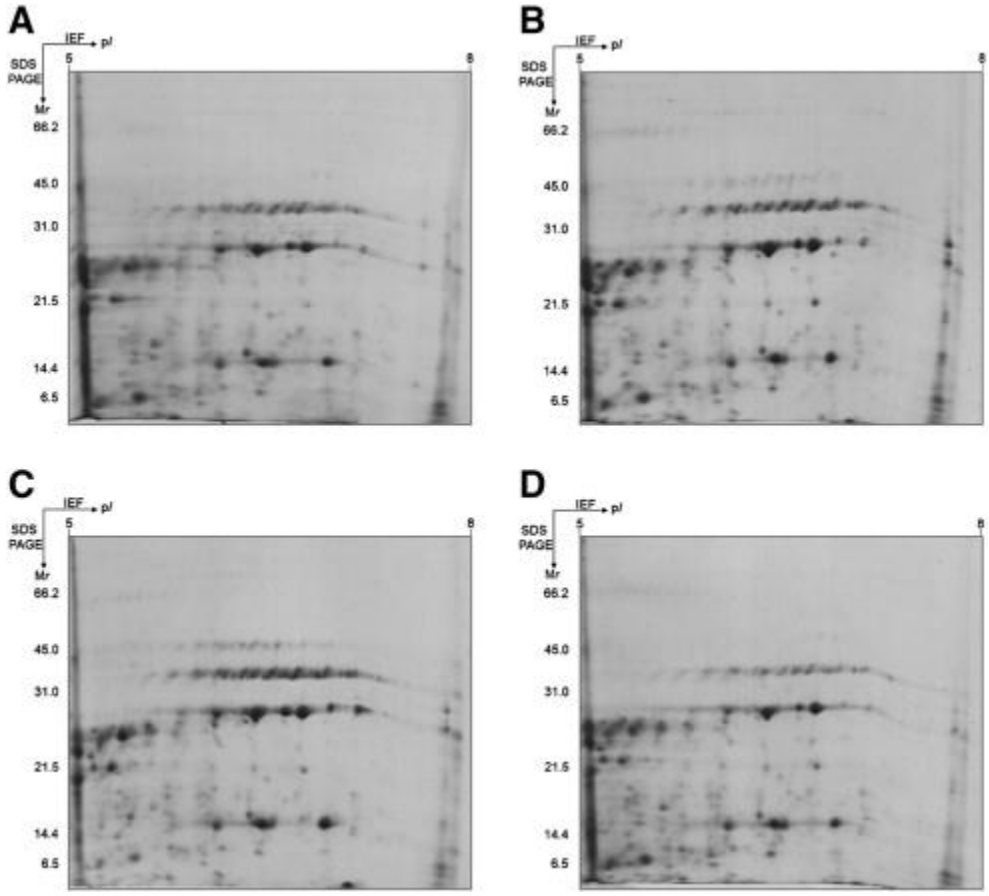
Supplementary Fig. 1. Spanish (inside square) and Andalusia maps with points indicating the locations where the ten populations were surveyed. For more data see Supplementary Table 1.



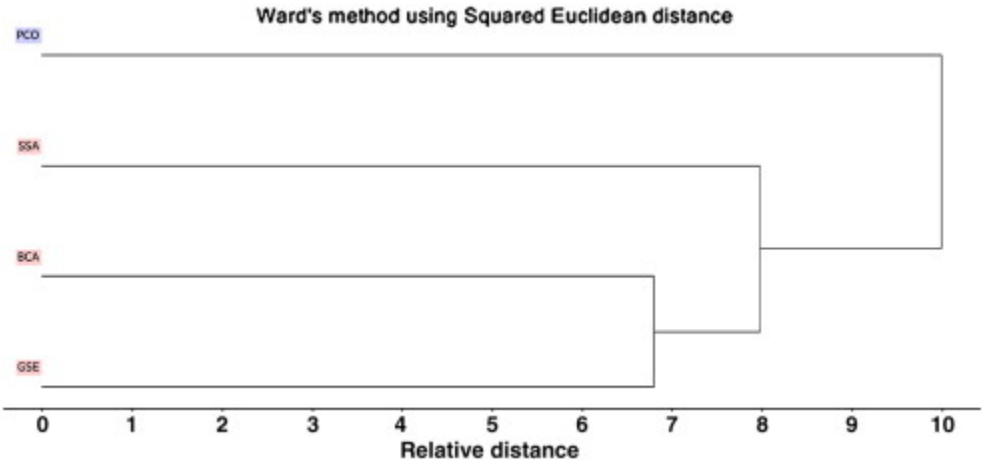
Supplementary Fig. 2. Picture of a representative acorn from the different populations (top), and flour powder after blade mill grinding (bottom). Population acronyms are indicated in Supplementary Table 1.



Supplementary Fig. 3. Representative 2-DE gel of *Quercus ilex* subsp. *ballota* flour extracts from mature seeds. Extracts were subjected to IEF on 7 cm 3–10 (A) or 5–8 (B) IPG strips (first dimension) and 12 (A) or 13 (B) % SDS-PAGE (second Dimension).



Supplementary Fig. 4. Real gels from extracts of seed flour from the four populations analyzed by 2-DE. (A) SSA, (B) GSE, (C) PCO, (D) BCA populations.

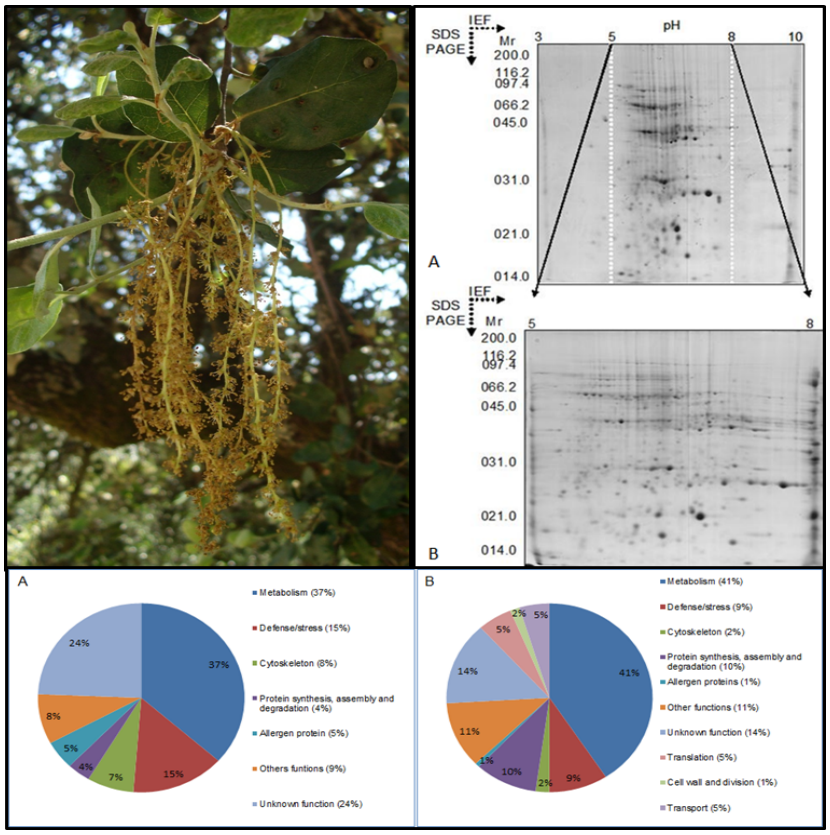


Supplementary Fig. 5. Graphical representation of the genetic distances between SDS-PAGE and 2-DE protein profiles established after Ward's method analysis using Squared Euclidian distance.

5. CAPÍTULO III

El estudio del proteoma de polen de encina.

Proteomic analysis of Holm oak (*Quercus ilex* subsp. *ballota* [Desf.] Samp.) pollen.



JOSÉ VALERO GALVÁN, LUIS VALLEDOR, RAQUEL GONZÁLEZ FERNÁNDEZ, RAFAEL Mª NAVARRO CERRILLO, JESUS V. JORRÍN-NOVO. 2012. Proteomic analysis of Holm oak (*Quercus ilex* subsp. *ballota* [Desf.] Samp.) pollen. Journal of Proteomics. 10.1016/j.jprot.2012.03.035.

JOURNAL OF PROTEOMICS XX (2012) XXX–XXX

Available online at www.sciencedirect.com

SciVerse ScienceDirect

www.elsevier.com/locate/jprot

Proteomic analysis of Holm oak (*Quercus ilex* subsp. *ballota* [Desf.] Samp.) pollen

José Valero Galván[a,d,*], *Luis Valledor*[b], *Raquel González Fernandez*[a], *Rafael Mª. Navarro Cerrillo*[c], *Jesus V. Jorrín-Novo*[a]

[a]*Agricultural and Plant Biochemistry and Proteomics Research Group, Dept. of Biochemistry and Molecular Biology, University of Córdoba, Agrifood Campus of International Excellence (ceiA3) 14071 Córdoba, Spain*
[b]*Department of Molecular Systems Biology (MOSYS), University of Vienna, Vienna, Austria*
[c]*Department of Forestry Engineering, ETSIAM, University of Córdoba, Agrifood Campus of International Excellence (ceiA3) 14071 Córdoba, Spain*
[d]*Department of Chemistry-Biology, Biomedical Sciences Institute, University of Juarez City, Chihuahua, Mexico*

ARTICLE INFO

Article history:
Received 13 December 2011
Accepted 21 March 2012

Keywords:
Holm oak
Quercus ilex
Pollen proteomics
Provenances variability

ABSTRACT

This paper presents an analysis of Holm oak pollen proteome, together with an evaluation of the potentiality that a proteomic approach may have in the provenance variability assessment. Proteins were extracted from pollen of four Holm oak provenances, and they were analyzed by gel-based (1- and 2-DE in combination with MALDI-TOF/TOF) and gel-free (nLC-LTQ Orbitrap MS) approaches. A comparison of 1- and 2-DE protein profiles of the four provenances revealed significant differences, both qualitative and quantitative, in abundance (18 bands and 16 spots, respectively). Multivariate statistical analysis carried out on bands and spots clearly showed distinct associations between provenances, which highlight their geographical origins. A total of 100 spots selected from the 402 spots observed on 2-DE gels were identified by MALDI-TOF/TOF. Moreover, a complementary gel-free shotgun approach was performed by nLC-LTQ Orbitrap MS. The identified proteins were classified according to biological processes, and most proteins in both approaches were related to metabolism and defense/stress processes. The nLC-LTQ Orbitrap MS analysis allowed us the identification of proteins belonging to the cell wall and division, transport and translation categories. Besides providing the first reference map of Holm oak pollen, our results confirm previous studies based on morphological observations and acorn proteomic analysis. Moreover, our data support the valuable use of proteomic techniques as phylogenetic tool in plant studies.

Abbreviations: SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; 1-DE, one dimensional electrophoresis; 2-DE, two dimensional electrophoresis; MALDI-TOF/TOF, matrix-assisted laser desorption/ionization time-of-flight/time-of-flight tandem mass spectrometry; nLC-LTQ Orbitrap MS, nano liquid chromatography electrospray ionization tandem mass spectrometry coupled to a Fourier Trap (Orbitrap); CHAPS, 3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonate; DTT, dithiothreitol; IEF, isoelectrofocusing; BSA, bovine serum albumin; TFA, trifluoroacetic acid; ACN, acetonitrile; CID, collision-induced dissociation; DW, dry weight; RG, Granada; GSE, Sevilla; BCA, Cádiz; SAA, Almería.

* *Corresponding author at:* Departamento de Bioquímica y Biología Molecular, Universidad de Córdoba, Campus de Rabanales, Ed. Severo Ochoa, E-14071 Córdoba, Spain. Fax: +34 957 218439.
E-mail address: bb2vagaj@uco.es (J. Valero Galván).
URL's: http://www.uco.es/investiga/grupos/probiveag/, http://www.uacj.mx/Paginas/UACJ.aspx (J. Valero Galván), http://www.univie.ac.at/mosys/ (L. Valledor), http://www.uco.es/investiga/grupos/probiveag/ (R. González Fernandez), http://www.uco.es/investiga/grupos/probiveag/ (J.V. Jorrín-Novo).

1874-3919/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.jprot.2012.03.035

1. Introduction

Holm oak (*Quercus ilex* subsp. *ballota* [Desf.] Samp.) is the dominant tree species in natural forest ecosystems over large areas of the Western Mediterranean Basin [1]. In Spain, it is widely used for forest systems and forest practice conservation, covering a surface of 2489000 ha approximately, which present estimated values of over 120 million Euro in fruit acorn production per year [2]. Holm oak acorns are a major component in the feeding systems of many Mediterranean wild and livestock species and, at the same time, they are the basic feed ingredient for domestically-bred pigs to obtain high quality meat [2].

Q. ilex subsp. *ballota*, like other oak species, has some features that favor a high intra-provenance genetic diversity: long life span, allogamy, wind pollination, monoecy, and a continuous geographical distribution [3]. Spanish Holm oak provenances have showed a prolonged isolation, in spite of being native and distributed throughout diverse forest areas with complex geographic patterns of genetic variation [4,5]. Some provenances have survived under the extreme dry climate conditions that prevail in southern Spain. However, and despite these differences, the classification of individual trees according to a given provenance based on phenotypes is quite challenging. In this context, proteomic studies have already proved their value to assess provenance variability in Andalusia Holm oaks. Proteomic approaches have been carried out to study in this species the following elements: i) the leaf proteome [6]; ii) the leaf proteome at several plant developmental stages with different provenances and under drought conditions [7,8]; and iii) the variability through acorn protein profile analysis [9]. Our results generally showed a remarkable difference between the different Andalusia provenances. According to these findings, we started a study focusing on the Holm oak pollen proteome, expected to be less variable than the leaf or acorn proteomes.

Quercus is considered a moderate cause of pollinosis in many areas of Europe, especially in those with high pollen levels [10]. Not only is the study of *Quercus* pollen useful because of their allergenic potential, but it is also essential for evaluating the ecological conservation of the Spanish woodland. Several pollen proteomic studies have been generally carried out in herbaceous plants [11–14], and to a lesser extent in forest trees and woody crops, focusing on allergen proteins [15–17]. An analysis of *Quercus ilex* subsp. *ballota* pollen proteome would help to gain insight into the molecular nature of allergens, as well as into other proteins involved in the functional specialization of oak pollen.

A comprehensive analysis of Holm oak pollen proteome has been performed by using complementary gel-based (1-DE and 2-DE with MALDI-TOF-TOF) and gel-free (nLC-LTQ Orbitrap MS) approaches. Moreover, we used the available protein, DNA and EST sequences accessible in public databases for protein identification. The potentiality of these approaches in assessing provenance variability has also been evaluated. Multivariate statistical analyses were performed to compare the 1- and 2-DE protein profiles of four provenances, and to establish phylogenetical distances among them. Holm oak pollen proteome has been partially described, with 355 protein species identified from peptide spectra techniques, allowing a functional protein characterization, as well as with the identification of provenance-specific protein markers.

2. Materials and methods

2.1. Plant material

Holm oak flowers were collected from two major Andalusia regions: south (SSA, RG and BCA) and northwest (GSE) (see reference [18] for details). Pollen grains were collected from freshly open flowers by shaking the anthers on a glass slide. Flower debris was removed using a microsieve, and pollen was visualized under a light microscope (Supplementary Fig. 1). Pollen was either used immediately, or stored at −70 °C after freezing in liquid nitrogen. Three different trees (biological replicates) per provenance were used for the protein extraction to observe their natural variability.

2.2. Protein extraction

Proteins were extracted from 500 mg of pollen using the TCA-phenol protocol [19,20]. The final pellet was solubilised in 100 μL of a solution of 7 M urea, 2 M thiourea, 4% (w/v) CHAPS, 0.5% (w/v) Triton X100, and 100 mM DTT. Insoluble material was removed by centrifugation, and the protein content in the supernatant was quantified according to the Bradford method [21] using BSA as standard. Samples were stored at −80 °C.

2.3. Gel electrophoresis

Protein extracts (70 μg BSA equivalent) were subjected to SDS-PAGE [22] on 13% polyacrylamide gels by using PROTEAN II cells (Bio-Rad, Hercules, USA). Gels were stained employing a Colloidal Coomassie procedure as reported in [23]. Images were digitized using a GS-800 Calibrated Densitometer (Bio-Rad, Hercules, USA), and then they were analyzed with Quantity One software (Bio-Rad, Hercules, USA).

Preliminary 2-DE experiments were made with the Mini-Protean 3 system (Bio-Rad, Hercules, USA), using 7-cm pH 3–10 linear gradient strips (Bio-Rad, Hercules, USA) and 13% polyacrylamide gels. The 2-DE analysis was carried out using the large system (17 cm). IEF was performed using 17-cm, 5–8 pH linear range, IPG strips (Bio-Rad, Hercules, USA). These strips were actively rehydrated applying 50 V for 12 h with 250 μL of a solution of 7 M urea, 2 M thiourea, 4% (w/v) CHAPS, 0.2% (v/v) IPG ampholite buffer pH 5–8, 100 mM DTT, and 0.01% (w/v) bromphenol blue, containing 300 μg BSA equivalent of proteins [24]. The strips were loaded onto a Protean IEF Cell system (Bio-Rad, Hercules, USA) and electrofocused at 20 °C using the following parameters: 250 V for 20 min, followed by 150 min linear gradient from 250 V to 10000 V, and finally focused on up to 40000 V at 10000 Vh. After IEF, the strips were immediately reduced and alkylated according to [24]. The second dimension was performed on 13% polyacrylamide gels using the Protean Dodeca Cell (Bio-Rad, Hercules, USA). The gels were run at 150 constant volts until the dye reached the bottom of the gel. Gels were stained employing

Please cite this article as: Valero Galván J, et al, Proteomic analysis of Holm oak (*Quercus ilex* subsp. *ballota* [Desf.] Samp.) pollen, J Prot (2012), doi:10.1016/j.jprot.2012.03.035

Colloidal Coomassie method [23]. Images were digitized using a GS-800 Calibrated Densitometer (Bio-Rad, Hercules, USA), and then they were analyzed with PD-Quest software v8.0 (Bio-Rad, Hercules, USA), using tenfold over background as the minimum criterion for presence/absence. The analysis was reevaluated by visual inspection, focusing on the spots which were present in all three biological replicates for each sample.

2.4. Phylogenetic and statistical analyses

Prior to statistical and phylogenetic analyses, the volume of pixels for each band or spot was normalized according to the total volume of bands detected (SDS-PAGE) or to the total volume of valid spots in each gel (2-DE) respectively, and then they were log-transformed. The web-based software NIA array analysis tool [25] (available at http://lgsun.grc.nia.nih.gov/anova/index.html) was used for statistical analysis, treatment and cluster analysis of protein abundance values was carried out by using and following the recommendations described by Valledor and Jorrin (2011) [26]. Data were statistically analyzed using the parameters defined by Brumbarova et al. [27]. This software tool selects statistically valid protein spots based on analysis of variance (ANOVA). After uploading the data and the indications about biological replicates and technical subreplicates, our data were statistically analyzed using the following settings: error model 'max (average, actual)', 0.01 proportion of highest variance values to be removed before variance averaging, 10° of freedom for the Bayesian error model, 0.05 FDR threshold, and zero permutations. A multivariate analysis was carried out in two steps: firstly, hierarchical clustering was performed to check the entire dataset, and the results were represented in dendrograms using the cluster function of the software; secondly, the entire dataset was analyzed by PCA. The settings used for the PCA analysis were: co-variance matrix type, three principal components, one fold change and 0.4 correlation threshold for clusters. PCA results are shown as a biplot (Fig. 2).

2.5. MALDI-TOF/TOF analysis

Spots were manually excised using a scalpel. Gel plugs were digested with modified porcine trypsin (sequencing grade; Promega), by using an automatic ProGest digestion station (Genomics Solution) according to Schevchenko et al. [28], with minor variations. The conditions were: twice destained steps for 30 min with 200 mM ammonium bicarbonate in 40% ACN at 37 °C; twice washed with 25 mM ammonium bicarbonate for 5 min and 25 mM ammonium bicarbonate in 50% ACN for 15 min respectively; dehydration with 100% ACN for 5 min and sample dried; hydration using 10 μL trypsin in a solution of 25 mM ammonium bicarbonate at a final concentration of 12.5 ng/μL for 10 min a room temperature, and the digestion was proceeded at 37 °C for 12 h. Subsequently, digestion was stopped adding 10 μL of a solution of 0.5% TFA in water. Tryptic peptides were purified in an automatic ProMS station (Genomic Solutions) by using a resin C18 microcolumn (ZipTip, Millipore), and they were eluted directly with a matrix solution (α-cyano hydroxycinnamic acid at a concentration of 5 mg/mL in 70% ACN/0.1% TFA) on MALDI plaque in 1 μL of final volume. After the cocrystallization on plaque, samples were analyzed by MALDI-TOF/TOF mass spectrometry to obtain the peptide mass fingerprinting (MS) in a 4800 Proteomics Analyzer (Applied Biosystems). The settings were: 800 to 4000 m/z range, with an accelerating voltage of 20 kV, in reflection mode, with delayed extraction set to "on", and an elapsed time of 120 ns. Spectra were internally calibrated with peptides from trypsin autolysis (M^+H^+=842.509, M^+H^+=2211.104) with an m/z precision of ± 20 ppm. Most abundant peptide ions were subjected to MS/MS analysis, providing information that can be used to define the peptide sequence.

A combined search (PMF and MS/MS) was performed with GPS Explorer™ software v3.5 (Applied Biosystems) over non-redundant NCBI databases using the MASCOT search engine (Matrix Science Ltd., London; http://www.matrixscience.com). The database search utilized the following parameters: taxonomy restrictions to *Viridiplantae* (02.03.2010), one missed cleavage sites, 100 ppm mass tolerance in MS and 0.5 Da for MS/MS 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, and confirmed by the accurate overlapping of the matched peptides with the major peaks of the mass spectrum.

2.6. nLC-LTQ Orbitrap MS

Moreover, an analysis was made of the protein content of pollen by a gel-free shotgun approach using a Finnigan Surveyor HPLC system, coupled with a LTQ-Orbitrap XL (Thermo Fisher Scientific, USA). In this case, a pollen pool was made with 25 mg of pollen from each replicate of the four provenances. Proteins obtained by the TCA-phenol method mentioned above were resolubilized in 100 μL of a solution of 6 M urea and 40 mM $(NH_4)HCO_3$. Fifty micrograms of proteins were adjusted to a final volume of 200 μL with a solution of 25 mM ammonium bicarbonate, reduced with 10 mM DTT and alkylated with 15 mM iodoacetamide. Proteins were digested using modified porcine trypsin (sequencing grade; Promega) at final concentration of 12.5 ng/μL, and at 37 °C overnight. Digestion was stopped adding a solution of TFA to a final concentration of 0.5%. Peptides were cleaned and concentrated in a C18 Cartridges Octadecyl C18/18% (Apllied Separations). Peptides were dried in a Speed-vac® Concentrator and resuspended in 50 μL of a 5% ACN and 0.1% formic acid solution. Five microliters of peptide mixtures were loaded onto a C18 column (BioBasic C18, 75 μm × 10 cm, particle size 5 μm, Thermo Fisher Scientific, USA) to achieve an elution flow of 0.3 μL min^{-1} using an autosampler. Peptides were eluted during a 0–40% gradient (Buffer A, 0.1% formic acid and 5% ACN; Buffer B, 0.1% formic acid and 95% ACN) over 60 min and detected in LTQ-Orbitrap equipped with a nanoelectrospray ion source (nESI). The general mass spectrometric conditions were: spray voltage, 2 kV; no sheath and auxiliary gas flow; ion transfer tube temperature was of 170 °C; the capillary voltage was set to 39 V and tube lens value was 180 V. The Orbitrap was operated at resolution settings of 30000 and full-scan spectra over the m/z range 400–1500 were acquired in the positive ion mode. The mass spectrometry method used was double play with CID activation; the normalized collision energy was 35; the activation time was 35 ms; the isolation width was 3.0; and the activation Q value

Please cite this article as: Valero Galván J, et al, Proteomic analysis of Holm oak (*Quercus ilex* subsp. *ballota* [Desf.] Samp.) pollen, J Prot (2012), doi:10.1016/j.jprot.2012.03.035

was 0.25. Fragmented more intense ions were 5 and dynamic exclusion was set to "on". Acquired data were analyzed with Proteome Discoverer v1.3 software (Thermo Fisher Scientific, USA). Since commercially available databases lack *Quercus* sequences, a new in-house database was designed for protein identification. This database consists in Uniprot ReSeq, *Arabidopsis* protein (Tair 11; www.arabidopsis.org), *Populus* and *Eucaliptus* protein (Phytozome 8; www.phytozome.org) and the six-frame translation of all of the EST sequences from *Quercus* genre available at Genebank (http://www.ncbi.nlm.nih.gov/genbank/) and The Gene Index (http://compbio.dfci.harvard.edu/tgi/). Peptide search was performed using Proteome Discovery 1.3 (Thermo) and SEQUEST algorithm with the following parameters: a maximum of 2 missed cleavage sites, acetylation of N-terminus, oxidation of methionine and carbamidomethyl cysteine formation, with a mass tolerance of 10 ppm for the parent ion and 0.8 Da for the fragment ion. Only proteins with at least one unique peptide with an XCor>charge state +0.5 (FDR<0.01) were considered as significant hits. Theoretical pIs were calculated using Compute pI/Mw tool from ExPAXy Bioinformatics Resource Portal (http://expasy.org/tools/). Unnamed or unknown proteins were subjected to a sequence similarity search using BLAST tool from UniprotKB database (http://www.uniprot.org/?tab=blast) with taxonomy restrictions to plants (UniProt release 2012_02).

3. Results and discussion

We are developing a research project focused on the study of natural variability in Holm oak provenances [6–9,18,29]. In this context, a high variability within trees, and between trees and provenances, has been reported when analyzing the leaf [8] and acorn [9,29] proteomes. We have collected pollen with the objective of characterizing and cataloging Andalusia Holm oak provenances. This is due to the fact that pollen is supposed to have a less variable proteome than other plant tissues, like leaves and acorns. Provenances sampled for this study are distributed geographically into two major Andalusia regions: south (SSA, RG and BCA) and northwest (GSE). Climate and physiographic variability are contrasted on these areas [18].

Provenances were compared according to protein yield, number and intensity of bands (SDS-PAGE) or spots (2-DE). Statistical analysis carried out on the protein yield ($p \leq 0.01$), and depending on the provenances, ranged from 8.10±0.28 mg/g DW (GSE provenances) to 14.60±0.44 mg/g DW (BCA provenances) (Fig. 1). These values were in the range of those reported for different plant species (including *Quercus* spp.), and they could represent just a small fraction (10%) of the total protein content in pollen, as reported in the literature or determined by Kjeldahl methodology [30,31]. Protein content, differently from that reported for acorns [9], did not show a significant correlation according to climate and geographical conditions where provenances are located.

3.1. SDS-PAGE protein profile

In total, a mean of 48 bands were resolved by SDS-PAGE in the 6–200 kDa Mr range (Supplementary Fig. 2). Mean relative intensity values for each band, as well as their SDs, are provided

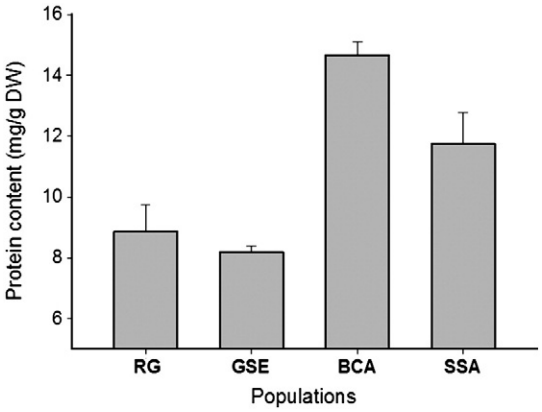


Fig. 1 – Differences in pollen protein contents from four Andalusia Holm oak provenances.

in Supplementary Table 1. Bands 2, 3, 5, 8, 10, 17, 26, 33 and 43 showed qualitative differences among provenances, that is, band intensity below detection limit in at least one provenance. Thus, band 2 was not detected in GSE provenances; bands 3, 5, and 10 were not detected in RG and GSE; while bands 8 and 26 were not detected in RG; and finally bands 17 and 33 were not detected in GSE. Bands 7, 9, 16, 32, 37, 39, 41, 45, and 49 showed quantitative differences (band common to all provenances but with different intensities; Duncan test, $p<0.05$) (Supplementary Fig. 2). Data were processed to select protein spots with significant and reproducible patterns across the analytical repetitions by using the web-based NIA array analysis software tool developed by Sharov and coworkers [25]. Data were subjected to multivariate statistical and clustering analysis in order to establish groups of provenances and distances among them. We used NIA array analysis tool for hierarchical clustering and PCA analyses (Fig. 2A). PCA results showed that PC1 and PC2 explained 42.52% and 38.87% of total variance, respectively. Both analyses showed that SSA and BCA provenances had similar protein expression profiles. However, RG provenance protein profiles tend to cluster to the previous group, while GSE provenance was separated from the rest of the provenances. Despite no many specific studies have been carried out on the use of pollen protein profiles to study variability in Holm oak provenances, an analysis of four distinct *A. thaliana* genotypes revealed no reproducible differences between samples derived from Col-0 wild-type plants and Atmlo5/Atmlo9 double mutants [32]. In contrast, significant differences between proteomic profiling of birch (*Betula verrucosa*) pollen extracts from different origins have been observed [33]. Our data were in agreement with the findings of Valero et al. [18], which reported the close relation of between these provenances based on acorn morphometry and chemical composition (see Fig. 4 of Valero et al. [18]). Hence, SSA, BCA and RG provenances showed a tendency to join in the same group, while the GSE provenance tended to cluster with the provenances that grow in the north of Andalusia. Furthermore, our results were in agreement with the study of Valero et al. [9] which, comparing the acorn protein profiles of these provenances, also showed their tendency to cluster in the same direction [9]. BCA provenance was clustered with SSA and RG

Please cite this article as: Valero Galván J, et al, Proteomic analysis of Holm oak (*Quercus ilex* subsp. *ballota* [Desf.] Samp.) pollen, J Prot (2012), doi:10.1016/j.jprot.2012.03.035

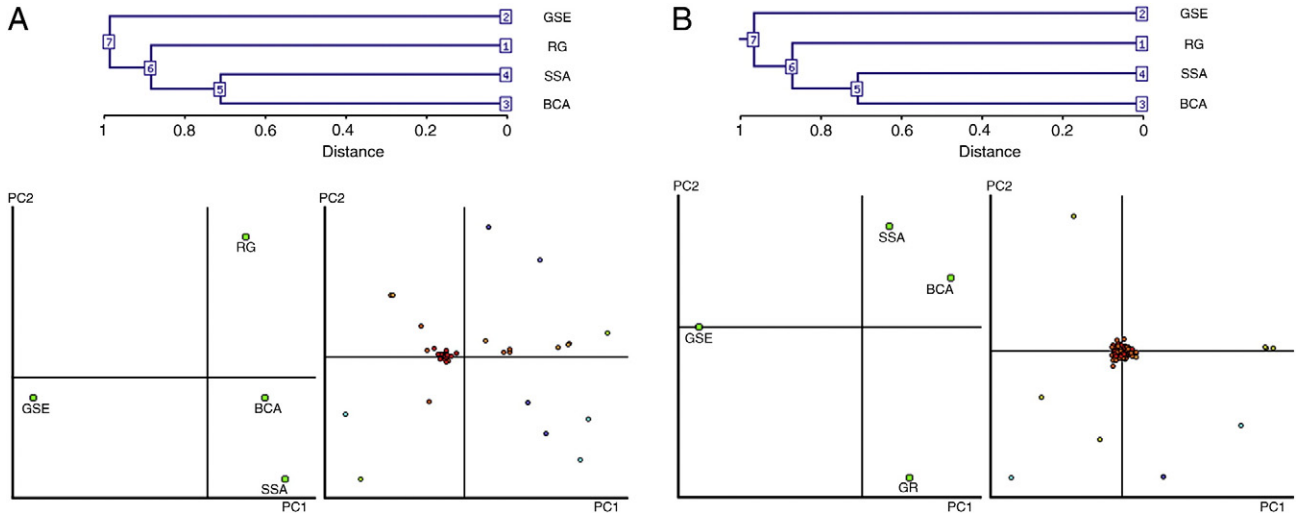


Fig. 2 – Associations between experimental samples and protein bands/spots generated by cluster analysis (top) and PCA (lower) from pollen protein SDS-PAGE (A) and 2-DE (B) data of the four Andalusia Holm oak provenances. PCA of samples (left graph) and protein bands/spots (right graph) was plotted in the first two component spaces. A short distance between samples and protein spots in the component space is indicative of similarity in their expression profiles.

group, while GSE showed a tendency to cluster with nearby provenances geographically (for most details see Fig. 3 of Valero et al. [9]).

3.2. *2-DE protein profile*

In preliminary 2-DE experiments with small 3–10 IPG range strips, up to 300 reproducible spots with p*I* from 4 to 10 and Mr up to 200 kDa were detected (Supplementary Fig. 3A), with most of them concentrated in the 5–8 pH region. Considering this, and aiming to increase the resolution by preventing protein co-migration as far as possible, IEF was performed in 17 cm gels by using 5–8 pH linear range strips (Supplementary Fig. 3B). Under these conditions, up to 600 spots were resolved. This observation differs from other pollen proteomic works published previously, in which 4–7 pH strips were used [11,13]. The figure for the Holm oak pollen was higher than that reported for tea (328 spots) [34] and for *Arabidopsis* (499 spots) [35], but lower than that reported in eastern white pine (645 spots) [36], tomato (960 spots) [14] and germinating rice pollen (2300 spots) [37]. These differences may be attributed to the methodology employed, although biological differences cannot be discarded.

The availability of a 2-DE reference map of Holm oak pollen is an unavoidable requisite to study the natural diversity of various provenances. According to this, we compared the pollen 2-DE profile of four different Andalusia provenances (representative gels are shown in Supplementary Fig. 4). A master gel containing all the resolved spots is presented in Fig. 3. For each provenance, the number of consistent spots (present in all the replicates) was similar (402±2) (Supplementary Table 2). Statistical analysis (ANOVA test) of the consistent spots showed that 9 spots had qualitative differences (spot intensity below detection limit in at least one provenances: spots 4, 5, and 10–16), while 7 spots had significant quantitative differences (common spots to all provenances but with different intensities: spots 74, 92, 252, 268, 385, 388 and 414) with a false discovery rate (FDR) <0.05. We used NIA array analysis tool for hierarchical clustering (Fig. 2B). Both analyses were consistent with SDS-PAGE distance data: SSA and BCA provenances had similar protein expression profiles, with which the RG provenance tends to cluster, while the GSE provenance was separated from the rest. PCA results showed that PC1 and PC2 explained 45.15% and 35.91% of total variance, respectively. PCA also provided information on the relevance of each protein related to the discrimination of different Holm oak populations. Correlationships between PCs and the different quantitative variable spots are indicated in Table 1.

3.3. *Protein identification*

A total of 16 variable spots, together with the most abundant constant spots covering all the Mr and p*I* range (84 spots), were excised from the gels and subjected to digestion with trypsin for further MALDI-TOF/TOF analysis. In addition, a gel-free shotgun (nLC-LTQ Orbitrap MS) strategy was employed for peptide analysis in protein extracts. A total of 77 spots provided confident matches after MALDI-TOF/TOF analysis (Supplementary Table 3, Supplementary Data 1 and 2 and Fig. 2), while 273 proteins were identified by nLC-LTQ Orbitrap MS (Supplementary Table 4). Confident matched proteins, grouped according to their biological functions, are included in both tables. We obtained, as happened in other orphan tree species [33,36,38], a low percentage of identification, which is related to the absence of protein sequences in databases. Nevertheless, in other studies, such as those on *Arabidopsis* pollen, a mean of 189 protein spots has been identified [35]. Predicted molecular masses and p*I*s for most of the identified proteins were generally consistent with the experimental masses and p*I*s (Supplementary Table 3). However, we observed some proteins

Please cite this article as: Valero Galván J, et al, Proteomic analysis of Holm oak (*Quercus ilex* subsp. *ballota* [Desf.] Samp.) pollen, J Prot (2012), doi:10.1016/j.jprot.2012.03.035

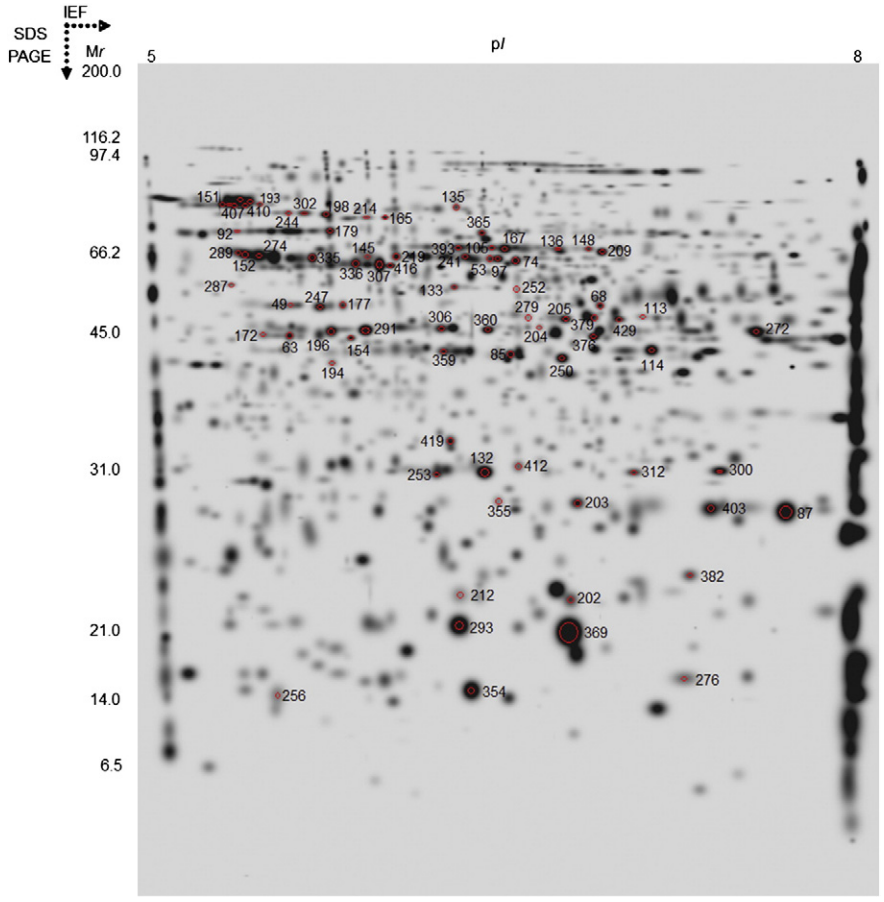


Fig. 3 – Master 2-DE gel of pollen extracts from *Quercus ilex* subsp *ballota*. Proteins were separated on 17 cm, pH 5–8 IPG strips (first dimension), and 13% SDS-PAGE (second dimension). On the left, the position of molecular weight marker proteins is shown. Identified protein spots are numbered according to Supplementary Table 3.

that had some deviations (e.g. spots 259, 266 and 350 that showed an apparent molecular mass higher than the mass of corresponding to identified proteins). We have also observed that some of the identified proteins were presented as two confident matches (e.g. spots 98, 298, 239, 29 and 152). Some of the identified proteins were present as multiple spots on 2-DE gels. These deviations in molecular masses and pIs, as well as the possibility of having multiple spots for the same protein, have been reported in other pollen proteome studies [11–13,32]. This situation could derive from various factors: i) errors in the database sequence (partial protein sequence) due to post-translational modifications, ii) protein degradation and partial synthesis of proteins during pollen maturation, iii) protein translation from alternatively spliced mRNAs, iv) protein homologues that may be unique to pollen and v) protein comigration.

Most of the variable spots among provenances provided no confident identifications, due to the fact that some of them produced poor quality spectra, or even no spectra or matches at all. Identification was possible for spot 252 (S-adenosylmethionine synthase), 268 (Hypothetical protein LOC100273659), 414

Table 1 – Correlationships between different quantitative variable spots and PCs.

Spots	Correlation	PC number	Direction	Populations			
				GR	GSE	BCA	SSA
4	−0.923	2	Negative	1.44	−10	−10	−10
5	0.973	1	Positive	1.71	−10	1.81	1.88
10	0.983	1	Positive	0.96	−10	1.51	0.91
11	0.976	1	Positive	1.02	−10	1.35	1.25
12	0.923	2	Positive	−10	2.17	2.07	2.15
13	0.71	1	Positive	0.63	−10	1.52	−10
14	−0.79	2	Negative	1.48	1.81	−10	−10
15	−0.741	3	Negative	1.83	1.12	1.31	−10
16	0.795	3	Positive	2.13	1.57	−10	1.65

Please cite this article as: Valero Galván J, et al, Proteomic analysis of Holm oak (*Quercus ilex* subsp. *ballota* [Desf.] Samp.) pollen, J Prot (2012), doi:10.1016/j.jprot.2012.03.035

(GDP-D-mannose 4,6 dehydratase 2), and 92 (Actin 3). These may be considered as provenance markers, but data must be validated by using a higher number of provenances, together with other approaches.

3.4. *Functional categories*

The proteins identified by both approaches were categorized into functional groups (Fig. 4). More than a half of the identified proteins by the MALDI-TOF/TOF analysis were grouped into metabolism (37%), defense/stress response (15%), and cytoskeleton (8%) categories (Fig. 4A), whereas those identified by nLC-LTQ Orbitrap MS were grouped into metabolism (41%), protein synthesis, assembly and degradation (10%), and defense/stress response (9%) (Fig. 4B). However, cell wall and division, transport, and transduction categories were identified by nLC-LTQ Orbitrap MS. Similar results have been reported in preliminary works on 2-DE based proteomic analysis of pollen from *A. thaliana* [12,13,31], tomato (*Lycopersicon esculentum*) [14], and rice (*O. sativa* L. ssp. japonica) [11,39], where most of the proteins identified were involved in metabolism, and defense against biotic and abiotic stresses [12,15,32].

Pollen germination and tube growth require a high rate of carbohydrate metabolism to meet energetic and biosynthetic demands [39,40]. In the present study, a high percentage of proteins were involved in carbohydrate and energy metabolisms (Supplementary Tables 3 and 4). Other pollen proteomic studies have also showed the presence of a large amount of energy-related proteins and metabolic enzymes in pollen [11,12,32]. Pollen grains are free-floating structures and are subjected to various abiotic and biotic stresses, including drought and extreme temperatures [14]. In this study, 12 proteins identified by MALDI-TOF–TOF and 26 by nLC-LTQ Orbitrap MS were related to defense/stress functions. These proteins were also reported in the pollen of *Camelia sinensis* [33], *Parietaria judaica* [41], *A. thaliana* [12,13], *L. esculentum* [14], and *Oryza sativa* [11]. Proteins corresponding to cytoskeleton have also been identified in Holm oak pollen (e.g. actin and profilin). Actin and profiling proteins also have been identified in the pollen of other species [11–13,32]. Pollen grains have been studied to identify allergen proteins in a number of species because of their allergenic activity in humans. *Quercus* is considered a moderate cause of pollinosis in many areas of Europe, especially in those with high pollen counts [10]. *Quercus* pollen production is very high, although it varies from species to species, and findings on the number of pollen grains per flower range may vary from 1.9×10^4 to 1.3×10^6 [42]. However, allergens in Holm oak pollen have been partly characterized and little information is available regarding to the extent of cross-reactivity between Holm oak pollen allergens [15,16,43]. In this study, four protein species were identified as pollen allergens: Que a 1 isoform (spot 229, 230 and 284), and Cas s 1 (spot 283). These allergenic proteins have also been characterized in *Quercus alba* [15,42,43]. However, a new allergen protein has been found by nLC-LTQ Orbitrap MS (allergenic isoflavone reductase-like protein Bet v 6.0102).

4. Conclusion

Q. ilex is the most representative and relevant species of the most significant Southern Europe and Andalusian forest ecosystem: “dehesa”. In this work, we present a study that is part of a research project which aims to understand the genetic, molecular and physiological bases of this species’ natural variability, its phenology, implantation, growth, and/or response to stresses. At the methodological level, a multidisciplinary approach is being performed, from physiology or classic biochemistry to the new -omic techniques (transcriptomics, proteomics and metabolomics), which define the biology of systems. Among these, high-throughput proteomic techniques allow a faster analysis of large numbers of proteins, and the use of complementary gel-based and gel-free approaches provides a better proteome characterization and comparison.

This paper reports a first approach to the proteome analysis of *Q. ilex* pollen. The combination of MALDI-TOF/TOF and nLC-LTQ Orbitrap MS analyses has enable to characterize the Holm oak pollen proteome partially. Many of these proteins have already been described from pollen of various flowering plants, while others are those typically associated with *Querqus* ssp. (e.g., Que a 1 isoform, Cas s 1, allergenic isoflavone reductase-like protein Bet v 6.0102). Hence, this study (along with others)

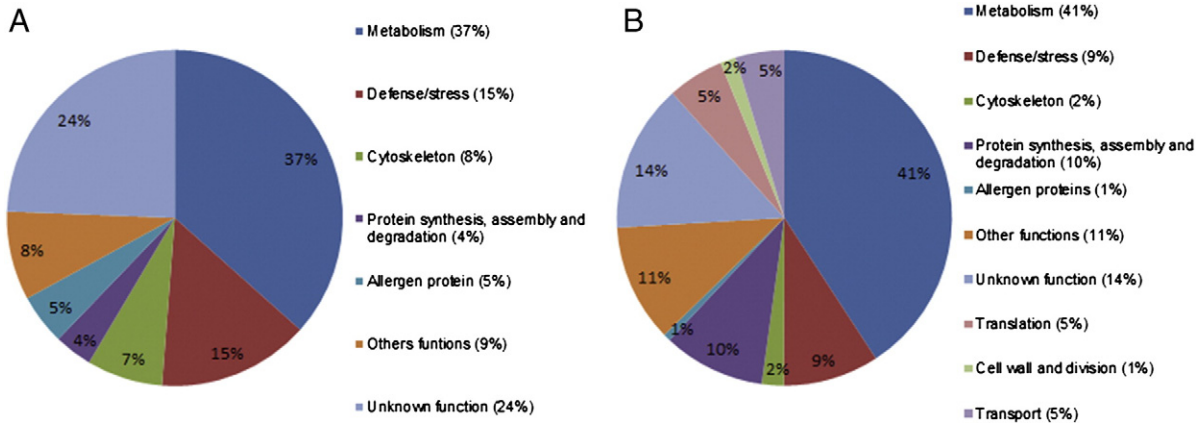


Fig. 4 – Functional categorization of proteins identified from mature Holm oak pollen. Classification of proteins identified by MALDI-TOF/TOF (A) and nLC-LTQ Orbitrap MS (B) analyses.

Please cite this article as: Valero Galván J, et al, Proteomic analysis of Holm oak (*Quercus ilex* subsp. *ballota* [Desf.] Samp.) pollen, J Prot (2012), doi:10.1016/j.jprot.2012.03.035

represents a significant contribution towards the construction of a comprehensive pollen proteome database encompassing many different species, which could serve as a valuable resource for researchers in plant biology in general, and in sexual plant reproduction in particular. Due to the fact that *Q. ilex* is an orphan tree species, the low percentage of identification obtained in this study is related to low spectral quality of some spots and/or the absence of protein sequences in databases. Current projects are concerned with *Q. ilex* genome sequencing. The more genes are sequenced, and these sequences annotated, the probability of characterizing the proteome will increase significantly.

Gel electrophoresis (1- and 2-DE) data, in combination with proper statistical tests (multivariate analysis), have allowed the establishment of groups and distances among provenances. Moreover, they are correlated to geographical tree location as it has been previously published [9,18]. These data support the valuable use of proteomic techniques as phylogenetic tools in plant studies, allowing the characterization and cataloging of Andalusia Holm oak provenances.

Supplementary materials related to this article can be found online at doi:10.1016/j.jprot.2012.03.035.

Acknowledgments

We would like to thank Dr. Samuel Ogueta, Dr. Consuelo Gómez, Carlos Fuentes and Raquel González (Proteomics Service, SCAI, University of Córdoba) for MALDI-TOF/TOF and nLC-LTQ Orbitrap MS analyses, and technical assistance.

Jose Valero was recipient of an Alban Program fellowship (I06D00010MX). This work is part of three research projects financed by the Spanish Ministry of Science and Innovation, cofinanced by the European Community FEDER funds: CGL2008-04503-C03-01/BOS, AGL2002-00530, and AGL2009-12243-C02-02.

REFERENCES

[1] Pulido FJ, Díaz M, Hidalgo de Trucios SJ. Size structure and regeneration of Spanish Holm oak *Quercus ilex* forests and dehesas: effects of agroforestry use on their long-term sustainability. Forest Ecol Manage 2001;146:1–13.

[2] Gea-Izquierdo G, Cañellas I, Montero G. Acorn production in Spanish Holm oak woodlands. Invest Agrar Sist Recur For 2006;15:339–54.

[3] Michaud H, Lumaret R, Romane F. Variation in the genetic structure and reproductive biology of holm oak populations. Plant Ecol 1992;99–100:107–13.

[4] Lumaret R, Mir C, Michaud H, Raynal V. Phylogeographical variation of chloroplast DNA in Holm oak (*Quercus ilex* L.). Mol Ecol 2002;11:2327–36.

[5] Petit RJ, Brewer S, Bordács S, Burg K, Cheddadi R, Coart E, et al. Identification of refugia and post-glacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence. Forest Ecol Manage 2002;156:49–74.

[6] Jorge I, Navarro Cerrillo RM, Lenz C, Ariza D, Porras C, Jorrin J. The Holm oak leaf proteome: analytical and biological variability in the protein expression level assessed by 2-DE and protein identification tandemmass spectrometry de novo sequencing and sequence similarity searching. Proteomics 2005;5:222–34.

[7] Echevarría-Zomeño S, Ariza D, Jorge I, Lenz C, Del Campo A, Jorrin 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.

[8] Jorge I, Navarro Cerrillo RM, Lenz C, Ariza D, Jorrin J. Variation in the Holm oak leaf proteome at different plant developmental stages, between provenances and in response to drought stress. Proteomics 2006;6(Suppl. 1):S207–14.

[9] Valero GJ, Valledor L, Navarro Cerrillo RM, Gil E, Jorrin JV. Studies of variability in Holm oak (*Quercus ilex* subsp. *ballota* [Desf.] Samp.) through acorn protein profile analysis. J Proteomics 2011;74:1244–55.

[10] Belmonte J, Roure J, March X. Aerobiology of Vigo, North-Western Spain: atmospheric pollen spectrum and annual dynamics of the most important taxa, and their clinical importance for allergy. Aerobiologia 1998;14:155–63.

[11] Dai S, Li L, Chen T, Chong K, Xue Y, Wang T. Proteomic analyses of *Oryza sativa* mature pollen reveal novel proteins associated with pollen germination and tube growth. Proteomics 2006;6:2504–29.

[12] Holmes-Davis R, Tanaka CK, Vensel WH, Hurkman WJ, McCormick S. Proteome mapping of mature pollen of *Arabidopsis thaliana*. Proteomics 2005;5:4864–84.

[13] Sheoran IS, Sproule K, Olson D, Ross A, Sawhney V. Proteome profile and functional classification of proteins in *Arabidopsis thaliana* (Landsberg erecta) mature pollen. Sex Plant Reprod 2006;19:185–96.

[14] Sheoran IS, Ross AR, Olson DJ, Sawhney VK. Proteomic analysis of tomato (*Lycopersicon esculentum*) pollen. J Exp Bot 2007;58:3525–35.

[15] Egger C, Focke M, Bircher AJ, Scherer K, Mothes-Luksch N, Horak F, et al. The allergen profile of beech and oak pollen. Clin Exp Allergy 2008;38:1688–96.

[16] Wallner M, Erler A, Hauser M, Klinglmayr E, Gadermaier G, Vogel L, et al. Immunologic characterization of isoforms of Car b 1 and Que a 1, the major hornbeam and oak pollen allergens. Allergy 2009;64:452–60.

[17] Napoli A, Aiello D, Di Donna L, Moschidis P, Sindona G. Vegetable proteomics: the detection of Ole e 1 isoallergens by peptide matching of MALDI MS/MS spectra of underivatized and dansylated glycopeptides. J Proteome Res 2008;7:2723–32.

[18] Valero GJ, Jorrín-Novo J, Gómez C, Ariza D, García-Olmo J, Navarro-Cerrillo RM. Population variability based on the morphometry and chemical composition of the acorn in Holm oak (*Quercus ilex* subsp. *ballota* [Desf.] Samp.). Eur J Forest Res 2011, doi:10.1007/s10342-011-0563–8.

[19] Wang W, Vignani R, Scali M, Cresti M. A universal and rapid protocol for protein extraction from recalcitrant plant tissues for proteomic analysis. Electrophoresis 2006;27:2782–6.

[20] Maldonado AM, Echevarría-Zomeño S, Jean-Baptiste S, Hernández M, Jorrín-Novo JV. Evaluation of three different protocols of protein extraction for Arabidopsis thaliana leaf proteome analysis by two-dimensional electrophoresis. J Proteomics 2008;71:461–72.

[21] Ramagli LS, Rodriguez LV. Quantitation of microgram amounts of protein in two-dimensional polyacrylamide gel electrophoresis sample buffer. Electrophoresis 1985;6:559–63.

[22] Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 1970;227:680–5.

[23] Mathesius U, Keijzers G, Natera SH, Weinman JJ, Djordjevic MA, Rolfe BG. Establishment of a root proteome reference map for the model legume *Medicago truncatula* using the expressed sequence tag database for peptide mass fingerprinting. Proteomics 2001;1:1424–40.

[24] Gorg A, Postel W, Baumer M, Weiss W. Two-dimensional polyacrylamide gel electrophoresis, with immobilized pH gradients in the first dimension, of barley seed proteins: discrimination of cultivars with different malting grades. Electrophoresis 1992;13:192–203.

Please cite this article as: Valero Galván J, et al, Proteomic analysis of Holm oak (*Quercus ilex* subsp. *ballota* [Desf.] Samp.) pollen, J Prot (2012), doi:10.1016/j.jprot.2012.03.035

[25] Sharov A, Dudekula D, Ko M. A web-based tool for principal component and significance analysis of microarray data. Bioinformatics 2005;21:2548–9.
[26] Valledor L, Jorrín J. Maximizing the information obtained by quantitative two dimensional gel electrophoresis analyses by an appropriate experimental design and statistical analyses. J Proteomics 2011;74:1–18.
[27] Brumbarova T, Matros A, Mock HP, Bauer P. A proteomic study showing differential regulation of stress, redox regulation and peroxidase proteins by iron supply and the transcription factor FER. Plant J 2008;54:321–34.
[28] Schevchenko A, Wilm M, Vorm O, Mann M. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. Anal Chem 1996;68:850–8.
[29] Martín M, Navarro-Cerrillo R, Ortega P, Alvarez J. The use of cotyledon proteins to assess the genetic diversity in sweet holm oak. J Forest Sci 2009;55:526–31.
[30] Saa-Otero MP, Díaz-Losada E, Fernández-Gómez E. Analysis of fatty acids, proteins and ethereal extract in honeybee pollen — considerations of their floral origin. Grana 2000;39: 175–81.
[31] Roulston TH, Cane JH, Buchmann SL. What governs protein content of pollen: pollinator preferences, pollen–pistil interactions, or phylogeny? Ecol Monogr 2000;70:617–43.
[32] Noir S, Bräutigam A, Colby T, Schmidt J, Panstruga R. A reference map of the *Arabidopsis thaliana* mature pollen proteome. Biochem Biophys Res Commun 2005;337:1257–66.
[33] Erler A, Hawranek T, Krückemeier L, Asam C, Egger M, Ferreira F, et al. Proteomic profiling of birch (*Betula verrucosa*) pollen extracts from different origins. Proteomics 2011;11: 1486–98.
[34] Li J, Chen J, Zhang Z, Pan Y. Proteome analysis of tea pollen (*Camellia sinensis*) under different storage conditions. J Agric Food Chem 2008;56:7535–44.
[35] Zou J, Song L, Zhang W, Wang Y, Ruan S, Wu WH. Comparative proteomic analysis of *Arabidopsis* mature pollen and germinated pollen. J Int Plant Biol 2009;51:438–55.
[36] Fernando DD. Characterization of pollen tube development in *Pinus strobus* (Eastern white pine) through proteomic analysis of differentially expressed proteins. Proteomics 2005;5(18): 4917–26.
[37] Dai S, Chen T, Chon K, Xue Y, Liu S, Wang T. Proteomics identification of differentially expressed proteins associated with pollen germination and tube growth reveals characteristics of germinated *Oryza sativa* pollen. Mol Cell Proteomics 2007;6:207–30.
[38] Abril N, Gion J-M, Kerner R, Müller-Starck G, Cerrillo RMN, Plomion C, et al. Proteomics research on forest trees, the most recalcitrant and orphan plant species. Phytochemistry 2011;72:1219–42.
[39] Dai S, Wang T, Yan X, Chen S. Proteomics of pollen development and germination. J Proteome Res 2007;6: 4556–63.
[40] Krichevsky A, Kozlovsky SV, Tian G-W, Chen M-H, Zaltsman A, Citovsky V. How pollen tubes grow. Dev Biol 2007;303: 405–20.
[41] Barranca M, Fontana S, Taverna S, Duro G, Zanella-Cleon I, Becchi M, et al. Proteomic analysis of *Parietaria judaica pollen* and allergen profiling by an immunoproteomic approach. Biotechnol Lett 2010;32:565–70.
[42] Molina RT, Rodríguez AM, Palaciso IS, López FG. Pollen production in anemophilous trees. Grana 1996;35:38–46.
[43] Niederberger V, Pauli G, Grönlundc H, Fröschla R, Rumpold H, Kraft D, et al. Recombinant birch pollen allergens (rBet v 1 and rBet v 2) contain most of the IgE epitopes present in birch, alder, hornbeam, hazel, and oak pollen: a quantitative IgE inhibition study with sera from different populations. J Allergy Clin Immunol 1998;102:579–91.

Please cite this article as: Valero Galván J, et al, Proteomic analysis of Holm oak (*Quercus ilex* subsp. *ballota* [Desf.] Samp.) pollen, J Prot (2012), doi:10.1016/j.jprot.2012.03.035

SUPPLEMENTARY TABLES

Supplementary Table 1. Normalized and transformed relative volumes of the 1-DE resolved bands, numbered according to Figure 1. Values are showed as the mean ± SD of three analytical replicates. n.d.: no detected band (qualitative differences). Mean values with the same letters ($^{a.b.c.d.e}$) indicate homogeneous subsets for α=0.05 according to Duncan test. (RG, Granada; GSE, Sevilla; BCA, Cádiz; SSA, Almería).

		Provenances				Statistical analysis	
Band number	Mr (kDa)	RG	GSE	BCA	SSA	*F*	*P*
1	155.353	2.56±0.24 ^{a}	3.26±1.03 ^{a}	2.46±0.35 ^{a}	2.25±0.78 ^{a}	1.23	0.35
2	142.913	1.13±0.05	n.d.	1.44±0.25	n.d.	102.46	0.00
3	132.901	n.d.	n.d.	1.45±0.32	0.47±0.81	7.35	0.01
4	121.265	2.95±0.13 ^{a}	3.11±0.80 ^{a}	2.65±0.38 ^{a}	2.20±0.61 ^{a}	1.62	0.25
5	112.395	n.d.	n.d.	1.78±0.19	1.48±0.22	127.16	0.00
6	105.117	2.88±0.28 ^{a}	3.40±0.90 ^{a}	3.45±0.19 ^{a}	3.23±0.60 ^{a}	0.62	0.62
7	99.821	3.00±0.59 ^{bc}	2.89±0.64 ^{b}	4.00±0.29 ^{c}	2.22±0.45 ^{a}	6.17	0.01
8	91.456	n.d.	0.91±0.83 ^{a}	1.71±1.48 ^{a}	0.48±0.83 ^{a}	1.76	0.23
9	83.510	3.14±0.47 ^{b}	3.10±0.61^{b}	3.78±0.45 ^{c}	0.81±1.40 ^{a}	7.46	0.01
10	76.229	n.d.	n.d.	3.17±2.80	2.66±2.31	2.60	0.12
11	74.171	3.97±0.80 ^{a}	4.08±0.89 ^{a}	4.48±0.24 ^{a}	4.01±0.97 ^{a}	0.27	0.84
12	67.130	3.45±0.43 ^{a}	4.33±1.07 ^{a}	4.11±0.18 ^{a}	3.61±0.61 ^{a}	1.16	0.38
13	63.525	1.70±0.87 ^{a}	0.99±1.71 ^{a}	1.96±0.41 ^{a}	1.69±0.55 ^{a}	0.50	0.69
14	60.566	2.78±0.59 ^{a}	3.37±0.64 ^{a}	2.80±0.21 ^{a}	2.16±0.21 ^{a}	3.46	0.07
15	58.343	3.78±0.16 ^{a}	4.35±0.49 ^{a}	3.90±0.22 ^{a}	3.25±0.91 ^{a}	2.16	0.17
16	53.321	5.08±0.19 ^{b}	5.58±0.26 ^{c}	5.08±0.03 ^{b}	4.49±0.58 ^{a}	5.45	0.02
17	51.787	1.85±3.20	n.d.	2.53±2.19	4.12±0.63	2.26	0.15
18	49.973	5.11±0.83 ^{a}	5.45±1.01 ^{a}	4.97±0.18 ^{a}	4.71±0.69 ^{a}	0.51	0.68
19	45.000	3.97±0.56 ^{a}	3.85±1.74 ^{a}	3.55±0.48 ^{a}	2.91±2.57 ^{a}	0.26	0.84
20	44.067	4.68±0.83 ^{a}	3.20±3.09 ^{a}	2.38±2.11 ^{a}	3.60±1.58 ^{a}	0.63	0.61
21	42.485	2.79±2.59 ^{a}	4.76±1.00 ^{a}	3.83±0.14 ^{a}	4.08±1.33 ^{a}	0.84	0.50
22	41.196	3.06±1.07 ^{a}	2.39±2.07 ^{a}	1.93±0.49 ^{a}	3.45±0.70 ^{a}	0.89	0.48
23	38.801	5.39±0.99 ^{a}	7.95±1.92 ^{a}	5.27±0.17 ^{a}	5.82±0.60 ^{a}	3.71	0.06
24	36.813	4.73±0.33 ^{a}	3.57±3.24 ^{a}	4.22±0.61 ^{a}	5.18±0.25 ^{a}	0.52	0.68
25	35.637	4.61±0.76 ^{a}	6.00±0.97 ^{a}	4.64±0.72 ^{a}	4.30±0.08 ^{a}	3.34	0.07
26	34.314	n.d.	4.52±0.70	1.29±2.24	4.29±0.44	10.50	0.00
27	32.896	4.60±0.10 ^{a}	2.97±2.59 ^{a}	4.36±0.66 ^{a}	3.97±0.45 ^{a}	0.84	0.50
28	31.494	2.98±0.13 ^{a}	3.68±0.84 ^{a}	3.50±0.13 ^{a}	3.11±0.58 ^{a}	1.21	0.36
29	31.139	3.74±0.38 ^{a}	4.75±1.46 ^{a}	3.92±0.63 ^{a}	3.46±0.06 ^{a}	1.40	0.31
30	27.687	3.92±0.33 ^{a}	4.08±0.51 ^{a}	4.66±0.39 ^{a}	3.58±0.65 ^{a}	2.54	0.12

31	26.939	3.83±0.08 [a]	3.17±0.89 [a]	3.47±0.43 [a]	2.39±0.57 [a]	3.42	0.07
32	25.682	2.22±0.33 [b]	0.46±0.79 [a]	2.57±0.35 [b]	2.02±0.67 [b]	7.99	0.00
33	25.127	2.55±0.18	n.d.	1.68±1.45	2.42±0.48	6.98	0.01
34	23.964	4.64±0.00 [a]	4.64±0.00 [a]	4.64±0.00 [a]	4.64±0.00 [a]	1.42	0.30
35	22.878	3.20±0.37 [a]	3.50±1.15 [a]	2.40±0.65 [a]	2.46±0.51 [a]	1.65	0.25
36	21.521	4.67±0.14 [a]	5.52±0.97 [a]	5.21±0.40 [a]	4.70±0.54 [a]	1.42	0.30
37	20.451	3.81±0.77 [b]	3.90±0.69 [b]	3.94±0.28 [b]	1.20±2.08 [a]	3.97	0.05
38	19.635	3.23±0.23 [a]	3.02±0.59 [a]	2.20±1.91 [a]	3.04±0.32 [a]	0.60	0.62
39	19.157	3.07±0.29 [c]	0.64±1.11 [a]	2.33±1.34 [b]	2.82±0.29 [b]	4.48	0.04
40	18.165	1.51±0.29 [a]	1.72±0.23 [a]	2.16±0.75 [a]	1.98±0.72 [a]	0.82	0.51
41	16.877	3.56±0.86 [c]	3.81±0.08 [c]	2.26±0.87 [b]	1.60±0.27 [a]	8.42	0.00
42	16.613	3.21±0.72 [a]	3.73±0.53 [a]	3.45±0.28 [a]	3.01±0.35 [a]	1.17	0.37
43	16.406	3.50±0.81	n.d.	n.d.	n.d.	55.38	0.00
44	16.166	4.80±1.49 [b]	6.05±1.04 [c]	4.70±0.12 [b]	3.23±0.62 [a]	4.32	0.04
45	15.804	5.10±0.99 [a]	6.92±1.17 [d]	6.01±0.08 [c]	4.67±0.12 [b]	5.10	0.02
46	14.988	7.09±0.54 [a]	3.89±2.08 [a]	4.37±2.20 [a]	6.13±0.94 [a]	2.59	0.12
47	13.889	4.19±1.07 [a]	4.00±0.48 [a]	4.69±0.26 [a]	4.01±0.84 [a]	0.57	0.64
48	13.507	5.98±0.23 [a]	4.53±0.76 [a]	5.12±1.22 [a]	5.78±0.77 [a]	1.93	0.20
49	12.321	4.96±0.45 [c]	3.38±1.28 [b]	2.21±0.75 [a]	3.37±0.80 [b]	5.01	0.03
50	11.381	2.10±0.49 [a]	2.73±1.04 [a]	2.25±0.60 [a]	3.07±0.21 [a]	1.39	0.31

Supplementary Table 2. Normalized and transformed relative volumes of the 2-DE resolved spots. Values are showed as the mean ± SD of three biological replicates. n.d. no detected spot (qualitative differences). (RG, Granada; GSE, Sevilla; BCA, Cádiz; SSA, Almería).

	Provenances				Statistical analysis	
Spot	RG	GSE	BCA	SSA	P	FDR
4	1.44±0.11	n.d.	n.d.	n.d.	0.00	0.00
5	1.71±0.48	n.d.	1.81±0.36	1.88±0.10	0.00	0.00
10	0.96±0.09	n.d.	1.52±0.34	0.91±0.32	0.00	0.00
11	1.02±0.12	n.d.	1.35±0.20	1.25±0.25	0.00	0.00
12	n.d.	2.17±0.08	2.08±0.12	2.16±0.07	0.00	0.00
13	0.63±0.05	n.d.	1.52±0.18	n.d.	0.00	0.00
14	1.49±0.42	1.81±0.27	n.d.	n.d.	0.00	0.00
15	1.84±0.06	1.12±0.70	1.32±0.13	n.d.	0.00	0.00
16	2.14±0.03	1.57±0.32	n.d.	1.66±0.04	0.00	0.00
17	2.45±0.10	2.50±0.09	2.12±0.13	1.81±0.30	0.00	0.06
18	2.27±0.17	2.29±0.27	2.26±0.27	2.25±0.24	1.00	1.00
19	2.46±0.15	2.36±0.19	2.39±0.07	2.47±0.10	0.93	1.00
20	2.41±0.06	1.74±0.56	2.17±0.36	1.82±0.67	0.27	0.89
21	2.05±0.17	2.05±0.33	1.99±0.25	1.97±0.07	0.98	1.00
22	2.33±0.14	2.23±0.41	2.31±0.16	2.36±0.07	0.94	1.00
23	2.35±0.11	2.49±0.40	1.90±0.18	2.46±0.17	0.01	0.16
24	2.27±0.09	1.77±0.79	1.99±0.08	1.91±0.64	0.68	1.00
25	2.43±0.16	2.70±0.03	2.32±0.08	2.38±0.04	0.23	0.88
26	2.01±0.15	1.57±0.74	1.81±0.30	2.00±0.11	0.51	0.95
27	2.26±0.32	2.00±0.59	2.10±0.27	2.21±0.12	0.83	1.00
28	1.86±0.28	1.13±0.59	1.46±0.81	1.30±0.44	0.43	0.93
29	2.13±0.17	1.84±0.47	1.51±0.24	1.34±0.84	0.22	0.87
30	2.47±0.06	2.26±0.19	2.25±0.18	2.36±0.07	0.67	1.00
31	1.57±0.73	1.99±0.65	1.83±0.37	1.75±0.51	0.85	1.00
32	2.62±0.21	2.77±0.25	2.65±0.30	2.80±0.08	0.74	1.00
33	2.36±0.11	2.55±0.34	2.28±0.18	2.56±0.19	0.39	0.93
34	2.31±0.09	2.39±0.38	2.18±0.04	2.26±0.25	0.75	1.00
35	1.74±0.11	1.25±0.24	1.25±0.15	1.20±0.46	0.07	0.53
36	1.96±0.04	2.08±0.15	1.75±0.26	1.97±0.57	0.63	0.99
37	1.09±0.34	0.51±0.38	0.81±0.10	0.75±0.43	0.21	0.84
38	1.12±1.16	1.23±1.94	1.53±0.25	2.43±0.44	0.50	0.95
39	2.34±0.05	2.47±0.06	2.46±0.28	2.57±0.33	0.70	1.00
40	1.94±0.48	2.39±0.45	1.87±0.43	1.88±0.18	0.33	0.91
41	1.99±0.42	2.37±0.23	2.26±0.14	2.64±0.19	0.03	0.31
42	1.95±0.08	1.85±0.22	1.91±0.33	1.72±0.10	0.75	1.00
43	1.71±0.29	1.49±0.31	1.63±0.08	1.55±0.48	0.85	1.00
44	2.08±0.16	2.49±0.14	2.19±0.17	2.39±0.25	0.16	0.80
45	2.05±0.32	1.91±0.54	1.99±0.34	2.07±0.43	0.96	1.00
46	2.71±0.19	2.14±0.62	2.66±0.37	3.11±0.12	0.02	0.24
47	0.86±0.21	0.86±0.25	0.75±0.65	0.98±0.18	0.76	1.00
48	1.16±0.28	0.98±0.26	1.14±0.11	1.01±0.22	0.83	1.00

49	1.29±0.08	1.13±0.26	1.28±0.15	1.26±0.11	0.90	1.00
50	1.51±0.01	1.44±0.13	1.62±0.03	1.69±0.09	0.68	1.00
51	1.81±0.35	2.23±0.35	2.13±0.15	2.14±0.03	0.22	0.87
52	2.14±0.40	2.10±0.09	1.95±0.34	2.29±0.25	0.55	0.95
53	2.27±0.24	2.16±0.32	2.21±0.06	2.33±0.16	0.85	1.00
54	2.13±0.25	1.92±0.12	1.97±0.12	2.18±0.10	0.55	0.95
55	1.87±0.18	1.82±0.14	1.66±0.13	1.81±0.03	0.81	1.00
56	1.93±0.12	1.86±0.09	1.83±0.08	1.99±0.03	0.90	1.00
57	1.78±0.31	2.03±0.10	2.06±0.17	1.11±1.03	0.30	0.89
58	1.85±0.08	1.92±0.64	1.96±0.14	1.84±0.16	0.96	1.00
59	1.72±0.05	1.17±0.22	1.40±0.11	1.45±0.25	0.13	0.72
60	1.97±0.03	2.03±0.16	1.81±0.23	2.01±0.07	0.77	1.00
61	2.19±0.02	1.98±0.78	2.45±0.13	2.27±0.16	0.55	0.95
62	1.86±0.12	1.60±0.12	1.63±0.28	1.91±0.10	0.39	0.93
63	1.69±0.26	1.29±0.70	1.42±0.16	1.71±0.11	0.47	0.94
64	1.76±0.14	1.80±0.45	1.88±0.18	1.76±0.37	0.96	1.00
65	1.47±0.12	0.90±0.75	1.28±0.22	0.79±0.33	0.18	0.81
66	1.73±0.27	1.86±0.22	1.58±0.03	1.67±0.46	0.69	1.00
67	1.55±0.72	2.54±0.11	2.24±0.14	1.91±0.16	0.01	0.16
68	2.08±0.14	1.91±0.42	1.72±0.04	1.84±0.27	0.44	0.94
69	1.98±0.01	1.92±0.22	1.61±0.20	1.53±0.22	0.11	0.66
70	1.64±0.15	1.60±0.10	1.46±0.04	1.90±0.17	0.24	0.89
71	1.99±0.10	1.79±0.37	1.73±0.38	1.86±0.20	0.72	1.00
72	1.47±0.06	1.66±0.07	1.52±0.70	1.56±0.16	0.93	1.00
73	2.23±0.03	2.03±0.02	1.96±0.22	2.00±0.12	0.58	0.97
74	1.54±0.23	1.71±0.30	1.48±0.64	0.56±0.33	0.00	0.06
75	1.63±0.15	1.20±0.55	1.63±0.09	1.78±0.10	0.08	0.57
76	2.15±0.12	2.08±0.14	1.94±0.15	1.83±0.10	0.45	0.94
77	1.08±0.28	1.65±0.40	1.46±0.01	1.60±0.17	0.06	0.51
78	1.52±0.05	1.51±0.22	1.65±0.12	1.64±0.20	0.88	1.00
79	1.53±0.15	1.90±0.48	1.74±0.06	1.65±0.56	0.67	1.00
80	1.68±0.18	1.46±0.13	1.71±0.45	1.78±0.07	0.52	0.95
81	2.03±0.15	2.07±0.16	2.05±0.13	2.11±0.04	0.98	1.00
82	1.85±0.05	1.33±0.81	1.74±0.12	1.81±0.10	0.40	0.93
83	1.43±0.10	1.79±0.17	1.68±0.24	1.49±0.16	0.35	0.93
84	1.65±0.17	2.01±0.15	1.95±0.07	1.75±0.04	0.34	0.91
85	1.74±0.12	2.04±0.08	1.59±0.67	1.58±0.27	0.38	0.93
86	1.53±0.16	1.51±0.47	1.90±0.23	1.64±0.12	0.31	0.89
87	1.79±0.26	2.07±0.06	2.37±0.08	2.20±0.16	0.04	0.36
88	1.83±0.17	1.79±0.34	1.81±0.18	1.87±0.22	0.99	1.00
89	1.75±0.32	1.74±0.37	1.87±0.06	1.83±0.23	0.92	1.00
90	1.93±0.14	2.22±0.11	2.05±0.14	1.83±0.39	0.31	0.89
91	1.12±0.14	1.21±0.20	1.58±0.09	1.32±0.34	0.24	0.89
92	1.09±0.14	2.22±0.09	2.15±0.02	2.11±0.02	0.00	0.05
93	1.63±0.17	1.92±0.12	1.59±0.14	1.63±0.11	0.40	0.93
94	1.70±0.08	2.09±0.17	1.86±0.22	1.72±0.22	0.27	0.89

95	1.11±0.37	0.99±0.29	1.25±0.07	1.40±0.21	0.37	0.93
96	1.19±0.27	1.00±0.45	1.29±0.17	1.36±0.14	0.48	0.94
97	1.92±0.35	2.30±0.10	2.19±0.18	2.16±0.03	0.29	0.89
98	2.47±0.10	2.71±0.05	2.52±0.05	2.51±0.01	0.60	0.97
99	2.63±0.08	2.62±0.13	2.63±0.04	2.65±0.08	1.00	1.00
100	2.28±0.18	2.50±0.17	2.35±0.11	2.43±0.03	0.72	1.00
101	2.44±0.32	2.39±0.05	2.21±0.08	2.31±0.09	0.68	1.00
102	1.90±0.03	1.72±0.75	1.77±0.22	1.86±0.28	0.95	1.00
103	1.47±0.21	1.84±0.20	1.71±0.21	1.34±0.41	0.11	0.66
104	1.73±0.16	1.83±0.03	1.89±0.33	2.10±0.06	0.40	0.93
105	1.80±0.02	1.69±0.26	1.84±0.55	1.92±0.17	0.86	1.00
106	1.69±0.15	1.89±0.24	1.94±0.13	1.88±0.04	0.68	1.00
107	1.76±0.07	1.98±0.08	1.97±0.08	2.08±0.08	0.52	0.95
108	1.47±0.16	1.44±0.33	1.52±0.10	1.60±0.03	0.91	1.00
109	1.82±0.10	1.39±0.09	1.51±0.25	1.90±0.15	0.07	0.53
110	1.93±0.03	2.09±0.11	2.08±0.06	2.18±0.07	0.69	1.00
111	1.26±0.08	1.60±0.06	1.49±0.33	1.44±0.06	0.52	0.95
112	1.79±0.07	1.93±0.33	1.65±0.27	2.33±0.24	0.02	0.22
113	1.91±0.03	0.56±0.95	1.77±0.23	1.70±0.18	0.00	0.07
114	1.96±0.13	0.74±0.99	1.58±0.21	1.29±0.72	0.11	0.66
115	-0.06±0.33	-0.15±0.49	0.31±0.79	-0.58±0.39	0.28	0.89
116	1.22±0.36	0.72±0.65	1.04±0.47	0.66±0.74	0.55	0.95
117	1.70±0.24	1.61±0.22	1.05±0.76	1.66±0.17	0.19	0.82
118	1.65±0.10	1.50±0.58	0.72±0.88	1.33±0.53	0.23	0.87
119	1.78±0.23	1.64±0.42	0.39±1.20	1.81±0.16	0.02	0.24
120	2.08±0.06	1.20±1.30	1.65±0.69	1.76±0.17	0.54	0.95
121	1.36±0.30	-0.03±0.19	0.62±0.34	0.79±0.39	0.00	0.00
122	2.32±0.03	1.95±0.39	1.79±0.33	1.77±0.23	0.06	0.51
123	2.24±0.18	2.84±0.11	2.44±0.30	2.41±0.07	0.02	0.22
124	2.06±0.08	1.26±1.13	1.58±0.28	1.06±0.73	0.31	0.89
125	1.36±0.11	1.18±0.83	1.05±0.10	1.33±0.69	0.89	1.00
126	1.58±0.13	1.66±0.81	1.83±0.29	1.50±0.13	0.82	1.00
127	1.00±0.06	1.00±0.47	1.15±0.07	1.01±0.19	0.92	1.00
128	1.13±0.23	1.47±0.06	1.31±0.06	1.38±0.10	0.54	0.95
129	1.29±0.27	1.75±0.26	1.43±0.18	1.76±0.15	0.09	0.61
130	1.57±0.01	1.40±0.16	1.46±0.05	1.48±0.10	0.91	1.00
131	1.67±0.06	1.68±0.13	1.58±0.15	1.93±0.15	0.42	0.93
132	2.16±0.13	1.86±0.27	2.02±0.17	2.06±0.08	0.57	0.97
133	1.52±0.20	1.27±0.31	1.43±0.07	1.73±0.08	0.24	0.89
134	2.23±0.13	1.82±0.12	1.93±0.23	1.99±0.18	0.28	0.89
135	1.56±0.20	1.51±0.31	1.95±0.23	1.63±0.53	0.39	0.93
136	2.17±0.09	2.35±0.03	2.37±0.02	2.52±0.08	0.38	0.93
137	1.83±0.06	1.54±0.04	1.86±0.11	1.95±0.12	0.30	0.89
138	1.79±0.14	1.31±1.14	1.68±0.35	1.57±0.24	0.42	0.93
139	2.53±0.04	2.58±0.09	2.76±0.10	2.65±0.07	0.67	1.00
140	2.40±0.16	2.35±0.11	2.60±0.05	2.55±0.10	0.53	0.95

141	1.46±0.26	1.52±0.07	1.63±0.09	1.75±0.05	0.57	0.97
142	2.04±0.05	1.87±0.01	2.27±0.01	2.12±0.01	0.26	0.89
143	1.51±0.29	1.27±0.17	1.55±0.19	1.57±0.09	0.53	0.95
144	1.42±0.17	1.07±0.67	1.63±0.07	0.91±0.38	0.11	0.66
145	2.28±0.08	2.37±0.10	2.45±0.13	2.37±0.13	0.86	1.00
146	1.15±0.90	2.10±0.14	2.22±0.15	1.93±0.29	0.03	0.33
147	2.27±0.06	2.34±0.08	2.25±0.13	2.39±0.07	0.89	1.00
148	2.12±0.01	2.53±0.05	2.29±0.01	2.24±0.00	0.21	0.84
149	2.04±0.02	1.45±0.21	1.97±0.21	2.08±0.27	0.02	0.22
150	1.09±0.14	1.01±0.25	1.38±0.10	1.35±0.22	0.33	0.91
151	1.80±0.12	1.81±0.20	2.06±0.05	2.11±0.18	0.36	0.93
152	2.03±0.04	1.90±0.02	2.20±0.01	2.34±0.03	0.15	0.80
153	1.19±0.49	1.48±0.09	1.36±0.39	1.38±0.27	0.77	1.00
154	2.05±0.11	1.80±0.07	2.07±0.09	2.16±0.24	0.37	0.93
155	1.63±0.06	1.39±0.36	1.74±0.15	1.65±0.12	0.44	0.93
156	1.46±0.12	1.40±0.05	1.40±0.30	1.36±0.11	0.98	1.00
157	1.24±0.28	1.43±0.21	1.66±0.27	1.53±0.24	0.33	0.91
158	1.34±0.18	0.99±0.66	1.19±0.30	1.12±0.42	0.79	1.00
159	1.76±0.04	1.51±0.32	1.85±0.31	1.99±0.16	0.17	0.81
160	1.20±0.06	1.14±0.26	1.06±0.49	1.06±0.37	0.94	1.00
161	1.22±0.16	1.13±0.28	1.22±0.22	1.23±0.19	0.97	1.00
162	1.35±0.22	1.16±0.13	1.11±0.16	1.09±0.95	0.19	0.82
163	0.98±0.25	1.14±0.05	1.25±0.35	1.09±0.07	0.73	1.00
164	1.38±0.21	1.31±0.14	1.09±0.12	1.35±0.21	0.61	0.98
165	1.35±0.09	0.84±0.66	1.30±0.42	1.46±0.19	0.25	0.89
166	2.22±0.05	1.63±0.15	2.02±0.19	2.20±0.03	0.02	0.24
167	1.93±0.11	1.55±0.33	1.93±0.21	1.87±0.24	0.25	0.89
168	1.89±0.27	1.19±0.65	1.69±0.18	1.88±0.14	0.07	0.53
169	2.36±0.02	1.55±0.69	1.58±0.24	1.65±0.30	0.04	0.36
170	1.96±0.03	2.03±0.08	2.01±0.26	1.84±0.37	0.81	1.00
171	2.05±0.04	1.97±0.28	2.09±0.09	2.05±0.15	0.96	1.00
172	1.83±0.12	1.95±0.14	1.64±0.16	1.87±0.50	0.58	0.97
173	1.27±0.10	1.25±0.28	1.27±0.14	1.44±0.13	0.85	1.00
174	1.80±0.20	1.36±0.31	1.37±0.10	1.34±0.27	0.14	0.76
175	1.92±0.09	2.05±0.13	1.98±0.18	2.21±0.23	0.55	0.95
176	1.06±0.48	0.97±0.22	1.30±0.19	1.25±0.27	0.52	0.95
177	1.26±0.25	1.24±0.30	1.22±0.39	1.36±0.07	0.94	1.00
178	1.54±0.18	0.92±0.28	1.28±0.02	1.40±0.04	0.06	0.51
179	1.20±0.17	1.36±0.28	1.27±0.35	1.04±0.22	0.60	0.97
180	1.53±0.15	1.15±0.42	1.41±0.05	1.34±0.11	0.42	0.93
181	1.10±0.15	0.72±0.10	0.78±0.51	1.26±0.13	0.09	0.61
182	0.96±0.14	0.65±1.18	1.35±0.08	0.90±0.25	0.55	0.95
183	1.40±0.04	1.43±0.14	1.48±0.10	1.47±0.46	0.98	1.00
184	1.97±0.19	1.88±0.73	1.75±0.19	1.76±0.12	0.89	1.00
185	1.17±0.50	1.29±0.18	1.37±0.43	1.10±0.68	0.90	1.00
186	2.07±0.10	2.17±0.05	2.24±0.17	2.17±0.20	0.88	1.00

187	2.08±0.18	2.21±0.04	2.27±0.24	1.97±0.29	0.46	0.94
188	1.46±0.07	1.28±0.18	1.56±0.11	1.49±0.25	0.69	1.00
189	1.70±0.01	1.12±0.57	1.55±0.22	1.28±0.29	0.15	0.80
190	2.07±0.11	2.20±0.25	2.11±0.08	1.76±0.30	0.18	0.81
191	1.86±0.08	1.60±0.19	1.95±0.13	1.38±0.37	0.05	0.44
192	1.74±0.06	1.74±0.18	1.77±0.36	1.50±0.07	0.59	0.97
193	2.17±0.01	1.99±0.07	2.12±0.15	1.97±0.14	0.71	1.00
194	1.29±0.12	0.72±0.71	1.20±0.04	1.05±0.16	0.29	0.89
195	1.54±0.14	1.48±0.21	1.60±0.04	1.57±0.14	0.96	1.00
196	1.49±0.10	1.44±0.45	1.82±0.11	1.51±0.25	0.33	0.91
197	1.53±0.57	1.55±0.11	1.40±0.30	1.31±0.09	0.79	1.00
198	1.69±0.06	1.74±0.10	1.77±0.01	1.64±0.13	0.94	1.00
199	1.58±0.09	1.17±0.22	1.64±0.24	1.38±0.40	0.18	0.81
200	1.51±0.16	1.72±0.02	1.63±0.18	1.64±0.15	0.82	1.00
201	1.64±0.03	1.27±0.35	1.55±0.18	1.46±0.15	0.43	0.93
202	1.83±0.05	1.69±0.23	1.98±0.02	1.91±0.02	0.58	0.97
203	1.30±0.25	1.93±0.23	1.60±0.47	1.65±0.40	0.18	0.81
204	1.41±0.23	1.25±0.32	1.31±0.08	1.25±0.09	0.90	1.00
205	1.25±0.08	0.90±0.15	1.41±0.09	1.10±0.20	0.19	0.82
206	2.55±0.09	2.33±0.16	2.48±0.20	2.50±0.09	0.71	1.00
207	1.26±0.11	1.53±0.29	1.65±0.15	1.69±0.12	0.21	0.84
208	2.04±0.03	1.94±0.14	1.96±0.08	2.04±0.07	0.95	1.00
209	1.84±0.03	1.70±0.31	1.72±0.18	1.90±0.02	0.78	1.00
210	2.09±0.04	2.28±0.12	2.39±0.03	2.35±0.09	0.46	0.94
211	2.04±0.07	2.12±0.19	2.12±0.06	2.09±0.08	0.97	1.00
212	1.79±0.11	1.80±0.26	1.88±0.16	1.70±0.17	0.88	1.00
213	2.18±0.04	2.03±0.04	2.32±0.12	2.23±0.08	0.55	0.95
214	1.80±0.05	1.71±0.19	1.62±0.25	1.65±0.30	0.86	1.00
215	1.51±0.50	1.31±0.25	1.26±0.30	1.35±0.16	0.80	1.00
216	1.62±0.07	1.50±0.24	1.84±0.05	1.84±0.01	0.32	0.89
217	1.37±0.15	1.31±0.43	1.64±0.17	1.25±0.25	0.37	0.93
218	1.42±0.15	1.03±0.98	1.36±0.15	1.38±0.12	0.77	1.00
219	1.43±0.26	1.77±0.23	1.52±0.02	1.52±0.11	0.46	0.94
220	1.77±0.14	1.75±0.17	1.76±0.13	1.88±0.14	0.94	1.00
221	1.49±0.16	1.83±0.06	1.85±0.11	1.68±0.09	0.34	0.92
222	1.89±0.04	1.74±0.09	1.88±0.07	1.98±0.10	0.77	1.00
223	1.45±0.27	1.27±0.21	1.70±0.14	1.58±0.22	0.26	0.89
224	1.09±0.24	0.92±0.33	0.93±0.13	1.04±0.13	0.89	1.00
225	1.31±0.07	1.17±0.27	1.48±0.17	1.37±0.10	0.60	0.97
226	0.97±0.16	1.05±0.41	0.99±0.21	1.03±0.14	0.98	1.00
227	2.75±0.06	2.75±0.12	2.59±0.19	2.74±0.05	0.79	1.00
228	2.19±0.02	1.69±0.47	1.55±0.36	1.44±0.24	0.02	0.28
229	2.86±0.02	3.10±0.03	2.96±0.18	3.00±0.02	0.67	1.00
230	2.04±0.13	2.34±0.11	2.02±0.34	1.87±0.61	0.43	0.93
231	1.77±0.24	1.54±0.58	0.52±0.48	1.01±0.61	0.01	0.16
232	1.69±0.08	1.67±0.06	1.82±0.05	1.84±0.02	0.82	1.00

233	1.58±0.09	0.46±0.47	1.20±0.47	1.36±0.33	0.00	0.06
234	1.60±0.11	1.67±0.37	1.16±0.33	1.40±0.39	0.21	0.84
235	2.19±0.10	2.05±0.09	1.71±0.18	1.10±0.43	0.00	0.00
236	2.09±0.01	2.10±0.11	2.12±0.10	2.19±0.06	0.96	1.00
237	1.52±0.15	1.69±0.48	1.68±0.15	1.84±0.15	0.54	0.95
238	1.56±0.06	1.37±0.42	1.60±0.14	1.66±0.04	0.59	0.97
239	2.67±0.07	2.23±0.02	2.53±0.04	2.34±0.02	0.11	0.66
240	1.80±0.05	2.04±0.22	1.95±0.26	1.96±0.34	0.77	1.00
241	1.52±0.18	1.33±0.18	1.45±0.21	1.54±0.20	0.80	1.00
242	1.38±0.02	1.39±0.24	1.52±0.09	1.51±0.17	0.89	1.00
243	1.69±0.17	1.54±0.26	1.00±0.19	1.72±0.08	0.00	0.11
244	2.18±0.11	2.07±0.18	2.10±0.22	2.28±0.10	0.73	1.00
245	2.20±0.05	1.83±0.22	1.87±0.34	1.85±0.22	0.29	0.89
246	2.15±0.14	2.17±0.07	2.08±0.21	1.89±0.18	0.52	0.95
247	2.22±0.03	2.09±0.05	2.24±0.18	2.13±0.16	0.84	1.00
248	1.36±0.10	1.53±0.15	1.73±0.19	1.41±0.14	0.37	0.93
249	1.91±0.07	1.56±0.24	2.12±0.06	1.81±0.16	0.09	0.61
250	1.43±0.14	1.46±0.21	1.71±0.16	1.35±0.22	0.42	0.93
251	2.10±0.20	1.83±0.18	1.79±0.10	1.81±0.22	0.48	0.94
252	1.90±0.08	1.00±0.03	2.05±0.03	1.97±0.03	0.00	0.05
253	2.03±0.14	1.93±0.30	2.23±0.05	2.27±0.32	0.31	0.89
254	1.59±0.05	1.73±0.20	1.73±0.03	1.66±0.23	0.91	1.00
255	1.37±0.04	1.16±0.29	1.39±0.22	1.44±0.22	0.65	1.00
256	2.32±0.01	1.85±0.38	2.26±0.39	2.20±0.19	0.18	0.81
257	2.11±0.19	2.01±0.21	2.05±0.05	1.96±0.06	0.91	1.00
258	1.61±0.18	1.81±0.08	1.71±0.08	1.76±0.04	0.82	1.00
259	2.28±0.21	1.97±0.18	2.11±0.19	2.08±0.13	0.51	0.95
260	1.00±0.56	0.57±0.46	0.60±0.11	1.08±0.04	0.19	0.82
261	1.68±0.59	1.58±0.39	2.12±0.38	2.03±0.08	0.28	0.89
262	2.02±0.05	2.01±0.04	2.01±0.11	2.17±0.08	0.85	1.00
263	1.86±0.11	1.48±0.21	1.80±0.19	1.95±0.13	0.17	0.81
264	2.09±0.03	2.04±0.02	2.23±0.01	2.18±0.03	0.79	1.00
265	2.30±0.05	1.94±0.21	2.24±0.26	2.14±0.09	0.31	0.89
266	1.93±0.02	1.36±0.14	1.98±0.00	2.06±0.02	0.01	0.12
267	1.74±0.24	1.70±0.25	1.80±0.10	1.86±0.15	0.89	1.00
268	0.50±0.07	1.48±0.19	1.88±0.06	1.84±0.03	0.00	0.05
269	1.57±0.29	1.12±0.48	1.63±0.09	1.89±0.04	0.01	0.16
270	1.97±0.02	1.95±0.06	2.06±0.06	2.13±0.07	0.84	1.00
271	1.32±0.18	1.40±0.21	1.19±0.71	1.51±0.23	0.80	1.00
272	1.21±0.13	1.22±0.38	1.37±0.18	1.36±0.42	0.87	1.00
273	1.32±0.19	1.42±0.23	1.48±0.11	1.60±0.05	0.67	1.00
274	1.63±0.08	1.27±0.24	1.42±0.37	1.75±0.08	0.16	0.80
275	1.31±0.26	1.24±0.22	1.49±0.28	1.33±0.51	0.83	1.00
276	1.54±0.06	1.42±0.28	1.80±0.06	1.52±0.16	0.35	0.93
277	1.39±0.06	0.91±0.32	1.39±0.15	1.42±0.12	0.09	0.61
278	1.96±0.10	1.79±0.32	1.59±0.08	1.59±0.39	0.27	0.89

279	2.16±0.17	1.64±0.44	0.69±1.14	0.65±0.81	0.03	0.31
280	2.08±0.06	1.21±0.77	1.64±0.44	1.17±1.01	0.30	0.89
281	1.78±0.06	1.66±0.42	1.63±0.59	1.71±0.45	0.97	1.00
282	2.72±0.06	2.71±0.11	2.33±0.20	2.52±0.13	0.13	0.75
283	3.24±0.09	3.20±0.18	3.29±0.08	3.34±0.05	0.90	1.00
284	2.21±0.07	2.43±0.18	2.39±0.20	2.54±0.05	0.41	0.93
285	2.19±0.13	2.35±0.24	2.23±0.03	2.01±0.25	0.40	0.93
286	0.83±0.40	1.33±0.61	1.47±0.07	0.99±0.24	0.15	0.80
287	1.87±0.08	1.96±0.13	1.90±0.04	1.85±0.00	0.97	1.00
288	1.58±0.12	1.65±0.12	1.51±0.10	1.37±0.50	0.63	0.99
289	1.64±0.10	0.92±0.57	1.64±0.18	1.68±0.17	0.01	0.14
290	1.81±0.12	1.94±0.13	1.86±0.04	1.49±0.27	0.20	0.82
291	2.25±0.06	1.97±0.11	2.00±0.08	2.04±0.11	0.53	0.95
292	2.01±0.19	1.81±0.34	1.87±0.18	1.88±0.23	0.85	1.00
293	1.11±0.36	1.33±0.27	1.20±0.16	1.10±0.51	0.85	1.00
294	2.14±0.11	2.05±0.18	1.80±0.35	2.18±0.18	0.26	0.89
295	1.66±0.09	2.00±0.23	1.72±0.20	1.63±0.19	0.34	0.91
296	1.57±0.04	1.55±0.03	1.48±0.16	1.29±0.09	0.62	0.99
297	1.44±0.03	1.93±0.17	1.77±0.36	1.33±0.15	0.03	0.28
298	1.43±0.03	1.64±0.17	1.52±0.05	1.25±0.10	0.40	0.93
299	1.54±0.33	1.78±0.21	1.65±0.04	1.77±0.37	0.67	1.00
300	1.67±0.20	2.07±0.37	1.71±0.24	1.60±0.19	0.15	0.80
301	1.86±0.12	1.67±0.32	1.79±0.17	1.90±0.07	0.74	1.00
302	2.20±0.08	2.05±0.13	2.24±0.08	2.39±0.06	0.41	0.93
303	1.19±0.32	1.09±0.64	1.07±0.72	0.89±0.35	0.92	1.00
304	1.55±0.06	1.86±0.42	1.64±0.21	1.62±0.10	0.54	0.95
305	1.26±1.09	2.15±0.10	2.09±0.18	2.15±0.14	0.67	1.00
306	1.50±0.07	1.66±0.09	1.68±0.19	1.83±0.06	0.53	0.95
307	1.75±0.23	1.49±0.62	1.91±0.32	1.73±0.50	0.72	1.00
308	1.82±0.15	1.07±0.90	1.55±0.20	1.44±0.26	0.29	0.89
309	1.76±0.27	1.79±0.26	1.80±0.08	1.78±0.09	1.00	1.00
310	1.30±0.04	1.12±0.20	1.46±0.03	1.26±0.23	0.57	0.96
311	1.61±0.10	1.36±0.15	1.57±0.18	1.49±0.04	0.68	1.00
312	1.79±0.20	1.98±0.31	2.16±0.13	1.85±0.17	0.35	0.93
313	2.25±0.06	1.68±0.22	2.01±0.24	1.90±0.11	0.06	0.51
314	1.99±0.28	1.60±0.48	1.93±0.19	1.90±0.08	0.38	0.93
315	1.71±0.06	1.89±0.05	1.84±0.06	1.87±0.17	0.86	1.00
316	1.52±0.09	1.50±0.06	1.74±0.16	1.15±1.00	0.61	0.98
317	1.20±0.03	1.23±0.26	1.21±0.13	1.23±0.03	1.00	1.00
318	1.53±0.05	1.61±0.12	1.77±0.13	1.75±0.15	0.68	1.00
319	1.17±0.16	1.14±0.33	1.39±0.07	1.36±0.15	0.63	0.99
320	1.60±0.24	1.28±0.28	1.86±0.09	1.57±0.19	0.08	0.57
321	1.33±0.36	1.37±0.02	1.58±0.07	1.53±0.07	0.64	1.00
322	1.69±0.04	1.44±0.12	1.57±0.09	1.78±0.13	0.45	0.94
323	1.23±0.19	1.47±0.09	1.62±0.05	1.52±0.21	0.36	0.93
324	1.03±0.09	1.35±0.42	1.11±0.11	1.29±0.33	0.51	0.95

325	1.22±0.09	1.28±0.42	0.88±0.77	1.39±0.30	0.91	1.00
326	2.51±0.03	2.07±0.37	2.03±0.28	1.90±0.07	0.02	0.24
327	1.91±0.11	1.44±0.35	1.70±0.37	1.04±0.85	0.17	0.81
328	1.70±0.15	0.75±1.25	1.47±0.07	1.10±0.62	0.37	0.93
329	2.07±0.08	1.80±0.69	1.19±0.74	1.89±0.27	0.19	0.82
330	1.99±0.05	1.33±0.59	1.72±0.06	1.88±0.08	0.04	0.36
331	2.24±0.24	2.00±0.39	1.90±0.24	1.60±0.80	0.43	0.93
332	2.24±0.42	2.28±0.40	2.24±0.26	1.95±0.43	0.70	1.00
333	1.67±0.06	1.63±0.31	1.67±0.07	1.64±0.27	1.00	1.00
334	1.42±0.05	0.77±1.16	1.38±0.12	1.31±0.30	0.52	0.95
335	1.53±0.09	1.82±0.03	1.50±0.17	1.57±0.30	0.46	0.94
336	1.95±0.13	1.78±0.32	1.77±0.22	1.71±0.21	0.72	1.00
337	1.41±0.19	1.53±0.23	1.61±0.12	1.55±0.26	0.85	1.00
338	2.13±0.16	2.17±0.12	2.19±0.18	1.87±0.43	0.40	0.93
339	2.10±0.10	2.27±0.16	2.22±0.20	2.29±0.14	0.78	1.00
340	2.10±0.03	1.93±0.29	2.12±0.11	2.10±0.31	0.77	1.00
341	2.21±0.04	2.00±0.11	2.08±0.13	1.91±0.30	0.51	0.95
342	1.83±0.22	2.18±0.06	2.02±0.10	2.05±0.19	0.43	0.93
343	1.62±0.04	2.09±0.13	1.92±0.10	1.98±0.29	0.17	0.81
344	2.44±0.07	2.07±0.18	2.32±0.14	2.09±0.22	0.18	0.81
345	1.68±0.08	1.83±0.17	1.90±0.08	1.86±0.04	0.78	1.00
346	2.00±0.22	1.89±0.30	2.06±0.07	2.05±0.05	0.85	1.00
347	1.20±0.49	2.03±0.03	1.46±0.16	1.44±0.24	0.00	0.08
348	2.15±0.02	2.00±0.21	2.25±0.09	2.08±0.10	0.66	1.00
349	1.52±0.08	1.41±0.22	1.76±0.10	1.60±0.16	0.46	0.94
350	1.77±0.29	1.23±0.28	1.73±0.11	1.46±0.14	0.06	0.50
351	1.34±0.06	1.26±0.32	1.19±0.13	1.28±0.28	0.94	1.00
352	1.51±0.02	1.23±0.20	1.49±0.01	1.59±0.20	0.42	0.93
353	1.28±0.06	1.34±0.53	1.68±0.26	0.91±0.36	0.06	0.51
354	1.81±0.26	1.18±0.16	1.10±0.95	1.63±0.18	0.04	0.36
355	1.81±0.09	1.47±0.17	1.86±0.08	1.80±0.13	0.28	0.89
356	2.14±0.04	1.80±0.11	2.06±0.14	2.09±0.10	0.41	0.93
357	1.76±0.06	1.40±0.31	1.59±0.24	1.62±0.16	0.47	0.94
358	2.02±0.04	1.60±0.26	1.93±0.11	1.80±0.04	0.25	0.89
359	1.68±0.18	1.36±0.60	1.74±0.24	1.13±0.31	0.15	0.80
360	1.43±0.14	1.27±0.27	1.44±0.16	1.43±0.12	0.86	1.00
361	1.31±0.13	1.09±0.27	1.25±0.53	1.51±0.29	0.48	0.94
362	1.81±0.18	1.76±0.10	1.69±0.03	1.86±0.29	0.90	1.00
363	1.49±0.14	1.20±0.32	1.57±0.03	1.62±0.07	0.26	0.89
364	2.04±0.13	1.92±0.26	1.78±0.19	2.09±0.05	0.48	0.94
365	0.74±0.40	1.20±0.36	1.37±0.11	1.17±0.54	0.23	0.87
366	2.26±0.05	2.02±0.20	2.18±0.18	2.26±0.12	0.60	0.97
367	1.68±0.06	0.44±1.10	1.40±0.16	1.01±0.88	0.06	0.50
368	1.28±0.22	0.60±0.62	1.47±0.15	0.88±1.21	0.41	0.93
369	2.12±0.04	2.18±0.05	2.18±0.08	2.33±0.14	0.75	1.00
370	2.52±0.05	2.48±0.45	2.40±0.26	2.39±0.12	0.92	1.00

371	1.89±0.07	0.90±0.96	1.77±0.34	1.56±0.27	0.00	0.06
372	2.21±0.15	1.97±0.16	2.12±0.07	2.06±0.16	0.67	1.00
373	2.11±0.37	1.93±0.10	2.02±0.18	1.30±0.96	0.23	0.87
374	2.03±0.31	1.99±0.57	2.01±0.30	1.91±0.26	0.98	1.00
375	1.79±0.15	1.97±0.04	1.82±0.22	1.79±0.08	0.83	1.00
376	1.84±0.17	1.74±0.43	2.06±0.15	1.97±0.19	0.48	0.94
377	1.87±0.22	2.21±0.17	1.73±0.34	1.70±0.30	0.09	0.61
378	2.11±0.11	1.93±0.12	2.34±0.17	2.05±0.15	0.26	0.89
379	1.52±0.48	1.78±0.52	1.70±0.40	1.56±0.24	0.86	1.00
380	1.88±0.38	1.96±0.12	2.04±0.07	1.93±0.08	0.90	1.00
381	2.20±0.09	1.70±0.37	2.10±0.16	1.89±0.26	0.09	0.61
382	2.44±0.11	2.25±0.04	2.50±0.05	2.47±0.06	0.58	0.97
383	2.13±0.06	2.06±0.20	2.18±0.12	1.96±0.22	0.73	1.00
384	1.61±0.08	1.33±0.24	1.55±0.03	1.50±0.28	0.62	0.99
385	2.33±0.07	1.48±0.61	2.47±0.29	2.41±0.33	0.00	0.05
386	1.15±0.05	1.41±0.17	0.74±0.64	1.12±0.20	0.59	0.97
387	1.64±0.14	1.51±0.25	1.90±0.21	1.66±0.12	0.37	0.93
388	2.66±0.09	2.71±0.04	2.03±0.25	2.42±0.09	0.00	0.05
389	2.93±0.19	2.30±0.24	2.50±0.20	2.61±0.20	0.01	0.16
390	2.68±0.17	2.77±0.03	2.52±0.18	2.75±0.18	0.55	0.95
391	2.93±0.17	2.85±0.12	2.84±0.17	2.74±0.30	0.81	1.00
392	2.15±0.58	1.79±0.55	1.60±0.66	2.19±0.28	0.47	0.94
393	2.89±0.11	2.83±0.19	2.54±0.31	2.77±0.31	0.31	0.89
394	2.05±0.14	1.39±1.22	0.97±0.26	1.65±0.65	0.29	0.89
395	2.87±0.07	2.40±0.16	2.66±0.30	2.56±0.17	0.10	0.63
396	3.19±0.09	2.99±0.19	3.18±0.06	3.10±0.02	0.70	1.00
397	2.66±0.28	2.29±0.09	2.47±0.17	2.55±0.14	0.27	0.89
398	2.64±0.05	2.25±0.61	2.45±0.03	2.55±0.17	0.47	0.94
399	2.63±0.13	2.64±0.15	2.70±0.11	2.77±0.15	0.88	1.00
400	2.47±0.13	2.73±0.15	2.69±0.17	2.79±0.28	0.36	0.93
401	2.33±0.01	2.42±0.27	2.47±0.04	2.50±0.02	0.85	1.00
402	2.78±0.30	2.17±0.63	2.47±0.10	2.67±0.05	0.16	0.80
403	1.59±1.50	2.48±0.12	2.55±0.13	2.68±0.08	0.78	1.00
404	2.92±0.12	3.07±0.06	2.73±0.15	2.77±0.14	0.28	0.89
405	2.71±0.23	2.81±0.15	2.62±0.14	2.71±0.04	0.82	1.00
406	1.90±0.26	2.03±0.18	2.14±0.09	1.33±1.06	0.29	0.89
407	2.31±0.25	2.10±0.25	2.30±0.21	2.37±0.23	0.56	0.96
408	2.56±0.21	2.58±0.06	2.71±0.05	2.48±0.38	0.71	1.00
409	2.87±0.01	2.68±0.09	2.89±0.08	2.80±0.26	0.70	1.00
410	3.13±0.01	2.74±0.27	3.17±0.01	3.19±0.13	0.07	0.52
411	1.84±0.09	1.68±0.26	2.04±0.13	1.84±0.05	0.46	0.94
412	2.58±0.22	2.34±0.23	2.10±0.11	2.32±0.33	0.13	0.72
413	2.91±0.13	2.79±0.09	2.98±0.19	2.81±0.17	0.74	1.00
414	1.00±0.03	2.60±0.01	2.08±0.10	2.65±0.15	0.00	0.05

Supplementary Table 3. List of Holm oak pollen proteins identified by using MALDI-TOF/TOF MS/MS. The proteins are listed under broad functional categories.

Number [a]	Mr/pI exp.[b]	Mr/pI theor. [c]	Name and function	Species	Accession number [d]	Protein score/matched peptides (% coverage)/ MS-MS ions (score) [e]
			Metabolism (30)			
259	60.63/6.35	53.79/5.69	Adenosylhomocysteinase	*Medicago truncatula*	gi|29691168	115/11(24)/ DQADYISVPVEGPYKPAHYR (61)
264	60.29/6.49	53.65/5.51	Adenosylhomocysteinase	*Nicotiana tabacum*	gi|54288757	350/17(22)/ GETLQEYWWCTER (51) SKFDNLYGCR (40) LSKDQADYISVPVEGPYKPAHYR (89) DQADYISVPVEGPYKPAHYR (91)
266	60.31/6.45	53.77/5.60	Adenosylhomocysteinase	*Petroselinum crispum*	gi|417744	184/15(24)/ GETLQEYWWCTER (56) SKFDNLYGCR (31) DQADYISVPVEGPYKPAHYR (49)
350	52.62/6.92	41.65/6.28	Alcohol dehydrogenase	*Alnus glutinosa*	gi|71793966	357/19(39)/ GQTPLFPR (45) FGVTEFVNPK (43) DHDKPVQEVLAEMTDGGVDR Oxidation (M) (44) GTFFGNYKPR (67) FITHSVPFSEINK (53)
346	50.79/6.90	47.70/8.21	Alcohol dehydrogenase	*Ricinus communis*	gi|255544387	310/15 (26)/ GHDDLCEDFFAYNR (69) KIQVIGSYGGR (76) IQVIGSYGGR (69) LAESGIFNLTDAVSR (29)
343	50.92/7.11	47.70/8.21	Alcohol dehydrogenase	*Ricinus*	gi|255544387	333/13(18)/

				communis		GHDDLCEDFFAYNR (88) GTLYDGETR (27) KIQVIGSYGGR (70) IQVIGSYGGR (79) LAESGIFNLTDAVSR (30)
152	68.63/5.75	63.48/5.31	ATP synthase alpha subunit vacuolar, putative	*Ricinus* *Ricinus communis*	gi|255544516	257/22(30)/ YSNSDTVVYVGCGER (101) LAEMPADSGYPAYLAAR Oxidation (M) (25) LGDLFYR (31)
257	62.28/6.31	55.79/6.02	ATP synthase subunit alpha, mitochondrial	*Helianthus annuus*	gi|114404	248/14(24)/ EAFPGDVFYLHSR (100) ISNFYTNFQVDEIGR (89)
265	62.11/6.52	44.63/6.54	ATP1	*Quercus subsericea*	gi|164685320	455/24(47)/ VVDALGVPIDGR (49) DNGMHALIIYDDLSK Oxidation (M) (17) EAFPGDVFYLHSR (98) GIRPAINVGLSVSR (26) LELAQYR (Ions score 18) EVAAFAQFGSDLDAATQALLNR (84) LTEVLKQPQYAPLPIEK (26)
358	62.12/6.74	30.70/8.37	ATPase alpha subunit	*Hedwigia ciliata*	gi|114216174	109/7(28)/ EAFPGDVFYLHSR (74)
136	48.83/5.77	31.46/5.85	Cytosolic glutamine synthetase	*Daucus carota*	gi|2454631	169/5(11)/ LTGKHETADINNFSWGVANR (79) HETADINNFSWGVANR (78)
139	48.98/5.92	31.46/5.85	Cytosolic glutamine synthetase	*Daucus carota*	gi|2454631	260/5(15)/ LTGKHETADINNFSWGVANR (134) HETADINNFSWGVANR (112)
355	62.06/6.88	49.99/6.06	Chain A,	*Pisum*	gi|9955321	119/12(25)/

			dihydrolipoamide dehydrogenase of glycine decarboxylase	*sativum*		ALLHSSHMYHEAKHSFANHGVK Oxidation (M) (40) VGKFPFMANSR Oxidation (M) (17) FPFMANSR Oxidation (M) (17)
306	55.12/6.57	50.89/9.27	Dihydrolipoamide succinyltransferase component of 2-oxoglutarate dehydrogenase, putative	*Ricinus communis*	gi|255578100	122/12(13)/ SATEPQLPPKER (32) GLVVPVIR (28) EAVFFLR (37)
59	65.31/5.37	61.48/5.54	Dihydroxyacetone kinase, putative	*Ricinus communis*	gi|255586153	149/8(15)/ NYTGDRLNFGLAAEQAK (56) SEGYKVETVIVGDDCALPPPR (43) YGGASAGYR (33)
148	60.62/5.93	47.63/5.46	Enolase	*Brassica rapa* subsp. *campestris*	gi|34597330	340/20(31)/ IVLPVPAFNVINGGSHAGNK (79) LAMQEFMILPVGASSFK 2 Oxidation (M) (40) AGWGVMASHR (23) AGWGVMASHR Oxidation (M)(21) YNQLLR (39) IEEELGSEAVYAGANFR(81)
210	62.34/6.04	47.63/5.46	Enolase	*Brassica rapa* subsp. *campestris*	gi|34597330	340/24(29)/ IVLPVPAFNVINGGSHAGNK (86) LAMQEFMILPVGASSFK 2 Oxidation (M) (37) EAMKMGVEVYHNLK 2 Oxidation (M) (20) YGQDATNVGDEGGFAPNIQENKEGLELLK (65) AGWGVMASHR (39)

						AGWGVMASHR Oxidation (M) (21) YNQLLR (43) IEEELGSEAVYAGANFR (98)
356	61.53/6.93	48.05/5.7	Enolase	*Annona cherimola*	gi|224482647	515/26(53)/ HITIFSPEGR (71) LFQVEYAFK (14) AAAVTSIGVR (46) LLDQTSVTHLFPITK (24) YLGLLATGMTADAR Oxidation (M) (30) TLVQQAR (42) NEAAEFR (19) ATSAGLKEQEAINFLEK (78) VLSTEEIDEHLTAISERD (97)
101	61.42/5.37	49.21/5.25	F1-ATP synthase, beta subunit	*Sorghum bicolor*	gi|4388533	282/17(33)/ VVDLLAPYQR (51) AHGGFSVFAGVGER (29) VGLTGLTVAEHFR (77) FTQANSEVSALLGR (47)
98	49.21/5.25	54.38/5.03	F1-ATP synthase, beta subunit	*Sorghum bicolor*	gi|4388533	361/16(41)/ VVDLLAPYQR (73) VGLTGLTVAEHFR (79) FTQANSEVSALLGR (124)
414	50.54/7.01	42.04/5.75	GDP-D-mannose 4,6-dehydratase 2	*Arabidopsis thaliana*	gi|186510910	105/15(34)/ VGLQTKLFLGNLQASR (2)
344	46.21/7.15	23.62/6.21	Malate dehydrogenase, cytoplasmic, putative	*Oryza sativa Japonica Group*	gi|110289264	283/14(43)/ VLVTGAAGQIGYALVPMIAR Oxidation (M) (30) MELVDAAFPLLK Oxidation (M) (32) VLVVANPANTNALILK (50) NAIIWGNHSSTQYPDVNHATVK (69) TPSGEKPVR (41)

298	42.51/5.83	49.26/7.64	Cytosolic phosphoglycerate kinase 1	*Populus nigra*	gi|3738257	244/11(22)/ VDLNVPLDDNFNITDDTR (101) YSLKPLVPR (33) IVAEIPEGGVLLLENVR (78)
155	68.03/5.92	61.31/5.39	Phosphoglyceromutase	*Mesembryanthemum crystallinum*	gi|3914394	325/13(20)/ GVDAQIASGGGR (69) TSGEYLVHNGVR (93) FGHVTFFWNGNR (54) DAILSGKFDQVR (77)
252	55.55/6.30	43.22/5.34	S-adenosylmethionine synthase	*Camellia sinensis*	gi|75311075	616/20(42)/ VLVNIEQQSPDIAQGVHGHLTK (95) KNGTCPWLRPDGK (55) NGTCPWLRPDGK (61) TQVTVEYYNEK (45) TIFHLNPSGR (67) FVIGGPHGDAGLTGR (97) SIVANGLAR (37) TAAYGHFGR (71)
336	30.29/7.07	27.92/5.90	Triosphosphate isomerase-like protein	*Solanum tuberosum*	gi|76573375	304/10(24)/ VIACVGETLEQR (60) VATPAQAQEVHFELR (99) VATPAQAQEVHFELRK (85) ELAAQPDVDGFLVGGASLKPEFIDIIK (23)
98	60.89/5.46	54.38/5.03	Vacular ATP synthase subunit B, putative	*Arabidopsis thaliana*	gi|15233891	390/26(47)/ DFEENGSMER Oxidation (M)(11) GYPGYMYTDLATIYER Oxidation (M)(40) QIYPPINVLPSLSR(75) KFVAQGAYDTR(81) TLDQFYSR(20)

97	61.07/5.40	54.38/5.03	Vacuolar ATP synthase subunit B, putative	*Arabidopsis thaliana*	gi|15233891	191/17(35)/ YQEIVNIR (16) GYPGYMYTDLATIYER Oxidation (M) (2) QIYPPINVLPSLSR (49) KFVAQGAYDTR (43)
206	59.25/5.98	53.39/5.37	Xylose isomerase	*Ricinus communis*	gi|255583291	379/11(18)/ HQYDWDAATAANFLR (66) LNIECNHATLSGHSCHHELETAR (159) NGGLAPGGFNFDAK (61) LIEDGSLAELVR (61)
160	59.48/5.88	53.39/5.37	Xylose isomerase	*Ricinus communis*	gi|255583291	300/10(18)/ HQYDWDAATAANFLR (74) LNIECNHATLSGHSCHHELETAR (83) NGGLAPGGFNFDAK (72) LIEDGSLAELVR (47)
			Defense/stress (12)			
239	30.32/6.43	24.96/5.36	Ascorbate peroxidase	*Eucalyptus camaldulensis*	gi|111434273	280/12(40)/ EDKPQPPPEGR (52) QVFGVQMGLSDKDIVALSGGHTLGR Oxidation (M) (116) ALLADPVFRPLVEK (60)
227	17.60/6.37	15.52/5.61	Cu/Zn superoxide dismutase	*Helianthus annuus*	gi|50978416	83/4(21)/ QIPLIGGQSIIGR (39) AVVVHADPDDLGKGGHELSK (33)
154	65.37/5.77	51.73/4.97	Cysteine protease Cp4	*Actinidia deliciosa*	gi|146216000	230/7(12)/ FEIFKDNLR (70) FADLTNEEYR (68) AVANQPVSVAIEAGGR (80)
339	48.04/6.90	58.80/8.41	Cytochrome P450 monooxygenase	*Medicago truncatula*	gi|84514165	67/10(21)/ EIKMMPFGAGR 2 Oxidation (M) (31)

			CYP89A28			
369	24.10/7.31	20.07/5.64	Glycine-rich protein 2	*Nicotiana sylvestris*	gi|121631	68/4(20)/ CGESGHFAR(53)
61	17.23/5.54	17.37/7.82	Glycine-rich RNA-binding protein	*Prunus avium*	gi|34851124	71/4(15)/ GFGFVTFSNEK (56)
104	70.68/5.46	71.38/5.14	Heat shock 70 kDA protein	*Triticum aestivum*	gi|2827002	356/28(34)/ TTPSYVAFTDTER (55) NQVAMNPTNTVFDAKR Oxidation (M) (5) NAVVTVPAYFNDSQR (43) MVNHFVQEFK Oxidation (M) (40) STVHDVVLVGGSTR (10) EQVFSTYSDNQPGVLIQVYEGER (73) NALENYAYNMR Oxidation (M) (27)
110	68.89/5.65	72.72/5.95	Heat shock 70 kDA protein	*Phaseolus vulgaris*	gi|399940	168/17(26)/ AVITVPAYFNDAQR (55) IINEPTAAALSYGMNNK Oxidation (M) (23) GVNPDEAVAMGAAIQGGILR Oxidation (M) (29)
107	68.94/5.59	72.72/5.95	Heat shock 70 kDa protein, mitochondrial	*Phaseolus vulgaris*	gi|399940	280/19(24)/ TTPSVVAFNQK (43) AVITVPAYFNDAQR (69) IINEPTAAALSYGMNNK Oxidation (M) (33) GVNPDEAVAMGAAIQGGILR Oxidation (M) (54) SQVFSTAADNQTQVGIK (18)
102	70.38/5.40	71.52/5.17	Similar to HSC70-1	*Vitis vinifera*	gi|225449497	496/35(42)/ TTPSYVAFTDTER (83) NAVVTVPAYFNDSQR (38)

						IINEPTAAAIAYGLDKK (62) MVNHFVQEFK Oxidation (M) (33) ARFEELNMDLFR Oxidation (M) (1) VQQLLQDFFNGK (9) EQVFSTYSDNQPGVLIQVYEGER (79) NALENYAYNMR Oxidation (M) (18)
58	70.53/5.35	71.52/.5.17	Similar to HSC70-1	*Vitis vinifera*	gi|225449497	459/27(34)/ TTPSYVAFTDTER (75) NAVVTVPAYFNDSQR (45) IINEPTAAAIAYGLDKK (67) MVNHFVQEFK Oxidation (M) (51) ARFEELNMDLFR Oxidation (M) (9) EQVFSTYSDNQPGVLIQVYEGER (82) NALENYAYNMR Oxidation (M) (I25)
57	70.64/5.30	71.52/5.17	Similar to HSC70-1	*Vitis vinifera*	gi|225449497	336/27(35)/ TTPSYVAFTDTER (58) NAVVTVPAYFNDSQR (54) MVNHFVQEFK Oxidation (M) (36) ARFEELNMDLFR Oxidation (M) (2) VQQLLQDFFNGK (Ions score 14) EQVFSTYSDNQPGVLIQVYEGER (62) NALENYAYNMR Oxidation (M) (11)
			Cytoskeleton (6)			
142	52.67/5.82	41.90/5.31	Actin	*Caragana korshinskii*	gi|218533930	700/29(60)/ AVFPSIVGRPR (59) VAPEEHPVLLTEAPLNPK (74) TTGIVLDSGDGVSHTVPIYEGYALPHAI LR (77) GYSFTTTAER (70) SYELPDGQVITIGAER (100) AEYDESGPSIVHR (95)

						AEYDESGPSIVHRK (58)
140	52.46/5.72	41.90/5.31	Actin 3	*Populus trichocarpa*	gi|224088196	771/ 28(59)/ AVFPSIVGRPR (55) VAPEEHPVLLTEAPLNPK (99) TTGIVLDSGDGVSHTVPIYEGYALPHAI LR (84) LDLAGRDLTDALMK Oxidation (M) (8) GYSFTTTAER (54) SYELPDGQVITIGAER (136) AEYDESGPSIVHR (98) AEYDESGPSIVHRK (72)
92	60.29/6.49	41.90/5.31	Actin 3	*Populus trichocarpa*	gi|224088196	758/32(22)/ AGFAGDDAPR(66) AVFPSIVGRPR(54) IWHHTFYNELR(68) VAPEEHPVLLTEAPLNPK(51) LDLAGRDLTDALMK Oxidation (M)(5) GYSFTTTAER(47) SYELPDGQVITIGAER(105) IKVVAPPER(31) AEYDESGPSIVHR(82) AEYDESGPSIVHRK(54)
247	49.04/6.44	40.08/6.24	Type IIIa membrane protein cp-wap 13	*Vigna unguiculata*	gi|2218152	380/23(44)/ VPEGFDYELYNR (94) YIYTIDDDCFVAK (65) YIYTIDDDCFVAKDPSGK (35) GYPFSLR (36) GTLFPMCGMNLAFDR 2 Oxidation (M) (5) ELIGPAMYFGLMGDGQPIGR 2 Oxidation (M) (14)

						ECTSVQKCYIELSK (20)
193	49.24/6.24	40.08/6.24	Type IIIa membrane protein cp-wap 13	*Vigna unguiculata*	gi|2218152	349/22(41)/ VPEGFDYELYNR (84) YIYTIDDDCFVAK (70) YIYTIDDDCFVAKDPSGK (36) GYPFSLR (21) GTLFPMCGMNLAFDR 2 Oxidation (M) (18) ELIGPAMYFGLMGDGQPIGR 2 Oxidation (M) (9)
313	59.95/6.56	51.45/5.81	UDP-glucose pyrophosphorylase	*Annona cherimola*	gi|224482653	226/15(33)/ DGWYPPGHGDVFPSLR (81) VQLLEIAQVPDEHVNEFK (16) VLQLETAAGAAIR (70)
			Protein synthesis and processing (3)			
296	50.83/6.62	49.60/6.25	Elongation factor Tu, putative	*Arabidopsis thaliana*	gi|15236220	265/17(35)/ AIAFDEIDKAPEEK (31) FPGDDIPIIR (Ions score 15) ILDNGQAGDNVGLLLR (102) HTAFFSNYRPQFYLR (42)
208	59.13/6.02	53.92/5.54	Mitochondrial processing peptidase alpha subunit.	*Ricinus communis*	gi|255549792	117/4(5)/ EVEAIGGNIGASASR (109)
287	30.70/6.57	27.55/5.92	Proteasome subunit alpha type, putative	*Ricinus communis*	gi|255550415	567/26(53)/ HITIFSPEGR (71) LFQVEYAFK (14) AAAVTSIGVR (46) LLDQTSVTHLFPITK (24) YLGLLATGMTADAR Oxidation (M) (30) TLVQQAR (42)

						NEAAEFR (19) ATSAGLKEQEAINFLEK (78) VLSTEEIDEHLTAISERD (97)
			Allergen proteins(4)			
283	21.24/6.79	17.50/5.25	Pollen allergen Cas s 1	*Castanea sativa*	gi|212291464	222/8(29)/ AFVLDSDNLIPK (91) KITFGEASK (55) ITFGEASK (42)
229	21.55/6.32	17.46/5.39	Pollen allergen Que a 1 isoform	*Quercus alba*	gi|167472849	118/5(25)/ AFVLDSDNLIPK (76) AVEAYLVAHPDLYK (25)
230	23.06/6.32	17.46/5.39	Pollen allergen Que a 1 isoform	*Quercus alba*	gi|167472849	72/4(16)/ AFVLDSDNLIPK (43) AVEAYLVAHPDLYK (20)
284	22.82/6.80	17.46/5.39	Pollen allergen Que a 1 isoform	*Quercus alba*	gi|167472849	86/4(16)/ AFVLDSDNLIPK (56) AVEAYLVAHPDLYK (20)
			Other functions(7)			
372	30.38/7.44	35.17/6.94	Carbonic anhydrase isoform 1	*Gossypium hirsutum*	gi|4754913	179/7(19)/ VCPSHVLDMQPGEAFVVR Oxidation (M) (61) EAVNVSLGNLLSYPFVR (59) GGYYDFVK (42)
270		63.71/6.35	Ketol-acid reducto-isomerase, chloroplast precursor	*Ricinus communis*	gi|255568281	323/12(24)/ QIGVIGWGSQGPAQAQNLR (Ions scor100) EINGAGINSSFAVHQDVDGR (99) GILLGAVHGIVESLFR (82)
366	18.30/7.29	16.30/6.3	Nucleoside diphosphate kinase, putative	*Ricinus communis*	gi|255571035	364/10(23)/ GLVGEIISR(50) GLVGEIISRFEK(46)

						GDFAIDIGR(54) NVIHGSDSVESAR(105) NVIHGSDSVESARK(73)
244	46.04/6.25	39.23/6.97	Nutrient reservoir, putative	*Ricinus communis*	gi|255580564	150/10(14)/ VQVVGVDGR(55) AGNLFIVPR(68)
137	47.87/5.86	40.28/6.58	Os05g0156300	*Oryza sativa* Japonica Group	gi|115462193	160/12(22)/ AALVEFYAPWCGHCK (54) YGVSGFPTLK (65)
239	27.21/5.21	24.96/5.36	Os07g0694700	*Oryza sativa*	gi|115474285	160/8(24)/ QDKPEPPPEGR (24) YAADEDAFFADYAEAHLK (108)
131	44.49/5.77	40.19/5.98	Protein disulfide isomerase-like protein	*Glycine max*	gi|49257115	611/8(9)/ YGVSGYPTLK (33)
			Unknown function (20)			
285	28.31/6.83	26.88/8.76	Unknown	*Populus trichocarpa*	gi|118489858	134/9(34)/ YTSIKPLGDR (64) EKPSIGTVIAVGPGPLDEEGNR (32)
175	30.13/6.22	16.38/6.29	Unknown	*Medicago truncatula*	gi|217071874	89/6(41)/ IQLLTPNIGVVYSGMGPDFR Oxidation (M) (1) LYKEPIPVTQLVR (59)
50	55.90/5.34	47.62/4.98	Unknown	*Glycine max*	gi|255637103	311/20(36)/ DSYLILDTLPSEYDSR (72) GVLLYGPPGTGK (34) LAGPQLVQMFIGDGAK Oxidation (M) (15) FDSEVSGDREVQR (25) KIEFPHPSEEAR (40)

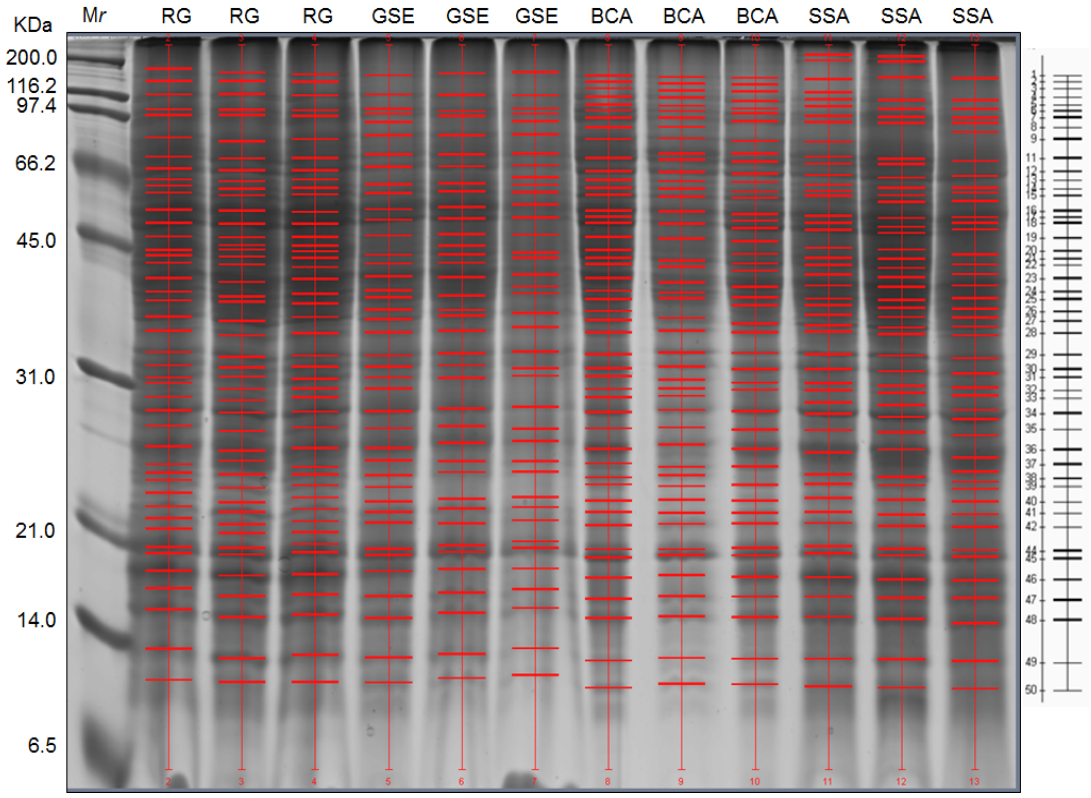
						KMNVHPDVNFEELAR Oxidation (M) (29) DATEVNHEDFNEGIIQVQAK (17)
232	28.42/6.49	22.87/6.88	Unknown	*Glycine max*	gi|255640620	145/6(36)/ GCVFTYDAVGSYER (99) VGYSSQGSGSTLIMPFLDNQLK Oxidation (M) (21)
291	45.12/6.76	33.70/5.58	Unnamed protein product	*Vitis vinifera*	gi|270228151	71/6(21)/ EAGILLSYDPNVR (47)
216	68.01/6.00	61.37/5.4	Predicted protein	*Populus trichocarpa*	gi|224140653	265/14(20)/ AHGTAVGLPSEDDMGNSEVGHNALGAGR Oxidation (M) (89) LPSHYLVSPPEIDR(72) FGHVTFFWNGNR(30) DAILSGKFDQVR(33)
84	48.31/5.48	37.84/5.07	Predicted protein	*Populus trichocarpa*	gi|224100781	268/8(16)/ SLIANLSAANCYK (77) ITVITQGADPVVVAEDGKVK (79)
87	48.15/5.59	37.84/5.07	Predicted protein	*Populus trichocarpa*	gi|224100781	277/8(17)/ SLIANLSAANCYK (79) ITVITQGADPVVVAEDGKVK (86) AGCYAANVIIQR (88)
268	70.10/6.31	70.55/7 6.4	Predicted protein	*Populus trichocarpa*	gi|224094759	484/23(29)/ AVIELENYGLPFSR (79) AFGGQSLNFGK (46) TGHALLHTLYGQAMK Oxidation (M) (92) GVGPLKDHIYLHLNHLPPEVLK (58) LGANSLLDIVVFGR (93) HTLGYWENEK (28)

						HTLGYWENEKVR (13)
298	49.36/6.66	49.26/7.64	Predicted protein	*Populus trichocarpa*	gi|224068811	293/13(29)/ RGEGPTLVECETYR (36)
302	50.58/6.78	49.26/7.64	Predicted protein	*Populus trichocarpa*	gi|224068811	293/14(24)/ TKPHVNVGTIGHVDHGK (72) AIAFDEIDKAPEEK (37) GITIATAHVEYETAKR (36) FPGDEIPIVR (59) HTAFFSNYRPQFYLR (44)
249	45.71/6.54	35.27/5.81	Predicted protein	*Populus trichocarpa*	gi|224127993	170/14(21)/ NGEREFMFYR Oxidation (M) (8) EFMFYR (33) EFMFYR Oxidation (M) (15) EAGILLSYDPNVR (81)
382	48.75/7.60	36.67/7.01	Predicted protein	*Populus trichocarpa*	gi|224055669	385/29 (21)/ IGINGFGR (57) YDTVHGQWK (47) FGIVEGLMTTVHSITATQK Oxidation (M) (61) AASFNIIPSSTGAAK (28) LTGMAFR Oxidation (M) (14) AGIALNDNFVK (35) LVSWYDNEWGYSSR (82)
236	35.42/6.28	29.49/6.18	Predicted protein	*Populus trichocarpa*	gi|224066191	156/9(32)/ LQGNYYFQEQLSR (105)
395	27.77/7.72	24.85/6.17	PREDICTED: hypothetical protein	*Vitis vinifera*	gi|225436699	122/3(12)/ NPFGQIPVLEDGDLTLFESR (113)
145	60.43/5.69	59.25/5.9	PREDICTED: hypothetical protein	*Vitis vinifera*	gi|225456079	781/28(46)/ IINVIGEPIDER (70) VVDLLAPYQR (67) AHGGFSVFAGVGER (75)

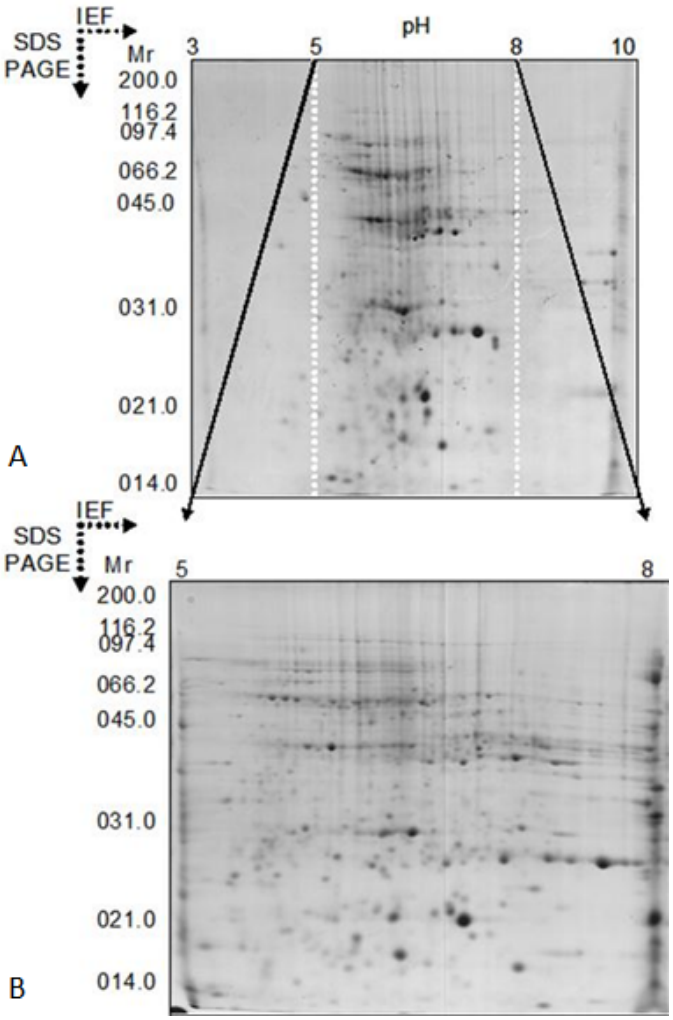
						CALVYGQMNEPPGAR Oxidation (M) (21) VGLTGLTVAEHFR (78) FTQANSEVSALLGR (120) IPSAVGYQPTLATDLGGLQER (96) QISELGIYPAVDPLDSTSR (65) MLSPHILGEEHYNTAR Oxidation (M) (41)
59	65.31/5.37	53.13/4.92	PREDICTED: hypothetical protein	*Vitis vinifera*	gi|225428879	138/7(11) FEIFKDNLR (35) FADLTNEEYR AVANQPVSVAIEAGGR
152	68.63/5.75	68.35/5.25	Hypothetical protein OsI_09127	*Oryza sativa*	gi|125541318	243/20(24)/ TTPSVVAFNQK (39) AVITVPAYFNDAQR (72) EVDEVLLVGGMTR Oxidation (M) (16) GVNPDEAVAMGAAIQGGILR Oxidation (M) (49)
370	27.96/7.40	24.85/6.17	PREDICTED: hypothetical protein	*Vitis vinifera*	gi|225436699	93/4(17)/ NPFGQIPVLEDGDLTLFESR (79)
263	62.28/6.46	53.48/7.6 6	Hypothetical protein LOC100273659	*Zea mays*	gi|226510596	113/9(18)/ KILVYGASFR (49) ILVYGASFR (39)

[a] Numbers correspond to Figure 3 and Supplementary Table 2.
[b] Molecular weight (kDa) and p*I* calculated by using molecular weight standards and the PD-Quest Advance v8.01 software.
[c] Molecular weight (kDa) and p*I* annotated in the NCBI database.
[d] NCBI database accession numbers.
[e] Mascot score (S=-10*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.

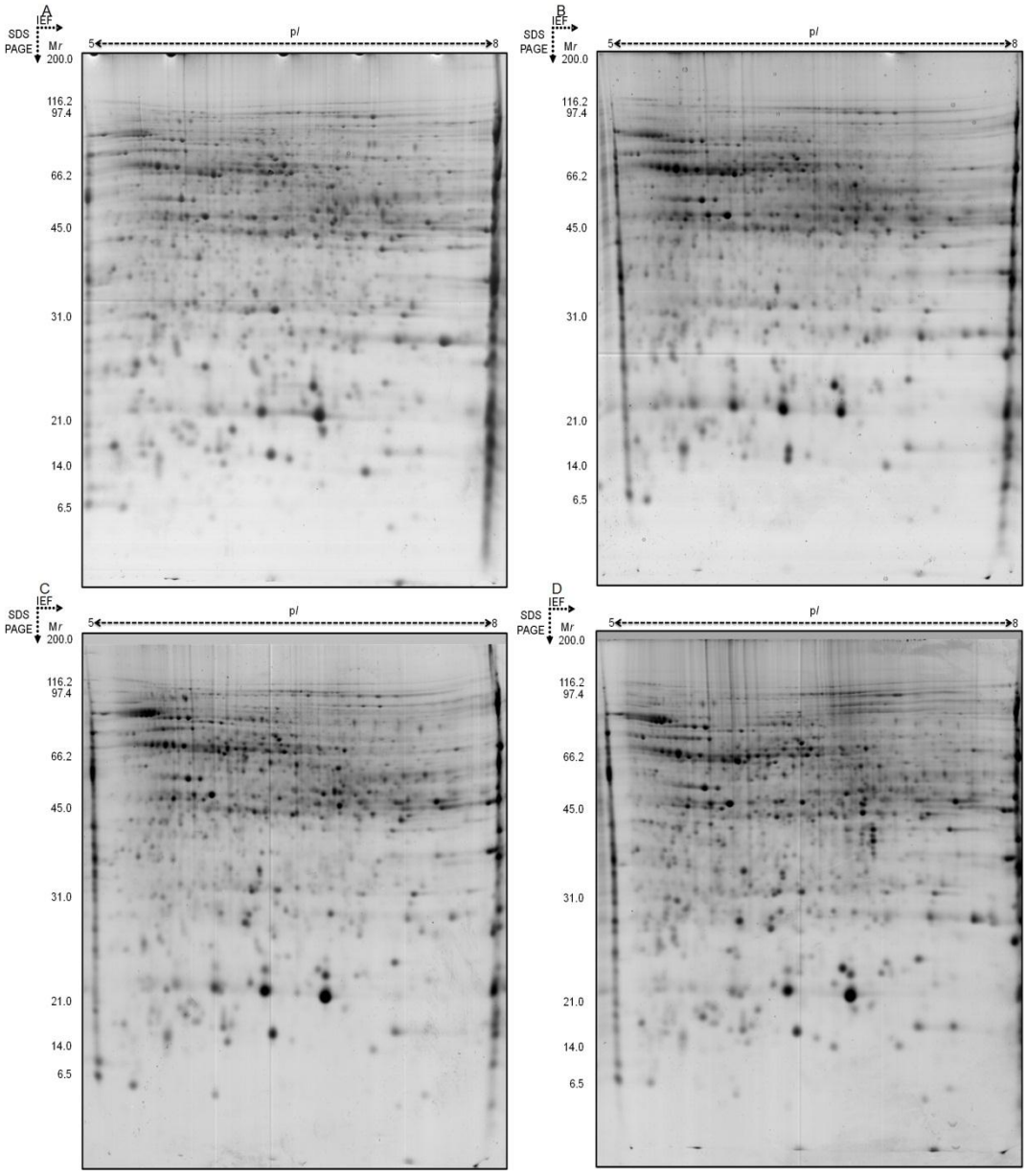
Supplementary Fig. 1. Picture of a representative Holm oak flower (A) and microscopic observation of fresh mature pollen from the four provenances (B) [RG, Granada (a); GSE, Sevilla (b); BCA, Cádiz (c); SSA, Almería (d)].



Supplementary Fig. 2. SDS-PAGE protein profile from *Quercus ilex* subsp. *ballota* pollen (three analytical replicates per provenances). On the left, it is showed the position of molecular weight protein markers; on the top of the gel, the four provenances studied; the lines show the position of protein bands; on the right, the diagram of the resolved bands.



Supplementary Fig. 3. Representative 2-DE gel of *Quercus ilex* subsp *ballota* pollen extracts. Protein extracts were subjected to IEF on 7 cm pH 3-10 (A) or 17 cm pH 5-8 (B) IPG strips (first dimension) and 13% SDS-PAGE (second dimension).



Supplementary Fig. 4. Real gels from pollen extracts of the four Holm oak provenances analyzed by 2-DE. (A) GR, (B) GSE, (C) BCA, and (D) SSA.

6. Discusión general

La encina es una especie forestal que forma parte de la dehesa, un sistema agrosilvopastoral tradicional de la Península Ibérica donde los usos agrícola, ganadero y forestal se integran en este sistema de explotación integral. La producción de bellota de la encina genera unos beneficios económicos de aproximadamente 120 millones de euros, ya que es un componente alimenticio importante en la dieta del cerdo ibérico. Otras características importantes de la encina es su participación en el secuestro del carbono, en el ciclo del agua, en la protección del suelo y como reservorio de biodiversidad, asimismo es una especie forestal usada en los programas de reforestación [79]. No obstante, la conservación de la dehesa es, en este momento, un objetivo prioritario, entendido como una forma de preservar, desarrollar y revalorizar su riqueza económica, biológica, ambiental, social y cultural, por lo que se hace necesario promover su gestión de una manera integral y sostenible; todo ello a través de su estudio y el desarrollo de nuevas estrategias y técnicas. El manejo y mantenimiento sostenible requerirá la identificación de rodales selectos, genotipos productivos y adaptados a condiciones ambientales adversas, para lo que habrá que explotar la variabilidad natural existente en encina (típica de especies alógamas y poco intervenidas por el hombre) [31].

El objetivo general de esta tesis fue analizar la variabilidad en poblaciones de encina localizados en las regiones forestales de Andalucía. Para alcanzar este objetivo hemos utilizado varias aproximaciones metodológicas, entre los que se encuentran el análisis morfométrico, la espectroscopia de infrarrojo cercano (NIRS) y la proteómica. Este estudio se ha realizado utilizando como material vegetal semilla (bellota) y polen de encina, debido a que en investigaciones anteriores se observó una gran variabilidad en el proteoma de las hojas [9]. Para lograr el objetivo general se plantearon tres objetivos específicos: i) el estudio de la variabilidad poblacional de la encina basada en la morfometría y composición química de la bellota; ii) el estudio de la variabilidad poblacional de la encina a través del análisis del perfil de proteínas de bellotas; y iii) el estudio del proteoma del polen de la encina. A continuación se presenta una discusión de los aspectos más importantes de la tesis.

Variabilidad poblacional de la encina basada en la morfometría y composición química de la bellota;

El primer objetivo de la tesis fue estudiar la variabilidad poblacional de la encina basada en la morfometría y composición química de la bellota. Para ello, se llevó a cabo una recolección de bellota de trece poblaciones distribuidas a lo largo de todo el territorio de la región de Andalucía (ver cap. 1, Figura 1 y Tabla 1). Las trece poblaciones analizadas en este estudio mostraron diferencias significativas en la morfometría de la bellota (ver cap. 1, Tabla 3), confirmando las conclusiones de estudios previos [47-49]. Cuando los datos morfométricos de la bellota fueron correlacionados con los datos geográficos y climáticos se observó que las bellotas más grandes (de mayor tamaño y peso) se encontraron en las poblaciones localizadas a baja altitud, en concreto las de las poblaciones de Huelva (CHU y CHT), Sevilla (GSE y APS), Córdoba (PCO) y Jaén (TJA). Por el contrario, las bellotas más pequeñas (menor tamaño y peso) se encontraron en las poblaciones localizadas a mayor altitud, en las provincias de Almería (SSA), Granada (RG), Cádiz (BCA y SCA), Málaga (RMA), Córdoba (CCO) y Jaén (VJA) (ver cap.1, Figura 2). Estos resultados fueron similares a datos publicados anteriormente, donde se comprobó que las características geográficas, como la latitud y altitud, se encuentran relacionadas con las características morfométricas de la bellota del género *Quercus* [47, 48, 80]. Por otra parte, las condiciones pluviométricas de las poblaciones de encina localizadas en el oriente de Andalucía son bastante más bajas si la comparamos con las precipitaciones que reciben las poblaciones que se desarrollan en las regiones occidentales. Aunque los resultados morfométricos analizados no mostraron una correlación clara con los datos climáticos de las poblaciones analizadas en este estudio, sí se observó que aquellas poblaciones con menos pluviometría (poblaciones orientales) presentaron los valores más bajos en peso, diámetro y longitud de bellota comparado con las poblaciones que se desarrollan en climas más húmedos (poblaciones occidentales). Se ha comprobado que efectivamente la pluviometría de una región es un factor determinante en las características morfométricas de la bellota. Así, las poblaciones de *Q. suber* que crecían en zonas más húmedas presentaban bellotas más grandes que las poblaciones que se desarrollaban en zonas con menos humedad [47].

Hasta la fecha, la composición química de la bellota de encina se ha estudiado poco y los datos que se han documentado son bastante variables. En esta dirección, se ha usado la metodología NIRS para determinar un análisis completo de la composición química de la harina de bellota haciendo énfasis en la comparación, catalogación y caracterización de las trece poblaciones de la región de Andalucía. Los resultados obtenidos demostraron que existen diferencias significativas en cuanto a la composición química entre las poblaciones analizadas (ver cap. 1, Tabla 5). Los datos obtenidos de los compuestos químicos analizados fueron similares a los trabajos previos publicados en bellotas de encina [30, 61, 62]. Sin embargo, los contenidos de grasa total y de los ácidos grasos oleico y palmítico fueron relativamente más elevados a los contenidos observados por otros autores [61, 62, 81, 82], mientras que los de ácido linoleico fueron más bajos que los encontrados por otros autores [61, 81-83]. El contenido calórico varió entre 19934,5 (±417,3) y 20398,7 (±449,3) kJ/kg en las muestras analizadas, sin embargo, no se han encontrado datos anteriores de dicho parámetro en la literatura para compararlos. Las correlaciones entre los datos geográficos y climáticos de los lugares de procedencia con respecto a los datos de la composición química de la bellota se encuentran muy poco estudiadas. En este trabajo, las bellotas de las poblaciones localizadas en el norte de Andalucía (Huelva, CHU-CHT y Sevilla, GSE-APS) presentaron valores relativamente más altos en contenido de cenizas y en ácidos grasos oleico y linoleico, mientras que las poblaciones localizadas en el sur de la región (Almería, SSA, Granada RG, Cádiz BCA-SCA, y Málaga RMA) mostraron valores relativamente más bajos de azúcares y de ácido palmítico. Aunque no se encontró una correlación clara con los datos de pluviometría, sí se observó que las poblaciones que se desarrollaron en climas más secos (Jaén VJA, Sevilla SAA, Córdoba CCO y Granada GR) mostraron una tendencia a presentar contenidos más altos en azúcares y ácido palmítico, mientras que las poblaciones que se desarrollaron en zonas mésicas (Málaga RMA, Cádiz BCA, y Jaén TJA) tendían a presentar valores de contenido calórico y ácido linoleico más elevados. Las poblaciones que se desarrollaron en climas intermedios (Huelva CTH-CHU, Sevilla APS-GSE y Cádiz SCA) presentaron valores más elevados en los contenidos de cenizas, grasa total y ácido oleico. En la literatura se encuentran varios estudios que sustentan el uso de la composición química de la bellota para caracterizar varios taxas del género *Quercus* [54-56, 81, 83-85]. Los resultados obtenidos en dichos estudios no

permiten concluir claramente si la variabilidad en la composición química es debida a la composición genética o al medio ambiente donde se desarrollan las poblaciones; sin embargo, está claro que el medio ambiente condiciona en gran medida la diversidad en la composición química del género *Quercus* [56]. Por tanto, las variaciones obtenidas en la morfología y composición química de la bellota podrían estar determinadas por varios factores asociados a i) las condiciones medioambientales, tales como las precipitaciones y la temperatura de la zona donde se localiza la población, ii) la fisiología de la bellota, como los procesos de maduración y germinación, y iii) a los estreses abióticos y bióticos, tales como la desecación y el ataque de plagas como *Curculio* spp. y *Cydia* spp. [30, 57, 58].

Variabilidad poblacional de la encina a través del análisis del perfil de proteínas de bellotas

En los últimos años, la proteómica se ha convirtiendo en una herramienta muy útil para realizar estudios de variabilidad poblacional en especies vegetales [67-73]. En esta dirección, nuestro grupo de investigación ha iniciado recientemente un proyecto para estudiar la variabilidad de poblaciones de encina en Andalucía. Los primeros trabajos han ido dirigidos a caracterizar el proteoma de hojas de la especie [9, 10]. Los resultados obtenidos indican la existencia de una gran variabilidad en el perfil proteico obtenido por 2-DE, incluso entre muestras de hojas obtenidas del mismo árbol, lo que parece lógico teniendo en cuenta, por un lado, que el proteoma es muy variable y, por otro, la enorme plasticidad y dinamismo de las plantas. Dichas investigaciones demostraron la gran variabilidad que presentan las hojas de la encina, lo que recomendó que se abordase el estudio de variabilidad en la encina mediante el análisis proteómico de un material con un proteoma más estable, como es el caso de la bellota. En este contexto, el segundo objetivo de esta tesis fue estudiar la variabilidad poblacional de la encina a través del análisis de los perfiles de proteínas de diez de las trece poblaciones analizadas a través de los datos morfométricos y composición química. Las poblaciones analizadas en este estudio fueron: en el sur (Almería SSA, Granada RG, Cádiz (SCA-BCA), en el noreste (Córdoba PCO, Jaén VJA-PJA), y en el noroeste (Sevilla GSE-APS y Huelva CHU).

En primer lugar, se realizó el análisis del perfil proteico mediante 1-DE. Las poblaciones mostraron diferencias significativas en el contenido de proteínas y número e

intensidad de las bandas proteicas (ver cap. 2 Tabla 1). El contenido de proteína varió de 2,9 a 5,9 mg g^{-1} de peso seco en las poblaciones SSA y CHU respectivamente. Estos valores estuvieron en el rango determinado previamente para diferentes especies del género *Quercus* spp. [86]. Los valores del contenido de proteínas fueron correlacionados positivamente con los determinados por NIRS (r=0.877, P<0.05). El contenido en proteína total determinado mediante 1-DE correspondió con el 5-10% del contenido total de proteína determinado por NIRS. La variación del contenido de proteínas entre las distintas poblaciones puede estar relacionado con las condiciones ambientales, como se ha observado en otras especies del género *Quercus* spp. [54]. Los resultados encontrados en este estudio reafirman que el contenido de proteína estaba correlacionado con las características morfológicas, geográficas (altitud y latitud) y climáticas (temperatura máxima mensual media de las zonas en donde se encuentran las poblaciones de encina analizadas) (ver cap. 2 Figure 1). Así, las poblaciones que presentaron bellotas más grandes (de mayor tamaño y peso) presentaron mayores concentraciones de proteína que aquellas con bellotas pequeñas (de menor tamaño y peso). Además, las poblaciones que se localizaban en el norte de la región mostraron una tendencia de presentar valores más altos en los contenidos de proteína que en las poblaciones localizadas en el sur. Asimismo, se observó que el contenido de proteína disminuía con respecto a la altitud y se incrementaba con respecto a la temperatura máxima mensual media de las áreas donde las poblaciones se desarrollaban.

Al analizar los datos normalizados de los perfiles proteicos de 1-DE (ver cap. 2, Figura 2) a través de un análisis multivariante (ver cap. 2, Figuras 3 y 4), se observó que las poblaciones de Sevilla (GSE y APS) se encuentran más estrechamente relacionadas con las poblaciones de Jaén (PJA) y Huelva (CHU). Por el contrario, las poblaciones de Cádiz (SCA y BCA), Córdoba (PCO), Jaén (VJA), Almería (SSA) y Granada (RG) mostraron una tendencia a separarse del grupo anterior, siendo más afines filogenéticamente entre ellas. Dentro de este grupo, las poblaciones de Almería (SSA) y Granada (RG) mostraron una mayor correlación que las poblaciones de Cádiz (SCA y BCA), Córdoba (PCO) y Jaén (VJA). Estos resultados fueron correlacionados con los datos climáticos y geográficos del sitio de origen de las poblaciones. Las distancias filogenéticas mostraron una correlación significativa con respecto a las precipitaciones (valor de P=0,046) y también con respecto a

la combinación de la altitud y las precipitaciones (P=0,052). Estas tendencias de agrupamiento de las poblaciones fueron muy similares a los observados mediante los datos morfométricos y de composición química de la bellota (ver cap. 1, Figura 4). Esta variabilidad dependiente de las condiciones geográficas y climáticas, junto con la introgresión en el género *Quercus*, ha sido ampliamente estudiada [24, 87-89]. Así, un estudio de los perfiles de proteínas de poblaciones de encina de las regiones de Sevilla y Córdoba concluyó que las poblaciones de Córdoba presentaron una mayor variabilidad que las poblaciones de Sevilla [78]. De igual forma, un análisis de las poblaciones de encina de diferentes regiones geográficas mediterráneas (Grecia, Italia, Norte de África y la Península Ibérica) sugirió que la diferenciación genética y la migración de la encina ocurrió de este a oeste [43] y que después de la glaciación la encina empezó a migrar a áreas localizadas más hacia el norte. Además, los árboles de encina de la Península Ibérica estuvieron más relacionados con los del Norte de África. Asimismo, en este estudio, los árboles de encina de Andalucía se han considerado como intermediarios entre los árboles distribuidos en Marruecos y en las otras regiones de la Península Ibérica [43].

En el análisis de los mapa proteicos de la 2-DE, se seleccionaron cuatro de las poblaciones de acuerdo a las distancias filogenéticas resultado del análisis 1-DE y las distancias geográficas. Las poblaciones analizadas fueron Sevilla (GSE), Córdoba (PCO), Almería (SSA) y Cádiz (BCA). Al analizar los datos normalizados de los perfiles proteicos de 2-DE (ver cap. 2, Tabla Suplementaria 3 y Figura 5), se pudieron determinar 56 especies proteicas diferenciales entre las distintas poblaciones (ver cap. 2, Tabla suplementaria 3). Tras realizar el análisis de agrupamiento (ver cap. 2 Figura 6), se observó que la población de Córdoba (PCO) fue la más distante de las cuatro poblaciones, mientras que la población de Almería (SSA) mostró una tendencia a unirse a las poblaciones de Cádiz (BCA) y Sevilla (GSE). Las poblaciones de Cádiz y Sevilla fueron las más relacionadas entre sí. Sin embargo, los datos normalizados del contenido relativo de proteína obtenido de los perfiles proteicos por 2-DE mantuvieron una correlación parcial con los agrupamientos de los datos de los perfiles proteicos obtenidos por 1-DE. Sólo los perfiles de proteína de las poblaciones de Almería (SSA) y Cádiz (BCA), mostraron una tendencia de agrupamiento similar a los datos filogenéticos resultado del análisis de los perfiles proteicos de 1-DE. Estas discrepancias pudieran deberse al reducido número de poblaciones analizadas a través

de 2-DE, y al mayor número de datos utilizados para hacer los árboles filogenéticos en 2-DE.

Una de las principales limitaciones en los estudios basados en aproximaciones proteómicas de especies huérfanas es la falta de bases de datos para hacer la búsqueda de las secuencias peptídicas. En la encina, esta es una de las principales desventajas. De ahí que la identificación de las proteínas en encina sea baja. Sin embargo, aun con estas limitaciones, en este trabajo de investigación se han identificado el 61% de las 23 bandas proteicas de bellota detectadas por 1-DE (ver cap. 2. Tabla 2) y el 30% de las manchas proteicas separadas por 2-DE (ver cap. 2 Tabla 3 y Figura 5). Las especies proteicas identificadas en la bellota fueron principalmente proteínas relacionadas con el almacenamiento y la defensa/respuesta a estrés.. Un total de 15 especies proteicas fueron identificadas como precursor de la legumina, las cuales variaron en M*r* de 12,8 a 29,3 kDA y de p*I* de 6 a 8. Dichas proteínas correspondieron al mismo número de acceso de proteína de *Q. robur* en la base de datos. Cuatro proteínas fueron identificadas como leguminas (bandas 16 y 20, y manchas proteicas 31 y 33), así como también dos proteínas fueron identificadas como globulinas 1S (ver cap. 3 Tablas 2 y 3). En numerosos trabajos de investigación se ha demostrado la enorme heterogeneidad que presentan las proteínas de almacenamiento. Esta heterogeneidad se ha atribuido i) al resultado de familias multigénicas que codifican para secuencias primarias, ii) al procesamiento proteolítico diferencial de los genes expresados y iii) a las diferencias en la glicosilación de las proteínas [90-95]. Muchas de las proteínas identificadas en el proteoma de la bellota mostraron M*r* y p*I* similares a las proteínas encontradas en *Ricinus communis* L., *Lens culinaris* L. y *Vicia faba* L. var. *minor*. Otra parte de las proteínas identificadas estaban involucradas en respuesta a respuesta o defensa a estrés (ver cap. 2 Tablas 2 y 3). Tres bandas (nº 3, 12, y 14) fueron identificadas como proteínas de choque térmico. Se ha observado que estas proteínas se incrementan en respuesta al estrés hídrico [96]. En este estudio también se ha identificado una 2-Cys peroxirredoxina BAS1 (banda 7). Las peroxirredoxina 2-Cys de plantas son post-translacionalmente dirigidos a los cloroplastos [97], donde se encargan de la protección de la membrana fotosintética contra el daño foto-oxidativo [98]. La expresión de estos genes depende de señales fotosintéticas de los cloroplastos hacia el núcleo, tales como la disponibilidad del aceptor en el fotosistema I y

ABA [99]. Por otra parte, la peroxirredoxina 2B (banda 13) han sido relacionada con la respuesta a estrés hídrico en estudios previos con hojas de encina [8, 9]. Las bandas 4 y 15 fueron identificadas como quitinasas, la cuales se han relacionado con la respuesta al ataques de patógenos [100, 101]. Otra proteína identificada fue la gliceraldehido-3-fosfato deshidrogenasa (bandas 1, 2 y 3) que asume un papel muy importante en el proceso de hidratación de las semillas [102]. Estas proteínas diferenciales, se han propuesto como marcadores moleculares de variabilidad.

Proteoma del polen de la encina.

Debido a la gran variabilidad entre los individuos de una misma población en los perfiles de proteínas encontrados previamente en hoja [10] y en esta tesis en bellota, se decidió analizar el perfil proteico de polen con el objetivo de caracterizar y catalogar las poblaciones de encina de Andalucía. Para llevar a cabo este estudio se analizaron muestras de polen de las poblaciones de Sevilla (GSE), Granada (RG), Cádiz (BCA) y Almería (SSA). Las poblaciones presentaron diferencias significativas en contenido de proteína ($p \leq 0,05$), variando de 8,1±0,3 mg/g de peso seco en la población de Sevilla (GSE) a 14,6±0,4 mg/g de peso seco en la población de Cádiz (BCA) (ver cap. 3, Figura 1). Estos valores estuvieron en el rango publicado para otras diferentes especies de plantas (incluida *Quercus* spp.) y correspondió al 10% del contenido total de proteína determinado por el método de Kjeldahl [103, 104].

El análisis proteómico de polen de encina fue realizado utilizando técnicas basadas en gel (1-DE y 2-DE acoplado a MALDI-TOF/TOF MS/MS) y técnicas de análisis masivo basadas en la separación de las proteínas por cromatografía líquida (nLC-LTQ Orbitrap MS), para tener mayor cobertura en la identificación del proteoma de polen. Los análisis filogenético y multivariante de los datos normalizados de los perfiles proteicos obtenidos por 1-DE mostraron que las poblaciones de Almería (SSA) y Cádiz (BCA) presentaron perfiles de expresión de proteínas similares. Sin embargo, la población de Granada (RG) mostró una tendencia a unirse al grupo de las poblaciones de SSA y BCA, mientras que la población de Sevilla (GSE) se separó del resto de las poblaciones analizadas, mostrando los perfiles proteicos más heterogéneos. Estas tendencias de agrupamiento de las poblaciones fueron muy similares a los observados a través de los datos normalizados de los perfiles de

proteína obtenidos por 2-DE. Cuando comparamos estos agrupamientos con los datos morfométricos y la composición química de la bellota (ver cap. 1, Figura 4) y con los datos de los perfiles proteicos de bellota (ver cap. 2 Figuras 3 y 4), encontramos las mismas tendencias de agrupamientos. Los resultados obtenidos en este estudio demostraron que el polen puede ser usado para establecer filogenia en encina. No obstante, existen resultados contradictorios en otras especies en cuanto al análisis de los perfiles proteicos. En *Arabidopsis thaliana* no se observaron diferencias significativas entre los perfiles proteicos de polen de cuatro ecotipos diferentes [105]. Sin embargo, en un estudio de polen de *Betula verrucosa* recogido en diferentes localizaciones mostraron diferencias significativas entre sus perfiles proteicos [106].

Mediante el análisis 2-DE, se encontraron un total de 16 manchas proteicas variables entre las cuatro poblaciones. Estas manchas variables junto con 84 manchas proteicas constantes, que cubrían todos los rangos de M*r* y p*I* (para identificar el mayor número posible de especies proteicas), se sometieron a digestión con tripsina para posteriormente ser analizadas mediante MALDI-TOF/TOF MS/MS. En total fueron identificadas 77 manchas proteicas (ver cap. 3 Tabla Suplementaria 3) mediante MALDI-TOF/TOF MS/MS, mientras que por medio del análisis masivo nLC-LTQ Orbitrap MS, se identificaron 273 especies proteicas (ver cap. 3, Tabla Suplementaria 4). Los valores de M*r* y p*I* fueron similares a los obtenidos experimentalmente (ver cap. 3 Tabla Suplementaria 3). No obstante, varias proteínas presentaron algunas desviaciones en M*r* y p*I* (por ejemplo, las proteínas 259, 266 y 350). Asimismo, algunas manchas proteicas presentaron dos proteínas distintas en la misma mancha proteica (por ejemplo las proteínas 98, 298, 239, 29 y 152). Por otra parte, algunas proteínas fueron identificadas varias veces en distintas manchas proteicas. Este tipo de características también se ha observado en varios estudios de proteomas de polen en otras especies de plantas [105, 107-109]. Estas características se han atribuido i) a errores en la secuencia (secuencias parciales de proteínas) debido a las modificaciones post-traduccionales, ii) a la posible degradación de las proteínas y síntesis parcial estas durante la maduración del polen, iii) a la traslación de proteínas después del procesamiento de ARN mensajero y a la co-emigración de proteínas. Más de la mitad de las proteínas identificadas por MALDI-TOF/TOF estuvieron relacionadas con las siguientes categorías: metabolismo (37%), respuesta a defensa/estrés (15%) y citoesqueleto (8%) (ver

cap. 3 Figura 4), mientras que las proteínas identificadas por nLC-LTQ Orbitrap MS fueron agrupadas de la siguiente manera: metabolismo (41%), síntesis de proteínas, ensamblaje y degradación (10%), y respuesta a defensa/estrés (9%) (ver cap. 3 Figure 4B). Asimismo, se han identificado otras categorías como división de la pared celular, transporte y traducción (ver cap. 3 Figura 4B). Resultados similares se han observado en trabajos preliminares sobre el análisis proteómico basado en geles 2-DE en *A. thaliana* [105, 108, 110], tomate (*Lycopersicon esculentum*) [109], y arroz (*O. sativa* L. spp. japónica) [107], donde las proteínas identificadas fueron relacionadas con dichas categorías funcionales [105, 108, 110]. La germinación y el crecimiento del tubo polínico requiere de una tasa alta de carbohidratos para satisfacer la demanda de la biocinética y energía requerida en el proceso [111]. En el presente estudio, un alto porcentaje de las proteínas identificadas estuvieron involucradas en dichos mecanismos (ver cap.3 Tabla suplementaria 3 y 4). Otros estudios de proteínas de polen también han identificado una gran cantidad de proteínas relacionadas a la metabolismo energético y encimas metabólicas [105, 107, 108]. Es bien conocido que el polen está sujeto a varios estreses abióticos y bióticos, entre los que destacan la sequía y las altas temperaturas [109]. En este estudio, 12 proteínas identificadas por MALDI-TOF/TOF y 26 por nLC-LTQ Orbitrap MS se han relacionado con funciones de respuesta a defensa/estrés. Estas proteínas también se ha encontrado en polen de *Camelia sinensis* [112], *Parietaria judaica* [113], *A. thaliana* [108, 110], *L. esculentum* [109], y *Oryza sativa* [107]. Otro de los grupos importantes de proteínas identificadas han sido las relacionadas al citoesqueleto, como son la actina y la profilina. Estas proteínas también se han identificado en otras especies [105, 107, 108]. Cabe destacar que en este estudio, cuatro proteínas fueron identificadas como alergénicas: tres isoformas de Que a 1 (proteínas 229, 230 y 284), y una Cas s 1 (proteína 283) (ver cap. 3, Tabla Suplementaria 3). Así, es la primera vez que se muestran datos de proteínas alérgenas en la encina. Estos resultados se refuerzan con los encontrados para *Q. alba* [114]. Junto a las cuatro proteínas descritas anteriormente, una nueva proteína alergénica se ha encontrado en la encina como es la proteína alergénica semejante a la isoflavona reductasa Bet v 6.0102 [115].

7. Conclusiones

Las conclusiones que se extraen de la presente tesis doctoral son las siguientes:

1. Se ha encontrado una alta variabilidad en cuanto a los parámetros morfométricos, composición química y perfil proteico entre poblaciones de encina en Andalucía.

2. Las poblaciones distribuidas en el norte de Andalucía son diferentes morfológicamente a las poblaciones del sur. Esta variabilidad estuvo influenciada por la latitud y altitud de las zonas donde se desarrollan las poblaciones.

3. Los datos de morfometría y composición química de la bellota establecieron dos grupos de poblaciones relacionadas filogenéticamente entre sí. El primer grupo estuvo formado por las poblaciones de Huelva (CHT y CHU), Sevilla (GSE y APS), Jaén (TJA), Cádiz (SCA) y Córdoba (PCO), mientras que el segundo grupo estuvo formado por las poblaciones de Almería (SSA), Jaén (VJA), Córdoba (CCO), Cádiz (BCA), Granada (RG) y Málaga (RMA).

4. El análisis de los perfiles proteicos de bellota a través de 1-DE permitió correlacionar las poblaciones con sus características morfométricas, localización (norte y sur) y con las condiciones climáticas (xéricas, mésicas e intermedias) del lugar de origen de las poblaciones.

5. El análisis de los perfiles proteicos de polen a través de 1-DE y 2-DE permitió correlacionar a las poblaciones con la localización del lugar de origen de las cuatro poblaciones analizadas. Por un lado las poblaciones de Almería (SSA) y Cádiz (BCA) se agruparon con la población de Granada (RG), mientras que la población de Sevilla (GSE) se separó de este grupo.

6. Se ha caracterizado parcialmente el proteoma de bellota y de polen de encina. Se han identificado 25 especies proteicas, correspondientes a 10 productos génicos para la bellota, mientras que para el polen se han identificado 350 especies proteicas, correspondiendo a 337 productos génicos para el polen.

7. Los análisis de los datos de morfometría, composición química y perfiles proteicos de bellota y polen demostraron que las poblaciones de encina de la regiones forestales de Huelva (CHT y CHU), Sevilla (GSE y APS), y Jaén (TJA) siempre mostraron tendencias similares en sus agrupamientos, este mismo fenómeno se observo para las poblaciones de Almería (SSA), Jaén (VJA), Córdoba (CCO), Cádiz (BCA), Granada (RG) y Málaga (RMA).

8. Bibliografía

[1] Ministerio de Medio Ambiente. 2002. Ministerio de Medio Ambiente, http://www.magrama.gob.es/es/.

[2] Consejería de Agricultura y Pesca. 2008. Consejería de Agricultura y Pesca, http://www.juntadeandalucia.es/agriculturaypesca/portal/.

[3] Consejería de Medio Ambiente. 2008. Consejería de Medio Ambiente. http://www.juntadeandalucia.es/medioambiente/site/portalweb/.

[4] Fernández-Cancio A, Navarro Cerrillo RM, Fernández Fernández R, Gil Hernández P, Manrique Menéndez E, Calzado Martínez C. Climate classification of *Abies pinsapo* Boiss Forests in Southern Spain. Investigación Agraria: Sistemas y Recursos Forestales 2007;16:222-229.

[5] Navarro Cerrillo RM. Valoración de los daños. Un ejemplo concreto: evaluación de daños de seca en Andalucía. En: Tuset, J y Sánchez, G. (eds.). La seca: el decaimiento de encinas, alcornoques y otros Quercus de España,Ministerio de Medio Ambiente, Madrid, España. 2004:341-359.

[6] Sánchez ME, Caetano P, Ferraz J, Trapero A. Phytophthora disease of Quercus ilex in south-western Spain. Forest Pathology 2002;32:5-18.

[7] Ruiz-Labourdette D, Nogués-Bravo D, Ollero HS, Schmitz MF, Pineda FD. Forest composition in Mediterranean mountains is projected to shift along the entire elevational gradient under climate change. Journal of Biogeography 2012;39:162-176.

[8] Echevarría-Zomeño S, Ariza D, Jorge I, Lenz C, Del Campo A, Jorrin JV, Navarro RM. Changes in the protein profile of *Quercus ilex* leaves in response to drought stress and recovery. Journal of Plant Physiology 2009;166:233-245.

[9] Jorge I, Navarro RM, Lenz C, Ariza D, Jorrin J. Variation in the holm oak leaf proteome at different plant developmental stages, between provenances and in response to drought stress. Proteomics 2006;6 Suppl 1:S207-214.

[10] Jorge I, Navarro RM, Lenz C, Ariza D, Porras C, Jorrin 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;5:222-234.

[11] Echevarría-Zomeño S, Abril N, Ruiz-Laguna J, Jorrín-Novo J, Maldonado-Alconada A. 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 2012;34:793-805.

[12] Blondel J, Aronson J. Biology and Wildlife of the Mediterranean Region. Oxford University Press. New York 1999:328 pp.

[13] Valladares F. El hábitat mediterráneo continental: un sistema humanizado, cambiante y vulnerable. En, Paracuelllos, M. (ed.). Ambientes mediterráneos. Funcionamiento, biodiversidad y conservación de los ecositemas. Madrid, España. 2007:216-239.

[14] Charco J. El Bosque Mediterráneo en el Norte de África. Biodiversidad y Lucha contra la Desertificación. Agencia Española de Cooperación Internacional. Madrid. 1999.

[15] Costa M, Morla C, Sainz H. Los Bosques de la Península Ibérica. Una Interpretación Geobotánica. Geoplaneta. Barcelona, España. 1998:497-505.

[16] Guidoto G, Regato P, Jiménez-Caballero S. The major forest types in the Mediterranean. WWF Mediterranean Programme Office. Rome. 1986.

[17] Valladares F, Peñuelas J, Calabuig EL. Ecosistemas terrestres. En, Moreno, J. M. (ed.) Evaluación de los impactos del cambio climático en España, Ministerio de Medio Ambiente. Madrid, España 2004b:68-112.

[18] Govarts R, Frodin DG. World checklist and bibliography of Fagales (Betulaceae, Corylaceae, Fagaceae and Tricodendraceae). Royal Botanical Gardens, Kew. Richmond, Surrey, U.K. 1998:414 pp.

[19] Lawrence GH. Taxonomy fo Vascular Plants. MacMillan, Nueva York. 1951.

[20] Nixon CK, Jensen RJ, Manos P, Muller CH. Flora of North America, North of Mexico. Vol. 3. Magnoliophyta: Magnoliidae and Hamamelidae. Oxford University Press, Nueva York. 1997.

[21] Gómez Manzaneque F. Los Bosques Ibéricos. Ed. Planeta. 1997:572 pp.

[22] Maldonado FJ, Sainz De Ollero H, Sánchez De Dios R. Distribución y estado de conservación de los bosques en España. UAM-WWF Adena. Documento no publicado. 1998.

[23] Franco J. *Quercus*. In Castroviejo, et al. (Eds.). Flora Ibérica, Consejo Sup. Inv. Cientif, Madrid, España. 1990;2:16-36.

[24] Barbero M, Loisel R, Quézel P. Biogeography, ecology and history of Mediterranean *Quercus ilex* ecosystems. Plant Ecology 1992;99-100:19-34.

[25] Ruíz De La Torre J. Flora Mayor. Organismo Autónomo Parques Nacionales. Madrid, España. 2006:1756 pp.

[26] Tuset JJ, Sánchez G. La seca: el decaimiento de encinas, alcornoques y otros *Quercus* en España. Organismo Autónomo de Parques Nacionales. Ministerio de Medio Ambiente, Madrid, España. 2004:419 pp.

[27] Jiménez MP, Díaz-Fernández PM, Iglesias S, De Tuero M, L. G. Las regiones de procedencia de *Quercus ilex* L. en España. ICONA. Madrid. 1996:93 pp.

[28] Rodríguez CI, Muños de Morales B. Fenología de *Quercus ilex* L. y *Quercus suber* L. en una dehesa del centro peninsular. Universidad Politécnica de Madrid. Escuela Universitaria de Ingeniería Técnica y Forestal. Madríd, España. 2009:100 pp.

[29] Martín BM. La dehesa. Agricultura 1996;762:44-49.

[30] Daza A, López-Bote CJ, Tomás Barberán FA, Espin JC, López Carrasco C, Olivares A, Rey AI. Effect of mediterranean forest parasite with *Curculio* sp. on nutritional value of acorn for Iberian pig feeding and fat characteristics. Meat Science 2007;76:316-320.

[31] Pulido FJ, Díaz M, Hidalgo de Trucios SJ. 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 2001;146:1-13.

[32] Jönsson U. Phytophthora species and oak decline – can a weak competitor cause significant root damage in a nonsterilized acidic forest soil? New Phytologist 2004;162:211-222.

[33] Gallego FJ, de Algaba AP, Fernandez-Escobar R. Etiology of oak decline in Spain. European Journal of Forest Pathology 1999;29:17-27.

[34] Abril N, Gion J-M, Kerner R, Müller-Starck G, Cerrillo RMN, Plomion C, Renaut J, Valledor L, Jorrin-Novo JV. Proteomics research on forest trees, the most recalcitrant and orphan plant species. Phytochemistry 2011;72:1219-1242.

[35] Confalonieri M, Balestrazzi A, Bisoffi S, Carbonera D. In vitro culture and genetic engineering of *Populus* spp: synergy for forest tree improvement. Plant Cell, Tissue and Organ Culture 2003;72:109-138.

[36] Burgarella C, Lorenzo Z, Jabbour-Zahab R, Lumaret R, Guichoux E, Petit RJ, Soto A, Gil L. Detection of hybrids in nature: application to oaks (*Quercus suber* and *Q. ilex*). Heredity 2009;102:442-452.

[37] Castro-Díez P, Villar-Salvador P, Pérez-Rontomé C, Maestro-Martínez M, Montserrat-Martí G. Leaf morphology and leaf chemical composition in three *Quercus* (Fagaceae) species along a rainfall gradient in NE Spain. Trees - Structure and Function 1997;11:127-134.

[38] Jimenez P, de Heredia UL, Collada C, Lorenzo Z, Gil L. High variability of chloroplast DNA in three Mediterranean evergreen oaks indicates complex evolutionary history. Heredity 2004;93:510-515.

[39] Lumaret R, Jabbour-Zahab R. Ancient and current gene flow between two distantly related Mediterranean oak species, *Quercus suber* and *Q. ilex*. Annals of Botany 2009;104:725-736.

[40] Lumaret R, Tryphon-Dionnet M, Michaud H, Sanuy A, Ipotesi E, Born C, Mir C. Phylogeographical variation of chloroplast DNA in cork oak (*Quercus suber*). Annals of Botany 2005;96:853-861.

[41] Toumi L, Lumaret R. Allozyme characterisation of four Mediterranean evergreen oak species. Biochemical Systematics and Ecology 2001;29:799-817.

[42] Vakkari P, Blom A, Rusanen M, Raisio J, Toivonen H. Genetic variability of fragmented stands of pedunculate oak (*Quercus robur*) in Finland. Genetica 2006;127:231-241.

[43] Lumaret R, Mir C, Michaud H, Raynal V. Phylogeographical variation of chloroplast DNA in holm oak (*Quercus ilex* L.). Molecular Ecology 2002;11:2327-2336.

[44] Michaud H, Lumaret R, Romane F. Variation in the genetic structure and reproductive biology of holm oak populations. Plant Ecology 1992;99-100:107-113.

[45] Ortego J, Bonal R, Muñoz A. Genetic consequences of habitat fragmentation in long-lived tree species: The case of the mediterranean holm oak (*Quercus ilex*, L.). Journal of Heredity 2010;101:717-726.

[46] Toumi L, Lumaret R. Allozyme characterisation of four Mediterranean evergreen oak species. Biochemical Systematics and Ecology 2001 29:799–817.

[47] Ramírez-Valiente JA, Valladares F, Gil L, Aranda I. Population differences in juvenile survival under increasing drought are mediated by seed size in cork oak (*Quercus suber* L.). Forest Ecology and Management 2009;257:1676-1683.

[48] Singh B, Saklani K, Bhatt B. Provenance variation in seed and seedlings attributes of *Quercus glauca* Thunb. in Garhwal Himalaya, India. Dendrobiology 2010;63:59–63.

[49] Vázquez FM, Espárrago F, López MJA, Jaraquemada F, Pérez MC. Descripción de la especie *Quercus rotundifolia* Lam. y sus formas para Extremadura. Colección información técnica agraria. Serie agricultura 17. Junta de Extremadura, Consejería de Agricultura y Comercio. Servicio de investigación agraria. Badajoz, España 1992:23 pp.

[50] Jiménez P, Gil L. Conservación de recursos genéticos de los *Quercus* mediterráneos en España. Investigación agraria. Sistemas y recursos forestales. Fuera de Serie. 2000;2:103-115.

[51] Leiva MJ, Fernández-Alés R. Variability in seedling water status during drought within a *Quercus ilex* subsp. *ballota* population, and its relation to seedling morphology. Forest Ecology and Management 1998;111:147-156.

[52] Rey AI, Daza A, López-Carrasco C, López-Bote CJ. Feeding Iberian pigs with acorns and grass in either free-range or confinement affects the carcass characteristics and fatty acids and tocopherols accumulation in Longissimus dorsi muscle and backfat. Meat Science 2006;73:66-74.

[53] Tejeda JF, Gandemer G, Antequera T, Viau M, García C. Lipid traits of muscles as related to genotype and fattening diet in Iberian pigs: total intramuscular lipids and triacylglycerols. Meat Science 2002;60:357-363.

[54] Özcan T. Total protein and amino acid compositions in the acorns of Turkish *Quercus* L. taxa. Genetic Resources and Crop Evolution 2006;53:419-429.

[55] Rafii ZA, Dodd RS, Pelleau Y. Biochemical diversity and systematics of Mediterranean evergreen oak from South East France. Biochemical Systematics and Ecology 1993;21:687-694.

[56] Rafii ZA, Zavarin E, Pelleau Y. Chemosystematic differentiation of *Quercus ilex* and *Q. rotundifolia* based on acorn fatty acids. Biochemical Systematics and Ecology 1991;19:163-166.

[57] Steele MA, Hadj-Chikh LZ, Hazeltine J. Caching and feeding decisions by *Sciurus carolinensis*: responses to weevil-infested acorns. Journal of Mammalogy 1996;77:305-314.

[58] Vázquez FM. Semillas del género *Quercus* L. (Biología, Ecología y Manejo). Servicio de Investigación y Desarrollo Tecnológico, Consejería de Agricultura y Comercio. Junta de Extremadura, Mérida, España. 1998:234 pp.

[59] Cozzolino D. Uso de la espectroscopía de reflectancia en el infrarrojo cercano (NIRS) en el análisis de alimentos para animales. Agrociencia 2002;6:25-32.

[60] Alomar D, Fuchslocher R. Fundamentos de la espectroscopía de reflectancia en el infrarrojo cercano (NIRS) como método de análisis de forrajes. Agrosur, Chile 26 1. 1998.

[61] Rodríguez-Estévez V, García A, Gómez AG. Characteristics of the acorns selected by free range Iberian pigs during the montanera season. Livestock Science 2009;122:169-176.

[62] Fernández I, Gómez A, Moreno P, Fernández P, ED DP, íaz E, Sánchez L. Characteristics of the acorns of oak (*Quercus ilex*) in different zones from the valley of the pedroches (Córdoba). Bromatologic inventory. Producciones agroganaderas: Gestión eficiente y conservación del medio natural. XLV RC de la SEEP. Gijón, Spain. 2005:377-385 pp.

[63] 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. From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and arnino acid analysis. Natute Biotechnology 1996;14:61-65.

[64] Cánovas FM, Dumas-Gaudot E, Recorbet G, Jorrin J, Mock H-P, Rossignol M. Plant proteome analysis. Proteomics 2004;4:285-298.

[65] Jorrín JV, Maldonado AM, Castillejo MA. Plant proteome analysis: A 2006 update. Proteomics 2007;7:2947-2962.

[66] Jorrín-Novo JV, Maldonado AM, Echevarría-Zomeño S, Valledor L, Castillejo MA, Curto M, Valero J, Sghaier B, Donoso G, Redondo I. 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 2009;72:285-314.

[67] Bahrman N, Petit RJ. Genetic polymorphism in maritime pine (*Pinus pinaster* Ait.) assessed by two-dimensional gel electrophoresis of needle, bud, and pollen proteins. Journal of Molecular Evolution 1995;41:231-237.

[68] Chevalier F, Martin O, Rofidal V, Devauchelle AD, Barteau S, Sommerer N, Rossignol M. Proteomic investigation of natural variation between *Arabidopsis* ecotypes. Proteomics 2004;4:1372-1381.

[69] Emre İ. Electrophoretic analysis of some *Lathyrus* L. species based on seed storage proteins. Genetic Resources and Crop Evolution 2009;56:31-38.

[70] Jacobsen S, Nesic L, Petersen M, Sondergaard I. Classification of wheat varieties: use of two-dimensional gel electrophoresis for varieties that can not be classified by matrix assisted laser desorpiton/ionization-time of flight-mass spectrometry and an artificial neural network. Electrophoresis 2001;22:1242-1245.

[71] Ostergaard O, Melchior S, Roepstorff P, Svensson B. Initial proteome analysis of mature barley seeds and malt. Proteomics 2002;2:733-739.

[72] Singh N, Matta N. Variation studies on seed storage proteins and phylogenetics of the genus *Cucumis*. Plant Systematics and Evolution 2008;275:209-218.

[73] Yüzbaşioğlu E, Açik L, Özcan S. Seed protein diversity among lentil cultivars. Biologia Plantarum 2008;52:126-128.

[74] Bonhomme L, Monclus R, Vincent D, Carpin S, Claverol S, Lomenech AM, Labas V, Plomion C, Brignolas F, Morabito D. Genetic variation and drought response in two *Populus* x *euramericana* genotypes through 2-DE proteomic analysis of leaves from field and glasshouse cultivated plants. Phytochemistry 2009;70:988-1002.

[75] Bonhomme L, Monclus R, Vincent D, Carpin S, Lomenech AM, Plomion C, Brignolas F, Morabito D. Leaf proteome analysis of eight *Populus xeuramericana* genotypes: genetic variation in drought response and in water-use efficiency involves photosynthesis-related proteins. Proteomics 2009;9:4121-4142.

[76] Petit RJ, Bahrman N, Baradat PH. Comparison of genetic differentiation in maritime pine (*Pinus pinaster* Ait.) estimated using isozyme, total protein and terpenic loci. Heredity 1995;75:382-389.

[77] Vienne DD, Burstin J, Gerber S, Leonardi A, Guilloux ML, Murigneux A, Beckert M, Bahrman N, Damerval C, Zivy M. Two-dimensional electrophoresis of proteins as a source of monogenic and codominant markers for population genetics and mapping the expressed genome. Heredity 1996;76:166-177.

[78] Martín M, Navarro-Cerrillo R, Ortega P, Alvarez J. The use of cotyledon proteins to assess the genetic diversity in sweet holm oak. Journal of Forest Science 2009;55:526-531.

[79] Montoya J. Técnicas de reforestación con encinas, alcornoques y otras especies de Quercus mediterráneos. In: Orozco-Bayo, E. & Monreal-Montoya, J. (eds.) Forestación en tierras agrícolas. Universidad de Castilla-La Mancha, Cuenca , España. 1997:199-213 pp.

[80] Aizen MA, Woodcock H. Latitudinal trends in acorn size in eastern North American species of *Quercus*. Canadian Journal of Botany 1992;70:1218-1222.

[81] Charef M, Yousfi M, Saidi M, Stocker P. Determination of the fatty acid composition of acorn (*Quercus*), *Pistacia lentiscus* seeds growing in Algeria. Journal of the American Oil Chemists' Society 2008;85:921-924.

[82] León-Camacho M, Viera-Alcaide I, Vicario I. Acorn (*Quercus spp.*) fruit lipids: Saponifiable and unsaponifiable fractions: A detailed study. Journal of the American Oil Chemists' Society 2004;81:447-453.

[83] Tejerina D, García-Torres S, Cabeza de Vaca M, Vázquez FM, Cava R. Acorns (*Quercus rotundifolia* Lam.) and grass as natural sources of antioxidants and fatty acids in the “montanera” feeding of Iberian pig: Intra- and inter-annual variations. Food Chemistry 2011;124:997-1004.

[84] Bruschi P, Grossoni P, Bussotti F. Within- and among-tree variation in leaf morphology of *Quercus petraea* (Matt.) Liebl. natural populations. Trees - Structure and Function 2003;17:164-172.

[85] Dodd RS, Rafii ZA, Bojovic S. Chemosystematic study of hybridization in californian live oak: Acorn steroids. Biochemical Systematics and Ecology 1993;21:467-473.

[86] Bonner F, Vozzo J. 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 1987:21 pp.

[87] Corine H, Serge R. An examination of the interaction between climate, soil and leaf area index in a *Quercus ilex* ecosystem. Annals Forest Science 2003;60:153-161.

[88] Ducousso A, Guyon JP, Kremer A. Latitudinal and altitudinal variation of bud burst in western populations of sessile oak (*Quercus petraea* (Matt) Liebl). Annals Forest Science 1996;53:775-782.

[89] Morales R, Vicente JA, Galán de Mera A. cpDNA evidence of introgressions in *Quercus* L. (Fagaceae). The influence of the phytotopographic position. Flora 2005;200:222-228.

[90] Casey R, Short MN. Variation in amino acid composition of legumin from *Pisum*. Phytochemistry 1981;20:21-23.

[91] Chileh T, Esteban-García B, Diego López A, García-Maroto F. Characterization of the 11S globulin gene family in the castor plant *Ricinus communis* L. Journal of Agricultural and Food Chemistry 2009;58:272-281.

[92] Derbyshire E, Wright DJ, Boulter D. Legumin and vicilin, storage proteins of legume seeds. Phytochemistry 1976;15:3-24.

[93] Maplestone P, Allison J, Hussein EHA, Gamal El-Din AE-KY, Gatehouse JA, Boulter D. Variation of the legumin seed storage protein amongst *Vicia* species. Phytochemistry 1985;24:1717-1723.

[94] Scippa GS, Rocco M, Ialicicco M, Trupiano D, Viscosi V, Di Michele M, Arena S, Chiatante D, Scaloni A. The proteome of lentil (*Lens culinaris* Medik.) seeds: Discriminating between landraces. Electrophoresis 2010;31:497-506.

[95] Wobus U, Bäumlein H, Bassüner R, Heim U, Jung R, Müntz K, Saalbach G, Weschke W. Characteristics of two types of legumin genes in the field bean (*Vicia faba* L.var. minor) genome as revealed by cDNA analysis. FEBS Letters 1986;201:74-80.

[96] Gallardo K, Le Signor C, Vandekerckhove J, Thompson RD, Burstin J. Proteomics of *Medicago truncatula* seed development establishes the time frame of diverse metabolic processes related to reserve accumulation. Plant Physiologists 2003;133:664-682.

[97] Baier M, Dietz KJ. The plant 2-Cys peroxiredoxin BAS1 is a nuclear-encoded chloroplast protein: its expressional regulation, phylogenetic origin, and implications for its specific physiological function in plants. The Plant Journal 1997;12:179-190.

[98] Baier M, Dietz KJ. protective function of chloroplast 2-Cysteine peroxiredoxin in photosynthesis. Evidence from transgenic Arabidopsis. Plant Physiology 1999;119:1407-1414.

[99] Baier M, E S, Dietz KJ. The acceptor availability at photosystem I and ABA control nuclear expression of 2-Cys peroxiredoxin-A in Arabidopsis thaliana. Plant Cell Physiology 2004;45 997-1006.

[100] Bestel-Corre G, Dumas-Gaudot E, Gianinazzi-Pearson V, Gianinazzi S. Mycorrhiza-related chitinase and chitosanase activity isoforms in *Medicago truncatula* Gaertn. Symbiosis 2002;32:173-194.

[101] Salzer P, Bonanomi A, Beyer K, Vögeli-Lange R, Aeschbacher RA, Lange J, Wiemken A, Kim D, Cook DR, Boller T. Differential Expression of Eight Chitinase Genes in Medicago truncatula Roots During Mycorrhiza Formation, Nodulation, and Pathogen Infection. Molecular Plant-Microbe Interactions 2000;13:763-777.

[102] Gallardo K, Job C, Groot SP, Puype M, Demol H, Vandekerckhove J, Job D. Proteomic analysis of arabidopsis seed germination and priming. Plant Physiologists 2001;126:835-848.

[103] Roulston TaH, Cane JH, Buchmann SL. What governs protein content of pollen: pollinator preferences, pollen–pistil interactions, or phylogeny? Ecological Monographs 2000;70:617-643.

[104] Saa-Otero MP, Díaz-Losada E, Fernández-Gómez E. Analysis of fatty acids, proteins and ethereal extract in honeybee pollen - Considerations of their floral origin. Grana 2000;39:175-181.

[105] Noir S, Bräutigam A, Colby T, Schmidt J, Panstruga R. A reference map of the *Arabidopsis thaliana* mature pollen proteome. Biochemical and Biophysical Research Communications 2005;337:1257-1266.

[106] Erler A, Hawranek T, Krückemeier L, Asam C, Egger M, Ferreira F, Briza P. Proteomic profiling of birch (*Betula verrucosa*) pollen extracts from different origins. Proteomics 2011;11:1486-1498.

[107] Dai S, Li L, Chen T, Chong K, Xue Y, Wang T. Proteomic analyses of *Oryza sativa* mature pollen reveal novel proteins associated with pollen germination and tube growth. Proteomics 2006;6:2504-2529.

[108] Holmes-Davis R, Tanaka CK, Vensel WH, Hurkman WJ, McCormick S. Proteome mapping of mature pollen of *Arabidopsis thaliana*. Proteomics 2005;5:4864-4884.

[109] Sheoran IS, Ross AR, Olson DJ, Sawhney VK. Proteomic analysis of tomato (*Lycopersicon esculentum*) pollen. Journal of Experimental Botany 2007;58:3525-3535.

[110] Sheoran I, Sproule K, Olson D, Ross A, Sawhney V. Proteome profile and functional classification of proteins in *Arabidopsis thaliana* (Landsberg erecta) mature pollen. Sexual Plant Reproduction 2006;19:185-196.

[111] Krichevsky A, Kozlovsky SV, Tian G-W, Chen M-H, Zaltsman A, Citovsky V. How pollen tubes grow. Developmental Biology 2007;303:405-420.

[112] Li J, Chen J, Zhang Z, Pan Y. Proteome analysis of tea pollen (*Camellia sinensis*) under different storage conditions. Journal of Agricultural and Food Chemistry 2008;56:7535-7544.

[113] Barranca M, Fontana S, Taverna S, Duro G, Zanella-Cleon I, Becchi M, De Leo G, Alessandro R. Proteomic analysis of *Parietaria judaica pollen* and allergen profiling by an immunoproteomic approach. Biotechnology Letters 2010;32:565-570.

[114] Wallner M, Erler A, Hauser M, Klinglmayr E, Gadermaier G, Vogel L, Mari A, Bohle B, Briza P, Ferreira F. Immunologic characterization of isoforms of Car b 1

and Que a 1, the major hornbeam and oak pollen allergens. Allergy 2009;64:452-460.

[115] Karamloo F, Wangorsch A, Kasahara H, Davin LB, Haustein D, Lewis NG, Vieths S. Phenylcoumaran benzylic ether and isoflavonoid reductases are a new class of cross-reactive allergens in birch pollen, fruits and vegetables. European Journal Biochemistry 2001;268:5310-5320.

9. Anexos

Chapter 2286

NATURAL VARIABILITY AND RESPONSES TO STRESSES IN ANDALUSIA HOLM OAK (QUERCUS ILEX SUBSP. BALLOTA [DESF.] SAMP.) POPULATIONS

***Jose Valero Galvan*[*1], *Besma Sghaier-Hammami*[1], *Rafael Mª Navarro Cerrillo*[2], *Jesus V. Jorrin-Novo*[1]**

Agricultural and Plant Biochemistry and Proteomics Research Group, Dept. of Biochemistry and Molecular Biology, University of Córdoba, Córdoba, Spain (http://www.uco.es/investiga/grupos/probiveag/)[1]

Department of Forestry Engineering, ETSIAM, University of Córdoba, Córdoba, Spain[2]

ABSTRACT

This work aims to highlight the importance of *Quercus ilex* in Mediterranean forestry. Holm oak fruits (acorns) are essential for wildlife, as well as for pig fattening in *dehesas*. In addition, Holm oak is acquiring more interest for Mediterranean reforestation. In this review we present studies related to the analysis of natural variability and responses to stresses in Andalusia Holm oak populations. We present data from our own research [1-7] and those found in the recent literature, emphasizing the uses of classical morphometry, together with the modern, holistic, -omics approaches and near-infrared spectroscopy. Furthermore, we discuss the difficulty of using a recalcitrant species as an experimental system and the limitations of some of the proposed techniques. By using different approaches, we obtain a deeper knowledge on natural variability and biological processes, including growth, development, organogenesis, and responses to stresses.

The present review chapter is organized according to four sections, namely: (i) the relevance, main problems and challenges related to conservation and use (Introduction); (ii) the research with this species (from the field to the lab), focusing on the last one; (iii) the study of variability (from morphometry to –omics- technologies); and (iv) and the responses to abiotic stresses, concretely to drought stress.

[*]Corresponding author: José Valero Galván, Departamento de Bioquímica y Biología Molecular, Universidad de Córdoba, Campus de Rabanales, Ed. Severo Ochoa, E-14071 Córdoba, Spain. e-mail: bb2vagaj@uco.es Fax: +34 957 218439

1. INTRODUCTION

Holm oak (*Quercus ilex*) is a monoecious, wind-pollinated oak tree included in the subgenes sclerophyllodrys, specie *ilex* that has two subspecies *ilex* and *ballota* [8]. The Holm oak is a widely distributed species throughout the Mediterranean basin [9], where it is present over a large area covering 6000 km on longitude from Portugal to Syria, and 150 km on latitude from Morocco and Algeria to France. In Spain, Holm oak woodlands cover around 3 million ha either in pure or mixed stands, commonly mixed with other *Quercus* species (*Quercus suber* L., *Quercus faginea* Lam., and *Quercus pyrenaica* Willd.). The most conspicuous species form the traditional agrosilvopastoral system called *dehesa* and the main populations of this species are distributed over the Valle del Duero, Levante, Aragón, Valle del Ebro, Castilla-La Mancha and Andalusia. They are located over a wide range of environmental conditions; mainly in the thermo- and meso- Mediterranean altitudinal levels, and from semiarid to humid Mediterranean climates (350-1200 mm of annual rainfall); from sea level to 2100 m a.s.l. In Andalusia region, Holm oaks occupy a territorial extension of approximately 735 671 ha and two types of oak woods have been described, the *luso-extremeño* populations and the *béticas* populations. The *luso-extremeño* Holm oaks communities are located in the western zone of Andalusia, in the provinces of Huelva, Seville and Cordoba (Fig. 1). When it is well preserved, those forests are quite homogeneous and very rich in plant species. In this type of Holm oak, savannah-like forests are dominant with structure of silvopastroal system known as *dehesa*. In the oriental area, the *béticas* communities located at Sierras Béticas and Penibéticas, Holm oak forests has a shrubtype structure with a high density. At the Iberian Peninsula, two subspecies of *Q. ilex* have been described [8]. *Quercus ilex* L., subsp. *ilex* is distributed in northeast littoral areas where the winter is relatively warm; whereas *Quercus ilex* L. subsp. *ballota* (Desf.) Samp. occupies continental areas with colder winters and hotter summers than those prevailing in littoral areas [8]. *Quercus ilex* subsp. *ilex* and *Q. ilex* subsp. *ballota* differ in their physiological characteristics, including vulnerability to water-stress-induced embolism [10].

Figure 1. Holm oak growing in deheza located in Pozoblanco, Córdoba, Spain.

Holm oak seeds are a major component in the feeding systems of many Mediterranean wild and livestock species with an estimated value of over 120 million Euros in fruit acorn production per year [11]. Moreover, it is also the basic feed ingredient for domestically-bred high quality meat pigs [11]. Today, there is an increasing interest in the restoration of savannah-like oak forests on abandoned agricultural land along the Mediterranean basin countries. Only in Andalusia (Spain), more than 75 000 ha were reforested from 1993 to 1999, thanks to the introduction of millions of new seedlings [12, 13]. However, in the *dehesas* of Mediterranean areas of western and south-western Spain, where grasslands reach 70±90% cover, tree regeneration is very low because of an inappropriate livestock management, together with a high acorn predation by a variety of vertebrate herbivores (i.e., cattle, deer, wild boar, mice and rabbits), water limitations, and the dramatic effect of the diseases which occur periodically [14-17]. Those factors are considered the major limitations for the seedling establishment, and subsequent growth of tree stands, and plant distribution Holm oak ecosystems. Criteria for tree selection for restoration programs are frequently the production of big and abundant acorns per individual tree. However, in order to create systems which are able to self-regenerate, it is important to take into account the seedling traits related to abiotic (drought) and biotic (pest and diseases) tolerance.

Q. ilex subsp. *ballota,* as well as other tree species, has characteristics that favor a high intrapopulation genetic diversity: long life span, allogamy, wind pollination, monoecy, and a continuous geographical distribution [18], which has led to a high intra-population variability [19]. Spanish Holm oak populations have showed a continue isolation and complex geographic patterns of genetic variation [18, 20], being native and distributed through diverse forest areas. Some populations have survived under the extreme dry climate conditions prevailing in southern Spain areas, so these individuals are expected to present efficient mechanisms of adaptability to these conditions. However, the characterization of individual trees from a given population is quite challenging. Extreme variability occurs especially in the populations of fairly broad geographic distributions. Several problems of tree typification, particularly in the wide-ranging species, are still unresolved, while modern techniques to catalogue and select genotypes among provenances are needed to improve phylogenetic analysis.

2. RESEARCH WITH THIS SPECIES (FROM THE FIELD TO THE LAB)

The research of Mediterranean oaks variability includes the characterization of natural diversity and the selection of superior phenotypes (plus trees *vs.* selected stands) in natural populations, as the bases for ulterior breeding programs. Major objectives of those programs range from yield improvement to adaptation to adverse environmental conditions or stresses. Currently, tree breeding is gaining new interest under the threat of global warming. Scientists carry out research and develop new tools aiming to improve the efficiency of tree breeding programs from a scientific rather than an empirical approach. Molecular markers can be correlated with growth, productivity, and environmental adaptation phenotypes and would help in conserving biodiversity, breeding programs and selecting elite mother trees. Nevertheless, forest tree research presents some challenges as experimental systems because of large genome, long life cycles, and low growth together with difficulties in sample

preparation analysis, recalcitrance for genetic transformation and *in vitro* regeneration [21]. These different factors constitute the main constraints and challenges in the development of classical and modern holistic -omic molecular studies, and the development of classical or biotechnological breeding programs [22]. Thus, an *ISI Web of Knowledge* search covering the period 1990–2011 (July), using as a search string "*Quercus*" genus generated only 11 244 hits compared with the generated by *Arabidopsis* (67 660), and *Oriza spp.* (75 115). Only 23% of these researches were focused on the study the tree Mediterranean evergreen *Quercus* species, including *Q. ilex* (1,613), *Q. suber* (704), and *Q. coccifera* (293). From a total of 1,613 hits generated, only 9% (145) were related to study Holm oak variability.

The existence of a high variability and polymorphism in *Quercus* spp. has been concluded from data obtained by using anatomical, chemical, and enzyme activity indicators, as well as DNA marker techniques, the latter also used for introgression and phylogeographical studies [23-29]. However, there is little information related to study the natural variability of Spain Holm oak forest trees [18, 19, 25, 30, 31]. In this context, Holm oak proteomics researches have been conducted to study: i) leaf proteome and different plant developmental stages [3], ii) in different populations and under drought conditions[1, 2]; and iii) seed proteome[7]. The results often indicated a remarkable difference between regional populations.

3. STUDY OF VARIABILITY (FROM MORPHOMETRY TO –OMICS-TECHNOLOGIES)

Acorns are a major feed resource for livestock and wild mammals, birds and insects in oak forests. Acorn development is very slow during spring and summer, and the fruits are still very small (less than 0.5 cm in diameter) in early September, when the fast growth period begins. Individual trees show highly variable intra- and inter-annual production, depending on several environmental and endogenous factors, at least from late summer and autumn of the previous year, until autumn in the dissemination year [32]. Many studies related to acorn production by *Q. ilex* have focused on crop quantification [11], but these researches left out acorn size and weight, which are important in calculating acorn yield, and it has been suggested as a good character for seed selection on seedling production and establishment [33, 34]. Most oak morphometry studies have characterized population variation using acorn and leaf size [35, 36]. However, those works have neither included a sufficiently wide-range survey populations nor complementary biochemical approaches for characterization of population differences. In order to obtain current information of acorn morphometry and chemical composition, 13 populations of Andalusia region (southern Spain and an area of interest for the Iberian pig rearing system) have been analyzed [6]. Populations located geographically in northern (humid) areas were expected to have higher length, diameter and weight acorns than southern (dry) populations. Moreover, it was postulated that acorn size would decrease with the location altitude as it has been observed with a significant inverse correlation of seed length and weight with altitude exhibited on *Quercus glauca* [36]. Selection of big acorns, as a consequence of selection of tall seedlings, is likely to be related to better response in water-limiting environments; indeed, a positive relationship among acorn size, seedling height and summer survival has been reported for *Q. ilex* [33]. Bigger acorns

lead to increased seedling growth, higher root and shoot ratios, which can improve seedling performance by developing and maintaining deep and extensive root system accessing deeper water layers within the soil profile [35, 36]. Seedlings from bigger acorns averaged a higher number of leaves and a higher leaf area which might be an advantage for young seedlings [37].

Acorn Chemical Composition

Acorn chemical composition of Holm oak has a high economic importance in the feeding of many Mediterranean wild and livestock species (i.e. Iberian pigs). Acorn chemical composition of oaks has shown to be a valuable tool for assessing variability [38-40], but few studies have been carried out on acorn chemical composition of Holm oak from a natural variability point of view (Table 1). Acorns have high fat content (4-12%), with a high content of oleic acid (> 63% of total fatty acids), and palmitic and linoleic acids concentrations (12-20%).

Table 1. Variation in chemical composition of different population of Holm oak. The descriptive statistics are presented according to the mean.

Chemical composition	References			
	[45]	[44]	[46]	[6]
Pulp DM (%)	56.72-60.57	68.70-66.00	56.00	63.90-83.60
Ash (%)[1]	1.99-1.89	1.50-1.50	1.80-2.00	1.34-2.02
Crude protein (%)[1]	04.48-04.8	4.60-5.40	3.70-5.00	3.90-5.94
Fiber (%)[1]	2.71-2.92	5.70-5.90	0.81-1.06	-
Starch (%)[1]	59.73-58.28	-	-	-
Sugars (%)[1]	5.89-6.84	-	-	8.95-12.47
Crude fat (%)[1]	10.05-10.76	5.60-7.00	9.50-10.40	9.14-14.95
C16 (%)[2]	13.97-14.28	-	16.30-20.60	12.15-14.65
C16:1 (%)[2]	0.12-0.11	-	-	-
C17 (%)[2]	0.10-0.10	0.10-0.10	-	-
C17:1 (%)[2]	0.09-0.1	-	-	-
C18 (%)[2]	3.15-3.28	3.60-3.40	3.30-4.20	-
C18:1 (%)[2]	64.99-65.45	56.70-62.90	53.64-62.50	64.98-67.81
C18:2 (%)[2]	15.21-15.31	17.70-18.50	12.20-18.20	15.62-17.14
C18:3 (%)[2]	0.73-0.81	1.10-1.70	-	-
C20 (%)[2]	0.44-0.46	0.30-0.30	-	-
C20:1 (%)[2]	0.56-0.57	-	-	-

[1]As percent of pulp DM; [2]As percent of crude fat

These compounds are very important in the pig diet because they affect the Iberian dry ham aging process [41]. Starch was the main acorn component (50%), in contrast with the low protein content (4-6%). Chemosystematic differentiation based on differences in acorn fatty acid composition between Italian and Spanish populations of *Q. ilex* yielded partial separations of the individual populations [40]. Such separation using acorn fatty acids for

native and hybrid populations have also been observed for *Q. agrifolia* and *Q. wislizenii* [42]. French Mediterranean evergreen oak populations were reported to be intermediate and heterogeneous in fatty acid profiles between the Spanish and the Italian oak, suggesting a zone of hybridization [39]. According to this, we have applied the near-infrared spectroscopy technique to predict a complete chemical analysis of acorn flour in an attempt to compare, catalog, and characterize natural populations of Holm oak from the Andalusia region. Results indicated that there were statistic significant differences in acorn chemical composition between the different Holm oak populations [6]. The analysis showed that northern populations had a tendency to present higher values in acorn weight, length and diameter, and in ash, protein, fat; and oleic acid content; while southern population showed lower values in sugar, energy, palmitic and linoleic parameters. Variations in fatty acid levels were observed in different taxa [38, 43] and populations of oaks [39, 40, 42]. Moreover, populations studied on this work seemed to integrate acorn morphology and chemical composition related to environmental conditions of the original provenances. Variations in the acorn chemical composition could be related to factors associated with (i) weather conditions, such as precipitation, temperature and temperature oscillations; (ii) acorn physiological conditions such as ripening or germination; and (iii) sanitary conditions, such as desiccation, dampening, rotting and attacks from pests, such as *Curculio* spp., *Cydia* spp. or phytophages [44].

Holm Oak Proteome

Given that genetic information is only indicative of the cell's potential and does not reflect the actual state in a particular cell at a given time, the concept of "proteome" (protein complement expressed by a genome [47]) has emerged. Proteome provides complementary and critical information by revealing the regulation, activities, quantities and interaction of protein in cells, as well as how their abundance responds to developmental and environmental signals. According to the research objective, different areas within proteomics can be defined (reviewed in [48]), namely: i) descriptive proteomics, including intracellular and subcellular proteomics; ii) differential expression proteomics, iii) post-translational modifications; iv) interactomics; and v) proteinomics (targeted or hypothesis-driven proteomics). The workflow of a standard proteomic experiment includes all or most of the following steps: experimental design, sampling, preparation of tissue, cell or organelle, protein extraction, fractionation or purification, labeling or modification, separation, MS analysis, protein identification and statistical analysis of data and validation (Fig. 2). Previous studies have demonstrated the complexity of protein extraction from vegetable tissues, and how the protein solubilization is critical for gel preparation with discrete protein spots that are suitable for MS analysis [21, 48-50]. The most appropriate protocols should be developed for several species [51], and must be optimized for biological systems (i.e. plant species, organ, tissue, cells), as well as for research objectives. In our experience, although phenol extraction is considered time-consuming, it generates high-quality protein extract from a large variety of plant species [51]. The importance of the extraction protocol in proteomic experiments may be summarized in the following statement: only if you can extract and solubilize a protein, you will have the chance to detect and identify it. This is even more important in plant tissues, due to the low protein content compared to other biological systems because of the presence of the cell wall

and vacuoles that account the majority of cell mass. Additionally, the presence of proteases and oxidative enzymes, together with the accumulation of large quantities of polysaccharides, lipids, phenolics, and other secondary metabolites must be taken into account. Currently, the separation and identification of proteins is possible through techniques such as the combination of SDS-PAGE, band cutting, trypsin digestion and LC separation of the resulting peptides which remains the proteomic platform capable of providing the best results in terms of protein coverage. Nevertheless, the predominant separation technology used in plant proteomics is still by far 2-DE, being continuously evaluated and improved. The technology 2-DE is used for separating and displaying the components of large protein complexes. It shows as main advantages the simplicity, reproducibility, a wide size range (10 kDa to 500 kDa) and the fact that both moderately hydrophobic and acidic-basic proteins can be isolated and visualized. It is important to point out the restrictions of gel based techniques, such as low abundance proteins, limited *pI* range and absence of membrane proteins may limit the broad mapping of proteins in plant samples. In order to approach the complexity of the protein functional machinery, continuous improvements in techniques and protocols for high-throughput proteomics are being made at all workflow stages. The increasing development of new protocols, platforms and workflows (some of them being quite complex and requiring sophisticated equipment and expertise) has entailed a huge amount of data generated. Some of these data have been deposited and organized in several databases that are available to researchers [48].

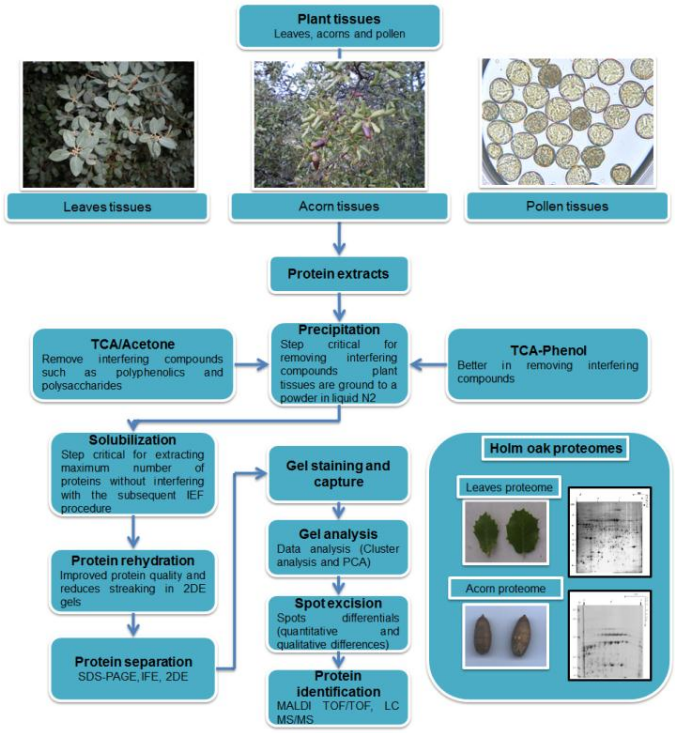


Figure 2. Conditions for the preparation of protein extract from plants tissues that are suitable for 2-DE, and different strategies of separation and identification of proteins.

Genes, visualized as precise protein bands or spots which reflect physiological status are good candidates for assessing variability, establishing genetic distances and phylogenetic relationships between different species and individuals, being used in studies with herbaceous species [52-59]. In forest tree, mainly in *Pinus* spp., 2-DE analysis has showed to be a valuable tool which provides informative molecular markers for assessing variability and the establishment of linkage maps [52, 60, 61]. A proteomic research program is being carried out with *Quercus ilex* subsp. *ballota* in order to study variability of Holm oaks populations in Andalusia [1-3]. Jorge et al. [3] reported the first description of the leaf proteome of Holm oak. The biological variance calculated for 100 major spots was of 56%, and it depended on leaf orientation (North, South, East and West), crown position (top, bottom), and the time of leaves collection. As a consequence, it has been almost impossible to use the leaf protein profile in cataloguing trees or populations since the intra-population variability can be greater than the inter-population variability. However, four protein spots were found to be significantly different among three provenances [3]. Spots 103 and 104 were only detected in Qi16a seedlings, and were identified as an ATP synthase alpha chain (spot 103) and 2, 3-bisphosphoglycerate-independent phosphoglycerate mutase (spot 104). The two unidentified spots, 101 and 102, were not detected in leaf extracts from the Qi11e provenance, and there were non-significant differences in protein amounts between Qi14a and Qi16a provenances. Furthermore, a study of variability in Holm oak seed has been carried out by using SDS-PAGE and 2-DE in ten populations distributed throughout the Andalusia region [7], which was expected to be less variable than the leaf proteome. Protein content was correlated with acorn size (weight, length and diameter), latitude and altitude date, and average monthly maximum temperature. The results showed that populations located geographically in the northern areas would be expected to have higher protein content than southern populations. Moreover, it would decrease with the altitude location. Finally, the protein content would increase with the mean monthly maximum temperature where the populations are growing. The clustering analysis of acorn protein profiles showed that northern populations separated of southern populations had a tendency to clustering together. The main restriction of the approach used is the low number of proteins identified, most of them corresponding to the reserve 11S globulin family of legumins which have a difficult interpretation from a biological point of view. Andalusia Holm oak populations showed a high variability according to the glutelin acorn protein fraction [4]. Finally, Holm oak pollen proteome has been studied as an alternative means to describe natural variability in different populations and interpret it in biological terms, for being a simpler and more stable organ than others such as leaf or fruit. A comparison of the protein profiles of pollen from four Holm oak population revealed that 17 protein spots were differentially expressed between them. The multivariate statistical analyses carried out on these variable expressed spots showed that 9 protein species were essential for population discrimination, hence determining their proposition as population markers.

4. PROTEOMICS IN THE STUDY OF RESPONSES TO ABIOTIC (DROUGHT)

Plant proteomic studies published to date have proven to be very useful in characterizing the response to drought stress in different forest tree species [1, 2, 62, 63]. In oak species, these studies have only been performed at the morphological and physiological level [17, 64-67]. Moreover, drought effects on morphological and physiological characteristics as predawn leaf water potential, leaf and root relative water content, height and root collar diameter growth and chlorophyll fluorescence were investigate in one-year-old-seedlings obtained from seed of Holm oak Andalusia populations [1, 2]. Drought stress is associated with a reduced water availability and cellular dehydration. Therefore, changes in cellular metabolism associated with an osmotic adjustment could be expected. Proteome changes upon drought have been studied in oaks [1, 2, 68, 69]. Some differentials proteins identified belong to enzymes of photosynthesis, and carbohydrate and nitrogen metabolism categories, as well as to the group of stress-related proteins. RubisCO and RubisCO activase were absent under drought condition. In contrast, RubisCO small subunit was only present under drought conditions. Furthermore, under drought conditions, globulin and HMW-glutenin and one enzyme involved in carbohydrate metabolism, b-amylase was induced. Moreover, one fructose–bisphosphate aldolase was enhanced in response to drought. This enzyme is one of the key regulatory glycolytic enzymes [2]. Additionally to the protein reported in the last work, three PSII oxygen evolving complex (PSII OEC) protein decreased in intensity in the seedlings which were under drought stress. Triosephosphate isomerases showed lower intensity values in drought and recovered plants than in watered seedlings. One peroxidase increased in intensity in plants stressed by drought, while one peroxiredoxin disappeared in water stressed plants. Other proteins involved in different metabolic pathways were identified, including a glutamine synthetase, quinoneoxido reductase and isoflavone reductase [1]. Changes in the protein profiles as a consequence of drought stress in two populations also have been observed after a 28-day-stress-preriod. Some of the identified proteins that present differences between well-watered and drought seedling were three ATP synthase beta subunit, one photosystem II oxygen- evolving complex protein 2, and LHCII type I chlorophyll a/b binding protein.

5. CONCLUSION

Different approaches can be used to characterize natural variability in plants species, from the morphological tools to the modern -omics techniques. However, these have been scarce used in *Quercus ilex*. Our research group is pioneer at this respect. Firstly, we have used a morphometry, and near infrared spectroscopy approach to study the natural variability in populations of southern Spain. Data showed that northern populations of *Q. ilex* are different from southern populations in their acorn morphology and chemical composition. Secondly, we have optimized the protein profiles of leaf, acorn and pollen to study Holm oak natural variability. Holm oak leaves present a great variability in the expression level of most of the major proteins, and has proven to be useful in the identification of protein spots that can be used as markers for differentiating holm oak provenances. The data obtained from

acorn protein profiles were used to discriminate and correlate protein profile with some climate and geographical date. Thirdly, we have used proteomics for studying plant responses to drought. Qualitative changes in the leaf 2-DE protein map have been observed under drought conditions, and these changes can be interpreted in terms of metabolic adaptations to water deficit conditions, including photosynthesis inhibition, mobilization of the protein and carbohydrate reserves, and increased glycolysis.

ACKNOWLEDGEMENTS

José Valero was a recipient of an Alban Programme fellowship (I06D00010MX). This work is part of research projects financed by the Spanish Ministry of Science and Innovation, and cofunded by FEDER: CGL2008-04503-C03-01/BOS, AGL2002-00530, and AGL2009-12243-C02-02.

REFERENCES

[1] Echevarría-Zomeño S, Ariza D, Jorge I, Lenz C, Del Campo A, Jorrin JV and Navarro RM. *Changes in the protein profile of Quercus ilex leaves in response to drought stress and recovery.* Journal of Plant Physiology. 2009;166(3):233-45.

[2] Jorge I, Navarro RM, Lenz C, Ariza D and Jorrin J. *Variation in the holm oak leaf proteome at different plant developmental stages, between provenances and in response to drought stress.* Proteomics. 2006;6 Suppl 1:S207-14.

[3] Jorge I, Navarro RM, Lenz C, Ariza D, Porras C and Jorrin 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(1):222-34.

[4] Martín M, Navarro-Cerrillo R, Ortega P and Alvarez J. *The use of cotyledon proteins to assess the genetic diversity in sweet holm oak.* Journal of Forest Science. 2009;55(11):526-31.

[5] Sghaier-Hammami B, Valero Galván, J, Romero Rodríguez MªC, Navarro Cerrillo RMª, Jorrín Novo *J 2-DE based proteomic analysis of the response of Spanish Holm oak to Phytophthora cinnamomi reveals variability among provenances.* Final COST FA0603 Meeting, Dijon, France. 2011:56.

[6] Valero-Galván J, Jorrín-Novo J, Gómez C, Ariza D, García-Olmo J and Navarro R. *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. 2011 b;DOI: 10.1007/s10342-011-0563-8.

[7] Valero-Galván J, Valledor L, RM N, Gil E and JV J-N. *Studies of variability in Holm oak (Quercus ilex subsp. ballota [Desf.] Samp.) through acorn protein profile analysis.* Journal of Proteomics. 2011 a;74:1244 –55.

[8] Franco J. *Quercus.* In: Castroviejo, et al. (Eds.), Flora Ibérica, Consejo Sup. Inv. Cientif, Madrid, pp. 1990;2:16-36.

[9] Barbero M, Loisel R and Quézel P. *Biogeography, ecology and history of Mediterranean Quercus ilex ecosystems*. Plant Ecology. 1992;99-100(1):19-34.

[10] Corcuera L, Camarero J and Gil-Pelegrín E. *Effects of a severe drought on Quercus ilex radial growth and xylem anatomy.* Trees - Structure and Function. 2004;18(1):83-92.

[11] Gea-Izquierdo G, Cañellas I and Montero G. *Acorn production in Spanish holm oak woodlands*. *Invest Agrar:* Sist Recur For 2006;15:339-54.

[12] Pausas J, Bladé C, Valdecantos A, Seva J, Fuentes D, Alloza J, Vilagrosa A, Bautista S, Cortina J and Vallejo R. *Pines and oaks in the restoration of Mediterranean landscapes of Spain: New perspectives for an old practice — a review.* Plant Ecology. 2004;171(1):209-20.

[13] Soto A, Lorenzo Z and Gil L. *Differences in fine-scale genetic structure and dispersal in Quercus ilex L. and Q. suber L.: consequences for regeneration of mediterranean open woods*. Heredity. 2007;99(6):601-7.

[14] Herrera J. *Acorn predation and seedling production in a low-density population of cork oak (Quercus suber L.).* Forest Ecology and Management. 1995;76(1-3):197-201.

[15] Ogaya R and Penuelas J. *Contrasting foliar responses to drought in Quercus ilex and Phillyrea latifolia.* Biologia Plantarum. 2006;50(3):373-82.

[16] Sork VL. *Examination of Seed Dispersal and Survival in Red Oak, Quercus Rubra (Fagaceae), Using Metal-Tagged Acorns*. Ecology. 1984;65(3):1020-22.

[17] Villar-Salvador P, Planelles R, Enríquez E and Rubira JP. *Nursery cultivation regimes, plant functional attributes, and field performance relationships in the Mediterranean oak Quercus ilex L.* Forest Ecology and Management. 2004;196(2-3):257-66.

[18] Lumaret R, Mir C, Michaud H and Raynal V. *Phylogeographical variation of chloroplast DNA in holm oak (Quercus ilex L.).* Molecular Ecology. 2002;11(11):2327-36.

[19] Michaud H, Lumaret R and Romane F. *Variation in the genetic structure and reproductive biology of holm oak populations*. Plant Ecology. 1992;99-100(1):107-13.

[20] Petit RJ, Brewer S, Bordács S, Burg K, Cheddadi R, Coart E, Cottrell J, Csaikl UM, van Dam B, Deans JD, Espinel S, Fineschi S, Finkeldey R, Glaz I, Goicoechea PG, Jensen JS, König AO, Lowe AJ, Madsen SF, Mátyás G, Munro RC, Popescu F, Slade D, Tabbener H, de Vries SGM, Ziegenhagen B, de Beaulieu J-L and Kremer A. *Identification of refugia and post-glacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence*. Forest Ecology and Management. 2002;156(1-3):49-74.

[21] Abril N, Gion J-M, Kerner R, Müller-Starck G, Cerrillo RMN, Plomion C, Renaut J, Valledor L and Jorrin-Novo JV. *Proteomics research on forest trees, the most recalcitrant and orphan plant species*. Phytochemistry. 2011;72(10):1219-42.

[22] Confalonieri M, Balestrazzi A, Bisoffi S and Carbonera D. I*n vitro culture and genetic engineering of Populus spp.: synergy for forest tree improvement.* Plant Cell Tissue Org. Cult. 2003;72:109–38.

[23] Burgarella C, Lorenzo Z, Jabbour-Zahab R, Lumaret R, Guichoux E, Petit RJ, Soto A and Gil L. *Detection of hybrids in nature: application to oaks (Quercus suber and Q. ilex).* Heredity. 2009;102(5):442-52.

[24] Castro-Díez P, Villar-Salvador P, Pérez-Rontomé C, Maestro-Martínez M and Montserrat-Martí G. *Leaf morphology and leaf chemical composition in three Quercus*

(Fagaceae) species along a rainfall gradient in NE Spain. Trees - Structure and Function. 1997;11(3):127-34.

[25] Jimenez P, de Heredia UL, Collada C, Lorenzo Z and Gil L. *High variability of chloroplast DNA in three Mediterranean evergreen oaks indicates complex evolutionary history.* Heredity. 2004;93(5):510-5.

[26] Lumaret R and Jabbour-Zahab R. *Ancient and current gene flow between two distantly related Mediterranean oak species, Quercus suber and Q. ilex.* Annals of Botany. 2009;104(4):725-36.

[27] Lumaret R, Tryphon-Dionnet M, Michaud H, Sanuy A, Ipotesi E, Born C and Mir C. *Phylogeographical variation of chloroplast DNA in cork oak (Quercus suber).* Annals of Botany 2005;96(5):853-61.

[28] Toumi L and Lumaret R. *Allozyme characterisation of four Mediterranean evergreen oak species.* Biochemical Systematics and Ecology. 2001;29(8):799-817.

[29] Vakkari P, Blom A, Rusanen M, Raisio J and Toivonen H. *Genetic variability of fragmented stands of pedunculate oak (Quercus robur) in Finland.* Genetica. 2006;127(1-3):231-41.

[30] Ortego J, Bonal R and Muñoz A. *Genetic Consequences of Habitat Fragmentation in Long-Lived Tree Species: The Case of the Mediterranean Holm Oak (Quercus ilex, L.).* Journal of Heredity. 2010;101(6):717-26.

[31] Toumi L and Lumaret R. *Allozyme characterisation of four Mediterranean evergreen oak species.* Biochemical Systematics and Ecology 2001 29:799–817.

[32] Rodrıguez-Estevez V, Garcıa A, Perea J, Mata C and Gomez A. *Produccion de bellota en la dehesa: factores influyentes.* Arch Zootec. 2007;56:25-43.

[33] Gómez J. *Bigger is not always better: conflicting selective pressures on seed size in Quercus ilex.* Evolution 2004;58:71-80.

[34] Leiva MJ and Fernández-Alés R. *Variability in seedling water status during drought within a Quercus ilex subsp. ballota population, and its relation to seedling morphology.* Forest Ecology and Management. 1998;111(2-3):147-56.

[35] Ramírez-Valiente JA, Valladares F, Gil L and Aranda I. *Population differences in juvenile survival under increasing drought are mediated by seed size in cork oak (Quercus suber L.).* Forest Ecology and Management. 2009;257(8):1676-83.

[36] Singh B, Saklani K and Bhatt B. *Provenance variation in seed and seedlings attributes of Quercus glauca Thunb. in Garhwal Himalaya, India.* Dendrobiology 2010;63:59–63.

[37] Díaz M, Møller AP and Pulido FJ. *Fruit abortion, developmental selection and developmental stability in Quercus ilex.* Oecologia. 2003;135(3):378-85.

[38] Özcan T. Total Protein and Amino acid *Compositions in the Acorns of Turkish Quercus L. Taxa.* Genetic Resources and Crop Evolution. 2006;53(2):419-29.

[39] Rafii ZA, Dodd RS and Pelleau Y. *Biochemical diversity and systematics of Mediterranean evergreen oak from South East France.* Biochemical Systematics and Ecology. 1993;21(6-7):687-94.

[40] Rafii ZA, Zavarin E and Pelleau Y. *Chemosystematic differentiation of Quercus ilex and Q. rotundifolia based on acorn fatty acids.* Biochemical Systematics and Ecology. 1991;19(2):163-66.

[41] Cantos E, Espín JC, López-Bote C, de la Hoz L, Ordóñez JA and Tomás-Barberán FA. *Phenolic compounds and fatty acids from acorns (Quercus spp.), the main dietary*

constituent of free-ranged Iberian pigs. Journal of Agricultural and Food Chemistry. 2003;51(21):6248-55.

[42] Dodd RS, Rafii ZA and Bojovic S. *Chemosystematic study of hybridization in californian live oak: Acorn steroids*. Biochemical Systematics and Ecology. 1993;21(4):467-73.

[43] Charef M, Yousfi M, Saidi M and Stocker P. *Determination of the fatty acid composition of acorn (Quercus), Pistacia lentiscus seeds growing in Algeria.* Journal of the American Oil Chemists' Society. 2008;85(10):921-24.

[44] Daza A, López-Bote CJ, Tomás Barberán FA, Espin JC, López Carrasco C, Olivares A and Rey AI. *Effect of mediterranean forest parasite with Curculio sp. on nutritional value of acorn for Iberian pig feeding and fat characteristics*. Meat Science. 2007;76(2):316-20.

[45] Rodríguez-Estévez V, García A and Gómez AG. *Characteristics of the acorns selected by free range Iberian pigs during the montanera season.* Livestock Science. 2009;122(2-3):169-76.

[46] Fernández I, Gómez A, Moreno P, Fernández P, De Pedro E, Díaz E and Sánchez L. *Characteristics of the acorns of oak (Quercus ilex) in different zones from the valley of the pedroches (Córdoba).* Bromatologic inventory. Producciones agroganaderas: Gestión eficiente y conservación del medio natural. XLV RC de la SEEP. Gijón, Spain. 2005;1:pp. 377-85.

[47] Wilkins MR, Pasquali C, Appel RD, Ou K, Golaz O, Sanchez J-C, Yan JX, Gooley AA, Hughes G, Humphery-Smith I, Williams KL and Hochstrasser DF. *From proteins to proteomes: Large scale protein identification by two-dimensional electrophoresis and arnino acid analysis*. Nat Biotech. 1996;14(1):61-65.

[48] Jorrín-Novo JV, Maldonado AM, Echevarría-Zomeño S, Valledor L, Castillejo MA, Curto M, Valero J, Sghaier B, Donoso G and Redondo I. *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. 2009;72(3):285-314.

[49] Cánovas FM, Dumas-Gaudot E, Recorbet G, Jorrin J, Mock H-P and Rossignol M. *Plant proteome analysis*. Proteomics. 2004;4(2):285-98.

[50] Jorrin JV, Maldonado AM and Castillejo MA. *Plant proteome analysis: a 2006 update.* Proteomics. 2007;7(16):2947-62.

[51] Maldonado AM, Echevarría-Zomeño S, Jean-Baptiste S, Hernández M and Jorrín-Novo JV. *Evaluation of three different protocols of protein extraction for Arabidopsis thaliana leaf proteome analysis by two-dimensional electrophoresis*. Journal of Proteomics. 2008;71(4):461-72.

[52] Bahrman N and Petit RJ. *Genetic polymorphism in maritime pine (Pinus pinaster Ait.) assessed by two-dimensional gel electrophoresis of needle, bud, and pollen proteins*. Journal of Molecular Evolution. 1995;41(2):231-37.

[53] Basha SM. *Identification of Cultivar Differences in Seed Polypeptide Composition of Peanuts (Arachis hypogaea L.) by Two-Dimensional Polyacrylamide Gel Electrophoresis*. Plant Physiology. 1979;63(2):301-6.

[54] Chevalier F, Martin O, Rofidal V, Devauchelle AD, Barteau S, Sommerer N and Rossignol M. *Proteomic investigation of natural variation between Arabidopsis ecotypes*. Proteomics. 2004;4(5):1372-81.

[55] Emre İ. *Electrophoretic analysis of some Lathyrus L. species based on seed storage proteins*. Genetic Resources and Crop Evolution. 2009;56(1):31-38.

[56] Jacobsen S, Nesic L, Petersen M and Sondergaard I. *Classification of wheat varieties: use of two-dimensional gel electrophoresis for varieties that can not be classified by matrix assisted laser desorpiton/ionization-time of flight-mass spectrometry and an artificial neural network.* Electrophoresis. 2001;22(6):1242-5.

[57] Ostergaard O, Melchior S, Roepstorff P and Svensson B. *Initial proteome analysis of mature barley seeds and malt.* Proteomics. 2002;2(6):733-9.

[58] Singh N and Matta N. *Variation studies on seed storage proteins and phylogenetics of the genus Cucumis*. Plant Systematics and Evolution. 2008;275(3):209-18.

[59] Yüzbaşioğlu E, Açik L and Özcan S. *Seed protein diversity among lentil cultivars*. Biologia Plantarum. 2008;52(1):126-28.

[60] Petit RJ, Bahrman N and Baradat PH. *Comparison of genetic differentiation in maritime pine (Pinus pinaster Ait.) estimated using isozyme, total protein and terpenic loci.* Heredity. 1995;75(4):382-89.

[61] Vienne DD, Burstin J, Gerber S, Leonardi A, Guilloux ML, Murigneux A, Beckert M, Bahrman N, Damerval C and Zivy M. *Two-dimensional electrophoresis of proteins as a source of monogenic and codominant markers for population genetics and mapping the expressed genome*. Heredity. 1996;76(2):166-77.

[62] Bonhomme L, Monclus R, Vincent D, Carpin S, Claverol S, Lomenech AM, Labas V, Plomion C, Brignolas F and Morabito D. *Genetic variation and drought response in two Populus x euramericana genotypes through 2-DE proteomic analysis of leaves from field and glasshouse cultivated plants*. Phytochemistry. 2009;70(8):988-1002.

[63] Bonhomme L, Monclus R, Vincent D, Carpin S, Lomenech AM, Plomion C, Brignolas F and Morabito D. *Leaf proteome analysis of eight Populus xeuramericana genotypes: genetic variation in drought response and in water-use efficiency involves photosynthesis-related proteins*. Proteomics. 2009;9(17):4121-42.

[64] Crescente MF, Gratani L and Larcher W. *Shoot growth efficiency and production of Quercus ilex L. in different climates*. Flora. 2002;197(1):2-9.

[65] Fotelli M, Radoglou K and Constantinidou H. *Water stress responses of seedlings of four Mediterranean oak species*. Tree Physiology. 2000;20(16):1065-75.

[66] Hódar JA. *Leaf fluctuating asymmetry of Holm oak in response to drought under contrasting climatic conditions*. Journal of Arid Environments. 2002;52(2):233-43.

[67] Ponton S, Dupouey J-L, Bréda N and Dreyer E. *Comparison of water-use efficiency of seedlings from two sympatric oak species: genotype × environment interactions*. Tree Physiology. 2002;22(6):413-22.

[68] Sergeant K, Spieß N, Renaut J, Wilhelm E and Hausman JF. *One dry summer: A leaf proteome study on the response of oak to drought exposure*. Journal of Proteomics. 2011;74(8):1385-95.

[69] Turco E, Close TJ, Fenton RD and Ragazzi A. *Synthesis of dehydrin-like proteins in Quercus ilex L. and Quercus cerris L. seedlings subjected to water stress and infection*

with Phytophthora cinnamomi. Physiological and Molecular Plant Pathology. 2004;65(3):137-44.

Estudio de la respuesta al estrés hídrico en dos poblaciones de encina (*Quercus ilex* subsp. *ballota* (Desf.) Samp.) mediante una aproximación de proteómica comparativa basada en electroforesis bidimensional

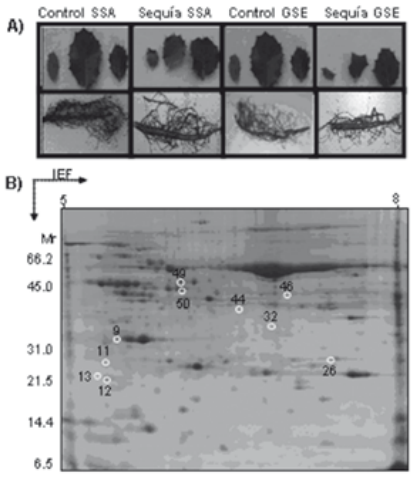
José Valero Galván[1], Rafael Mª Navarro Cerrillo[2], Mª Cristina Romero Rodríguez[1], David Ariza[2], Jesús Jorrín Novo[1]

[1] Grupo de Bioquímica, Proteómica Vegetal y Agrícola. Departamento de Bioquímica y Biología Molecular. Córdoba, España. [2] Departamento de Ingeniería Forestal. ETSIAM, Universidad de Córdoba. Córdoba, España

Resumen

La encina *(Quercus ilex subsp. ballota)* es una de las especies forestales más importantes del Bosque Mediterráneo, siendo ampliamente utilizada en programas de conservación forestal, reforestación y otras practicas silvícolas. El éxito de dichas prácticas depende, en gran medida, de la supervivencia de las plántulas, siendo el estrés por sequía la principal causa de mortalidad tras el transplante. Es por ello que la selección de "individuos plus" con tolerancia a estrés hídrico es de gran importancia. Los mecanismos de respuesta a estrés en encina y su variabilidad poblacional han sido muy poco estudiados. En el presente trabajo se ha evaluado, mediante una aproximación de proteómica comparativa basada en geles bidimensionales, la respuesta diferencial de dos poblaciones de encina a la sequía. Aunque dicha aproximación revela diferencias en la intensidad de manchas entre tratamientos, no lo hace entre poblaciones, sin que, por otro lado, hayamos podido asociar las diferencias observadas al carácter más o menos tolerante de la población utilizada. Ello puede ser debido a la intensidad y duración del estrés, a las dificultades del sistema experimental o limitaciones de la técnica.

Nuestro grupo de investigación esta trabajando en un proyecto de investigación (Variabilidad, catalogación, respuesta a estreses y propagación clonal de encina (*Q. ilex*), DECOVA, AGL2009-12243-C02-02), dirigido, entre otros objetivos, a estudiar la respuesta de la encina a estrés producido por sequía utilizando una aproximación multidisciplinar, proteómica incluida [1, 2]. La proteómica en especies forestales no es fácil, debido a su dificultad como sistema experimental (variabilidad genética alta, ciclo de vida largo, ausencia de entradas en bases de datos de DNA genómico, ESTs o proteínas). No obstante, su estudio esta justificado dada la impor-



Numero de identificación	Nombre de la proteína
Metabolismo energético	
44	ATP sintetasa, Cloroplástica
9	Complejo productor de oxígeno
11	LHCII tipo I
Glicólisis	
26	Triosafosfato isomerasa
46	Fosfoglicerato quinasa, Citosólica
Proteínas de respuesta a estrés	
12	2-cys peroxiredoxina basal 1
13	Peroxiredoxina
32	β-1,3-glucanasa
Metabolismo de los aminoácidos	
49	S-adenosil-metionina sintetasa
50	Glutamina sintetasa

Figura 1. *A) Imagen de hojas y raíz de plantas de las dos poblaciones utilizadas a los 28 días del tratamiento de sequía. SSA (población almeriense), GSE (población sevillana). B) Gel 2-DE representativo de extractos de hoja de encina. La imagen corresponde a una muestra de la población SSA a los 28 días del tratamiento de sequía. En la tabla se indican las proteínas identificadas correspondientes a manchas diferenciales entre tratamientos.*

tancia económica y medioambiental de la especie, componente principal de la "dehesa" y el Bosque Mediterráneo. Mediante una aproximación de proteómica de expresión diferencial basada en electroforesis bidimensional se han analizados cambios en el perfil proteico de dos poblaciones de encina [Almería (SSA) y Sevilla (GSE)] [3] sometidas a estrés hídrico mediante riego deficitario (28 días). Previamente se caracterizó la respuesta de las dos poblaciones mediante parámetros de crecimiento, potencial hídrico, humedad relativa en raíz y hojas, y eficiencia fotosintética (resultados no publicados). A tenor de los resultados obtenidos se concluyó que la población SSA fue más tolerante, mientras que la población GSE fue la más susceptible (Figura 1A).

Las proteínas de hoja (300 mg) de plantas de un año de crecimiento (control, plantas regadas cada tres días; tratamiento de sequía, plantas no regadas a los 28 días) fueron extraídas usando el protocolo de precipitación TCA-FENOL [4]. Las condiciones de electroforesis, tinción, análisis de los geles y estadístico e identificación de proteínas se realizaron como lo indica Maldonado et al. [4].

Después de la tinción de los geles y análisis de las imágenes se detectaron alrededor de 350-370 manchas. Tras el análisis estadístico se concluyo que existían 40 manchas diferenciales entre tratamientos y tan solo 4 entre poblaciones (muestras control). Dado el carácter de especie huérfana de encina, el porcentaje de identificación fue bajo, menos del 50 % de las manchas diferenciales. En general, se observó como consecuencia del tratamiento de sequía un descenso en la intensidad de manchas correspondientes a proteínas fotosintéticas, fosfoglicerato-quinasa (glicólisis), peroxiredoxinas (estrés oxidativo) y una glucanasa (proteína de defensa frente a patógenos) (Figura 1B). Por el contrario se observó un aumento de intensidad de la glutamina sintetasa y la s-adenosil-metionina sintetasa. No se encontraron diferencias estadísticamente significativas entre las dos poblaciones analizadas, tanto en controles como en tratamiento de sequía.

Estos resultados son similares a los obtenidos en plántulas de encina sometidas a diferentes niveles de estés hídrico [1, 2] y a otras especies forestales como el caso del genero Populus [5], sin que hasta la fecha, y en especies forestales, se haya concluido de forma contundente la base molecular de la respuesta a estrés, y si las consecuencias del estrés.

Referencias

[1] Jorge I, Navarro-Cerrillo RM, Lenz C, Ariza D, Jorrín J, 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.

[2] Echevarría-Zomeño S, Ariza D, Jorge I, Lenzc C, Campo A, Jorrín J, Navarro-Cerrillo, RM. Changes in the protein profile of *Quercus ilex* leaves in response to drought tress and in response to drought stress and recovery. J. Plant Physiol.2009; 166:233–245.

[3] Valero GJ, Navarro-Cerrillo RM, Gómez CA, Ariza D, Jorrín J. Population variability and mother tree selection in holm oak (*Quercus ilex* L. subsp. *ballota* [Desf.] Samp.) based on acorn morphometry and analysis by near infrared reflectance spectroscopy (NIRS). 2009, (Submited).

[4] Maldonado AM, Echevarria-Zomeno S, Jean-Baptiste S, Hernandez M, Jorrin-Novo JV. Evaluation of three different protocols of protein extraction for *Arabidopsis thaliana* leaf proteome analysis by two-dimensional electrophoresis. J. Proteomics, 2008; 71:461–72.

[5] Bonhomme L, Monclus R, Vincent D, Carpin S, Claverol S, Lomenech MA, Labas V, Plomion C, Brignolas F, Morabito D. Genetic variation and drought response in two *Populus euramericana* genotypes through 2-DE proteomic analysis of leaves from field and glasshouse cultivated plants. Phytochemistry, 2009; 70: 988–1002.

JOURNAL OF PROTEOMICS 72 (2009) 285–314

available at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/jprot

Review

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

Jesús V. Jorrín-Novo[a,*], *Ana M. Maldonado*[a], *Sira Echevarría-Zomeño*[a], *Luis Valledor*[b], *Mari A. Castillejo*[a], *Miguel Curto*[a], *José Valero*[a], *Besma Sghaier*[a], *Gabriel Donoso*[a], *Inmaculada Redondo*[a]

[a]*Agricultural and Plant Biochemistry and Proteomics Research Group, Dept. of Biochemistry and Molecular Biology, University of Córdoba, Córdoba, Spain*

[b]*EPIPHYSAGE Research Group, Área de Fisiología Vegetal, Departamento B.O.S., Universidad de Oviedo, Oviedo, Spain*

ARTICLE DATA

Keywords:
Subcellular Proteomics
Differential Expression Proteomics
Posttranslational Modifications
Interactomics

ABSTRACT

This review is the continuation of three previously published articles [Jorrin JV, Maldonado AM, Castillejo MA. Plant proteome analysis: a 2006 update. Proteomics 2007; 7: 2947–2962; Rossignol M, Peltier JB, Mock HP, Matros A, Maldonado AM, Jorrin JV. Plant proteome analysis: a 2004–2006 update. Proteomics 2006; 6: 5529–5548; Canovas FM, Dumas-Gaudot E, Recorbet G, Jorrin J, Mock HP, Rossignol M. Plant proteome analysis. Proteomics 2004; 4: 285–298] and aims to update the contribution of Proteomics to plant research between 2007 and September 2008 by reviewing most of the papers, which number approximately 250, that appeared in the Plant Proteomics field during that period. Most of the papers published deal with the proteome of *Arabidopsis thaliana* and rice (*Oryza sativa*), and focus on profiling organs, tissues, cells or subcellular proteomes, and studying developmental processes and responses to biotic and abiotic stresses using a differential expression strategy. Although the platform based on 2-DE is still the most commonly used, the use of gel-free and second-generation Quantitative Proteomic techniques has increased. Proteomic data are beginning to be validated using complementary -omics or classical biochemical or cellular biology techniques. In addition, appropriate experimental design and statistical analysis are being carried out in accordance with the required Minimal Information about a Proteomic Experiment (MIAPE) standards. As a result, the coverage of the plant cell proteome and the plant biology knowledge is increasing. Compared to human and yeast systems, however, plant biology research has yet to exploit fully the potential of proteomics, in particular its applications to PTMs and Interactomics.

* *Corresponding author*. Fax: +34957218439.
E-mail address: bf1jonoj@uco.es (J.V. Jorrín-Novo).

1874-3919/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jprot.2009.01.026

Contents

1. Introduction

1.1. *Where is (Plant) Proteomics and where should it be?*

"How", "where", "when", and "what for" are the several hundred thousand of individual protein species produced in a living organism? How do they interact with one another and with other molecules to construct the cellular building? How do they work in order to fit in with programmed growth and development, and to interact with their biotic and abiotic environment? To answer all of these questions is the objective, first, of Proteomics, and ultimately of Systems Biology [4].

In order to approach the above questions experimentally, continuous improvements in techniques and protocols for High-throughput Proteomics are being made at all workflow stages, starting from the laboratory (tissue and cell fractionation, protein extraction, depletion, purification, separation, MS analysis) and ending at the computer (algorithms for protein identification and bioinformatics tools for data analysis, databases and repositories). Since no single approach can fully unravel the complexity of living organisms, and experimental results need to be validated, Proteomics, like any methodology, should be considered as part of a multidisciplinary integrative analysis at different levels, extending from the gene to the phenotype through proteins. This analysis should involve -omics (Genomics, Transcriptomics, Proteomics, Metabolomics) as well as classical biochemical and cellular biology techniques [5–7]. Integrative transcriptomic and proteomic studies, and depending on the biological system, may or may not show good correlationship between the level of protein and mRNA. Biological or methodological explanations are normally forwarded to explain such discrepancies [8–11].

Despite the technological achievements in Proteomics, only a tiny fraction of the cell proteome has been characterized so far, and only for a few biological systems (human, fruit fly, *Arabidopsis*, rice). Even for these organisms, the function of quite a number of proteins remains to be investigated [12]. Proteomic techniques have a number of limitations, such as sensitivity, resolution and speed of data capture. They also face a number of

challenges, such as deeper proteome coverage, Proteomics of unsequenced "orphan" organisms [13], Top-down Proteomics [14], protein quantitation [15], PTMs and Interactomics. Most of these limitations and challenges reflect the difficulty of working with the biological diversity of proteins and their range of physicochemical properties.

One of the unanswered biological questions is to determine the number of protein species that can be translated from a single gene as a result of alternative splicing or PTMs, and the biological meaning of each one [16,17]. The percentage of plant genes with at least one alternative transcript [18] is estimated to be 20%, and while this figure is likely to be an underestimate [19], it is lower than that estimated for humans. An example of alternative splicing in plants is represented by the pre-mRNAs of *Arabidopsis thaliana* genes that encode serine/arginine-rich proteins, a conserved family of splicing regulators in eukaryotes [20]. Remarkably, up to 95 transcripts are produced from only 15 genes, resulting in a six-fold increase of the transcriptome. Sequence analysis of the splice variants reveals that the proteins predicted for most of these variants either lack one or more modular domains or contain truncated domains. The alternative splicing of Ser/Arg-rich genes is controlled in a developmental and tissue-specific manner, and is affected by various hormones and abiotic stresses.

Alternative transcripts arise with the use of multiple transcription start sites that produce in frame protein species differing only in their N-terminal sequence. This phenomenon has been observed for *A. thaliana* glutathione S-transferase GSTF8 [21]. Analysis of GSTF8-L and GSTF8-S proteins showed that the former is targeted to plastids, whereas the latter is cytoplasmic. *In silico* analyses revealed that GSTF8-S is conserved across a wide range of plants, while GSTF8-L is confined to the Brassicaceae. These studies support the fact that alternative translation start sites of the GSTF8 promoter are used to create tissue-specific and stress-responsive expression patterns as well as target the same protein to two different subcellular localisations. Another example of multiple transcription start sites is represented by two genes of *Arabidopsis* encoding the L-isoaspartate methyltransferase enzymes PIMT1 and 2 (EC 2.1.1.77). PIMT2 produces at least eight transcripts depending on which of the four potential start sites is used; this results in three different initiating methionines, and two possibilities, 5′ and 3′, for alternative splice site selection of the first intron [22]. The entire array of transcripts produces mature proteins capable of converting L-isoaspartate to L-aspartate using small peptide substrates. However, the location of each of the corresponding protein forms–nucleus, cytoplasm, endo-membrane system, chloroplasts, or mitochondria–depends on the transcript from which they are produced.

Proteomics has experienced an explosion of new protocols, platforms and workflows; some of them are quite complex and require sophisticated equipment and expertise, with each technique has its own features, advantages and limitations. A huge amount of data is being generated, and some of it is being deposited and organised in several databases available to the scientific community: the PPDB, http://ppdb.tc.cornell.edu [23]; the PODB, http://proteome.dc.affrc.go.jp/Soybean/; the Organellome, http://podb.nibb.ac.jp/Organellome [24]; the UniProt knowledgebase reported in this issue by Schneider et al.; and others mentioned throughout the manuscript (Table 1).

Exploiting the full potential of Proteomics surpasses the possibilities of individual laboratories and requires large-scale open collaborations [25,26] such as the Human Proteome Organization (HUPO) (http://www.hupo.org/). Along these lines, the multinational coordinated project called the "Green Proteome" is one of the major challenges for Plant Proteomics in the near future. In this respect, the Proteomics Subcommittee of the Multinational *Arabidopsis* Steering Committee has begun to operate; this subcommittee was created to help to coordinate International Proteomics Research in *Arabidopsis* (http://www.masc-Proteomics.org/) [27]. One of its most important benefits has been the establishment of a searchable database of MS/MS reference spectra derived from *Arabidopsis*, *Chlamydomonas reinhardtii*, *Medicago truncatula*, *Solanum tuberosum* (potato), *Solanum lycopersicum* (tomato) and other plants (ProMEX; http://promex.mpimp-golm.mpg.de/home.shtml) [28]. This database allows reliable protein identification through a genome-independent approach, since newly generated MS/MS spectra can be matched against previous experimental MS/MS spectra. This approach allows semiquantitative analysis at the same time as spectrum matching. Other initiatives that deserve to be mentioned is the Plant Proteomics in Europe" (COST Action FA0603).

After nearly 10 years of Proteomics research [29], it is possible to look back at previous research and publications, identifying errors, which also allows us to avoid repeating them in new studies. These errors come from the experimental design, the analysis and the interpretation of the data [30]. In addition, data validation should always aspire to more than merely description or speculation. It is not uncommon to find low-confidence protein identification in the literature, especially in the case of unsequenced organisms; and inappropriate statistical analyses of results have often been performed. It is interesting to see how many manuscripts contain the term "Proteome" when probably only a tiny fraction of the total proteome has been analysed. Along these lines, HUPO's Proteomic Standard Initiative has developed guidance modules [31] that have been translated into Minimal Information about a Proteomic Experiment (MIAPE) documents. The MIAPE documents recommend proteomic techniques that should be considered and followed when conducting a proteomic experiment. Proteomics journals should be, and in fact are, extremely strict when recommending that investigators follow the MIAPE standards for publishing a proteomic experiment.

The huge amount of data generated in a proteomic experiment has led many journals to recommend, and some to require, that original data be submitted to public repositories. The repository options available and analysis of the data in the pipeline are presented in the review by Mead et al. [32], which discusses recent papers in greater detail [33]. A shift in the protein identification paradigm is currently underway, moving from sequencing and database searching to spectrum searching in spectral libraries. This underscores the importance of repositories for Proteomics [28,34,35].

Different areas within Proteomics can be defined: i) Descriptive Proteomics, including Subcellular Proteomics; ii) Differential Expression Proteomics; iii) Posttranslational

Table 1 – Useful online resources and Plant Proteome Databases.

Web site	Name/description
Genome databases	
[a]http://www.ncbi.nlm.nih.gov/	National Center for Biotechnology Information
[a]http://www.ncbi.nlm.nih.gov/Genbank	NIH genetic sequence database
[a]http://compbio.dfci.harvard.edu/tgi/plant.html	The gene index project. The Computational Biology and Functional Genomics Laboratory, and the Dana-Farber Institute and Public School of Public Health
http://www.arabidopsis.org/	TAIR, The *Arabidopsis* Information Resource
[a]http://www.legoo.org/	Legoo, a bioinformatics gateway towards integrative legume biology
[a]http://medicago.toulouse.inra.fr/Mt/EST/	MENS, *Medicago* EST Navigation System
[a]http://www.medicago.org/genome/	*Medicago truncatula* sequencing project
[a]http://www.sgn.cornell.edu/about/tomato_sequencing.pl	International Tomato Sequencing Project
[a]http://www.maizesequence.org/index.html	Maize Genomes Sequencing Project
[a]http://www.potatogenome.net	The Potato Genome Sequencing Consortium
[a]http://www.phytozome.net/sorghum	The Sorghum bicolor Genome Project
http://pgrc.ipk-gatersleben.de/cr-est/index.php	The IPK Crop EST Database
http://chloroplast.cbio.psu.edu/	A database for fully sequenced plastid genomes
http://genomics.msu.edu/plant_specific/	Plant Specific Database
http://www.plantgdb.org/	The Plant Genome Database (GDB)
http://rice.plantbiology.msu.edu	The Rice Genome Annotation Project
http://mips.gsf.de/proj/plant/jsf/rice/index.jsp	MosDB, the MIPS *Oryza sativa* database
http://ricegaas.dna.affrc.go.jp/	RiceGAAS, Rice Genome Automated Annotation System
http://cgf.ucdavis.edu/home/	CGF, College of Agricultural and Environmental Sciences. Genomics Facilities
http://harvest.ucr.edu/	HarvEST, EST database-viewing software. Oriented to comparative genomics
Proteome databases	
http://ca.expasy.org/	The ExPASy (Expert Protein Analysis System) proteomics server of the Swiss Institute of Bioinformatics (SIB) Analysis of protein sequences, structures and 2-D PAGE
[a]http://www.hupo.org/	HUPO, Human Proteome Organization
http://mips.gsf.de/	MIPS, Munich Information Center for Protein Sequences
http://www.ebi.ac.uk/pride/	The PRIDE, PRoteomics IDEntifications Database. EMBL-EBI (European Bioinformatic Institute)
http://www.ebi.ac.uk/integr8/EBI-Integr8-HomePage.do	Integr8, Integrated information about deciphered genomes and their corresponding proteomes. EMBL-EBI (European Bioinformatic Institute)
http://organelledb.lsi.umich.edu/	Organelle DB, A database of organelle proteins, and subcellular structures/complexes
http://www.compbio.dundee.ac.uk/SNAPPI/downloads.jsp	SNAPPIVIEW (Structures,iNterfaces and Alignments for Protein–Protein Interactions)
http://cbm.bio.uniroma2.it/phospho3d/	Phospho3.Database of three-dimensional structures of phosphorylation sites
http://www.cs.ualberta.ca/~bioinfo/PA/GOSUB/	Proteome Analyst PA-GOSUB 2.5. Sequences, predicted GO molecular functions and subcellular localisations
http://www.rcsb.org/pdb/home/home.do	RCSB, The Research Collaboratory for Structural Bioinformatics. Protein Database (PDB)
http://wwwmgs.bionet.nsc.ru/mgs/gnw/pdbsite/	PDB-Site. Comprehensive structural and functional information on PTMs, catalytic active sites, ligand binding, protein–protein, protein–DNA and protein–RNA interactions) in the Protein Data Bank (PDB)
http://wolfpsort.org/	WoLF PSORT, Protein Subcellular Localization Prediction
http://www.rostlab.org/db/NMPdb/	NMPdb, Nuclear Matrix Associated Proteins

JOURNAL OF PROTEOMICS 72 (2009) 285–314

http://www.cbs.dtu.dk/services/TargetP/	TargetP, predicts the subcellular location of eukaryotic proteins, based on the predicted presence of the N-terminal presequences
http://www.cbs.dtu.dk/services/SecretomeP/	The SecretomeP, Predictions of protein secretion and information on various PTMs and localizational localisational aspects of the protein
http://www.yass.sdu.dk/	VEMS, Virtual Expert Mass Spectrometrist. Program for integrated proteome analysis.
http://www.cbs.dtu.dk/services/NetPhos/	The NetPhos server produces neural network predictions for serine, threonine and tyrosine phosphorylation sites in eukaryotic proteins
Plant Proteome Databases	
[a]http://ca.expasy.org/sprot/ppap/ppap_stat.html	PPAP, the UniProtKB/Swiss-Prot Plant Proteome Annotation Program
[a]http://www.costfa0603.com/index.php	Plant Proteomics in Europe, COST F0603
[a]http://www.masc-proteomics.org/	The Multinational *Arabidopsis* Steering Committee, Proteomics Subcommittee
Standard for sample handling and data interpretation	
[a]http://fgcz-atproteome.unizh.ch/index.php	*Arabidopsis* proteome database
http://proteomics.arabidopsis.info/	NASC Proteomics database for *Arabidopsis* data. Proteomics experiments and their related data
http://gabi.rzpd.de/projects/Arabidopsis_Proteomics/	Gabipd, Proteomic data. *Arabidopsis thaliana* and *Brassica napus*
[a]http://ppdb.tc.cornell.edu	The PPDB (Plant Proteome DataBase) for *Arabidopsis thaliana* and maize
[a]http://proteome.dc.affrc.go.jp/Soybean/	Soybean proteome Database (2-DE)
http://www.noble.org/medicago/proteomics.html	Center for *Medicago* Genomics Research: Proteomics
http://gene64.dna.affrc.go.jp/RPD/	Rice Proteome Database (2-DE)
Plant Proteome Specialized databases	
[a]http://promex.mpimp-golm.mpg.de/home.shtml	ProMEX (Protein Mass spectra Extraction) MSMS spectral database
[a]http://podb.nibb.ac.jp/Organellome	PODB2, The Plant Organelles Database, 2
[a]www.suba.bcs.uwa.edu.au	SUBA, Subcelllular location of *Arabidopsis* proteins
http://www.araperox.uni-goettingen.de/	AraPerox, Database of Putative Proteins of *Arabidopsis* Peroxisomes
http://andrewschein.com/pclr/index.html	PCLR, Chloroplast Localization Prediction
http://www.cbs.dtu.dk/services/ChloroP/	The ChloroP server predicts the presence of chloroplast transit peptides (cTP) in and the location of potential cTP cleavage sites
http://www.gartenbau.uni-hannover.de/genetik/AMPP	The *Arabidopsis* Mitochondrial Proteome Project
http://www.plantenergy.uwa.edu.au/ampdb/	*Arabidopsis* Mitochondrial Protein Database
http://www.plprot.ethz.ch/	plprot — a plastid protein database
http://www.seed-proteome.com/	*Arabidopsis* Seed proteome
http://aramemnon.botanik.uni-koeln.de/	Aramemnon, Plant membrane protein databases
http://plantrbp.uoregon.edu/	Putative Orthologous Groups and Plant RNA Binding Protein Database
http://www.coiled-coil.org/arabidopsis/index.html	ARABI-COIL, *Arabidopsis* Coiled-Coil Protein Database
http://bioinf.scri.sari.ac.uk/cgi-bin/atnopdb/home	AtNoPDB, *Arabidopsis* Nucleolar Protein Database
[a]http://phosphat.mpimp-golm.mpg.de/	*Arabidopsis* phosphoproteome database
http://digbio.missouri.edu/p3db/	P3DB, Plant protein phosphorilation phosphorylation database

[a] Cited in the text.

Modifications; iv) Interactomics; and v) Proteinomics (targeted or hypothesis-driven Proteomics). In plants, PTMs–with the exception of the phosphoproteome [36]–and Interactomics remain the main challenges.

Deeper proteome coverage has become possible by combining electrophoresis and liquid chromatography techniques, resulting in the report of nearly 50% of all *Arabidopsis* predicted gene models, including some not represented in The *Arabidopsis* Information Resource (TAIR) (AtProteome; http://fgcz-atproteome.unizh.ch/index.php) [37]. The combination of subcellular fractionation techniques with the use of modern mass spectrometry equipment has allowed extensive characterization of the plant subcellular proteome, which has led to the discovery of new metabolic pathways [38]. An expanded update of the subcellular localisation of *Arabidopsis* proteins has been carried out (SUBA; www.suba.bcs.uwa.edu.au) [39], and the most comprehensive study on the *Arabidopsis* chloroplast proteome has been published, including information on sorting signals, PTMs and protein abundance. This information is accessible at the PPDB (http://ppdb.tc.cornell.edu) [40]. Despite these advances, some recalcitrant proteomes, as is the case with membrane proteins or other highly hydrophobic proteins, remain elusive, and protein trafficking is still an important unresolved issue [41,42]. In fact, some proteins have been identified at the "apparently wrong" location [43].

Plant Proteomics is beginning to make some practical contributions to applied fields including biomedicine, through the identification and characterization of allergens [44; Fasoli et al., this issue]; agronomy, through studies of the equivalence of transgenic crops [45], genotyping [46,47], studies of heterosis [48] and the environment [45]; and food science, through studies of food quality control and traceability [49].

Knowledge at the molecular level on relevant plant-derived food allergens is scanty, and their detailed and comprehensive characterization can be carried out using proteomics. This is the case of three papers that appeared during the period reviewed; these papers analyse allergens from wheat flour [50], maize pollen [51] and olive tree pollen [52].

The proteomic and transcriptomic analysis of seeds from wheat transgenic plants overexpressing a low-molecular-weight glutenin subunit by Scossa et al. [53] revealed that an increase in the amount of transgenic protein is accompanied by a reduction in the endogenous levels of the glutenin subunit, all subclasses of gliadins and metabolic enzymes, as well as chloroform/methanol-soluble proteins. These results support the hypothesis that a global compensatory metabolic response is occurring. Similarly, at least three additional papers have reported changes in the protein profile as a result of transgenesis, including maize [54], grape [55] and tobacco [56].

Proteomics could be a good complementary approach to molecular breeding, as shown for *Brassica napus* [57], grape [58], sunflower [47] and strawberry [46]. The genetic diversity among 10 Iranian bread wheat (*Triticum aestivum*) genotypes has been analysed by testing 12 quality traits, 320 amplified fragment length polymorphisms (AFLP), 491 simple sequence repeat (SSR) alleles and 294 proteome markers. The average genetic diversity based on quality traits (0.684 with a range of 0.266–0.997) is higher than that based on AFLP (0.502 with a range of 0.328–0.717), SSR (0.503 with a range of 0.409–0.595) and proteome markers (0.464 with a range of 0.264–0.870) [59].

This review is a continuation of three ones previously published [1–3] and aims to update the contribution of Proteomics to plant research during 2007–September 2008 by reviewing most of the papers that appeared on this field throughout this period. We have tried not to be repetitive with recently published reviews (Table 2), and instead we have sought to emphasise new contributions or improvements published since the three previous review periods. Two Plant Proteomics books have recently appeared covering the literature up to 2006, with chapters devoted to technologies, including bioinformatics tools, as well as discussions of Descriptive/Expression Proteomics studies in *Arabidopsis*, rice, legumes, cereals, oilseed crops, Subcellular Proteomics, PTMs, Interactomics, responses to stresses and Structural Proteomics [60,61]. In addition, an excellent and extensive review on Rice Proteomics by Agrawal, Jwa, and Rakwal has recently appeared in Proteomics [62]. In this issue two reviews on Crop Proteomics appear, by Komatsu and Ashan on soybean, and by Finnie and Svensson on barley. Because of their novelty and originality, the reviews by Weckerth [5] and Farrokhi et al. [63] should be mentioned. The former covers methodological aspects of Metabolomics and Proteomics, including statistical analysis and data integration. It reports a strategy combining metabolic and protein profiling with multivariate exploratory data mining. The latter review introduces the field of peptidomics and its application to the study of plant peptides. This field, developed in the last few years, has been used primarily to study neuropeptides in animals and protease degradation products. A number of plant peptides that are both extracellular and intracellular and that involve a number of functions, including signalling and defence, have been identified and characterized through biochemical and genetic studies. Nevertheless, a systematic throughput analysis using Proteomics has so far not been reported.

2. Publications for period from 2007 to September 2008

2.1. *Plant Proteomics: still trailing behind Human and Yeast Proteomics*

The use of Proteomics in plant biology research has increased significantly in the period from 2007 to September 2008, becoming a routine methodology in a number of plant laboratories worldwide. During this period, both qualitative and quantitative improvements in Plant Proteomics have occurred, ushering in a new phase of "Second-Generation Plant Proteomics". During this period, research has moved towards the use of new platforms (DIGE and other Quantitative Proteomic approaches such as iTRAQ, SILAC and gel-free), the performance of suitable experimental designs and data analyses and the use of complementary approaches for data validation, such as Western blotting, transcriptomics, enzyme activity, reverse genetics, and those of Cell Biology. MIAPE requirements are being observed in most of the papers

Table 2 – Plant Proteomics reviews that have appeared in the 2007–September 2008 period [a].

Weckwerth W. Integration of metabolomics and proteomics in molecular plant physiology—coping with the complexity by data-dimensionality reduction. Physiol Plant 2008; 132: 176–189 [5].	Commented in the text
Salekdeh GH and Komatsu S. Crop proteomics: aim at sustainable agriculture of tomorrow. Proteomics 2007; 7:2976–2996 [45].	Review of proteomics of agricultural crops (rice, soybean, wheat, barley, maize), with emphasis on the use of the technique in food trazeability, substantial equivalence, allergen characterization, genotype cataloguing, assisting plant breeding. Proteomics of plant responses to stress. The omnipresent proteins: RubisCO, and other photosynthetic proteins, stress-responsive
Agrawal GK, Jwa N-S and Rakwal R. Rice proteomics: ending phase I and the beginning of phase II. Proteomics 2009; (in press) [62].	Excellent review on rice proteomics
Komatsu S, Toorchi M and Yukawa K. Soybean Proteomics. Curr Proteomics 2007; 4:182–186 [256].	Discuss the strength and weaknesses of proteomics technologies and limitations of current techniques for soybean biology studies.
Carpentier SC, Panis B, Vertommen A, Swennen R, Sergeant K, Renaut J, Laukens K, Witters E, Samyn B and Devreese B. Proteome analysis of non-model plants: a challenging but powerful approach. Mass Spectrom Rev 2008; 27: 354–377 [13].	Excellent review on proteomics of "orphan", non-model, non-sequenced plant species, but of interest from an agricultural point of view. Methodological considerations. Emphasis on protein identification. Banana as the species the authors work with.
Farrokhi N, Whitelegge JP and Brusslan JA. Plant peptides and peptidomics. Plant Biotechnol J 2008; 6:105–134 [63].	Commented in the text
Thelen JJ and Peck SC. Quantitative proteomics in plants: choices in abundance. Plant Cell 2007; 19: 3339–3346 [149].	Overview of the methodological options for Quantitative Proteomics, including label-free approaches, *in vivo* and *in vitro* isotopic labeling, iTRAQ and AQUA
Haynes PA and Roberts TH. Subcellular shotgun proteomics in plants: looking beyond the usual suspects. Proteomics 2007;7: 2963–2975 [185].	Subcellular Proteomics, assigning (bioinformatics) and determining (microscopy) subcellular location of proteins. Methodology: organelle purification, sample preparation, 2-DE, MudPIT, SDS/IEF-LC
Ito J, Heazlewood JL and Millar AH. The plant mitochondrial proteome and the challenge of defining the Posttranslational Modifications responsible for signalling and stress effects on respiratory functions. Physiol Plant 2007; 129: 207–224 [187].	Plant mitochondrial proteins and PTMs (phosphorylation, oxidation). Comparative study between plants (*Arabidopsis*), yeast and mammals (human).
Komatsu S, Konishi H and Hashimoto M. The proteomics of plant cell membranes. J Exp Bot 2007; 58: 103–112 [181].	Characterization of membrane proteins is still challenging. Proteomics of membrane proteins in *Arabidopsis* and rice.
Jamet E, Albenne C, Boudart G, Irshad M, Canut H and Pont-Lezica R. Recent advances in plant cell wall proteomics. Proteomics 2008; 8: 893–908 [186].	Methodology employed to study plant cell wall proteome. Around 400 cell wall proteins (CWPs) of *Arabidopsis*, representing about one fourth of its estimated cell wall proteome, have been described. From the proteins identified the biology of the cell wall can be reconstructed.
Qureshi MI, Qadir S and Zolla L. Proteomics-based dissection of stress-responsive pathways in plants. J Plant Physiol 2007; 164: 1239–1260 [257].	Comparison of the proteins that are induced or overexpressed under abiotic stress, including those identified by classical biochemical approaches and high throughput proteomics. Classical approaches: antioxidant enzymes, signalling proteins (salt-overly sensitive, jasmonate, abscisic acid pathways, MAPK, calcium, GABA). Proteomics of water, temperature, heavy metals, ozone, light, nutrient, stresses.
De la Fuente van Bentem S and Hirt H. Using phosphoproteomics to reveal signalling dynamics in plants. Trends Plant Sci 2007; 12:404–411 [221].	Research on plant signalling network through mass spectrometry and protein and peptide microrrays for identifying and characterizing phophorylated proteins and sites and to identify kinase activities. Specific reference to flagellin signalling in *Arabidopsis*.
Mehta A, Magalhaes BS, Souza DS, Vasconcelos EA, Silva LP, Grossi-de-Sa MF, Franco OL, da Costa PH and Rocha TL. Rooteomics: the challenge of discovering plant defence-related proteins in roots. Curr Protein Pept Sci 2008; 9:108–116 [258].	A new term coined: rooteomics. Root, probably the most recalcitrant organ. Recent developments, limitations of the current techniques, and technological perspectives for root proteomics aiming at the identification of resistance-related proteins are discussed.
Zhang Q and Riechers DE. Proteomics: An Emerging Technology for Weed Science Research Weed Science 2008;56: 306–313 [259].	Proteomics can be used to understand mechanisms of herbicide tolerance and weed resistance

(continued on next page)

Table 2 (*continued*)

Mehta A, Brasileiro AC, Souza DS, Romano E, Campos MA, Grossi-de-Sa MF, Silva MS, Franco OL, Fragoso RR, Bevitori R and Rocha TL. Plant–pathogen interactions: what is proteomics telling us? Febs J 2008;275:3731–3746 [260].	A review of papers covering plant–pathogen (virus, bacteria, fungi, nematodes) interactions with list of pathogen responsive proteins identified.
Nesatyy VJ and Suter MJ. Analysis of environmental stress response on the proteome level. Mass Spectrom Rev 2008. 27:556–574 [261].	Potential of proteomics to the analysis of environmental stress, defining a new field: environmental proteomics, mainly directed at studying the effect of toxic chemicals, on living organisms. General remarks on proteomics, experimental methods and technologies. released to the environment

[a] The books "Plant Proteomics. Technologies, strategies, and applications", edited by G.K. Agrawal and R. Rakwal, and "Plant Proteomics", edited by J. Samaj and J.J. Thelen, have not been included. A number of reviews, no listed, appear in this issue of the Journal of Proteomics: barley seed proteomics (Finnie and Svenson), soybean proteomics (Komatsu and Ahsan), root–microbe interactions (Mathesius), abiotic environmental stress in *Arabidopsis thaliana* (Taylor et al.), disulfide proteome, thioredoxin targets (Montrichard et al., and Lindahl1 and Kieselbach).

published most recently. However, plant biology research is far from fully exploiting the potential of Proteomics and, as is a general rule for plant biology research, progress in Plant Proteomics continues to lag behind that of Human and Yeast Proteomics (Fig. 1). Advances in Plant Genomics facilitate proteomic research, increasing confidence in protein identification and characterization. For example, among the 6,212,793 entries in the UniProtKB/TrEMBL (Release 39.1 of 02-Sep-2008), only 488,400 (7.86%) correspond to plants (Fig. 1A) (also see Schneider et al., in this issue). During this period, the complete genome sequences of grape [64], *Populus trichocarpa* [65] and the seedless plant *Physcomitrella patens* [66] have been reported [67], and genome sequencing projects for several important agricultural crops are in progress: barrel medic (http://www.medicago.org/, http://www.legoo.org/), tomato (http://www.sgn.cornell.edu/about/tomato_sequencing.pl), maize (http://www.maizesequence.org/index.html), potato (http://www.potatogenome.net), sorghum (http://www.jgi.doe.gov/sequencing/why/CSP2006/sorghum.html), and soybean (http://www.ncbi.nlm.nih.gov/Genbank). The number of ESTs reported at the Dana Farber (http://compbio.dfci.harvard.edu/tgi/plant.html) is indicated in Table 3. In all, at least 1.8 million plant-derived ESTs exist in the public domain. Sputnik (The Sputnik Exhaustive and Comprehensive Analysis of Plant-derived EST Sequences Database, http://www.plant-genomic-papers.com) has been implemented as a largely automated pipeline for processing, clustering and annotating large numbers of EST

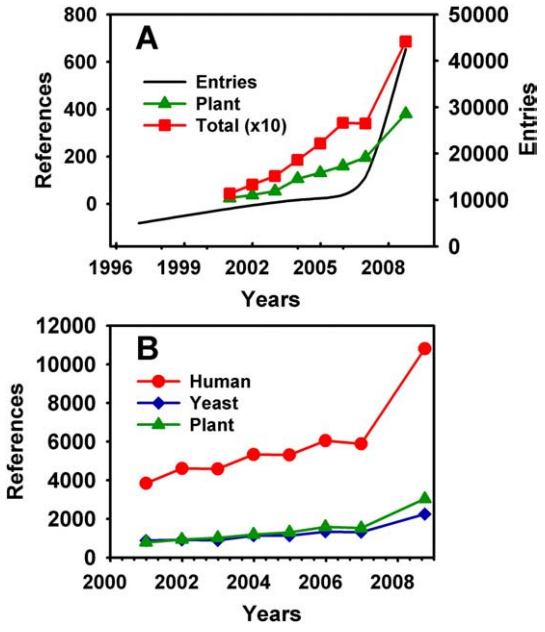


Fig. 1 – Number of references reported in the ISI Web of Knowledge during the 2000–2008 (up to September 19th) period. The search was performed with the key words: "proteomic(s)" or "proteome(s)" and "proteomic(s)" or "proteome(s)"plus "plant", "human" or yeast (A); "genomic (s)" or "genome (s)" plus "plant", "human" or "yeast" (B). In A, the number of entries reported at the UniProtKB/TrEMBL is also shown (http://us.expasy.org/sprot/ppap_stat.htm1).

Table 3 – Number of ESTs entries for some plant spp. by September 2008 [a].

Plant spp.	Dana Faber/NCBI
Apple (*Malus domestica*)	251 233/263 977
*Arabidopsis thaliana**	619 908/1 788 657
Barley (*Hordeum vulgare*)	479 742/517 057
Grape (*Vitis vinifera*)	347 879/209 596
Letuce (*Lactuca sativa*)	80 598/80854
Lotus (*Lotus japonicus*)	148 338/1581 37
Maize (*Zea mays*)	1 151 181/1 466 402
*Medicago truncatula**	259 642/262 363
Nicotiana benthamiana	41 922/43 668
Pine (*Pinus*)	355 326/509 731
Popplar (*Populus*)	411 781/417 092
Rice (*Oryza sativa*)	1 163 134/1 400 741
Soybean (*Glycine max*)	381 524/906 215
Sunflower (*Helianthus annuus*)*	93 283/94 562
Tobacco (*Nicotiana tabacum*)	163 524/244 400
Tomato (*Solanum lycopersicum*)	330 396/260 951
Wheat (*Triticum aestivum*)	1 034 368/1 067 423
Pea (*Pisum sativum*)*	No entries/6327
Holm oak (*Quercus ilex*)*	No entries/no entries
Radiate pine (*Pinus radiata*)*	No entries/151
Aleppo pine (*Pinus halepensis*)*	No entries/no entries
Pineaple (*Ananas comosus*)*	No entries/5649
Date palm (*Phoenix dactylifera*)*	No entries/4
Notro (*Embothrium coccineum*)*	No entries/no entries
Pteris vitata fern*	No entries/no entries

[a] Taken from The Gene Index Project at the Dana Farber Cancer Institute (http://compbio.dfci.harvard.edu/tgi/plant.html) and NCBI (http://www.ncbi.nlm.nih.gov/). With asterisk the plant systems our group is/has working/worked with.

sequences. Peptide sequences can be derived from the non-redundant cluster set, allowing large-scale genomic analyses on the basis of the partially reconstructed proteome. In fact, the term "Genomeless Genomics" has been coined to describe the application of complete-genomic analyses to these reconstructed genomes.

A total of 380 reports were found in the 'ISI Web of Knowledge' during this period by entering "Plant" and "Proteome(s)" or "Proteomic(s)" as keywords. This figure represents just a small proportion of the 6855 references on Proteomic(s) or Proteome(s) found in ISI; most of these references in ISI are devoted to Yeast and Human Proteomes (Fig. 1).

During this period there have not been large differences with respect to the previously revised ones (Table 1 of Rossignol et al., 2006, Jorrin et al., 2007 [1,2]) concerning the plant species under investigation, with *Arabidopsis* and rice, being the subject of around 75% of the original papers published. Nevertheless, proteomic papers have recently appeared on new species which include several herbaceous non-crops and woody plants (fruit and forest trees, these last ones being the species least addressed): wheatgrass *Elymus elongatum* [68], jasmine tobacco *Nicotiana alata* [69], the resurrection plants *Boea hygrometrica* [70] and *Xerophyta viscosa* [71], chaco potato *Solanum chacoense* [72], wild watermelon *Citrullus lanatus* [73], citrus *Citrus sinensis* [74], Austrian pine *Pinus nigra* [75], rose gum eucalyptus *Eucalyptus grandis* [76], sitka spruce *Picea sitchensis* [77], California pine *Pinus radiata* [78], and conference pears *Pyrus communis* [79].

As has been discussed before [1,2], proteomic analyses have been carried out with different individuals, plant developmental stages, organs, tissues, cells, and subcellular fractions. With the exception of the use of cell suspensions, protoplasts [80], pollen [81], and xylem/phloem saps [82], most Differential Expression Proteomics papers have used plant organs. It is important to note that organs contain a mixture of different cell types; each one with its own proteome signature, and only a small fraction of the cell types are likely to respond to chemical or biological stimuli. The use of laser microdissection techniques [83] for plant cell sampling has been reported in the transcriptomic and proteomic analysis of non-dividing pericycle cells of primary root of maize (*Zea mays*) [84]. The protein profile of these cells is compared with that of the rest of the transverse nonpericycle cells. In this study, approximately 30 μg of protein were extracted from 200,000 cells in 2-DE sample buffer, then separated by 2-DE (small gels, pH 4–7, silver staining). Altogether 20 proteins were identified after n-HPLC-ESI-Q-TOF and after searching through the nrNCBI database using MASCOT.

The proteome, by definition, is quite dynamic, and this is true even for clonal cells, as shown by Sigal et al. [85] using the non-small cell lung carcinoma line Hl299. They studied the variability of protein levels as well as the temporal dynamics of this variability using *in silico* synchronised cells. They used time-lapse fluorescence microscopy to follow the fluctuations in the level of 20 endogenous YFP-tagged proteins in individual cells over multiple cell cycles under controlled conditions. They found that the average coefficient of variation of the total fluorescence depended on the protein measured and ranged from 0.12 to 0.28, and that it remained nearly constant for a given protein throughout the cell cycle. Little is known about the genetic basis of variation in protein levels in genetically diverse populations. Using a label-free, mass spectrometry-based approach to measure protein levels in total, unfractionated cellular proteins, the genetic basis of variation in protein abundance was studied in a cross between two diverse strains of yeast [86]. Loci that influenced protein abundance differed from those that influenced transcript levels, a finding that highlights the importance of analysing the proteome directly to understand fluctuations in protein levels. Plants respond to small changes in the environment by adapting their metabolism, and this response is part of their survival strategy. Consistent with this idea, we found in our analysis of the *Quercus ilex* leaf proteome from mature trees in the field that the 2-DE protein profile depended on leaf orientation (North, South, East and West), position (top, bottom), and the time of day when the leaves were collected [87]. What originally was considered to be homogeneous material, since all the leaves came from the same tree, turned out to be far from true. As a consequence, it has been almost impossible to use the leaf protein profile in cataloguing trees or populations, since the intrapopulation variability can be greater than the interpopulation variability. Thus we changed the focus of our analyses to an organ with a more stable proteome, namely the seed (Valero et al., unpublished results).

In the present update, a total of 152 original papers have been reviewed. Descriptive Proteomics papers (13) deal with the proteome of whole plants; the proteome of major organs, leaves, roots, seeds, fruit, or pollen; or the proteome of extracellular xylem and phloem sap. Subcellular Proteomic papers (11) are devoted primarily to characterizing the proteomes of the chloroplast/plastid, mitochondria, and membrane fraction. A Differential Expression Proteomics strategy has been used to study developmental processes (18), responses to biotic (19) and abiotic stresses (31), and hormones (6). A number of papers deal with cataloguing plant genotypes or populations, comparative analysis of mutant/transgenic plants and their wild parental, and the characterization of allergens (14). PTM studies have characterized the phosphoproteome (9), and to a lesser extent the redox proteome (disulfide, nitrosylation, glutathionylation; 5), with only one paper each published on ubiquitination [88], glycosylation [89], acetylation [90], N-terminal modification, and protein sorting [91]. Only five papers deal with Interactomics block, and they have sought to identify Trx [92], calmodulin [93], 14-3-3 protein [94], and rice OsG1 (an ortholog of GIGANTEA that regulates photoperiod flowering in *Arabidopsis*) targets [95], and photosystem II light-harvesting complexes [96]. A number of papers appearing in the Proteinomics or Targeted Proteomics approach group study seed proteins and proteases, among others (16).

3. Methodology

3.1. *Second-generation techniques and appropriate experimental design and data analysis*

The workflow of a standard proteomic 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, and statistical analysis of data and validation. New insights and advances in the different steps of the workflow in the literature reviewed are discussed below. The most appropriate protocols to be used depends on and must be optimised for the biological system (i.e. plant species, organ, tissue, cells), as well as the objectives of the research (descriptive, comparative, PTMs, interactions, targeted Proteomics).

3.2. *Experimental design*

A good experimental design is crucial for the success of any proteomic experiment. Erykson and Fenyo [97] have developed a simulation tool for evaluating the success of current designs and for predicting the performance of future, better-designed proteomics experiments [96]. The simulation gives a holistic view of a general analytical experiment and attempts to identify the factors that affect the success rate. It has been used to predict the success of proteome analyses of Human tissue and body fluids that use various experimental design principles. Several parameters are required to simulate the steps of a proteome analysis: i) the distribution of protein amounts in the sample analysed; ii) the loss of analyte material and the maximal limit of the amount loaded at each step of sample manipulation (e.g. separation, digestion, and chemical modification); iii) the dynamic range, the detection limit and the losses associated with MS analysis. Depending on what experiment is being modelled, the detection limit used in a simulation can represent either protein identification only (lower identification limit) or protein identification with quantification (lower quantification limit).

The establishment of an adequate number of replicates is essential for any Differential Expression Proteomics experiment. This number should be set up while taking into account the dynamic nature of the proteome, and a good number will allow correct interpretation of the results and the confident assignment of any protein to the group of variable ones. This makes sense when searching for proteins that can be used as markers of disease or when looking for protein markers to develop plant breeding programmes. In our work to characterize the *P. radiata* needle proteome [78], we previously determined the analytical and biological variability. Tests of the analytical variability examined both the experimental procedures (protein extraction, IEF, SDS-PAGE, gel staining-destaining) and the accuracy of the hardware and software in acquiring and analysing images. Tests of the biological variability looked at several different samples from different trees. Tests performed using 10 and 12 analytical and biological replicates from a representative set of 250 spots gave values for the analytical and biological variability, respectively, of 31% and 42%. This is in contrast to the analytical and biological variability reported for other systems, which is as low as 16% and 24%, respectively. In our work using *P. radiata*, we found differences in the standard error of mean spot quantity, depending on the number of replicates; the error ranged from 111 and 115 ng for two analytical and biological replicates, to 58 and 59 ng for 10 analytical and 12 biological gel replicates. Using more than six biological replicates did not significantly reduce the standard error, so this figure should be optimal for comparative proteomic experiments. Since normally this is not feasible, most of papers in our literature review used only three biological replicates. Given the susceptibility of the data to variation, we were restrictive in our study of *P. radiata* when deciding whether a spot showed variation. First, all the spots considered had to be consistent, i.e., present or absent in all the biological replicates of the particular stage in question; second, when not qualitative (presence vs. absence), differences had to be statistically significant ($p \leq 0.05$, ANOVA); finally, the variance with respect to a control (in this case, embryos from mature seeds) had to be higher than the average biological coefficient of variance determined for a representative set of 150 spots.

3.3. *Sample preparation*

The importance of the extraction protocol in a proteomic experiment can be summarised in the following statement: only if you can extract and solubilise a protein you have the chance to detect and identify it. This sentence summarises the importance of the extraction protocol in a proteomic experiment. This is even more important in the case of plant tissues, due to the low protein content relative to other systems, the presence of the cell wall and vacuoles that account for the majority of the cell mass, the presence of proteases and oxidative enzymes, and the accumulation of large quantities of polysaccharides, lipids, phenolics and other secondary metabolites [98]. Two main types of protocols have been used with plant material. The first type involves tissue homogenisation in buffer-based media, while the second one uses organic solvent media (TCA-acetone, phenol, precipitation protocols), although both protocols can be combined. Since no single protein extraction protocol can capture the full proteome, the chosen protocol should be optimised for the particular plant tissue and research objective. The ideal method should be highly reproducible and should extract the greatest number of protein species, while at the same time reducing the level of contaminants and minimising artifactual protein degradation and modification. Advantages and limitations of each extraction media have been evaluated and discussed in recent publications [99–104]. Now, hydrophobic proteins require specific protocols. Thus, for example, n-dodecyl β-D-maltoside favours the capturing and analysis of hydrophobic proteins, and in contrast to SDS, it can be removed prior to MS analysis. Recently, Toorchi et al. have reported the use of acoustic technology for tissue and protein pellet homogenisation, which stands as an alternative to the classic mortar and pestle or sonic baths [105]. This technology performs far better than water bath sonication at producing high-quality 2D gels and minimising the processing time required for High-throughput Proteomics research.

In order to evaluate the effectiveness of a given extraction protocol, protein quantification is needed. This makes the total absence of such data from some recent papers all the more striking. When using standard colorimetric methods for protein quantitation, namely the Bradford, Lowry and BCA assays, we must consider that some of the components of the solubilisation media may interfere with the colour development reaction [106].

Another issue to consider is the extreme complexity of the proteome and the large dynamic range in protein abundance, which overwhelm the capability of all currently available analytical platforms. Sample prefractionation is a good approach to reduce the complexity of the proteome sample and decrease the dynamic range. Jiang et al. [107] reviewed different protein fractionation technologies and methods based on structural and functional characteristics of the proteins, including the isolation of peptides containing rare amino acids, terminal peptides, PTM peptides and endogenous peptides. Recently, a number of papers have reported the applicability of ribulose-1,5-bisphosphate carboxylase/oxygenase depletion columns for leaf proteomic analysis [108]. In our hands, the method has given excellent results. Polyethylene glycol fractionation has also proven to be valuable for depleting RubisCO from samples [109,110]. Another possibility is the EQUALIZER technology, based on a combinatorial library of hexameric peptide (combinatorial peptide ligand libraries or CPLL) ligands bound to porous beads. Each bead contains billions of copies of a unique hexapeptide ligand distributed throughout its porous structure, and each bead potentially has a ligand different from that of every other bead. With a population of millions of individual peptide ligands obtained by combinatorial chemistry, any protein present in the starting material could theoretically interact with one or a few particular beads [111]. This technology has not yet been reported in studies on plants.

3.4. *One and two-dimensional electrophoresis*

One-dimensional electrophoresis, SDS-PAGE, in combination with appropriate software, is a simple and reliable technique for finger-printing crude plant extracts [112], and is especially useful in the case of hydrophobic and low-molecular-weight proteins. We are using it to characterize and catalogue populations of holm oak (Valero et al., unpublished results). It can also be used to quantify changes in the abundance of a specific protein or changes in specific modifications of a protein using in-gel stable isotope labelling. The combination of SDS-PAGE, band cutting, trypsin digestion and LC separation of the resulting peptides remains the proteomic technique capable of providing the greatest protein coverage [113,114]. Electrophoretic methods can be made compatible with chemical labelling in order to allow Quantitative Proteomics using such techniques as iTRAQ [115].

2-DE is by far the predominant separation technology, and it is continuously being evaluated and improved in the areas of separation of hydrophobic proteins [116], gel staining [117–121], image capture and analysis [122–126] and automation [127]. One of the major criticisms of 2-DE is its low precision, with relative standard deviations reported to fall in the range of 15–70%. Major sources of variability for this technique may include the transfer between the first and the second dimension, the analyst's expertise and the detection of separated proteins [128,129]. One of the main error sources can be attributed to the irregular changes in the background signal from gel to gel, which can be solved using fluorescence labelling.

To compare 2-DE maps between samples, which could be analytical or biological replicates, cells, tissues, organs or treatments, plant proteomic papers prior to the reviewed period used arbitrary criteria (fold ratios) or univariate parametric and non-parametric statistical tests, namely Student's *t*-test or Mann–Whitney *U*-test to compare two groups and ANOVA or Kruskal–Wallis to compare more than two groups. These tests analyse individual spots instead of the complete set, omitting information about correlated variables. Multivariate data analysis methods, such as principal component analysis (PCA), are now used to pinpoint spots that differ between samples. These multivariate methods focus not only on differences in individual spots, but also on the covariance structure between proteins [122]. However, the results of these methods are sensitive to data scaling, and they may fail to produce valid multivariate models due to the high number of spots in the gels that do not contribute to the discrimination process [130]. One of the limitations of PCA analysis is that it does not allow missing values, a problem that can be avoided by imputing them when possible (if enough replicates are available) [131]. In contrast to these multivariate methods, univariate tests increase the number of false positives, and for each species, a measure of significance reflecting rates of false discovery should be calculated. This measure of significance is called a *q* value, and it has been described by Karp et al. [132].

Blue native polyacrylamide gel electrophoresis (BN-PAGE) has become very popular for analysing membrane proteins and protein complexes, but in some cases, as in the analysis of chloroplast light-harvesting complexes (LHC), it shows poor resolution, especially for proteins in the range of 22–25 kDa [133]. In order to preserve the advantages of the technique and use it with in-gel fluorescent detection and in-gel catalytic activity assays, Coomassie dye in the cathode buffer can be replaced with non-coloured mixtures of anionic and neutral detergents [134]. D'Amici et al. have reported a new protocol for native 3-D electrophoresis that allows exhaustive separation and identification of membrane proteins [135]. The method was applied first to the thylakoid membrane of spinach, which contains four large protein complexes. The protocol includes native liquid phase isoelectrofocusing (N-LP-IEF) of protein complexes in the first dimension, followed by blue native polyacrylamide gel electrophoresis (BN-PAGE) in the second dimension. Finally, the individual components of the complex are resolved using denaturing electrophoresis (SDS-PAGE).

3.5. *Second-generation techniques*

The use of differential 2-DE (DIGE) for comparative studies is reported in a significant number of papers published during the reviewed period (Heinemeyer et al., this issue). Problems can arise with this methodology when two different species for a given protein, one labelled and the other unlabeled, differ in masses. As a consequence, in the case of automated spot picking according to CyDye coordinates, the exact positions of the maximum amounts of the proteins under study are not accurate, resulting in a loss of sensitivity in the subsequent MS analyses. To solve this, differences in the migration of labelled and unlabelled species should be quantified in order to obtain more reliable protein identification [136].

The use of gel-free, LC-based approaches [137; Stevenson et al., this issue], including MudPIT [138,139] still remains anecdotal in Plant Proteomics studies [140].

Second-generation MS technologies for Quantitative Proteomics [141,142] have begun to be applied to Plant Proteomics research. These technologies include stable isotope labelling, ICAT [143], iTRAQ [144,145], SILAC [146,147], or label-free methods (peak integration, spectral counting). The latest trends and applications of the technique, as well as its potential, have been previously reviewed [148–150; Oeljeklaus et al., this issue). Combining different approaches increases the proteome coverage. In this way, Hebeler et al. [151] carried out a quantitative analysis of changes in protein abundance related to early leaf senescence in *Arabidopsis* using a double reverse labelling strategy, DIGE and ^{15}N-labeling of buffer extracted proteins followed by HPLC/ESI-MS/MS (Q STAR XL) and protein identification based on the *Arabidopsis* EBI protein sequence database (SEQUEST). The combination of gel and gel-free techniques allowed identification of different sets of variable spots.

3.6. *Mass spectrometry*

Mass spectrometry of proteins is constantly evolving, with protocols, machines and software continuously improving, such that the technology is approaching the limit of its capabilities. In the past few years, the development of Orbitrap and new dissociation methods such as electron-transfer dissociation, have opened up new possibilities in proteome analysis. Although Bottom-up Proteomics (analysis of proteolytic peptide mixtures) remains the predominant platform, top-down strategies (analysis of intact proteins) should allow a more complete characterization of the proteome, including protein isoforms and post-translational modifications. All these aspects have been discussed in detail in recent reviews [12,152–155].

Using classical quadrupole and ion trap mass analysers, intact protein masses can be determined with standard deviations in the range of 2–5 kDa. In some cases, these approaches, based on nonisotopically resolved related ion deconvolution using average masses from protein spectra, show differences between measured and calculated masses that exceed the standard deviations. This occurs due to the complexity of mixtures of protein isoforms and the presence of post-translational modifications. These limitations can be overcome by using highly accurate mass determinations from Fourier transform ion cyclotron resonance mass spectrometry [156,157]. The use of a hybrid ion trap Fourier transform mass spectrometer has been reported in the analysis of the citrus fruit proteome [158] and of the chloroplast envelope of pea and maize [148].

3.7. *Different techniques are complementary*

The literature makes clear that the different techniques, platforms, and workflows are complementary, and all of them are necessary for more complete coverage of the proteome. This is clearly shown in the paper by Lee et al. [159], which analysed several different *Arabidopsis* leaf protein fractions: i) TCA–acetone precipitates solubilised in urea, urea–SDS and urea-n-dodecyl β-D-maltoside buffer; ii) membrane proteins; iii) hydrophobic proteins; and iv) organellar proteins (nuclei, microsomal membranes, plastids, and mitochondria). They analysed these fractions using two shotgun approaches, MudPIT and 1-D SDS-PAGE-LC-MS, with a quadrupole IT mass spectrometer. In all, 2342 nonredundant proteins were identified, and 68–545 proteins were identified by only one of the two methods. For example, in the case of whole-leaf extract solubilised by SDS buffer, 70 proteins were identified only by MudPIT and 109 only by 1-D LC-MS. For both approaches, the number of transmembrane proteins ranged from 57 to 77. One disadvantage of 1-D gel-LC-MS/MS may be that the limits of protein discovery are reached after a few replicate experiments. This is not the case for MudPIT, which requires multiple analyses in order to survey all the proteins in a sample.

3.8. *Algorithms*

Proteomic investigation is facilitated, and the proteomic scientist's life is complicated, by the increasing number of protein databases containing protocols, search and quantification algorithms, protein sequences, predicted and verified functional information, structural and functional classifications, phylogeny and ontology [159]. It is outside the scope of this review to discuss all of these aspects in detail, and we refer the reader to recent key reviews or original papers [30,160–167]. Of special relevance to the community is the characterization of proteomes using homology-driven proteomics, since the genomes of most plant species of interest have not been sequenced and in fact databases lack any entries for some of them [168,169].

4. Descriptive Proteomics

4.1. *Increasing proteome coverage, at least in Arabidopsis and rice*

Within the Descriptive Proteomics area, the work by Baerenfaller et al. should be highlighted [37], since it represents the most exhaustive proteome map reported for plants. This study supports that a combination of SDS-PAGE, band cutting, trypsin digestion and LC separation of the resulting peptides constitutes the most powerful technique in terms of protein coverage [113,114]. In all, 1354 linear ion trap runs were conducted with protein extracts from six different organs and cell suspensions. The outcome was 13,029 proteins identified, comprising 10,902 from plant organs and 8698 from cell suspensions, using PeptideProphet and PepSplice algorithms, with 86,456 unique peptides originating from 790,181 tandem MS peaks, with a rate of false discoveries below 1%. This figure corresponds to nearly 50% of all predicted *Arabidopsis* gene products, not including protein variants. Despite its success, this work is a clear example of the limitations of the proteomic approach. Proteins involved in transcriptional regulation and signalling are underrepresented, while housekeeping proteins, involved in primary metabolism (mainly carbohydrate metabolisms), are overrepresented. This analysis preferentially detected large proteins and proteins for which the transcripts are expressed at high levels. Major modifications of peptides that were detected included Met or Trp oxidation and Cys carbamidomethylation, all of which supposedly took place during *in vitro* manipulation.

When protein and transcript levels were compared, values were found to range from 0.52 (seeds) to 0.68 (leaves). Comparing the MS/MS spectra with the TAIR7 genome database identified peptides from genomic regions with no annotated protein-coding capacity. Such comparison also identified organ-specific biomarkers, allowing the compilation of an organ-specific set of peptides corresponding to 4105 proteins that will facilitate targeted, Quantitative Proteomic surveys in the future.

Using a (descriptive) proteomic approach, Catusse et al. reported a 2-DE reference map of sugarbeet seeds, root, stem, cotyledons and perisperm tissues. Some of the proteins identified were not reported previously in seeds, and the results allowed the proposal of novel mechanisms for translation initiation based on both cap-independent and dependent processes. Seed metabolism could be reconstructed from the proteins identified, which corresponded to 121 biochemical functions, and some pathways such as glycolysis were thoroughly represented [170]. In a similar work, Katz et al. studied the metabolism of the citrus fruit, with a focus on citrate and sugar metabolism, as well as vesicular trafficking [158].

Due to their simple proteome, some plant samples are very good candidates for proteomic analysis, with their low protein content posing the major difficulty to complete coverage. Aki et al. [171] reported a shotgun analysis of the proteome of phloem and xylem saps from rice. In that work, they took advantage of the complete genomic information available for this plant. Xylem sap was prepared using the root pressure method, whereas phloem sap was prepared using the insect laser method. The technical difficulties caused by the limited amount of protein in these samples were overcome by the use of nano-flow liquid chromatography linked to a mass spectrometer. Two alternative methods were used: i) SDS-PAGE, band cutting and digestion with the tryptic peptides subjected to an integrated microfluidic device (HPLC-Chip for a nano scale LC) coupled to an ion trap MS; and ii) 2-D LC of the sample tryptic peptides (SCX-HPLC-chip), with the chip coupled to the IT. The number of proteins identified using both strategies was high: 118 different proteins and eight different peptides in xylem sap, 107 different proteins and five different peptides in phloem sap. The identified proteins included signal transduction proteins, putative transcription factors, stress response factors, and metabolic enzymes. In phloem sap, three of the proteins identified showed significant similarity to the *Arabidopsis* TFL1 and FT proteins that play important roles in controlling the inflorescence architecture and the transition from the vegetative phase to the reproductive phase. The expression pattern of genes coding for the FT-like proteins identified was examined by qRT-PCR, revealing a tissue-specific protein pattern. The existence of peptide signals that act over long distances by moving through the phloem or xylem is a matter of controversy and speculation. Consequently, this study also carried out MS analysis of peptides, but no known plant peptide hormones were detected, such as systemin, phytosulfokine or CLAVATA3.

In addition to these studies, the proteomes of several other plants were analysed in the papers published during the review period: *Arabidopsis* leaves, roots, and flowers [172], *P. radiata* needles [78], wheat roots [173], tomato pollen [174], maritime pine xylem [6], soybean xylem/apoplast [140], maize root pericycle cells from primary roots [84] and *M. truncatula* protoplasts [80].

5. Subcellular Proteomics

5.1. *Imminent completion of the chloroplast and mitochondrial proteomes*

Proteomics provides a powerful tool for characterizing the biochemistry of organelles. However, information obtained through proteomics can only be valid if the organelle sample is not heavily cross-contaminated. Hence, although proteomic analysis can give an indication of possible cross-contamination, biochemical analyses comprising tests of enzyme activity, Western blotting, and metabolite analysis must always be performed [7,175]. Significant effort has been devoted to the development of protocols for isolating highly purified preparations of organelles and sub-organellar compartments for protein profiling. These organelles and compartments include: chloroplasts [176,177], mitochondria [178], vacuoles [179], cell wall [180], membranes [181,182], and embryoplasts [139].

The technique known as LOPIT, which stands for localisation of organelle proteins by isotope tagging, assesses the distribution patterns of organelles by measuring the relative abundance of proteins between fractions along density gradients using stable isotopic tags, and has been used to assign proteins to different endomembrane organelles [183].

Subcellular Proteomics in plants has been extensively reviewed [38,181,184–187]. In this issue, H. Millar and coworkers review a collection of papers in which plant responses to abiotic stresses were evaluated in *Arabidopsis* using whole-tissue and Subcellular Proteomics approaches. Therefore, we will comment on just some of the papers that have appeared on this topic during the review period.

A detailed analysis and characterization of the *Arabidopsis* ribosomal proteome, both cytosolic and mitochondrial, has been reported by Carroll et al. [188]. They performed *in silico* digestion of the complete ribosomal proteome using an in-house, modified version of Proteogest in order to define targets for data acquisition and to establish a strategy for data collection that maximises recognition of peptides derived from ribosomal proteins. Approximately 13,000 unique peptide sequences were obtained, with 9400 predicted to come from only one of the 409 protein sequences (gene-specific peptides) corresponding to 80–249 proteins. Ribosomal proteins were separated by SDS-PAGE, and bands (21) or ribosomal extracts containing low-molecular-weight proteins were digested with trypsin, or with both chymotrypsin and pepsin in order to increase the coverage. Peptide match information from MASCOT files was extracted and filtered, and only the highest quality unambiguous spectra that matched exclusively certain *in silico* predicted gene-specific peptides were accepted. This procedure, based on high confidence criteria, allows what are called gene-specific identifications. A total of 1446 high-quality MS/MS spectra matching 795 peptide sequences was reported, allowing the identification of peptides from five gene families previously unidentified, and providing experimental data on 79 of the 80

different types of ribosomal subunits. The analysis identified 31 small subunit proteins, 46 large subunit proteins, and seven P-proteins, all of them gene-specific. Several PTMs were identified using chemically mild sample preparation conditions, electrophoretic prefractionation and phosphoprotein/phosphopeptide enrichment, including initiator methionine removal, N-terminal acetylation, N-terminal methylation, lysine N-methylation and phosphorylation.

Several studies have provided new insights into the role of sphingolipid/sterol-rich domains, detergent-insoluble membranes (DIM), and so-called "lipid rafts" of the plasma membrane (PM) in mammalian cells and in leaves, cell cultures and seedlings of higher plants [189]). Lipid rafts are defined as a Triton X-100-insoluble, low-density membrane fraction. DIM fractions prepared from *M. truncatula* root PMs [190] were characterized by structure-electron microscopy and the levels of lipid, sterol, sphingolipid and protein analysed. They were found to be enriched in sphingolipids and Delta (7)-sterols, with spinasterol being the main component, although steryl glycosides and acyl-steryl glycosides were also present. A total of 270 proteins were identified following SDS-PAGE, gel slice digestion, Q-TRAP nLC-MS/MS, and MASCOT searches of the nonredundant NCBI subdatabase *M. truncatula*, the *M. truncatula* local EST database (http://medicago.toulouse.inra.fr/Mt/EST/) and the *M. truncatula* genomic database (http://www.medicago.org/genome/). Among the proteins identified, receptor kinases and other proteins related to signalling, cellular trafficking and cell wall functioning were well represented, whereas those involved in transport and metabolism were poorly represented. The results also showed evidence for the existence of a complete PM redox system in lipid rafts.

Other subcellular proteomic studies, which are not mentioned in detail here, deal with chloroplast envelope membranes from pea and maize [148], stroma lamellae from 10 different plant species [191], potato tuber amyloplasts [192], soybean peroxisomes [193] and plasma membrane and tonoplast from rice [194,195] and *Arabidopsis* [196].

6. Differential Expression Proteomics

6.1. *The first option to get into Proteomics and dominated by 2-DE*

The biological meaning of a protein can be inferred from its differential presence in specific genotypes (i.e. wild, transgenic, mutant), cell type, developmental stage and external conditions. Half of the original plant proteomic papers published during this review period used this strategy to identify and characterize proteins or genes involved in growth, development and plant responses to hormones, stresses, and other events including pollen germination, embryo formation, seed development and germination, fruit development and ripening, leaf senescence, programmed cell death, wood formation, the effect of auxins, giberellins, abscisic and jasmonic acids, mineral nutrition, symbioses, biotic stress (pathogen, insects), abiotic stress (drought, osmotic shock, temperature, UV light, ozone, heavy metals, waterlogging), and oxidative stress (Table 4).

Although 2-DE coupled to MS is by far the most frequently used platform for Differential Expression Proteomics, there are some exceptions. For example, Larson et al. [197] and Wen et al. [198] used MudPIT to study the reaction of sugar beet to *Fusarium oxysporum* and the reaction of root pea to *Nectria haematococca*, respectively. Other exceptions are the works of Palmblad et al. and Patterson et al. The former reported the use of metabolic stable isotope ^{15}N-labelling to identify heat-responsive proteins in *Arabidopsis* [146]. The latter used the iTRAQ approach to identify boron-responsive proteins in barley roots [199].

6.2. *Stress responses*

Proteomics has been used widely during the review period to gain a deeper knowledge of the processes that take place in an organism subjected to stress. The type of stress most studied is pathogenesis, followed by drought stress, metal toxicity, and salt-osmotic stress. Although diversity is the rule, most studies examining these stresses have focused on *Oryza sativa* (rice). The plant tissues used in these papers have been as diverse as leaves and cotyledons, roots, fruits, phloem and xylem saps, apoplastic fluid, entire seedlings, shoots, stem segments, seeds, nuclear fractions, gametophores and meristematic tissue. Among the proteins observed to respond to stress are those involved in detoxifying reactive oxygen species (ROS), which makes sense considering that most stressors increase production of ROS in plants. Hajheidari et al. [200] studied the effect of drought stress in the seeds of three wheat genotypes that differ in their tolerance of water shortage. Soluble proteins were buffer-extracted and separated by 2-DE, and significant differences were assessed by two-way ANOVA. A total of 121 drought-responsive spots were analysed by MALDI TOF/TOF, and 57 proteins were identified. Of these, 38 have been identified as putative Trx targets, indicating a link between drought tolerance and redox homeostasis.

Kang et al. [201] carried out an interesting research aimed at characterizing the phenotype of the lesion-mimic mutant *spotted leaf 6* (*spl6*) rice. Proteins present at different levels in the leaf blades of wild-type and mutant plants were identified by MALDI-TOF using ProFound. Northern blot assays were performed to correlate protein expression with the levels of the corresponding transcripts. Microscopy observations allowed the detection of degradation of the thylakoid membrane in the nonspotted sections of the mutant, which explained the absence of ATP synthase, RuBisCO large subunit, and RuBisCO activase in this genotype. The authors propose the absence of TPX and PDI as the reason why the cells cannot resist the oxidative burst that results in the degradation of the thylakoid membranes, programmed cell death and lesion development.

Other proteins that are often found in response to stress are the pathogenesis-related proteins (PRs). Many papers that were published before the present review period and that examine plant development and responses to stresses limit their proteomic analysis to one time point. As a consequence, information on multiple quantitative plant responses and the sequential events that mediate them is missing. Time-course analysis can provide this missing information, and has been useful, for example, in showing that plant resistance

Table 4 – Differential Expression Proteomics papers appeared during the reviewed period.

Topic	Biological system	Ref.
Seed development	*Medicago truncatula* seeds (seed coat, endosperm and embryo) at five developmental stages	[205]
	Zea mays kernel or endosperm	[206]
Seed germination	*Fagus sylvatica* seeds imbibed in water or treated with ABA or GA_3	[207]
	Oryza sativa seeds sampled at different germination times	[203]
	Arabidopsis thaliana seeds with a controlled deterioration treatment	[204]
De-etiolation	*Oryza sativa* etiolated seedlings	[262]
Fruit development and ripening	*Solanum lycopersicum* var. *cerasiforme* fruits collected from anthesis to maturity	[263]
	Vitis vinifera L. cv. Cabernet Sauvignon grape skin collected at different days after anthesis	[264]
	Vitis vinifera cv. Nebbiolo Lampia grapes sampled from 1 month after flowering to complete ripe stage	[265]
Tillering	*Oryza sativa* basal nodes sampled from two cultivars with different numbers of tillers	[266]
Senescence (leaves)	*Arabidopsis thaliana* mutant plants displaying advanced leaf senescence	[151]
Pollen germination	*Oryza sativa* mature and germinated pollen grains	[81]
Microspore derived embryo	*Brassica napus* microscospore-derived embryos in two different culture systems	[10]
Juvenile wood formation	*Eucalyptus grandis* stems cambial region sampled at different ages	[76]
Chloroplast differentiation	*Oryza sativa* seedling-derived plastids with different periods of light exposure	[267]
	Pisum sativum L. protein complexes from etioplasts, etio-chloroplast and chloroplasts	[268]
Callus differentiation	*Oryza sativa* seed-derived callus at different differentiation stages	[269]
Plant hormones	*Oryza sativa* seeds treated with gibberellins GA and ABA	[270]
	Oryza sativa leaf sheats treated with gibberellin GA_3	[271]
	Oryza sativa leaves with different ABA treatments	[272]
	Oryza sativa leaf sheaths treated with auxin 2,4-D or auxin repressor PCIB	[273]
	Oryza sativa seedling-derived shoots and roots with different concentrations of jasmonic acid	[274]
Signal transduction	Brassinosteroid response in *Arabidopsis thaliana*. BR-deficient mutant *det2-1* and BR-insensitive mutant *bri1-116*. brassinolide	[208,209, 275]
Nitrogen source	Roots from *Lolium perenne* plants grown in liquid medium with different N supplies (ammonium, nitrate, glycine)	[276]
Mineral nutrition: Fe	Roots from tomato plants (*fer* mutant, wild-type and overexpressing transgenic line) grown with different amounts of Fe	[277]
Programmed cell death	Rice seedlings (wild and lesion mimic mutants) inoculated with *Magnaporthe grisea*.	[278]
Symbiosis: *Sinorhizobium meliloti*	*Medicago truncatula* roots (wild type and supernodulation mutant) inoculated with the *Sinorhizobium meliloti*.	[279]
	Nodules from 10-weeks old *Medicago truncatula* inoculated with *Sinorhizobium meliloti* (control and drought stressed)	[280]
Pathogens	Leaves of blackleg-susceptible *Brassica napus* and blackleg-resistant *Brassica carinata* inoculated with *Leptosphaeria maculans*.	[281]
	Leaves of canola (*Brassica napus* L.) plants infected with *Sclerotinia sclerotiorum*	[282]
	Roots of *Brassica napus* infected by *Plasmodiophora brassicae*	[283]
	Leaves of *Capsicum chinense* plants harbouring the L^3 gene infected by two strains (compatible, incompatible) of pepper mild mottle virus	[202]
	Spikelets of six barley genotypes of varying resistance inoculated with *Fusarium graminearum*.	[284]
	Leaves of *Populus euramericana* inoculated with *Marssonina brunnea* f. sp. *multigermtubi* at different after inoculation times	[285]
	Leaves of *Arabidopsis* wild-type and *CaHIR1*-overexpressing transgenic plants inoculated with *P. syringae* pv. *tomato*, *Hyaloperonospora parasitica* and *Botrytis cinerea*.	[286]
	Peach fruit inoculated with *Penicillium expansum* and treated with SA or *Pichia membranefaciens*.	[287]
	Sugar beet genotypes resistant (R) and susceptible (S) inoculated with *Fusarium oxysporum* (F-19)	[197]
	Roots of pea inoculated with *Nectria haematococca*.	[198]
	Xylem sap of tomato plants infected with *Fusarium oxysporum*	[82]
	Phloem of Austrian pine inoculated with *Diplodia scrobiculata* and *Sphaeropsis sapinea*	[75]
	Leaves of *spl6* mutant and wild-type rice plants were used	[201]
	Apoplastic fluid of *Arabidopsis* seedlings, control or treated with oligogalacturonids.	[288]
	Seedlings of *O. sativa* subsp. *japonica* treated with probenazole and inoculated with bacteria	[289]
Trichoderma induced resistance	Cotiledons of *Cucumis sativus* L. plants inoculated with *T. asperellum* strain T34+ and *Pseudomonas syringae* pv. *lachrymans* (Psl).	[290]
	Root and shoot tissues of maize (*Zea mays*) treated with *Trichoderma harzianum* T22	[291]
Insects	Stem segments of Sitka spruce (*P. sitchensis*) seedlings mechanically wounded and weevils inoculated	[77]

(continued on next page)

Table 4 (*continued*)

Topic	Biological system	Ref.
Drought	Droughted seeds of three spring wheat genotypes differing in its drought tolerance.	[200]
	Roots of *Zea mays* water stressed	[292]
	Xylem sap of water stressed *Zea mays* plants	[293]
	Nuclear fraction of chickpea seedlings subjected to progressive dehydration	[294]
	Extracellular matrix (ECM) of eight chickpea (*Cicer arietinum* L.) cultivars subjected to dehydration.	[295]
	Dried and rehydrated leaves from *Boea hygrometrica* plants.	[70]
	Leaves of droughted tall wheatgrass (*Elymus elongatum* Host) plants	[68]
Salt stress	Roots of NaCl treated *Triticum aestivum* plants.	[296]
	Gametophores of *Physcomitrella patens* were exposed to high-salinity (250, 300, and 350 mM NaCl).	[67]
	Roots of *Arabidopsis* plants treated with NaCl.	[297]
Osmoticum stress:	Meristematic tissue of Cachaco (ABB cooking banana) and Mbwazirume (AAA highland banana) cultured in high-sucrose medium.	[298]
	Rice leaf sheath treated with mannitol, drought and cold	[299]
Heat	Leaves of *Arabidopsis*	[146]
	Leaves of rice	[300]
Cold	Rice leaf and roots	[109,110]
Shadow avoidance	Leaves of shadowed tomato plants	[301]
UV light	Leaves of soybean lines differing in flavonoid contents	[302]
Ozone	Rice leaves	[303]
Heavy metals	Leaves of poplar clones treated with cadmium	[304]
	Roots of *Cannabis sativa* treated with cupper	[305]
	Second oldest trifoliate leaf of cowpea treated with manganese	[306]
	Roots and leaves of B-tolerant barley plants treated with boron	[307]
	Roots of Al-resistant and susceptible plants	[308]
	Roots of rice treated with Al	[309]
Nutritional limitation	Roots tomato wild and mutant plants	[310]
Agrochemicals: safeners	Root, leaf, and coleoptile of *Triticum tauschii* seedlings treated with herbicide and/or safener	[311]
Waterlogging	Leaves of tomato plants subjected to waterlogging	[312]
Oxidative stress	Roots of rice treated with GSH, DPI or ascorbate	[313]
	Leaves of rice treated with different concentrations of H_2O_2	[314]

to pathogens is associated with rapid induction of the expression of specific genes. As an example, Yoshimura et al. [73] reported changes in the proteome (2-DE) of root watermelon (*C. lanatus* L.), a xerophyte species, in response to water withholding. Comparative analysis revealed that many proteins induced in the early stage of drought stress are involved in root morphogenesis and carbon/nitrogen metabolism, which may contribute to drought avoidance by enhancing root growth. In contrast, the majority of lignin synthesis-related proteins, which enhance tolerance of physical desiccation, and molecular chaperones, which maintain protein integrity were induced at a later stage of drought stress. These data suggest that this xerophyte switches survival strategies from drought avoidance to drought tolerance during progression through drought stress, and this switch involves temporal regulation of the root proteome.

Using pepper plants harbouring the *L3* gene, which confers resistance to pepper mild motle virus (PMMoV), Elvira et al. [202] analysed the PRs induced in response to virus inoculation using either a virulent (PMMoV-I) or avirulent (PMMoV-S) strain. Acid-soluble proteins from leaves and apoplastic fluid proteins were buffer-extracted and separated by 2-DE. Proteins were analysed by Western blots and identified by Edman sequencing and BLAST searching. Northern blot analysis of RNA coding for PR proteins was also performed. From their results, they concluded that, although a few PR proteins are specifically induced in the incompatible interaction, both reactions are qualitatively similar, differing only in the degree and timing of the response. Thus, the earlier and higher accumulation of PR proteins and mRNA was detected when plants were inoculated with the avirulent strain.

6.3. *Embryogenesis, seed maturation and germination*

Embryogenesis, seed maturation and germination are the developmental processes most often studied by proteomics, with reports published on rice [203], *Arabidopsis* [204], *M. truncatula* [205], maize [206], rapeseed [10] and *Fagus sylvatica* [207]. Using a comparative proteomic approach we analysed changes in the protein profile in date palm zygotic embryos during development, maturation, and germination (Sghaier et al., 2008, Proteomics, in press). The workflow used consisted of protein extraction, 2-DE, image analysis, and differential spot identification by MS. At the analysed stages (7), up to 194 spots showed qualitative or quantitative differences. By performing a multivariate analysis such as PCA, we were able to group samples and determine the most discriminant spots, and we recommend this approach for other proteomic studies. In addition, hierarchical clustering analyses can be performed to reveal the existence of groups of proteins showing similar evolution patterns [206].

6.4. Signalling

The series of papers by Wang and coworkers on brassinosteroid (BR) responses and signalling in *Arabidopsis* is an excellent example of the use of different approaches to validate data and gain a better understanding of key biological processes. Also, they cover different proteomic areas: descriptive, PTM and Interactomics. The proteomic study of BR-regulated proteins was performed using DIGE and LC-MS/MS [208]. A total of 42 BR-regulated proteins were identified. Analyses of the BR-insensitive mutant *bri1-116* and BR-hypersensitive mutant *bzr1-1D* led to the identification of five proteins (PATL1, PATL2, THI1, AtMDAR3 and NADP-ME2) affected by BR treatment in the mutants, suggesting their importance in BR action. Selected proteins were further studied using insertion knockout mutants or immunoblotting assays. Interestingly, approximately 80% of the BR-responsive proteins had not been identified in previous microarray studies, and direct comparison of protein and RNA changes in BR mutants revealed extremely weak correlation. RT-PCR analysis of selected genes revealed gene-specific kinetic relationships between changes in RNA and protein levels. Furthermore, BR-regulated Posttranslational Modifications of BiP2 protein were detected based on spot shifts in 2-D DIGE.

Proteomics has also been used to identify elements of the brassinosteroid signal transduction cascade that interact with BRl1, the major membrane receptor-like kinase of BRs [209]. Seedlings of the BR-deficient mutant *det-21* were treated with brassinolide and total or plasma membrane protein preparations were subjected to DIGE. Mass spectrometry analysis of differential abundant control- and brassinolide-treated seedling proteins allowed the identification of two BR-signalling kinases, BSK1 and 2, which are members of the receptor-like cytoplasmic kinase subfamily RLCK-XII. In response to BR, these two proteins are phosphorylated and shift in position from the basic to the acidic part of the gel. Proteomic data in that study were validated using transgenic plants expressing a fusion protein of BSK1 with yellow fluorescent protein (YFP). The analysis of the *bri1-5* mutant showed that BSK phosphorylation is BRl1-dependent. A sequence analysis revealed the existence of putative N-terminal myristoylation sites that target them to the membrane, as shown by confocal microscopy of BSK1-YFP. MS analysis revealed Ser^{230} to be the phosphorylation site, both *in vitro* and *in vivo*. *In vivo* interaction of BRI1 and BSK1 was demonstrated using bimolecular fluorescence complementation and co-immunoprecipitation assays. T-DNA insertion mutants and transgenic plants overexpressing BSK3 were used to determine the function of BSKs.

7. PTMs

7.1. Phospho and, to some extent, the redox proteomes

Mass spectrometry is the key technology for detecting, mapping and quantifying the great array of Posttranslational Modifications occurring on proteins, with as many as 300 being reported [17]. Such chemical modifications determine the activity status, localisation, turnover and interactions of proteins. The biological motivation for studying these modifications is their relevance for modulating and determining nearly all aspects of cell biology. MS strategies developed and successfully used for large-scale analysis of PTMs in other biological systems laid the foundation for plant biology studies, but only a few of the available protocols have been used to explore PTMs in plant systems, and most of these modifications remain unexplored.

In this section we discuss the recent advances done in this area, focusing on protein phosphorylation, the most extensively studied PTM, as well as on redox modifications. Some reviews on these topics have appeared recently [187]. Except for two papers using maize and *S. chacoense* [72], a relative of tomato, all the published works deal with model systems: *Arabidopsis* (15), rice (3) and *M. truncatula* (1). Most of the papers are descriptive, while some of them study changes in the PTM pattern in response to environmental stress [210,211].

7.2. Phosphoproteome

The current challenges in Phosphoproteomics can be summarised as the following three. First, efficient enrichment of phosphoproteins/phosphopeptides is needed due to the low abundance of such regulatory phosphoproteins, which is too low for the sensitivity of most mass spectrometers. Second, quantitative analysis is necessary because reversible protein phosphorylation/dephosphorylation events control many physiological processes. Third, plant phosphoprotein databases are needed in order to integrate current knowledge on plant phosphoproteins and bioinformatic tools to predict phosphorylation motifs. The papers described below address some of these issues using a wide variety of biochemical and analytical chemical approaches to enrich, detect and characterize protein and peptide phosphorylation on a proteome-scale. These techniques have advantages and drawbacks, but the overall conclusion is that a combination of all of them would increase the coverage of the proteome studied and the knowledge of its dynamics and implications for plant physiology.

Large-scale identification of *in vivo* phosphorylation sites by MS is possible as a result of the development of different techniques for the enrichment of phosphopeptides and phosphoproteins [212–215]. This achievement, together with the use of highly sensitive and accurate mass spectrometers, has led to the identification and mapping of thousands of phosphopeptides. Using Fe-IMAC and HAMMOC prior to the nLC-MS analysis, Sugiyama et al. [212] made possible the identification of 2172 phosphorylation sites in 1346 *Arabidopsis* proteins, stressing the relevance of tyrosine phosphorylation in plant physiology. Tan et al. [214] use SCX-RP coupled with ESI IT to identify 269 phosphoproteins in rice that are associated with chromatin, including transcription factors and histone modification proteins. Organelle purification [215,216] and prefractionation of protein or peptide samples are necessary in order to increase the coverage of the phosphoproteome. In addition, a number of chemical derivatisation strategies, in which the phosphate group can be replaced by a more stable moiety, have been developed. Using a Ser- and Thr-biotin-tagged approach, Kwon et al. [217] identified 31 protein spots from RuBisCO-depleted fractions from *Arabidopsis* seedlings.

The work by Meimoun et al. [218] describes the use of a commercial phosphoprotein affinity matrix (Qiagen), and this was tested on sorghum and *Arabidopsis* extracts. Phosphoenolpyruvate carboxylase (PEPC) enzymatic assays and Western blots, as well as 3-D mass spectrometry analysis of immunoprecipitated PEPC showed that the column efficiently binds P-PEPC with no contamination by non-P-PEPC. In fact, the column captured 80% of the proteins labelled *in vivo* using ^{32}P-phosphate, and the majority were subsequently found in the elution fraction (88%). This was also visualised by SDS-PAGE (1D and 2D gels), followed by Pro-Q diamond staining. Analysis of the P-protein fraction by 1-D gels and liquid chromatography/tandem mass spectrometry allowed the identification of 250 proteins with a variety of functions.

One of the papers reviewed analyses the proteome and phosphoproteome during ovule fertilisation in *S. chacoense*. It compares the efficiency and sensitivity of different methods for detection and purification of phosphopeptides and proteins: immunodetection, *in vivo* labelling and phosphoprotein-specific staining [72]. That study concludes that there is a small overlap among the three methods, which is in agreement with the idea that different methods must be combined to achieve complete coverage of the proteome.

Phospho Proteomics in plants is moving from being descriptive to quantitative. Concerning mass spectrometry-based quantitative Phospho Proteomics, we highlight two recent papers, both of which study the dynamics of *Arabidopsis* protein phosphorylation in response to the microbial elicitor fla22, but they use different quantitative strategies [215,219]. Benschop et al. used ^{14}N/^{15}N metabolic labelling to compare phosphorylation levels of plasma membrane proteins from treated and control cells, and the quantitative information is acquired at the MS stage, while Nushe et al. obtain quantitative data at the MS2 stage using iTRAQ labelling. Interestingly, both studies show excellent agreement in the proteins they identify as being involved in the regulation of many of the flagellin-induced phosphorylation sites, including kinases and regulatory proteins potentially involved in defence. The differences in the number of identifications may reflect the different equipment used and the statistical analyses performed. These two papers are an encouraging example of how different approaches to Quantitative Phospho Proteomics can lead to similar results. In addition, these results emphasise the potential of proteomic technology in quantifying the dynamics of phosphorylation, in which case it would join other large-scale genomic approaches that serve as valuable tools in determining regulatory mechanisms.

Recently, Kruger et al. [220] presented an alternative label-free method for Quantitative Phospho Proteomics studies that allows the determination of phosphorylation sites and phosphorylation stoichiometries. The method couples capillary liquid chromatography (capLC) with inductively coupled plasma-mass spectrometry (ICP-MS) to perform quantitative phosphodetermination of protein extracts based on their phosphorus content. As an internal control for protein amount, the sulphur content is also determined. The protein phosphorylation stoichiometry can be determined via the phosphorus to sulphur ratio, given that the number of amino acids containing sulphur (cysteine and methionine) is known. Applying this method to *Arabidopsis* and *C. reinhardtii* has enabled description of their global phosphorylation states, indicating differences in the average protein phosphorylation levels between organisms. In addition, it shows differences in the proteomic phosphorylation patterns and stoichiometry of *Arabidopsis* in a tissue- and development-dependent manner, indicating a different adjustment of the kinase/phosphatase system in each specific situation.

A recent review has summarised the large-scale mapping of *in vivo* phosphorylation sites using a combination of mass spectrometry-based techniques and protein chip-based methods, and their potential to unravel plant signal transduction pathways [221]. De la Fuente van Bentem et al. [213] used this methodology to unravel phosphoproteome dynamics in control and H_2O_2 treated-*Arabidopsis* cells by examining cytoplasmic and nuclear fractions. This analysis led to the identification of 303 *in vivo* phosphorylation sites, including 21 different protein kinases. Validation of some of these results was based on immunoblotting and mutational analysis. In addition, quantitative analysis of the phosphorylation status of proteins (kinome profiling) was estimated using peptide arrays generated using peptides corresponding to *in vivo* phosphorylation sites. The investigators concluded that different protein kinases operate in cytosolic and nuclear compartments.

These protein/peptide array experiments provide a large-scale inventory of the site specificity of multiple protein kinases, revealing specific phosphomotifs for different kinases [213]. To facilitate the experimental validation of kinase-substrate pairs identified by MS, it would be useful to have software applications that integrate all the information available and enable prediction of upstream kinases for substrate phosphorylation sites based on data from plant systems. Along these lines, Heazlewood et al. [36] have developed a method for the prediction of phosphorylation sites in *Arabidopsis*. They describe the *Arabidopsis* Protein Phosphorylation Site Database (PhosPhAt) (http://phosphat.mpimpgolm.mpg.de) as a valuable resource for plant researchers since it integrates current knowledge of MS-based identified phosphorylation sites in *Arabidopsis* and site prediction based on phosphorylation motifs. They used 802 experimentally validated serine phosphorylation sites to develop prediction software for serine phosphorylation, resulting in 27,782 predicted sites in 17,035 proteins. This information is accessible for each peptide, together with the experimental conditions such as tissue sampled and the phosphopeptide enrichment method. An integrative approach combining large-scale mass spectrometry-based mapping of phosphorylation sites with protein and peptide chip analysis and bioinformatic predictive tools, will reveal protein kinases and phosphatases, their substrates and their positions within signalling webs.

7.3. *Redox proteome*

Although it has not been studied nearly so extensively as the phosphoproteome, the number of plant redox proteome studies has increased over the past few years. This is due mainly to the continuous development of new, more sensitive chemical derivatisation reactions, affinity purification media, specific detection reagents and the use of highly sensitive mass spectrometers that allow direct detection of such modifications [222–225]. A recent review [226] summarises the progress made on the study of redox regulation from a systems biology

perspective, and it illustrates the central role of redox regulation in plant networks. We discuss below some of the recent progress made in describing the thiol/disulfide proteome, the redox-dependent and cross-talking signalling pathways, and the target genes/proteins of redox regulation. The specificity of these modifications is an intriguing issue that is still far from being understood.

Two studies use 2-DE oxidant/reductant diagonal-SDS-PAGE to analyse the dynamics of the thiol–disulphide redox proteome from mitochondrial [227] and chloroplast fractions [228]. The mitochondrial analysis resulted in the identification of 18 proteins, from both the soluble and membrane fractions, among them known glutaredoxin/thioredoxin targets. In addition, comparison of the identified protein sequences with homologues from other species identified specific Cys residues that may be responsible for plant-specific redox modulations of mitochondrial proteins [227]. The analysis of the chloroplast fractions (thylakoid, luminal and RuBisCO-depleted stroma) resulted in the identification of 22 novel proteins, not previously reported as being part of the redox proteome [228].

For several years Buchanan and coworkers have investigated the central role of the regulatory disulfide protein thioredoxin (Trx) in various aspects of cereal physiology, allowing them to identify potential Trx targets [229]. They have extended such studies to dicot species, and they have analysed the redox state of seed proteins during germination in *M. truncatula* [92].

After one decade of investigating nitric oxide (NO) functions in plant physiology, it is well established that NO modulates the activity of proteins through nitrosylation and probably tyrosine nitration (reviewed in Besson-Bard et al. [230]) [229]. Lindermayr et al. (2005) reported in *Arabidopsis* the first study and catalogue of nitrosylated proteins in plants, using the "biotin switch" combined with LC-MS, and this paper was singled out as "high impact" in Plant Physiology Journal. Using this approach, Romero-Puertas et al. [231] analysed changes in S-nitrosylated proteins in *Arabidopsis* during the hypersensitive disease resistance response. This analysis identified 16 proteins belonging to different functional categories such as metabolic enzymes, signalling and antioxidant defence. We have used biotin switch-affinity chromatography but we have coupled it to a nano-HPLC-LTQ mass spectrometer to identify protein targets of S-nitrosylation in both *Arabidopsis* leaves and cell suspension cultures during compatible and incompatible interaction with the bacteria *Pseudomonas syringae*. This analysis resulted in the identification of nearly 200 candidate proteins, some of which had already been described in *Arabidopsis* and in animal systems in the context of redox regulation. Among the proteins identified, we found some involved in defence and stress-related responses, redox-related proteins, cytoskeleton proteins, metabolic enzymes and signalling/regulating proteins (Maldonado et al., unpublished).

The experimental evidence for protein functional regulation and the identification of the residues involved are critical to understanding the molecular mechanisms by which NO exerts its action. This should be approached in a step-by-step manner [231,232]. For example, the molecular mechanism for S-nitrosylation of peroxiredoxin II was investigated using biochemical and reverse genetics, revealing a novel regulatory mechanism for peroxiredoxins in which NO modification inhibits the hydroperoxide-reducing peroxidase activity and the $ONOO^-$ detoxification activity of PrxII E. This system regulates the effects of its own radicals through the S-nitrosylation of crucial components of the antioxidant defence [231]. Reported studies evidenced the need for sufficiently sensitive methods for detecting and identifying low levels of S-nitrosylated proteins in complex protein mixtures in order to fully appreciate the range, extent and selectivity of this modification under both physiological and pathological conditions. However this has been limited by the lack of rapid and accurate methods for the detection of these S-nitrosylated proteins and the exact modification sites. In fact, the identification of the Cys residue involved is critical to understanding the biological significance of such modification on modulating protein function, and it will give clues into NO signal specificity. Recent papers have reviewed the methodology available for the proteomic analysis of protein S-nitrosylation, and they have described new chemical approaches to detect and purify S-nitrosothiols [222–224,233]. These advances have been developed in other biological systems, and we hope it is only a matter of time before they can be used in plants. For example, Han and Chen [223] describe an improvement in the classical biotin switch method using a urea-based, detergent-free protocol combined with LC-MS/MS that allows simultaneous identification of S-nitrosylated sites and their cognate proteins in S-nitrosoglutathione(GSNO)-treated HeLa cell extracts. This considerably reduces the amount of starting material needed. The use of extracted ion chromatography (XIC) enables quantitative comparison of S-nitrosylation levels between control and treated samples [223]. Camerini et al. [222] reported the "His-tag switch", a novel strategy for the purification and identification of S-nitrosylated proteins that has the added advantage that it identifies the cysteine residues undergoing S-nitrosylation. Pursuing the same objective, Han et al. [233] developed the "AMCA switch method" that involves labelling the S-nitrosylated cysteines with the fluorophore 7-amino-4-methylcoumarin-3-acetic acid (AMCA). This label allows in-gel detection of S-nitrosylated proteins after UV exposure, and identification of modified residues by liquid chromatography-tandem mass spectrometry (LC-MS/MS) using the AMCA tag signal in the MS spectra.

Even though S-glutathionylation, the reversible formation of mixed disulfides between glutathione and cysteinyl residues, constitutes an important redox-mechanism for dynamic, posttranslational regulation of a broad range of proteins, no high-throughput analysis of this modification has appeared during the review period. Dalle-Donne et al. have summarised the current knowledge on S-glutathionylation, providing a list of S-glutathionylated proteins in both animal and plant systems, as well as the proteomic and analytical methods available for its study [234]. We are using a similar approach to that used for S-nitrosothiols to identify proteins that can be glutathionylated in *Arabidopsis* leaves, using the biotin moiety GSSG-biotin to tag the proteins followed by streptavidin-agarose capture.

Proteomic technology has led to an understanding of how key enzymes are regulated at the molecular level and has

allowed the assignment of novel additional functions to well-known enzymes. A good example is the glycolytic enzyme glyceraldehyde 3-P dehydrogenase (GAPDH). The presence of two critical Cys residues in the active site of the enzyme makes it susceptible to reversible modulation by glutathionylation and nitrosylation. Combining MS analysis and analyses of mutant and GFP-recombinant proteins, Holtgrefe et al. [235] reported the occurrence of such modifications in purified GADPH, its translocation to the nucleus and its DNA-binding properties. Based on the data obtained the authors suggest that this enzyme is part of the system that protects the cell against ROS and RNS.

7.4. *Other PTMs*

Most of the other PTMs have yet to be studied so far, despite the existence of well-established protocols [236,237]. The study of ubiquitination has emerged as one of the most active areas in proteomic research. The development of a GST-tagged approach for affinity purification of ubiquitinated proteins has allowed the large-scale analysis of the ubiquitinated proteome in *Arabidopsis* cell suspension culture and its identification using MudPIT technology. In all, 294 proteins have been identified, and 85 ubiquitinated lysine residues have been confirmed in 56 proteins [88]. This will mean a significant advance in understanding the role of ubiquitination in the functional regulation and degradation of many classes of proteins. A variety of mass spectrometric methods has been described for the identification of ubiquitination sites and of ubiquitin substrates, although they still have not been tested on plant systems [238–240].

Minic et al. [89] reported a global view of *Arabidopsis* cell wall N-glycosylated proteins by combining affinity chromatography on Concanavalin A Sepharose, 2-DE and nanoHPLC/MS/MS and MALDI-TOF/MS. These analyses resulted in the identification of 102 glycoproteins and, interestingly, *in silico* predictions that they localise to the secretory pathway.

Casati et al. [90] have studied changes in the maize nuclear proteome induced by UV using genotypes with a different UV tolerance. Differential accumulation of chromatin proteins, particularly histones, constituted the largest class identified among the responsive proteins. UV-B-tolerant lines showed a greater acetylation on the N-terminal tails of histones H3 and H4 after irradiation. These acetylated histones are enriched in the promoter and transcribed regions of the two UV-B-upregulated genes examined.

Most proteins in all organisms undergo crucial N-terminal modifications involving N-terminal methionine excision, N-α-acetylation, N-myristoylation, or S-palmitoylation. The occurrence of these poorly annotated modifications has been investigated in *Arabidopsis*. In the past, experimental data for the N-terminal sequences of animal, fungi, and archaeal proteins were used to build predictive tools, and *in vitro* N-myristoylation was performed using both plant and animal N-myristoyltransferases. N-terminal modifications from the sequenced genome of *Arabidopsis* were determined by MS, resulting in the identification of 105 new modified protein N-termini. Proteins that had undergone both N-terminal methionine (Met) cleavage and N-acetylation were found to be strongly overrepresented among the most abundant proteins, in contrast to those retaining their original, unblocked Met [241].

Since all PTMs regulate signalling events, data for each of these studies, whether phosphoproteomic, redox proteomic, or others, should be integrated in order to achieve a comprehensive view of systems biology that makes experimentally testable predictions. Finally, protein–protein interaction screens, gene expression profiling, and mutant screens and analyses, are still necessary to validate those data unequivocally.

8. Interactomics

8.1. *The major challenge of Plant Proteomics*

A key step in the advance from genomics through functional genomics towards systems biology is the definition of protein interactions in living cells [242], which is one of the main challenges in plant research for the next few years [243]. Berggard et al. [244] and Miernyk and Thelen [245] have recently reviewed methods for studying protein interactions. Five biochemical approaches can be used for isolating and characterizing *in vitro* protein complexes: co-immunoprecipitation, blue native gel electrophoresis, *in vitro* binding assays, protein cross-linking, microarrays, and rate-zonal centrifugation. These techniques have been rarely used in plant research, with only four reports found during the past 18 months.

Analysis of thioredoxin (Trx) targets is a classic in Interactomic studies. In cereals, seed germination is accompanied by extensive changes in the redox state of seed proteins, with a conversion of oxidised proteins of dry seeds to a reduced form after imbibition; Trx appears to play a key role in this process [246]. Trx targets have been characterized in *M. truncatula* seeds [92] using two complementary gel-based approaches. Proteins were (1) labelled with a thiol-specific probe, monobromobimane (mBBr), following *in vitro* reduction by an NADP/Trx system, or (2) isolated on a mutant Trx affinity column. Altogether, 111 Trx-linked proteins were identified after 1- or 2-DE/ESI-Q-TOF. The X! TANDEM software was used to match MS/MS spectra against all plant proteins in the transcript assembly of *M. truncatula* (MtGI), all plant sequences in the NCBI database, and the common Repository of Adventitious Proteins known to occur as unavoidable contamination. Identified proteins function in major processes of the seed, including metabolism, cell structure, protein biogenesis and degradation, storage proteins, binding proteins, response to stress, signal transduction, as well as unknown processes.

Calmodulin (CaM) is one of the best-studied calcium sensors, which react by binding free Ca and initiating multiple physiological responses. Popescu et al. reported the use of protein microarrays to detect calmodulin targets [93]. *Arabidopsis* clones were generated containing 1133 ORFs fused to a tandem affinity purification (9xMYC epitope, His-6, a rhinovirus 3C protease cleavage site, and the 2× IgG binding domain of protein A). Clones were transformed into *Agrobacterium tumefaciens*, and transformants were used to infiltrate *Nicotiana benthamiana*. The 1133 proteins represented putative and known protein kinases (404), transcription factors (291), protein degradation-related proteins (113), heat-shock proteins (63), cytochrome P-450 (58), CaMs and CMs like (51), RNA-

binding proteins (35), ATP/GTP binding proteins (10), and proteins with unknown functions (108). Each protein was arrayed onto FAST slides (Schleicher & Schuell, Keene, NH), and arrays incubated with three CaMs and four CMs-like antibodies, scanned. Images were processed and analysed with Matlab, to obtain protein–protein interaction data. The array proved the existence of more than 173 novel *in vitro* binding partners. Analysis of these targets revealed remarkable divergence in the binding of many of the CaMs/CMs-like, with each protein binding to unique targets. This study demonstrates the feasibility of the array technology for characterization of the interactome.

Over 150 target proteins of five 14-3-3 isoforms have been described in barley using yeast two-hybrid screens and proteomics, with 10% of the identified genes having been previously reported [94]. Protein targets were affinity-purified using His-tagged 14-3-3 recombinant proteins, separated by SDS-PAGE, and gel slices were digested with trypsin. Peptides were subjected to nLC-ESI-MS/MS (Q-TOF) and spectral data were used to search against nrNCBI using MASCOT. Both approaches showed differences in the identified proteins: the two-hybrid and affinity purification strategies identified 132 and 30 interactors, respectively, with an overlap of only 10 proteins. More than one-third of the proteins identified in the two-hybrid system belonged to the signalling functional group, while it represented only 10% of the proteins identified using proteomics, with metabolic enzymes the most abundant (42%). Carbonic anhydrase and six enzymes of the Calvin cycle, as well as enzymes of the sucrose-glycolysis pathway, were shown to be 14-3-3 targets. A number of proteins with a well-described function in hormonal signalling were identified using the two-hybrid system, but not by affinity purification.

Abe et al. [95] reported the use of tandem affinity purification to isolate proteins that interact with rice OsG1, which is an ortholog of GIGANTEA that regulates photoperiod flowering in *Arabidopsis*. Seven proteins, including dynamin, were identified as OsGI-interacting proteins. The interaction of OsGI with dynamin was verified by co-immunoprecipitation using a myc-tagged version of OsGI.

Finally, the paper by Luthje et al. included in this issue, is one more example of the use of Proteomics to characterize protein complexes-in this case, the plant plasma membrane redox complex.

9. Proteinomics

9.1. *Proteomics is more than high throughput*

Proteomic techniques are used to characterize a specific protein or a structural or functional group of proteins. This is what we can call “Hypothesis-driven Proteomics”, “Targeted Proteomics”, or “Proteinomics.” This type of study, in which Proteomics is merely part of a multiapproach strategy in combination with microscopy, Genomics, Transcriptomics, Metabolomics, and classical biochemical techniques, provides relevant information on protein structure and function, isoforms, organs, cells and subcellular location and trafficking, processing, signal peptides, PTMs, expression kinetics and correlation with RNA and metabolites. At the same time it is a method to validate data obtained using one specific approach. Proteomic platforms have been used to investigate specific proteins: seed storage [11,122], chlorophyllase [247], glucanase [248], α-amylases [249], vacuolar sorting receptors [250], esterase [251], xylanase-inhibitors [252], and peroxidases [253]. Such platforms have also been used to investigate processes, such as mRNA processing and degradation [254], and proteolysis [255]. The stability and degradation of mRNA have been studied in chloroplasts from light-grown and dark-adapted spinach plants using biochemical approaches, including RNA processing/degradation, UV-crosslinking, TLC and polyribonucleotide phosphorylase assays; as well as proteomic approaches [254]. Soluble protein extracts that correctly reproduced *in vitro* the differential mRNA stability observed *in vivo* were prepared from chloroplasts and separated by DEAE and Mono Q ion exchange chromatography. This purification strategy was designed to enrich proteins that bind nucleic acids, including those involved in transcription and translation. Protein fractions were subjected to SDS-PAGE and MS analysis after tryptic digestion (LC-ESI-MS/MS). Tandem mass spectra were used to search the non-redundant NCBI database using the SEQUEST algorithm and a QUALSCORE tool developed in-house. In total, 234 proteins were identified, a number of which had not previously been reported in other plastid proteome studies. The 234 proteins included proteins involved in RNA stability, such as nucleases, RNA-binding proteins, and ATP-dependent RNA helicase. They also included proteins involved in transcription and translation, and in metabolism (photosynthesis, energy, amino acid, redox). On the basis of the proteins identified and the *in vitro* characterization of the RNA degradation, the investigators proposed the existence of two different pathways that determine the fate of mRNA: a processing and stabilisation pathway and a degradation pathway. Shifting light-grown spinach plants to darkness for 48 h induces the mRNA degradation pathway, which is inactive or less active under normal light conditions [254].

10. Concluding remarks

In the current scientific scenery Proteomics must be understood as part of a multidisciplinary approach. A combination of high throughput -omics, and classical biochemical and cellular biology techniques should be used for data validation and to deepen in the knowledge of living organisms.

Despite the continuous development and improvement of powerful proteomic techniques, protocols, equipments and bioinformatic tools, just a minimal fraction of the cell proteome, and for only a few organisms, has been characterized so far. This is mainly related to the enormous diversity and complexity of proteomes, and to technical limitations in quantitation, sensitivity, resolution, speed of data capture and analysis. Some recalcitrant proteomes, for example membrane proteins or other highly hydrophobic proteins, remain elusive. Nevertheless, some questions are starting to be answered, including the potential number of protein species per gene as a result of Postranscriptional and Posttranslational Modifications, protein trafficking and interactions events.

Exploiting the full potential of Proteomics surpasses the possibilities of individual laboratories and any “Green

Proteome Initiative" will require large-scale transnational collaborations. In this respect, the Proteomics Subcommittee of the Multinational *Arabidopsis* Steering Committee has begun to operate in an attempt to coordinate International Proteomics Research in *Arabidopsis*.

After nearly 10 years of Proteomics research, looking back at previous publications it is possible to identify errors derived from incorrect experimental design, data analysis and interpretation. In addition to that, proteomic data should be validated in order to go beyond description or speculation. It is not uncommon to find in the literature low-confidence protein identification (especially in the case of unsequenced organisms, one of the main challenges of Proteomics), and inappropriate statistical analyses of the results. At this respect HUPO's Proteomic Standard Initiative has developed the MIAPE documents. The establishment of repositories containing MS/MS reference spectra will be very useful and will contribute to facilitate protein identification and quantification via a genome-independent approach, especially in the case of orphan species.

The use of Proteomics in plant biology research has increased significantly in the period from 2007 to September 2008 (380 reports at the ISI Web of Knowledge), becoming a routine methodology in a number of plant laboratories worldwide. During this period, both qualitative and quantitative improvements in Plant Proteomics have occurred, ushering in a new phase, namely "Second-Generation Plant Proteomics", in which quantitative and gel-free, proteomic techniques have started to be used. However, following the general rule for plant biology research, progress in Plant Proteomics continues to lag behind that of Human and Yeast Proteomics. During the revised period there have not been large differences with respect to the previously revised ones concerning the plant species under investigation, being *Arabidopsis* and rice the subject of 75% of the original papers published.

In the present update, a total of 152 original papers have been reviewed, distributed in the following areas: Descriptive Proteomics (13), Subcellular Proteomic (11), Differential Expression Proteomics of developmental processes (18), responses to biotic (19) and abiotic stresses (31), hormones (6), plant genotypes or populations, (14), PTMs- phosphoproteome (9), -redox proteome (5), -ubiquitination (1), -glycosylation (1), -acetylation (1), -N-terminal modification, and protein sorting (1), Interactomics (5). The last two areas constitute the main challenge of Plant Proteomics in the near future. Plant Proteomics is beginning to make some practical contributions to applied fields including biomedicine, through the identification and characterization of allergens agronomy, through studies of the equivalence of transgenic crops, genotyping, studies of heterosis and food science and through studies of food quality control and traceability.

Acknowledgements

Thanks to the Spanish "Ministerio de Ciencia e Innovacion" (Project BIO-2006-14790), a la "Junta de Andalucía" y "Universidad de Córdoba" for financial support.

REFERENCES

[1] Jorrin JV, Maldonado AM, Castillejo MA. Plant proteome analysis: a 2006 update. Proteomics 2007;7:2947–62.
[2] Rossignol M, Peltier JB, Mock HP, Matros A, Maldonado AM, Jorrin JV. Plant proteome analysis: a 2004–2006 update. Proteomics 2006;6:5529–48.
[3] Canovas FM, Dumas-Gaudot E, Recorbet G, Jorrin J, Mock HP, Rossignol M. Plant proteome analysis. Proteomics 2004;4:285–98.
[4] Bruggeman FJ, Westerhoff HV. The nature of systems biology. Trends Microbiol 2007;15:45–50.
[5] Weckwerth W. Integration of metabolomics and proteomics in molecular plant physiology—coping with the complexity by data-dimensionality reduction. Physiol Plant 2008;132:176–89.
[6] Paiva JA, Garces M, Alves A, Garnier-Gere P, Rodrigues JC, Lalanne C, et al. Molecular and phenotypic profiling from the base to the crown in maritime pine wood-forming tissue. New Phytol 2008;178:283–301.
[7] Jaquinod M, Villiers F, Kieffer-Jaquinod S, Hugouvieux V, Bruley C, Garin J, et al. A Proteomics dissection of *Arabidopsis thaliana* vacuoles isolated from cell culture. Mol Cell Proteomics 2007;6:394–412.
[8] Lu P, Vogel C, Wang R, Yao X, Marcotte EM. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. Nat Biotechnol 2007;25:117–24.
[9] Nie L, Wu G, Culley DE, Scholten JC, Zhang W. Integrative analysis of transcriptomic and proteomic data: challenges, solutions and applications. Crit Rev Biotechnol 2007;27:63–75.
[10] Joosen R, Cordewener J, Supena ED, Vorst O, Lammers M, Maliepaard C, et al. Combined transcriptome and proteome analysis identifies pathways and markers associated with the establishment of rapeseed microspore-derived embryo development. Plant Physiol 2007;144:155–72.
[11] Li Q, Wang BC, Xu Y, Zhu YX. Systematic studies of 12S seed storage protein accumulation and degradation patterns during *Arabidopsis* seed maturation and early seedling germination stages. J Biochem Mol Biol 2007;40:373–81.
[12] Cravatt BF, Simon GM, Yates III JR. The biological impact of mass-spectrometry-based proteomics. Nature 2007;450:991–1000.
[13] Carpentier SC, Panis B, Vertommen A, Swennen R, Sergeant K, Renaut J, et al. Proteome analysis of non-model plants: a challenging but powerful approach. Mass Spectrom Rev 2008;27:354–77.
[14] Han X, Jin M, Breuker K, McLafferty FW. Extending top-down mass spectrometry to proteins with masses greater than 200 kilodaltons. Science 2006;314:109–12.
[15] Cox J, Mann M. Is proteomics the new genomics? Cell 2007;130:395–8.
[16] Kim E, Magen A, Ast G. Different levels of alternative splicing among eukaryotes. Nucl Acids Res 2007;35:125–31.
[17] Witze ES, Old WM, Resing KA, Ahn NG. Mapping protein post-translational modifications with mass spectrometry. Nat Methods 2007;4:798–806.
[18] Barbazuk WB, Fu Y, McGinnis KM. Genome-wide analyses of alternative splicing in plants: opportunities and challenges. Genome Res 2008;18:1381–92.
[19] Simpson CG, Lewandowska D, Fuller J, Maronova M, Kalyna M, Davidson D, et al. Alternative splicing in plants. Biochem Soc Trans 2008;036:508–10.
[20] Palusa SG, Ali GS, Reddy ASN. Alternative splicing of pre-mRNAs of *Arabidopsis* serine/arginine-rich proteins:

regulation by hormones and stresses. Plant J 2007;49:1091–107.
[21] Thatcher LF, Carrie C, Andersson CR, Sivasithamparam K, Whelan J, Singh KB. Differential gene expression and subcellular targeting of *Arabidopsis* glutathione S-transferase F8 is achieved through alternative transcription start sites. J Biol Chem 2007;282:28915–28.
[22] Dinkins RD, Majee SM, Nayak NR, Martin D, Xu Q, Belcastro MP, et al. Changing transcriptional initiation sites and alternative 5- and 3-splice site selection of the first intron deploys *Arabidopsis* protein isoaspartyl methyltransferase2 variants to different subcellular compartments. Plant J 2008;55:1–13.
[23] Sun Q, Zybailov B, Majeran W, Friso G, Olinares PD, van Wijk KJ. PPDB, the database at Cornell. Nucleic Acids Res 2009;37: D969–74.
[24] Mano S, Miwa T, Nishikawa S, Mimura T, Nishimura M. The plant organelles database (PODB): a collection of visualized plant organelles and protocols for plant organelle research. Nucleic Acids Res 2008;36:D929–937.
[25] Dowsey AW, Guang-Zhong Y. The future of large-scale collaborative Proteomics. Proc IEEE 2008;96:1292–309.
[26] Brunner E, Ahrens CH, Mohanty S, Baetschmann H, Loevenich S, Potthast F, et al. A high-quality catalog of the *Drosophila melanogaster* proteome. Nat Biotechnol 2007;25:576–83.
[27] Weckwerth W, Baginsky S, van Wijk K, Heazlewood JL, Millar H. The multinational *Arabidopsis* steering subcommittee for Proteomics assembles the largest proteome database resource for plant systems biology. J Proteome Res 2008;7:4209–10.
[28] Hummel J, Niemann M, Wienkoop S, Schulze W, Steinhauser D, Selbig J, et al. ProMEX: a mass spectral reference database for proteins and protein phosphorylation sites. BMC Bioinformatics 2007;8:216.
[29] Bini L, Pallini V, Hochstrasser DF, Sanchez J-C. From genome to proteome: back to the future. Proteomics 2007;7:1561–3.
[30] Nesvizhskii AI, Vitek O, Aebersold R. Analysis and validation of proteomic data generated by tandem mass spectrometry. Nat Methods 2007;4:787–97.
[31] Taylor CF, Paton NW, Lilley KS, Binz PA, Julian Jr RK, Jones AR, et al. The minimum information about a proteomics experiment (MIAPE). Nat Biotechnol 2007;25:887–93.
[32] Mead JA, Shadforth IP, Bessant C. Public proteomic MS repositories and pipelines: available tools and biological applications. Proteomics 2007;7:2769–86.
[33] Deutsch EW, Lam H, Aebersold R. PeptideAtlas: a resource for target selection for emerging targeted proteomics workflows. EMBO Rep 2008;9:429–34.
[34] Lam H, Deutsch EW, Eddes JS, Eng JK, Stein SE, Aebersold R. Building consensus spectral libraries for peptide identification in proteomics. Nat Methods 2008;5:873–5.
[35] Falth M, Savitski MM, Nielsen ML, Kjeldsen F, Andren PE, Zubarev RA. SwedCAD, a database of annotated high-mass accuracy MS/MS spectra of tryptic peptides. J Proteome Res 2007;6:4063–7.
[36] Heazlewood JL, Durek P, Hummel J, Selbig J, Weckwerth W, Walther D, et al. PhosPhAt: a database of phosphorylation sites in *Arabidopsis thaliana* and a plant-specific phosphorylation site predictor. Nucleic Acids Res 2008;36:D1015–1021.
[37] Baerenfaller K, Grossmann J, Grobei MA, Hull R, Hirsch-Hoffmann M, Yalovsky S, et al. Genome-scale proteomics reveals *Arabidopsis thaliana* gene models and proteome dynamics. Science 2008;320:938–41.
[38] Dunkley TPJ, Hester S, Shadforth IP, Runions J, Weimar T, Hanton SL, Griffin JL, Bessant C, Brandizzi F, Hawes C, Watson RB, Dupree P and Lilley KS. Mapping the *Arabidopsis* organelle proteome. Proc Natl Acad Sci USA 2006;103:6518-6523.
[39] Heazlewood JL, Verboom RE, Tonti-Filippini J, Small I, Millar AH. SUBA: the *Arabidopsis* Subcellular Database. Nucleic Acids Res 2007;35:D213–218.
[40] Zybailov B, Rutschow H, Friso G, Rudella A, Emanuelsson O, Sun Q, et al. Sorting signals, N-terminal modifications and abundance of the chloroplast proteome. PLoS ONE 2008;3: e1994.
[41] Lunn JE. Compartmentation in plant metabolism. J Exp Bot 2007;58:35–47.
[42] Jarvis P. Targeting of nucleus-encoded proteins to chloroplasts in plants. New Phytol 2008;179:257–85.
[43] Millar AH, Whelan J, Small I. Recent surprises in protein targeting to mitochondria and plastids. Curr Opin Plant Biol 2006;9:610–5.
[44] Gonzalez-Buitrago JM, Ferreira L, Isidoro-Garcia M, Sanz C, Lorente F, Davila I. Proteomic approaches for identifying new allergens and diagnosing allergic diseases. Clin Chim Acta 2007;385:21–7.
[45] Salekdeh GH, Komatsu S. Crop proteomics: aim at sustainable agriculture of tomorrow. Proteomics 2007;7:2976–96.
[46] Alm R, Ekefjard A, Krogh M, Hakkinen J, Emanuelsson C. Proteomic variation is as large within as between strawberry varieties. J Proteome Res 2007;6:3011–20.
[47] Hajduch M, Casteel JE, Tang S, Hearne LB, Knapp S, Thelen JJ. Proteomic analysis of near-isogenic sunflower varieties differing in seed oil traits. J Proteome Res 2007;6:3232–41.
[48] Hoecker N, Lamkemeyer T, Sarholz B, Paschold A, Fladerer C, Madlung J, et al. Analysis of nonadditive protein accumulation in young primary roots of a maize (*Zea mays* L.) F(1)-hybrid compared to its parental inbred lines. Proteomics 2008;8:3882–94.
[49] Saz JM, Marina ML. High performance liquid chromatography and capillary electrophoresis in the analysis of soybean proteins and peptides in foodstuffs. J Sep Sci 2007;30:431–51.
[50] Sotkovsky P, Hubalek M, Hernychova L, Novak P, Havranova M, Setinova I, et al. Proteomic analysis of wheat proteins recognized by IgE antibodies of allergic patients. Proteomics 2008;8:1677–91.
[51] Petersen A, Dresselhaus T, Grobe K, Becker WM. Proteome analysis of maize pollen for allergy-relevant components. Proteomics 2006;6:6317–25.
[52] Napoli A, Aiello D, Di Donna L, Moschidis P, Sindona G. Vegetable proteomics: the detection of Ole e 1 isoallergens by peptide matching of MALDI MS/MS spectra of underivatized and dansylated glycopeptides. J Proteome Res 2008;7:2723–32.
[53] Scossa F, Laudencia-Chingcuanco D, Anderson OD, Vensel WH, Lafiandra D, D'Ovidio R, et al. Comparative proteomic and transcriptional profiling of a bread wheat cultivar and its derived transgenic line overexpressing a low molecular weight glutenin subunit gene in the endosperm. Proteomics 2008;8:2948–66.
[54] Zolla L, Rinalducci S, Antonioli P, Righetti PG. Proteomics as a complementary tool for identifying unintended side effects occurring in transgenic maize seeds as a result of genetic modifications. J Proteome Res 2008;7:1850–61.
[55] Sauvage FX, Pradal M, Chatelet P, Tesniere C. Proteome changes in leaves from grapevine (*Vitis vinifera* L.) transformed for alcohol dehydrogenase activity. J Agric Food Chem 2007;55:2597–603.
[56] Prins A, van Heerden PD, Olmos E, Kunert KJ, Foyer CH. Cysteine proteinases regulate chloroplast protein content and composition in tobacco leaves: a model for dynamic interactions with ribulose-1,5-bisphosphate carboxylase/

oxygenase (Rubisco) vesicular bodies. J Exp Bot 2008;59:1935–50.
[57] Devouge V, Rogniaux H, Nesi N, Tessier D, Gueguen J, Larre C. Differential proteomic analysis of four near-isogenic *Brassica napus* varieties bred for their erucic acid and glucosinolate contents. J Proteome Res 2007;6:1342–53.
[58] Vincent D, Ergul A, Bohlman MC, Tattersall EA, Tillett RL, Wheatley MD, et al. Proteomic analysis reveals differences between *Vitis vinifera* L. cv. Chardonnay and cv. Cabernet Sauvignon and their responses to water deficit and salinity. J Exp Bot 2007;58:1873–92.
[59] Eivazi, Naghavi, Hajheidari, Pirseyedi, Ghaffari, Mohammadi, et al. Assessing wheat (*Triticum aestivum* L.) genetic diversity using quality traits, amplified fragment length polymorphisms, simple sequence repeats and proteome analysis. Ann Appl Biol 2008;152:81–91.
[60] Agrawal GK, Rakwal R. Technologies, strategies, and applications. New Jersey: John Wiley & Sons Inc.; 2008.
[61] Samaj J, Thelen JJ. Plant Proteomics. Germany: Springer-Verlag; 2007.
[62] Agrawal GK, Jwa N-S and Rakwal R. Rice Proteomics: ending phase I and the beginning of phase II. Proteomics 2009;9:935–63.
[63] Farrokhi N, Whitelegge JP, Brusslan JA. Plant peptides and peptidomics. Plant Biotechnol J 2008;6:105–34.
[64] The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 2007;449:463–7.
[65] Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science 2006;313:1596–604.
[66] Rensing SA, Lang D, Zimmer AD, Terry A, Salamov A, Lucas S, et al. The physcomitrella genome reveals evolutionary insights into the conquest of land by plants. Science 2008;319:64–9.
[67] Wang X, Yang P, Gao Q, Liu X, Kuang T, Shen S, et al. Proteomic analysis of the response to high-salinity stress in *Physcomitrella patens*. Planta 2008;228:167–77.
[68] Gazanchian A, Hajheidari M, Sima NK, Salekdeh GH. Proteome response of *Elymus elongatum* to severe water stress and recovery. J Exp Bot 2007;58:291–300.
[69] Brownfield L, Ford K, Doblin MS, Newbigin E, Read S, Bacic A. Proteomic and biochemical evidence links the callose synthase in *Nicotiana alata* pollen tubes to the product of the NaGSL1 gene. Plant J 2007;52:147–56.
[70] Jiang G, Wang Z, Shang H, Yang W, Hu Z, Phillips J, et al. Proteome analysis of leaves from the resurrection plant *Boea hygrometrica* in response to dehydration and rehydration. Planta 2007;225:1405–20.
[71] Ingle RA, Schmidt UG, Farrant JM, Thomson JA, Mundree SG. Proteomic analysis of leaf proteins during dehydration of the resurrection plant *Xerophyta viscosa*. Plant Cell Environ 2007;30:435–46.
[72] Vyetrogon K, Tebbji F, Olson DJ, Ross AR, Matton DP. A comparative proteome and phosphoproteome analysis of differentially regulated proteins during fertilization in the self-incompatible species *Solanum chacoense* Bitt. Proteomics 2007;7:232–47.
[73] Yoshimura K, Masuda A, Kuwano M, Yokota A, Akashi K. Programmed proteome response for drought avoidance/tolerance in the root of a C3 xerophyte (wild watermelon) under water deficits. Plant Cell Physiol 2008;49:226–41.
[74] Lliso I, Tadeo FR, Phinney BS, Wilkerson CG, Talón M. Protein changes in the albedo of citrus fruits on postharvesting storage. J Agric Food Chem 2007;55:9047–53.
[75] Wang D, Eyles A, Mandich D, Bonello P. Systemic aspects of host–pathogen interactions in Austrian pine (*Pinus nigra*): a Proteomics approach. Physiol Mol Plant Pathol 2006;68:149–57.
[76] Fiorani Celedon PA, de Andrade A, Meireles KG, Gallo de Carvalho MC, Caldas DG, Moon DH, et al. Proteomic analysis of the cambial region in juvenile *Eucalyptus grandis* at three ages. Proteomics 2007;7:2258–74.
[77] Lippert D, Chowrira S, Ralph SG, Zhuang J, Aeschliman D, Ritland C, et al. Conifer defense against insects: proteome analysis of Sitka spruce (*Picea sitchensis*) bark induced by mechanical wounding or feeding by white pine weevils (*Pissodes strobi*). Proteomics 2007;7:248–70.
[78] 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.
[79] Pedreschi R, Vanstreels E, Carpentier S, Hertog M, Lammertyn J, Robben J, et al. Proteomic analysis of core breakdown disorder in conference pears (*Pyrus communis* L.). Proteomics 2007;7:2083–99.
[80] de Jong F, Mathesius U, Imin N, Rolfe BG. A proteome study of the proliferation of cultured *Medicago truncatula* protoplasts. Proteomics 2007;7:722–36.
[81] Dai S, Chen T, Chong K, Xue Y, Liu S, Wang T. Proteomics identification of differentially expressed proteins associated with pollen germination and tube growth reveals characteristics of germinated *Oryza sativa* pollen. Mol Cell Proteomics 2007;6:207–30.
[82] Houterman PM, Speijer D, Dekker HL, De Koster CG, Cornelissen BJC, Rep M. The mixed xylem sap proteome of *Fusarium oxysporum*-infected tomato plants. Mol Plant Pathol 2007;8:215–21.
[83] Gutstein HB, Morris JS. Laser capture sampling and analytical issues in Proteomics. Expert Rev Proteomics 2007;4:627–37.
[84] Dembinsky D, Woll K, Saleem M, Liu Y, Fu Y, Borsuk LA, et al. Transcriptomic and proteomic analyses of pericycle cells of the maize primary root. Plant Physiol 2007;145:575–88.
[85] Sigal A, Milo R, Cohen A, geva-Zatorsky N, Klein Y, Liron Y, et al. Variability and memory of protein levels in human cells. Nature 2006;444:643–5.
[86] Foss EJ, Radulovic D, Shaffer SA, Ruderfer DM, Bedalov AM, Goodlett DR, et al. Genetic basis of protein variation in yeast. Nat Genetics 2007;39:1369–75.
[87] Jorge I, Navarro RM, Lenz C, Ariza D, Porras C, Jorrin 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.
[88] Maor R, Jones A, Nuhse TS, Studholme DJ, Peck SC, Shirasu K. Multidimensional protein identification technology (MudPIT) analysis of ubiquitinated proteins in plants. Mol Cell Proteomics 2007;6:601–10.
[89] Minic Z, Jamet E, Negroni L, Arsene der Garabedian P, Zivy M, Jouanin L. A sub-proteome of *Arabidopsis thaliana* mature stems trapped on Concanavalin A is enriched in cell wall glycoside hydrolases. J Exp Bot 2007;58:2503–12.
[90] Casati P, Campi M, Chu F, Suzuki N, Maltby D, Guan S, et al. Histone acetylation and chromatin remodeling are required for UV-B-dependent transcriptional activation of regulated genes in maize. Plant Cell 2008;20:827–42.
[91] Rutschow H, Ytterberg AJ, Friso G, Nilsson R, van Wijk KJ. Quantitative proteomics of a chloroplast SRP54 sorting

mutant and its genetic interactions with CLPC1 in *Arabidopsis*. Plant Physiol 2008;148:156–75.
[92] Alkhalfioui F, Renard M, Vensel WH, Wong J, Tanaka CK, Hurkman WJ, et al. Thioredoxin-linked proteins are reduced during germination of *Medicago truncatula* seeds. Plant Physiol 2007;144:1559–79.
[93] Popescu SC, Popescu GV, Bachan S, Zhang Z, Seay M, Gerstein M, et al. Differential binding of calmodulin-related proteins to their targets revealed through high-density *Arabidopsis* protein microarrays. Proc Natl Acad Sci USA 2007;104:4730–5.
[94] Schoonheim PJ, Veiga H, Pereira Dda C, Friso G, van Wijk KJ, de Boer AH. A comprehensive analysis of the 14-3-3 interactome in barley leaves using a complementary proteomics and two-hybrid approach. Plant Physiol 2007;143:670–83.
[95] Abe M, Fujiwara M, Kurotani K, Yokoi S, Shimamoto K. Identification of dynamin as an interactor of rice GIGANTEA by tandem affinity purification (TAP). Plant Cell Physiol 2008;49:420–32.
[96] Galetskiy D, Susnea I, Reiser V, Adamska I, Przybylski M. Structure and dynamics of photosystem II light-harvesting complex revealed by high-resolution FTICR mass spectrometric proteome analysis. J Am Soc Mass Spectrom 2008;19:1004–13.
[97] Eriksson J, Fenyo D. Improving the success rate of proteome analysis by modeling protein-abundance distributions and experimental designs. Nat Biotechnol 2007;25:651–5.
[98] Isaacson T, Damasceno CM, Saravanan RS, He Y, Catala C, Saladie M, et al. Sample extraction techniques for enhanced proteomic analysis of plant tissues. Nat Protoc 2006;1:769–74.
[99] Maldonado AM, Echevarria-Zomeno S, Jean-Baptiste S, Hernandez M, Jorrin-Novo JV. Evaluation of three different protocols of protein extraction for *Arabidopsis thaliana* leaf proteome analysis by two-dimensional electrophoresis. J Proteomics 2008;71:461–72.
[100] Zheng Q, Song J, Doncaster K, Rowland E, Byers DM. Qualitative and quantitative evaluation of protein extraction protocols for apple and strawberry fruit suitable for two-dimensional electrophoresis and mass spectrometry analysis. J Agric Food Chem 2007;55:1663–73.
[101] Canas B, Pineiro C, Calvo E, Lopez-Ferrer D, Gallardo JM. Trends in sample preparation for classical and second generation proteomics. J Chromatogr A 2007;1153:235–58.
[102] Sarma AD, Oehrle NW, Emerich DW. Plant protein isolation and stabilization for enhanced resolution of two-dimensional polyacrylamide gel electrophoresis. Anal Biochem 2008;379:192–5.
[103] Wang W, Tai F, Chen S. Optimizing protein extraction from plant tissues for enhanced proteomics analysis. J Sep Sci 2008;31:2032–9.
[104] Pirovani CP, Carvalho HA, Machado RC, Gomes DS, Alvim FC, Pomella AW, et al. Protein extraction for proteome analysis from cacao leaves and meristems, organs infected by *Moniliophthora perniciosa*, the causal agent of the witches' broom disease. Electrophoresis 2008;29:2391–401.
[105] Toorchi M, Nouri M-Z, Tsumura M, Komatsu S. Acoustic technology for high-performance disruption and extraction of plant proteins. J Proteome Res 2008;7:3035–41.
[106] Kao SH, Wong HK, Chiang CY, Chen HM. Evaluating the compatibility of three colorimetric protein assays for two-dimensional electrophoresis experiments. Proteomics 2008;8:2178–84.
[107] Jiang X, Ye M, Zou H. Technologies and methods for sample pretreatment in efficient proteome and peptidome analysis. Proteomics 2008;8:686–705.
[108] Cellar NA, Kuppannan K, Langhorst ML, Ni W, Xu P, Young SA. Cross species applicability of abundant protein depletion columns for ribulose-1,5-bisphosphate carboxylase/oxygenase. J Chromatogr B Analyt Technol Biomed Life Sci 2008;861:29–39.
[109] Lee DG, Ahsan N, Lee SH, Kang KY, Lee JJ, Lee BH. An approach to identify cold-induced low-abundant proteins in rice leaf. C R Biol 2007;330:215–25.
[110] Hashimoto M, Komatsu S. Proteomic analysis of rice seedlings during cold stress. Proteomics 2007;7:1293–302.
[111] Righetti PG, Boschetti E. Sherlock Holmes and the proteome—a detective story. FEBS J 2007;274:897–905.
[112] Supek F, Peharec P, Krsnik-Rasol M, Smuc T. Enhanced analytical power of SDS-PAGE using machine learning algorithms. Proteomics 2008;8:28–31.
[113] de Godoy LM, Olsen JV, de Souza GA, Li G, Mortensen P, Mann M. Status of complete proteome analysis by mass spectrometry: SILAC labeled yeast as a model system. Genome Biol 2006;7:R50.
[114] Tribl F, Lohaus C, Dombert T, Langenfeld E, Piechura H, Warscheid B, et al. Towards multidimensional liquid chromatography separation of proteins using fluorescence and isotope-coded protein labelling for quantitative proteomics. Proteomics 2008;8:1204–11.
[115] Lengqvist J, Uhlen K, Lehtio J. iTRAQ compatibility of peptide immobilized pH gradient isoelectric focusing. Proteomics 2007;7:1746–52.
[116] Braun RJ, Kinkl N, Beer M, Ueffing M. Two-dimensional electrophoresis of membrane proteins. Anal Bioanal Chem 2007;389:1033–45.
[117] van den Broeck HC, America AH, Smulders MJ, Gilissen LJ, van der Meer IM. Staining efficiency of specific proteins depends on the staining method: wheat gluten proteins. Proteomics 2008;8:1880–4.
[118] Witzel K, Surabhi GK, Jyothsnakumari G, Sudhakar C, Matros A, Mock HP. Quantitative proteome analysis of barley seeds using ruthenium(II)-tris-(bathophenanthroline-disulphonate) staining. J Proteome Res 2007;6:1325–33.
[119] Harris LR, Churchward MA, Butt RH, Coorssen JR. Assessing detection methods for gel-based proteomic analyses. J Proteome Res 2007;6:1418–25.
[120] Wang X, Li X, Li Y. A modified Coomassie Brilliant Blue staining method at nanogram sensitivity compatible with proteomic analysis. Biotechnol Lett 2007;29:1599–603.
[121] Chakravarti B, Ratanaprayul W, Dalal N, Chakravarti DN. Comparison of SYPRO Ruby and Deep Purple using commonly available UV transilluminator: wide-scale application in proteomic research. J Proteome Res 2008;7:2797–802.
[122] Jacobsen S, Grove H, Jensen KN, Sorensen HA, Jessen F, Hollung K, et al. Multivariate analysis of 2-DE protein patterns—practical approaches. Electrophoresis 2007;28:1289–99.
[123] Berth M, Moser FM, Kolbe M, Bernhardt J. The state of the art in the analysis of two-dimensional gel electrophoresis images. Appl Microbiol Biotechnol 2007;76:1223–43.
[124] Daszykowski M, Stanimirova I, Bodzon-Kulakowska A, Silberring J, Lubec G, Walczak B. Start-to-end processing of two-dimensional gel electrophoretic images. J Chromatogr A 2007;1158:306–17.
[125] Maurer MH. Software analysis of two-dimensional electrophoretic gels in proteomic experiments. Current Bioinformatics 2006;1:255–62.
[126] Yang Y, Thannhauser TW, Li L, Zhang S. Development of an integrated approach for evaluation of 2-D gel image analysis: impact of multiple proteins in single spots on comparative proteomics in conventional 2-D gel/MALDI workflow. Electrophoresis 2007;28:2080–94.

[127] Demianova Z, Shimmo M, Poysa E, Franssila S, Baumann M. Toward an integrated microchip sized 2-D polyacrylamide slab gel electrophoresis device for proteomic analysis. Electrophoresis 2007;28:422–8.
[128] Schroder S, Zhang H, Yeung ES, Jansch L, Zabel C, Watzig H. Quantitative gel electrophoresis: sources of variation. J Proteome Res 2008;7:1226–34.
[129] Valcu CM, Valcu M. Reproducibility of two-dimensional gel electrophoresis at different replication levels. J Proteome Res 2007;6:4677–83.
[130] Jensen KN, Jessen F, Jorgensen BM. Multivariate data analysis of two-dimensional gel electrophoresis protein patterns from few samples. J Proteome Res 2008;7:1288–96.
[131] Pedreschi R, Hertog ML, Carpentier SC, Lammertyn J, Robben J, Noben JP, et al. Treatment of missing values for multivariate statistical analysis of gel-based proteomics data. Proteomics 2008;8:1371–83.
[132] Karp NA, McCormick PS, Russell MR, Lilley KS. Experimental and statistical considerations to avoid false conclusions in proteomics studies using differential in-gel electrophoresis. Mol Cell Proteomics 2007;6:1354–64.
[133] Gomez SM, Nishio JN, Faull KF, Whitelegge JP. The chloroplast grana proteome defined by intact mass measurements from liquid chromatography mass spectrometry. Mol Cell Proteomics 2002;1:46–59.
[134] Van Leene J, Stals H, Eeckhout D, Persiau G, Van De Slijke E, Van Isterdael G, et al. A tandem affinity purification-based technology platform to study the cell cycle interactome in *Arabidopsis thaliana*. Mol Cell Proteomics 2007;6:1226–38.
[135] D'Amici GM, Timperio AM, Zolla L. Coupling of native liquid phase isoelectrofocusing and blue native polyacrylamide gel electrophoresis: a potent tool for native membrane multiprotein complex separation. J Proteome Res 2008;7:1326–40.
[136] Hrebicek T, Durrschmid K, Auer N, Bayer K, Rizzi A. Effect of CyDye minimum labeling in differential gel electrophoresis on the reliability of protein identification. Electrophoresis 2007;28:1161–9.
[137] Gevaert K, Van Damme P, Ghesquiere B, Impens F, Martens L, Helsens K, et al. A la carte proteomics with an emphasis on gel-free techniques. Proteomics 2007;7:2698–718.
[138] Ye ML, Jiang XG, Feng S, Tian RJ, Zou HF. Advances in chromatographic techniques and methods in shotgun proteome analysis. Trends Analyt Chem 2007;26:80–4.
[139] Jain R, Katavic V, Agrawal GK, Guzov VM, Thelen JJ. Purification and proteomic characterization of plastids from *Brassica napus* developing embryos. Proteomics 2008;8:3397–405.
[140] Djordjevic MA, Oakes M, Li DX, Hwang CH, Hocart CH, Gresshoff PM. The glycine max xylem sap and apoplast proteome. J Proteome Res 2007;6:3771–9.
[141] Lau KW, Jones AR, Swainston N, Siepen JA, Hubbard SJ. Capture and analysis of quantitative proteomic data. Proteomics 2007;7:2787–99.
[142] Bantscheff M, Schirle M, Sweetman G, Rick J, Kuster B. Quantitative mass spectrometry in proteomics: a critical review. Anal Bioanal Chem 2007;389:1017–31.
[143] Shiio Y, Aebersold R. Quantitative proteome analysis using isotope-coded affinity tags and mass spectrometry. Nat Protoc 2006;1:139–45.
[144] Wiese S, Reidegeld KA, Meyer HE, Warscheid B. Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research. Proteomics 2007;7:340–50.
[145] Gan CS, Chong PK, Pham TK, Wright PC. Technical, experimental, and biological variations in isobaric tags for relative and absolute quantitation (iTRAQ). J Proteome Res 2007;6:821–7.
[146] Palmblad M, Mills DJ, Bindschedler LV. Heat-shock response in *Arabidopsis thaliana* explored by multiplexed quantitative proteomics using differential metabolic labeling. J Proteome Res 2008;7:780–5.
[147] Nelson CJ, Huttlin EL, Hegeman AD, Harms AC, Sussman MR. Implications of ^{15}N-metabolic labeling for automated peptide identification in *Arabidopsis thaliana*. Proteomics 2007;7:1279–92.
[148] Brautigam A, Hofmann-Benning S, Weber AP. Comparative Proteomics of chloroplast envelopes from c3 and c4 plants reveals specific adaptations of the plastid envelope to c4 photosynthesis and candidate proteins required for maintaining c4 metabolite fluxes. Plant Physiol 2008;148:568–79.
[149] Thelen JJ, Peck SC. Quantitative Proteomics in plants: choices in abundance. Plant Cell 2007;19:3339–46.
[150] America AH, Cordewener JH. Comparative LC-MS: a landscape of peaks and valleys. Proteomics 2008;8:731–49.
[151] Hebeler R, Oeljeklaus S, Reidegeld KA, Eisenacher M, Stephan C, Sitek B, et al. Study of early leaf senescence in *Arabidopsis thaliana* by quantitative proteomics using reciprocal ^{14}N/^{15}N labeling and difference gel electrophoresis. Mol Cell Proteomics 2008;7:108–20.
[152] Han X, Aslanian A, Yates III JR. Mass spectrometry for proteomics. Curr Opin Chem Biol 2008;12:483–90.
[153] Zubarev R, Mann M. On the proper use of mass accuracy in Proteomics. Mol Cell Proteomics 2007;6:377–81.
[154] Molina H, Horn DM, Tang N, Mathivanan S, Pandey A. Global proteomic profiling of phosphopeptides using electron transfer dissociation tandem mass spectrometry. Proc Natl Acad Sci USA 2007;104:2199–204.
[155] Good DM, Wirtala M, McAlister GC, Coon JJ. Performance characteristics of electron transfer dissociation mass spectrometry. Mol Cell Proteomics 2007;6:1942–51.
[156] Marshall AG, Hendrickson CL, Emmett MR, Rodgers RP, Blakney GT, Nilsson CL. Fourier transform ion cyclotron resonance: state of the art. Eur J Mass Spectrom 2007;13:57–9.
[157] Meng F, Wiener MC, Sachs JR, Burns C, Verma P, Paweletz CP, et al. Quantitative analysis of complex peptide mixtures using FTMS and differential mass spectrometry. J Am Soc Mass Spectrom 2007;18:226–33.
[158] Katz E, Fon M, Lee Y, Phinney B, Sadka A, Blumwald E. The citrus fruit proteome: insights into citrus fruit metabolism. Planta 2007;226:989–1005.
[159] Lee J, Garrett WM, Cooper B. Shotgun proteomic analysis of *Arabidopsis thaliana* leaves. J Sep Sci 2007;30:2225–30.
[160] Mulder NJ, Kersey P, Pruess M, Apweiler R. In silico characterization of proteins: UniProt, InterPro and Integr8. Mol Biotechnol 2008;38:165–77.
[161] Matthiesen R. Methods, algorithms and tools in computational Proteomics: a practical point of view. Proteomics 2007;7:2815–32.
[162] Song Z, Chen L, Ganapathy A, Wan XF, Brechenmacher L, Tao N, et al. Development and assessment of scoring functions for protein identification using PMF data. Electrophoresis 2007;28:864–70.
[163] Balgley BM, Laudeman T, Yang L, Song T, Lee CS. Comparative evaluation of tandem MS search algorithms using a target-decoy search strategy. Mol Cell Proteomics 2007;6:1599–608.
[164] Edwards NJ. Novel peptide identification from tandem mass spectra using ESTs and sequence database compression. Mol Syst Biol 2007;3:102.
[165] Higgs RE, Knierman MD, Freeman AB, Gelbert LM, Patil ST, Hale JE. Estimating the statistical significance of peptide identifications from shotgun proteomics experiments. J Proteome Res 2007;6:1758–67.
[166] Brosch M, Swamy S, Hubbard T, Choudhary J. Comparison of Mascot and X!Tandem performance for low and high

accuracy mass spectrometry and the development of an adjusted Mascot threshold. Mol Cell Proteomics 2008;7:962–70.
[167] Joo WA, Lee JB, Park M, Lee JW, Kim HJ, Kim CW. Comparison of search engine contributions in protein mass fingerprinting for protein identification. Biotechnol Bioproc E 2007;12:125–30.
[168] Waridel P, Frank A, Thomas H, Surendranath V, Sunyaev S, Pevzner P, et al. Sequence similarity-driven proteomics in organisms with unknown genomes by LC-MS/MS and automated de novo sequencing. Proteomics 2007;7:2318–29.
[169] Junqueira M, Spirin V, Balbuena TS, Thomas H, Adzhubei I, Sunyaev S, et al. Protein identification pipeline for the homology-driven proteomics. J Proteomics 2008;71:346–56.
[170] Catusse J, Strub J-M, Job C, Van Dorsselaer A, Job D. Proteome-wide characterization of sugarbeet seed vigor and its tissue specific expression. Proc Natl Acad Sci USA 2008;105:10262–7.
[171] Aki T, Shigyo M, Nakano R, Yoneyama T, Yanagisawa S. Nano scale proteomics revealed the presence of regulatory proteins including three FT-like proteins in phloem and xylem saps from rice. Plant Cell Physiol 2008;49:767–90.
[172] Espagne C, Martinez A, Valot B, Meinnel T, Giglione C. Alternative and effective proteomic analysis in *Arabidopsis*. Proteomics 2007;7:3788–99.
[173] Song X, Ni Z, Yao Y, Xie C, Li Z, Wu H, et al. Wheat (*Triticum aestivum* L.) root proteome and differentially expressed root proteins between hybrid and parents. Proteomics 2007;7:3538–57.
[174] Sheoran IS, Ross AR, Olson DJ, Sawhney VK. Proteomic analysis of tomato (*Lycopersicon esculentum*) pollen. J Exp Bot 2007;58:3525–35.
[175] Salvi D, Rolland N, Joyard J, Ferro M. Assessment of organelle purity using antibodies and specific assays: the example of the chloroplast envelope. Methods Mol Biol 2008;432:345–56.
[176] Block M, Douce R, Joyard J, Rolland N. Chloroplast envelope membranes: a dynamic interface between plastids and the cytosol. Photosynthesis Res 2007;92:225–44.
[177] Salvi D, Rolland N, Joyard J, Ferro M. Purification and proteomic analysis of chloroplasts and their sub-organellar compartments. Methods Mol Biol 2008;432:19–36.
[178] Lee CP, Eubel H, O'Toole N, Millar AH. Heterogeneity of the mitochondrial proteome for photosynthetic and non-photosynthetic *Arabidopsis* metabolism. Mol Cell Proteomics 2008;7:1297–316.
[179] Robert S, Zouhar J, Carter C, Raikhel N. Isolation of intact vacuoles from *Arabidopsis* rosette leaf-derived protoplasts. Nat Protocols 2007;2:259–62.
[180] Jamet E, Boudart G, Borderies G, Charmont S, Lafitte C, Rossignol M, et al. Isolation of plant cell wall proteins. Methods Mol Biol 2008;425:187–201.
[181] Komatsu S, Konishi H, Hashimoto M. The Proteomics of plant cell membranes. J Exp Bot 2007;58:103–12.
[182] Mitra SK, Gantt JA, Ruby JF, Clouse SD, Goshe MB. Membrane proteomic analysis of *Arabidopsis thaliana* using alternative solubilization techniques. J Proteome Res 2007;6:1933–50.
[183] Lilley KS, Dunkley TP. Determination of genuine residents of plant endomembrane organelles using isotope tagging and multivariate statistics. Methods Mol Biol 2008;432:373–87.
[184] Andersen JS, Mann M. Organellar proteomics: turning inventories into insights. EMBO Rep 2006;7:874–9.
[185] Haynes PA, Roberts TH. Subcellular shotgun proteomics in plants: looking beyond the usual suspects. Proteomics 2007;7:2963–75.
[186] Jamet E, Albenne C, Boudart G, Irshad M, Canut H, Pont-Lezica R. Recent advances in plant cell wall Proteomics. Proteomics 2008;8:893–908.
[187] Ito J, Heazlewood JL, Millar AH. The plant mitochondrial proteome and the challenge of defining the posttranslational modifications responsible for signalling and stress effects on respiratory functions. Physiol Plant 2007;129:207–24.
[188] Carroll AJ, Heazlewood JL, Ito J, Millar AH. Analysis of the *Arabidopsis* cytosolic ribosome proteome provides detailed insights into its components and their post-translational modification. Mol Cell Proteomics 2008;7:347–69.
[189] Mongrand S, Morel J, Laroche J, Claverol S, Carde J-P, Hartmann M-A, et al. Lipid rafts in higher plant cells: purification and characterization of Triton X-100-insoluble microdomains from tobacco plasma membrane. J Biol Chem 2004;279:36277–86.
[190] Lefebvre B, Furt F, Hartmann MA, Michaelson LV, Carde JP, Sargueil-Boiron F, et al. Characterization of lipid rafts from *Medicago truncatula* root plasma membranes: a proteomic study reveals the presence of a raft-associated redox system. Plant Physiol 2007;144:402–18.
[191] Zolla L, Rinalducci S, Timperio AM. Proteomic analysis of photosystem I components from different plant species. Proteomics 2007;7:1866–76.
[192] Stensballe A, Hald S, Bauw G, Blennow A, Welinder KG. The amyloplast proteome of potato tuber. FEBS J 2008;275:1723–41.
[193] Arai Y, Hayashi M, Nishimura M. Proteomic analysis of highly purified peroxisomes from etiolated soybean cotyledons. Plant Cell Physiol 2008;49:526–39.
[194] Natera SH, Ford KL, Cassin AM, Patterson JH, Newbigin EJ, Bacic A. Analysis of the *Oryza sativa* plasma membrane proteome using combined protein and peptide fractionation approaches in conjunction with mass spectrometry. J Proteome Res 2008;7:1159–87.
[195] Whiteman SA, Nuhse TS, Ashford DA, Sanders D, Maathuis FJ. A proteomic and phosphoproteomic analysis of *Oryza sativa* plasma membrane and vacuolar membrane. Plant J 2008;56:146–56.
[196] Whiteman SA, Serazetdinova L, Jones AM, Sanders D, Rathjen J, Peck SC, et al. Identification of novel proteins and phosphorylation sites in a tonoplast enriched membrane fraction of *Arabidopsis thaliana*. Proteomics 2008;8:3536–47.
[197] Larson RL, Hill AL, Nunez A. Characterization of protein changes associated with sugar beet (*Beta vulgaris*) resistance and susceptibility to *Fusarium oxysporum*. J Agric Food Chem 2007;55:7905–15.
[198] Wen F, VanEtten HD, Tsaprailis G, Hawes MC. Extracellular proteins in pea root tip and border cell exudates. Plant Physiol 2007;143:773–83.
[199] Patterson J, Ford K, Cassin A, Natera S, Bacic A. Increased abundance of proteins involved in phytosiderophore production in boron-tolerant barley. Plant Physiol 2007;144:1612–31.
[200] Hajheidari M, Eivazi A, Buchanan BB, Wong JH, Majidi I, Salekdeh GH. Proteomics uncovers a role for redox in drought tolerance in wheat. J Proteome Res 2007;6:1451–60.
[201] Kang SG, Matin MN, Bae H, Natarajan S. Proteome analysis and characterization of phenotypes of lesion mimic mutant spotted leaf 6 in rice. Proteomics 2007;7:2447–58.
[202] Elvira MI, Galdeano MM, Gilardi P, Garcia-Luque I, Serra MT. Proteomic analysis of pathogenesis-related proteins (PRs) induced by compatible and incompatible interactions of pepper mild mottle virus (PMMoV) in *Capsicum chinense* L3 plants. J Exp Bot 2008;59:1253–65.
[203] Yang P, Li X, Wang X, Chen H, Chen F, Shen S. Proteomic analysis of rice (*Oryza sativa*) seeds during germination. Proteomics 2007;7:3358–68.
[204] Rajjou L, Lovigny Y, Groot SP, Belghazi M, Job C, Job D. Proteome-wide characterization of seed aging in *Arabidopsis*: a comparison between artificial and natural aging protocols. Plant Physiol 2008;148:620–41.

[205] Gallardo K, Firnhaber C, Zuber H, Hericher D, Belghazi M, Henry C, et al. A combined proteome and transcriptome analysis of developing *Medicago truncatula* seeds: evidence for metabolic specialization of maternal and filial tissues. Mol Cell Proteomics 2007;6:2165–79.
[206] Mechin V, Thevenot C, Le Guilloux M, Prioul J-L, Damerval C. Developmental analysis of maize endosperm proteome suggests a pivotal role for pyruvate orthophosphate dikinase. Plant Physiol 2007;143:1203–19.
[207] Pawlowski TA. Proteomics of European beech (*Fagus sylvatica* L.) seed dormancy breaking: influence of abscisic and gibberellic acids. Proteomics 2007;7:2246–57.
[208] Deng Z, Zhang X, Tang W, Oses-Prieto JA, Suzuki N, Gendron JM, et al. A proteomic study of brassinosteroid response in *Arabidopsis*. Mol Cell Proteomics 2007;6:2058–71.
[209] Tang W, Kim TW, Oses-Prieto JA, Sun Y, Deng Z, Zhu S, et al. BSKs mediate signal transduction from the receptor kinase BRI1 in *Arabidopsis*. Science 2008;321:557–60.
[210] Chitteti BR, Peng Z. Proteome and phosphoproteome differential expression under salinity stress in rice (*Oryza sativa*) roots. J Proteome Res 2007;6:1718–27.
[211] El-Khatib RT, Good AG, Muench DG. Analysis of the *Arabidopsis* cell suspension phosphoproteome in response to short-term low temperature and abscisic acid treatment. Physiol Plant 2007;129:687–97.
[212] Sugiyama N, Nakagami H, Mochida K, Daudi A, Tomita M, Shirasu K, et al. Large-scale phosphorylation mapping reveals the extent of tyrosine phosphorylation in *Arabidopsis*. Mol Syst Biol 2008;4:193.
[213] de la Fuente van Bentem S, Anrather D, Dohnal I, Roitinger E, Csaszar E, Joore J, et al. Site-specific phosphorylation profiling of *Arabidopsis* proteins by mass spectrometry and peptide chip analysis. J Proteome Res 2008;7:2458–70.
[214] Tan F, Li G, Chitteti BR, Peng Z. Proteome and phosphoproteome analysis of chromatin associated proteins in rice (*Oryza sativa*). Proteomics 2007;7:4511–27.
[215] Benschop JJ, Mohammed S, O'Flaherty M, Heck AJ, Slijper M, Menke FL. Quantitative phospho proteomics of early elicitor signaling in *Arabidopsis*. Mol Cell Proteomics 2007;6:1198–214.
[216] Lee J, Xu Y, Chen Y, Sprung R, Kim SC, Xie S, et al. Mitochondrial phosphoproteome revealed by an improved IMAC method and MS/MS/MS. Mol Cell Proteomics 2007;6:669–76.
[217] Kwon SJ, Choi EY, Seo JB, Park OK. Isolation of the *Arabidopsis* phosphoproteome using a biotin-tagging approach. Mol Cells 2007;24:268–75.
[218] Meimoun P, Ambard-Bretteville F, Colas-des Francs-Small C, Valot B, Vidal J. Analysis of plant phosphoproteins. Anal Biochem 2007;371:238–46.
[219] Nuhse TS, Bottrill AR, Jones AM, Peck SC. Quantitative phosphoproteomic analysis of plasma membrane proteins reveals regulatory mechanisms of plant innate immune responses. Plant J 2007;51:931–40.
[220] Kruger R, Wolschin F, Weckwerth W, Bettmer J, Lehmann WD. Plant protein phosphorylation monitored by capillary liquid chromatography-element mass spectrometry. Biochem Biophys Res Commun 2007;355:89–96.
[221] de la Fuente van Bentem S, Hirt H. Using phospho proteomics to reveal signalling dynamics in plants. Trends Plant Sci 2007;1(2):404–11.
[222] Camerini S, Polci ML, Restuccia U, Usuelli V, Malgaroli A, Bachi A. A novel approach to identify proteins modified by nitric oxide: the HIS-TAG switch method. J Proteome Res 2007;6:3224–31.
[223] Han P, Chen C. Detergent-free biotin switch combined with liquid chromatography/tandem mass spectrometry in the analysis of S-nitrosylated proteins. Rapid Commun Mass Spectrom 2008;22:1137–45.
[224] Kettenhofen NJ, Broniowska KA, Keszler A, Zhang Y, Hogg N. Proteomic methods for analysis of S-nitrosation. J Chromatogr B Analyt Technol Biomed Life Sci 2007;851:152–9.
[225] Xu H, Zhang L, Freitas MA. Identification and characterization of disulfide bonds in proteins and peptides from tandem MS data by use of the MassMatrix MS/MS search engine. J Proteome Res 2008;7:138–44.
[226] Wormuth D, Heiber I, Shaikali J, Kandlbinder A, Baier M, Dietz KJ. Redox regulation and antioxidative defence in *Arabidopsis* leaves viewed from a systems biology perspective. J Biotechnol 2007;129:229–48.
[227] Winger AM, Taylor NL, Heazlewood JL, Day DA, Millar AH. Identification of intra- and intermolecular disulphide bonding in the plant mitochondrial proteome by diagonal gel electrophoresis. Proteomics 2007;7:4158–70.
[228] Ströher E, Dietz K-J. The dynamic thioldisulphide redox proteome of the *Arabidopsis thaliana* chloroplast as revealed by differential electrophoretic mobility. Physiologia Plantarum 2008;133:566–83.
[229] Balmer Y, Vensel WH, Cai N, Manieri W, Schurmann P, Hurkman WJ, et al. A complete ferredoxin/thioredoxin system regulates fundamental processes in amyloplasts. Proc Natl Acad Sci USA 2006;103:2988–93.
[230] Besson-Bard A, Pugin A, Wendehenne D. New insights into nitric oxide signaling in plants. Annu Rev Plant Biol 2008;59:21–39.
[231] Romero-Puertas MC, Laxa M, Matte A, Zaninotto F, Finkemeier I, Jones AM, et al. S-nitrosylation of peroxiredoxin II E promotes peroxynitrite-mediated tyrosine nitration. Plant Cell 2007;19:4120–30.
[232] Belenghi B, Romero-Puertas MC, Vercammen D, Brackenier A, Inze D, Delledonne M, et al. Metacaspase activity of *Arabidopsis thaliana* is regulated by S-nitrosylation of a critical cysteine residue. J Biol Chem 2007;282:1352–8.
[233] Han P, Zhou X, Huang B, Zhang X, Chen C. On-gel fluorescent visualization and the site identification of S-nitrosylated proteins. Anal Biochem 2008;377:150–5.
[234] Dalle-Donne I, Rossi R, Giustarini D, Colombo R, Milzani A. S-glutathionylation in protein redox regulation. Free Radical Bio Med 2007;43:883–98.
[235] Holtgrefe S, Gohlke J, Starmann J, Druce S, Klocke S, Altmann B, et al. Regulation of plant cytosolic glyceraldehyde 3-phosphate dehydrogenase isoforms by thiol modifications. Physiol Plant 2008;133:211–28.
[236] Wan J, Roth AF, Bailey AO, Davis NG. Palmitoylated proteins: purification and identification. Nat Protoc 2007;2:1573–84.
[237] Gamblin DP, van Kasteren SI, Chalker JM, Davis BG. Chemical approaches to mapping the function of post-translational modifications. FEBS J 2008;275:1949–59.
[238] Denis NJ, Vasilescu J, Lambert JP, Smith JC, Figeys D. Tryptic digestion of ubiquitin standards reveals an improved strategy for identifying ubiquitinated proteins by mass spectrometry. Proteomics 2007;7:868–74.
[239] Tomlinson E, Palaniyappan N, Tooth D, Layfield R. Methods for the purification of ubiquitinated proteins. Proteomics 2007;7:1016–22.
[240] Mollah S, Wertz IE, Phung Q, Arnott D, Dixit VM, Lill JR. Targeted mass spectrometric strategy for global mapping of ubiquitination on proteins. Rapid Commun Mass Spectrom 2007;21:3357–64.
[241] Martinez A, Traverso JA, Valot B, Ferro M, Espagne C, Ephritikhine G, et al. Extent of N-terminal modifications in

cytosolic proteins from eukaryotes. Proteomics 2008;8:2809–31.
[242] Kocher T, Superti-Furga G. Mass spectrometry-based functional Proteomics: from molecular machines to protein networks. Nat Methods 2007;4:807–15.
[243] Gingras AC, Gstaiger M, Raught B, Aebersold R. Analysis of protein complexes using mass spectrometry. Nat Rev Mol Cell Biol 2007;8:645–54.
[244] Berggard T, Linse S, James P. Methods for the detection and analysis of protein–protein interactions. Proteomics 2007;7:2833–42.
[245] Miernyk JA, Thelen JJ. Biochemical approaches for discovering protein–protein interactions. Plant J 2008;53:597–609.
[246] Lemaire SD, Michelet L, Zaffagnini M, Massot V, Issakidis-Bourguet E. Thioredoxins in chloroplasts. Curr Genet 2007;51:343–65.
[247] Shemer TA, Harpaz-Saad S, Belausov E, Lovat N, Krokhin O, Spicer V, et al. Citrus chlorophyllase dynamics at ethylene-induced fruit color-break: a study of chlorophyllase expression, posttranslational processing kinetics, and in situ intracellular localization. Plant Physiol 2008;148:108–18.
[248] Levy A, Erlanger M, Rosenthal M, Epel BL. A plasmodesmata-associated beta-1,3-glucanase in *Arabidopsis*. Plant J 2007;49:669–82.
[249] Bak-Jensen KS, Laugesen S, Ostergaard O, Finnie C, Roepstorff P, Svensson B. Spatio-temporal profiling and degradation of alpha-amylase isozymes during barley seed germination. FEBS J 2007;274:2552–65.
[250] Avila EL, Brown M, Pan S, Desikan R, Neill SJ, Girke T, et al. Expression analysis of *Arabidopsis* vacuolar sorting receptor 3 reveals a putative function in guard cells. J Exp Bot 2008;59:1149–61.
[251] Gershater MC, Cummins I, Edwards R. Role of a carboxylesterase in herbicide bioactivation in *Arabidopsis thaliana*. J Biol Chem 2007;282:21460–6.
[252] Croes E, Gebruers K, Robben J, Noben JP, Samyn B, Debyser G, et al. Variability of polymorphic families of three types of xylanase inhibitors in the wheat grain proteome. Proteomics 2008;8:1692–705.
[253] Gabaldón C, Gómez-Ros L, Núñez-Flores M, Esteban-Carrasco A, Barceló A. Post-translational modifications of the basic peroxidase isoenzyme from *Zinnia elegans*. Plant Molr Biol 2007;65:43–61.
[254] Baginsky S, Grossmann J, Gruissem W. Proteome analysis of chloroplast mRNA processing and degradation. J Proteome Res 2007;6:809–20.
[255] Schilling O, Overall CM. Proteomic discovery of protease substrates. Curr Opin Chem Biol 2007;11:36–45.
[256] Komatsu S, Toorchi M, Yukawa K. Soybean Proteomics. Curr Proteomics 2007;4:182–6.
[257] Qureshi MI, Qadir S, Zolla L. Proteomics-based dissection of stress-responsive pathways in plants. J Plant Physiol 2007;164:1239–60.
[258] Mehta A, Magalhaes BS, Souza DS, Vasconcelos EA, Silva LP, Grossi-de-Sa MF, et al. Rooteomics: the challenge of discovering plant defense-related proteins in roots. Curr Protein Pept Sci 2008;9:108–16.
[259] Zhang Q, Riechers DE. Proteomics: an emerging technology for Weed Sci Research. Weed Sci 2008;56:306–13.
[260] Mehta A, Brasileiro AC, Souza DS, Romano E, Campos MA, Grossi-de-Sa MF, et al. Plant–pathogen interactions: what is Proteomics telling us? Febs J 2008;275:3731–46.
[261] Nesatyy VJ, Suter MJ. Analysis of environmental stress response on the proteome level. Mass Spectrom Rev 2008;27:556–74.
[262] Yang P, Chen H, Liang Y, Shen S. Proteomic analysis of de-etiolated rice seedlings upon exposure to light. Proteomics 2007;7:2459–68.
[263] Faurobert M, Mihr C, Bertin N, Pawlowski T, Negroni L, Sommerer N, et al. Major proteome variations associated with cherry tomato pericarp development and ripening. Plant Physiol 2007;143:1327–46.
[264] Deytieux C, Geny L, Lapaillerie D, Claverol S, Bonneu M, Doneche B. Proteome analysis of grape skins during ripening. J Exp Bot 2007;58:1851–62.
[265] Giribaldi M, Perugini I, Sauvage FX, Schubert A. Analysis of protein changes during grape berry ripening by 2-DE and MALDI-TOF. Proteomics 2007;7:3154–70.
[266] Yeu S-Y, Park B-S, Sang W-G, Choi Y-D, Kim M-C, Song J-T, et al. The serine proteinase inhibitor OsSerpin is a potent tillering regulator in rice. J Plant Biol 2007;50:600–4.
[267] Kleffmann T, von Zychlinski A, Russenberger D, Hirsch-Hoffmann M, Gehrig P, Gruissem W, et al. Proteome dynamics during plastid differentiation in rice. Plant Physiol 2007;143:912–23.
[268] Kanervo E, Singh M, Suorsa M, Paakkarinen V, Aro E-M, Battchikova N, et al. Expression of protein complexes and individual proteins upon transition of etioplasts to chloroplasts in pea (*Pisum sativum*). Plant Cell Physiol 2008;49:396–410.
[269] Yin L, Tao Y, Zhao K, Shao J, Li X, Liu G, et al. Proteomic and transcriptomic analysis of rice mature seed-derived callus differentiation. Proteomics 2007;7:755–68.
[270] Kim ST, Kang SY, Wang Y, Kim SG, Hwang du H, Kang KY. Analysis of embryonic proteome modulation by GA and ABA from germinating rice seeds. Proteomics 2008;8:3577–87.
[271] Takasaki H, Mahmood T, Matsuoka M, Matsumoto H, Komatsu S. Identification and characterization of a gibberellin-regulated protein, which is ASR5, in the basal region of rice leaf sheaths. Mol Genet Genomics 2008;2 (79):359–70.
[272] He H, Li J. Proteomic analysis of phosphoproteins regulated by abscisic acid in rice leaves. Biochem Biophys Res Commun 2008;371:883–8.
[273] Shi F, Takasaki H, Komatsu S. Quantitative analysis of auxin-regulated proteins from basal part of leaf sheaths in rice by two-dimensional difference gel electrophoresis. Phytochemistry 2008;69:637–46.
[274] Cho K, Agrawal GK, Shibato J, Jung YH, Kim YK, Nahm BH, et al. Survey of differentially expressed proteins and genes in jasmonic acid treated rice seedling shoot and root at the proteomics and transcriptomics levels. J Proteome Res 2007;6:3581–603.
[275] Tang W, Deng Z, Oses-Prieto JA, Suzuki N, Zhu S, Zhang X, et al. Proteomics studies of brassinosteroid signal transduction using prefractionation and two-dimensional DIGE. Mol Cell Proteomics 2008;7:728–38.
[276] Thornton B, Osborne SM, Paterson E, Cash P. A proteomic and targeted metabolomic approach to investigate change in Lolium perenne roots when challenged with glycine. J Exp Bot 2007;5(8):1581–90.
[277] Brumbarova T, Matros A, Mock HP, Bauer P. A proteomic study showing differential regulation of stress, redox regulation and peroxidase proteins by iron supply and the transcription factor FER. Plant J 2008;54:321–34.
[278] Kim ST, Kim SG, Kang YH, Wang Y, Kim JY, Yi N, et al. Proteomics analysis of rice lesion mimic mutant (spl1) reveals tightly localized probenazole-induced protein (PBZ1) in cells undergoing programmed cell death. J Proteome Res 2008;7:1750–60.
[279] van Noorden GE, Kerim T, Goffard N, Wiblin R, Pellerone FI, Rolfe BG, et al. Overlap of proteome changes in *Medicago truncatula* in response to auxin and *Sinorhizobium meliloti*. Plant Physiol 2007;144:1115–31.
[280] Larrainzar E, Wienkoop S, Weckwerth W, Ladrera R, Arrese-Igor C, Gonzalez EM. *Medicago truncatula* root nodule

proteome analysis reveals differential plant and bacteroid responses to drought stress. Plant Physiol 2007;144:1495–507.
[281] Sharma N, Hotte N, Rahman MH, Mohammadi M, Deyholos MK, Kav NN. Towards identifying *Brassica* proteins involved in mediating resistance to *Leptosphaeria* maculans: a proteomics-based approach. Proteomics 2008;8:3516–35.
[282] Liang Y, Srivastava S, Rahman MH, Strelkov SE, Kav NN. Proteome changes in leaves of *Brassica napus* L. as a result of *Sclerotinia sclerotiorum* challenge. J Agric Food Chem 2008;56:1963–76.
[283] Cao T, Srivastava S, Rahman MH, Kav NNV, Hotte N, Deyholos MK, et al. Proteome-level changes in the roots of *Brassica napus* as a result of *Plasmodiophora brassicae* infection. Plant Sci 2008;174:97–115.
[284] Geddes J, Eudes F, Laroche A, Selinger LB. Differential expression of proteins in response to the interaction between the pathogen *Fusarium graminearum* and its host, *Hordeum vulgare*. Proteomics 2008;8:545–54.
[285] Yuan K, Zhang B, Zhang Y, Cheng Q, Wang M, Huang M. Identification of differentially expressed proteins in poplar leaves induced by *Marssonina brunnea* f. sp. Multigermtubi. J Genet Genomics 2008;35:49–60.
[286] Jung HW, Lim CW, Lee SC, Choi HW, Hwang CH, Hwang BK. Distinct roles of the pepper hypersensitive induced reaction protein gene CaHIR1 in disease and osmotic stress, as determined by comparative transcriptome and proteome analyses. Planta 2008;227:409–25.
[287] Chan Z, Qin G, Xu X, Li B, Tian S. Proteome approach to characterize proteins induced by antagonist yeast and salicylic acid in peach fruit. J Proteome Res 2007;6:1677–88.
[288] Casasoli M, Spadoni S, Lilley KS, Cervone F, De Lorenzo G, Mattei B. Identification by 2-D DIGE of apoplastic proteins regulated by oligogalacturonides in *Arabidopsis thaliana*. Proteomics 2008;8:1042–54.
[289] Lin YZ, Chen HY, Kao R, Chang SP, Chang SJ, Lai EM. Proteomic analysis of rice defense response induced by probenazole. Phytochemistry 2008;69:715–28.
[290] Segarra G, Casanova E, Bellido D, Odena MA, Oliveira E, Trillas I. Proteome, salicylic acid, and jasmonic acid changes in cucumber plants inoculated with *Trichoderma asperellum* strain T34. Proteomics 2007;7:3943–52.
[291] Shoresh M, Harman GE. The molecular basis of shoot responses of maize seedlings to *Trichoderma harzianum* T22 inoculation of the root: a proteomic approach. Plant Physiol 2008;147:2147–63.
[292] Zhu J, Alvarez S, Marsh EL, Lenoble ME, Cho IJ, Sivaguru M, et al. Cell wall proteome in the maize primary root elongation zone. II. Region-specific changes in water soluble and lightly ionically bound proteins under water deficit. Plant Physiol 2007;145:1533–48.
[293] Alvarez S, Marsh EL, Schroeder SG, Schachtman DP. Metabolomic and proteomic changes in the xylem sap of maize under drought. Plant Cell Environ 2008;31:325–40.
[294] Pandey A, Chakraborty S, Datta A, Chakraborty N. Proteomics approach to identify dehydration responsive nuclear proteins from chickpea (*Cicer arietinum* L.). Mol Cell Proteomics 2008;7:88–107.
[295] Bhushan D, Pandey A, Choudhary MK, Datta A, Chakraborty S, Chakraborty N. Comparative Proteomics analysis of differentially expressed proteins in chickpea extracellular matrix during dehydration stress. Mol Cell Proteomics 2007;6:1868–84.
[296] Wang MC, Peng ZY, Li CL, Li F, Liu C, Xia GM. Proteomic analysis on a high salt tolerance introgression strain of *Triticum aestivum/Thinopyrum ponticum*. Proteomics 2008;8:1470–89.
[297] Jiang Y, Yang B, Harris NS, Deyholos MK. Comparative proteomic analysis of NaCl stress-responsive proteins in *Arabidopsis* roots. J Exp Bot 2007;58:3591–607.
[298] Carpentier SC, Witters E, Laukens K, Van Onckelen H, Swennen R, Panis B. Banana (*Musa* spp.) as a model to study the meristem proteome: acclimation to osmotic stress. Proteomics 2007;7:92–105.
[299] Zang X, Komatsu S. A Proteomics approach for identifying osmotic-stress-related proteins in rice. Phytochemistry 2007;68:426–37.
[300] Lee DG, Ahsan N, Lee SH, Kang KY, Bahk JD, Lee IJ, et al. A proteomic approach in analyzing heat-responsive proteins in rice leaves. Proteomics 2007;7:3369–83.
[301] Hattrup E, Neilson KA, Breci L, Haynes PA. Proteomic analysis of shade-avoidance response in tomato leaves. J Agric Food Chem 2007;55:8310–8.
[302] Xu C, Sullivan JH, Garrett WM, Caperna TJ, Natarajan S. Impact of solar Ultraviolet-B on the proteome in soybean lines differing in flavonoid contents. Phytochemistry 2008;69:38–48.
[303] Cho K, Shibato J, Agrawal GK, Jung YH, Kubo A, Jwa NS, et al. Integrated transcriptomics, Proteomics, and metabolomics analyses to survey ozone responses in the leaves of rice seedling. J Proteome Res 2008;7:2980–98.
[304] Kieffer P, Dommes J, Hoffmann L, Hausman JF, Renaut J. Quantitative changes in protein expression of cadmium-exposed poplar plants. Proteomics 2008;8:2514–30.
[305] Bona E, Marsano F, Cavaletto M, Berta G. Proteomic characterization of copper stress response in *Cannabis sativa* roots. Proteomics 2007;7:1121–30.
[306] Führs H, Hartwig M, Molina LE, Heintz D, Van Dorsselaer A, Braun HP, et al. Early manganese-toxicity response in *Vigna unguiculata* L.—a proteomic and transcriptomic study. Proteomics 2008;8:149–59.
[307] Patterson J, Ford K, Cassin A, Natera S, Bacic A. Increased abundance of proteins involved in phytosiderophore production in boron-tolerant barley. Plant Physiol 2007;144:1612–31.
[308] Zhen Y, Qi JL, Wang SS, Su J, Xu GH, Zhang MS, et al. Comparative proteome analysis of differentially expressed proteins induced by Al toxicity in soybean. Physiol Plant 2007;131:542–54.
[309] Yang Q, Wang Y, Zhang J, Shi W, Qian C, Peng X. Identification of aluminum-responsive proteins in rice roots by a proteomic approach: cysteine synthase as a key player in Al response. Proteomics 2007;7:737–49.
[310] Li J, Wu XD, Hao ST, Wang XJ, Ling HQ. Proteomic response to iron deficiency in tomato root. Proteomics 2008;8:2299–311.
[311] Zhang Q, Xu F, Lambert KN, Riechers DE. Safeners coordinately induce the expression of multiple proteins and MRP transcripts involved in herbicide metabolism and detoxification in *Triticum tauschii* seedling tissues. Proteomics 2007;7:1261–78.
[312] Ahsan N, Lee DG, Lee SH, Kang KY, Bahk JD, Choi MS, et al. A comparative proteomic analysis of tomato leaves in response to waterlogging stress. Physiol Plant 2007;131:555–70.
[313] Kim SG, Kim ST, Kang SY, Wang Y, Kim W, Kang KY. Proteomic analysis of reactive oxygen species (ROS)-related proteins in rice roots. Plant Cell Rep 2008;27:363–75.
[314] Wan XY, Liu JY. Comparative Proteomics analysis reveals an intimate protein network provoked by hydrogen peroxide stress in rice seedling leaves. Mol Cell Proteomics 2008;7:1469–88.

Pemán, J. y Navarro R.M. 1998. Ed. Servei de Publicaciones Univ. Lleida. 102 pp.

Rossignol. M, Peltier. Jb, Mock. H.P, Matros. A, Maldonado. A, Jorrín. J. 2006. Proteomics. 6:5529-5548

Proteómica en *Quercus ilex*: aplicación al estudio de la variabilidad poblacional y la respuesta a estrés hídrico

Echevarría-Zomeño S[1, 2], Valero J[1, 2], Ariza J[1], Lenz C[3], Navarro RM[1], Jorrín JV.[2]

[1] Dpto. de Ingeniería Forestal, ETSIAM, Universidad de Córdoba; [2] Bioquímica y Proteómica Vegetal y Agrícola, Dpto. de Bioquímica y Biología Molecular, Universidad de Córdoba; [3] Applied Biosystems Deutschland, Darmstadt, Germany

Introducción

La encina (*Quercus ilex* L.) es la especie forestal más representativa del bosque mediterráneo. Muestra gran variabilidad fenotípica, algo característico de especies alógamas poco intervenidas por el hombre, debido a su origen y condiciones ambientales (Jiménez et al., 1999). En España, es una especie ampliamente utilizada en programas de reforestación (Navarro y Calzado, 2008, en prensa), y aunque se considera tolerante a la sequía, ésta constituye la primera causa de mortalidad de individuos tras el transplante (Navarro et al., 1998a, Navarro et al., 1998b).

En nuestros grupos de investigación se está llevando a cabo un proyecto multidisciplinar con la especie *Q. ilex* utilizando la proteómica como herramienta clave. Los primeros trabajos pusieron de manifiesto la existencia de gran variabilidad en el perfil proteico 2-DE de hojas, tanto intra e interpoblacional, como en un mismo individuo (Jorge et al., 2005), y pusieron de manifiesto cambios en el mapa proteico en respuesta a estrés moderado (Jorge et al., 2006).

Tras estos trabajos preliminares, se realizaron nuevos experimentos con un doble objetivo: (i) caracterizar la variabilidad genética de las poblaciones para su posterior catalogación utilizando bellotas, ya que su proteoma se considera más estable que el de hojas y (ii) dilucidar los mecanismos moleculares responsables de la respuesta a estrés hídrico a partir de la identificación de cambios en el patrón de proteínas en condiciones de sequía.

Material y métodos

Para los estudios de variabilidad se utilizaron bellotas de *Q. ilex* de cuatro poblaciones de Andalucía geográficamente distantes. Para los ensayos de estrés hídrico, se usaron hojas de plantones de *Q. ilex* de la procedencia Extremadurense-Sierra Morena Occidental sometidos a tres tratamientos: i) riego a capacidad de campo, ii) sequía (no riego) durante 14 días y iii) sequía durante 7 días más 7 días de riego. La extracción de proteínas se llevó a cabo por precipitación con TCA/acetona (Damerval et al., 1986). Las SDS-PAGE se realizaron en geles al 13% de poliacrilamida. Los IEF se llevaron a cabo en tiras IPG 5-8 de 17 cm. Las imágenes de los geles capturadas con un densitómetro (GS-800, BioRad) se analizaron con los programas QuantityOne o PDQuest (BioRad). Los spots diferenciales se escindieron de los geles y analizaron por MALDI-TOF/TOF o LC-MS/MS. La identificación de las proteínas se realizó utilizando los motores de búsqueda MASCOT o ProteinPilot (Applied Biosystems).

Resultados

Las distintas poblaciones presentan patrones de bandas (SDS-PAGE) o spots (2-DE) característicos y diferenciales, estableciéndose los agrupamientos filogenéticos mostrados en la Figura 1. Algunos de los spots diferenciales se identificaron como pertenecientes al grupo de las leguminas. En respuesta a sequía, se observó un descenso en la expresión de proteínas de la fotosíntesis (PSII OEC 1 y 2) y ruta glicolítica (triosafosfato isomerasa, fructosa bifosfato aldolasa). En el tratamiento de recuperación, las proteínas del

metabolismo fotosintético tienen valores similares a los del control, no así las de la glicolisis, que no se recuperan tras este corto periodo de riego.

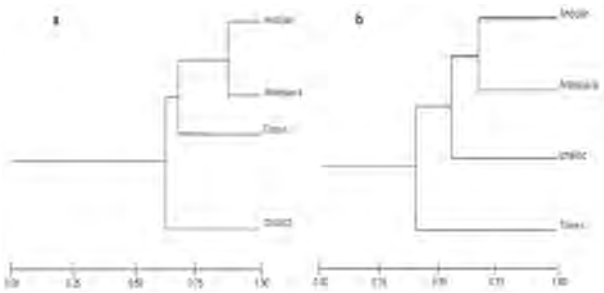


Figura 1. *Relaciones filogenéticos de las cuatro poblaciones de encinas andaluzas calculadas con los patrones de bandas (SDS-PAGE) (a) y de spots (2-DE) (b).*

Proceso/proteínas	Sequía	Recuperación
Fotosíntesis		
PSII OEC 1	↓	↔
PSII OEC 2	↓	↔
Glicolisis		
Triosafosfato isomerasa	↓	↓
Fructosa-1,6-bifosfato aldolasa	↔	↓
Estrés		
Peroxidasa	↑	↔
Peroxirredoxina	O	↔
Otras		
Glutamina sintetasa	↔	↓
Isoflavona reductasa	↓	↓
F23N19.10	↓	↔

Tabla 1. *Proteínas diferenciales identificadas y sus correspondientes efectos en sequía y recuperación con respecto al control. ↓, el spot correspondiente disminuye en intensidad; ↑, el spot aumenta en intensidad; O, el spot desaparece; ↔, el spot presenta una intensidad similar a la del control.*

Conclusiones

1. Es posible discriminar entre poblaciones de encina mediante electroforesis SDS-PAGE y 2-DE. Un aspecto clave de esta línea de investigación es el uso de órganos de proteoma estable, como son las semillas.

2. En condiciones de sequía moderada, se observó un descenso en la expresión de enzimas de la fotosíntesis y la glicolisis. Las primeras tienen un nivel similar al control en las plantas recuperadas mientras que las segundas no llegan a recuperarse tras el corto periodo de 7 días de riego.

Bibliografía

Jiménez, P., Gil, L. and Petit, R. J., in: Espinel, S. and Ritter, D. (Eds.), Congreso Internacional sobre Aplicación de la Biotecnología a la Genética Forestal (BIOFOR-99), Diputación Foral de Alava, Vitoria, Spain 1999, pp. 141-142.

Navarro, R. M., Campo, A., Alejano, R. and Alvarez, L., Informaciones Técnicas, Junta de Andalucía. Consejería de Agricultura y Pesca. D.G. de Investigación y Formación Agraria. Servicios de Publicaciones y Divulgación 1998a, p. 60.

Navarro, R. M., Del Campo, A., Gálvez, C. and Contreras, V., Caracterización del cultivo de planta forestal en contenedor, Conserjería de Agricultura y Pesca, Junta de Andalucía, Sevilla, Spain 1998b, p. 190.

Jorge, I., Navarro, R. M., Lenz, C., Ariza, D., et al., Proteomics 2005, 5, 222-234.

Jorge, I., Navarro, R. M., Lenz, C., Ariza, D. and Jorrin, J., Proteomics 2006, 6 Suppl 1, S207-214.

Damerval, C., de Vienne, D., Zivy, M. and Thiellement, H., Electrophoresis 1986, 7, 52-54.