



**INSTITUTO POTOSINO DE INVESTIGACIÓN
CIENTÍFICA Y TECNOLÓGICA, A.C.**

POSGRADO EN CIENCIAS EN BIOLOGIA MOLECULAR

**Comparative Proteomic Analysis Amongst Seeds
of Wild and Cultivated Amaranth Species by a Gel
Based Approach**

Tesis que presenta

Esaú Bojórquez Velázquez

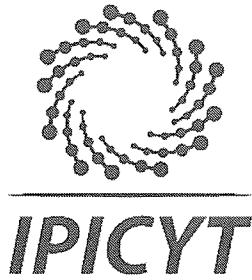
Para obtener el grado de

Doctor en Ciencias en Biología Molecular

Director de la Tesis:

Dr. Ana Paulina Barba de la Rosa

San Luis Potosí, S.L.P., noviembre de 2019



Constancia de aprobación de la tesis

La tesis “**Comparative Proteomic Analysis Amongst Seeds of Wild and Cultivated Amaranth Species by a Gel Based Approach**” presentada para obtener el Grado de Doctor en Ciencias en Biología Molecular fue elaborada por **Esaú Bojórquez Velázquez** y aprobada el ocho de noviembre del dos mil diecinueve por los suscritos, designados por el Colegio de Profesores de la División de Biología Molecular del Instituto Potosino de Investigación Científica y Tecnológica, A.C.

Dra. Ana Paulina Barba de la Rosa
Directora de la tesis

Dr. Samuel Lara González
Miembro del Comité Tutoral

Dr. Angel Gabriel Alpuche Solís
Miembro del Comité Tutoral

Dr. Alfredo Herrera Estrella
Miembro del Comité Tutoral

Dr. Eduardo Espitia Rangel
Miembro del Comité Tutoral

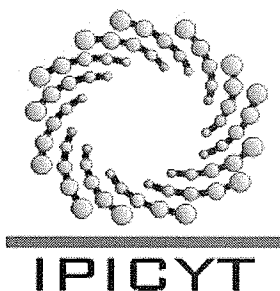


Créditos Institucionales

Esta tesis fue elaborada en el Laboratorio de Proteómica y Biomedicina Molecular de la División de Biología Molecular del Instituto Potosino de Investigación Científica y Tecnológica, A.C., bajo la dirección de la Dra. Ana Paulina Barba de la Rosa.

Durante la realización del trabajo el autor recibió una beca académica del Consejo Nacional de Ciencia y Tecnología (No. de registro 421670) y del Instituto Potosino de Investigación Científica y Tecnológica, A. C.

El desarrollo del presente trabajo fue financiado con el proyecto “Amaranto en la soberanía alimentaria: caracterización molecular de especies silvestres para el mejoramiento de cultivos comerciales”, dentro de la convocatoria Proyectos de Desarrollo Científico para Atender Problemas Nacionales 2014 con clave PDCPN2014-01, propuesta No. 248415.



Instituto Potosino de Investigación Científica y Tecnológica, A.C.

Acta de Examen de Grado

El Secretario Académico del Instituto Potosino de Investigación Científica y Tecnológica, A.C., certifica que en el Acta 108 del Libro Segundo de Actas de Exámenes de Grado del Programa de Doctorado en Ciencias en Biología Molecular está asentado lo siguiente:

En la ciudad de San Luis Potosí a los 8 días del mes de noviembre del año 2019, se reunió a las 11:10 horas en las instalaciones del Instituto Potosino de Investigación Científica y Tecnológica, A.C., el Jurado integrado por:

Dr. Samuel Lara González	Presidente	IPICYT
Dr. Ángel Gabriel Alpuche Solís	Secretario	IPICYT
Dr. Alfredo Heriberto Herrera Estrella	Sinodal externo	LANGEBIO
Dra. Ana Paulina Barba de la Rosa	Sinodal	IPICYT
Dr. Eduardo Espitia Rangel	Sinodal externo	INIFAP

a fin de efectuar el examen, que para obtener el Grado de:

DOCTOR EN CIENCIAS EN BIOLOGÍA MOLECULAR

sustentó el C.

Esaú Bojórquez Velázquez

sobre la Tesis intitulada:

Comparative Proteomic Analysis Amongst Seeds of Wild and Cultivated Amaranth Species by a Gel Based Approach

que se desarrolló bajo la dirección de

Dra. Ana Paulina Barba de la Rosa

El Jurado, después de deliberar, determinó

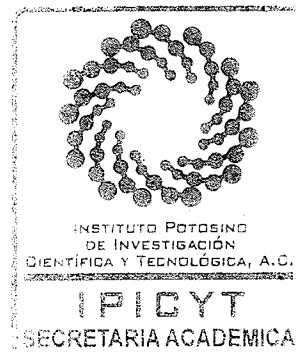
APROBARLO

Dándose por terminado el acto a las 13:15 horas, procediendo a la firma del Acta los integrantes del Jurado. Dando fe el Secretario Académico del Instituto.

A petición del interesado y para los fines que al mismo convengan, se extiende el presente documento en la ciudad de San Luis Potosí, S.L.P., México, a los 8 días del mes de noviembre de 2019.

Dr. Marcial Bonilla Marín
Secretario Académico

Mtra. Ivonne Lizette Cuevas Vélez
Jefa del Departamento del Posgrado



Dedicatoria

A mi familia, amigos, profesores y asesores, todos han contribuido en mi formación profesional y personal, por todo su tiempo y el apoyo brindado, muchas gracias.

A mi esposa, Miriam Livier, por el soporte tan fuerte que tu cariño me ha dado.

Agradecimientos

Gracias al Consejo Nacional de Ciencia y Tecnología (CONACYT) y al Instituto Potosino de Investigación Científica y Tecnológica (IPICYT).

Gracias a la Dra. Ana Paulina Barba de la Rosa por aceptarme como su estudiante. Gracias a los miembros de mi comité, el Dr. Eduardo Espitia Rangel, el Dr. Alfredo Herrera Estrella, el Dr. Ángel Gabriel Alpuche Solís y el Dr. Samuel Lara González, por sus enseñanzas, orientación y apoyo durante el transcurso de esta investigación, especialmente gracias por la formación científica que me han brindado.

Gracias al M. en C. Alberto Barrera Pacheco (Laboratorio de Proteómica y Biomedicina Molecular) por el soporte técnico en el análisis de espectrometría de masas, a la Dra. Olga Araceli Patrón Soberano (LINAN, IPICYT) por la asistencia técnica en microscopía electrónica, a la Dra. Aida J. Velarde Salcedo por los ensayos de inhibición enzimática, al Ing. Manuel Guillermo Silva Díaz (CNS) por la administración del servidor MASCOT y la gestión de bases de datos, y gracias al Dr. Antonio De León Rodríguez y al Dr. Víctor Emmanuel Balderas Hernández por su apoyo en el análisis GC-MS.

Gracias a la Mtra. Ivonne Lizette Cuevas Vélez, a la Ing. Edith Rodríguez Delgadillo y a la Lic. Teresa Casas Soubervielle del departamento de posgrado del IPICYT por todas sus atenciones y apoyo en la realización de trámites oficiales.

Gracias a todos mis amigos y colegas del Laboratorio de Proteómica y Biomedicina Molecular, por su ayuda y todos los buenos momentos durante estos años.

Gracias a la convocatoria CONACYT-Problemas Nacionales por el financiamiento del proyecto "Amaranto en la Soberanía Alimentaria"-Número de registro 248415.

Contenido

CONSTANCIA DE APROBACIÓN DE LA TESIS.....	II
CRÉDITOS INSTITUCIONALES.....	III
ACTA DE EXAMEN.....	IV
DEDICATORIA.....	V
AGRADECIMIENTOS.....	VI
RESUMEN.....	IX
ABSTRACT.....	X
CHAPTER 1.....	1
MORPHOLOGICAL, NUTRITIONAL, AND NUTRACEUTICAL CHARACTERIZATION OF WILD AND CULTIVATED AMARANTH SEEDS.....	1
1.1 INTRODUCTION.....	2
1.2 MATERIALS AND METHODS.....	3
1.2.1 Amaranth genotypes.....	3
1.2.2 Morphological and structural characterization of amaranth seeds.....	3
1.2.3 Amaranth flours proximate composition.....	3
1.2.4 Amaranth seeds protein extraction and electrophoretic profile.....	4
1.2.5 In-gel digestion and nLC-MS/MS protein identification.....	4
1.2.6 In vitro gastrointestinal digestion.....	5
1.2.7 Inhibition of dipeptidyl peptidase IV (DPPIV) activity.....	6
1.2.8 Inhibition of angiotensin converting enzyme (ACE) activity.....	6
1.2.9 Lipid extraction, GC-MS analysis, and squalene quantification.....	6
1.2.10 Statistical analysis.....	7
1.3 RESULTS.....	7
1.3.1 Morphological characterization.....	7
1.3.2 Proximal composition of amaranth seeds flours.....	9
1.3.3 Electrophoretic pattern and protein identification.....	14
1.3.4 Amaranth peptides with inhibitory activity against DPPIV and ACE.....	14
1.3.5 Lipids characterization of amaranth seeds.....	19
1.4 DISCUSSION.....	19
1.5 CONCLUSIONS.....	28
CHAPTER 2.....	29
CHARACTERIZATION OF HYDROPHILIC AND HYDROPHOBIC PROTEIN FRACTIONS OF WILD AND CULTIVATED AMARANTH SEEDS BY 1-DE-NLC-MS/MS.....	29
2.1 INTRODUCTION.....	30
2.2 MATERIALS AND METHODS.....	32
2.2.1 Plant materials.....	32
2.2.2 Protein extraction.....	32
2.2.3 1D-SDS-PAGE profile of amaranth proteins.....	33
2.2.4. Statistical analysis.....	33
2.2.5. In-gel digestion and nLC-MS/MS analysis.....	34
2.2.6. Protein identification using MS/MS data sets and database searching.....	34
2.2.7. Bioinformatic analysis.....	35
2.3 RESULTS.....	36
2.3.1. Hydrophobic fraction impacts on protein content and electrophoretic profile.....	36
2.3.2. Differentially accumulated proteins reflect the relationships amongst amaranth species.....	39
2.3.3. LEA proteins are species-specific.....	39
2.3.4. Differential accumulation of GBSSI and oil bodies related proteins amongst species.....	47

2.3.5. Identification of new paralogs of amaranth globulins	48
2.3.6. In silico molecular characterization of amaranth 11S globulins paralogs.....	51
2.4. DISCUSSION.....	58
2.5. CONCLUSIONS.....	63
CHAPTER 3	64
COMPARATIVE PROTEOMIC ANALYSIS AMONGST SEEDS OF WILD AND CULTIVATED AMARANTH SPECIES BY 2-DE-NLC-MS/MS	64
3.1. INTRODUCTION	65
3.2. MATERIALS AND METHODS	67
3.2.1 Biological material.....	67
3.2.2 Protein extraction.....	67
3.2.3 Two-Dimensional Electrophoresis (2-DE) and image analysis	67
3.2.4 In-gel digestion and nLC-MS/MS analysis	68
3.2.5 Protein identification using MS/MS data sets and database searching	69
3.3 RESULTS.....	69
3.3.1 Hydrophilic fraction	70
3.3.2 Hydrophobic fraction.....	93
3.3.3 Abundant proteins in amaranth wild species.....	113
3.4 DISCUSSION.....	115
3.4.1 11S globulin and GBSSI isoforms	117
3.4.2 Amaranth wild species characteristic proteins	117
3.5 CONCLUSIONS.....	123
GENERAL CONSIDERATIONS AND PROSPECTIVE	124
ANNEXES	126
APPENDIX 1. SCIENTIFIC PRODUCTION.....	127
REFERENCES	128

Resumen

Análisis proteómico comparativo entre semillas de especies silvestres y cultivadas de amaranto mediante un enfoque basado en electroforesis en gel

El amaranto se ha descrito como un sistema prometedor para la producción de alimentos, ya que proporciona semillas de excelente calidad nutricional, principalmente debido a sus proteínas con un equilibrio adecuado de aminoácidos esenciales y un contenido despreciable de prolaminas, así como a la presencia de compuestos nutracéuticos, como péptidos bioactivos con diversas funciones, entre las que destaca la inhibición de la dipeptidil peptidasa IV (DPPIV) y la enzima convertidora de angiotensina (ECA). Las semillas de amaranto también contienen una fracción de lípidos rica en escualeno, un hidrocarburo insaturado, al que se le han atribuido diversos efectos beneficiosos para la salud. Aunque las especies cultivadas destinadas a la producción de granos presentan características interesantes como la capacidad de resistir ciertos tipos de estrés abiótico, no se han realizado estudios sobre especies silvestres, que son una fuente potencial de genes con aplicabilidad agrobiotecnológica en la mejora de cultivos. Con base en lo anterior, este trabajo se enfoca en la comparación de las características bioquímicas y morfológicas entre las semillas de especies de amaranto silvestres, *A. hybridus* y *A. powellii*, y especies cultivadas, *A. cruentus* y *A. hypochondriacus* (cultivares tipo ceroso y no-ceroso). La primera parte incluye la caracterización microscópica, la determinación de la composición proximal y los perfiles totales de lípidos y proteínas de las semillas, así como la evaluación de péptidos bioactivos. En la segunda sección, se empleó un enfoque de extracción secuencial basado en polaridad para la comparación de los perfiles electroforéticos de proteínas en una dimensión, su identificación por nLC-MS/MS y análisis *in silico* de las proteínas diferenciales. El tercer enfoque viene dado por el análisis de proteínas hidrofílicas e hidrofóbicas mediante electroforesis en dos dimensiones (2-DE). Se observó que la estructura del perisperma de las semillas depende de la composición del almidón y correlaciona con las variaciones en los perfiles electroforéticos de las proteínas totales. Las enzimas DPPIV y ECA fueron inhibidas por péptidos de amaranto en una relación dosis-respuesta. Sintetas de almidón acopladas a granulo (GBSSI, Granule Bound Starch Synthase I), proteínas de reserva y proteínas abundantes de la embriogénesis tardía, se identificaron con acumulación diferencial heterogénea entre las especies. También se identificaron varios parálogos de globulinas 7S y 11S, algunos de ellos no reportados hasta ahora. El análisis 2-DE sugiere que las globulinas 11S y GBSSI están sujetas a modificaciones postraduccionales, principalmente fosforilaciones. Algunas proteínas relacionadas con metabolismo energético y de carbohidratos, polisacáridos de la pared celular, respuesta estrés y daño y con regulación génica, mostraron una alta acumulación solo en especies silvestres.

PALABRAS CLAVE: Estructura de almidón, Péptidos bioactivos, Escualeno, Proteínas de reserva de semillas, Globulinas 11S, LEAs, GBSSI, 2-DE-nLC-MS/MS.

Abstract

Comparative Proteomic Analysis Amongst Seeds of Wild and Cultivated Amaranth Species by a Gel Based Approach

Amaranth has been outlined as a promising system for food production since it provides seeds of excellent nutritional quality, mainly due to their proteins with adequate balance of essential amino acids and negligible prolamins content, and the presence of nutraceutical compounds, like encrypted peptides with several biological functions, amongst which the inhibition of dipeptidyl peptidase IV (DPPIV) and angiotensin converting enzyme (ACE) stands out. Amaranth seeds also contain an oily fraction rich in squalene, an unsaturated hydrocarbon, which has been attributed to diverse beneficial health effects. Although the cultivated species destined to grain production have interesting characteristics such as the ability to resist certain types of abiotic stresses, no studies have been carried out on wild species, which are a potential source of genes with agrobiotechnological applicability for crop improvement. Based on the foregoing, this work focuses on the comparison of biochemical and morphological characteristics amongst seeds of wild, *A. hybridus* and *A. powellii*, and cultivated, *A. cruentus* and *A. hypochondriacus* (waxy and non-waxy cultivars), amaranth species. The first part includes the microscopic characterization, the determination of the proximal composition and the overall total lipid and protein profiles of the seeds, as well as the evaluation of bioactive peptides. In the second section, a polarity based sequential extraction approach was employed for the comparison of one-dimensional protein electrophoretic profiles, their identification by nLC-MS/MS and *in silico* analysis of the differential accumulated proteins. The third part was given by the analysis of hydrophilic and hydrophobic proteins by two-dimensional electrophoresis (2-DE). Seeds perisperm structure was observed to be dependent on the starch composition and correlates with variations in electrophoretic profiles of total proteins. DPPIV and ACE activity were inhibited by amaranth peptides in a dose-response relationship. Granule bound starch synthase I (GBSSI), storage and late embryogenesis abundant proteins were identified with heterogeneous differential accumulation amongst species. Several paralogs of 7S and 11S globulins were also identified, some of them not reported so far. 2-DE analyses suggest that 11S globulins and GBSSI are subject to post-translational modifications, mainly phosphorylations. A set of carbohydrate and energy metabolism, cell wall polysaccharides, damage and stress response, and genic regulation related proteins displays high accumulation only in wild species.

KEYWORDS: Starch structure, Bioactive peptides, Squalene, Seed storage proteins, 11S globulins, Late embryogenesis abundant proteins, GBSSI, 2-DE-nLC-MS/MS.

CHAPTER 1

Morphological, nutritional, and nutraceutical characterization of wild and cultivated amaranth seeds

1.1 Introduction

Amaranth (*Amaranthus* spp.) is one of the oldest cultivated plants, which had great importance for the Aztec, Mayan, and Incas. Amaranth was grown as staple crop together with corn but was banned during the Spanish Conquest. Since the 70's amaranth resurged as an alternative crop not only due to its high nutritional value (high lysine and methionine content) but also because amaranth prolamins content is negligible (Huerta-Ocampo & Barba de la Rosa, 2011), which are the seed storage proteins responsible for the manifestation of celiac disease and cerebropathias. In this new century amaranth gained renewed importance due to its nutraceutical properties; amaranth proteins contain encrypted peptides, amongst the most studied are those with antihypertensive action (Huerta-Ocampo & Barba de la Rosa, 2011). Furthermore, the inhibitory peptides in amaranth seed proteins against dipeptidyl peptidase IV (DPPIV) and angiotensin converting enzyme (ACE) activity have been identified and characterized (Barba de la Rosa et al., 2010; Velarde-Salcedo et al., 2013). The oily fraction of amaranth seeds is rich in squalene, an unsaturated hydrocarbon to which hypocholesterolaemic properties have been attributed (Chaturvedi, Sarojini, & Devi, 1993).

In addition to nutritional characteristics, amaranth plants have attractive agronomic features. They grow where other cereals and vegetables cannot such as dry soils, high altitudes, and high temperatures (Huerta-Ocampo & Barba de la Rosa, 2011). Amaranth cultivation has increased worldwide, and breeders produced a large number of new varieties adapted to different environments. However, some of these new varieties are only new names for old varieties or landraces, hence the use of wild amaranth species with remarkable tolerance to several abiotic stresses such as *A. hybridus* and *A. powellii* are of great interest. Although these wild species are proposed as the ancestors of the main cultivated species used for seed production such as *A. cruentus* and *A. hypochondriacus*, their molecular relationships have not been established. The aim of the present study was to compare the morphological characteristics and bioactive compounds content of cultivated and wild amaranth species.

1.2 Materials and methods

1.2.1 Amaranth genotypes

Amaranth seeds of wild (*A. hybridus* and *A. powellii*) species as well as the most cultivated and studied species, *A. hypochondriacus* cv Nutrisol, and *A. cruentus* cv Amaranteca, were provided by the National Institute for Forestry, Agriculture and Livestock Research (INIFAP), Mexico. Two more cultivars of *A. hypochondriacus* were included in the study; Cristalina (non-waxy type) and Opaca (waxy type), which are derived from a heterozygous plant for this character by six generations of single seed descendent and were collected from Atzitzintla, Tlaxcala, Mexico.

1.2.2 Morphological and structural characterization of amaranth seeds

Seed weight was calculated by weighing 100 seeds on an electronic balance DV215CD Discovery (Ohaus, Parsippany, NJ, USA) with 0.01/0.1 mg accuracy. The weight of 100 seeds was extrapolated to 1000 seeds. Seed dimensions (diameter and width) were taken with a SteREO Discovery V8 (Carl-Zeiss, Oberkochen, GE). All measurements were done in triplicates. Images of whole seeds and cross-sections were obtained with the same stereoscope. Cross-sections were stained with an iodine solution (2% KI (w/v), 1% I₂ (w/v) for 30 s, washed with distilled water for 1 min and observed at the stereoscope. Cross- and paradermal-sections were visualized by scanning electron microscopy (SEM) with an ESEM model Quanta 200 (FEI, Hillsboro, OR, USA), from the National Laboratory of Nanosciences and Nanotechnology Research (LINAN) IPICYT.

1.2.3 Amaranth flours proximate composition

Seeds were cleaned, frozen in liquid nitrogen and milled using a KRUPS GX4100 (Solingen, GE) milled to obtain fine flour. Flour samples were stored in plastic tubes

at -80°C until analysis. Total nitrogen content was determined by micro-Kjeldahl method (AOAC, 2007, method 12.960.52), and total protein content was calculated using a 5.85 factor. Fat content was determined by the Soxhlet method (AOAC, 2007, 996.01 method). Crude fibre and ash contents were obtained according to AOAC (2007) methods 991.43 and 900.02, respectively. All determinations were made at least in triplicates.

1.2.4 Amaranth seeds protein extraction and electrophoretic profile

Total protein extracts, from the six amaranth species studied, were obtained by mixing 0.1 g of flour with 2 mL of a solution containing 7 M urea, 2 M thiourea, 2% (v/v) Triton X-100 and 0.05 M DTT. Suspensions were mixed by vortexing for 15 min at 4°C and centrifuged at $17,000\times g$ at 20°C , supernatants were recovered, and protein quantified using the Bradford protein assay (Bio-Rad, Hercules, CA, USA). Each sample was analysed by denaturing polyacrylamide gel electrophoresis (SDS-PAGE) in a discontinuous Tris-glycine system. The stacking and resolving gels were at 4% and 13.5% of acrylamide (29:1, acrylamide:bisacrylamide), respectively. Protein (15 μg) was loaded onto the gel and separated in a Mini-Protean III system (Bio-Rad), gel was run at 10 mA/gel for 30 min followed by 25 mA/gel until bromophenol blue reached the bottom of the gel. After electrophoresis, the gels were stained with a 0.05% Coomassie blue R-250 (USB Corporation, Cleveland, OH, USA) in 40% methanolic solution containing 10% acetic acid.

1.2.5 *In-gel* digestion and nLC-MS/MS protein identification

Protein bands were manually excised from gels, destained, reduced with 10 mM dithiothreitol and alkylated with 55 mM iodoacetamide. Protein digestion was carried out with sequencing-grade trypsin (Promega, Madison, WI, U.S.A.). Tryptic peptides were analysed with a nanoACQUITY UPLC System (Waters, Milford, MA, U.S.A.) coupled to a SYNAPT-HDMS Q-TOF (Waters) mass spectrometer. MS/MS spectra data sets were used to generate PKL files using Protein Lynx Global Server v2.4

(PLGS, Waters). Proteins were then identified using the MASCOT search engine v2.5 (Matrix Science, London, U.K.). Searches were conducted against the Viridiplantae subset of the NCBI nr protein database (6 686 534 sequences, May 2018). Trypsin was used as the specific protease, and one missed cleavage was allowed. The mass tolerance for precursor and fragment ions was set to 50 ppm and 0.1 Da, respectively. Carbamidomethyl cysteine was set as fixed modification and oxidation of methionine was specified as variable modification. The protein identification criteria included at least two MS/MS spectra matched at 99% level of confidence, and identifications were considered successful when significant MASCOT scores >50 were obtained, indicating the identity or extensive homology at $p < 0.01$ and the presence of a consecutive y ion series of more than three amino acids.

1.2.6 *In vitro* gastrointestinal digestion

A simulated gastrointestinal digestion *in vitro* model was carried out as reported before (Velarde-Salcedo et al., 2013). Briefly, 1 g of amaranth defatted flour was resuspended in 20 mL of 0.03 M NaCl pH 2.0. In order to inactivate proteases, the suspensions were heated in a water bath at 80 °C for 5 min and allowed to cool down at room temperature. Porcine pepsin (Sigma–Aldrich, St. Louis, MI, USA) previously dissolved in 0.03 M NaCl pH 2.0 was added in a 1:20 ratio (w/w enzyme to substrate). Samples were digested at constant pH for 3 h at 37 °C and pH was then adjusted to 7.5. A mixture of trypsin (Sigma–Aldrich) and pancreatin (Sigma–Aldrich) was prepared (1:1 w/w trypsin:pancreatin ratio in 0.1 N NaHCO₃), added to the digestive solution and incubated at constant pH for an additional 3 h period (1:20 w/w enzyme to substrate ratio for both the enzymes, trypsin and pancreatin). Digestion was stopped by heating the suspensions at 75 °C for 20 min and centrifuged at 13,000×g for 30 min. Peptides were ultra-filtrated through 10 kDa filters (Amicon Ultra-10 centrifugal filters, Sigma-Aldrich). The peptides concentration was determined by the Lowry-based DC Protein Assay (Bio-Rad) using BSA as a standard, and then stored at –20 °C until analysis.

1.2.7 Inhibition of dipeptidyl peptidase IV (DPPIV) activity

DPPIV activity was measured using the chromogenic substrate Gly-Pro-pNitroanilide (Sigma-Aldrich) as previously reported (Velarde-Salcedo et al., 2013). Briefly, 10 μ l of 100 ng/mL of DPPIV (Sigma-Aldrich) were added to 40 μ L amaranth peptides dissolved in 100 mM Tris pH 8.0 and 50 μ L of 1 mM Gly-Pro-pNitroanilide dissolved in Tris buffer. Mixture was incubated at 37 °C for 1 h. Absorbance was measured at 415 nm in a Multiskan Go plate reader (Thermo Fisher Scientific Inc., Waltham, Massachusetts, USA). Results were expressed as μ mol of nitroaniline/min based on a p-nitroaniline (Sigma-Aldrich) standard curve.

1.2.8 Inhibition of angiotensin converting enzyme (ACE) activity

Peptides with inhibitory activity against ACE were measured by spectrophotometric assay as reported before (Barba de la Rosa et al., 2010). Briefly 20 μ L of sample were added to 0.1 mL of 0.1 M potassium phosphate buffer (pH 8.3) containing 0.3 M NaCl and 5 mM hippuryl–histidyl–leucine (HHL, Sigma). ACE (5 mU) (EC 3.4.15.1, 5.1 U/mg, Sigma) was added and the reaction mixture was incubated at 37 °C for 30 min. The reaction was terminated by the addition of 0.1 ml of 1 M HCl. The hippuric acid formed was extracted with ethyl acetate, heat-evaporated at 95 °C for 10 min, dissolved in distilled water and measured in spectrophotometer at 228 nm. The activity of each sample was tested in triplicate. Captopril was used as a positive control. The IC₅₀ value was defined as the peptide concentration (mg/mL) needed to inhibit 50% ACE activity; it was calculated by an ACE inhibition (%) vs. log peptide concentration (mg/mL) linear regression.

1.2.9 Lipid extraction, GC-MS analysis, and squalene quantification

Lipids from amaranth seeds were extracted by mixing 0.1 g of amaranth flour with 1.75 ml of hexane, in constant agitation at room temperature by 5 h. Samples were centrifuged (17,000 \times g, 20 min, 25 °C) and supernatants transferred into new tubes.

Extracts were analysed by GC-MS using a 7820A/5977E System (Agilent Technologies, Santa Clara, California, USA), with a HP-5ms capillary column (Agilent Technologies) of 30 m length, 250 µm of inner diameter and 0.25 µm-film thickness. Samples, 3 µL, were injected in splitless mode. The column was held at 80 °C for 1 min after injection, the temperature programmed at 20 °C/min to 210 °C and held for 10 min more, then 10 °C/min to 280 °C and held for 35 min. Helium was used as carrier gas, at a constant column flow rate of 1 mL/min. The injector temperature was 250 °C and the detector temperature was 230 °C. The mass spectrometer was operated under Electron Impact Ionization at 70 eV with a mass range from 30 to 500 amu. Lipids were identified comparing their retention times and the mass spectra against NIST Mass Spectral Library v2.2. Results were expressed as the individual relative percentage of each lipid present in the sample. For squalene absolute quantification, an analytical standard (Sigma) was used to construct a calibration curve from 1 to 10 mg/L ($r^2 = 0.997$).

1.2.10 Statistical analysis

An analysis of variance (ANOVA) was carried out using the Sigma Plot software analysis v12.3 (Systat Software, Inc., San Jose, CA, USA) with Holm-Sidak test for paired analysis and considering $p < 0.05$ for statistically significant differences.

1.3 Results

1.3.1 Morphological characterization

A. powellii showed the lowest value for thousand seeds weight (TSW) with only 0.45 g and *A. hypochondriacus* cv Cristalina presented the highest TWS of 0.90 g (Table 1.1). *A. hypochondriacus* cv Nutrisol have the smallest seed diameter, 1.19 mm, and the smallest width was observed in *A. powellii* (0.88 mm). The largest seeds were those of *A. hybridus* and *A. hypochondriacus* cv Cristalina with dimensions of 1.32 × 1.15 mm and 1.33 × 1.13 mm, respectively. Although different

Table 1.1 Physical parameters of wild and cultivated amaranth species

Amaranth species	TSW (g)	Diameter (mm)	Width (mm)	D/W
<i>A. hybridus</i>	0.71 ±0.00 ^c	1.32 ±0.07 ^a	1.15 ±0.06 ^a	1.15
<i>A. powellii</i>	0.45 ±0.00 ^d	1.27 ±0.08 ^b	0.88 ±0.07 ^c	1.15
<i>A. cruentus</i>	0.81 ±0.00 ^b	1.29 ±0.08 ^a	1.16 ±0.06 ^a	1.11
<i>A. hypochondriacus</i> cv Opaca	0.86 ±0.00 ^{a,b}	1.22 ±0.10 ^{a,b}	1.11 ±0.07 ^a	1.15
<i>A. hypochondriacus</i> cv Cristalina	0.90 ±0.00 ^a	1.33 ±0.09 ^a	1.13 ±0.08 ^b	1.17
<i>A. hypochondriacus</i> cv Nutrisol	0.68 ±0.00 ^c	1.19 ±0.05 ^c	1.04 ±0.05 ^b	1.14

TSW, Thousand Seed Weight; **D/W**, Diameter to Width ratio. Mean values of three replicates ± standard deviation. Different superscript letter by column indicate statistically significant differences.

in diameter and width, most species conserve the same diameter/width ratio, with exception of *A. hypochondriacus* cv Cristalina, which present more oval shaped seeds (1.17 D/W ratio) and *A. cruentus* with the most rounded seeds (1.11 D/W ratio).

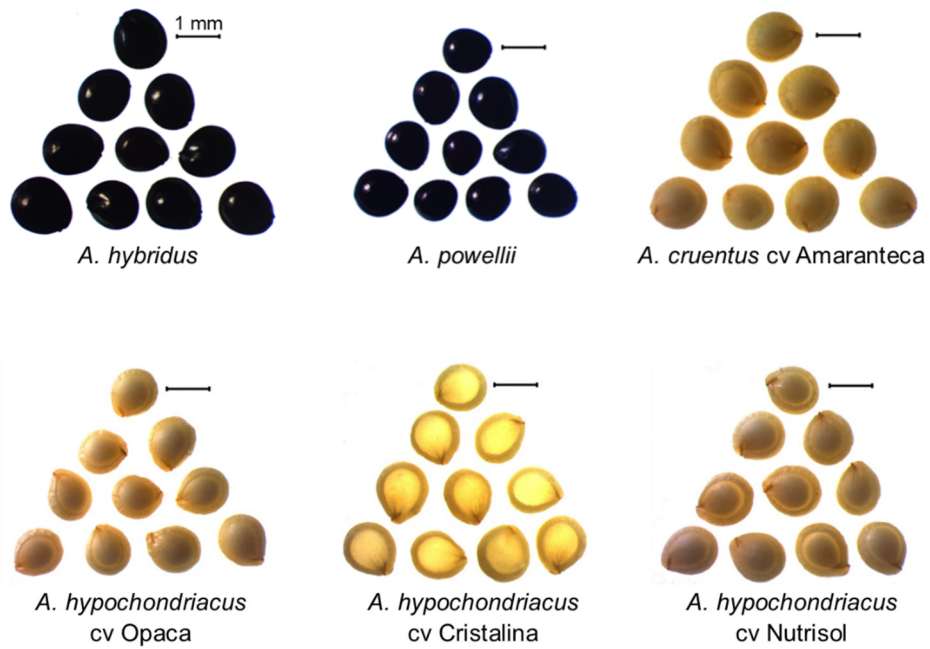
Phenotypic differences in amaranth seeds, which are characteristic of each species, were observed (Figure 1.1). The wild species are bright black in colour, while seeds of cultivated species are cream light. *A. powellii* have the smallest seeds while *A. hybridus* and *A. cruentus* are the largest ones. Seeds cross-cuts showed that the wild species *A. hybridus* and *A. powellii* are translucent; the cultivated species *A. cruentus* has opaque seeds while *A. hypochondriacus* cultivars were distinguished due to their translucent or opaque characteristics (Figure 1.2A). This vitreous characteristic has been related with the type, degree of cross-linking, and molecular weight distribution of proteins and starch in seeds, which is confirmed by iodine staining that highlighted the structures within the starch perisperm (Figure 1.2B).

Wild species and *A. hypochondriacus* cv Cristalina stained purple-blue corresponding to non-waxy lines with high amylose content, while the opaque species stained red-brown corresponding to waxy lines with low amylose content. Seeds cross-sections and paradermal cuts were observed by SEM microscopy (Figure 1.3) showing that in fact, *A. hybridus*, *A. powellii*, and *A. hypochondriacus* cv Cristalina have polyhedral structures in their vitreous perisperm, whereas the perisperms of *A. cruentus* cv Amaranteca, *A. hypochondriacus* cvs Nutrisol and Opaca did not displays the typical polyhedral structure of amaranth starch granules.

1.3.2 Proximal composition of amaranth seeds flours

The flours proximal composition from wild and cultivated amaranth seeds is shown in table 1.2. Although *A. powellii* produces the smallest seeds, this species has the highest protein (17.8%) and fat (8.1%) contents. *A. cruentus* is the species with the lowest protein content (14.8%), but the highest starch content (73.0%). On the other hand, *A. hybridus* and *A. hypochondriacus* cv Cristalina, with the largest seeds, are

A)



B)

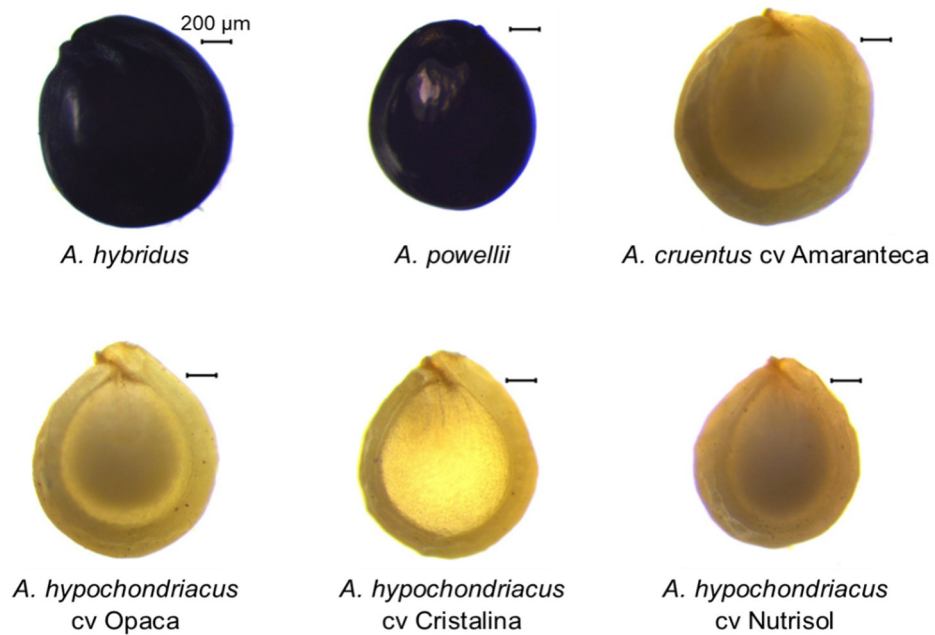


Figure 1.1. Global morphological characteristics of wild and cultivated amaranth species visualized in a group of seeds (**A**), and appreciated with detail by zooming in one single individual (**B**).

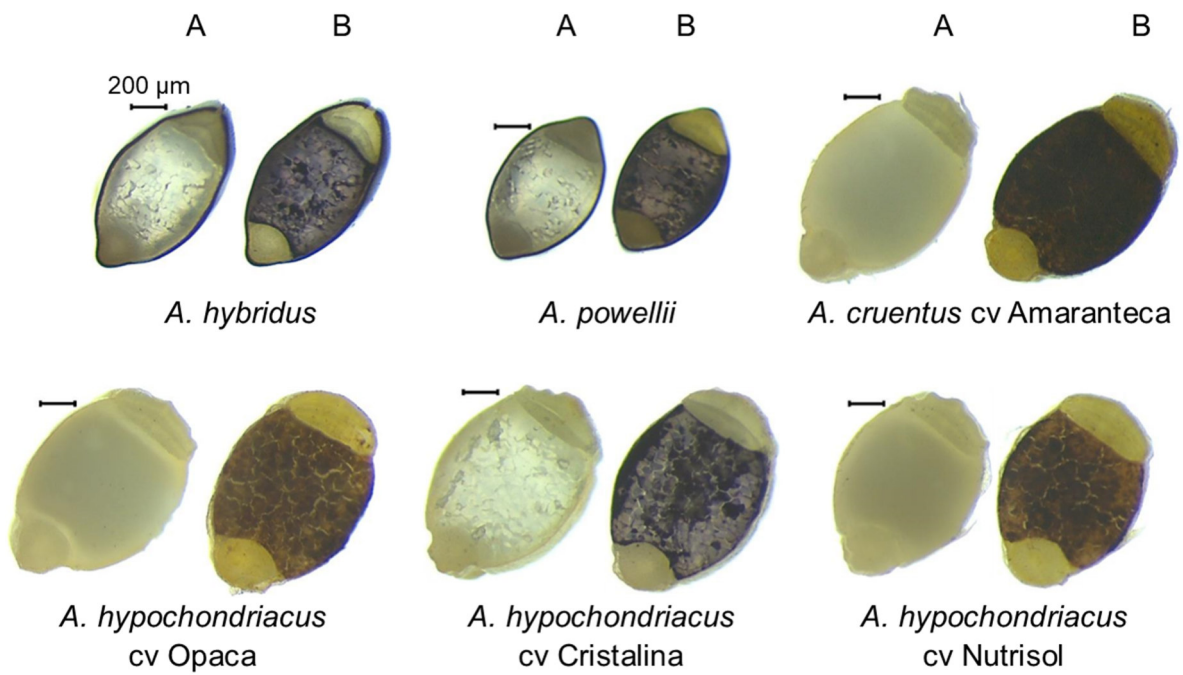
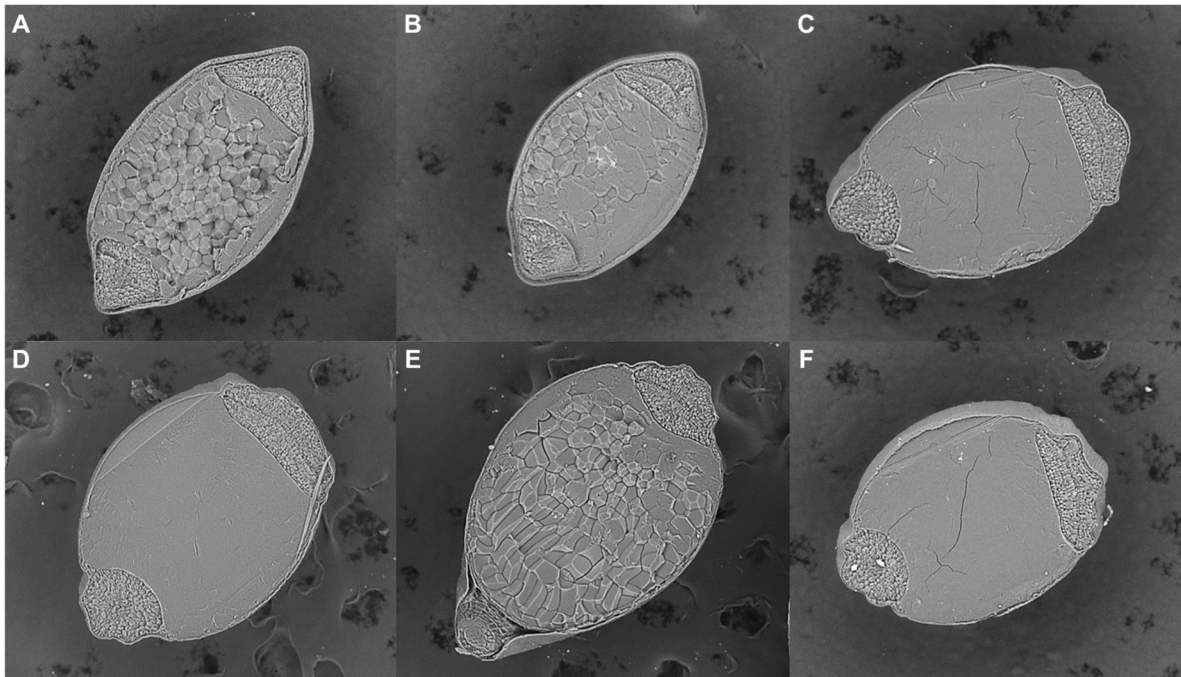


Figure 1.2. Cross-sections of seeds from wild and cultivated amaranth species before (A) and after (B) iodine staining.

A)



B)

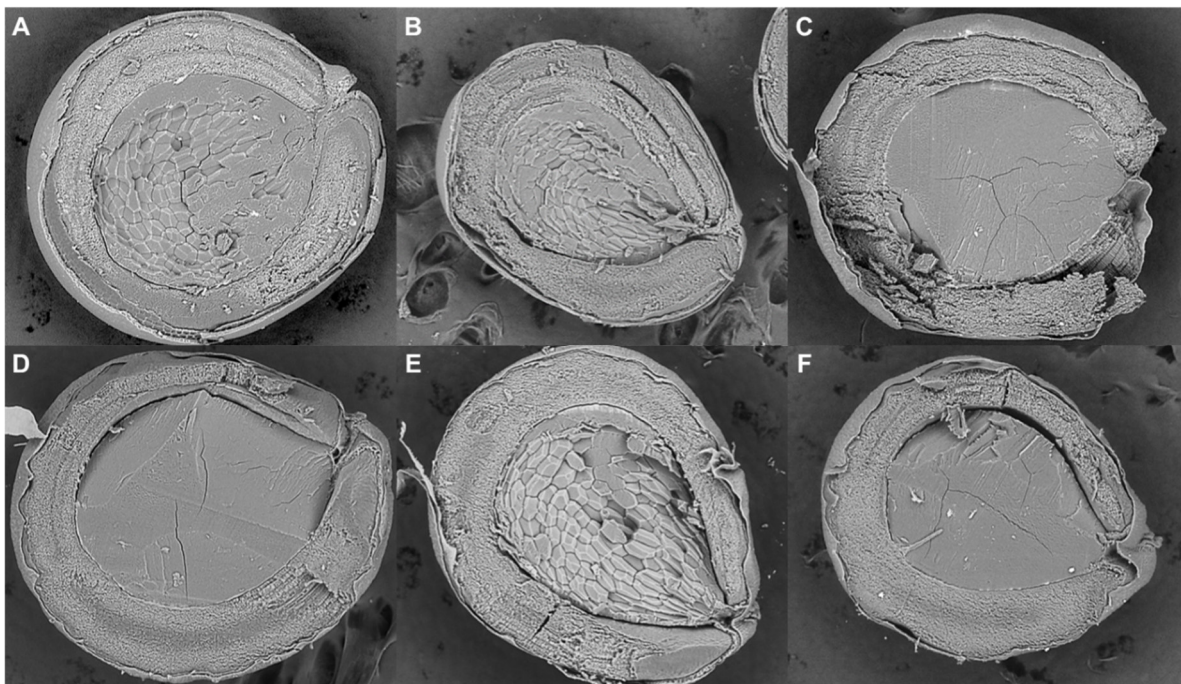


Figure 1.3. Scanning electron microscopy images of transversal (A) and paradermal (B) sections of amaranth seeds. A, *A. hybridus*; B, *A. powellii*; C, *A. cruentus* cv *Amaranteca*; D, *A. hypochondriacus* cv *Opaca* (waxy); E, *A. hypochondriacus* cv *Cristalina* (non-waxy); and F, *A. hypochondriacus* cv *Nutrisol*.

Table 1.2. Proximate composition of wild and cultivated amaranth species (%db).

Amaranth species	Protein ¹	Fat	Crude Fibre	Ash	Carbohydrates ²
<i>A. hybridus</i>	15.9±0.2 ^b	5.9±0.0 ^c	6.1±0.0 ^a	3.7±0.1 ^a	68.5±0.3 ^d
<i>A. powellii</i>	17.8±0.1 ^a	8.2±0.0 ^a	5.9±0.0 ^b	3.6±0.1 ^a	64.5±0.2 ^e
<i>A. cruentus</i> cv Amaranteca	14.8±0.4 ^c	6.9±0.0 ^b	2.5±0.1 ^e	2.8±0.1 ^b	73.0±0.6 ^a
<i>A. hypochondriacus</i> cv Opaca	16.7±0.8 ^b	6.9±0.1 ^b	3.5±0.1 ^d	3.0±0.0 ^b	69.9±0.8 ^c
<i>A. hypochondriacus</i> cv Cristalina	16.7±0.1 ^b	5.7±0.2 ^c	3.9±0.0 ^c	2.9±0.0 ^b	70.9±0.3 ^b
<i>A. hypochondriacus</i> cv Nutrisol	15.8±0.1 ^b	6.9±0.1 ^b	2.4±0.1 ^f	3.6±0.2 ^a	71.4±0.2 ^b

¹N×5.85; ²By difference; Mean values of three replicates ± standard deviation; different superscript letters by column indicate statistically significant differences at $p<0.05$.

the species with the lowest fat content (5.9 and 5.7%, respectively). Interestingly, *A. hybridus* has the highest crude fibre (6.1%) and ash (3.7%) contents. It is interesting that amongst the *A. hypochondriacus* species, the most commercial cultivar, Nutrisol, showed less protein content (15.8%) in comparison with cvs Opaca and Cristalina (16.7%).

1.3.3 Electrophoretic pattern and protein identification

Total proteins were analysed by SDS-PAGE (Figure 1.4). In all species and cultivars bands located at 35–37 kDa and 18–20 kDa were observed, which represent the acidic and basic subunits of the canonical 11S globulins. The most remarkable differences amongst species and cultivars analysed were observed in the range of 50–70 kDa. Both wild species as well as *A. hypochondriacus* cv Cristalina have a characteristic band around 65 kDa. *A. powellii* and *A. cruentus* share a band of 60 kDa. *A. hybridus* and all *A. hypochondriacus* cultivars showed a 55 kDa band. These three bands were cut from gel, analysed by nLC-MS/MS and identified as a Granule Bound Starch Synthase I (GBSSI, Figure 1.5 and Table 1.3).

1.3.4 Amaranth peptides with inhibitory activity against DPPIV and ACE

A simulated gastrointestinal digestion *in vitro* method was used to release the encrypted peptides from all amaranth samples. The capacity of released amaranth peptides to inhibit both DPPIV and ACE enzymes was measured. DPPIV inhibition increased in a dose-response relationship (Figure 1.6A), higher inhibition activity was detected at the highest tested concentration (3.2 mg/mL). At this concentration, *A. hypochondriacus* cv Opaca rendered the highest inhibitory activity reaching a 60% of DPPIV inhibition with an IC₅₀ of 1.6 mg/mL. *A. hypochondriacus* cv Cristalina and *A. powellii* displays the least DPPIV inhibition reaching only 40% at the highest tested peptide concentration. A similar ACE inhibitory activity profile was observed (Figure 1.6B). *A. hypochondriacus* cv Opaca peptides presented the highest activity reaching 80% inhibition at 3.2 mg/mL with an IC₅₀ of 0.6 mg/mL. *A. hypochondriacus*

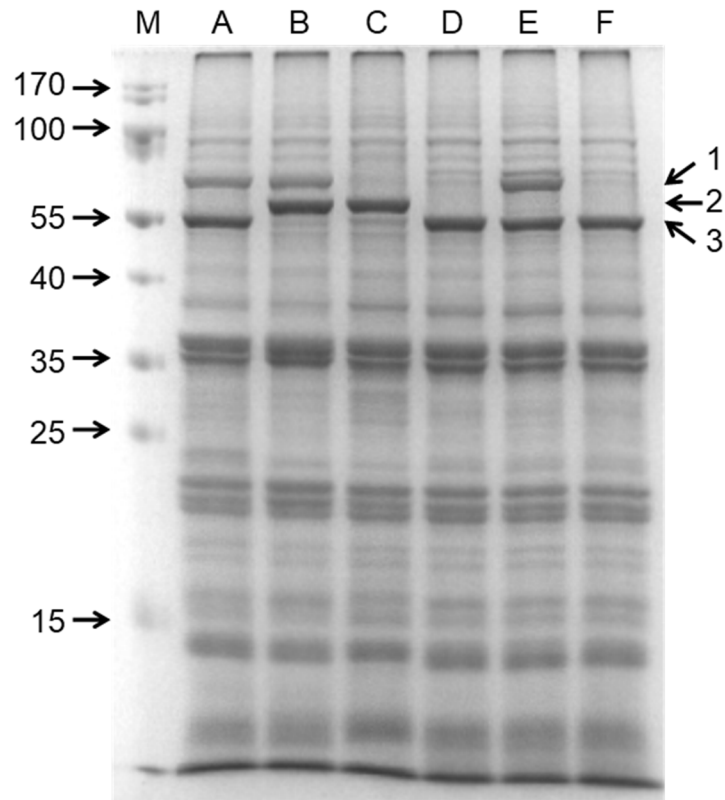


Figure 1.4. Electrophoretic pattern of total proteins extracted from amaranth seeds. Lanes: **M**, Molecular weight marker (kDa); **A**, *A. hybridus*; **B**, *A. powellii*; **C**, *A. cruentus* cv *Amaranteca*; **D**, *A. hypochondriacus* cv *Opaca* (waxy); **E**, *A. hypochondriacus* cv *Cristalina* (non-waxy); **F**, *A. hypochondriacus* cv *Nutrisol*. Arrows: 1 = 65 kDa; 2 = 60 kDa; 3 = 55 kDa.

A)

1	METVTSSHFV	SNFANTAMGS	SDPKLTLANN	ALKSNQMSTH	NGLRPLMSNI
51	DMLRLSNPK	STTVELRKER	FHAPFIRSGM	NVVFVGAEVA	PWSKTGGLGD
101	VLGGLPPALA	ARGHRVMTVS	PRYDQYRDGW	DTSVTVEFQV	GNRTETVRYF
151	HTYKRGVDRI	FVDHPLFLAR	VWGITGSKLY	GPK AGADYED	NQLRF SLLCQ
201	AALEAPRVLN	LNNNPNFSGP	YGENVVFIAN	DWHTALLPAY	LKAIYQPKGI
251	YNNAK VAFCI	HNIVYQGR F	LADYPRLHLP	EELRPVFEM	DGYDRPIKGR
301	KINWM AGIL	QSDR VVTVSP	YYAQELISGV	ER GVELDDVV	RQTG VTGIVN
351	GMDVQEWNP	TDKYIGINFN	ITTVMTAKPL	IK EALQAEVG	LPVDR NIPLI
401	GFIGRLEEQK	GSDILAEAIP	RFIKENVQIV	VLGTGKEVME	KQIEQLEILY
451	PEKARGVTKF	NSPLAHMIVA	GADEFMLIPSR	FEPGLIQLY	SMRYGTVPV
501	ASTGGLVDTV	KEGYTGFHMG	RFSANCDMVD	PADISAVETT	VHR ALTTYNS
551	PAMR EMVINC	MTQDFSWKEP	ARKWEELLS	LGVAGSRPGF	EGTESIPLAT
601	ENIATP				

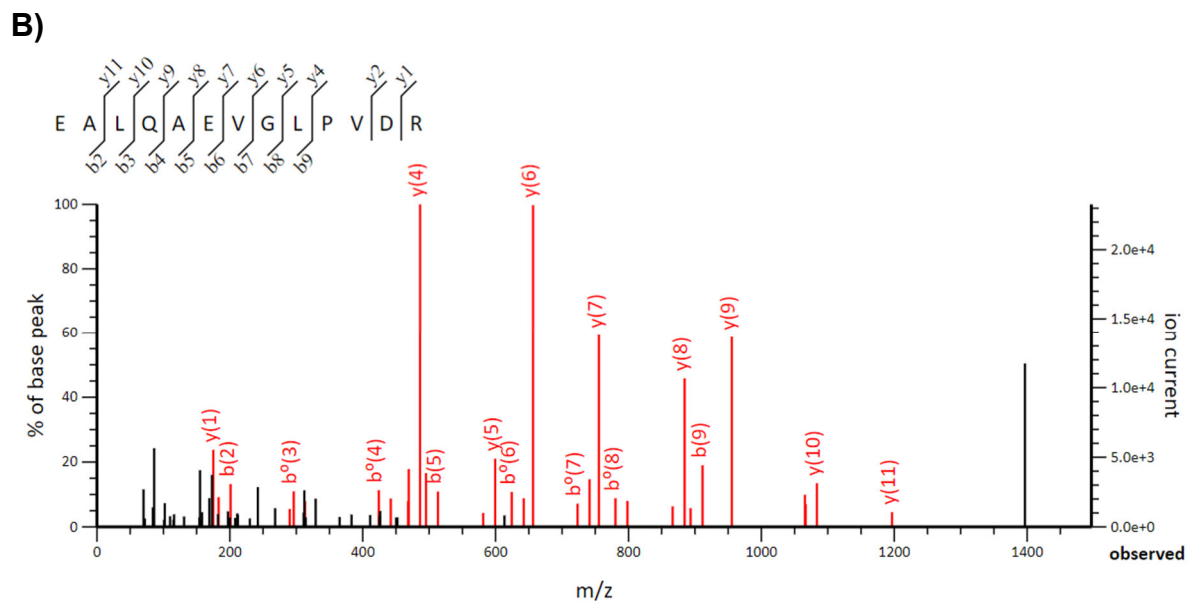


Figure 1.5. Identification of GBSSI by nLC-MS/MS. **A**, Peptides with significant scores that contribute to protein sequence coverage. **B**, Representative fragmentation pattern of one identified peptide.

Table 1.3. Identification of differentially accumulated bands amongst wild and domesticated amaranth species.

Band No. ^a	Protein name	Species ^b	Accession Number ^c	Exp Mr ^d	Theor Mr ^e	Sequence Coverage ^f (%)	Peptides ^g	Peptide Score ^h
1	Granule bound starch synthase I	<i>A. hypochondriacus</i>	BAJ09328	65	67.6	19	K.AGADYEDNQLR.F	79
							K.VAFCIHNIVYQGR.F	71
							K.AGILQSDR.V	63
							R.GVELDDVVR.Q	72
							K.EALQAEVGLPVDR.N	86
							K.GSDILAEAIPIR.F	81
							K.ENVQIVVLGTGK.E	77
							K.FNSPLAHMIVAGADFMLIPSR.F	78
							K.EGYTGFMGR.F	79
							R.ALTTYNSPAMR.E + Oxidation (M)	60
2	Granule bound starch synthase I	<i>A. hypochondriacus</i>	BAJ09328	60	67.6	16	K.AGADYEDNQLR.F	80
							R.FSLLCQAALAPR.V	104
							K.VAFCIHNIVYQGR.F	75
							R.VVTVSPYYAQELISGVER.G	80
							R.GVELDDVVR.Q	79
							K.EALQAEVGLPVDR.N	82
							K.GSDILAEAIPIR.F	73
							R.ALTTYNSPAMR.E	81
3	Granule bound starch synthase I	<i>A. hypochondriacus</i>	BAJ09328	55	67.6	5	K.AGADYEDNQLR.F	56
							K.EALQAEVGLPVDR.N	54
							K.ENVQIVVLGTGK.E	54

^aBand number as indicated in Figure 1.4. ^bSpecies matching sequence. ^cAccession number according to NCBI protein database. ^dExperimental mass (kDa) of identified proteins. ^eTheoretical mass (kDa) of identified proteins retrieved from the database. ^fSequence coverage (%). ^gIdentified peptide sequences. ^hMASCOT score for each of identified peptides.

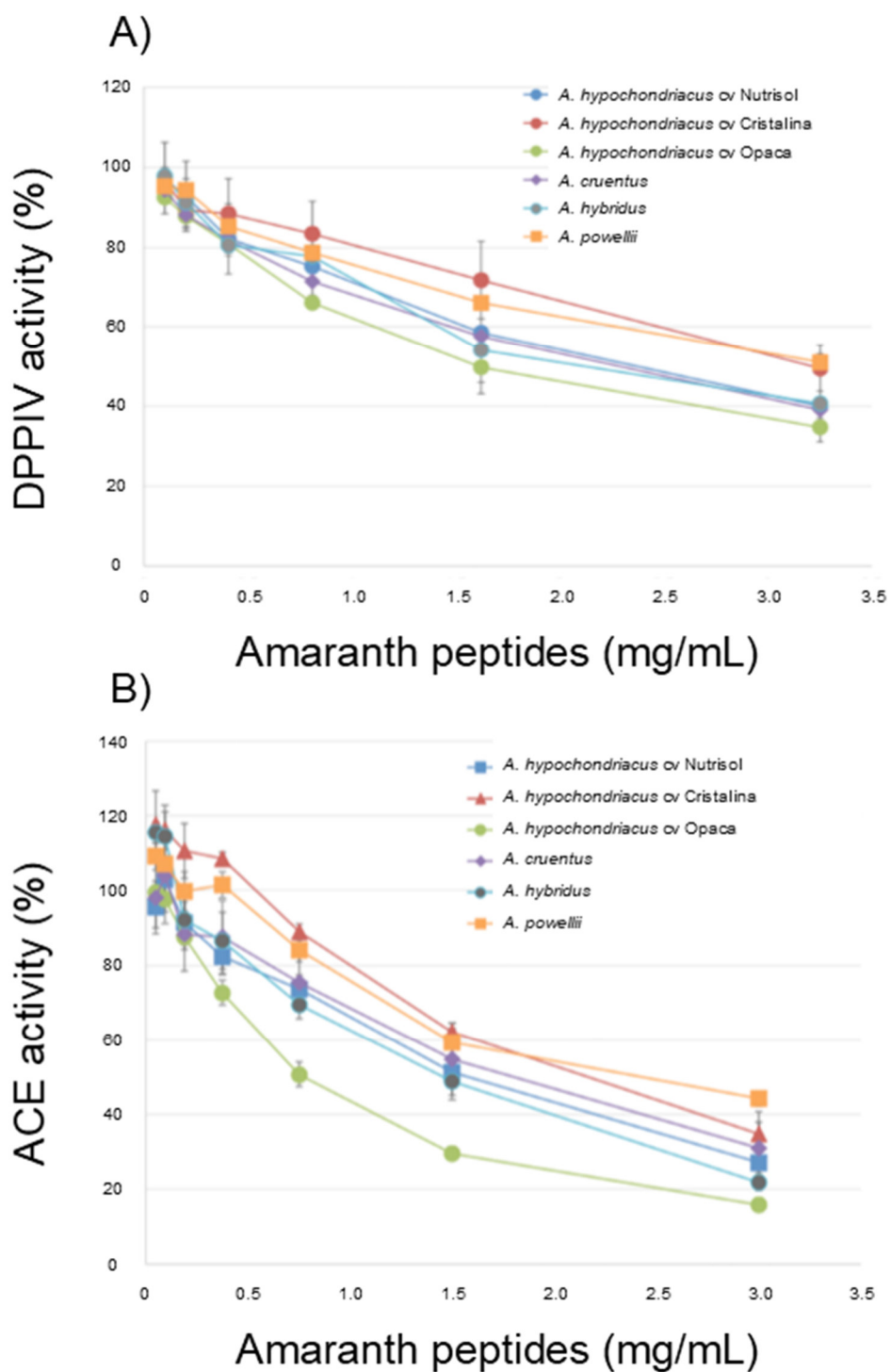


Figure 1.6. Inhibitory activity of amaranth peptides released by simulated gastrointestinal digestion in vitro against **A)** DPPIV and **B)** ACE.

cv Nutrisol, *A. hybridus* and *A. cruentus* showed an IC₅₀ of 1.5 mg/mL while *A. hypochondriacus* cv Cristalina and *A. powellii* the IC₅₀ was of 2.5 mg/mL.

1.3.5 Lipids characterization of amaranth seeds

The lipid composition analysed by GC-MS showed the presence of palmitic and linoleic acids in all species. Linoleic ethyl esters were present only in *A. hybridus* and *A. hypochondriacus* cvs Cristalina and Nutrisol, while oleic acid ethyl ester was only present in *A. hypochondriacus* cvs Cristalina and Nutrisol. Butyl ester of palmitic and stearic acids were detected in all samples. Stigmasterol, an important phytosterol, was detected in higher abundance in the wild species *A. powellii*, followed by *A. hypochondriacus* cvs Cristalina, and *A. hybridus*. Squalene, an unsaturated hydrocarbon, was detected in all samples but interestingly the highest abundance was detected in *A. cruentus* (Figure 1.7 and Table 1.4). Because of its importance and since relative abundance does not reflect the real quantity present in samples, squalene was quantified. Results showed that squalene concentration ranged from 0.197 to 0.335 g/100 g of seeds and these values in relation to oil content ranged from 2.85 to 4.86 g/100 g oil (Table 1.5).

1.4 Discussion

Wild ancestors of common cereals, such as rice and wheat, have been used as resources for quality improvement of cultivated grains (Cooper, 2015). However, despite the potential of a several crop ancestors to face the challenges of modern agriculture, there are few collections of wild relatives and even the available wild genetic resources are still under-utilized (McCouch et al., 2013).

Mexico is rich in genetic diversity of amaranth species such as *A. hybridus* and *A. powellii* and (Espitia-Rangel, 2012). These wild accessions have been considered as the ancestors of the cultivated species *A. cruentus* and *A. hypochondriacus* (Sauer, 1967), but concerns have been raised about this hypothesis. Hence, this morphological and molecular analysis of wild and cultivated

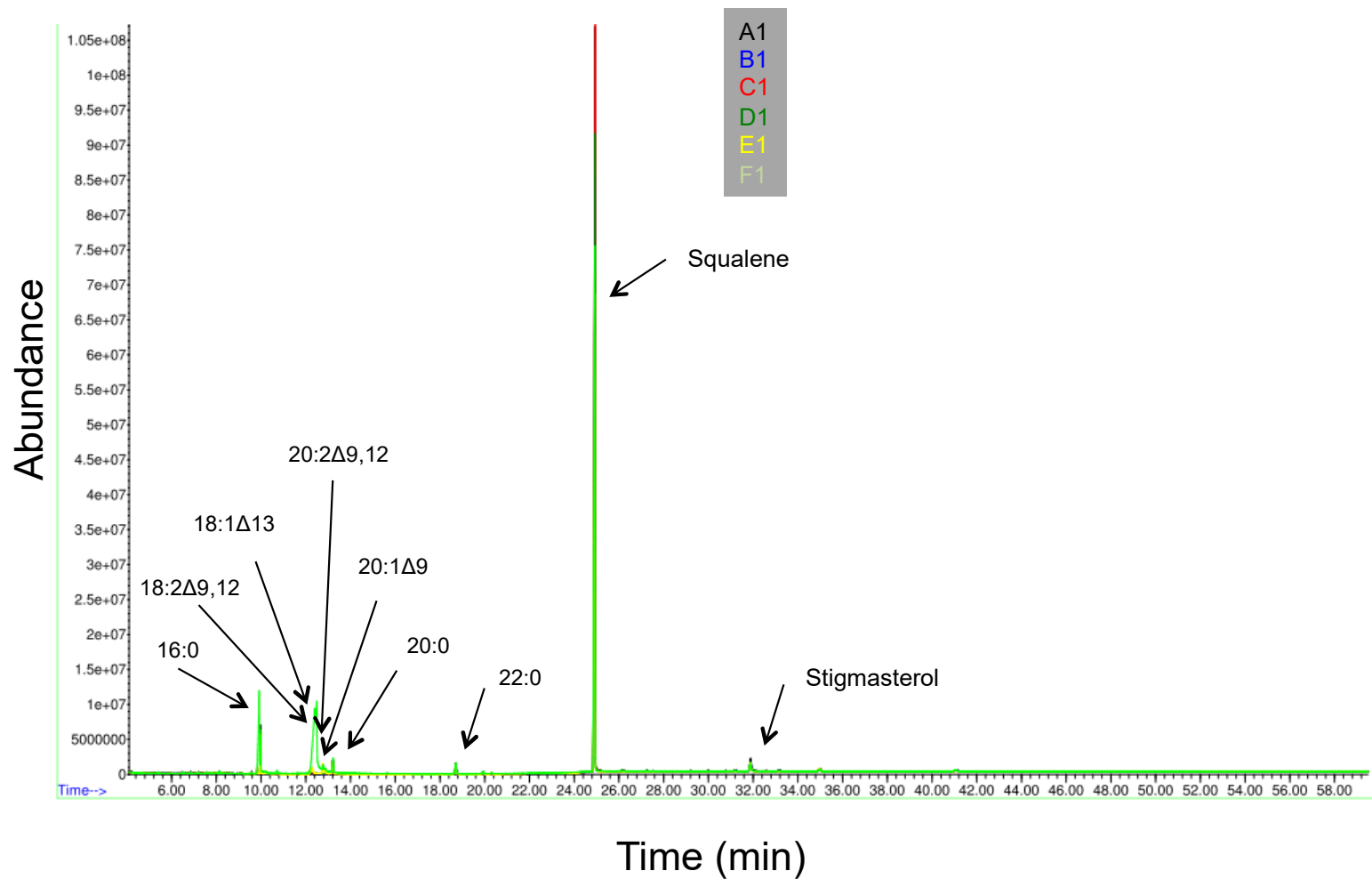


Figure 1.7. Lipid profile of amaranth total oil hexane fraction analysed by GC-MS. A1, *A. hybridus*; B1, *A. powellii*; C1, *A. cruentus* cv Amaranteca; D1, *A. hypochondriacus* cv Opaca (waxy); E1, *A. hypochondriacus* cv Cristalina (non-waxy); F1, *A. hypochondriacus* cv Nutrisol.

Table 1.4. Lipid composition of amaranth species total oil hexane fraction analysed by GC-MS.

Compound	RT (min)	Lipids (relative abundance) <i>Amaranthus</i> species/cultivars					
		<i>hybridus</i>	<i>powellii</i>	<i>cruentus</i>	Opaca	Cristalina	Nutrisol
Palmitic acid (16:0)	9.91	0.18 ±0.00 ^d	0.60 ±0.14 ^{bcd}	0.39 ±0.03 ^{cd}	1.11 ±0.08 ^b	0.91 ±0.06 ^{bc}	9.91 ±0.62 ^a
Linoleic acid (18:2Δ9,12)	12.41	0.15 ±0.01 ^d	0.65 ±0.07 ^{cd}	0.29 ±0.03 ^{cd}	0.89 ±0.04 ^{bc}	1.27 ±0.09 ^b	18.82 ±0.54 ^a
cis-13-Octadecenoic acid 18:1Δ13	12.49	0.23 ±0.02 ^d	1.02 ±0.21 ^c	0.58 ±0.03 ^d	1.07 ±0.05 ^c	1.77 ±0.08 ^b	11.71 ±0.24 ^a
Linoleic acid ethyl ester (20:2Δ9,12)	12.78	0.11 ±0.006 ^c	ND	ND	ND	1.93 ±0.03 ^a	1.33 ±0.07 ^b
Oleic acid, ethyl ester (20:1Δ9)	12.88	ND	ND	ND	ND	0.99 ±0.10 ^{bc}	1.00 ±0.04 ^a
Palmitic acid, butyl ester (20:0)	13.23	1.91 ±0.14 ^{bc}	2.49 ±0.18 ^a	1.53 ±0.05 ^d	2.02 ±0.17 ^b	2.49 ±0.075 ^a	1.64 ±0.02 ^{cd}
Stearic acid, butyl ester (22:0)	18.71	1.77 ±0.12 ^b	2.32 ±0.25 ^a	1.38 ±0.02 ^c	1.85 ±0.15 ^b	2.27 ±0.098 ^a	1.49 ±0.04 ^{bc}
SQUALENE	24.92	92.97 ±0.33 ^a	89.75 ±1.10 ^b	94.16 ±0.14 ^a	90.61 ±0.57 ^b	85.68 ±0.54 ^c	52.64 ±1.32 ^d
Stigmasterol	31.88	2.68 ±0.08 ^b	3.17 ±0.01 ^a	1.68 ±0.01 ^c	2.44 ±0.20 ^b	2.69 ±0.17 ^{ab}	1.44 ±0.04 ^c

Results expressed as the mean of relative abundance ± SD (n=3). Means values with different superscript letter by line indicate statistically significant differences at $p < 0.05$.

Table 1.5. Squalene quantification by GC-MS in wild and cultivated amaranth species.

Amaranth species	Squalene (g/100g)	
	in seeds	in oil
<i>A. hybridus</i>	0.246±0.03 ^b	4.17±0.45 ^b
<i>A. powellii</i>	0.252±0.02 ^b	3.12±0.27 ^c
<i>A. cruentus</i>	0.335±0.02 ^a	4.86±0.31 ^a
<i>A. hypochondriacus</i> cv Opaca	0.271±0.01 ^b	3.93±0.08 ^b
<i>A. hypochondriacus</i> cv Cristalina	0.217±0.00 ^c	3.80±0.05 ^b
<i>A. hypochondriacus</i> cv Nutrisol	0.197±0.01 ^d	2.85±0.12 ^c

Mean values ± SD of three determinations, different superscript letters by column indicate statistically significant differences at $p < 0.05$.

species could help validate the amaranth phylogeny and evolutionary relationships analyses (Espitia-Rangel, 2012).

For years, plant wild species have survived to abiotic and biotic stresses. Seeds have used dark or bright colours as a signal of toxic materials, as protective action against predators (Lev-Yadun, 2016). These pigmentations are due to polyphenols, plant metabolites that play a role in the protection of plants against ultraviolet radiation, pathogens, and herbivores (Alvarez-Jubete, Wijngaard, Arendt, & Gallagher, 2010). The absence of these pigments in cultivated amaranths is considered as a trait of domestication. However, there are black seeds such as *Pisum humile* and *P. fulvum*, which are highly edible but mimic various toxic seeds of legumes that grow in the same region such as *Lathyrus ochrus* (Lev-Yadun, 2016), and wild black seeded amaranths (a hybrid between *A. hypochondriacus* and *A. hybridus*), are grown in Michoacán-Mexico to make special black tamales (Sauer, 1967). Another characteristic of wild seeds is the hardness of their testa as protective tissues for mechanical defences against granivores attack. Hence, a light colour seed with soft testa has been the target for domestication. This can also be observed in amaranth, light seeds with soft and thin testa were selected for cultivation and domestication (Figures 1.1, 1.2 and 1.3).

Seed size is another characteristic related to the profitability of agricultural operations. Selection of big seeds, in terms of genetic changes, is related to breakdown of seed dispersal and seed dormancy (Fernández-Marín et al., 2014). The reported size for amaranth seeds is 0.9–1.7 mm diameter and TSW ranged from 0.6 to 1.0 g (Assad, Reshi, Jan, & Rashid, 2017), values that agree with our results, except for *A. powellii* that have the smallest diameter and TSW of 0.88 mm 0.45 g, respectively. Interestingly, the cultivated species *A. hypochondriacus* cv Cristalina, *A. cruentus*, and the wild *A. hybridus* bear the largest seeds. Genotypes with small seeds are correlated with low seed quality, since larger seed size is probably advantageous because of their better standability under agricultural conditions, and because of the greater plantlets size arising from them, and particularly important for crops with edible seeds (Espitia-Rangel, 2012; Lush & Wien, 1980). Weight has also

been widely used as characteristic for seed improvement and increased yield. Our results showed that the cultivated species *A. hypochondriacus* cvs Cristalina and Opaca, as well *A. cruentus* are the cultivars that presented the highest values of TSW.

Although several morphological characteristics have been studied in relation to seeds breeding, less attention has been paid and it is poorly understood how agricultural selection and cultivation affected the nutritional quality of seeds, which are important traits for crop improvement and meet with Food Sovereignty (Muñoz, Liu, Kan, Li, & Lam, 2017).

The scarce literature available on cultivars and their wild relatives are focused on protein and amino acid contents and very little information exists on other important nutritional traits such oils or starch. In general, a decrease in protein, fibre, and minerals is related with the increase in carbohydrates. Seeds with higher carbohydrate content are bigger and with higher TSW. This agrees with our results, *A. powellii*, wild species with the smallest size and TSW presented the highest protein and lowest carbohydrate contents (17.8% and 64.5%, respectively). Amongst the cultivated species, which produce bigger seeds, *A. cruentus* have the lowest protein but highest carbohydrate content (14.8 and 73.9%, respectively).

Amaranth lipid content varies from 5.7% to 8.1% (Table 1.2), values that are in the range of reported values from 6 to 20% (Assad et al., 2017). In soybean, oil content is higher in domesticated seeds as compared with its wild counterpart (Zong et al., 2017). However, in amaranth this relationship is not clear. The wild *A. powellii*, is the species with the highest fat content but *A. hybridus*, also a wild species, showed the lowest fat content. Our results are in agreement with those of Fernández-Marín et al. (2014), who reported that in some legumes such as soybean, peanut, lens, amongst others, domestication caused a decreased in total carotenoid content, especially a reduction in α - and γ -tocopherol was detected as domestication increased. In amaranth a reduction in abundance of stigmaterol from 3.17% presented in the wild species *A. powellii*, to 1.44% in the most domesticated species, *A. hypochondriacus* cv Nutrisol was observed (Table 1.4).

Amaranth grain is considered a good source of crude fibre, content that is higher than in rice, sorghum, oat, barley, and potato (USDA Food Composition Databases). *A. hybridus* showed the highest crude fibre content (6.1%) and the lowest content was found in *A. cruentus* and *A. hypochondriacus* cv Nutrisol, species with the highest starch content. Fibre is an important part of human nutrition; enough fibre intake is related with prevention of colon cancer. It has been reported that fibre in amaranth could be responsible for the control of blood cholesterol level, preventing the development of atherosclerosis and its complications (Caselato-Sousa, Ozaki, De Almeida, & Amaya-Farfan, 2014).

Carbohydrates in amaranth are of especial attention due to the very small size of the starch granules (0.5–2 µm), which gives functional characteristics of great interest in food applications (Kong, Bao, & Corke, 2009). Carbohydrates were very important during amaranth domestication that is thought to have occurred during the prehispanic times. Aztecs used sticky grain amaranths to make cakes as part of religious ceremonies (Sauer, 1967). The sticky grain selection was the origin of so-called waxy varieties of cereals and other starch-producing crops (Hunt, Denyer, Packman, Jones, & Howe, 2010). Amaranth waxy types were selected and nowadays are the cultivated species (*A. cruentus*, *A. hypochondriacus*). Sticky starch type is characterized by very low content of amylose, a character that modifies the glycaemic index. Waxy starch types generally have higher glycaemic loads, its consumption is related with a better physical performance and fast recovery associated with intense physical activity and with the reloading of glycogen storages after exercise (Wright, 2005). Therefore, higher-glycaemic cereal grains, which may have been an advantage and better tasting, treat in past cultures, today may impose a disadvantage for modern civilization where exercise and physical activity has decreased, and glycaemic loads are related with type-2 diabetes risk (WHO, 2016). So non-waxy varieties should be reconsidered for the generation of new amaranth cultivars and wild species are an important source of non-waxy starches and high protein contents.

The waxy and non-waxy amaranth types are related with the morphological observation, the wild species showed a well-defined polyhedral perisperm (Figure

1.3), but also with the differential accumulation of proteins (Figure 1.4). The differentially accumulated proteins were identified as GBSSI (Figure 1.5 and Table 1.3), enzyme responsible for amylose synthesis. It is important to mention that only the species that presented polyhedral structures observed by SEM (Figure 1.3), have the GBSSI band at 65 kDa (Figure 1.4). GBSSI isoforms of 60 and 55 kDa could be not functional leading to the synthesis of starches with different ratios of amylose/amylopectin, and therefore different rheological and physicochemical characteristics, which may be highly valued in the food and beverage science and technology industry.

In relation to nutraceutical characteristic, although limited data are available in this sense, it has been reported that ancient wheat are not healthier than modern wheat (Shewry & Hey, 2015). In this work we present evidence about that one cultivated species, *A. hypochondriacus* cv Opaca showed the highest inhibition against the DPPIV and ACE activities.

Type-2 diabetes is a chronic metabolic disorder considered as one of the major global health problems (WHO, 2016). The actual therapies to lower the hyperglycaemic state in patients with diabetes are based on the inhibition of DPPIV, enzyme responsible for degradation and inactivation of glucose-dependent insulinotropic polypeptide (GIP) and glucagon-like peptide-1 (GLP-1). GIP and GLP-1 are incretin hormones, which have a function to induce insulin secretion in the β -pancreatic cells. Synthetic DPPIV inhibitors are used as drugs therapies, however, there is a risk of side effects, DPPIV has several other functions than incretins inhibition (Matteucci & Giampietro, 2011) and natural DPPIV inhibitors are of great interest as therapy to promote a healthy life (Siró, Kápolna, Kápolna, & Lugasi, 2008). The possible mechanisms of action of amaranth peptides with inhibitory activity against DPPIV have been described previously (Velarde-Salcedo et al., 2013).

On the other hand, hypertension is one of the main risk factors for cardiovascular diseases; hypertension might affect 1 billion individuals worldwide. ACE plays an important role in the regulation of blood pressure by catalysing the production of the vasoconstrictor Angiotensin II and inactivating the vasodilator

Bradykinin. Thus, ACE-inhibitory drugs are commonly used to control high blood pressure in hypertensive subjects. It is reported that the main source of ACE inhibitory peptides is fermented milk with IC₅₀ values ranged from 0.47 to 1.70 mg/mL (Gonzalez-Gonzalez, Tuohy, & Jauregi, 2011). Amaranth IC₅₀ value was 0.6 mg/mL, which is in the range of milk peptides. Peptides YP, LPP, LRP, VPP, and IKP have been detected in amaranth seed protein hydrolyzates, the IKP peptide has been described as one of the most potent inhibitor of ACE activity (Huerta-Ocampo & Barba de la Rosa, 2011).

Interestingly, although cultivated amaranth species presented higher carbohydrates content, the amount of fat was not dramatically decreased (5.7–8.1%). We found that the oil composition in amaranth was characterized for the high levels of squalene, reaching values from 2.85 g/100g oil in *A. hypochondriacus* cv Nutrisol and up to 4.86 g/100g oil in *A. cruentus* (Table 1.2). He and Corke (2003) reported values from 1.0% to 7.3% squalene/oil, depending on the amaranth cultivar/specie analysed, while D'Amico and Schoenlechner (2017) reported concentrations from 2.26 to 11.19%.

Amaranth can modulate cholesterol levels in serum, which is due to its content of squalene (Chaturvedi et al., 1993; D'Amico & Schoenlechner, 2017). The recommended squalene intake (0.25–0.5 mg a day) may lower blood cholesterol levels reducing the risk of atherosclerosis and heart attack (Reddy & Couvreur, 2009).

Palmitic, linoleic, and cis-octadecanoic acids were detected (Figure 1.7 and Table 1.4). The ethyl esters of linoleic and oleic acids and butyl esters of palmitic and stearic acids were detected as minor components. In this regard, it was reported that in carob (*Ceratonia siliqua* L.) seeds, the most abundant fatty acids were the methyl-esters of oleic acid (C18:1), linoleic acid (C18:2n6), palmitic acid (C16:0), and stearic acid (C18:0) (Gubbuk, Kafkas, Guven, & Gunes, 2010).

Several clinical studies have shown that a high trans-fatty acid diet causes adverse changes in the plasma lipoprotein profile, with an increase in LDL and a decrease in HDL (Siddhuraju & Becker, 2001). In the present study, no trans-fatty acids such as elaidic and linolelaidic, myristic, behenic, erucic and lignoceric acids were detected.

1.5 Conclusions

Based on these results, we propose to *A. powellii* as an interesting option to generate amaranth cultivars with higher protein contents in their grains. *A. hybridus* showed the highest crude fibre content (6.1%), while *A. cruentus* and *A. hypochondriacus* cv Nutrisol had the highest starch content. Wild species and *A. hypochondriacus* cv Cristalina presented a perisperm with polyhedral well-defined structures and share the presence of a 65 kDa band corresponding to GBSSI; while *A. hypochondriacus* cvs Nutrisol and Opaca and *A. cruentus* cv Amaranteca showed a starch with low or no amylose content. The higher inhibition of DPPIV activity was detected at the highest concentration of peptides (3.2 mg/mL). *A. hypochondriacus* cv Opaca rendered the highest activity reaching until 60% of DPPIV inhibition with an IC₅₀ of 1.6 mg/ml. Regarding ACE inhibitory activity, also *A. hypochondriacus* cv Opaca showed the highest activity reaching of 80% inhibition at 3.2 mg/mL with an IC₅₀ of 0.6 mg/mL. Lipids in amaranth varied amongst species and cultivars, squalene highest concentrations were detected in *A. cruentus* followed by *A. hybridus*. This knowledge could be useful for the improvement of amaranth phenotyping with special focus on food quality and health-promoting compounds, hence wild species rediscovery will provide more information to support amaranth breeding.

CHAPTER 2

Characterization of hydrophilic and hydrophobic protein fractions of wild and cultivated amaranth seeds by 1-DE-nLC-MS/MS

2.1 Introduction

Food security is threatened by both the growing human population, estimated to reach around 9.3 billion by the year 2050, and the loss of crops due to climate changes and soil deterioration (Leprince, Pellizzaro, Berriri, & Buitink, 2016; Lobell, Schlenker, & Costa-Roberts, 2011). Seeds are the centre to crop production, human nutrition, and food security (McCouch et al., 2013; Muñoz et al., 2017), they contain the full genetic complement of the plant, allowing it to survive even under prolonged periods of stress conditions (Finch-Savage & Bassel, 2016; Wozny, Kramer, Finkemeier, Acosta, & Koornneef, 2018). Then it is of important concern to collect and preserve the germplasm of commercial species as well as their wild relatives, which have survived several climate changes and are valuable resources of genetic information that could be useful in the development of crop breeding strategies to solve current and future agricultural challenges (Lobell et al., 2011; McCouch et al., 2013; Muñoz et al., 2017).

Orthodox seeds can survive the removal of most of their cellular water and can be stored in dry state for a long period of time. Desiccation tolerance and maintenance of seeds quiescent state are associated with wide range of systems related with cell protection, detoxification, and repair (Finch-Savage & Bassel, 2016; Nguyen, Cueff, Hegedus, Rajjou, & Bentsink, 2015). The presence of proteins such as the late embryogenesis abundant (LEA) proteins, heat shock proteins (HSPs), and seed storage proteins (SSPs) confer seeds desiccation tolerance, allowing them to survive in dry state preserving their germination ability and propagation after long-term storage conditions (Righetti et al., 2015; Zinsmeister et al., 2016).

LEA proteins are suggested to play an important role in seed desiccation tolerance (Tunnacliffe & Wise, 2007), they are known to stabilize membranes against the deleterious effects of drying. Further, LEAs can prevent protein aggregation during freezing and drying and interact with and stabilize liposomes in the dry state (Thalhammer, Hundertmark, Popova, Seckler, & Hinch, 2010). Some LEAs can stabilize sugar glasses (Shimizu et al., 2010) suggesting that they play a role in longevity, which is a crucial factor for the conservation of genetic resources and to ensure proper seedling establishment and crop yield (Hundertmark, Buitink,

Leprince, & Hinch, 2011). On the other hand, SSPs are a major source of dietary protein for human nutrition. SSPs beyond serving as a nutrient reservoir they may play specific functions during seed formation (Nguyen et al., 2015; Shah et al., 2015) and could have a key role in seed longevity (Müntz, Belozersky, Dunaevsky, Schlereth, & Tiedemann, 2001). SSPs play a fundamental role in germination and seedling growth (Mouzo, Bernal, López-Pedrouso, Franco, & Zapata, 2018). Due to their abundance and high propensity to oxidation, SSPs are considered a powerful reactive oxygen species (ROS) scavenging system that could protect cellular components that are important for embryo survival (Davies, 2005; Sano et al., 2016).

Amaranth is a crop that had great importance for Aztec, Mayan, and Inca cultures. However, Spaniards prohibited its cultivation due to its link with pagan ceremonies (Sauer, 1967). Nevertheless, during the past two decades, reports on amaranth nutritional and nutraceutical characteristics have increased, leading to a new era in the history of amaranth cultivation (Huerta-Ocampo & Barba de la Rosa, 2011). The importance of amaranth as a crop for human nutrition is due to the high quality of its proteins. Amaranth seed proteins contain an adequate balance of essential amino acids (Bressani & García-Vela, 1990), with values close to nutritional human requirements, being particularly rich in lysine and methionine, which are deficient in cereals and legumes, respectively (Huerta-Ocampo & Barba de la Rosa, 2011; Valcárcel-Yamani & Caetano Da Silva Lannes, 2012). Furthermore, the content of prolamins, the SSPs fraction responsible for the manifestation of celiac disease, is negligible or practically null (Janssen et al., 2017). The genus *Amaranthus* consists of about 70 species distributed in very diverse habitats in terms of climatic conditions and geographical location (Aguilar-Hernández et al., 2011; Huerta-Ocampo et al., 2014), of which only three species, *A. caudatus*, *A. cruentus*, and *A. hypochondriacus* are cultivated as grain amaranths for human consumption, the last two being native to Mexico (Espitia-Rangel, Mapes-Sánchez, Nuñez-Colín, & Escobedo-López, 2010). The most probable ancestors or wild relatives of these species are *A. hybridus* and *A. powellii*, which grow under harsh conditions throughout the Mexican territory. The wide natural variation in amaranth offers the opportunity to identify markers that could be important for the nutrition, protection

and longevity of seeds, which would result in the development of high productivity cultivars.

The aim of this study was to characterize the protein electrophoretic profiles of seeds from wild amaranths *A. hybridus* and *A. powellii* and compared them with the cultivated species such as *A. cruentus* and *A. hypochondriacus*, carried out using 1-D-SDS-PAGE (One dimension-sodium dodecyl sulphate-polyacrylamide gel electrophoresis) and nLC-MS/MS (Nano liquid chromatography coupled to tandem mass spectrometry) as well as *in silico* analyses.

2.2 Materials and methods

2.2.1 Plant materials

Seeds of two black-seeded wild species *A. hybridus* and *A. powellii*, and two cream-seeded cultivated species *A. cruentus* cv Amaranteca and *A. hypochondriacus* cultivars Cristalina, Opaca, and Nutrisol, for a total of six samples, were submitted to analysis. Biological materials were kindly provided by the National Institute for Forest, Agricultural and Livestock Research (INIFAP, Mexico).

2.2.2 Protein extraction

For protein extraction seeds were frozen in liquid nitrogen and milled using a KRUPS GX4100 (Solingen, GE) mill to obtain fine flour. Flours were defatted with hexane in a 1:10 (w/v) ratio. The flour-hexane mixture was homogenized using vortex at maximum speed for 15 min at 4 °C, then centrifuged at 15,000×g for 30 min at 4 °C in a Beckman Avanti J-26S XPI centrifuge (Beckman, California, USA). The supernatant was discarded, and the precipitate air-dried. Proteins of polar nature were extracted from the defatted flour using 0.1 M Tris base, pH 8.5, containing 10% (v/v) glycerol and 2 mM PMFS (Sigma–Aldrich, St. Louis, MI, USA) at 1:20 (w/v) ratio. Mixture was agitated by vortex for 15 min at 4 °C and centrifuged at 17,000×g for 30 min at 4 °C. For extraction of hydrophobic proteins (including non-polar,

membrane, and cell wall proteins), the residue resulting from the hydrophilic fraction was resuspended in a solution of 7 M urea, 2 M thiourea, 2% (w/v) CHAPS, 2% (v/v) Triton X-100, mixed and centrifuged as mentioned above. Protein concentration was determined using the Protein Assay reagent (Bio-Rad), and bovine serum albumin as standard. All extractions and measurements were carried out in triplicates. Protein extracts of three independent biological replicates were applied to 1D-SDS-PAGE as described below.

2.2.3. 1D-SDS-PAGE profile of amaranth proteins

The hydrophilic and hydrophobic protein fractions were analysed by 1D-SDS-PAGE in a discontinuous Tris-glycine gels using 4 and 13.5% of acrylamide final concentration for the stacking and resolving gels, respectively. Protein extracts (50 µg) from each sample were loaded and separated in a SE 600 Ruby chamber (GE Healthcare, Little Chalfont, Buckinghamshire, UK) at 10 mA/gel for 1 h followed by 25 mA/gel for 4 h. After electrophoresis, gels were stained with a 0.05% Coomassie Brilliant Blue R-250 (USB Corporation, Cleveland, OH, USA) in 40% methanolic solution containing 10% acetic acid and destained with the same solution without the dye. Gels were digitalized in a Gel Doc XR+ Imaging System apparatus (Bio Rad) and densitometry analysis was performed with Quantity One software v4.5 (Bio Rad).

2.2.4. Statistical analysis

Densitometric data was submitted to an analysis of variance (ANOVA) with Holm-Sidak test using the Sigma Plot software v12.3 (Systat Software, Inc., San Jose, CA, USA), considering $p < 0.05$ for statistically significant differences. Bands with statistically different intensities for at least one species were selected for mass spectrometry analysis. Principal Component Analysis (PCA) and Agglomerative Hierarchical Clustering (AHC) were done using XLSTAT software (Addinsoft, Paris, France).

2.2.5. *In-gel* digestion and nLC-MS/MS analysis

Differentially accumulated protein bands were excised manually from the gels, destained, reduced and alkylated as described by Huerta-Ocampo et al. (2014). Protein digestion was carried out overnight at 37 °C with sequencing-grade trypsin (Promega). Nanoscale LC separation of tryptic peptides was performed with a nanoACQUITY UPLC System (Waters, Milford, MA, USA) equipped with a Symmetry C18 precolumn (5 µm, 20 mm × 180 µm, Waters) and a BEH130 C18 (1.7 µm, 100 mm × 100 µm, Waters) analytical column. The lock mass compound, [Glu1]-Fibrinopeptide B (Sigma-Aldrich), was delivered by the auxiliary pump of the nanoACQUITY UPLC System at 200 nL/min at a concentration of 100 fmol/mL to the reference sprayer of the Nano-Lock-Spray source of the mass spectrometer. Mass spectrometric analysis was carried out in a SYNAPT-HDMS Q-TOF (Waters). The spectrometer was operated in V-mode, and analyses were performed in positive mode ESI. The TOF analyser was externally calibrated with [Glu1]-Fibrinopeptide B from m/z 50 to 2422. The data were lock-mass corrected post-acquisition using the doubly protonated monoisotopic ion of [Glu1]-Fibrinopeptide B. The reference sprayer was sampled every 30s. The RF applied to the quadrupole was adjusted such that ions from m/z 50–2000 were efficiently transmitted. MS and MS/MS spectra were acquired alternating between low-energy and elevated-energy mode of acquisition (MS^e).

2.2.6. Protein identification using MS/MS data sets and database searching

MS/MS spectra data sets were used to generate PKL files using Protein Lynx Global Server v2.4 (Waters). Proteins were then identified using PKL files and the MASCOT search engine v2.5 (Matrix Science) against the *A. hypochondriacus* transcriptome and proteome data base v1.0 (23,054 sequences) available at <https://phytozome.jgi.doe.gov/> (Clouse et al., 2016). Trypsin was used as the specific protease, and one missed cleavage was allowed. The mass tolerance for precursor and fragment ions was set to 50 ppm and 0.1 Da, respectively. Carbamidomethyl

cysteine was set as fixed modification and oxidation of methionine was specified as variable modification. The protein identification criteria included at least two MS/MS spectra matched at 99% level of confidence, and identifications were considered successful when significant MASCOT individual ion scores > 33 were detected, indicating identity or extensive homology statistically significant at $p < 0.01$. Identifications were considered true only for peptide matches above identity threshold false discovery rate (FDR) $\leq 5\%$. To estimate the relative abundance of each protein per band, it was used the exponentially modified protein abundance index (emPAI) (Ishihama et al., 2005). BLAST algorithm was used for homology search against the Viridiplantae and *Arabidopsis thaliana* subsets of the UniProtKB database (<https://www.uniprot.org/blast/>).

2.2.7. Bioinformatic analysis

WebLogo's were constructed using 73 sequences of 11S globulins including Amaranthaceae, Brassicaceae, Chenopodiaceae, Cucurbitaceae, Fabace, Pedaliaceae, Poaceae and Polygonaceae families, downloaded from the Viridiplantae subset of the NCBI protein sequence repository (<http://weblogo.berkeley.edu/>; Crooks, Hon, Chandonia, & Brenner, 2004); (<https://www.ncbi.nlm.nih.gov/protein/>; NCBI Resource Coordinators, 2013). Search for conserved domains was done in different servers and databases, SMART (<http://smart.embl.de>; Letunic & Bork, 2018), PROSITE (<http://prosite.expasy.org/>; Sigrist et al., 2013), Pfam (<http://pfam.xfam.org>; Finn et al., 2016), InterPro (<http://www.ebi.ac.uk/interpro/>; Finn et al., 2017) and the NCBI's CDD (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=cdd>; Marchler-Bauer et al., 2005). Protein domains architecture images were generated with the PROSITE MyDomains-Image Creator tool (<https://prosite.expasy.org/mydomains/>; Hulo et al., 2008). Multiple sequence alignments were performed using Clustal Omega with default settings (<https://www.ebi.ac.uk/Tools/msa/clustalo/>; Sievers et al., 2011). Phylogenetic analysis and percentage amino acid composition were estimated with MEGA software v7.0.21 (Kumar, Stecher, & Tamura, 2016), the phylogenetic tree

was constructed with the neighbour-joining method and a bootstrap test of 1000 replicates and edited with iTOL (Letunic & Bork, 2016). For structural modelling, protein sequences were submitted to the I-TASSER server (<https://zhanglab.ccmb.med.umich.edu/I-TASSER/>; Roy, Kucukural, & Zhang, 2010), PDB files visualization and molecular graphics were performed with the UCSF Chimera package v1.11.2 (Pettersen et al., 2004).

2.3 Results

2.3.1. Hydrophobic fraction impacts on protein content and electrophoretic profile

To achieve greater coverage of seed proteins for analysis, extraction was carried out using a sequential approach based on protein polarity (Romero- Rodríguez, Maldonado-Alconada, Valledor, & Jorriin-Novo, 2014). Results showed that *A. hypochondriacus* cvs Opaca and Nutrisol had more hydrophilic proteins (Figure 2.1). However, differences in total protein content is reflected by the amount of hydrophobic protein fraction, hence that *A. powellii* has the highest protein content (173.5 mg/g), followed by *A. hypochondriacus* cv Cristalina and *A. hybridus* (147.9 and 140.8 mg/g, respectively). *A. cruentus* was the species with the lowest total protein content (108.8 mg/g).

Electrophoretic profile of the hydrophilic fractions showed protein bands throughout all the separation range from below 10 kDa to above 220 kDa (Figure 2.2 A). The most intense bands were observed at 33, 37, and 52 kDa. In contrast, the hydrophobic fraction showed lower number of bands, which were represented mainly by three groups, one between 20 to 24 kDa, the second from 32 to 35 kDa, and the last group, a highly variable region was formed with bands from 50 to 70 kDa (Figure 2.2B). In this fraction the presence or absence of bands (marked with a black arrow) amongst species was more evident than in the hydrophilic fraction. The histograms represent the differences in accumulation of some selected protein bands.

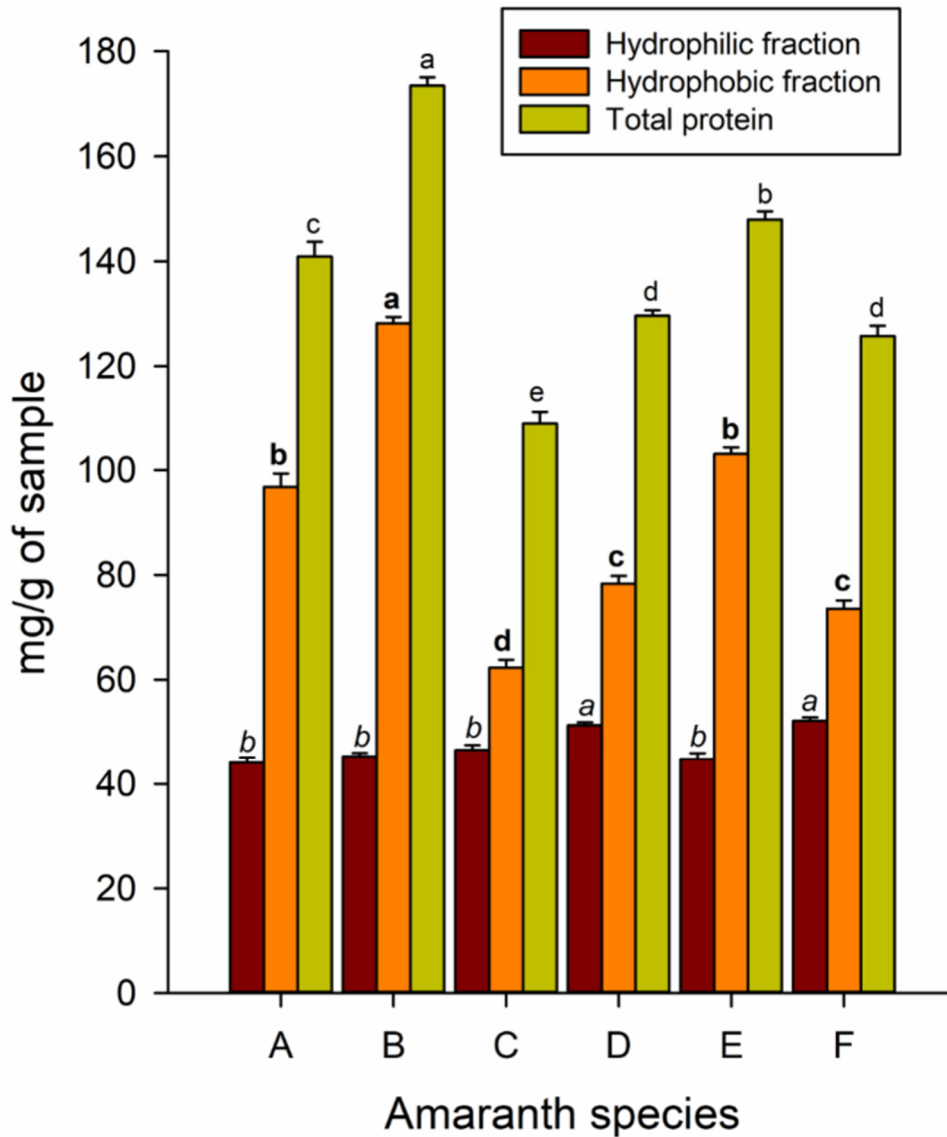


Figure 2.1. Bradford protein quantification of hydrophilic and hydrophobic proteins extracted from flour of wild and domesticated amaranth species. Protein quantification was carried out using the Bradford method. **A**, *A. hybridus*; **B**, *A. powellii*; **C**, *A. cruentus* cv Amaranteca; **D**, *A. hypochondriacus* cv Opaca (waxy); **E**, *A. hypochondriacus* cv Cristalina (non-waxy); **F**, *A. hypochondriacus* cv Nutrisol. Different letter above the bars indicates statistically differences at $p < 0.05$.

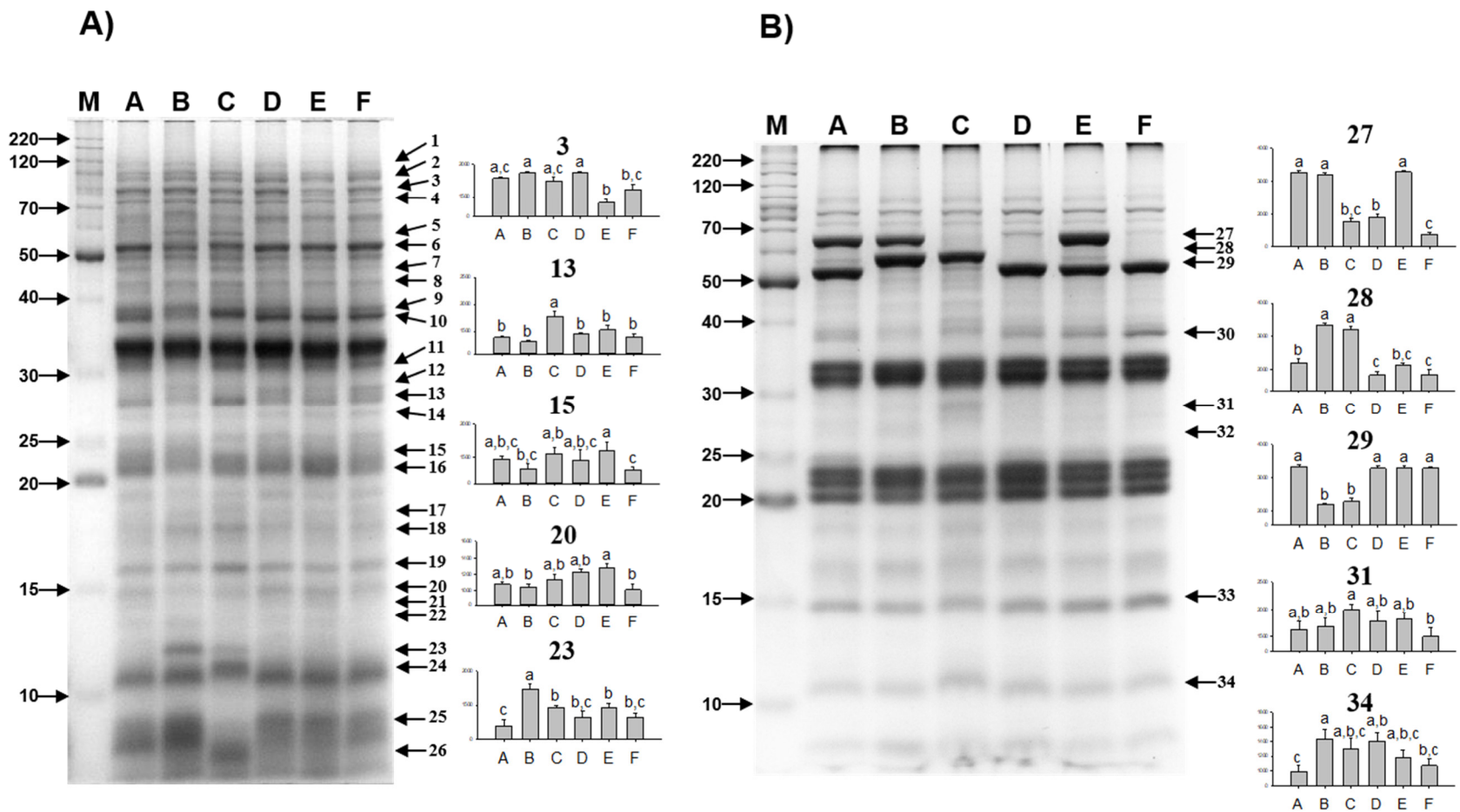


Figure 2.2. 1D-SDS-PAGE profile of amaranth seed proteins. **A)** Hydrophilic proteins, **B)** Hydrophobic proteins. Lanes: **M**, molecular weight marker (kDa); **A**, *A. hybridus*; **B**, *A. powellii*; **C**, *A. cruentus* cv *Amaranteca*; **D**, *A. hypochondriacus* cv *Opaca* (waxy); **E**, *A. hypochondriacus* cv *Cristalina* (non-waxy); **F**, *A. hypochondriacus* cv *Nutrisol*. Arrows at the right side of each gel indicates the differentially accumulated protein bands selected for nLC-MS/MS identification. Densitometric analyses from selected bands are shown in graphics. Different letter above the bars indicates statistically differences at $p < 0.05$.

2.3.2. Differentially accumulated proteins reflect the relationships amongst amaranth species

Differentially accumulated protein bands were excised from gels (Figure 2.2) and successfully identified by nLC-MS/MS (Table 2.1). In most of the cases more than one protein was identified by band. The identified proteins were classified according to the Gene Ontology (GO) biological process annotation (Figure 2.3). In the hydrophilic fractions the differentially accumulated proteins were related with several functions being seed development and germination, carbohydrate metabolism, and response to stress and defence the most abundant. The differentially accumulated protein bands in the hydrophobic fraction were represented by proteins related with seed development and germination, carbohydrate metabolism, biosynthesis of amino acids, steroids, and auxin homeostasis.

With the information of protein content in seeds and the differentially accumulated bands intensity, PCA and AHC analyses were carried out. PCA maps showed that two principal components accounted for 63.34% of variation (Figure 2.4A). These two main components grouped the wild species in the same quadrant, *A. cruentus* was located alone in one quadrant near to *A. hypochondriacus* (Opaca and Cristalina) and the most cultivated species *A. hypochondriacus* cv Nutrisol was the most distant from the rest of the species. The AHC dendrogram clearly indicates that *A. powellii* and *A. cruentus* have a close relationship as well as *A. hybridus* and *A. hypochondriacus* cv Cristalina (Figure 2.4B).

2.3.3. LEA proteins are species-specific

Different paralogs of late embryogenesis abundant proteins (LEAs) were identified (Tables 2.1 and 2.2). In band 3, which was down accumulated in *A. hypochondriacus* cv Cristalina, was detected one LEA (013747); in band 4 (up accumulated in *A. powellii* and *A. cruentus*) was detected the Embryonic DC-8 like (000638), and in band 6, which was observed accumulated in *A. hybridus* and diminished in *A. powellii*, the LEA (001171) was detected. Two LEA proteins (006906 and 016810)

Table 2.1. Amaranth proteins identified in differentially accumulated bands.

Band No.^a	Protein	Accession No.^b	Mr^c Exp.	Mr^d Theor.	Mascot Score^e	PM/SC^f (%)	emPAI^g
1	Elongation factor 2	001926	97.28	94.87	116	8/11	0.30
	Ribonuclease TUDOR 1	004841		109.14	58	18/20	0.38
	Low-temperature-induced-like	018897		87.08	51	13/20	0.22
2	Phosphoenolpyruvate carboxylase 3	004468		108.39	46	5/6	0.08
	Alpha-xylosidase 1	020003	87.98	93.92	233	9/11	0.30
	Poly [ADP-ribose] polymerase 3	003773		82.76	166	12/23	0.29
	Starch branching enzyme I	000673		100.56	160	16/20	0.41
	Elongation factor 2	001926		94.87	132	10/17	0.34
	Chaperone 1	008070		100.77	81	9/12	0.19
	Aminopeptidase M1	006828		95.06	58	5/7	0.12
3	Methionine synthase	017360	78.85	78.49	529	18/30	0.39
	Late embryogenesis abundant protein	013747		72.37	304	13/23	0.47
	Methionine synthase	022179		89.82	219	9/16	0.17
	Sucrose synthase	021141		69.92	124	8/17	0.28
	Alpha-xylosidase 1	010666		100.26	111	7/9	0.19
	Disulfide isomerase	015532		56.01	51	5/14	0.13
4	Vicilin-like	018839	72.74	61.21	191	8/18	0.25
	Embryonic protein DC-8-like	000638		65.20	61	6/10	0.23
5	11S globulin	021282	57.85	78.10	186	16/32	0.31
	Vicilin-like	018839		61.21	163	8/18	0.50
	Catalase	007232		52.51	112	6/12	0.31
	Glucose-6-phosphate isomerase	013135		57.20	80	5/12	0.28
6	Vicilin-like	018839	52.81	61.21	409	14/31	1.07
	11S globulin	021282		78.10	356	18/34	0.69
	Late embryogenesis abundant protein	001171		45.86	265	8/24	0.95

	UTP-glucose-1-phosphate uridylyltransferase	008585		48.70	176	9/25	0.63
	Enolase 1	001183		45.11	156	9/29	1.13
	ATP synthase subunit mitochondrial-like	001716		59.31	145	7/15	0.50
	Adenosylhomocysteinase 1	009349		53.89	119	8/16	0.56
	Leucine aminopeptidase 1-like	014952		63.04	43	5/10	0.18
7	Vicilin-like	018839	47.21	61.21	96	3/5	0.37
	Eukaryotic initiation factor 4A-9	003448		47.07	45	3/9	0.15
8	Phosphoglycerate kinase	006883	43.28	87.94	573	19/27	0.51
	Phosphoglycerate kinase	019107		42.55	379	14/38	1.13
	11S globulin	021282		78.10	140	14/25	0.15
	Vicilin-like	018839		61.21	133	6/13	0.19
	Actin-7	019031		41.93	104	7/19	0.40
9	Glyceraldehyde-3-phosphate dehydrogenase	011043	38.41	31.58	410	9/37	0.44
	Vicilin-like	018839		61.21	403	10/22	0.46
	Aldose 1-epimerase-like	015176		32.27	132	6/24	0.27
	Dehydrin Rab18-like	003168		26.47	58	5/27	0.33
	11S globulin	021282		78.10	37	5/10	0.10
10	Vicilin-like	018839	37.42	61.21	437	13/25	0.37
	Lactoylglutathione lyase	011906		30.92	143	4/14	0.28
	Malate dehydrogenase	021284		36.13	124	6/29	0.19
11	Vicilin-like	006304	31.21	62.07	534	15/34	0.47
	Late embryogenesis abundant protein (SMP)	006906		28.70	190	6/33	0.60
	Late embryogenesis abundant protein (SMP)	016810		22.64	167	5/34	1.10
	11S globulin	001411		55.75	130	4/10	0.13
	11S globulin	021282		78.10	108	7/16	0.09
	60S ribosomal protein L6-3	005418		25.60	94	5/29	0.70
	Agglutinin	007409		30.40	80	4/18	0.56
12	Oil body-associated protein 1A	009953	29.19	26.80	255	11/35	1.22

	Vicilin-like	006304		62.07	186	3/9	0.12
	60S ribosomal protein L7-4	008528		28.35	180	5/18	0.65
	Elongation factor 1-beta 1	002577		24.87	170	3/20	0.53
	Oil body-associated protein 2A	004342		25.76	98	7/29	0.99
	Protein synthesis inhibitor PD-S2-like	011528		30.52	94	4/19	0.42
13	Protein synthesis inhibitor PD-S2-like	011528	27.81	30.52	362	13/42	1.39
	Oil body-associated protein 2A	004342		25.76	193	6/25	0.80
14	Cysteine proteinase inhibitor 6	021786	26.87	27.78	270	11/52	1.44
	Vicilin-like	018839		61.21	76	5/11	0.26
	11S globulin	021282		78.10	46	5/10	0.15
15	11S globulin	001411	22.98	55.75	377	8/19	0.47
16	11S globulin	021282	21.73	78.10	290	10/16	0.51
	11S globulin	001411		55.75	242	5/12	0.38
17	17.6 kDa class I heat shock protein 3	013876	17.93	17.92	182	7/46	1.18
	Oleosin 5	013707		20.73	72	5/28	0.96
	Oleosin 5	015343		20.46	69	4/21	0.67
18	Cyclophilin	002428	17.43	17.21	85	2/13	0.26
	17.4 kDa class I heat shock protein 3	012223		17.36	34	6/33	0.25
19	Late embryogenesis abundant protein	008005	15.82	9.66	152	5/59	2.50
20	Vicilin-like	018839	14.82	61.21	84	4/8	0.30
23	Histone H4	005348	11.74	11.40	62	5/41	1.51
24	Late embryogenesis abundant protein	019862	10.73	8.53	49	3/49	0.47
27	GBSSI, chloroplastic amyloplastic	011500	63.25	63.00	1112	27/58	3.08
	11S globulin	021282		78.10	287	14/26	0.42
	Vicilin-like	018839		61.21	231	10/23	0.48
	Indole-3-aceticacid-amido synthetase	011444		70.17	65	4/6	0.21
28	GBSSI, chloroplastic amyloplastic	011500	56.53	63.00	837	19/43	2.36
	11S globulin	021282		78.10	475	17/38	1.04
	Vicilin-like	018839		61.21	38	3/6	0.12

29	11S globulin	021282	52.41	78.10	412	20/43	1.22
	GBSSI, chloroplastic amyloplastic	011500		63.00	122	8/13	0.47
	Vicilin-like	018839		61.21	119	10/23	0.66
	ATP synthase subunit mitochondrial-like	001716		59.31	116	9/21	0.59
	Late embryogenesis abundant protein	001171		45.86	114	5/13	0.35
	Elongation factor 1-alpha 1	001308		50.96	66	7/17	0.50
	Serine hydroxymethyltransferase 4	009350		59.74	66	4/12	0.12
30	Vicilin-like	018839	37.95	61.21	333	9/23	0.50
	11S globulin	021282		78.10	210	13/23	0.50
	Aldose 1-epimerase-like	015176		32.27	140	5/21	0.38
	11-beta-hydroxysteroid dehydrogenase 1B	004692		74.56	109	9/18	0.21
	Glyceraldehyde-3-phosphate dehydrogenase	013553		31.71	83	6/28	0.73
31	Vicilin-like	006304	19.47	62.10	90	7/17	0.20
	Oil body-associated protein 1A	009953		26.79	89	7/35	1.60
32	Vicilin-like	006304	27.11	62.10	90	3/5	0.06
	Protein synthesis inhibitor PD-S2-like	011528		30.52	82	8/30	0.81
	Oil body-associated protein 2A	004342		25.76	45	4/20	0.32
33	Vicilin-like	018839	14.75	61.21	137	4/9	0.27
	Vicilin-like	006202		59.16	109	5/7	0.28
	Nucleoside diphosphate-kinase1	014404		16.19	45	3/17	0.22
	11S globulin	001411		55.75	42	3/6	0.55
34	11S globulin	001411	10.72	55.75	109	6/15	0.22
	Vicilin-like	006202		59.16	54	5/7	0.28

^aBand numbers according Figure 2.2. ^bAccession number according to the database reported by Clouse et al. 2016. ^cExperimental molecular weight (kDa). ^dTheoretical molecular weight (kDa). ^eMASCOT Score, individual ion scores > 33 were statistically significant ($p < 0.01$), only identifications with peptide matches above identity threshold when FDR ≤ 5% were considered true. ^fPeptides Matched/Sequence Coverage. ^gExponentially Modified Protein Abundance Index. Protein names in bold letters are discussed in the text.

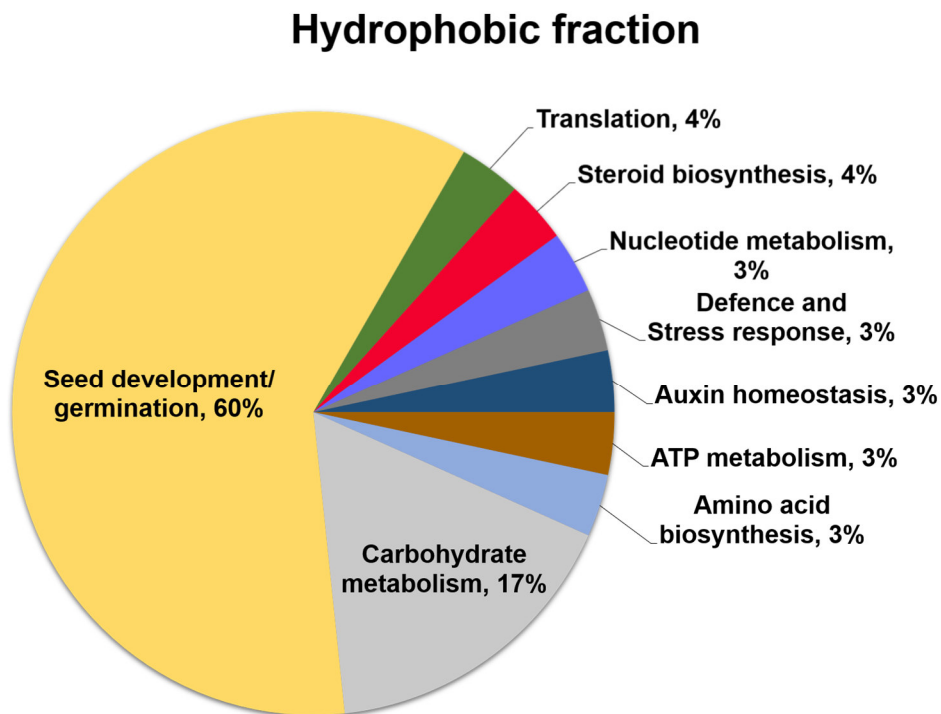
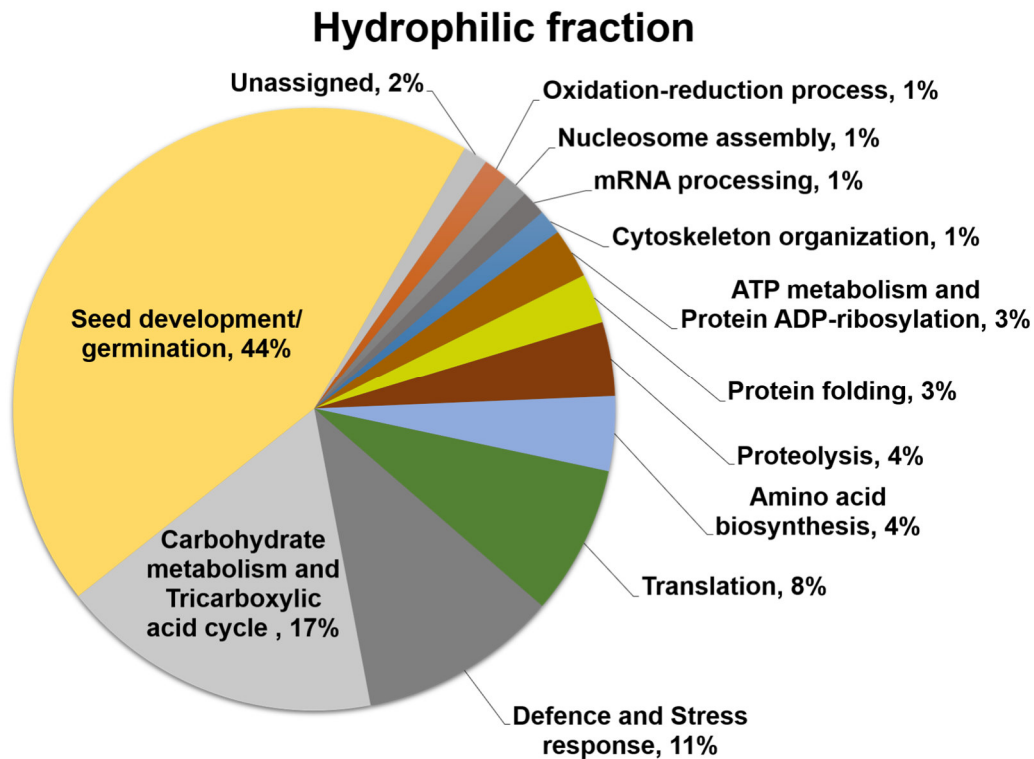


Figure 2.3. Classification of the proteins identified by nLC-MS/MS. The pie charts show the distribution into their biological process in percentage according to Gene Ontology Classification.

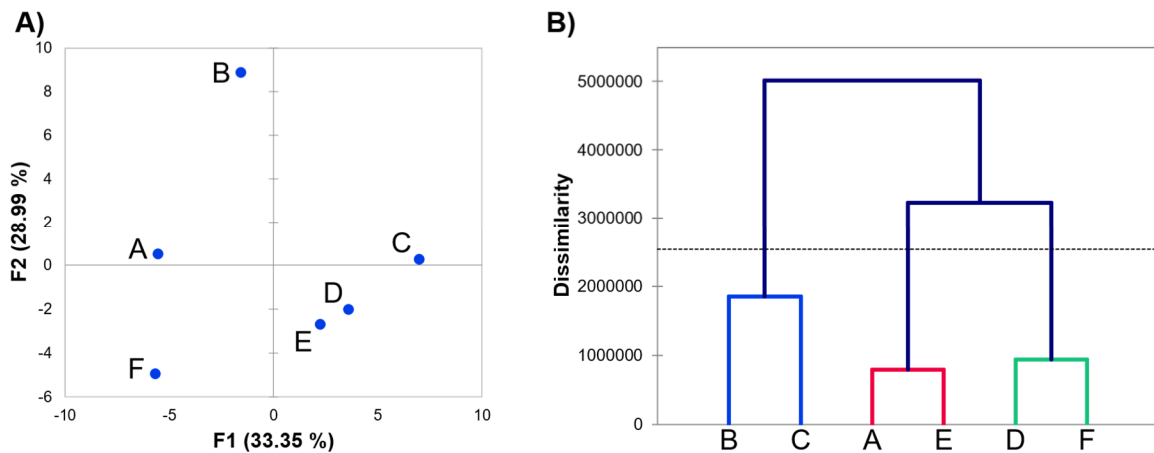


Figure 2.4. Principal Components Analysis (PCA) and Agglomerative Hierarchical Clustering (AHC). **A**, PCA score plot for the data set. The first two components account for 62.34% of the total variation. Each axis is labelled with the percent of total variance and the absolute eigenvalue. **B**, AHC dendrogram grouped amaranth species according to their similarity on protein profiles. Letters correspond to amaranth species: **A**, *A. hybridus*; **B**, *A. powellii*; **C**, *A. cruentus* cv Amaranteca; **D**, *A. hypochondriacus* cv Opaca (waxy); **E**, *A. hypochondriacus* cv Cristalina (non-waxy); **F**, *A. hypochondriacus* cv Nutrisol.

Table 2.2. Late embryogenesis abundant proteins reported in the amaranth genome database.^a

Protein name	<i>A. hypochondriacus</i> proteome accession number	Closer orthologue species/accession number	Pfam domain
Embryonic DC-8	AHYPO_000638-RA	<i>B. vulgaris</i>/XP_010683930.1	LEA 4
ECP63-like -X1	AHYPO_001171-RA	<i>C. quinoa</i>/XP_021737795.1	LEA 4
D-29-like	AHYPO_004157-RA	<i>C. quinoa</i> /XP_021740545.1	Neuromodulin N
DC-8 isoform X2	AHYPO_011345-RA	<i>B. vulgaris</i> /XP_010686551.1	Neuromodulin N
LEA protein	AHYPO_011345-RA	<i>C. quinoa</i> /XP_021714528.1	-
LEA hydroxyproline-rich glycoprotein	AHYPO_002268-RA	<i>C. quinoa</i> /XP_021768891.1	-
LEA hydroxyproline-rich glycoprotein	AHYPO_002278-RA	<i>C. quinoa</i> /XP_021774037.1	-
LEA protein, 3-like	AHYPO_002961-RA	<i>C. quinoa</i> /XP_021760058.1	-
LEA protein 2-like	AHYPO_002962-RA	<i>Q. suber</i> /XP_023883403.1	-
LEA hydroxyproline-rich glycoprotein	AHYPO_003750-RA	<i>S. oleracea</i> /XP_021856806.1	LEA 2
LEA protein	AHYPO_004102-RA	<i>B. vulgaris</i> /XP_010690833.1	-
LEA protein	AHYPO_005092-RA	<i>A. cruentus</i> /AQQ72603.1	-
LEA hydroxyproline-rich glycoprotein	AHYPO_005259-RA	<i>C. quinoa</i> /XP_021760362.1	LEA 2
LEA protein D-34	AHYPO_006906-RA	<i>B. vulgaris</i>/XP_010679058.1	SMP
LEA protein 31	AHYPO_006907-RA	<i>B. vulgaris</i> /XP_010679062.1	SMP
LEA protein 31	AHYPO_006909-RA	<i>C. quinoa</i> /XP_021762725.1	SMP
LEA D-34-like	AHYPO_006910-RA	<i>C. quinoa</i> /XP_021764266.1	SMP
LEA Lea5-like	AHYPO_007836-RA	<i>C. quinoa</i> /XP_021763205.1	LEA 3
Stress induced	AHYPO_008005-RA	<i>B. vulgaris</i>/XP_010676772.1	LEA 5
LEA hydroxyproline- rich glycoprotein	AHYPO_009141-RA	<i>C. quinoa</i> /XP_021841865.1	-
hydroxyproline-rich glycoprotein	AHYPO_009731-RA	<i>S. oleracea</i> /XP_021853522.1	LEA 2
LEA hydroxyproline-rich glycoprotein	AHYPO_010288-RA	<i>B. vulgaris</i> /XP_010682143.1	LEA 2
LEA D-29-like	AHYPO_010481-RA	<i>C. quinoa</i> /XP_021752787.1	-
LEA protein	AHYPO_011548-RA	<i>C. quinoa</i> /XP_021767182.1	DUF4149
LEA protein 47-like	AHYPO_011838-RA	<i>C. quinoa</i> /XP_021771823.1	SMP
LEA hydroxyproline-rich glycoprotein	AHYPO_012283-RA	<i>B. vulgaris</i> /XP_010689396.1	LEA 2
LEA protein group 6	AHYPO_013245-RA	<i>B. vulgaris</i> /XP_010679579.1	LEA 6
LEA hydroxyproline-rich glycoprotein	AHYPO_013450-RA	<i>C. quinoa</i> /XP_021716626.1	LEA 2, Why
LEA protein, 2	AHYPO_013934-RA	<i>C. quinoa</i> /XP_021774726.1	LEA 2, Why
LEA 47-like	AHYPO_014549-RA	<i>S. oleracea</i> /XP_021843088.1	SMP
D-34-like	AHYPO_014550-RA	<i>S. oleracea</i> /XP_021841918.1	-
LEA hydroxyproline-rich glycoprotein	AHYPO_016193-RA	<i>C. quinoa</i> /XP_021733033.1	-
D-34-like	AHYPO_016810-RA	<i>S. oleracea</i>/XP_021853558.1	SMP
LEA protein	AHYPO_019517-RA	<i>B. vulgaris</i> /XP_010674693.1	YtxH
P8B6	AHYPO_019862-RA	<i>C. quinoa</i>/XP_021768105.1	LEA 5
LEA, group 3	AHYPO_020199-RA	<i>C. quinoa</i> /XP_021763489.1	-
LEA protein 2-like	AHYPO_020201-RA	<i>Q. suber</i> /XP_023883403.1	-
LEA hydroxyproline-rich glycoprotein	AHYPO_021817-RA	<i>C. quinoa</i> /XP_021723671.1	LEA 2
LEA	AHYPO_013747-RA	<i>C. quinoa</i>/XP_021717409.1	-

^aProteins that were identified by nLC-MS/MS in differentially accumulated protein bands are in bold red. SMP, Seed maturation protein; Why, Water Stress and Hypersensitive response.

containing the Seed Maturation Protein (SMP) motif were identified in band 11, whose accumulation decreased in wild species. In bands 19 and 24, from *A. cruentus* and *A. powellii*, was identified only one protein corresponding to LEA 008005 and 019862, respectively. These two proteins showed the LEA_5 domain, which is one of the most hydrophilic LEAs (Hundertmark & Hinch, 2008). Interestingly the previously characterized AcLEA protein (005092), was not detected in any differentially accumulated protein band, which agrees with the observation that this LEA is very conserved amongst wild and cultivated amaranth species (Saucedo et al., 2017).

2.3.4. Differential accumulation of GBSSI and oil bodies related proteins amongst species

The most striking differences in protein profiles amongst amaranth species were detected in the hydrophobic fraction, especially in bands 27, 28, and 29 (Figure 2.2B, Table 2.1). In those bands, different proteoforms of the granule-bound starch synthase I (GBSSI, 011500) were identified. The accumulation of band 27 only in wild species (*A. hybridus* and *A. powellii*) as well as in *A. hypochondriacus* cv Cristalina, correlates with the observation that these species are classified as non-waxy type (Figure 1.2). However, band 28 is representative of *A. powellii* and *A. cruentus* cv Amaranteca, which are non-waxy and waxy phenotypes, respectively. By contrary, band 29 was detected in *A. hybridus* as well as in all *A. hypochondriacus* cultivars. As observed, only the GBSSI of higher molecular weight (band 27) correlates with the non-waxy phenotype (Figures 1.2 and 1.3), thus this protein could be the functional waxy enzyme.

In band 17, up accumulated in *A. cruentus*, were identified two paralogs of oleosin 5 (013707 and 015343). Accumulation of band 12 was observed in *A. hybridus* and *A. powellii*, in this band was identified two paralogs of oil body-associated proteins (OBAPs), OBAP1 (009953) and OBAP2 (004342); while in protein band 13 more accumulated in *A. cruentus* was detected another OBAP2. A vicilin isoform was also identified in band 12, which agrees with Zhao et al. (2016),

who reported that during oil body extraction in soybean, glycinin and β -conglycinin are co-purified.

2.3.5. Identification of new paralogs of amaranth globulins

Different paralogs of 7S and 11S globulins were detected in different protein bands (Tables 2.1 and 2.3). The canonical 7SB (006304) containing the β -barrel or cupin structural domain, which function as nutrient reservoir, was detected down accumulated in wild species (band 11) as well as in *A. hypochondriacus* cv. Nutrisol (band 31). The vicilin, containing antimicrobial peptide domain (006202), was accumulated in *A. hybridus* (band 33) and *A. powellii* and *A. cruentus* (band 34). The 7SD globulin (18839) containing both cupin and vicilin domains, was identified preferentially accumulated in *A. powellii* and *A. cruentus* (bands 4 to 12, and 14) as well as in *A. hypochondriacus* cv Cristalina and Nutrisol (bands 20, 30, and 33). The presence of this protein in different molecular weights could be explained by post-translational proteolytic processing during the deposition and storage process (Shewry, Napier, & Tatham, 1995).

The 11S globulin Ah11SB (001411) accumulated less in *A. hybridus* than in *A. powellii* (band 34) but more in *A. hypochondriacus* cv Cristalina (band 15). The legumin (021282), named as Ah11SHMW due to its unusual high molecular weight, was found more accumulated in *A. hybridus* and *A. hypochondriacus* (band 29). A fourth 11S globulin, named Ah11SPheRich (006768), was found by searching in the proteome database, but it was not differentially accumulated amongst amaranth species.

The phylogenetic tree constructed with 7S and 11S globulins from amaranth and members from other Caryophyllales belonging to the cupin superfamily, which is characterized by the presence of β -barrel structural domains (Dunwell, Khuri, & Gane, 2000), revealed that Ah11SA and Ah11SB are very close, however Ah11SHMW and AhPheRich are more similar to *Beta vulgaris* orthologs and it is very clear that 7S globulins formed another branch on the tree (Figure 2.5).

Table 2.3. Classification of amaranth 7S globulins (vicilin) according to the presence of specific structural domains. Proteins that were identified by nLC-MS/MS in differentially accumulated bands are in bold red.

Protein\Domain	Cupin	Xylanase Inhibitor	Vicilin
006202			X
010140(7SA)	X		
006304(7SB)	X		
007944(7SC)	X		X
018839(7SD)	X		X
003828		X	
005737		X	
007735		X	
011849		X	
011850		X	
011853		X	
011854		X	
016318		X	

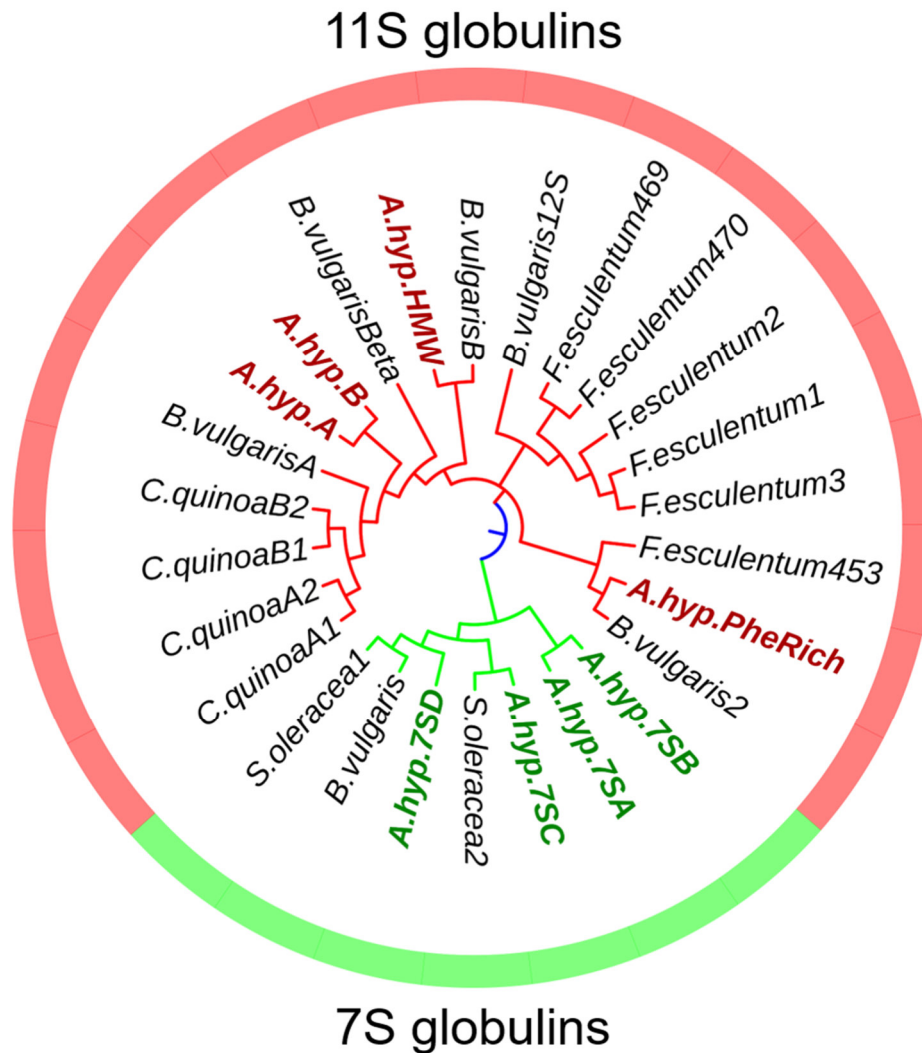


Figure 2.5. Phylogenetic relationships of seed storage proteins belonging to the cupin superfamily of the order Caryophyllales. Phylogenetic tree was constructed with the neighbour-joining method and a bootstrap test for 1000 replicates. Sequences names and NCBI or Phytozome identification numbers: *Beta vulgaris* (XP_010679084.1); *Spinacia oleracea* 1 (XP_021843200.1); *S. oleracea* 2 (XP_021861035.1); *A. hypochondriacus* A (3QAC); *A. hyp* B (AHYPO_001411-RA); *A. hyp* PheRich (AHYPO_006768-RA); *A. hyp* HMW (AHYPO_021282-RA); *Chenopodium quinoa* A1 (AAS67036.1); *C. quinoa* A2 (ABI94735.1); *C. quinoa* B1 (AAS67037.1); *C. quinoa* B2 (XP_021770181.1); *B. vulgaris* Beta (XP_021770181.1); *B. vulgaris* 2 (XP_010679299.1); *B. vulgaris* A (XP_010679302.1); *B. vulgaris* B (XP_010671027.1); *B. vulgaris* 12S (XP_010671026.1); *Fagopyrum esculentum* 1 (O23878.1); *F. esculentum* 2 (O23880.1); *F. esculentum* 3 (Q9XFM4.1); *F. esculentum* 453 (AAP15457.1); *F. esculentum* 470 (BAO50869.1); *A. hyp* 7SA (AHYPO_010140-RA); *A. hyp* 7SB (AHYPO_006304-RA); *A. hyp* 7SC (AHYPO_007944-RA); *A. hyp* 7SD (AHYPO_018839-RA).

2.3.6. *In silico* molecular characterization of amaranth 11S globulins paralogs

Clustal analysis for amaranth 11S globulins compared against the canonical and well-known soybean 11S globulins was carried out (Figure 2.6). All globulins present highly conserved structural features, as the proteolytic site Asn-Gly that is cleaved by a specific asparaginil endopeptidase generating the acidic and basic subunits linked by a disulphide bond, each one containing a cupin β -barrel domain (Figure 2.7). However, some differences in structure were observed when compared with the canonical Ah11SA (Figure 2.8). Ah11SB has a larger acidic chain and a short basic chain. Globulin denominated as Ah11SPheRich because at primary structure level shows high percentage of Phe (17.1%) in comparison with the other globulins (2.8 to 5.2%) (Figure 2.9). The Ah11SHMW is a globulin paralog of high molecular weight showing the largest acidic chain (Figure 2.8). The analysis of Ah11SHMW primary structure exhibit a segment of 18 amino acid residues: G-S-E(Q)-W(R)-D(E)-P-R(S)-Y-P-G-H-G(E)-S-Q(E)-R-P-A(G/T)-H that is repeated 9 times within the acidic subunit (Figure 2.10). This segment was identified in SMART and Pfam servers as CTD domain, which is known to be involved in the regulation of transcript elongation process and mRNA processing, but until now, there are no reports about an 11S globulin containing this domain neither about its biological function.

In amaranth only the canonical 11S globulin, one of the most abundant proteins in the hydrophobic fraction, has been characterized at structural level by X-ray crystallography and named Ah11SA with PDB identifier 3QAC (Tandang-Silvas et al., 2012) (Figure 2.11). Three-dimensional structures of all amaranth 11S globulin paralogs were generated by homology modelling and compared with Ah11SA. The models presented the β -barrel and α -helix distinctive domains of legumin monomers. When compared with Ah11SA, the RMSD values for Ah11SB, Ah11SHMW and Ah11SPheRich were of 0.382, 0.777, and 0.820, respectively, indicating that these proteins are structural homologs. Yellow circles in models represent the intra- (IA) and inter- (IE) chain disulphide bonds. The orange non-structured region in Ah11SHMW represented the highly exposed CTD-like domain. The hydrophobicity and coulombic surfaces of both faces (IA and IE) of amaranth globulins structures

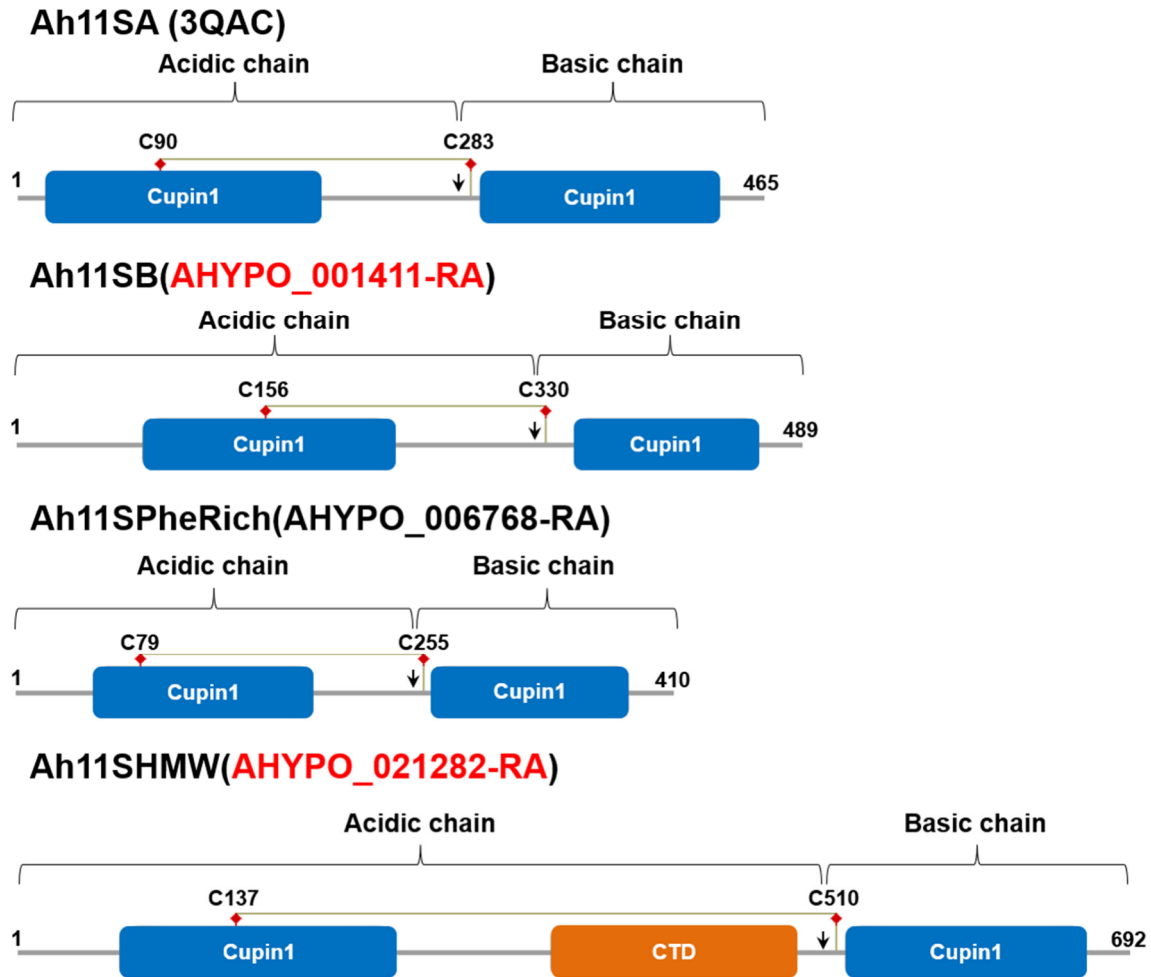


Figure 2.8. Conserved domains in amaranth 11S globulins. All monomers have two cupin domains. Cysteine residues involved in the formation of the disulfide bond between the acidic and the basic subunits are indicated. The arrow in each diagram indicates the proteolytic processing Asn-Gly site to which 11S globulins are subjected during its synthesis and deposition, giving rise to the subunits.

	Ala	Cys	Asp	Glu	Phe	Gly	His	Ile	Lys	Leu	Total
Ah11SA	6.67	1.08	3.87	9.03	5.16	8.60	2.37	6.24	2.58	6.67	465.00
Ah11SB	6.34	1.43	3.48	8.59	6.75	8.59	2.66	6.54	2.86	5.93	489.00
Ah11SPheRich	4.15	2.44	3.41	7.07	17.07	7.80	3.17	5.61	1.71	6.34	410.00
Ah11SHMW	4.91	1.30	4.34	7.66	3.18	10.55	3.61	5.06	2.75	6.79	692.00
GmA1aB1b	5.67	1.68	3.57	8.61	4.20	7.35	1.68	5.46	5.04	6.93	476.00
GmA1bB2	6.22	2.07	3.11	7.88	5.60	7.05	1.24	4.98	3.73	7.88	482.00
GmA2B1a	6.80	2.06	3.71	7.63	4.54	7.22	0.82	4.74	3.92	7.84	485.00
GmA3B4	3.65	1.22	4.67	8.72	3.25	7.91	3.25	3.65	3.65	6.90	493.00
GmA5A4B3	4.26	1.42	5.68	9.77	2.84	6.75	2.66	3.73	4.80	7.99	563.00

	Met	Asn	Pro	Gln	Arg	Ser	Thr	Val	Trp	Tyr	Total
Ah11S	1.94	5.38	4.30	6.88	9.46	6.88	4.09	5.59	0.86	2.37	465.00
Ah11SB	2.25	5.11	4.50	6.54	8.18	7.16	4.50	5.32	0.82	2.45	489.00
Ah11SPheRich	1.95	3.90	3.41	4.15	6.34	9.27	2.68	5.37	0.73	3.41	410.00
Ah11SHMW	0.87	4.19	8.24	7.08	8.96	8.09	3.18	4.48	1.73	3.03	692.00
GmA1aB1b	1.26	7.77	6.09	10.08	5.67	6.72	4.20	4.83	0.84	2.31	476.00
GmA1bB2	1.45	7.47	5.60	9.75	6.43	7.05	3.94	5.60	0.62	2.28	482.00
GmA2B1a	1.65	8.25	5.36	10.52	5.98	6.60	3.71	5.57	0.82	2.27	485.00
GmA3B4	0.61	6.69	7.30	9.53	6.29	7.71	3.85	7.10	0.81	3.25	493.00
GmA5A4B3	0.53	5.68	6.75	9.06	6.39	8.17	3.55	6.22	1.07	2.66	563.00

Figure 2.9 Amino acid composition of 11S globulins from amaranth (*A. hypochondriacus*, Ah) and soybean (*Glycine max*, Gm). Red square indicates the percentage of phenylalanine.


```

>Ah11SHMW
1 MARHRSARVLVPLALTLVLILSPTS LAQQWGSFNPPFLPAGQSSQSQTRLTRDTQCRIDC 60
61 QIDQLSANEPNIRIQAEAGVNEIWDPREQKEFQCAGVTVVRTQVEPNGLFLPHYNNAPSI 120
121 SYVIRGKALLGVTNPGCPETFEYGSSEPFSSERDLRRPGHKFERPGREFESIRDQHQKIR 180
181 RVYQGHIVALPAGVSKWIFYNDGQDRLLTIVTLFDLTLNNQNQLDDILRSFFLAGNPQGREGA 240
241 QGGKGSQRIFSENNILSGFDRQLLSQAFGIEPETVSKIQQQNDDRGAIIRVEGDLGLLIP 300
301 EWDREESRRPSESYRPGQGSEWDPRYPGHGSQRPTHGSEWDPRYPGHESQRPAHGSERDP 360
361 RYPGHGSQRPTHGSEWDPRYPGHESQRPAHGSEWDPRYPGHGSQRPTHGSEWDPRYPGHG 420
421 SORPAHGSEWDPRYPGHGSQRPAHGSEWDPRYPGHGSQRPGHGSQWEPSYPGHGSERP 480
481 QGQERICGGRRICEENGVCKPNGIETLCSVRITENIDDEKADVFNPPQGGRLTSLNSQ 540
541 KLPILNLYQLSAEKVNLYQNAIMAPNWKINAHSIIFYFTKGNRQVQIAGHEGRLVFDDMVQ 600
601 EGQLLVVPQNFVVLKKGQEGLEWVAFLLTSDAMISPLAGRISAIRGLPEQVVMNSYGLS 660
661 REEA KRLKYGRQELTVFSPSEEFQRKGYAIM 692

1 GSEWDPRYPGHGSQRPTH 18
2 GSEWDPRYPGHESQRPAH 18
3 GSERDPRYPGHGSQRPTH 18
4 GSEWDPRYPGHESQRPAH 18
5 GSERDPRYPGHGSQRPTH 18
6 GSEWDPRYPGHGSQRPAH 18
7 GSERDPRYPGHGSQRPAH 18
8 GSEWDPRYPGHGSQRPGH 18
9 GSQWEPSYPGHGSERP 18
** : : * * * * * * : * * *

```

Figure 2.10. Ah11SHMW amino acid sequence. The cupin β -barrel domains of 11S globulins are shown in green. The red and blue bold letters indicate the 9 repeated sequences that form the CTD-like domain and the alignment of this sequences are displayed.

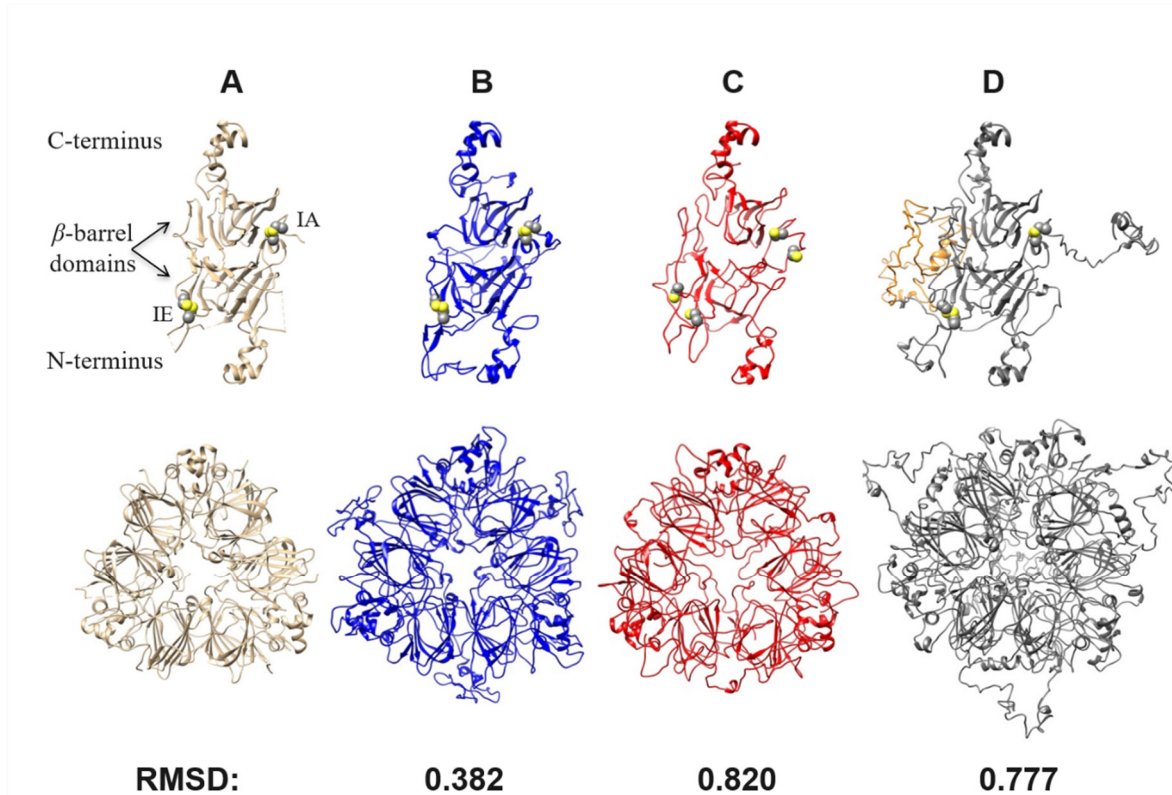


Figure 2.11. **A**, Experimental reported structure for the canonical 11S globulin monomer of *A. hypochondriacus* (Ah11SA, PDB 3QAC) and structural models generated from 11S globulin paralog sequences. **B**, Ah11SB (001411); **C**, Ah11SPheRich (006768); **D**, Ah11SHMW (021282). The low RMSD values indicate that all globulins are structural homologues. All globulins present the two β -barrel domains characteristic of these proteins, the highly conserved cysteines are shown in yellow spheres, which are involved in the formation of intra- (IA) and inter-chain (IE) disulphide bonds. The orange region in the model of Ah11SHMW delimits the CTD-like domain exclusive of this paralog, which is not present in any other 11S globulin reported so far.

are shown in figure 2.12. 11S globulins hydrophobic residues are located mainly on the central part of the IA face (orange region), but the hydrophobicity surface changes amongst the distinct paralogs being the Ah11SPheRich the more hydrophobic, which correlates with its high Phe content.

2.4. Discussion

Amaranth has greatly gained attention due to its agronomical and nutraceutical characteristics. However, only a few species, from various available, are cultivated for seeds production. Amaranth wild relatives have survived for thousand years growing under different environments such as very saline soils, high temperatures, UV radiation, and water deficit (Espitia-Rangel et al., 2010). Accordingly, they are considered important reservoirs of useful genes/proteins involved in plant resistance (Aguilar-Hernández et al., 2011; Huerta-Ocampo et al., 2014). However, information about morphological and molecular characteristics of wild amaranth species has not been reported.

Although *A. powellii* produces the smallest seed, this is the species with the highest protein content, while *A. cruentus*, one of the cultivated species, is the one with the lowest values. Thus, *A. powellii* represents an interesting option as a source of information that could be used to increase protein content in cultivated ones. Similar results have been reported for rice species (*Oryza* spp.) indicating that wild species contained higher protein amounts than the domesticated species, differences that were attributed to the glutelins fraction (Jiang et al., 2014). It is also known that glutelins in amaranth are an important seed storage protein fraction accounting for 23 to 42% of the total seed protein, depending on the extraction conditions (Barba de la Rosa, Gueguen, Paredes-López, & Viroben, 1992). The group of bands between 50 and 70 kDa has previously been detected as differentially accumulated in varieties of cultivated amaranths. Consequently, this protein fraction was suggested as a tool for identification of amaranth accessions (Barba de la Rosa et al., 2009; Džunková, Janovská, Čepková, Prohasková, & Kolář, 2011).

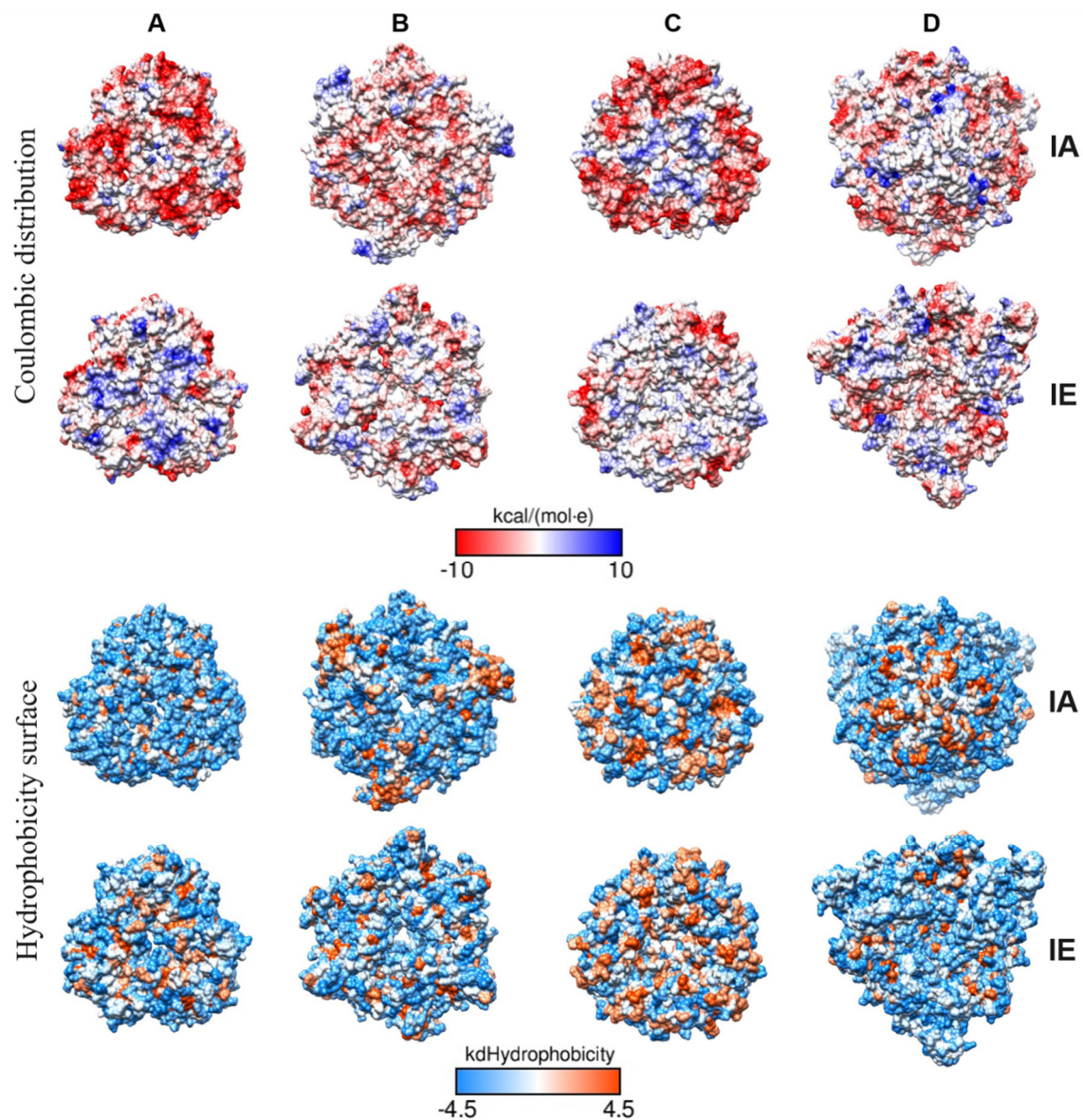


Figure 2.12. Coulombic distribution and hydrophobicity surface of IA and IE faces trimeric structures of 11S globulins paralogs from *A. hypochondriacus*. A, Ah11A; B, Ah11SB (001411); C, Ah11SPheRich (006768); D, Ah11SHMW (021282).

In orthodox seeds, LEA proteins have been associated with desiccation tolerance and maintenance in a quiescent state. LEAs are classified based on amino acid sequence and conserved motifs into five to nine sub-classes (Shih, Hoekstra, & Hsing, 2008). A good correlation between the abundance of certain LEAs and seed longevity has been reported (Rajjou & Debeaujon, 2008; Wozny et al., 2018). By searching in amaranth database, 39 LEA protein sequences with particular motifs were found (Table 2.2), but only some of them were identified differentially accumulated amongst species. The Embryonic DC-8 and LEA_5 group were detected preferentially accumulated in *A. powellii* and *A. cruentus*. DC-8 protein has been detected during embryogenesis and in cell walls of endosperm tissues, however its function is still unclear (Franz, Hatzopoulos, Jones, Krauss, & Sung, 1989; Tnani, López, Jouenne, & Vicient, 2012). LEA_5 and SMPs are proteins related with water stress tolerance (Silva Artur, Zhao, Ligterink, Schranz, & Hilhorst, 2019), SMP was less accumulated in wild species (*A. hybridus* and *A. powellii*) but preferentially accumulated in cultivated species, this is interesting since *A. cruentus* contain both LEA_5 and SMP proteins and is one species that can grow under severe water deficit (Espitia-Rangel et al., 2010; Huerta-Ocampo et al., 2014).

OBAPs (bands 13, 31, 32) as well as two paralogs of oleosins (band 17) were more abundant in *A. cruentus*. It has been reported that OBAPs are involved in oil bodies biogenesis, stability, trafficking, and mobilization (Lopez-Ribera et al., 2014). Oleosins act as natural emulsifiers and protect plant lipid reserves against oxidation and hydrolysis until seed germination and seedling establishment (Frandsen, Mundy, & Tzen, 2001). The putative role of some oleosins is related to controlling lipid body size and maintenance of its integrity (Tzen & Huang, 1992). It has been reported that an *A. thaliana* mutant deficient in OBAP1 shows changes in fatty acid composition, reduction of germination rate, and seed triacylglycerols content (Purkrtova, Jolivet, Miquel, & Chardot, 2008). Therefore, the differential accumulation of OBAP1 and OBAP2 could be related with the quantity and quality fat composition amongst amaranth species. These observations correlated with the relative abundance of fatty acids and hydrocarbons, such as squalene, reported for wild and cultivated amaranth species (Bojórquez-Velázquez et al., 2018).

GBSSI, also known as waxy protein, is a glucosyltransferase and the only enzyme responsible for elongation of amylose polymers in nutrient storage tissues. Park and Nishikawa (2012) analysed the Waxy locus in amaranth showing that a nonsense mutation in the coding region at exon 6 in *A. cruentus* and exon 10 in *A. hypochondriacus* prematurely ends translation and causes complete loss of gene function, leading to a waxy phenotype. Then the GBSSI identified in bands 28 and 29 could correspond to the non-functional truncated enzyme. Ahuja, Jaiswal, Hucl, and Chibbar (2014), reported that during wheat development, GBSSI considerably affects starch accumulation and glucan chain length distribution. It is known that high amylose contents could contribute to resistant starch (RS) through the formation of inclusion complexes with lipids (Raigond, Ezekiel, & Raigond, 2015). Zhou et al. (2016), have proposed a mechanism in which the deficiency in sucrose synthase III (SSIIIa) and the presence of GBSSI could be the responsible for RS accumulation.

SSPs are accumulated during seed development to serve as source of amino acids during germination and early seedling growth and represent the main source of protein for food and feed consumption. Globulins are the most abundant SSPs in dicotyledonous plants and are classified in two groups based on their sedimentation coefficients in 7S or vicilins and 11S or legumins (Shewry et al., 1995). The *A. hypochondriacus* database contains 13 different 7S globulin protein sequences, with members belonging to the three different types, which are classified based on their structural domains (Table 2.3). Only three of them were differentially accumulated amongst amaranth species. The 7S containing the antimicrobial domain was representative in wild species as well as in *A. cruentus*, and the canonical cupin-type was more representative in *A. hypochondriacus* species.

11S globulins or legumins are the more widely distributed SSPs in nature and are encoded by multigenic families. The soybean 11S globulin or glycinin, is composed by five different monomers, each encoded by a different gene (Li & Zhang, 2011). In amaranth only the canonical 11S globulin has been reported and characterized (Barba De La Rosa, Herrera-Estrella, Utsumi, & Paredes-López, 1996). Here we have detected two more paralogs differentially accumulated amongst wild and cultivated amaranths. The CTD-like domain, identified by

database searching in Ah11SHMWglobulin, has some special features: all those repeats have conserved Ser and Tyr that could be involved in signalling process by phosphorylation; His and Arg, positively charged amino acids that affect de solubility and assembly of a protein depending of pH variations; and two Pro, amino acid known as secondary structure breaker. It is possible that this domain suffers some post-translational modifications and has some biological activity in seeds, but further work should be done in this direction.

Recently the importance of SSPs has increased due to the presence of different paralogs and the fact that some of them do not only are nutrient reservoirs but are also involved in other functions during seed development or germination. A novel function for 11S globulins as auxin transporters has been reported, in which during the germination process, the change in pH induces the hexamer dissociation and its release, suggesting globulins as novel players in hormone homeostasis (Kumar et al., 2017). New roles of SSPs have been reported as buffer proteins against oxidative stress that might imply an important functions in seed development and longevity (Mouzo et al., 2018; Nguyen et al., 2015; Sano et al., 2016; Shah et al., 2015).

The surface properties of a protein, mainly hydrophobicity and charge distribution are very important since they dictate the physicochemical functionality of the molecule (Withana-Gamage & Wanasundara, 2012). Three-dimensional structure models of amaranth legumins displays similar features to the canonical 11S globulins, but they show some particular characteristics, variation in the superficial charge and hydrophobic residues distribution for example, which can confer differentiated functional properties to each legumin, like solubility or the ability to form interactions with other molecules. These physicochemical variations between amaranth 11S globulins paralogs are of relevance for two topics, first the application of the proteins as additives for the stabilization of food systems, and second, the implications in biological processes like seed development and germination.

2.5. Conclusions

Seed 1D-SDS-PAGE patterns have been a very powerful tool in detecting differential accumulation of several proteins amongst wild and cultivated amaranth species. It is interesting to highlight that protein accumulation profile indicates that *A. powellii* is more closely related to *A. cruentus*. LEAs could be potential targets for seed resistance and defence traits. OBAPs and oleosins could be targets to increase squalene and/or specific desired lipids content in seeds. Overall our results suggest that there are many new types of globulins paralogs and precursors in wild species, thus, wild amaranth species are very important genetic resources for improving the nutritional quality of amaranth seeds. New paralogs of 11S globulins were detected and structurally characterized *in silico*. Further work is needed to understand the biological functions of the newly identified globulins in amaranth seeds.

CHAPTER 3

Comparative proteomic analysis amongst seeds of wild and cultivated amaranth species by 2-DE-nLC-MS/MS

3.1. Introduction

Solving food demand worldwide has become an increasingly serious problem due to population growth and changes in crop conditions due to climate change, which has led to the search and establishment of strategies for the development of crops with increased characteristics in terms of production yields, their resistance to different types of stress, both biotic and abiotic, and its applicability in biotechnological processes (Komatsu, 2008; McCouch et al., 2013; Pandey, Irulappan, Bagavathiannan, & Senthil-Kumar, 2017). One such approach is given by research focused on the identification of the molecular mechanisms of certain plant species, which allow them to develop under conditions commonly considered unfavourable, for example, extreme temperatures, limited availability of water or nutrients, high salinity or the presence of predators or pathogens, whether viruses, bacteria or fungi (Di Silvestre, Bergamaschi, Bellini, & Mauri, 2018; Hu, Rampitsch, & Bykova, 2015).

Since proteins are the final effectors of most biochemical reactions, and give each organism unique qualities, proteomics has been established as a powerful technique for the identification of molecular targets with possible application in crop improvement (Aslam, Basit, Nisar, Khurshid, & Rasool, 2017; Cho, 2007; Eldakak, Milad, Nawar, & Rohila, 2013; Hood et al., 2012; Komatsu, Mock, Yang, & Svensson, 2013); contrasting the protein profiles of plants susceptible and resistant to certain conditions, comparing cultivars with desired characteristics, or genetically modified crops against their original counterpart, and through the comparison of wild and cultivated or domesticated species (Dong et al., 2017; C. Y. Gong & Wang, 2013; Kosová, Vítámvás, Urban, Prášil, & Renaut, 2018; Pichereaux et al., 2016). One of the most used proteomic approaches for these purposes is that based on two-dimensional electrophoresis (2-DE), since in addition to identifying the variation in the levels of protein accumulation, it allows the visualization of isoforms and post-translational modifications (PTMs) (Oliveira, Coorssen, & Martins-de-Souza, 2014; Rogowska-Wrzesinska, Le Bihan, Thaysen-Andersen, & Roepstorff, 2013). This type of analysis has been carried out in crops such as soybean and jatropha focused

on the identification of proteins involved with lipid content levels (León-Villanueva, Huerta-Ocampo, Barrera-Pacheco, Medina-Godoy, & Barba de la Rosa, 2018; Liu et al., 2013; Min et al., 2015), in barley to determine which are the best performing cultivars in the process of malting for beer production (Herrera-Díaz, Jelezova, Cruz-García, & Dinkova, 2018), in susceptible and resistant plants of cabbage and rice against *Xanthomonas campestris* infection and the attack of small brown grasshoppers, respectively (Dong et al., 2017; Villet et al., 2016), and in countless works with the purpose of elucidating the response to abiotic stress in monocot plants such as barley, maize, rice and wheat, and dicotyledons such as *A. thaliana* and soybean (Gong, Hu, & Wang, 2015).

Amaranth is a crop of great interest because of the good quality of its seeds protein and is rich in other macro- and micronutrients such as carbohydrates, lipids, minerals, and vitamins. The plant has shown resistance to abiotic stress such as salinity drought (Aguilar-Hernández et al., 2011; Huerta-Ocampo et al., 2014). It is estimated that there are about 70 species of amaranth, of which only three are destined in the cultivation for grain production, the rest are wild species distributed in the nature that develop under the exposition of diverse environmental conditions (Assad et al., 2017; Velarde-Salcedo, Bojórquez-Velázquez, & Barba de la Rosa, 2019). To date, descriptive studies of the seed proteome of one of the amaranth species cultivated for grain and its response to abiotic stress have been carried out (Huerta-Ocampo et al., 2014; Klubicová, Szabová, Skultety, Libiaková, & Hricová, 2016; Maldonado-Cervantes et al., 2014), however, there are no studies related to the comparison of protein profiles between wild and cultivated species by 2-DE based proteomics. Therefore, the aim of this work was to carry out a comparative proteomic analysis of the seeds of wild and domesticated species of amaranth through 2-DE, for the identification of proteins that can contribute to the improvement of the nutritional and agronomic qualities of the crop.

3.2. Materials and methods

3.2.1 Biological material

Seeds of two wild amaranth species, *A. hybridus* and *A. powellii*, and three varieties of two grain cultivated amaranth species, *A. cruentus* cv Amaranteca and *A. hypochondriacus* cvs Opaca and Cristalina, were provided by the National Institute for Forestry, Agriculture and Livestock Research (INIFAP), Mexico.

3.2.2 Protein extraction

Amaranth seed proteins were extracted using a polarity-based approach as described before in section 2.2.2. Ten volumes of chilled acetone were added to the hydrophilic and hydrophobic protein extracts, mixed with vortex and incubated at -20 °C by 12 h. Samples were centrifuged at 17,000×g for 30 min at 4 °C, supernatants were discarded, precipitated proteins were washed with 80% acetone and centrifuged once, supernatants were discarded again and proteins air dried. Dried proteins were resuspended with 7 M urea, 2 M thiourea, 4% (w/v) CHAPS and 0.05 M DTT and, before their application to isoelectric focussing, quantified using the Protein Assay reagent (Bio-Rad) with bovine serum albumin as standard.

3.2.3 Two-Dimensional Electrophoresis (2-DE) and image analysis

Hydrophilic and hydrophobic proteins were analysed by 2-DE. For the first dimension, isoelectric focusing (IF) was carried out onto 24 cm IPG linear gradient strips of pH 5-8 (Bio-Rad), rehydrated with 1.5 mg of protein. Focusing was conducted at 20 °C with an Ettan IPGphor system (GE Healthcare) at constant 100 mA per strip under the following conditions: 1, 250 V Step and Hold by 2 h; 2, 500 V gradient until 10 Vh; 3, 2000 V gradient by 2 h; 4, 4000 V gradient by 2 h; 5, 6000 V gradient by 2 h; 6, 8000 Step and Hold until 100000 Vh. After IF, the IPG strips were incubated for 15 min in equilibration buffer (6 M urea, 30% glycerol, 2% SDS, 0.05

M Tris-HCl pH 8.8, 1% DTT) in gentle agitation, and once again with equilibration buffer plus 2.5% iodoacetamide instead of DTT. In the second dimension, focused proteins were resolved in 13% polyacrylamide-SDS gels using the Ettan Daltsix Electrophoresis unit (GE Healthcare) at 10 mA/gel by 27 h. Once the electrophoretic run was finished, gels were stained with colloidal Coomassie G-250 by 12 h and destained with ultrapure water. Images were acquired with the Pharos FX Plus Molecular Imager (Bio-Rad) at 100 μm resolution and analysed with the Melanie software v9.2 (Gene Bio, SIB Swiss Institute of Bioinformatics). The densitometric data were submitted to a one-way ANOVA and spots were considered as differentially accumulated only if they presented both, a fold change ≥ 2.0 and $p \leq 0.001$.

3.2.4 *In-gel* digestion and nLC-MS/MS analysis

Differentially accumulated protein spots were manually excised from the gels and destained, reduced with 10 mM DTT in 25 mM ammonium bicarbonate, followed by protein alkylation with 55 mM iodoacetamide. Protein digestion was carried out overnight at 37 °C with sequencing-grade trypsin (Promega). Nanoscale LC separation of tryptic peptides was performed with a nanoACQUITY UPLC System (Waters) equipped with a Symmetry C18 precolumn (5 μm , 20 mm \times 180 μm , Waters) and a BEH130 C18 (1.7 μm , 100 mm \times 100 μm , Waters) analytical column. The lock mass compound, [Glu1]-Fibrinopeptide B (Sigma-Aldrich), was delivered by the auxiliary pump of the nanoACQUITY UPLC System at 200 nL/min at a concentration of 100 fmol/mL to the reference sprayer of the Nano-Lock-Spray source of the mass spectrometer. Mass spectrometric analysis was carried out in a SYNAPT-HDMS Q-TOF (Waters). The spectrometer was operated in V-mode, and analyses were performed in positive mode ESI. The TOF analyzer was externally calibrated with [Glu1]-Fibrinopeptide B from m/z 50 to 2422. The data were lock-mass corrected postacquisition using the doubly protonated monoisotopic ion of [Glu1]- Fibrinopeptide B. The reference sprayer was sampled every 30 s. The RF applied to the quadrupole was adjusted such that ions from m/z 50–2000 were

efficiently transmitted. MS and MS/MS spectra were acquired alternating between low-energy and elevated-energy mode of acquisition (MS^e).

3.2.5 Protein identification using MS/MS data sets and database searching

MS/MS spectra data sets were used to generate PKL files using Protein Lynx Global Server v2.4 (Waters). Proteins were then identified using PKL files and the MASCOT search engine v2.5 (Matrix Science) against the *A. hypochondriacus* transcriptome and proteome data base v1.0 (23,054 sequences) available at <https://phytozome.jgi.doe.gov/> (Clouse et al., 2016). Trypsin was used as the specific protease, and one missed cleavage was allowed. The mass tolerance for precursor and fragment ions was set to 20 ppm and 0.1 Da, respectively. Carbamidomethyl cysteine was set as fixed modification and oxidation of methionine was specified as variable modification. The protein identification criteria included at least two MS/MS spectra matched at 99% level of confidence, and identifications were considered successful when significant MASCOT individual ion scores > 33 were obtained, indicating identity or extensive homology statistically significant at $p < 0.01$. Identifications were considered true only for peptide matches above identity threshold FDR $\leq 5\%$. To estimate the relative abundance of each protein per spot, the exponentially modified protein abundance index (emPAI) was used (Ishihama et al., 2005). BLAST algorithm was used for homology search against the Viridiplantae and *Arabidopsis thaliana* subsets of the UniProtKB database (<https://www.uniprot.org/blast/>).

3.3 Results

Proteomic maps of amaranth seeds were successfully established. 2-DE patterns generated for the hydrophilic and hydrophobic fractions of each of the analysed amaranth species displays a good spots resolution. In the overall analysis, differential identified proteins are principally represented by a 38% of seed storage proteins (SSPs, 7S and 11S globulins), 10% of granule bound starch synthase I

(GBSSI), 6% of agglutinin, and 4% of 5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase.

3.3.1 Hydrophilic fraction

For the hydrophilic fraction, variations in protein accumulation profiles is accentuated in the low molecular weight region, where nearly to 10 kDa, it could be observed how some spots belongs only to certain species (Figure 3.1). In this fraction, 152 spots were detected as differentially accumulated amongst the amaranth species, from which, 104 were successfully identified by nLC-MS/MS, for a total of 73 unique proteins (Table 3.1). Some proteins were identified in several spots, for example, Vicilin-like seed storage protein (018839) found in 31 spots, 11S globulin (021282) in 17 spots, Agglutinin (007409) in 11 spots, Vicilin-like seed storage protein (006304) in 8 spots, 5-methyltetrahydropteroyltriglutamate-homocysteine methyltransferase (022179, 017357 and 017360) 7 spots, Vicilin-like seed storage protein At2g18540 (010140) 5 spots, Seed biotin-containing protein SBP65 (013747) 3 spots, and 18.3 kDa class I heat shock protein (HSP, 013876) in 3 spots.

Principal component analysis (PCA) was carried out by two approaches using the densitometric data of the 152 differentially accumulated spots in this fraction. In the first one, the images of each proteomic map, that is, each replicate, was taken as a dependent variable (observations), and in the second one, the intensity of each spot was the output variable (Figure 3.2). In the first case it is possible to visualize how the proteomic maps behave according to the global distribution of the spots, and as can be seen in figure 3.2A, the gels are grouped correctly with their respective homologs. It can be observed in the graph how the cultivars of the species *A. hypochondriacus* are grouped very close in the same quadrant, *A. cruentus* is positioned separately from these two cultivars, and the wild species are far from the cultivated ones. Now, when using the general information of the intensity profiles in the proteomic maps as independent variable, we can visualize the behaviour of each spot and determine in which of the species is preferably accumulated, for example, spots 25, 614, 659, and 682 are accumulated more in *A. cruentus* (Figure 3.2B).

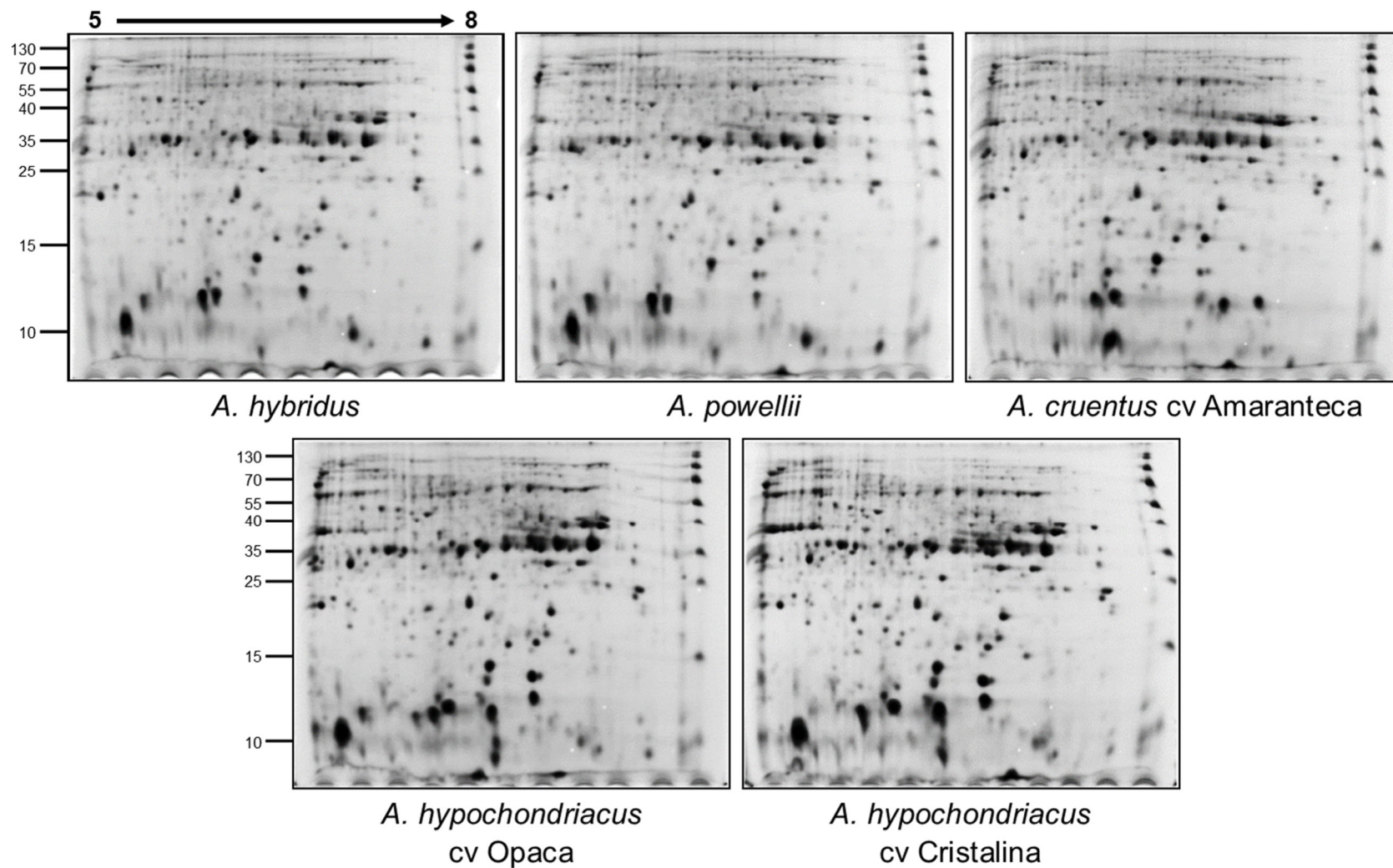
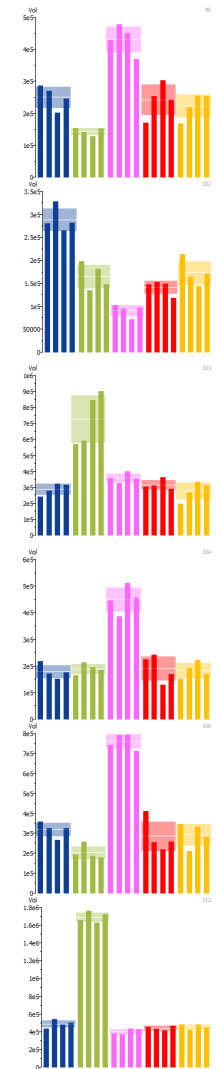


Figure 3.1. Representative proteomic maps of the hydrophilic fraction from amaranth seeds species.

Table 3.1. Hydrophilic amaranth seed proteins identified by nLC-MS/MS in differentially 2-DE accumulated spots.

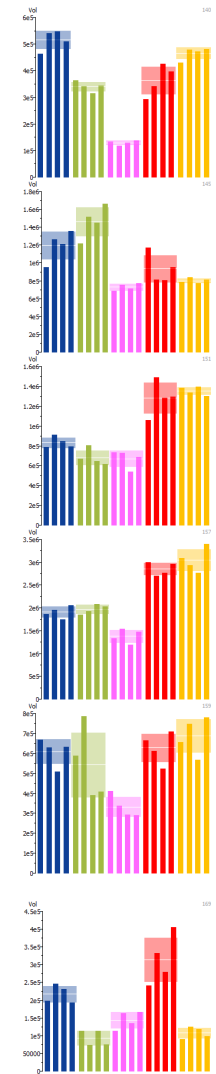
Spot No. ^a	Protein	Accession No. ^b	Ortholog ^c	Mr(kDa)/pI Exp. ^d	Mr(kDa)/pI Theo. ^e	Mascot Score ^f	PM/SC (%) ^g	emPAI ^h	Spot accumulation change ⁱ				
									A	B	C	D	E
28	Histone H4	005348	H4_SOYBN	12.8/5.1	11.4/11.5	176	5/50	3.37					
31	Vicilin-like seed storage protein	018839	AMP22_MACIN	13.4/6.4	60.9/6.6	124	4/8	0.27					
49	Glutathione S-transferase DHAR2	018688	DHAR2_ARATH	23.8/6.3	23.5/6.0	455	8/40	2.56					
53	Glutathione S-transferase	021773	GSTF_SILVU	23.3/7.4	23.9/6.7	527	9/54	2.64					
	Cys peroxiredoxin PER1	014116	REHY_ARATH		21.2/6.3	86	3/22	0.62					
72	18.3 kDa class I heat shock protein	013876	HSP11_OXYRB	24.6/7.6	17.9/5.8	126	2/17	-					

95	Vicilin-like seed storage protein	006304	VCL22_ARATH	25.3/5.2	61.9/5.9	524	6/12	0.46
102	Vicilin-like seed storage protein	010140	VCL21_ARATH	28.7/6.3	67.2/5.4	154	2/2	0.11
103	Late embryogenesis abundant protein 31	006906	LEA31_ARATH	27.6/5.4	28.6/5.0	273	4/18	0.64
	Vicilin-like seed storage protein	006304	VCL22_ARATH		61.9/5.9	188	4/9	0.26
	Probable 6-phosphogluconolactonase 4, chloroplastic	011478	6PGL4_ORYSJ		55.7/5.1	104	3/7	0.21
104	Vicilin-like seed storage protein	006304	VCL22_ARATH	28.1/7.4	61.9/5.9	271	3/8	0.19
106	Oil body-associated protein 2A	004342	OBP2A_ARATH	28.8/6.9	25.5/7.1	439	7/27	1.75
112	Oil body-associated protein 1A	009953	OBP1A_ARATH	29.5/6.7	26.6/6.2	477	6/30	1.89
	Proteasome subunit alpha type-6	008388	PSA6_TOBAC		27.3/6.1	79	3/14	0.47



125	Vicilin-like seed storage protein At2g18540	010140	VCL21_ARATH	29.8/6.0	67.2/5.4	246	4/3	0.26	
	Hydroxyacylglutathione hydrolase cytoplasmic	015200	GLO2C_ARATH		28.5/5.6	195	2/8	0.31	
126	Vicilin-like seed storage protein	006304	VCL22_ARATH	30.2/7.8	61.9/5.9	1177	6/13	0.42	
	Agglutinin	007409	Q38719_AMAHP		30.1/6.5	381	3/14	0.42	
130	Late embryogenesis abundant protein 31	006906	LEA31_ARATH	30.0/5.1	28.6/5.0	782	9/42	2.06	
	Vicilin-like seed storage protein	006304	VCL22_ARATH		61.9/5.9	229	4/7	0.26	
	40S ribosomal protein S3-1	002178	RS31_ARATH		26.8/9.6	128	2/12	0.30	
	Transcription factor TGA7	007998	TGA7_ARATH		25.1/4.8	70	3/18	0.53	
133	Vicilin-like seed storage protein	006304	VCL22_ARATH	31.0/7.4	61.9/5.9	254	5/13	0.34	
	Thiamine thiazole synthase 2, chloroplastic	004627	THI42_VITVI	32.0/5.5	36.7/5.1	324	5/21	0.79	
	11S globulin	021282	13SB_FAGES		77.6/7.0	202	5/8	0.26	
137	Vicilin-like seed storage protein	006304	VCL22_ARATH	32.1/5.8	61.9/5.9	710	9/19	0.67	

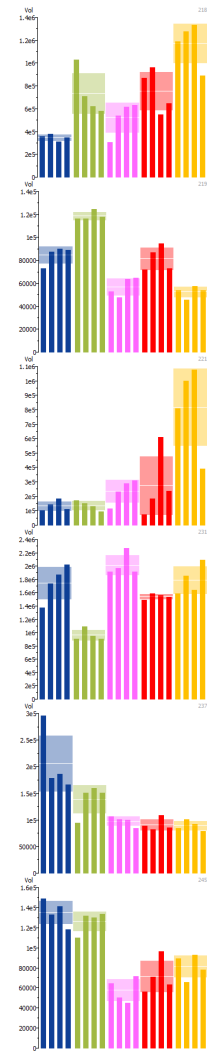
140	11S globulin	021282	13SB_FAGES	32.1/5.7	77.6/7.0	239	9/15	0.56
145	Glucose and ribitol dehydrogenase	010964	GRDH_DAUCA	33.5/7.0	31.5/6.5	819	11/33	2.29
151	11S globulin	001411	CRU1_RAPSA	34.1/6.0	55.4/6.3	94	3/5	0.23
157	11S globulin	001411	CRU1_RAPSA	35.2/6.2	55.4/6.3	890	6/15	0.47
	Agglutinin	007409	Q38719_AMAHP		30.1/6.5	80	2/9	0.26
	NADPH-dependent aldehyde reductase 1, chloroplastic	010965	ADRC1_ARATH		23.4/5.4	62	2/11	0.35
159	Vicilin-like seed storage protein At2g18540	010140	VCL21_ARATH	35.0/5.8	67.2/5.4	372	5/7	0.29
169	14-3-3-like protein B	001919	1433B_VICFA	35.3/5.1	29.7/4.9	172	3/18	0.45
	Vicilin-like seed storage protein	006304	VCL22_ARATH		61.9/5.9	75	2/5	0.13
	Eukaryotic translation initiation factor 3 subunit J	020855	EIF3J_NEMVE		24.4/4.7	73	2/9	0.35
	14-3-3-like protein B	010257	1433B_VICFA		29.6/5.0	63	2/11	0.28
	Vicilin-like seed storage protein	018839	AMP22_MACIN		60.9/6.6	63	4/7	0.28



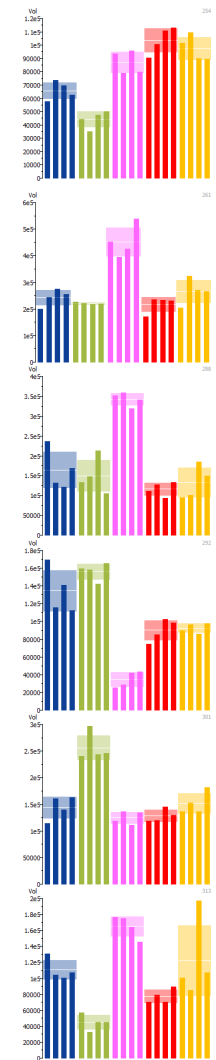
175	Agglutinin	007409	Q38719_AMAHP	35.8/6.8	30.1/6.5	559	5/27	0.83	
	Glucose and ribitol dehydrogenase	010964	GRDH_DAUCA		31.5/6.5	267	5/18	0.78	
176	Agglutinin	007409	Q38719_AMAHP	35.3/7.1	30.1/6.5	828	6/30	1.06	
	Glucose and ribitol dehydrogenase	010964	GRDH_DAUCA		31.5/6.5	358	6/21	1.00	
179	Agglutinin	007409	Q38719_AMAHP	36.1/6.6	30.1/6.5	181	4/21	0.62	
183	11S globulin	021282	13SB_FAGES	36.3/5.9	77.6/7.0	107	4/5	0.22	
189	Glutelin type-D 1	000876	GLUD1_ORYSJ	37.5/6.4	38.7/5.4	427	5/17	0.64	
	Glutelin type-D 1	015698	GLUA1_ORYSJ		36.1/6.3	383	3/13	0.38	
	11S globulin	021282	13SB_FAGES		77.6/7.0	111	2/4	0.11	
	Phosphoglycerate kinase, chloroplastic	006883	PGKH_SPIOL		87.8/7.6	102	3/5	0.14	
	UDP-arabinopyranose mutase 1	003560	RGP1_ORYSJ		40.4/5.7	93	2/5	0.21	
193	Vicilin-like seed storage protein	018839	AMP22_MACIN	37.3/6.9	60.9/6.6	99	4/8	0.30	

194	Malate dehydrogenase 1, mitochondrial	012920	MDHM1_ARATH	37.1/6.7	36.1/8.5	569	5/18	0.84	
	Agglutinin	007409	Q38719_AMAHP		30.1/6.5	298	4/21	0.79	
196	Malate dehydrogenase, mitochondrial	004479	MDHM_CITLA	37.7/6.4	36.2/8.4	110	2/8	0.24	
197	Agglutinin	007409	Q38719_AMAHP	37.3/6.6	30.1/6.5	270	4/21	0.60	
	Vicilin-like seed storage protein	018839	AMP22_MACIN		60.9/6.6	110	5/11	0.34	
198	Agglutinin	007409	Q38719_AMAHP	37.3/6.7	30.1/6.5	237	3/14	0.46	
202	Omega-amidase, chloroplastic	006839	NILP3_ARATH	37.6/6.5	40.4/8.6	106	2/8	0.21	
	Agglutinin	007409	Q38719_AMAHP		30.1/6.5	96	2/9	0.28	
213	Vicilin-like seed storage protein	018839	AMP22_MACIN	37.8/7.3	60.9/6.6	2433	10/20	0.90	
	Late embryogenesis abundant protein Lea14-A	013934	LEA14_GOSHI		36.4/5.0	127	3/11	0.34	
	40S ribosomal protein SA	011970	RSSA_SOYBN		30.1/5.0	83	2/9	0.26	

218	Vicilin-like seed storage protein	018839	AMP22_MACIN	37.7/5.2	60.9/6.6	602	9/18	0.69
	40S ribosomal protein SA	011970	RSSA_SOYBN		30.1/5.0	362	3/11	0.42
219	Vicilin-like seed storage protein	018839	AMP22_MACIN	38.1/6.8	60.9/6.6	254	5/10	0.38
	Annexin D2	020669	ANXD2_ARATH		36.0/6.1	235	6/22	0.92
	Malate dehydrogenase, mitochondrial	004479	MDHM_CITLA		36.2/8.4	165	4/15	0.54
221	Vicilin-like seed storage protein	018839	AMP22_MACIN	38.0/5.4	60.9/6.6	367	7/15	0.50
231	Vicilin-like seed storage protein	018839	AMP22_MACIN	38.2/7.0	60.9/6.6	669	10/20	0.79
237	Sorbitol dehydrogenase	006190	DHSO_ARATH	39.3/6.2	35.3/6.2	62	2/6	0.24
245	Glutelin type-D 1	000876	GLUD1_ORYSJ	39.6/6.0	38.7/5.4	224	5/17	0.63
	Glutelin type-D 1	015698	GLUA1_ORYSJ		36.1/6.3	187	4/17	0.52
	UDP-arabinopyranose mutase 1	003560	RGP1_ORYSJ		40.4/5.7	87	3/8	0.33

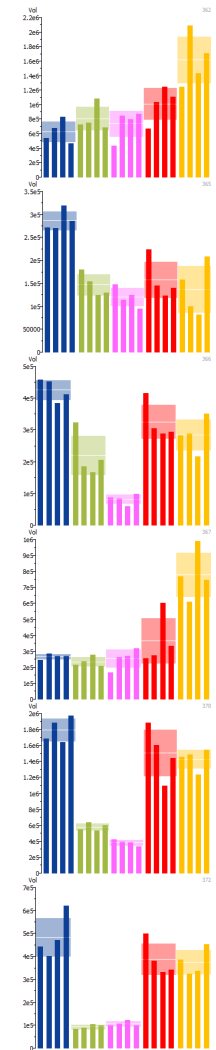


254	Probable protein disulfide-isomerase A6	018548	PDIA6_MEDSA	39.8/5.7	29.5/6.7	148	3/12	0.46
	Alpha-galactosidase	000222	AGAL_COFAR		66.9/9.1	98	4/7	0.25
	UDP-arabinopyranose mutase 1	003560	RGP1_ORYSJ		40.4/5.7	91	3/10	0.32
	Caffeic acid 3-O-methyltransferase	000613	COMT1_PRUDU		40.7/5.4	82	2/6	0.20
261	11S globulin	021282	13SB_FAGES	40.0/6.4	77.6/7.0	176	5/9	0.26
	Beta-galactosidase 8	006310	BGAL8_ARATH		87.5/6.4	171	3/5	0.13
	Alcohol dehydrogenase 1	005892	ADH1_PETHY		38.2/6.2	70	2/5	0.21
288	11S globulin	021282	13SB_FAGES	44.3/6.7	77.6/7.0	363	8/12	0.46
	SNF1-related protein kinase regulatory subunit gamma-like PV42b	005043	PV42B_ARATH		41.6/6.8	100	2/7	0.19
	Vignain	000235	CYSEP_RICCO		39.6/6.1	82	2/7	0.20
292	Isocitrate dehydrogenase [NADP]	001339	IDHC_TOBAC	45.3/6.4	46.2/5.8	113	5/12	0.53
301	UDP-D-apiiose/UDP-D-xylose synthase 2	014290	AXS2_ARATH	48.2/6.2	43.7/5.7	110	3/11	0.29
313	Vignain	000235	CYSEP_RICCO	48.6/6.7	39.6/6.1	160	3/11	0.34



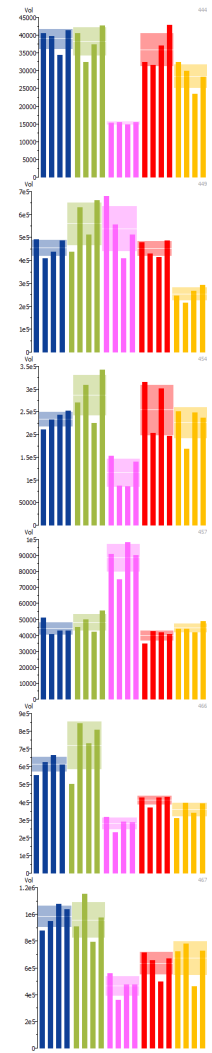
316	11S globulin	021282	13SB_FAGES	49.4/6.9	77.6/7.0	153	4/8	0.22	
	SNF1-related protein kinase regulatory subunit gamma-like PV42b	005043	PV42B_ARATH		41.6/6.8	56	2/5	0.20	
347	11S globulin	021282	13SB_FAGES	55.7/5.1	77.6/7.0	1046	9/17	0.59	
	Agglutinin	007409	Q38719_AMAHP		30.1/6.5	186	2/11	0.30	
	Vicilin-like seed storage protein	018839	AMP22_MACIN		60.9/6.6	146	5/12	0.39	
348	Enolase	001182	ENO_MESCR	58.4/5.9	48.2/5.5	1032	9/28	1.06	
352	Vicilin-like seed storage protein	018839	AMP22_MACIN	56.3/5.1	60.9/6.6	843	12/25	1.07	
	11S globulin	021282	13SB_FAGES		77.6/7.0	275	5/9	0.27	
355	Vicilin-like seed storage protein	018839	AMP22_MACIN	56.6/5.2	60.9/6.6	973	12/27	1.03	
	11S globulin	021282	13SB_FAGES		77.6/7.0	95	3/6	0.15	
356	Vicilin-like seed storage protein	018839	AMP22_MACIN	56.9/5.2	60.9/6.6	1132	3/28	1.15	

362	Vicilin-like seed storage protein	018839	AMP22_MACIN	57.3/5.3	60.9/6.6	231	6/14	0.43
	Elongation factor 1-alpha	001308	EF1A_ORYSJ		50.6/9.0	162	3/12	0.24
365	11S globulin	021282	13SB_FAGES	58.9/7.3	77.6/7.0	86	2/3	0.10
	Catalase	007232	CATA_IPOBA		52.3/6.8	82	3/6	0.23
	Vicilin-like seed storage protein	018839	AMP22_MACIN		60.9/6.6	70	2/3	0.13
366	Vicilin-like seed storage protein	018839	AMP22_MACIN	59.1/7.3	60.9/6.6	329	4/9	0.29
	11S globulin	021282	13SB_FAGES		77.6/7.0	102	4/5	0.22
367	Vicilin-like seed storage protein	018839	AMP22_MACIN	58.3/5.4	60.9/6.6	696	11/24	0.91
	Elongation factor 1-alpha	001308	EF1A_ORYSJ		50.6/9.0	164	4/11	0.33
370	11S globulin	021282	13SB_FAGES	58.5/7.1	77.6/7.0	1082	10/19	0.61
	Vicilin-like seed storage protein	018839	AMP22_MACIN		60.9/6.6	371	9/20	0.72
372	11S globulin	021282	13SB_FAGES	59.2/7.1	77.6/7.0	472	8/14	0.46
	Vicilin-like seed storage protein	018839	AMP22_MACIN		60.9/6.6	424	7/17	0.53



403	Granule-bound starch synthase 1, chloroplastic/amyloplastic	011500	SSG1_MANES	66.0/5.8	62.7/6.5	89	3/5	0.20	
413	Vicilin-like seed storage protein	018839	AMP22_MACIN	67.9/5.4	60.9/6.6	155	6/14	0.47	
425	2,3-bisphosphoglycerate-independent phosphoglycerate mutase	016738	PMGI_MESCR	72.8/5.9	60.1/5.5	252	7/14	0.51	
	Pyruvate decarboxylase 1	019658	PDC1_ARATH		44.9/5.6	104	2/6	0.17	
426	Granule-bound starch synthase 1, chloroplastic/amyloplastic	011500	SSG1_MANES	70.0/6.8	62.7/6.5	570	13/27	1.24	
	NADP-dependent malic enzyme	020870	MAOX_VITVI		65.6/6.0	117	4/8	0.27	
427	Granule-bound starch synthase 1, chloroplastic/amyloplastic	011500	SSG1_MANES	70.3/6.7	62.7/6.5	144	6/10	0.45	
	Vicilin-like seed storage protein	018839	AMP22_MACIN		60.9/6.6	75	3/6	0.21	
436	NADP-dependent malic enzyme	020870	MAOX_VITVI	72.5/6.5	65.6/6.0	273	10/13	0.48	

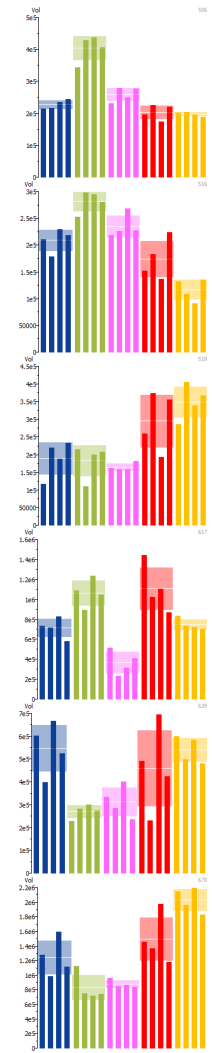
444	5-methyltetrahydropteroyltryptophan--homocysteine methyltransferase	017360	METE_MESCR	82.8/6.4	84.6/6.1	299	5/7	0.26
	Glycine--tRNA ligase, mitochondrial 1	003493	SYGM1_ARATH		76.0/5.9	80	2/3	0.11
449	Heat shock 70 kDa protein, mitochondrial	010311	HSP7M_PHAVU	79/5.5	33.3/5.0	149	4/12	0.56
454	Embryonic protein DC-8	000638	LEAD8_DAUCA	79.0/7.6	65.1/6.6	90	2/3	0.13
457	Embryonic protein DC-8	000638	LEAD8_DAUCA	85.7/6.5	65.1/6.6	238	6/10	0.43
	Agglutinin	007409	Q38719_AMAHP		30.1/6.5	70	2/9	0.29
466	Seed biotin-containing protein SBP65	013747	SBP65_PEA	90.6/7.3	72.4/6.7	84	2/5	0.11
467	Seed biotin-containing protein SBP65	013747	SBP65_PEA	89.1/7.3	72.4/6.7	96	3/5	0.17



472	5- methyltetrahydropteroyltriglutamate--homocysteine methyltransferase	017357	METE_MESCR	94.2/6.7	84.7/6.1	396	7/10	0.38	
	5- methyltetrahydropteroyltriglutamate--homocysteine methyltransferase	017360	METE_MESCR		84.6/6.1	365	7/10	0.38	
	5- methyltetrahydropteroyltriglutamate--homocysteine methyltransferase	022179	METE_CATRO		89.4/6.3	248	4/5	0.19	
478	5- methyltetrahydropteroyltriglutamate--homocysteine methyltransferase	017360	METE_MESCR	95.8/6.6	84.6/6.1	953	19/30	1.24	
	5- methyltetrahydropteroyltriglutamate--homocysteine methyltransferase	017357	METE_MESCR		84.7/6.1	885	18/28	1.15	
	5- methyltetrahydropteroyltriglutamate--homocysteine methyltransferase	022179	METE_CATRO		89.4/6.3	291	6/8	0.27	
479	5- methyltetrahydropteroyltriglutamate--homocysteine methyltransferase	017360	METE_MESCR	97.5/6.5	84.6/6.1	547	11/18	0.59	
	5- methyltetrahydropteroyltriglutamate--homocysteine methyltransferase	017357	METE_MESCR		84.7/6.1	539	11/18	0.59	
	Glycine--tRNA ligase, mitochondrial 1	003493	SYGM1_ARATH	76.0/5.9		302	7/11	0.39	
480	5- methyltetrahydropteroyltriglutamate--homocysteine methyltransferase	017360	METE_MESCR	99.6/6.4	84.6/6.1	517	8/12	0.45	
	5- methyltetrahydropteroyltriglutamate--homocysteine methyltransferase	017357	METE_MESCR		84.7/6.1	517	8/12	0.45	

483	Superoxide dismutase [Fe] 2, chloroplastic	009159	SODF2_ARATH	23.5/5.6	28.3/6.3	153	5/30	0.87	
	Vicilin-like seed storage protein	018839	AMP22_MACIN		60.9/6.6	122	3/6	0.19	
488	Seed biotin-containing protein SBP65	013747	SBP65_PEA	105.3/5.1	72.4/6.7	70	2/3	0.11	
492	Poly [ADP-ribose] polymerase 3	003773	PARP3_SOYBN	109.9/5.9	81.9/5.2	529	10/18	0.69	
494	5- methyltetrahydropteroyltrig lutamate--homocysteine methyltransferase	017360	METE_MESCR	109.5/5.4	84.6/6.1	94	3/4	0.13	
497	Poly [ADP-ribose] polymerase 3	003773	PARP3_SOYBN	106.2/5.4	81.9/5.2	259	7/9	0.45	
503	5- methyltetrahydropteroyltrig lutamate--homocysteine methyltransferase	017360	METE_MESCR	112.3/6.0	84.6/6.1	145	5/7	0.26	
	5- methyltetrahydropteroyltrig lutamate--homocysteine methyltransferase Poly [ADP-ribose] polymerase 3	017357	METE_MESCR		84.7/6.1	129	5/7	0.26	
		003773	PARP3_SOYBN		81.9/5.2	112	4/6	0.21	

506	Elongation factor 2	001926	EF2_BETVU	115.3/6.5	93.9/5.9	240	5/7	0.23
516	Chaperone protein ClpB1	008070	CLPB1_ARATH	116.8/6.4	100.7/5.7	585	15/21	0.72
	Chaperone protein ClpB1	012598	CLPB1_ARATH		81.4/6.4	248	8/13	0.43
	Low-temperature-induced 65 kDa protein	018897	LTI65_ARATH		87.0/5.7	162	3/4	0.13
	Elongation factor 2	001926	EF2_BETVU		93.9/5.9	63	2/2	0.08
518	Poly [ADP-ribose] polymerase 3	003773	PARP3_SOYBN	125.1/5.4	81.9/5.2	248	7/10	0.40
	Cell division cycle protein 48 homolog	012429	CDC48_SOYBN		90.7/5.2	131	4/5	0.19
	Heat shock protein 83	010680	HSP83_IPONI		39.9/4.5	115	2/7	0.21
617	11S globulin	021282	13SB_FAGES	9.5/5.4	77.6/7.0	179	4/6	0.21
639	Glutaredoxin	015561	GLRX_RICCO	10.5/6.9	13.9/7.7	161	3/30	1.15
670	Seed maturation protein	021176	A0A072VAH4_M EDTR	12.3/6.3	10.4/6.1	147	2/26	0.90



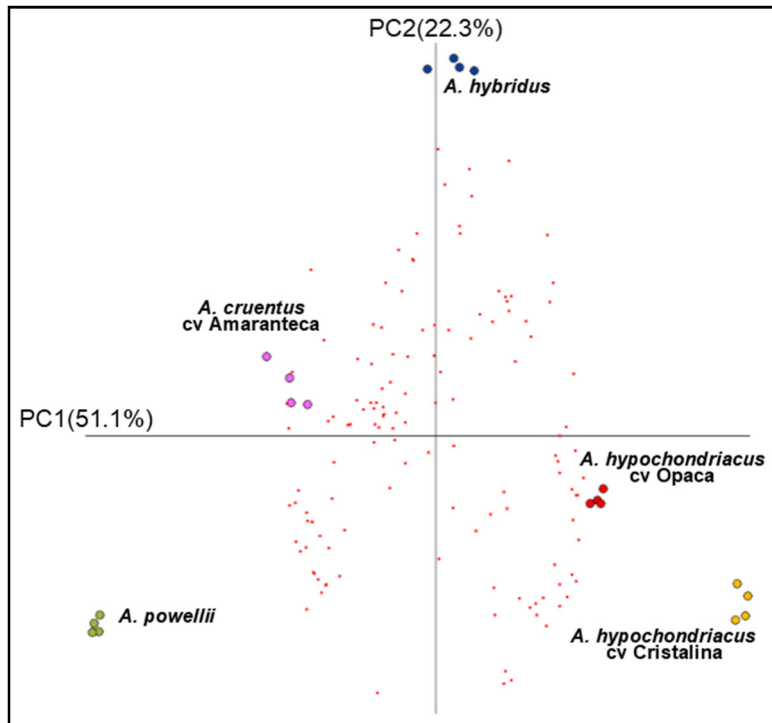
682	Vicilin-like seed storage protein	018839	AMP22_MACIN	13.2/6.0	60.9/6.6	1220	4/8	0.27	
683	Vicilin-like seed storage protein	018839	AMP22_MACIN	13.0/5.7	60.9/6.6	188	3/8	0.21	
689	Vicilin-like seed storage protein	018839	AMP22_MACIN	13.3/6.7	60.9/6.6	1091	3/8	0.19	
692	Vicilin-like seed storage protein	018839	AMP22_MACIN	13.7/6.0	60.9/6.6	474	3/8	0.19	
693	Vicilin-like seed storage protein	018839	AMP22_MACIN	13.5/7.2	60.9/6.6	304	2/5	0.13	
694	Vicilin-like seed storage protein	018839	AMP22_MACIN	13.6/5.8	60.9/6.6	221	2/5	0.13	

696	Major allergen Mal d 1	006247	MAL12_MALDO	13.7/5.7	17.9/6.2	210	6/46	2.32	
	Vicilin-like seed storage protein	018839	AMP22_MACIN		60.9/6.6	168	2/5	0.13	
737	Major allergen Mal d 1	006247	MAL12_MALDO	15.9/6.4	17.9/6.2	301	5/38	1.63	
	Dessication-induced 1VOC superfamily protein	014874	Q9LQP1_ARATH		11.0/5.0	167	2/20	0.85	
	4-hydroxy-4-methyl-2-oxoglutarate aldolase	013028	RRAA3_ARATH		17.0/5.1	44	2/12	0.50	
755	Vicilin-like seed storage protein	018839	AMP22_MACIN	17.3/6.6	60.9/6.6	419	3/8	0.21	
760	Peroxiredoxin-2B	015627	PRX2B_ARATH	17.4/6.1	17.5/5.9	564	9/39	4.62	
767	CBS domain-containing protein CBSX3, mitochondrial	009852	CBSX3_ARATH	18.1/6.6	22.4/7.9	148	5/18	1.28	
	CBS domain-containing protein CBSX3, mitochondrial	007851	CBSX3_ARATH		20.1/6.6	89	4/19	1.07	
771	18.3 kDa class I heat shock protein	013876	HSP11_OXYRB	17.7/6.0	17.9/5.8	547	6/43	3.24	

772	18.3 kDa class I heat shock protein	013876	HSP11_OXYRB	18.1/5.6	17.9/5.8	413	3/24	0.88	
784	11S globulin	021282	13SB_FAGES	19.0/5.7	77.6/7.0	495	7/9	0.39	
798	11S globulin	021282	13SB_FAGES	19.6/6.4	77.6/7.0	107	3/6	0.15	
	Vicilin-like seed storage protein	018839	AMP22_MACIN		60.9/6.6	70	2/5	0.13	
817	Vicilin-like seed storage protein At2g18540	010140	VCL21_ARATH	20.8/5.3	67.2/5.4	142	2/3	0.11	
821	Vicilin-like seed storage protein At2g18540	010140	VCL21_ARATH	20.0/5.1	67.2/5.4	324	5/12	0.30	

^aSpot number assigned by Melanie software. ^bAccession number according to the database reported by Clouse et al. 2016. ^cUniProtKB/Swiss-Prot ortholog identifier assigned by Trinotate. ^dExperimental mass and isoelectric point. ^eTheoretical mass and isoelectric point. ^fMASCOT Score, individual ion scores >33 were statistically significant ($p < 0.01$), only identifications with peptide matches above the identity threshold when $FDR \leq 5\%$ were considered true. ^gPeptides Matched/Sequence Coverage. ^hExponentially Modified Protein Abundance Index. ⁱProtein spot accumulation change histograms ($p \leq 0.001$ and fold change ≥ 2.0): A, *A. hybridus*; B, *A. powellii*; C, *A. cruentus* cv Amaranteca; D, *A. hypochondriacus* cv Opaca; E, *A. hypochondriacus* cv Cristalina.

A)



B)

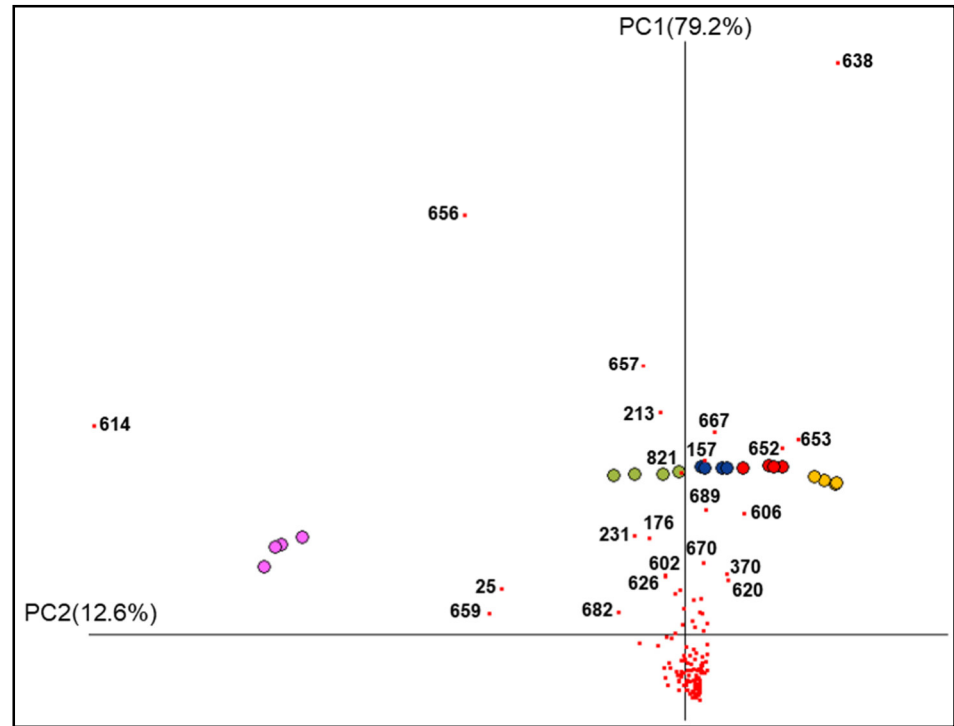


Figure 3.2. PCA constructed with the densitometric data of 152 differentially accumulated spots in the hydrophilic fraction. A) Images as observations, B) Spots as observations. Coloured dots represent gel replicates (Blue, *A. hybridus*; Green, *A. powellii*; Pink, *A. cruentus* cv Amaranteca; Red, *A. hypochondriacus* cv Opaca; Yellow, *A. hypochondriacus* cv Cristalina.) and each red square indicate one spot.

As mentioned previously the differences in the protein accumulation patterns in this fraction are noticeable in the low molecular weight region, reflected mainly by the spots 25, 31, 614, 638, 652, 653, 656, 657, 659, 667, 670, and 689 (Figure 3.3). Of these, only spots 31 and 670 were identified, as Vicilin-like seed storage protein and Seed maturation protein, respectively, the former accumulated mainly in cultivated species and the second, present in *A. hypochondriacus* cultivars and in the wild species *A. hybridus*. Most of the spots in the hydrophilic fraction which contribute to the differentiated grouping of the species, reflected by the PCA (Figure 3.2B), and that are highlighted in figure 3.3, were identified as isoforms of Legumins (11S globulins) and Vicilins (7S globulins) seed storage proteins. Although with variable proportions, in general these proteins are up accumulated in the wild species *A. hybridus* and in the *A. hypochondriacus* cultivars. 11S globulin (001411) in spots 151 and 157 have higher levels of accumulation in the two *A. hypochondriacus* cultivars, Vicilin-like seed storage protein (018839) in spots 218 and 362 are predominant in *A. hypochondriacus* cv Cristalina, and Vicilin-like seed storage protein At2g18540 (010140) in spot 821 is mainly present in *A. hybridus*. Vicilin-like (018839) and 11S globulin (021282) were identified in spots 213 and 370, both proteins present in spot 370 but in 213 only Vicilin-like was found, the highest abundance on these spots is shared between *A. hybridus* and *A. hypochondriacus* cultivars.

In addition to storage proteins, the analysis of the soluble fraction reveals other interesting and clearly differential proteins amongst the amaranth species. Oil body-associated proteins (OBAPs) were up accumulated in *A. cruentus* and *A. powellii* but as different homologues, OBAP 1A (009953), spot 112, have higher levels in *A. powellii*, and OBAP 2A (004342), spot 106, in *A. cruentus*. In the same way as storage proteins, agglutinin (007409) shows variable accumulation patterns, in spot 126 is up accumulated only in *A. cruentus*; presents higher levels in the cultivated species in spots 175, 176 and 179; and have elevated levels in all species, except *A. cruentus*, in spots 197, 198 and 202. Seed biotin-containing protein SBP65 (013747) is increased in wild species in spots 466 and 467 but have mayor levels in the *A. hypochondriacus* cultivars in spot 488. Three proteoforms of 5- methyltetrahy-

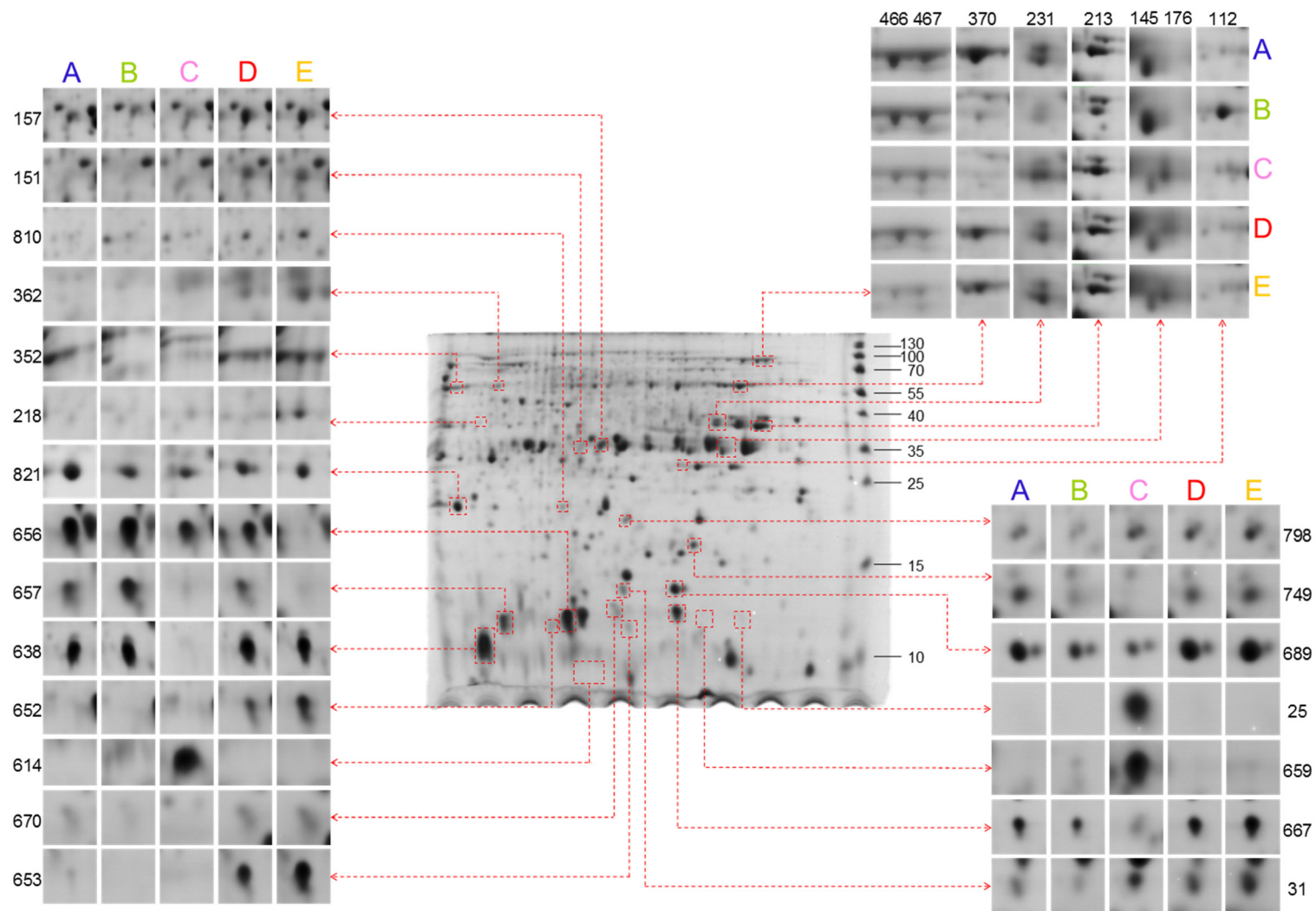


Figure 3.3. Representative proteomic map of the hydrophilic fraction and comparative accumulation profiles of the spots indicated in figure 3.3 B. A, *A. hybridus*; B, *A. powellii*; C, *A. cruentus* cv Amaranteca; D, *A. hypochondriacus* cv Opaca; E, *A. hypochondriacus* cv Cristalina.

dropteroyltriglutamate-homocysteine methyltransferase were found in seven differential spots; in spots 444 (017360) and 472 (022179, 017357 and 017360) down accumulated only in *A. cruentus*; with higher levels in *A. powellii* in spots 478 (022179, 017357 and 017360), 479 and 480 (017357 and 017360), accumulated more in *A. hypochondriacus* Cristalina in spot 494 (017360) and in *A. hybridus* in spot 503 (017357 and 017360). Finally, in this fraction the 18.3 kDa class I HSP (013876) presents a greater intensity in spots 72, 771 and 772 uniquely in *A. cruentus*, and the Major allergen Mal d 1 have higher levels in *A. powellii* (spot 737) and *A. cruentus* (spot 696).

3.3.2 Hydrophobic fraction

The hydrophobic fraction also reflects evident differences on protein accumulation amongst species, but in contrast to the hydrophilic fraction, is not found in the low molecular weight region but in the 55-70 kDa range, towards the alkaline end of the pH gradient, where a variable region stands out, which divides the analysed species into two groups based on their similarity of the spots profiles (Figure 3.4). The first group is formed by the wild species *A. hybridus* and *A. powellii* as well as the cultivated species *A. hypochondriacus* cv Cristalina, which shares two lines of spots in the mentioned region. In the second group are included the cultivated species *A. cruentus* cv Amaranteca y *A. hypochondriacus* cv Opaca that present only one of these patterns. These profiles in the hydrophobic fraction are in concordance with the previous finding by 1-D-SDS-PAGE described for bands 27, 28 and 29 in section 2.3.1 (Figure 2.2), unveiling that each of these bands are constituted by several proteo- or isoelectric forms.

The proteomic maps analysis of this fraction reveals 120 differentially accumulated spots, 95 of them were identified, for a total of 42 unique proteins (Table 3.2). In the PCA of this fraction *A. hypochondriacus* cv Cristalina is grouped together with wild species in both, when replicates or individual spots are established as dependent variables (Figure 3.5). This is because of, amongst the cultivable species, only *A. hypochondriacus* cv Crystalline shares the two sets of spots in the region of

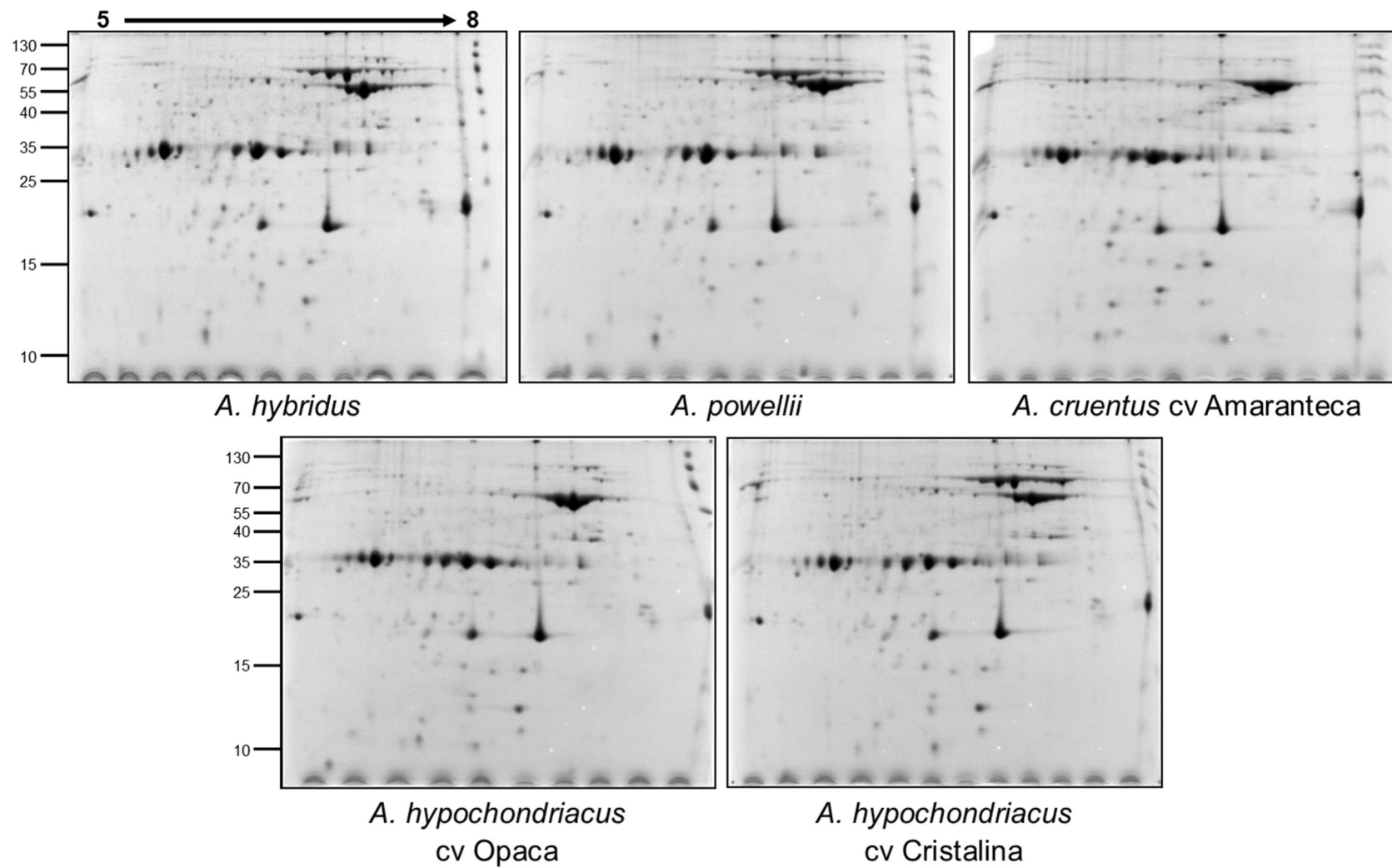
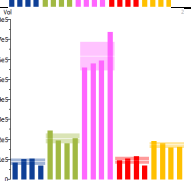
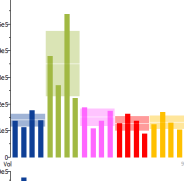
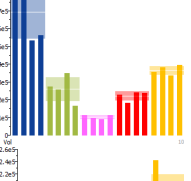
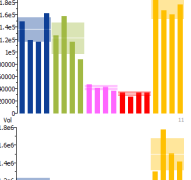
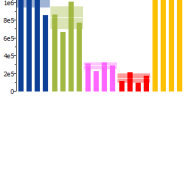
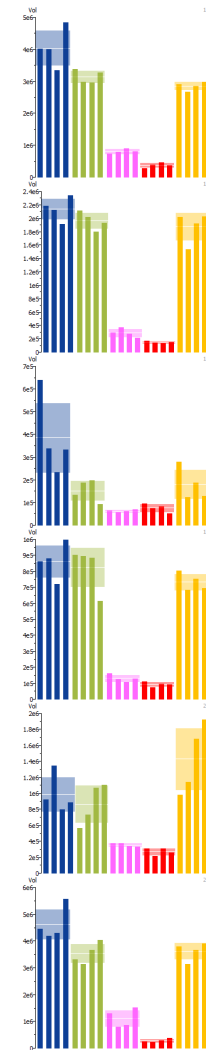


Figure 3.4. Representative proteomic maps of the hydrophobic fraction from amaranth seeds species.

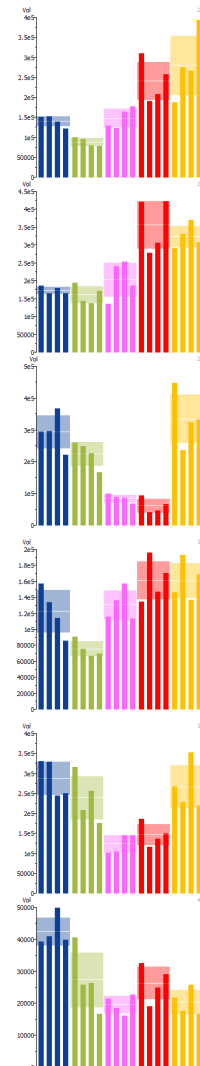
Table 3.2. Hydrophobic amaranth seed proteins identified by nLC-MS/MS in differentially 2-DE accumulated spots.

Spot No. ^a	Protein	Accession No. ^b	Ortholog ^c	Mr(kDa)/pI Exp. ^d	Mr(kDa)/pI Theo. ^e	Mascot Score ^f	PM/SC (%) ^g	emPAI ^h	Spot accumulation change ⁱ
									A B C D E
2	11S globulin	021282	13SB_FAGES	55.4/7.0	77.6/7.0	287	5/9	0.33	
8	11S globulin	021282	13SB_FAGES	64.3/7.6	77.6/7.0	369	9/15	0.53	
9	Granule-bound starch synthase I, chloroplastic/amyloplastic	011500	SSG1_MANES	73.6/7.4	62.7/6.5	988	19/42	2.23	
	11S globulin	021282	13SB_FAGES		77.6/7.0	549	10/18	0.69	
10	Granule-bound starch synthase I, chloroplastic/amyloplastic	011500	SSG1_MANES	73.3/6.6	62.7/6.5	1096	18/38	2.38	
11	Granule-bound starch synthase I, chloroplastic/amyloplastic	011500	SSG1_MANES	74.2/7.3	62.7/6.5	1349	22/50	3.09	
	11S globulin	021282	13SB_FAGES		77.6/7.0	180	4/6	0.21	

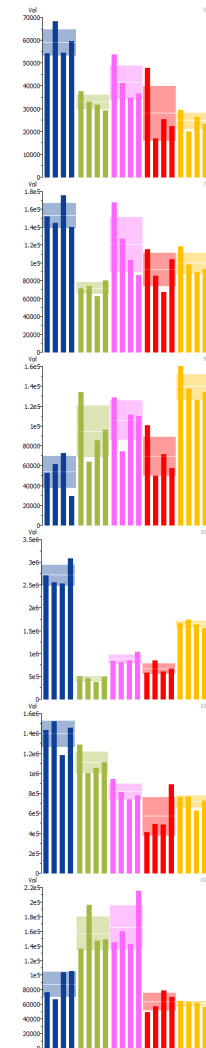
12	Granule-bound starch synthase I, chloroplastic/amyloplastic	011500	SSG1_MANES	69.2/7.0	62.7/6.5	1775	20/43	2.98
13	Granule-bound starch synthase I, chloroplastic/amyloplastic	011500	SSG1_MANES	70.8/6.8	62.7/6.5	1508	19/43	2.56
15	Granule-bound starch synthase I, chloroplastic/amyloplastic	011500	SSG1_MANES	71.9/7.5	62.7/6.5	1298	23/52	3.34
	11S globulin	021282	13SB_FAGES		77.6/7.0	331	6/9	0.33
16	Granule-bound starch synthase I, chloroplastic/amyloplastic	011500	SSG1_MANES	72.7/6.7	62.7/6.5	1666	22/45	3.41
20	Granule-bound starch synthase I, chloroplastic/amyloplastic	011500	SSG1_MANES	75.0/7.2	62.7/6.5	1295	21/45	3.60
25	Granule-bound starch synthase I, chloroplastic/amyloplastic	011500	SSG1_MANES	68.7/7.0	62.7/6.5	1346	16/33	1.58



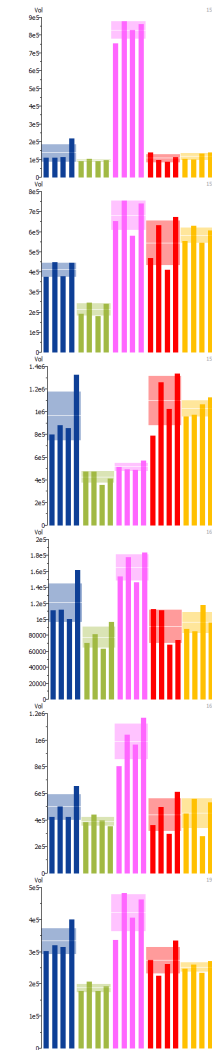
27	Granule-bound starch synthase I, chloroplastic/amyloplastic	011500	SSG1_MANES	73.8/5.7	62.7/6.5	852	10/25	0.74
	11S globulin	021282	13SB_FAGES		77.6/7.0	687	12/22	0.71
	Chaperonin CPN60, mitochondrial	003944	CH60A_ARATH		57.7/6.4	462	12/25	1.05
28	11S globulin	021282	13SB_FAGES	75.3/5.7	77.6/7.0	597	13/22	0.84
	Granule-bound starch synthase I, chloroplastic/amyloplastic	011500	SSG1_MANES		62.7/6.5	324	7/16	0.50
29	Granule-bound starch synthase I, chloroplastic/amyloplastic	011500	SSG1_MANES	74.7/6.5	62.7/6.5	1380	22/49	3.68
30	11S globulin	021282	13SB_FAGES	74.4/5.6	77.6/7.0	367	11/20	0.68
	Granule-bound starch synthase I, chloroplastic/amyloplastic	011500	SSG1_MANES		62.7/6.5	270	6/14	0.42
	Chaperonin CPN60, mitochondrial	003944	CH60A_ARATH		57.7/6.4	242	11/24	1.00
	Chaperonin CPN60, mitochondrial	014580	CH60A_ARATH			197	8/15	0.61
36	Granule-bound starch synthase I, chloroplastic/amyloplastic	011500	SSG1_MANES	77.3/6.4	62.7/6.5	1162	19/41	2.82
40	Granule-bound starch synthase I, chloroplastic/amyloplastic	011500	SSG1_MANES	85.5/6.3	62.7/6.5	132	3/8	0.20



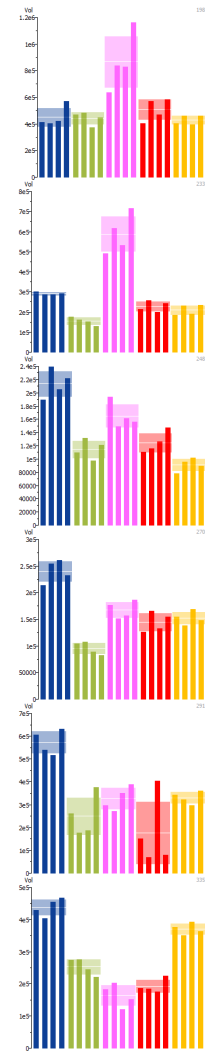
52	NADH dehydrogenase [ubiquinone] iron-sulphur protein 1, mitochondrial	001360	NDUS1_SOLTU	93.7/6.12	81.0/6.1	370	10/17	0.63
72	Alpha-xylosidase 1	005558	XYL1_ARATH	100.8/6.1	104.0/5.9	120	4/4	0.17
92	11S globulin	021282	13SB_FAGES	124.3/7.1	77.6/7.0	158	5/9	0.27
103	11S globulin	021282	13SB_FAGES	54.0/7.1	77.6/7.0	1121	10/18	0.59
105	Agglutinin	007409	Q38719_AMAHP	37.0/7.0	30.1/6.5	354	6/30	1.24
108	11S globulin	021282	13SB_FAGES	54.0/7.1	77.6/7.0	251	6/11	0.33



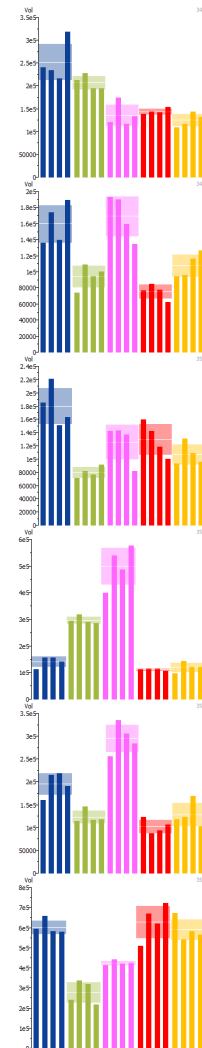
151	Vicilin-like seed storage protein	018839	AMP22_MACIN	13.0/5.9	60.9/6.6	183	2/5	0.13
152	Vicilin-like seed storage protein	018839	AMP22_MACIN	13.1/6.3	60.9/6.6	975	3/8	0.20
156	Vicilin-like seed storage protein	018839	AMP22_MACIN	13.4/6.7	60.9/6.6	1147	4/8	0.30
163	Vicilin-like seed storage protein	018839	AMP22_MACIN	13.8/6.0	60.9/6.6	116	2/5	0.14
168	Late embryogenesis abundant protein B19.3	008005	LE193_HORVU	14.1/6.3	9.7/5.9	147	2/26	1.06
	Protein SLE2	019862	SLE2_SOYBN		8.5/5.9	62	2/30	1.25
197	17.9 kDa class II heat shock protein	006834	HSP21_SOYBN	15.8/6.5	18.2/6.2	222	4/30	1.23



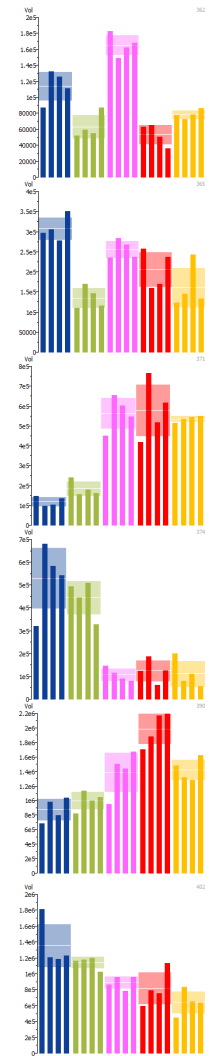
198	17.4 kDa class I heat shock protein;	012223	HSP17_ARATH	16.2/6.8	17.4/6.3	178	3/21	0.83
233	18.3 kDa class I heat shock protein	013876	HSP11_OXYRB	17.9/5.9	17.9/5.8	326	5/32	1.66
248	17.4 kDa class I heat shock protein	013881	HSP17_ARATH	19.4/5.7	19.1/5.6	170	4/25	1.55
270	Vicilin-like seed storage protein At2g18540	010140	VCL21_ARATH	21.2/5.3	67.2/5.4	115	2/3	0.12
291	Vicilin-like seed storage protein	006304	VCL22_ARATH	22.8/7.6	61.9/5.9	391	6/14	0.42
335	Vicilin-like seed storage protein	010140	VCL21_ARATH	26.7/5.9	67.2/5.4	171	4/5	0.25



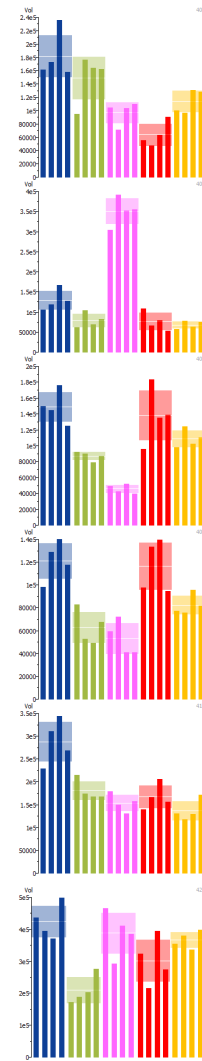
345	Vicilin-like seed storage protein At2g18540	010140	VCL21_ARATH	28.3/6.3	67.2/5.4	154	4/8	0.26
348	Peroxygenase	016889	PXG_SESIN	29.4/5.7	26.1/5.5	331	6/30	1.61
	Granule-bound starch synthase I, chloroplastic/amyloplastic	011500	SSG1_MANES		62.7/6.5	134	3/6	0.19
	Triosephosphate isomerase, cytosolic	017821	TPIS_COPJA		18.7/5.2	120	3/20	0.76
353	Oil body-associated protein 1A	009953	OBP1A_ARATH	31.1/7.0	26.6/6.2	130	3/16	0.66
354	Oil body-associated protein 1A	009953	OBP1A_ARATH	30.6/6.8	26.6/6.2	387	10/52	3.92
355	Oil body-associated protein 1A	009953	OBP1A_ARATH	29.7/6.7	26.6/6.2	159	2/11	0.51
	Proteasome subunit alpha type-6	008388	PSA6_TOBAC		27.3/6.1	99	3/14	0.49
357	Oil body-associated protein 1A	009953	OBP1A_ARATH		26.6/6.2	379	6/33	1.65
	11S globulin	021282	13SB_FAGES		77.6/7.0	149	2/3	0.10



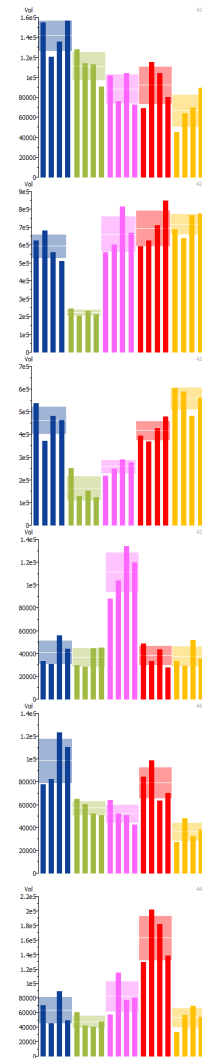
362	Vicilin-like seed storage protein	006304	VCL22_ARATH	30.5/7.3	61.9/5.9	112	3/8	0.20
365	Peroxygenase	016889	PXG_SESIN	31.0/6.0	26.1/5.5	184	4/21	1.01
	Triosephosphate isomerase, cytosolic	017821	TPIS_COPJA		18.7/5.2	62	2/14	0.47
371	Vicilin-like seed storage protein	006304	VCL22_ARATH	31.5/5.3	61.9/5.9	184	4/9	0.28
374	Vicilin-like seed storage protein	006304	VCL22_ARATH	31.7/5.3	61.9/5.9	633	5/11	0.44
390	11S globulin	001411	CRU1_RAPSA	35.8/6.0	55.4/6.3	205	3/5	0.24
402	Agglutinin	007409	Q38719_AMAHP	35.8/7.0	30.1/6.5	641	5/27	1.00
	11S globulin	021282	13SB_FAGES		77.6/7.0	232	4/8	0.20
	Glucose and ribitol dehydrogenase	010964	GRDH_DAUCA		31.5/6.5	86	2/8	0.25



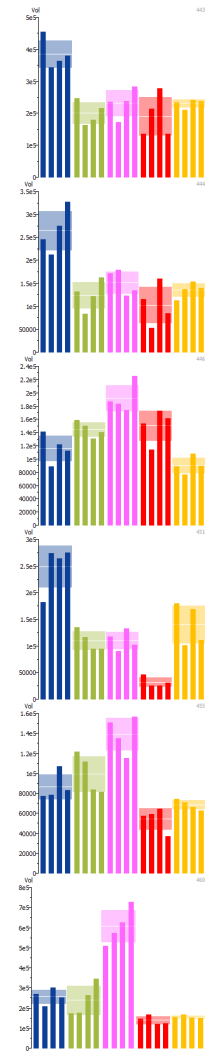
404	Annexin-like protein RJ4	021089	ANX4_FRAAN	36.6/7.3	24.3/7.8	219	7/30	2.16
	Agglutinin	007409	Q38719_AMAHP		30.1/6.5	149	3/14	0.42
405	11S globulin	021282	13SB_FAGES	36.7/5.8	77.6/7.0	575	9/17	0.61
408	Agglutinin	007409	Q38719_AMAHP	38.9/6.7	30.1/6.5	131	3/14	0.42
	Vicilin-like seed storage protein	018839	AMP22_MACIN		60.9/6.6	71	3/6	0.19
	Malate dehydrogenase, mitochondrial	004479	MDHM_CITLA		36.2/8.4	57	2/7	0.21
409	Agglutinin	007409	Q38719_AMAHP	37.8/6.6	30.1/6.5	127	3/14	0.44
411	Agglutinin	007409	Q38719_AMAHP	37.7/6.5	30.1/6.5	80	2/9	0.30
421	Vicilin-like seed storage protein	018839	AMP22_MACIN	38.3/7.4	60.9/6.6	381	8/15	0.63
	11S globulin	021282	13SB_FAGES		77.6/7.0	132	2/4	0.10
	Bifunctional UDP-glucose 4-epimerase and UDP-xylose 4-epimerase 1	012350	UGE1_PEA		37.6/6.5	77	2/6	0.22
	Malate dehydrogenase, glyoxysomal	013009	MDHG_CUCSA		37.6/8.1	67	2/7	0.22



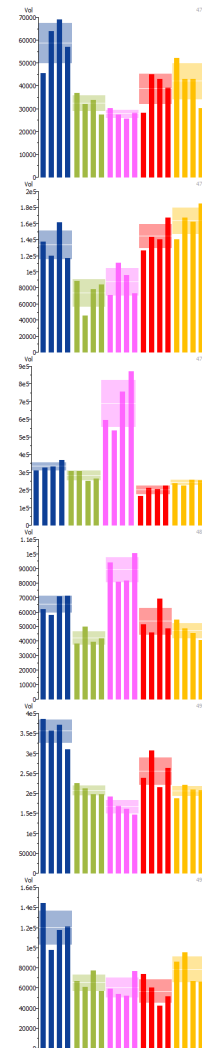
422	60S acidic ribosomal protein P0-1	013703	RLA01_ARATH	38.5/5.6	34.3/5.2	142	4/15	0.56
	11-beta-hydroxysteroid dehydrogenase 1B	004692	HSD1B_ARATH		73.7/5.7	90	3/5	0.17
424	Vicilin-like seed storage protein	018839	AMP22_MACIN	38.7/7.3	60.9/6.6	602	9/18	0.79
425	Vicilin-like seed storage protein	018839	AMP22_MACIN	38.9/7.3	60.9/6.6	720	8/15	0.69
429	Vicilin-like seed storage protein	018839	AMP22_MACIN	39.2/6.6	60.9/6.6	246	5/10	0.42
	Agglutinin	007409	Q38719_AMAHP		30.1/6.5	111	3/14	0.42
	Annexin D2	020669	ANXD2_ARATH		36.0/6.1	97	2/6	0.22
	Annexin D2	020326	ANXD2_ARATH		32.1/6.1	60	2/6	0.25
	2-alkenal reductase (NADP(+)-dependent)	016106	DBR_TOBAC		40.6/6.3	49	2/5	0.19
440	11-beta-hydroxysteroid dehydrogenase 1B	004692	HSD1B_ARATH	39.3/5.3	73.7/5.7	60	3/5	0.18
442	Vicilin-like seed storage protein	018839	AMP22_MACIN	40.5/7.0	60.9/6.6	252	6/13	0.47
	Agglutinin	007409	Q38719_AMAHP		30.1/6.5	102	2/11	0.29
	Glyceraldehyde-3-phosphate dehydrogenase	011043	G3P_ATRNU		31.6/6.7	48	2/9	0.27



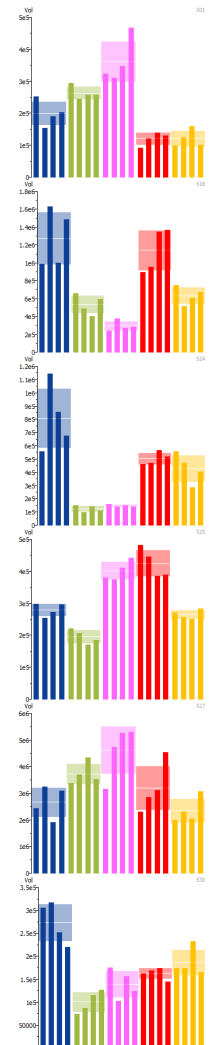
443	Glyceraldehyde-3-phosphate dehydrogenase	011043	G3P_ATRNU	39.3/7.6	31.6/6.7	395	10/36	2.58
444	Bifunctional UDP-glucose 4-epimerase and UDP-xylose 4-epimerase 1	012350	UGE1_PEA	38.8/7.7	37.6/6.5	70	2/5	0.23
446	Glyceraldehyde-3-phosphate dehydrogenase	011043	G3P_ATRNU	41.3/7.1	31.6/6.7	241	6/19	1.33
451	Fructose-bisphosphate aldolase, cytoplasmic isozyme	000665	ALF_SPIOL	41.2/6.8	38.3/6.2	203	4/18	0.51
	Alcohol dehydrogenase 1	005892	ADH1_PETHY		38.2/6.2	165	5/13	0.67
455	11S globulin	021282	13SB_FAGES	41.8/6.2	77.6/7.0	657	12/21	0.76
	Aldose 1-epimerase	015176	GALM_PIG		33.2/9.7	91	2/7	0.25
	Sorbitol dehydrogenase	006190	DHSO_ARATH		35.3/6.2	71	2/6	0.23
460	11S globulin	021282	13SB_FAGES	41.1/6.4	77.6/7.0	856	13/22	0.84



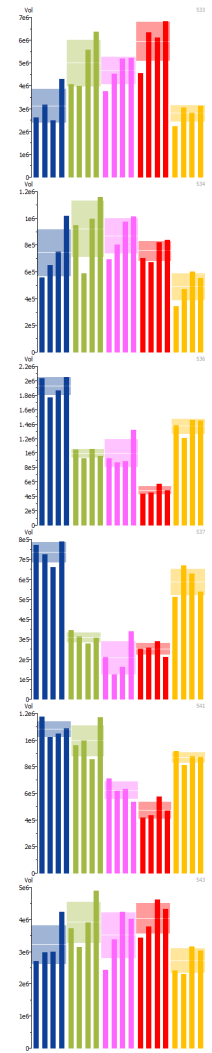
472	11S globulin	021282	13SB_FAGES	44.3/6.6	77.6/7.0	394	9/20	0.53
	Alcohol dehydrogenase 1	005892	ADH1_PETHY		38.2/6.2	149	2/5	0.21
473	Aspartate aminotransferase, cytoplasmic	013661	AATC_DAUCA	44.5/7.3	44.3/6.6	91	3/9	0.28
477	11S globulin	021282	13SB_FAGES	47.3/6.7	77.6/7.0	571	13/22	1.00
481	11S globulin	021282	13SB_FAGES	47.3/6.6	77.6/7.0	439	9/14	0.51
491	S-adenosylmethionine synthase 4	018844	METK4_ATRNU	49.6/5.9	43.3/5.5	448	8/26	1.06
	Granule-bound starch synthase I, chloroplastic/amyloplastic	011500	SSG1_MANES		62.7/6.5	116	4/8	0.25
	Caffeine synthase 1	013708	TCS1_CAMSI		37.7/5.7	75	3/11	0.32
494	11S globulin	021282	13SB_FAGES	52.1/7.1	77.6/7.0	181	4/9	0.22



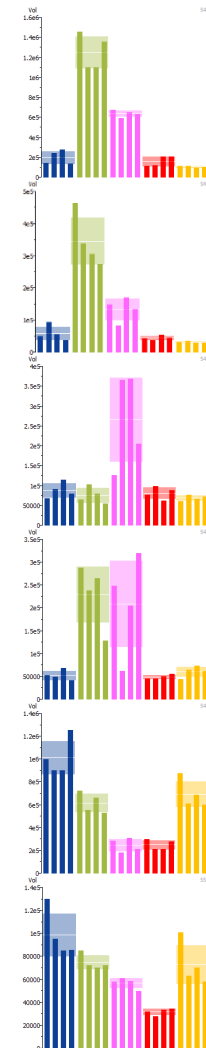
501	11S globulin	021282	13SB_FAGES	52.9/7.0	77.6/7.0	465	12/21	0.76
518	11S globulin	021282	13SB_FAGES	58.9/7.4	77.6/7.0	944	14/26	1.03
	Serine hydroxymethyltransferase 4	009350	GLYC4_ARATH		59.1/7.9	149	4/11	0.28
524	11S globulin	021282	13SB_FAGES	59.7/7.5	77.6/7.0	494	10/15	0.60
	Serine hydroxymethyltransferase 4	009350	GLYC4_ARATH		59.1/7.9	114	3/8	0.20
525	Granule-bound starch synthase I, chloroplastic/amyloplastic	011500	SSG1_MANES	60.2/6.5	62.7/6.5	101	2/3	0.12
	11S globulin	021282	13SB_FAGES		77.6/7.0	77	2/4	0.10
527	11S globulin	021282	13SB_FAGES	60.0/7.3	77.6/7.0	672	13/21	0.84
530	11S globulin	021282	13SB_FAGES	61.8/6.8	77.6/7.0	267	6/11	0.33
	Vicilin-like seed storage protein	018839	AMP22_MACIN		60.9/6.6	134	4/9	0.27
	Granule-bound starch synthase I, chloroplastic/amyloplastic	011500	SSG1_MANES		62.7/6.5	87	3/6	0.19



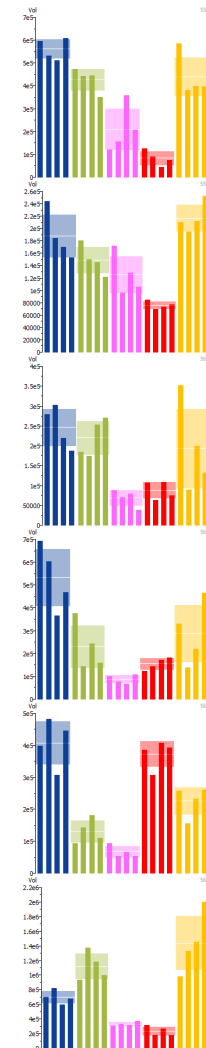
533	11S globulin	021282	13SB_FAGES	60.0/7.2	77.6/7.0	993	11/18	0.66
534	11S globulin	021282	13SB_FAGES	63.8/7.1	77.6/7.0	690	13/23	0.84
536	11S globulin	021282	13SB_FAGES	59.1/7.0	77.6/7.0	1184	12/22	0.88
537	11S globulin	021282	13SB_FAGES	61.7/6.9	77.6/7.0	594	11/18	0.69
	Granule-bound starch synthase I, chloroplastic/amyloplastic	011500	SSG1_MANES	62.7/6.5		496	11/24	0.91
	Vicilin-like seed storage protein	018839	AMP22_MACIN	60.9/6.6		446	8/19	0.72
541	11S globulin	021282	13SB_FAGES	61.0/7.0	77.6/7.0	768	15/26	1.03
	Granule-bound starch synthase I, chloroplastic/amyloplastic	011500	SSG1_MANES	62.7/6.5		185	5/10	0.34
	Vicilin-like seed storage protein	018839	AMP22_MACIN	60.9/6.6		142	3/7	0.20
543	11S globulin	021282	13SB_FAGES		77.6/7.0	825	13/23	0.84



544	11S globulin	021282	13SB_FAGES	63.3/7.4	77.6/7.0	574	11/19	0.68
546	11S globulin	021282	13SB_FAGES	64.3/7.5	77.6/7.0	295	8/15	0.46
	Granule-bound starch synthase I, chloroplastic/amyloplastic	011500	SSG1_MANES		62.7/6.5	85	2/4	0.12
547	Enolase	001182	ENO_MESCR	63.4/5.9	48.2/5.5	159	4/13	0.36
	UTP--glucose-1-phosphate uridylyltransferase	008585	UGPA_SOLTU		48.5/5.8	77	3/8	0.26
548	11S globulin	021282	13SB_FAGES	64.0/7.5	77.6/7.0	565	11/21	0.68
549	Granule-bound starch synthase I, chloroplastic/amyloplastic	011500	SSG1_MANES	62.3/6.4	62.7/6.5	1279	16/38	2.23
550	Granule-bound starch synthase I, chloroplastic/amyloplastic	011500	SSG1_MANES	64.3/6.5	62.7/6.5	928	14/30	1.49



551	Granule-bound starch synthase I, chloroplastic/amyloplastic	011500	SSG1_MANES	63.5/6.3	62.7/6.5	1063	19/46	2.91
553	Granule-bound starch synthase I, chloroplastic/amyloplastic	011500	SSG1_MANES	64.4/6.5	62.7/6.5	986	17/36	1.82
565	Granule-bound starch synthase I, chloroplastic/amyloplastic	011500	SSG1_MANES	71.4/7.6	62.7/6.5	992	19/40	2.28
566	Granule-bound starch synthase I, chloroplastic/amyloplastic	011500	SSG1_MANES	71.4/7.5	62.7/6.5	846	16/35	1.56
	11S globulin	021282	13SB_FAGES		77.6/7.0	353	8/12	0.46
567	Granule-bound starch synthase I, chloroplastic/amyloplastic	011500	SSG1_MANES	73.6/7.4	62.7/6.5	1106	17/36	1.88
	11S globulin	021282	13SB_FAGES		77.6/7.0	404	8/13	0.46
568	Granule-bound starch synthase I, chloroplastic/amyloplastic	011500	SSG1_MANES	74.7/7.1	62.7/6.5	1685	23/52	4.09



^aSpot number assigned by Melanie software. ^bAccession number according to the database reported by Clouse et al. 2016. ^cUniProtKB/Swiss-Prot ortholog identifier assigned by Trinotate. ^dExperimental mass and isoelectric point. ^eTheoretical mass and isoelectric point. ^fMASCOT Score, individual ion scores >33 were statistically significant ($p < 0.01$), only identifications with peptide matches above the identity threshold when $FDR \leq 5\%$ were considered true. ^gPeptides Matched/Sequence Coverage. ^hExponentially Modified Protein Abundance Index. ⁱProtein spot accumulation change histograms ($p \leq 0.001$ and fold change ≥ 2.0): A, *A. hybridus*; B, *A. powellii*; C, *A. cruentus* cv Amaranteca; D, *A. hypochondriacus* cv Opaca; E, *A. hypochondriacus* cv Cristalina.

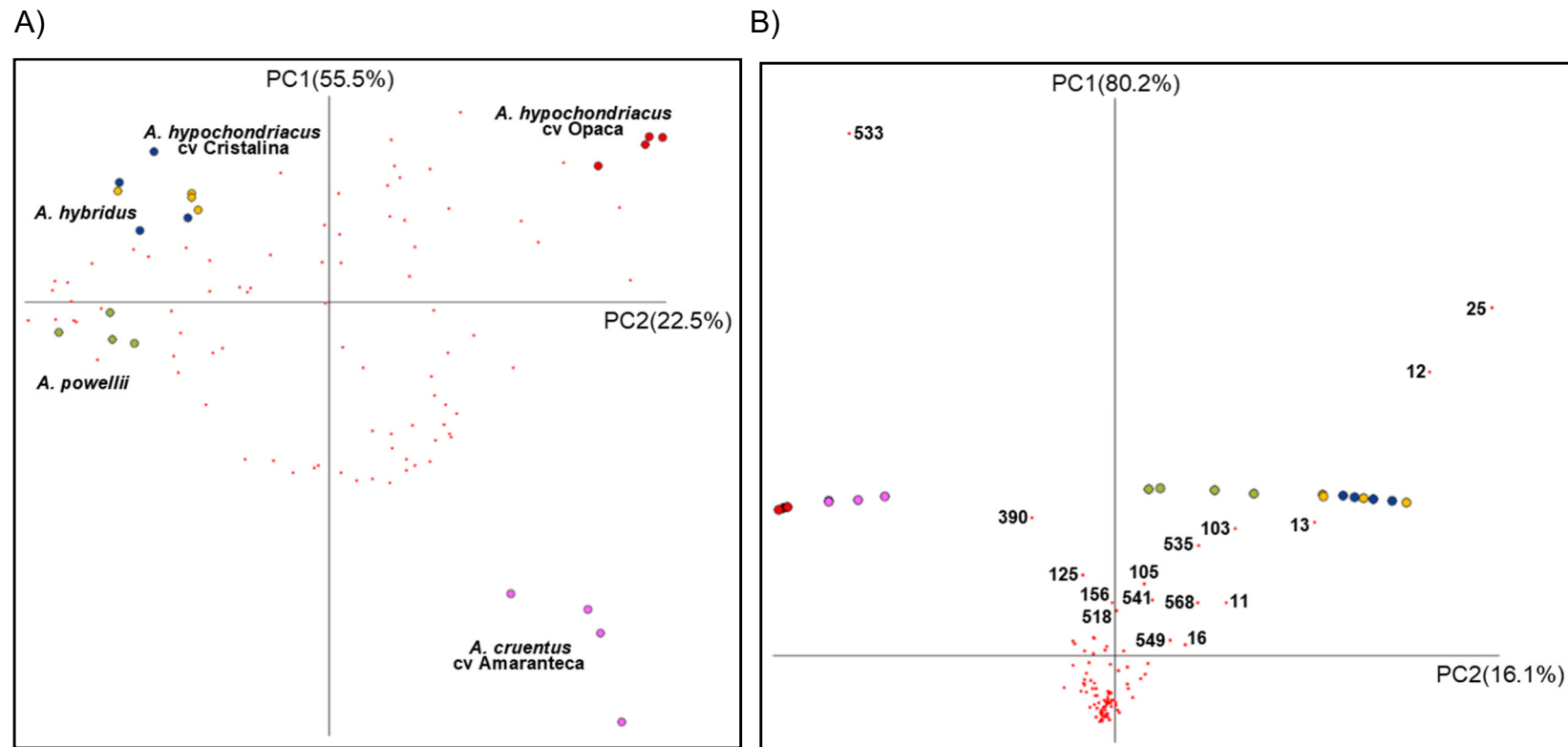


Figure 3.5. PCA constructed with the densitometric data of 120 differentially accumulated spots in the hydrophobic fraction. A) Images as observations, B) Spots as observations. Coloured dots represent gel replicates (Blue, *A. hybridus*; Green, *A. powellii*; Pink, *A. cruentus* cv Amaranteca; Red, *A. hypochondriacus* cv Opaca; Yellow, *A. hypochondriacus* cv Cristalina.) and each red square indicate one spot.

high variability with the wild species, as can be seen in detail in Figure 3.6.

The hydrophobic fraction is characterized by identification of three main proteins, broadly distributed in several spots; 11S globulin (021282) in 38 spots, granule-bound starch synthase I (GBSSI, 011500) in 30 spots, and Vicilin-like seed storage protein (018839) in 13 spots. In the case of 11S globulin and GBSSI, the large number of spots in which they were identified is since they are responsible for the region of high variability this fraction, consisting of 27 differential spots distributed in the form of beads on a string, or “trains” located in the upper part of the alkaline region of the proteomic maps, defined as the High Variation Region (HVR) (Figure 3.6). In the previous analysis of this region by 1D-SDS-PAGE, both proteins were identified in the same bands (Section 2.3.2, Table 2.1, Bands 27, 28 and 29). With the resolution of this fraction through 2-DE, a more detailed analysis of the protein species present in these bands was achieved. In the spots 9, 11, 12, 13, 15, 16, 20, 25, 566 and 568, located in the upper string, only GBSSI was identified (Table 3.2). In spots 518, 527, 533, 536, 537, 541, 543, corresponding to the lower string, 11S globulin was identified as the main protein, only in spots 537 and 541, the less intense spots of the string, GBSSI was also identified.

A group of four low molecular weight HSPs were identified, 17.4 kDa class I (spots 198 and 248), 17.9 kDa class II (spot 197) and 18.3 kDa class I (spot 233), all with higher abundance in *A. cruentus* and two also in *A. hybridus* (spots 197 and 248). OBAP 1A displays variable profiles, is accumulated more in *A. hybridus* in spot 353, in *A. powellii* and *A. cruentus* in spot 354, in *A. hybridus* and *A. cruentus* in spot 355, and have greater levels in *A. hybridus* and *A. hypochondriacus* cultivars in spot 357.

3.3.3 Abundant proteins in amaranth wild species

The proteins described in the previous sections have differential accumulation in at least one species with respect to the others, however, in most cases, the intensity of the spots is shared between wild and someone of the cultivated species. It is necessary to focus on proteins with accumulation levels increased only in wild spe-

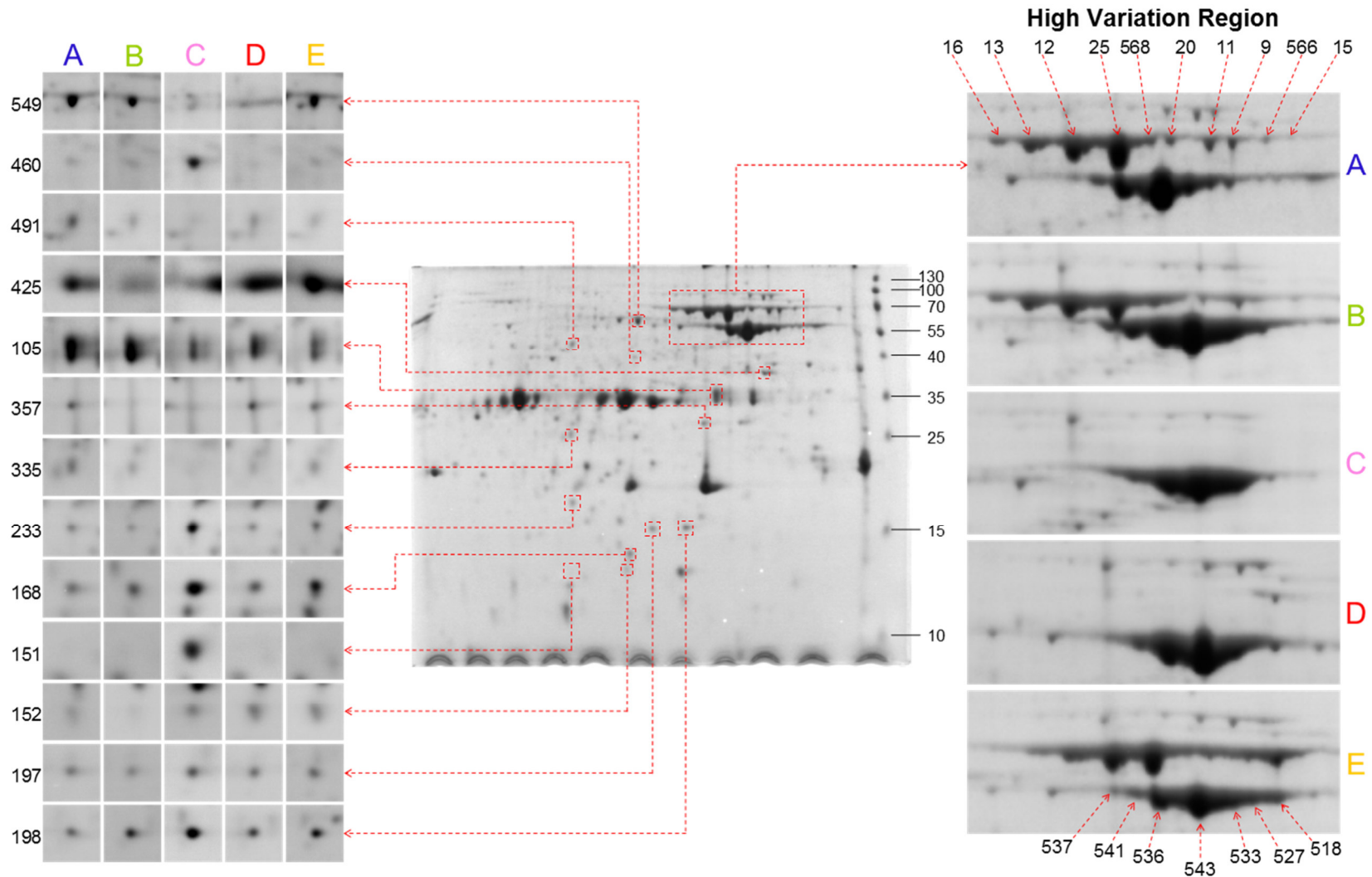


Figure 3.6. Representative proteomic map of the hydrophobic fraction and comparative accumulation profiles of some selected spots. A, *A. hybridus*; B, *A. powellii*; C, *A. cruentus* cv Amaranteca; D, *A. hypochondriacus* cv Opaca; E, *A. hypochondriacus* cv Cristalina.

cies to define what characterizes them from the cultivated ones at this molecular level. Proteins that meet this condition are involved in four biological processes or components; carbohydrate and energy metabolism, cell wall polysaccharides, damage and stress response, and genic regulation.

Concerning to carbohydrate and energy metabolism, in hydrophilic fraction Spots 145, 237 and 292 were identified as glucose and ribitol dehydrogenase, sorbitol dehydrogenase and isocitrate dehydrogenase, respectively, and are up accumulated in both wild species, whereas cytoplasmic fructose-bisphosphate aldolase is elevated in *A. hybridus* hydrophobic fraction (spot 451). Mitochondrial malate dehydrogenase (spot 196), 2,3-bisphosphoglycerate-independent phosphoglycerate mutase (spot 425) and NADP-dependent malic enzyme (spot 436) are increased only in *A. powellii* hydrophilic fraction, and NADH dehydrogenase [ubiquinone] iron-sulphur protein 1 displays higher levels in hydrophobic fraction of *A. hybridus* (spot 52).

Three proteins related with the establishment of cell wall integrity and dynamism were detected; UDP-D-apiose/UDP-D-xylose synthase 2 with high levels in *A. powellii*, hydrophilic fraction spot 301, and Alpha-xylosidase 1 and Bifunctional UDP-glucose 4-epimerase/UDP-xylose 4-epimerase 1, increased in *A. hybridus* hydrophobic fraction in spots 72 and 444, respectively. Regarding to damage and stress response, both species have high levels in proteins Late embryogenesis abundant protein 31 (LEA 31) and annexin-like protein RJ4, the former in spots 130 and 103 of the hydrophilic fraction of *A. hybridus* and *A. powellii*, respectively, and the second in spot 404 of the hydrophobic fraction. Finally, genic regulation related proteins specific of wild species where identified only in hydrophilic fraction, histone H4 (spot 28) and Elongation factor 2 (spot 506) with markedly increased abundance only in *A. powellii*.

3.4 Discussion

The comparative proteomics approach using 2-DE has been widely used for the evaluation of different varieties and plant species, both model, such as *A. thaliana*

(Pang et al., 2010), and crops of agronomic and economic relevance, for example barley (Jin et al., 2014, 2013; Witzel et al., 2010), rice (Jiang et al., 2014) wheat (Liu et al., 2012; Nemati et al., 2019; Pompa et al., 2013; Rocco et al., 2019), and soybean (Gomes et al., 2014; Min et al., 2015; Natarajan, Xu, Bae, Caperna, & Garrett, 2006; Natarajan et al., 2007; Xu et al., 2007). However, although there are reports about the proteome of amaranth seed (Klubíková et al., 2016; Maldonado-Cervantes et al., 2014), these are focused on a descriptive outlook using total protein extracts of only one cultivar of *A. cruentus*, and until now do not exist works that compare the proteomes of wild and cultivated species using differential extraction and 2-DE.

SSPs have been reported as the main variable proteins in the comparison of cultivars. For amaranth species they account for 38% of variations of the differential proteins, given by only three proteins, 11S globulin, Vicilin-like and Vicilin-like At2g18540. These behaviour have been observed in soybean and wheat, where SSPs represent up to 54% and 40% of variations, respectively (Gomes et al., 2014; Rocco et al., 2019). There are an heterogeneous distribution of SSPs, the same protein is more abundant in different spots amongst species, for example in the hydrophobic fraction, 11S globulin is characteristic of *A. hybridus* in spots 103, 494 and 536; *A. powellii* in spot 544; and *A. cruentus* in spots 405, 460, 477, 481 and 501, behaviour that has been observed in *A. thaliana* and soybean (Wan, Ross, Yang, Hegedus, & Kermode, 2007; Xu et al., 2007). This performance is reflected also for other proteins like Seed biotin-containing protein SBP65, OBAP 1A, agglutinin, and 5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase. This may be since homologous versions of the proteins for each species show variations in their amino acid sequence, which have an impact on their *pI* and/or molecular weight, placing them in different positions on the proteomic map. Another alternative may be the presence of post-translational modifications or discrepancy on proteolytic processing, which has been widely described for SSPs (Otegui, Herder, Schulze, Jung, & Staehelin, 2006; Shewry et al., 1995).

3.4.1 11S globulin and GBSSI isoforms

Some relevant features of SSPs and GBSSI has been discussed in sections 2.3 and 2.4, therefore, it is appropriate to focus on the aspects that can be appreciated only by 2-DE, namely, the identification of proteo- or isoforms, and the zone in the proteomic maps of hydrophobic fraction named as the High Variation Region is a clear example of this. The presence of 11S globulin is widely along the ranges of molecular weight and *pI* in the proteomics maps, but in the HVR displays an interesting pattern and their higher intensity together with GBSSI. These beads on a string or “trains” shaped spots in 2-DE profiles are characteristic of changes in *pI* due to post-translational modifications, mainly phosphorylation (Carter, Southwick, Lukov, Willardson, & Thulin, 2004; Deng et al., 2012). Phosphorylation and other charge modifying PTMs, like methylation or acetylation, can play important roles in the regulation of the activity of the proteins in question, for example, in the case of GBSSI the degree of phosphorylation would be a mechanism to modulate the rate of amylose synthesis, and in regard to 11S globulin, may be involved in the stabilization or breakdown of the quaternary structure of the protein, or function as a signalling mechanism for transport and mobilization during germination when it is required to use stored nutrients (Chen et al., 2016; López-Pedrouso, Alonso, & Zapata, 2014; Quiroga et al., 2013; Wan et al., 2007).

3.4.2 Amaranth wild species characteristic proteins

Carbohydrate and energy metabolism

Three dehydrogenases conserve high accumulation levels in both wild species, glucose-ribitol, sorbitol and isocitrate dehydrogenases. Glucose and ribitol dehydrogenase are expressed specifically during barley and lupin embryogenesis, with increased expression at the seed maturation stage, and is induced in carrot somatic embryos when treated with abscisic acid as signal of exposure to abiotic stress (Shiota et al., 2004). Sorbitol dehydrogenase catalyses the NAD⁺ dependent oxidation of sorbitol, ribitol and xylitol leading fructose, ribulose and xylulose, respectively, but may perform the reverse reaction generating sorbitol from fructose

in order to maintain the redox balance during the seed development (Aguayo et al., 2013; de Sousa, Paniago, Arruda, & Yunes, 2008). During the acquisition of desiccation tolerance in seeds, sugars accumulate, maintaining the stability of membranes and functional proteins by replacing the water molecules at the charged surfaces, then these two enzymes might function as a short alcohol-polyol-sugar dehydrogenases, possibly related to carbohydrate metabolism and the acquisition of desiccation tolerance (Shiota et al., 2004). Isocitrate dehydrogenase oxidatively decarboxylate isocitrate to 2-oxoglutarate, which is the carbon skeleton required for ammonia assimilation via the glutamine synthetase/glutamate synthase pathway. An alternative function for isocitrate dehydrogenase is in the production of NADPH to promote redox signalling or homeostasis in response to oxidative stress (Hodges, Flesch, Gálvez, & Bismuth, 2003; Mhamdi et al., 2010).

In addition to these dehydrogenases, some enzymes belonging to the central carbon and energy metabolism were also identified in abundant spots of wild species. Fructose biphosphate aldolase and NADH dehydrogenase [ubiquinone] iron-sulphur protein 1 stood increased only in *A. hybridus*. The former catalyses the reversible aldol cleavage of fructose-1,6-bisphosphate into dihydroxyacetone phosphate and glyceraldehyde-3-phosphate and is involved in glycolysis, gluconeogenesis in the cytoplasm and in the Calvin cycle in plastids (Lv et al., 2017); and the second is the central subunit of the mitochondrial membrane respiratory chain NADH dehydrogenase, also known as NADH:ubiquinone oxidoreductase and usually called Complex I, that is responsible of NADH generated in catabolic pathways coupled to mitochondrial oxidative phosphorylation, generating ATP, and is believed to belong to the minimal assembly required for catalysis (Brandt, 2006; Garmier et al., 2008).

Mitochondrial malate dehydrogenase, 2,3-bisphosphoglycerate-independent phosphoglycerate mutase, and NADP-dependent malic enzyme were characteristic of *A. powellii*. Mitochondrial malate dehydrogenase is a component of the tricarboxylic acid cycle and catalyses the reversible oxidation of malate to oxaloacetate coupled to the reduction of the NAD⁺, supplies malate to pyruvate via NAD⁺malic enzyme to provide CO₂ for fixation in the C₄ photosynthesis pathway

and has been reported that accumulates significantly during *A. thaliana* seed germination (Fu et al., 2005; Sew, Ströher, Fenske, & Millar, 2016). 2,3-bisphosphoglycerate-independent phosphoglycerate mutase catalyses the interconversion of 2- and 3-phosphoglycerate and are essential for glucose metabolism in most organisms (Rigden, Lamani, Mello, Littlejohn, & Jedrzejewski, 2003). NADP-dependent malic enzyme catalyses the oxidative decarboxylation of malate, producing pyruvate, CO₂, and NADPH in the presence of a divalent cation. In C₄ plants, the enzyme plays a specialized role in bundle sheath chloroplasts, where it provides CO₂ for fixation by RuBisCO. Non photosynthetic isoforms have been found in varied tissues of C₃ plants and in plastids and cytosol of C₄ plants, involved in lipid biosynthesis by providing carbon skeletons and reducing power (Drincovich, Casati, & Andreo, 2001; Gerrard Wheeler et al., 2005).

Except for 2,3-bisphosphoglycerate-independent phosphoglycerate mutase, besides to their central role as catalysts, all these enzymes have been related with the achievement of alternative and non-catalytic functions, so they can be considered as moonlight proteins. In *A. thaliana* eight genes coding for fructose biphosphate aldolase have identified, all of them are repressed or overexpressed to a greater or lesser extent in function of different types of abiotic stress (Lu et al., 2012). The absence of mitochondrial malate dehydrogenase in seeds, reduces the reserve accumulation and accelerates the aging process in seeds, causing the decline of seed biomass and viability, and led to germination delays (Sew et al., 2016). In seedlings treated with abscisic acid, NaCl or mannitol is induced the accumulation of NADP-dependent malic enzyme, and the *nadp-me1* mutant seeds displays loss viability earlier and are less sensitive than wild type seeds to abscisic acid-mediated repression of the germination.

The elevated accumulation of carbohydrate and energy metabolism related proteins, and their alternative roles in response of abiotic stress, may reflect some of the adaptation mechanisms against environmental adverse conditions, drought or UV-radiation for example, that seeds of wild amaranths conserve respect to cultivated species, to counteract the damage caused by oxidative stress at which are exposed in nature.

Cell wall integrity and dynamism

UDP-D-apiose/UDP-D-xylose synthase 2 catalyse the formation of UDP-D-apiose, monosaccharide present in the pectic polisaccharide rhamnogalacturonan-II (Mølhøj, Verma, & Reiter, 2003), Alpha-xylosidase 1 specifically released the unsubstituted side chain xylosyl residue attached to the backbone glucosyl residue situated farthest from the reducing end of the, xyloglucan, the main hemicellulosic polysaccharide present in the primary cell walls of dicotyledonous plants, that crosslinks cellulose microfibrils controlling the rate of cell expansion (Sampedro, Sieiro, Revilla, González-Villa, & Zarra, 2001). Bifunctional UDP-glucose 4-epimerase/UDP-xylose 4-epimerase 1, catalyse interconversions between both UDP-glucose and UDP-galactose, and UDP-xylose and UDP-arabinose being a central enzyme for free galactose and arabinose release during the turnover of cell wall polysaccharides (Kotake et al., 2009). These proteins could be related to the presence of a thicker testa in wild species and the need for higher enzymatic activity to break it or redirect carbohydrates during germination.

Damage and stress response

As mentioned in section 2.4, LEA proteins play an important role in resistance to abiotic stress in a variety of both prokaryotic and eukaryotic organisms, mainly against desiccation and water stress (Hundertmark & Hincha, 2008). LEA 31 was found remarkably accumulated in *A. powellii*. In *A. thaliana*, this protein is encoded by the gene At3g22490 termed as *Atrab28* (Responsive to Abscisic Acid 28), which overexpression confers faster germination rate than wild type, under control and stress conditions by salts or other osmolytes (Borrell et al., 2002). Like the ortholog of *A. thaliana*, LEA 31 of amaranth have three seed maturation protein motifs at primary structure level, that could imply that has a similar role during the seed development and germination.

Annexins are cytosolic proteins involved in signal transduction pathways that can attach or insert into plasma- or endo-membranes depending on cytosolic free Ca^{2+} concentration, pH, lipid composition or membrane voltage. In plants have been found in almost all types of tissues including seeds, roots, stems and leaves, and

are implicated in a wide range of processes as exocytosis, cell elongation, wall synthesis, nodulation, and fruit ripening, but mainly with Ca^{2+} and reactive oxygen species homeostasis due to its inference against abiotic stress conditions, for example cold, oxidative, saline, and abscisic acid response (Laohavisit & Davies, 2009; Mortimer et al., 2008). The annexin found here have greater levels of accumulation in both wild amaranth species and was identified in the hydrophobic fraction (spot 404), which can imply that, in seeds, this protein is linked to membranes. In seven days-old plants, the transcript level of the *A. thaliana* orthologue for this protein (AnnAt8) was reported to have an increase of 175 and 434 times in dehydration and exposure to 250 mM NaCl, respectively, these stress conditions are analogous to those of a seed due to its low water content, then, amaranth annexin might perform some protective role until germination begins (Cantero et al., 2006).

Genic regulation

Nucleosomes, the structural unit of chromatin packing, are formed of octamers of core histones (H2A, H2B, H3 and H4) around which DNA is wrapped. Chromatin tight and relaxed states are largely mediated by post-translational modifications that take place at the amino-terminal, in the “tails”, of core histones. Methylation, acetylation, phosphorylation and ubiquitination are the more extensively studied PTMs, crucial for the regulation of diverse cellular functions like DNA replication and repair, transcription, and gene expression (Peterson & Laniel, 2004; Zhang, Chen, Zhang, & Zhao, 2009). It would be expected that if a change in one of the histones is observed, a variation of the same magnitude should be reflected in the rest of them, as reported for their transcript levels up-regulation under drought conditions (Huang, Wu, Abrams, & Cutler, 2008), but interestingly, only one of the four core histones, histone H4, was detected differentially accumulated, with higher abundance in *A. powellii*. Histones are small and highly basic proteins with isoelectric points of 10-12, so they should be left out of the 2-DE analysis since IPG strips with a range of 5-8 were used in this work, however, the experimental isoelectric point observed for histone H4 was of 5.1, probably due to the presence of PTMs that

impact on this property. Nevertheless, despite the technical limitations of 2-DE, it should be noted that only histone H4 from *A. powellii* seeds has a PTMs profile that confers an acidic isoelectric point to this protein and could be related with a specific way of epigenetic regulation.

3.5 Conclusions

The polarity-based fractionation approach of amaranth seed proteins allows to generate proteomic maps with an adequate definition of spots, since it was possible to eliminate the masking effect generated by the presence of abundant proteins that normally results when working with total extracts. The heterogeneous distribution of storage proteins may indicate the high gene variation of sequences coding for 7S and 11S globulins between amaranth species, as well as a possible marked differential execution of gene regulation processes by proteolysis or incorporation of post-translational modifications during seed development or germination, this is also supported by the profiles observed in the HVR at the hydrophobic fraction, which are characteristic for the presence of phosphorylations. Another interesting aspect observed in this region is that the intensity of the spots in which GBSSI was identified is comparable with that of the spots for 11S globulin; without a doubt this is due to the high demand for starch synthesis during the development of the seed, however, the presence of this enzyme at levels as higher as those of the most abundant storage protein in mature seeds, suggests that in addition to its role in amylose synthesis, it can serve as a reserve or regulation protein during germination, and therefore being a moonlight protein. Over accumulated proteins exclusively in wild species had not previously been reported in studies comparing cultivated or domesticated and wild lines in other crops, however, the fact that they have been identified in works related with abiotic stress resistance indicates that they can represent key points in the regulation of important agronomic characteristics such as seed viability or longevity. The following investigations should focus on the identification and monitoring of post-translational modifications, as well on the evaluation of the proteins that have arisen as possibly responsible for conferring resistance to wild species.

General considerations and prospective

Amaranth is an interesting biological system in both scientific and agronomic fields, since it has outstanding characteristics, such as its ability to develop in various conditions of abiotic stress, rapid nutrient assimilation, and production of grains of high nutritional quality for human consumption. Despite being a well-established crop in various countries, our understanding about molecular characteristics of amaranth is limited. This work contributes to two main aspects of the knowledge of amaranth biochemistry. First, it establishes the scientific support on the importance of considering the germplasm of wild species for the improvement and obtaining of new varieties depending on the nutritional profile that is required, high protein content, a starch composition enriched in amylose or amylopectin, or favour the levels of certain lipids. In the second instance, with the 1-D-SDS-PAGE analysis and the establishment of proteomic maps using 2-DE, proteins of agrobiotechnological and scientific interest were unveiled. On the one hand, we have those related to resistance to abiotic stress and maintenance of cell homeostasis, represented by LEA proteins and annexins, on the other hand, there are Oleosins and OBAPs, linked to the transport and proper storage of lipids. GBSSI and 11SHMW globulin stand out, both for a large number of spots on which they were identified and for their characteristic profiles in 2-DE, which suggest post-translational processing by phosphorylation, and in the case of GBSSI for the presence of incomplete versions of the enzyme, which, being non-functional, results in the synthesis of a starch lacking amylose. It is also important to mention the identification of other 7S and 11S globulin paralogs, which may be related to other functions beyond their canonical role as reserve proteins. Therefore, now that differential proteins between wild and cultivated amaranth species have been established, the next step must be to characterize their functionality through *in vitro* and *in vivo* experiments. The following work should focus on the cloning, expression, purification and biochemical, biophysical and structural characterization of the proteins of interest, especially LEAs, GBSSI and 7S and 11S globulins, as well as in the post-translational modifications that regulate their functions, establish their molecular targets by

protein-protein interaction assays, and elucidate their signalling pathways. Another interesting alternative that must be considered is the silencing of the genes of interest, as well as their expression in model systems to evaluate how it impacts the metabolism and plant physiology. In addition, although the establishment of gel-based proteomes has allowed visualizing differences between seeds of the studied amaranth species, with the intention of broadening this panorama, we cannot rule out to carry out shotgun proteomics by multi-dimensional protein identification technologies.

Annexes

Appendix 1. Scientific production

Publications generated from this work

1. **Bojórquez-Velázquez E**, Velarde-Salcedo AJ, De León-Rodríguez A, Jiménez-Islas H, Pérez-Torres JL, Herrera-Estrella A, Espitia-Rangel E, Barba de la Rosa AP. 2018. Morphological, proximal composition, and bioactive compounds characterization of wild and cultivated amaranth (*Amaranthus* spp.) species. Journal of Cereal Science. 83:222-228. DOI: <https://doi.org/10.1016/j.jcs.2018.09.004>.

2. **Bojórquez-Velázquez E**, Barrera-Pacheco A, Espitia-Rangel E, Herrera-Estrella A, Barba de la Rosa AP. 2019. Protein analysis reveals differential accumulation of late embryogenesis abundant and storage proteins in seeds of wild and cultivated amaranth species. BMC Plant Biology. 19:59. DOI: <https://doi.org/10.1186/s12870-019-1656-7>.

Other contributions

3. Saucedo AL, Hernández-Domínguez EE, de Luna-Valdez LA, Guevara-García AA, Escobedo-Moratilla A, **Bojórquez-Velázquez E**, del Río-Portilla F, Fernández-Velasco DA and Barba de la Rosa AP. 2017. Insights on structure and function of a late embryogenesis abundant protein from *Amaranthus cruentus*: an intrinsically disordered protein involved in protection against desiccation, oxidant conditions, and osmotic stress. 2017. Frontiers in Plant Science. 8:497. <https://doi.org/10.3389/fpls.2017.00497>.

4. Hernández-Domínguez EE, Vargas-Ortiz E, **Bojórquez-Velázquez E**, Barrera-Pacheco A, Santos-Díaz MS, Camarena-Rangel NG, Barba de la Rosa AP. 2019. Molecular characterization and in vitro interaction analysis of Op14-3-3 μ protein from *Opuntia ficus-indica*: identification of a new client protein from shikimate pathway. Journal of Proteomics 198:151-162. DOI: <https://doi.org/10.1016/j.jprot.2019.01.013>.

5. Velarde-Salcedo AJ, **Bojórquez-Velázquez E**, Barba de la Rosa AP. 2019. Amaranth. In Whole Grains and their Bioactives: Composition and Health, First Edition. Jodee Johnson and Taylor Wallace, Editors. John Wiley & Sons Ltd. DOI: <https://doi.org/10.1002/9781119129486.ch8>.

References

- Aguayo, M. F., Ampuero, D., Mandujano, P., Parada, R., Muñoz, R., Gallart, M., ... Handford, M. (2013). Sorbitol dehydrogenase is a cytosolic protein required for sorbitol metabolism in *Arabidopsis thaliana*. *Plant Science*, 205–206, 63–75. <https://doi.org/10.1016/j.plantsci.2013.01.012>
- Aguilar-Hernández, H. S., Santos, L., León-Galván, F., Barrera-Pacheco, A., Espitia-Rangel, E., De León-Rodríguez, A., ... Barba de la Rosa, A. P. (2011). Identification of calcium stress induced genes in amaranth leaves through suppression subtractive hybridization. *Journal of Plant Physiology*, 168(17), 2102–2109. <https://doi.org/10.1016/j.jplph.2011.06.006>
- Ahuja, G., Jaiswal, S., Hucl, P., & Chibbar, R. N. (2014). Wheat genome specific granule-bound starch synthase i differentially influence grain starch synthesis. *Carbohydrate Polymers*, 114, 87–94. <https://doi.org/10.1016/j.carbpol.2014.08.004>
- Alvarez-Jubete, L., Wijngaard, H., Arendt, E. K., & Gallagher, E. (2010). Polyphenol composition and in vitro antioxidant activity of amaranth, quinoa buckwheat and wheat as affected by sprouting and baking. *Food Chemistry*, 119(2), 770–778. <https://doi.org/10.1016/j.foodchem.2009.07.032>
- Aslam, B., Basit, M., Nisar, M. A., Khurshid, M., & Rasool, M. H. (2017). Proteomics: Technologies and Their Applications. *Journal of Chromatographic Science*, 55(2), 182–196. <https://doi.org/10.1093/chromsci/bmw167>
- Assad, R., Reshi, Z. A., Jan, S., & Rashid, I. (2017). *Biology of Amaranths*. *Botanical Review* (Vol. 83). The Botanical Review. <https://doi.org/10.1007/s12229-017-9194-1>
- Barba de la Rosa, A. P., Barba Montoya, A., Martínez-Cuevas, P., Hernández-Ledesma, B., León-Galván, M. F., De León-Rodríguez, A., & González, C. (2010). Tryptic amaranth glutelin digests induce endothelial nitric oxide production through inhibition of ACE: Antihypertensive role of amaranth peptides. *Nitric Oxide - Biology and Chemistry*, 23(2), 106–111. <https://doi.org/10.1016/j.niox.2010.04.006>
- Barba de la Rosa, A. P., Fomsgaard, I. S., Laursen, B., Mortensen, A. G., Olvera-Martínez, L., Silva-Sánchez, C., ... De León-Rodríguez, A. (2009). Amaranth (*Amaranthus hypochondriacus*) as an alternative crop for sustainable food production: Phenolic acids and flavonoids with potential impact on its nutraceutical quality. *Journal of Cereal Science*, 49(1), 117–121. <https://doi.org/10.1016/j.jcs.2008.07.012>
- Barba de la Rosa, A. P., Gueguen, J., Paredes-López, O., & Viroben, G. (1992). Fractionation Procedures, Electrophoretic Characterization, and Amino Acid Composition of Amaranth Seed Proteins. *Journal of Agricultural and Food Chemistry*, 40(6), 931–936. <https://doi.org/10.1021/jf00018a002>
- Barba De La Rosa, A. P., Herrera-Estrella, A., Utsumi, S., & Paredes-López, O. (1996). Molecular characterization, cloning and structural analysis of a cDNA encoding an amaranth globulin. *Journal of Plant Physiology*, 149(5), 527–532. [https://doi.org/10.1016/S0176-1617\(96\)80329-4](https://doi.org/10.1016/S0176-1617(96)80329-4)
- Bojórquez-Velázquez, E., Velarde-Salcedo, A. J., De León-Rodríguez, A., Jimenez-

- Islas, H., Pérez-Torres, J. L., Herrera-Estrella, A., ... Barba de la Rosa, A. P. (2018). Morphological, proximal composition, and bioactive compounds characterization of wild and cultivated amaranth (*Amaranthus* spp.) species. *Journal of Cereal Science*. <https://doi.org/10.1016/j.jcs.2018.09.004>
- Borrell, A., Cutanda, M. C., Lumbreras, V., Pujal, J., Goday, A., Culiáñez-Macià, F. A., & Pagès, M. (2002). *Arabidopsis thaliana* Atrab28: A nuclear targeted protein related to germination and toxic cation tolerance. *Plant Molecular Biology*, *50*(2), 249–259. <https://doi.org/https://doi.org/10.1023/A:1016084615014>
- Brandt, U. (2006). Energy Converting NADH: Quinone oxidoreductase (Complex I). *Annual Review of Biochemistry*, *75*(1), 69–92. <https://doi.org/10.1146/annurev.biochem.75.103004.142539>
- Bressani, R., & García-Vela, L. A. (1990). Protein fractions in amaranth grain and their chemical characterization. *Journal of Agricultural and Food Chemistry*, *38*(5), 1205–1209. <https://doi.org/10.1021/jf00095a010>
- Cantero, A., Barthakur, S., Bushart, T. J., Chou, S., Morgan, R. O., Fernandez, M. P., ... Roux, S. J. (2006). Expression profiling of the *Arabidopsis* annexin gene family during germination, de-etiolation and abiotic stress. *Plant Physiology and Biochemistry*, *44*(1), 13–24. <https://doi.org/10.1016/j.plaphy.2006.02.002>
- Carter, M. D., Southwick, K., Lukov, G., Willardson, B. M., & Thulin, C. D. (2004). Identification of phosphorylation sites on phosphocyanin-like protein by QTOF mass spectrometry. *Journal of Biomolecular Techniques*, *15*(4), 257–264. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2291704/pdf/0150257.pdf>
- Caselato-Sousa, V. M., Ozaki, M. R., De Almeida, E. A., & Amaya-Farfan, J. (2014). Intake of heat-expanded amaranth grain reverses endothelial dysfunction in hypercholesterolemic rabbits. *Food and Function*, *5*(12), 3281–3286. <https://doi.org/10.1039/c4fo00468j>
- Chaturvedi, A., Sarojini, G., & Devi, N. L. (1993). Hypocholesterolemic effect of amaranth seeds (*Amaranthus esculantus*). *Plant Foods for Human Nutrition*, *44*(1), 63–70. <https://doi.org/10.1007/BF01088483>
- Chen, G.-X., Zhou, J.-W., Liu, Y.-L., Lu, X.-B., Han, C.-X., Zhang, W.-Y., ... Yan, Y.-M. (2016). Biosynthesis and regulation of wheat amylose and amylopectin from proteomic and phosphoproteomic characterization of granule-binding proteins. *Scientific Reports*, *6*(1), 33111. <https://doi.org/10.1038/srep33111>
- Cho, W. C. S. (2007). Proteomics technologies and challenges. *Genomics, Proteomics & Bioinformatics*, *5*(2), 77–85. [https://doi.org/10.1016/S1672-0229\(07\)60018-7](https://doi.org/10.1016/S1672-0229(07)60018-7)
- Clouse, J. W., Adhikary, D., Page, J. T., Ramaraj, T., Deyholos, M. K., Udall, J. A., ... Maughan, P. J. (2016). The amaranth genome: Genome, transcriptome, and physical map assembly. *The Plant Genome*, *9*(1), 0. <https://doi.org/10.3835/plantgenome2015.07.0062>
- Cooper, R. (2015). Re-discovering ancient wheat varieties as functional foods. *Journal of Traditional and Complementary Medicine*, *5*(3), 138–143. <https://doi.org/10.1016/j.jtcme.2015.02.004>
- Crooks, G., Hon, G., Chandonia, J.-M., & Brenner, S. (2004). WebLogo: A sequence logo generator. *Genome Research*, *14*, 1188–1190. <https://doi.org/10.1101/gr.849004.1>
- D'Amico, S., & Schoenlechner, R. (2017). Amaranth: Its unique nutritional and

- health-promoting attributes. *Gluten-Free Ancient Grains*, 131–159. <https://doi.org/10.1016/B978-0-08-100866-9.00006-6>
- Davies, M. J. (2005). The oxidative environment and protein damage. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1703(2), 93–109. <https://doi.org/10.1016/j.bbapap.2004.08.007>
- de Sousa, S. M., Paniago, M. D. G., Arruda, P., & Yunes, J. A. (2008). Sugar levels modulate sorbitol dehydrogenase expression in maize. *Plant Molecular Biology*, 68(3), 203–213. <https://doi.org/10.1007/s11103-008-9362-0>
- Deng, X., Hahne, T., Schröder, S., Redweik, S., Nebija, D., Schmidt, H., ... Wätzig, H. (2012). The challenge to quantify proteins with charge trains due to isoforms or conformers. *ELECTROPHORESIS*, 33(2), 263–269. <https://doi.org/10.1002/elps.201100321>
- Di Silvestre, D., Bergamaschi, A., Bellini, E., & Mauri, P. (2018). Large scale proteomic data and network-based systems biology approaches to explore the plant world. *Proteomes*, 6(2), 27. <https://doi.org/10.3390/proteomes6020027>
- Dong, Y., Fang, X., Yang, Y., Xue, G.-P., Chen, X., Zhang, W., ... Chen, J. (2017). Comparative proteomic analysis of susceptible and resistant rice plants during early infestation by small brown planthopper. *Frontiers in Plant Science*, 8(October), 1–14. <https://doi.org/10.3389/fpls.2017.01744>
- Drincovich, M. F., Casati, P., & Andreo, C. S. (2001). NADP-malic enzyme from plants: a ubiquitous enzyme involved in different metabolic pathways. *FEBS Letters*, 490(1–2), 1–6. [https://doi.org/10.1016/S0014-5793\(00\)02331-0](https://doi.org/10.1016/S0014-5793(00)02331-0)
- Dunwell, J. M., Khuri, S., & Gane, P. J. (2000). Microbial relatives of the seed storage proteins of higher plants: Conservation of structure and diversification of function during evolution of the cupin superfamily. *Microbiology and Molecular Biology Reviews*, 64(1), 153–179. <https://doi.org/10.1128/MMBR.64.1.153-179.2000>
- Džunková, M., Janovská, D., Čepková, P. H., Prohasková, A., & Kolář, M. (2011). Glutelin protein fraction as a tool for clear identification of Amaranth accessions. *Journal of Cereal Science*, 53(2), 198–205. <https://doi.org/10.1016/j.jcs.2010.12.003>
- Eldakak, M., Milad, S. I. M., Nawar, A. I., & Rohila, J. S. (2013). Proteomics: A biotechnology tool for crop improvement. *Frontiers in Plant Science*, 4(February), 1–12. <https://doi.org/10.3389/fpls.2013.00035>
- Espitia-Rangel, E. (Ed.). (2012). *Amaranto: Ciencia y Tecnología. Libro Científico No. 2. Amaranto: Ciencia y Tecnología* (1st ed.). México: INIFAP/SINAREFI.
- Espitia-Rangel, E., Mapes-Sánchez, E. C., Nuñez-Colín, C. A., & Escobedo-López, D. (2010). Geographical distribution of cultivated species of *Amaranthus*. *Revista Mexicana de Ciencias Agrícolas*, 1, 427–437.
- Fernández-Marín, B., Milla, R., Martín-Robles, N., Arc, E., Kranner, I., Becerril, J. M., & García-Plazaola, J. I. (2014). Side-effects of domestication: cultivated legume seeds contain similar tocopherols and fatty acids but less carotenoids than their wild counterparts. *BMC Plant Biology*, 14(1), 1599. <https://doi.org/10.1186/s12870-014-0385-1>
- Finch-Savage, W. E., & Bassel, G. W. (2016). Seed vigour and crop establishment: extending performance beyond adaptation. *Journal of Experimental Botany*, 67(3), 567–591. <https://doi.org/10.1093/jxb/erv490>

- Finn, R. D., Attwood, T. K., Babbitt, P. C., Bateman, A., Bork, P., Bridge, A. J., ... Mitchell, A. L. (2017). InterPro in 2017-beyond protein family and domain annotations. *Nucleic Acids Research*, *45*(D1), D190–D199. <https://doi.org/10.1093/nar/gkw1107>
- Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., ... Bateman, A. (2016). The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Research*, *44*(D1), D279–D285. <https://doi.org/10.1093/nar/gkv1344>
- Frandsen, G. I., Mundy, J., & Tzen, J. T. C. (2001). Oil bodies and their associated proteins, oleosin and caleosin. *Physiologia Plantarum*, *112*(3), 301–307. <https://doi.org/10.1034/j.1399-3054.2001.1120301.x>
- Franz, G., Hatzopoulos, P., Jones, T. J., Krauss, M., & Sung, Z. R. (1989). Molecular and genetic analysis of an embryonic gene, DC 8, from *Daucus carota* L. *MGG Molecular & General Genetics*, *218*(1), 143–151. <https://doi.org/10.1007/BF00330577>
- Fu, Q., Wang, B.-C., Jin, X., Li, H.-B., Han, P., Wei, K., ... Zhu, Y. (2005). Proteomic analysis and extensive protein identification from dry, germinating *Arabidopsis* seeds and young seedlings. *BMB Reports*, *38*(6), 650–660. <https://doi.org/10.5483/BMBRep.2005.38.6.650>
- Garnier, M., Carroll, A. J., Delannoy, E., Vallet, C., Day, D. A., Small, I. D., & Millar, A. H. (2008). Complex I dysfunction redirects cellular and mitochondrial metabolism in *Arabidopsis*. *Plant Physiology*, *148*(3), 1324–1341. <https://doi.org/10.1104/pp.108.125880>
- Gerrard Wheeler, M. C., Drincovich, F. M., Andreo, C. S., Flugge, U., Tronconi, M. a, & Maurino, V. G. (2005). A comprehensive analysis of the NADP-Malic enzyme gene family of *Arabidopsis*. *Plant Physiology*, *139*(September), 39–51. <https://doi.org/10.1104/pp.105.065953.1>
- Gomes, L. S., Senna, R., Sandim, V., Silva-Neto, M. A. C., Perales, J. E. A., Zingali, R. B., ... Fialho, E. (2014). Four conventional soybean [*Glycine max* (L.) Merrill] seeds exhibit different protein profiles as revealed by proteomic analysis. *Journal of Agricultural and Food Chemistry*, *62*(6), 1283–1293. <https://doi.org/10.1021/jf404351g>
- Gong, C. Y., & Wang, T. (2013). Proteomic evaluation of genetically modified crops: current status and challenges. *Frontiers in Plant Science*, *4*(March), 1–8. <https://doi.org/10.3389/fpls.2013.00041>
- Gong, F., Hu, X., & Wang, W. (2015). Proteomic analysis of crop plants under abiotic stress conditions: where to focus our research? *Frontiers in Plant Science*, *6*(June), 1–5. <https://doi.org/10.3389/fpls.2015.00418>
- Gonzalez-Gonzalez, C. R., Tuohy, K. M., & Jauregi, P. (2011). Production of angiotensin-I-converting enzyme (ACE) inhibitory activity in milk fermented with probiotic strains: Effects of calcium, pH and peptides on the ACE-inhibitory activity. *International Dairy Journal*, *21*(9), 615–622. <https://doi.org/10.1016/j.idairyj.2011.04.001>
- Gubbuk, H., Kafkas, E., Guven, D., & Gunes, E. (2010). Physical and phytochemical profile of wild and domesticated carob (*Ceratonia siliqua* L.) genotypes, *8*(4), 1129–1136.
- He, H. P., & Corke, H. (2003). Oil and squalene in *Amaranthus* grain and leaf.

- Journal of Agricultural and Food Chemistry*, 51(27), 7913–7920. <https://doi.org/10.1021/jf030489q>
- Herrera-Díaz, J., Jelezova, M. K., Cruz-García, F., & Dinkova, T. D. (2018). Protein Disulfide Isomerase (PDI1-1) differential expression and modification in Mexican malting barley cultivars. *PLOS ONE*, 13(11), e0206470. <https://doi.org/10.1371/journal.pone.0206470>
- Hodges, M., Flesch, V., Gálvez, S., & Bismuth, E. (2003). Higher plant NADP+-dependent isocitrate dehydrogenases, ammonium assimilation and NADPH production. *Plant Physiology and Biochemistry*, 41(6–7), 577–585. [https://doi.org/10.1016/S0981-9428\(03\)00062-7](https://doi.org/10.1016/S0981-9428(03)00062-7)
- Hood, L. E., Omenn, G. S., Moritz, R. L., Aebersold, R., Yamamoto, K. R., Amos, M., ... Locascio, L. (2012). New and improved proteomics technologies for understanding complex biological systems: Addressing a grand challenge in the life sciences. *PROTEOMICS*, 12(18), 2773–2783. <https://doi.org/10.1002/pmic.201270086>
- Hu, J., Rampitsch, C., & Bykova, N. V. (2015). Advances in plant proteomics toward improvement of crop productivity and stress resistance. *Frontiers in Plant Science*, 6(APR), 1–15. <https://doi.org/10.3389/fpls.2015.00209>
- Huang, D., Wu, W., Abrams, S. R., & Cutler, A. J. (2008). The relationship of drought-related gene expression in *Arabidopsis thaliana* to hormonal and environmental factors. *Journal of Experimental Botany*, 59(11), 2991–3007. <https://doi.org/10.1093/jxb/ern155>
- Huerta-Ocampo, J. Á., & Barba de la Rosa, A. P. (2011). Amaranth: A pseudo-cereal with nutraceutical properties. *Current Nutrition & Food Science*, 7, 1–9. <https://doi.org/10.2174/157340111794941076>
- Huerta-Ocampo, J. Á., Barrera-Pacheco, A., Mendoza-Hernández, C. S., Espitia-Rangel, E., Mock, H. P., & Barba de la Rosa, A. P. (2014). Salt stress-induced alterations in the root proteome of *Amaranthus cruentus* L. *Journal of Proteome Research*, 13(8), 3607–3627. <https://doi.org/10.1021/pr500153m>
- Hulo, N., Bairoch, A., Bulliard, V., Cerutti, L., Cuče, B. A., De castro, E., ... Sigrist, C. J. A. (2008). The 20 years of PROSITE. *Nucleic Acids Research*, 36(SUPPL. 1), 245–249. <https://doi.org/10.1093/nar/gkm977>
- Hundertmark, M., Buitink, J., Leprince, O., & Hinch, D. K. (2011). The reduction of seed-specific dehydrins reduces seed longevity in *Arabidopsis thaliana*. *Seed Science Research*, 21(3), 165–173. <https://doi.org/10.1017/S0960258511000079>
- Hundertmark, M., & Hinch, D. K. (2008). LEA (Late Embryogenesis Abundant) proteins and their encoding genes in *Arabidopsis thaliana*. *BMC Genomics*, 9, 1–22. <https://doi.org/10.1186/1471-2164-9-118>
- Hunt, H. V., Denyer, K., Packman, L. C., Jones, M. K., & Howe, C. J. (2010). Molecular basis of the waxy endosperm starch phenotype in broomcorn millet (*Panicum miliaceum* L.). *Molecular Biology and Evolution*, 27(7), 1478–1494. <https://doi.org/10.1093/molbev/msq040>
- Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J., & Mann, M. (2005). Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Molecular & Cellular Proteomics: MCP*, 4(9), 1265–1272.

- <https://doi.org/10.1074/mcp.M500061-MCP200>
- Janssen, F., Pauly, A., Rombouts, I., Jansens, K. J. A., Deleu, L. J., & Delcour, J. A. (2017). Proteins of amaranth (*Amaranthus* spp.), buckwheat (*Fagopyrum* spp.), and quinoa (*Chenopodium* spp.): A food science and technology perspective. *Comprehensive Reviews in Food Science and Food Safety*, 16(1), 39–58. <https://doi.org/10.1111/1541-4337.12240>
- Jiang, C., Cheng, Z., Zhang, C., Yu, T., Zhong, Q., Shen, J. Q., & Huang, X. (2014). Proteomic analysis of seed storage proteins in wild rice species of the *Oryza* genus. *Proteome Science*, 12(1). <https://doi.org/10.1186/s12953-014-0051-4>
- Jin, Z., Cai, G. L., Li, X. M., Gao, F., Yang, J. J., Lu, J., & Dong, J. J. (2014). Comparative proteomic analysis of green malts between barley (*Hordeum vulgare*) cultivars. *Food Chemistry*, 151, 266–270. <https://doi.org/10.1016/j.foodchem.2013.11.065>
- Jin, Z., Li, X. M., Gao, F., Sun, J. Y., Mu, Y. W., & Lu, J. (2013). Proteomic analysis of differences in barley (*Hordeum vulgare*) malts with distinct filterability by DIGE. *Journal of Proteomics*, 93, 93–106. <https://doi.org/10.1016/j.jprot.2013.05.038>
- Klubíková, K., Szabová, M., Skultety, L., Libiaková, G., & Hricová, A. (2016). Revealing the seed proteome of the health benefitting grain amaranth (*Amaranthus cruentus* L.). *Chemical Papers*, 70(10), 1322–1335. <https://doi.org/10.1515/chempap-2016-0065>
- Komatsu, S. (2008). Crop Proteomics and Its Application to Biotechnology. *Journal of Proteome Research*, 7(6), 2183–2183. <https://doi.org/10.1021/pr0837688>
- Komatsu, S., Mock, H.-P., Yang, P., & Svensson, B. (2013). Application of proteomics for improving crop protection/artificial regulation. *Frontiers in Plant Science*, 4(DEC), 522. <https://doi.org/10.3389/fpls.2013.00522>
- Kong, X., Bao, J., & Corke, H. (2009). Physical properties of *Amaranthus* starch. *Food Chemistry*, 113(2), 371–376. <https://doi.org/10.1016/j.foodchem.2008.06.028>
- Kosová, K., Vítámvás, P., Urban, M. O., Prášil, I. T., & Renaut, J. (2018). Plant abiotic stress proteomics: The major factors determining alterations in cellular proteome. *Frontiers in Plant Science*, 9(February), 1–22. <https://doi.org/10.3389/fpls.2018.00122>
- Kotake, T., Takata, R., Verma, R., Takaba, M., Yamaguchi, D., Orita, T., ... Tsumuraya, Y. (2009). Bifunctional cytosolic UDP-glucose 4-epimerases catalyse the interconversion between UDP-D-xylose and UDP-L-arabinose in plants. *Biochemical Journal*, 424(2), 169–177. <https://doi.org/10.1042/BJ20091025>
- Kumar, P., Kesari, P., Dhindwal, S., Choudhary, A. K., Katiki, M., Neetu, ... Kumar, P. (2017). A novel function for globulin in sequestering plant hormone: Crystal structure of *Wrightia tinctoria* 11S globulin in complex with auxin. *Scientific Reports*, 7(1), 4705. <https://doi.org/10.1038/s41598-017-04518-7>
- Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, 33(7), 1870–1874. <https://doi.org/10.1093/molbev/msw054>
- Laohavisit, A., & Davies, J. M. (2009). Multifunctional annexins. *Plant Science*, 177(6), 532–539. <https://doi.org/10.1016/j.plantsci.2009.09.008>

- León-Villanueva, A., Huerta-Ocampo, J. A., Barrera-Pacheco, A., Medina-Godoy, S., & Barba de la Rosa, A. P. (2018). Proteomic analysis of non-toxic *Jatropha curcas* byproduct cake: Fractionation and identification of the major components. *Industrial Crops and Products*, 111(October 2017), 694–704. <https://doi.org/10.1016/j.indcrop.2017.11.046>
- Leprince, O., Pellizzaro, A., Berriri, S., & Buitink, J. (2016). Late seed maturation: drying without dying. *Journal of Experimental Botany*, 68(4), erw363. <https://doi.org/10.1093/jxb/erw363>
- Letunic, I., & Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research*, 44(W1), W242–W245. <https://doi.org/10.1093/nar/gkw290>
- Letunic, I., & Bork, P. (2018). 20 years of the SMART protein domain annotation resource. *Nucleic Acids Research*, 46(D1), D493–D496. <https://doi.org/10.1093/nar/gkx922>
- Lev-Yadun, S. (2016). Seed Camouflage. In S. Lev-Yadun (Ed.), *Defensive (anti-herbivory) coloration in land plants* (pp. 41–49). Springer International Publishing. <https://doi.org/10.1007/978-3-319-42096-7>
- Li, C., & Zhang, Y.-M. (2011). Molecular evolution of glycinin and β -conglycinin gene families in soybean (*Glycine max* L. Merr.). *Heredity*, 106(4), 633–641. <https://doi.org/10.1038/hdy.2010.97>
- Liu, H., Wang, C., Komatsu, S., He, M., Liu, G., & Shen, S. (2013). Proteomic analysis of the seed development in *Jatropha curcas*: From carbon flux to the lipid accumulation. *Journal of Proteomics*, 91, 23–40. <https://doi.org/10.1016/j.jprot.2013.06.030>
- Liu, W., Zhang, Y., Gao, X., Wang, K., Wang, S., Zhang, Y., ... Yan, Y. (2012). Comparative proteome analysis of glutenin synthesis and accumulation in developing grains between superior and poor quality bread wheat cultivars. *Journal of the Science of Food and Agriculture*, 92(1), 106–115. <https://doi.org/10.1002/jsfa.4548>
- Lobell, D. B., Schlenker, W., & Costa-Roberts, J. (2011). Climate trends and global crop production since 1980. *Science*, 333(6042), 616–620. <https://doi.org/10.1126/science.1204531>
- López-Pedrouso, M., Alonso, J., & Zapata, C. (2014). Evidence for phosphorylation of the major seed storage protein of the common bean and its phosphorylation-dependent degradation during germination. *Plant Molecular Biology*, 84(4–5), 415–428. <https://doi.org/10.1007/s11103-013-0141-1>
- Lopez-Ribera, I., La Paz, J. L., Repiso, C., Garcia, N., Miquel, M., Hernandez, M. L., ... Vicient, C. M. (2014). The evolutionary conserved oil body associated protein OBAP1 participates in the regulation of oil body size. *Plant Physiology*, 164(3), 1237–1249. <https://doi.org/10.1104/pp.113.233221>
- Lu, W., Tang, X., Huo, Y., Xu, R., Qi, S., Huang, J., ... Wu, C. (2012). Identification and characterization of fructose 1,6-bisphosphate aldolase genes in *Arabidopsis* reveal a gene family with diverse responses to abiotic stresses. *Gene*, 503(1), 65–74. <https://doi.org/10.1016/j.gene.2012.04.042>
- Lush, W. M., & Wien, H. C. (1980). The importance of seed size in early growth of wild and domesticated cowpeas. *The Journal of Agricultural Science*, 94(1), 177–182. <https://doi.org/10.1017/S0021859600028033>

- Lv, G.-Y., Guo, X.-G., Xie, L.-P., Xie, C.-G., Zhang, X.-H., Yang, Y., ... Xu, H. (2017). Molecular characterization, gene evolution, and expression analysis of the Fructose-1, 6-bisphosphate aldolase (FBA) gene family in wheat (*Triticum aestivum* L.). *Frontiers in Plant Science*, 8(June), 1–18. <https://doi.org/10.3389/fpls.2017.01030>
- Maldonado-Cervantes, E., Huerta-Ocampo, J. A., Montero-Morán, G. M., Barrera-Pacheco, A., Espitia-Rangel, E., & Barba de la Rosa, A. P. (2014). Characterization of *Amaranthus cruentus* L. seed proteins by 2-DE and LC/MS-MS: Identification and cloning of a novel late embryogenesis-abundant protein. *Journal of Cereal Science*, 60(1), 172–178. <https://doi.org/10.1016/j.jcs.2014.02.008>
- Marchler-Bauer, A., Anderson, J. B., Cherukuri, P. F., DeWeese-Scott, C., Geer, L. Y., Gwadz, M., ... Bryant, S. H. (2005). CDD: A Conserved Domain Database for protein classification. *Nucleic Acids Research*, 33(DATABASE ISS.), 192–196. <https://doi.org/10.1093/nar/gki069>
- Matteucci, E., & Giampietro, O. (2011). Dipeptidyl Peptidase-4 Inhibition: Linking Chemical properties to clinical safety. *Current Medicinal Chemistry*, 18(31), 4753–4760. <https://doi.org/10.2174/092986711797535290>
- McCouch, S., Baute, G. J., Bradeen, J., Bramel, P., Bretting, P. K., Buckler, E., ... Zamir, D. (2013). Feeding the future. *Nature*, 499(7456), 23–24. <https://doi.org/10.1038/499023a>
- Mhamdi, A., Mauve, C., Houda, G., Saindrenan, P., Hodges, M., & Noctor, G. (2010). Cytosolic NADP-dependent isocitrate dehydrogenase contributes to redox homeostasis and the regulation of pathogen responses in *Arabidopsis* leaves. *Plant, Cell & Environment*, 33(7), 1112–1123. <https://doi.org/10.1111/j.1365-3040.2010.02133.x>
- Min, C. W., Gupta, R., Kim, S. W., Lee, S. E., Kim, Y. C., Bae, D. W., ... Kim, S. T. (2015). Comparative biochemical and proteomic analyses of soybean seed cultivars differing in protein and oil content. *Journal of Agricultural and Food Chemistry*, 63(32), 7134–7142. <https://doi.org/10.1021/acs.jafc.5b03196>
- Mølhøj, M., Verma, R., & Reiter, W.-D. (2003). The biosynthesis of the branched-chain sugar d-apiose in plants: functional cloning and characterization of a UDP-d-apiose/UDP-d-xylose synthase from *Arabidopsis*. *The Plant Journal*, 35(6), 693–703. <https://doi.org/10.1046/j.1365-313X.2003.01841.x>
- Mortimer, J. C., Laohavisit, A., Macpherson, N., Webb, A., Brownlee, C., Battey, N. H., & Davies, J. M. (2008). Annexins: Multifunctional components of growth and adaptation. *Journal of Experimental Botany*, 59(3), 533–544. <https://doi.org/10.1093/jxb/erm344>
- Mouzo, D., Bernal, J., López-Pedrouso, M., Franco, D., & Zapata, C. (2018). Advances in the biology of seed and vegetative storage proteins based on Two-Dimensional Electrophoresis coupled to Mass Spectrometry. *Molecules*, 23(10), 2462. <https://doi.org/10.3390/molecules23102462>
- Muñoz, N., Liu, A., Kan, L., Li, M.-W., & Lam, H.-M. (2017). Potential Uses of Wild Germplasms of Grain Legumes for Crop Improvement. *International Journal of Molecular Sciences*, 18(2), 328. <https://doi.org/10.3390/ijms18020328>
- Müntz, K., Belozersky, M. A., Dunaevsky, Y. E., Schlereth, A., & Tiedemann, J. (2001). Stored proteinases and the initiation of storage protein mobilization in

- seeds during germination and seedling growth. *Journal of Experimental Botany*, 52(362), 1741–1752. <https://doi.org/10.1093/jexbot/52.362.1741>
- Natarajan, S. S., Xu, C., Bae, H., Caperna, T. J., & Garrett, W. M. (2006). Characterization of storage proteins in wild (*Glycine soja*) and cultivated (*Glycine max*) soybean seeds using proteomic analysis. *Journal of Agricultural and Food Chemistry*, 54(8), 3114–3120. <https://doi.org/10.1021/jf052954k>
- Natarajan, S., Xu, C., Bae, H., Bailey, B. A., Cregan, P., Caperna, T. J., ... Luthria, D. (2007). Proteomic and genetic analysis of glycinin subunits of sixteen soybean genotypes. *Plant Physiology and Biochemistry*, 45(6–7), 436–444. <https://doi.org/10.1016/j.plaphy.2007.03.031>
- NCBI Resource Coordinators. (2013). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, 41(D1), D8–D20. <https://doi.org/10.1093/nar/gks1189>
- Nemati, M., Piro, A., Norouzi, M., Moghaddam Vahed, M., Nisticò, D. M., & Mazzuca, S. (2019). Comparative physiological and leaf proteomic analyses revealed the tolerant and sensitive traits to drought stress in two wheat parental lines and their F6 progenies. *Environmental and Experimental Botany*, 158, 223–237. <https://doi.org/10.1016/j.envexpbot.2018.10.024>
- Nguyen, T. P., Cuff, G., Hegedus, D. D., Rajjou, L., & Bentsink, L. (2015). A role for seed storage proteins in *Arabidopsis* seed longevity. *Journal of Experimental Botany*, 66(20), 6399–6413. <https://doi.org/10.1093/jxb/erv348>
- Oliveira, B. M., Coorssen, J. R., & Martins-de-Souza, D. (2014). 2DE: The Phoenix of Proteomics. *Journal of Proteomics*, 104, 140–150. <https://doi.org/10.1016/j.jprot.2014.03.035>
- Otegui, M. S., Herder, R., Schulze, J., Jung, R., & Staehelin, L. A. (2006). The proteolytic processing of seed storage proteins in *Arabidopsis* embryo cells starts in the multivesicular bodies. *The Plant Cell*, 18(10), 2567–2581. <https://doi.org/10.1105/tpc.106.040931>
- Pandey, P., Irulappan, V., Bagavathiannan, M. V., & Senthil-Kumar, M. (2017). Impact of combined abiotic and biotic stresses on plant growth and avenues for crop improvement by exploiting physio-morphological traits. *Frontiers in Plant Science*, 8(April), 1–15. <https://doi.org/10.3389/fpls.2017.00537>
- Pang, Q., Chen, S., Dai, S., Chen, Y., Wang, Y., & Yan, X. (2010). Comparative proteomics of salt tolerance in *Arabidopsis thaliana* and *Thellungiella halophila*. *Journal of Proteome Research*, 9(5), 2584–2599. <https://doi.org/10.1021/pr100034f>
- Park, Y. J., & Nishikawa, T. (2012). Characterization and expression analysis of the starch synthase gene family in grain amaranth (*Amaranthus cruentus* L.). *Genes Genet Syst*, 87(5), 281–289. <https://doi.org/10.1266/ggs.87.281>
- Peterson, C. L., & Laniel, M. A. (2004). Histones and histone modifications. *Current Biology: CB*, 14(14), 546–551. <https://doi.org/10.1016/j.cub.2004.07.007>
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13), 1605–1612. <https://doi.org/10.1002/jcc.20084>
- Pichereaux, C., Hernández-Domínguez, E. E., Santos-Díaz, M. del S., Reyes-Agüero, A., Astello-García, M., Guéraud, F., ... Barba de la Rosa, A. P. (2016).

- Comparative shotgun proteomic analysis of wild and domesticated *Opuntia* spp. species shows a metabolic adaptation through domestication. *Journal of Proteomics*, 143, 353–364. <https://doi.org/10.1016/j.jprot.2016.04.003>
- Pompa, M., Giuliani, M. M., Palermo, C., Agriesti, F., Centonze, D., & Flagella, Z. (2013). Comparative analysis of gluten proteins in three durum wheat cultivars by a proteomic approach. *Journal of Agricultural and Food Chemistry*, 61(11), 2606–2617. <https://doi.org/10.1021/jf304566d>
- Purkrtova, Z., Jolivet, P., Miquel, M., & Chardot, T. (2008). Structure and function of seed lipid body-associated proteins. *Comptes Rendus - Biologies*, 331(10), 746–754. <https://doi.org/10.1016/j.crvi.2008.07.016>
- Quiroga, I., Regente, M., Pagnussat, L., Maldonado, A., Jorrín, J., & de la Canal, L. (2013). Phosphorylated 11S globulins in sunflower seeds. *Seed Science Research*, 23(3), 199–204. <https://doi.org/10.1017/S0960258513000160>
- Raigond, P., Ezekiel, R., & Raigond, B. (2015). Resistant starch in food: A review. *Journal of the Science of Food and Agriculture*, 95(10), 1968–1978. <https://doi.org/10.1002/jsfa.6966>
- Rajjou, L., & Debeaujon, I. (2008). Seed longevity: Survival and maintenance of high germination ability of dry seeds. *Comptes Rendus Biologies*, 331(10), 796–805. <https://doi.org/10.1016/j.crvi.2008.07.021>
- Reddy, L. H., & Couvreur, P. (2009). Squalene: A natural triterpene for use in disease management and therapy. *Advanced Drug Delivery Reviews*, 61(15), 1412–1426. <https://doi.org/10.1016/j.addr.2009.09.005>
- Rigden, D. J., Lamani, E., Mello, L. V., Littlejohn, J. E., & Jedrzejewski, M. J. (2003). Insights into the catalytic mechanism of cofactor-independent phosphoglycerate mutase from x-ray crystallography, simulated dynamics and molecular modeling. *Journal of Molecular Biology*, 328(4), 909–920. [https://doi.org/10.1016/S0022-2836\(03\)00350-4](https://doi.org/10.1016/S0022-2836(03)00350-4)
- Righetti, K., Vu, J. L., Pelletier, S., Vu, B. L., Glaab, E., Lalanne, D., ... Buitink, J. (2015). Inference of longevity-related genes from a robust coexpression network of seed maturation identifies regulators linking seed storability to biotic defense-related pathways. *The Plant Cell*, 27(10), 2692–2708. <https://doi.org/10.1105/TPC.15.00632>
- Rocco, M., Tartaglia, M., Izzo, F. P., Varricchio, E., Arena, S., Scaloni, A., & Marra, M. (2019). Comparative proteomic analysis of durum wheat shoots from modern and ancient cultivars. *Plant Physiology and Biochemistry*, 135(July 2018), 253–262. <https://doi.org/10.1016/j.plaphy.2018.12.010>
- Rogowska-Wrzesinska, A., Le Bihan, M.-C., Thaysen-Andersen, M., & Roepstorff, P. (2013). 2D gels still have a niche in proteomics. *Journal of Proteomics*, 88, 4–13. <https://doi.org/10.1016/j.jprot.2013.01.010>
- Romero-Rodríguez, M. C., Maldonado-Alconada, A. M., Valledor, L., & Jorrin-Novo, J. V. (2014). Back to Osborne. Sequential protein extraction and LC-MS analysis for the characterization of the holm oak seed proteome. In *Plant Proteomics* (Vol. 1072, pp. 379–389). <https://doi.org/10.1002/9780470988879>
- Roy, A., Kucukural, A., & Zhang, Y. (2010). I-TASSER: A unified platform for automated protein structure and function prediction. *Nature Protocols*, 5(4), 725–738. <https://doi.org/10.1038/nprot.2010.5>
- Sampedro, J., Sieiro, C., Revilla, G., González-Villa, T., & Zarra, I. (2001). Cloning

- and expression pattern of a gene encoding an α -xylosidase active against xyloglucan oligosaccharides from *Arabidopsis*. *Plant Physiology*, 126(2), 910–920. <https://doi.org/10.1104/pp.126.2.910>
- Sano, N., Rajjou, L., North, H. M., Debeaujon, I., Marion-Poll, A., & Seo, M. (2016). Staying alive: Molecular aspects of seed longevity. *Plant and Cell Physiology*, 57(4), 660–674. <https://doi.org/10.1093/pcp/pcv186>
- Saucedo, A. L., Hernández-Domínguez, E. E., de Luna-Valdez, L. A., Guevara-García, A. A., Escobedo-Moratilla, A., Bojorquéz-Velázquez, E., ... Barba de la Rosa, A. P. (2017). Insights on structure and function of a Late Embryogenesis Abundant protein from *Amaranthus cruentus*: An Intrinsically disordered protein involved in protection against desiccation, oxidant conditions, and osmotic stress. *Frontiers in Plant Science*, 8(April), 1–15. <https://doi.org/10.3389/fpls.2017.00497>
- Sauer, J. D. (1967). The grain amaranths and their relatives: A revised taxonomic and geographic survey. *Annals of the Missouri Botanical Garden*, 54(2), 103. <https://doi.org/10.2307/2394998>
- Sew, Y. S., Ströher, E., Fenske, R., & Millar, A. H. (2016). Loss of mitochondrial malate dehydrogenase activity alters seed metabolism impairing seed maturation and post-germination growth in arabidopsis. *Plant Physiology*, 171(2), 849–863. <https://doi.org/https://doi.org/10.1104/pp.16.01654>
- Shah, M., Soares, E. L., Carvalho, P. C., Soares, A. A., Domont, G. B., Nogueira, F. C. S., & Campos, F. A. P. (2015). Proteomic analysis of the endosperm ontogeny of *Jatropha curcas* L. seeds. *Journal of Proteome Research*, 14(6), 2557–2568. <https://doi.org/10.1021/acs.jproteome.5b00106>
- Shewry, P. R., & Hey, S. (2015). Do “ancient” wheat species differ from modern bread wheat in their contents of bioactive components? *Journal of Cereal Science*, 65, 236–243. <https://doi.org/10.1016/j.jcs.2015.07.014>
- Shewry, P. R., Napier, J. A., & Tatham, A. S. (1995). Seed storage proteins: structures and biosynthesis. *The Plant Cell*, 7(7), 945–956. <https://doi.org/10.1105/tpc.7.7.945>
- Shih, M.-D., Hoekstra, F. A., & Hsing, Y.-I. C. (2008). Late Embryogenesis Abundant Proteins. In *Advances in Botanical Research* (pp. 211–255). [https://doi.org/10.1016/S0065-2296\(08\)00404-7](https://doi.org/10.1016/S0065-2296(08)00404-7)
- Shimizu, T., Kanamori, Y., Furuki, T., Kikawada, T., Okuda, T., Takahashi, T., ... Sakurai, M. (2010). Desiccation-induced structuralization and glass formation of group 3 Late Embryogenesis Abundant protein model peptides. *Biochemistry*, 49(6), 1093–1104. <https://doi.org/10.1021/bi901745f>
- Shiota, H., Yang, G., Shen, S., Eun, C.-H., Watabe, K., Tanaka, I., & Kamada, H. (2004). Isolation and characterization of six abscisic acid-inducible genes from carrot somatic embryos. *Plant Biotechnology*, 21(4), 309–314. <https://doi.org/10.5511/plantbiotechnology.21.309>
- Siddhuraju, P., & Becker, K. (2001). Species/variety differences in biochemical composition and nutritional value of Indian tribal legumes of the genus *Canavalia*. *Nahrung - Food*, 45(4), 224–233. [https://doi.org/10.1002/1521-3803\(20010801\)45:4<224::AID-FOOD224>3.0.CO;2-V](https://doi.org/10.1002/1521-3803(20010801)45:4<224::AID-FOOD224>3.0.CO;2-V)
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., ... Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence

- alignments using Clustal Omega. *Molecular Systems Biology*, 7(539). <https://doi.org/10.1038/msb.2011.75>
- Sigrist, C. J. A., De Castro, E., Cerutti, L., Cuche, B. A., Hulo, N., Bridge, A., ... Xenarios, I. (2013). New and continuing developments at PROSITE. *Nucleic Acids Research*, 41(D1), 344–347. <https://doi.org/10.1093/nar/gks1067>
- Silva Artur, M. A., Zhao, T., Ligterink, W., Schranz, E., & Hilhorst, H. W. M. (2019). Dissecting the genomic diversification of Late Embryogenesis Abundant (LEA) protein gene families in plants. *Genome Biology and Evolution*, 11(2), 459–471. <https://doi.org/10.1093/gbe/evy248>
- Siró, I., Kápolna, E., Kápolna, B., & Lugasi, A. (2008). Functional food. Product development, marketing and consumer acceptance—A review. *Appetite*, 51(3), 456–467. <https://doi.org/10.1016/j.appet.2008.05.060>
- Tandang-Silvas, M. R., Cabanos, C. S., Carrazco Peña, L. D., De La Rosa, A. P. B., Osuna-Castro, J. A., Utsumi, S., ... Maruyama, N. (2012). Crystal structure of a major seed storage protein, 11S proglubulin, from *Amaranthus hypochondriacus*: Insight into its physico-chemical properties. *Food Chemistry*, 135(2), 819–826. <https://doi.org/10.1016/j.foodchem.2012.04.135>
- Thalhammer, A., Hundertmark, M., Popova, A. V., Seckler, R., & Hinch, D. K. (2010). Interaction of two intrinsically disordered plant stress proteins (COR15A and COR15B) with lipid membranes in the dry state. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1798(9), 1812–1820. <https://doi.org/10.1016/j.bbamem.2010.05.015>
- Tnani, H., López, I., Jouenne, T., & Vicient, C. M. (2012). Quantitative subproteomic analysis of germinating related changes in the scutellum oil bodies of *Zea mays*. *Plant Science*, 191–192, 1–7. <https://doi.org/10.1016/j.plantsci.2012.02.011>
- Tunnacliffe, A., & Wise, M. J. (2007). The continuing conundrum of the LEA proteins. *Naturwissenschaften*, 94(10), 791–812. <https://doi.org/10.1007/s00114-007-0254-y>
- Tzen, J., & Huang, A. (1992). Surface structure and properties of plant seed oil bodies. *The Journal of Cell Biology*, 117(2), 327–335. <https://doi.org/10.1083/jcb.117.2.327>
- USDA Food Composition Databases. (n.d.). Retrieved June 7, 2019, from <https://ndb.nal.usda.gov/ndb/search/list>
- Valcárcel-Yamani, B., & Caetano Da Silva Lannes, S. (2012). Applications of quinoa (*Chenopodium quinoa* Willd.) and amaranth (*Amaranthus* Spp.) and their influence in the nutritional value of cereal based foods. *Food and Public Health*, 2(6), 265–275. <https://doi.org/10.5923/j.fph.20120206.12>
- Velarde-Salcedo, A. J., Barrera-Pacheco, A., Lara-González, S., Montero-Morán, G. M., Díaz-Gois, A., González De Mejía, E., & Barba De La Rosa, A. P. (2013). In vitro inhibition of dipeptidyl peptidase IV by peptides derived from the hydrolysis of amaranth (*Amaranthus hypochondriacus* L.) proteins. *Food Chemistry*, 136(2), 758–764. <https://doi.org/10.1016/j.foodchem.2012.08.032>
- Velarde-Salcedo, A. J., Bojórquez-Velázquez, E., & Barba de la Rosa, A. P. (2019). Amaranth. In *Whole Grains and their Bioactives* (pp. 209–250). Chichester, UK: John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781119129486.ch8>
- Villeth, G. R. C., Carmo, L. S. T., Silva, L. P., Santos, M. F., de Oliveira Neto, O. B., Grossi-de-Sá, M. F., ... Mehta, A. (2016). Identification of proteins in susceptible

- and resistant *Brassica oleracea* responsive to *Xanthomonas campestris* pv. *campestris* infection. *Journal of Proteomics*, 143, 278–285. <https://doi.org/10.1016/j.jprot.2016.01.014>
- Wan, L., Ross, A. R. S., Yang, J., Hegedus, D. D., & Kermode, A. R. (2007). Phosphorylation of the 12 S globulin cruciferin in wild-type and *abi1-1* mutant *Arabidopsis thaliana* (thale cress) seeds. *Biochemical Journal*, 404(2), 247–256. <https://doi.org/10.1042/BJ20061569>
- WHO. (2016). World health organization. Diabetes fact sheet. Retrieved June 7, 2019, from <https://www.who.int/en/news-room/fact-sheets/detail/diabetes>
- Withana-Gamage, T. S., & Wanasundara, J. P. D. (2012). Molecular modelling for investigating structure-function relationships of soy glycinin. *Trends in Food Science and Technology*, 28(2), 153–167. <https://doi.org/10.1016/j.tifs.2012.06.014>
- Witzel, K., Weidner, A., Surabhi, G. K., Varshney, R. K., Kunze, G., Buck-Sorlin, G. H., ... Mock, H. P. (2010). Comparative analysis of the grain proteome fraction in barley genotypes with contrasting salinity tolerance during germination. *Plant, Cell and Environment*, 33(2), 211–222. <https://doi.org/10.1111/j.1365-3040.2009.02071.x>
- Wozny, D., Kramer, K., Finkemeier, I., Acosta, I. F., & Koornneef, M. (2018). Genes for seed longevity in barley identified by genomic analysis on near isogenic lines. *Plant, Cell & Environment*, 41(8), 1895–1911. <https://doi.org/10.1111/pce.13330>
- Wright, H. H. (2005). The glycaemic index and sports nutrition. *Sajcn*, 18(3), 222–228.
- Xu, C., Caperna, T. J., Garrett, W. M., Cregan, P., Bae, H., Luthria, D. L., & Natarajan, S. (2007). Proteomic analysis of the distribution of the major seed allergens in wild, landrace, ancestral, and modern soybean genotypes. *Journal of the Science of Food and Agriculture*, 87(13), 2511–2518. <https://doi.org/10.1002/jsfa.3017>
- Zhang, K., Chen, Y., Zhang, Z., & Zhao, Y. (2009). Identification and verification of lysine propionylation and butyrylation in yeast core histones using PTMap software. *Journal of Proteome Research*, 8(2), 900–906. <https://doi.org/10.1021/pr8005155>
- Zhao, L., Chen, Y., Chen, Y., Kong, X., & Hua, Y. (2016). Effects of pH on protein components of extracted oil bodies from diverse plant seeds and endogenous protease-induced oleosin hydrolysis. *Food Chemistry*, 200, 125–133. <https://doi.org/10.1016/j.foodchem.2016.01.034>
- Zhou, H., Wang, L., Liu, G., Meng, X., Jing, Y., Shu, X., ... Li, J. (2016). Critical roles of soluble starch synthase SSIIIa and granule-bound starch synthase Waxy in synthesizing resistant starch in rice. *Proceedings of the National Academy of Sciences*, 113(45), 12844–12849. <https://doi.org/10.1073/pnas.1615104113>
- Zinsmeister, J., Lalanne, D., Terrasson, E., Chatelain, E., Vandecasteele, C., Vu, B. L., ... Leprince, O. (2016). ABI5 is a regulator of seed maturation and longevity in legumes. *The Plant Cell*, 28(11), 2735–2754. <https://doi.org/10.1105/tpc.16.00470>
- Zong, Y., Yao, S., Crawford, G. W., Fang, H., Lang, J., Fan, J., ... Jiang, H. (2017). Selection for oil content during soybean domestication revealed by x-ray

tomography of ancient beans. *Scientific Reports*, 7(January), 1–10.
<https://doi.org/10.1038/srep43595>