The Journal of Agricultural Science

cambridge.org/ags

Comparative proteomic analysis of hybrid maize MR2008 and its parental lines

E. E. Osawa-Martínez[1], B. Minjarez[1], Y. Rodríguez-Yáñez[1], E. E. Reza-Zaldivar[2], A. A. Canales-Aguirre[2] and S. Mena-Munguía[1]

[1]Universidad de Guadalajara, Centro Universitario de Ciencias Biológicas y Agropecuarias, Guadalajara, Mexico and [2]Unidad de Evaluación Preclínica, Biotecnología Médica y Farmacéutica, CONACYT Centro de Investigación y Asistencia en Tecnología y Diseño del Estado de Jalisco (CIATEJ), Guadalajara, México

Crops and Soils Research Paper

Cite this article: Osawa-Martínez EE, Minjarez B, Rodríguez-Yáñez Y, Reza-Zaldivar EE, Canales-Aguirre AA, Mena-Munguía S (2021). Comparative proteomic analysis of hybrid maize MR2008 and its parental lines. *The Journal of Agricultural Science* **159**, 570–579. https://doi.org/10.1017/S0021859621000812

Received: 9 November 2020
Revised: 23 June 2021
Accepted: 2 August 2021

Keywords:
Bioinformatics; heterosis; kernel; maize; proteomics

Author for correspondence:
S. Mena-Munguía, E-mail: smena@cucba.udg.mx

Abstract

Maize is one of the three staple foods in the world. The white variety represents 60% of the maize importation with a world consumption of 1125 million tons in 2019/2020. Currently, new technologies could contribute to the analysis of this seed, supporting quality control and improvement. This study aims to carry out the morphological and proteomic comparison between the hybrid MR2008 and its parental lines LUG03 and CML491 through mass spectrometry and bioinformatics analysis. Herein, we identified that 34.8% of the hybrid proteome differs from the parental proteome. Also, ontological and morphological analyses determined that the hybrid exhibits more characteristics related to CML491 than LUG03, for example, metabolic pathways and enzymes, such as anthocyanidin 3-O-glucosyltransferase (UniProt P16166). This analysis allowed the identification of dominant characters, metabolic pathways and confirms the utility of this methodology in agricultural practices, mainly in processes of selection and quality control of a crop.

Introduction

Maize is a widely studied species and an important food staple around the world. According to reports of the 2015/2016 cycle, the global maize production was approximately 1393 million tons. USA, China, Brazil, Argentina and India are the top five producing countries (FAOSTAT, 2017; Croptust.org, 2019). Maize is mainly used as food, flour, grits, hominy, animal feed, ethanol production, high fructose corn syrup and even bio-based plastics, while the corn starch is used as a binder, tablets additive for drugs and medications, as well as vitamins (Foley, 2013; Cummans, 2015). Maize is the main staple food of more than 1.2 billion people in Sub-Saharan Africa and Latin America (Prassad, 2008; CIMMYT, 2019). In Latin America, maize consumption is approximately 6.78 g/person/day; in Mexico, it is 25.4 g/person/day (FAOSTAT, 2014). Some reports established a per capita consumption of 330 g/day, representing 32–55% of the daily protein intake. In Mexico, the use of maize is estimated at 45 500 million tons, becoming one of the five major importing countries of maize for human consumption (Indexmundi.com, 2019), followed by Japan, the Republic of Korea, Spain and Vietnam (SAGARPA, 2016; FAOSTAT, 2017).

On the other hand, low and middle-income countries are exposed to external vulnerabilities related to production and trading with the rest of the world, mostly primary commodities. It is imperative to generate novel advances in plant biology knowledge, analysis and introduction of genetic variation, cytogenetics, quantitative genetics, molecular biology, biotechnology and, most recently, genomics. These approaches have been successfully applied to the plant breeding processes, increasing the production and yield of crops, as well as the scientific plant knowledge (Moose and Mumm, 2008).

Genomics, transcriptomics, proteomics and metabolomics have established new tools for the genetic manipulation, development and analysis of improved cultivars for food quality enhancement. Hybrid maize is one of the first examples of gene theory successfully applied to food production. Several studies demonstrated that hybrid maize possesses an increased yield. For example, in the USA, in the early 1930s, a yield of 1.6 ton/ha was documented; currently, it is approximately 9.5 ton/ha, this is a result of the development and application of technologies including hybrid maize, fertilizers and farm machinery (Duvick, 2005; Edgerton, 2009). In the same way, the average yield in the irrigation districts of Mexico is growing, with a yield fluctuation between 7 and 8 ton/ha, unlike areas that use native varieties with yields between 2 and 3 ton/ha (Turrent *et al.*, 2012).

Today, heterosis is accepted as a mechanism that revolutionized maize plant breeding and production. Since the first use of hybrids in the 1930s, maize yield has increased sixfold or more. Despite its importance, molecular mechanisms related to heterosis, such as hybrid vigour, are poorly understood (Holá *et al.*, 2017; Chen, 2013; Xing *et al.*, 2016). However, this

phenomenon is still present and observed in new hybrids exceeding the average yield production of their predecessors (Duvick, 2005).

All approaches to enhance crop quality have a high impact on human health. Human nutrition research is focused on quality-related aspects of food ingredients and extracts, and the specific diet formulation outcome. The underlying food matrix and the multiple interactions of their components could activate several physiological responses in the organism, including the growth and homeostatic regulation, and hyper-sensibility reactions (Sauer and Luge, 2014).

Currently, food production and distribution are a globalized issue. International and local regulatory agencies have established essential guidelines for food product labelling. However, it is challenging to determine the real origin of most of the components of some food products, mainly processed products. In this line, proteomics can contribute to essential areas of modern nutrition research. The use of robust mass spectrometric detectors in combination with liquid chromatography plays a vital role in food safety and quality issues. This technique is exceptionally well-suited for food contaminants analysis (Malik *et al*., 2010). Herein, the main objective was to carry out the morphological and proteomic comparison of a hybrid MR2008 and the parental lines LUG03 and CML491 through tandem mass spectrometry and bioinformatics analysis.

Materials and methods

Biological materials

Good agricultural practices for maize cultivation were employed. Seeds were grown in the tropical zone of La Huerta Jalisco, Mexico (19°30′N and 104°32′W) in the cycle fall/winter 2016/2017 on Haplic Phaeozem soil, pH 7.52, at 500 MASL, with semi-dry to dry weather during winter and spring, a temperature range of 16–33°C, and controlled pollination. Afterwards, they were harvested and transported to genebank at the University Center of Biological and Agricultural Sciences (CUCBA in Spanish) in Zapopan, Jalisco, Mexico. We randomly selected the seeds for morphological and proteomic analysis. The female parent line LUG03 and the MR2008 hybrid were kindly provided by the University of Guadalajara, and the male line CML491 (CIMMYT Maize Line) was obtained from the International Maize and Wheat Improvement Center (CIMMYT in Spanish) in Texcoco, Estado de Mexico, Mexico (CIMMYT, 2013).

The seed for the experiment is generally produced in the secondary growing season, in order to ensure isolation and gain a production cycle.

Morphology description

The morphological description was carried out in 50 plants randomly selected for each genotype. A total of 23 ordinal and nominal morphological variables were analysed employing the SNICS (National Service of Inspection and Certification of Seeds, species *Zea mays L.*) graphic manual (SNICS, 2010). The plants were grown in an experimental field in CUCBA, Zapopan, Jalisco, Mexico (20°44′34N, 103°33′W) during spring/summer rain-fed 2017, with a temperature range of 5–32°C; on Eutric Regosol soil, pH 6.76; at 1648 MASL, and an annual rainfall of 976.6 mm and relative humidity of 60.6%. For all genotypes, the data were collected and analysed with SIMPER (similarity percentages) to evaluate the contribution of each variable in the average resemblance between groups. The PCoA (Principal Coordinates Analysis), using the Gower coefficient for inclusion of qualitative variables, and PERMANOVA (Permutational Multivariate Analysis of Variance), for pairwise test comparisons in non-parametric data, were used to identify a significant difference between genotypes. The generated matrix had quantitative, qualitative and semi-quantitative morphological variables. This analysis was performed in PRIMER V6 software (Anderson *et al*., 2008).

Protein extraction

Seed samples for each genotype (5 g) were ground with a portable grinding mill. Afterwards, 100 mg of each sample were solubilized with 500 µl of 8 M urea and vortexed for 1 min at maximum speed. A 150 µl aliquot was added to a new microtube to be delipidated (three times) with a mixture of methanol-chloroform-water (4:1:3; v:v:v), vortexed 1 min at the highest speed, centrifuged at 12 000 × ***g*** for 5 min at room temperature, and finally methanol-precipitated (for details see Minjarez *et al*., 2013). Protein pellets were re-solubilized in 100 µl of 8 M urea by sonication. Next, 10 µl of each sample was combined with 20 µl of a mix of 8 M urea and 6 M guanidine hydrochloride. Then, the samples were precipitated with a TCA/acetone mixture (5°C), shaken vigorously, and placed on ice for 15 min. Centrifugation was carried out for 20 min at 4°C and 14 000 × ***g***. The supernatants were discarded, and the pellets were washed with cold acetone (−18°C), shaken, and centrifuged again for 20 min at 4°C maximum speed. The acetone was removed, and the pellets were air-dried. Finally, they were dissolved with a 400 µl of U/T/C mixture (7 M urea, 2 M thiourea, 2% CHAPS). The protein concentration was determined by the Bradford method (Bradford, 1976). Samples were stored at −20°C until use.

Trypsin digestion of polypeptides

As a result of the sample's low solubility, and to reduce polysaccharide concentration, 38 µl of the supernatant and the precipitate from each sample were loaded on a gel and stained with Coomassie Brilliant R250 Blue (Bio-Rad). First, an appropriate volume of 4× Laemmli Sample Buffer with β-mercaptoethanol was added for each sample and then denatured at 95°C for 5 min. Finally, the bands of each sample (precipitate and supernatants) were cut into fragments of approximately 1 mm, and the proteins were prepared for liquid chromatography and tandem mass spectrometry (LC-MS/MS) analysis.

Samples were digested with sequencing grade trypsin (Promega) as described by Shevchenko *et al*. (1996). Briefly, every sample was reduced with 10 mM of dithiothreitol for 20 min at 60°C. The thiol groups were alkylated, adding 55 mM of iodoacetamide for 30 min at room temperature in the dark. Then, 1000 ng of trypsin was added, and digestions were performed overnight. All the reagents were prepared in 50 mM ammonium bicarbonate buffer. The protein digestion was stopped with 12 µl of 10% trifluoroacetic acid (TFA) in water (Cf = 1%). The digestion mixture was dried in a vacuum centrifuge and resuspended in 20 µl of 2% acetonitrile (ACN), 0.1% TFA.

Liquid chromatography and tandem mass spectrometry

The resolution of sample complexity was achieved with high-performance liquid chromatography with a reverse-phase

https://doi.org/10.1017/S0021859621000812 Published online by Cambridge University Press

column. Five microlitres of each sample were loaded into a trap column (LC Column, 12 nm, 3 µm Triart-C18, 0.5 × 5.0 mm; YMC) and desalted with 0.1% TFA at 10 µl/min for 5 min. Peptides were loaded into an analytical column (LC Column, Luna Omega 3 µm Polar C18, 150 × 0.3 mm, Capillary Phenomenex) equilibrated with 3% ACN 0.1% formic acid (FA). Elution was carried out with a linear gradient of 3–35% B in A for 20 min (A: 0.1% FA; B: ACN, 0.1% FA) at a flow rate of 300 nl/min.

The mass spectrometry analysis in microESI qQTOF (6600plus TripleTOF, ABSCIEX) was carried out in the Proteomics Unit of the University of Valencia, Valencia, Spain (member of the PRB2-ISCIII ProteoRed Proteomics Platform). Samples were ionized in a source type: optiflow 1–50 µl micro applying 4.5 kV to the spray emitter. The analysis was carried out in a data-dependent mode. Mass spectrometer (MS1), were acquired scans from 400–1250 m/z for 200 ms. The quadrupole resolution was set to 'UNIT' for mass spectrometer in tandem (MS2), acquired at 100–1500 m/z for 20 ms in high sensitivity mode. The following switch criteria were used: charge: 2+ to 4 +; the minimum intensity of 300 counts per second (cps). Up to 75 ions were selected for fragmentation after each survey scan. Dynamic exclusion was set to 10 s. The system sensitivity was controlled with 2 µg trypsin digestion; 1726 proteins were identified in 20 min gradient.

Protein identification

The Protein Pilot default parameters were used to generate a peak list directly from 5600 Triple Tof wiff files. The Paragon algorithm (Shilov *et al.*, 2007) and the Protein Pilot v5.0 software were used to search the UniProtKB/Swiss-Prot database (version 01-2019) with the following established parameters: trypsin specificity, cys-alkylation, taxonomy restricted to *Z. mays L.* Proteins showing unused score >1.3 were identified with confidence ≥95% according to the following equation: ProtScore = –log (1–(per cent confidence/100)); per cent confidence 95%, expressed in ProtScore units 2.0.

Once the search was carried out, the data of the precipitate and the supernatant were combined because they belong to the same sample.

The Pro Group™ algorithm made protein grouping, a set of proteins that share some physical evidence. Unlike sequence alignment analyses, which compare full-length theoretical sequences, the formation of protein groups in Pro Group™ is made entirely by observed peptides. Because the observed peptides are determined from an experimentally acquired spectrum, the grouping is made directly from the spectrum. Thus, unobserved regions of protein sequence play no role.

Bioinformatics analysis

Protein identification was carried out using the UniProtKB/Swiss-Prot database (http://www.ebi.ac.uk/swiss-prot/) with the default parameters per page.

The list of protein names with 95% of confidence to *Z. mays L.*, UniProtKB/Swiss-Prot codes, m/z and pI was used in the Protein Annotation Through Evolutionary Relationship (PANTHER) classification system v15.0 (http://www.pantherdb.org/). Proteins were classified according to their role in molecular function and biological processes (Mi *et al.*, 2012). The Kyoto Encyclopedia of Genes and Genomes (KEGG) was employed to identify function levels and metabolic pathways (https://www.genome.jp/kegg/).

Results

Morphological characterization of genotypes

The analysis started and we identified significant morphological differences between genotypes through PERMANOVA with a P level <0.0001. The SIMPER analysis was performed to interpret the differences or similarities between group descriptors (morphological variables). The PCoA using the Gower coefficient (Fig. 1) identifies the contribution and explains the total variation of morphological characteristics evaluated.

Longer superior internodes, taller plants, an anthocyanin colouration at the base of glumes, a more prominent cob with a larger diameter and stigmas with less colouration characterize the hybrid (MR2008) compared to parental. The female parent line LUG03 exhibits a lower plant, shorter internode length, lack of stigma colouration, thinner upper cob leaves, as well as shorter cob length. In contrast, the male parent CML491 showed stigmas with anthocyanin colouration, leaf sheaths, adventitious roots with slight colouration, as well as in glume bases in the tassel; besides, it shows a different kernel shape and diameter and length of the cob.

Initially, data showed that the hybrid MR2008 shares more phenotypic characters with the male parent CML491 than the female parent LUG03, highlighting the stigma and the glume base anthocyanin colouration. Also, MR2008 is characterized by taller plants, longer superior internodes, as well as wider and longer ears.

Proteomics analysis: intersection proteins

To analyse the maize proteome, we developed a strategy using LC-MS/MS. First, the solubilized proteins were delipidated with a mixture of methanol-chloroform-water. Subsequently, proteins were reduced and alkylated and analysed through a hybrid quadrupole time-of-flight mass spectrometer.

The ACN gradient elution of the total ion count (Supplementary Figs 1(a), (b) and (c) show three tandem mass spectra for MR2008, LUG03 and CML491, respectively) exposed 2086 spectra for hybrid MR2008; 1908 for LUG03 and 1716 for CML491 (data not shown). Here, we identified a total of 500 polypeptides. The polypeptides were determined with at least two peptides in some genotypes with ≥95% confidence (Supplementary Table 1). As an example, there are three tandem mass spectra shown for anthocyanidin 3-O-glucosyltransferase (P16166), glutelin 2 (P04706) and malate synthase, glyoxysomal (P49081) with the sequence GVPAAHLSTFSAAATSFVVHC, GVGSTPILGQC[CAM]VEFLR and HGAALDAGGVEVR, respectively (Supplementary Figs 2(a), (b) and (c)).

The proteins identified in the three genotypes were grouped in a Venn-Euler diagram (Venny 2.1). Figure 2 shows the specific and shared proteins between all genotypes. From the 500 polypeptides, 290 were identified in MR2008, 256 in CML491, and 233 in LUG03. The 15% of proteins (75 polypeptides) are shared in all genotypes, a 5.2% (26 polypeptides) were present in both, MR2008 and CML491, and only 3% (15 polypeptides) are shared between MR2008 and LUG03. The hybrid (MR2008) had an exclusive expression of 174 proteins.

https://doi.org/10.1017/S0021859621000812 Published online by Cambridge University Press

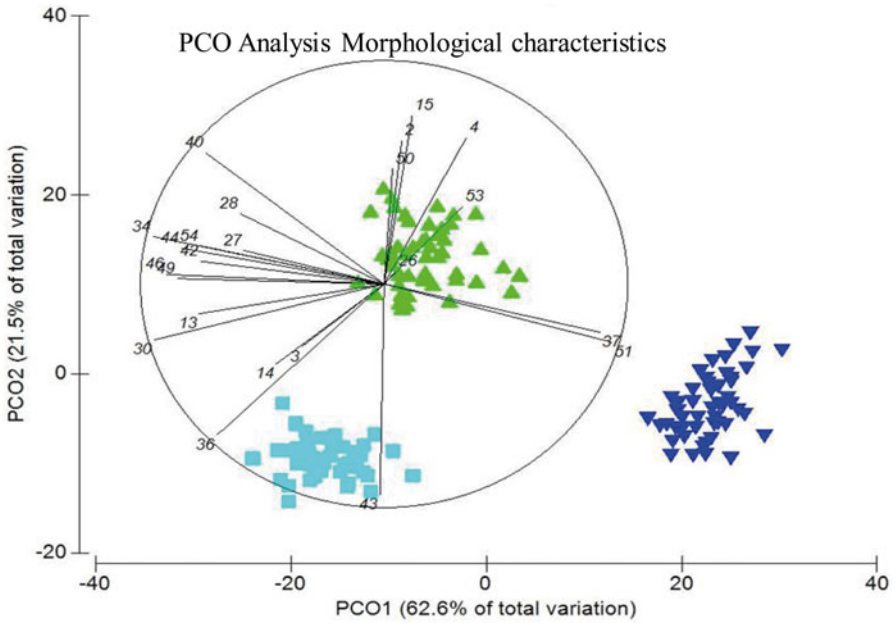

Fig. 1. Colour online. Principal Coordinates Analysis. Morphological variables explained 84% of total variation. Variables: 2. First leaf: length, 3. First leaf: width, 4. First leaf: length/width, 13. Stem: average length lower internodes, 14. Stem: diameter, 15. Stem: average length upper internodes, 26. Tassel: peduncle length, 27. Tassel length, 28. Tassel: main axis length, 30. Tassel: lateral branches attitude, 34. Tassel: anthocyanin colouration at the base of glumes, 36. Tassel: anthocyanin colouration of anthers, 37. Tassel: covering by the flag leaves, 40. Ear: intensity stigma colouration, 42. Tassel: length of lateral branches, 43. Plant: length, 44. Plant: height of ear, 46. Leaf: width of the blade, 49. Ear: length, 50. Ear: diameter, 51. Ear: shape, 53. Ear: number of grain rows, 54. Ear: number of grains per row. MR2008, LUG03 and CML491.

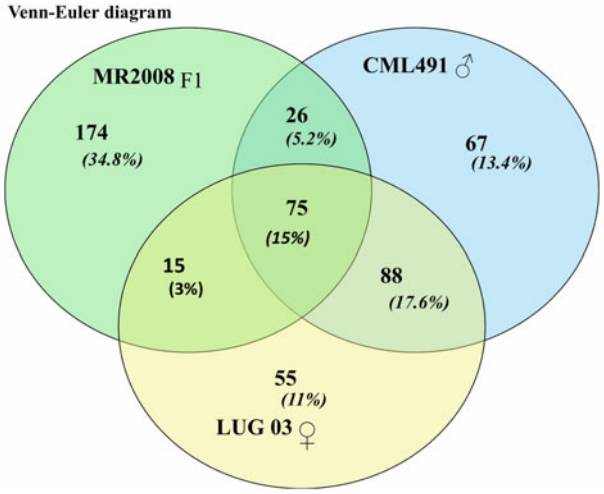

Fig. 2. Colour online. Venn diagram for three proteomes; shows the specific and the shared proteins in all genotypes. A total of 500 polypeptides were submitted; 290 for MR2008; 256 for CML491; 233 for LUG03; 75 shared between the three genotypes; 26 for MR2008 and CML491; 15 for MR2008 and LUG03, 174 differential proteins in MR2008.

Functional classification of polypeptides

The proteins identified were further classified according to molecular functions and biological processes using the PANTHER Classification System. Only those proteins shared between all genotypes were used to perform the functional classification (75 polypeptides). It is worth mentioning that there might be a difference between the access code numbers compared with the number of genes identified in the platform because a protein could not be described yet, or because the protein has more than one function.

Protein molecular functions were grouped into four categories, with 43 hits each. The category with more members was the catalytic activity (GO:0003824). This category possesses 25 polypeptides which represent 58.10% (Fig. 3a). Here, there are members such as the endochitinase A (P29022), aspartate aminotransferase (B4F9G1) and malate synthase, glyoxysomal (P49081) that are implicated in the hydrolase, transferase and oxidoreductase activity, respectively. The second category was the binding proteins (GO:0005488) with 13 members representing 30.2%; some members are the tubulin α-3 chain (P22275); 16.9 kDa class I heat shock protein 1 (B4G197) and phosphoglycerate kinase (B4G0K4); they belong to tubulin, chaperone and carbohydrate kinase classes, respectively.

In the classification of the biological processes, we obtained six categories with 60 hits each. The most representative biological process was the cellular process (GO:0009987), with 28 polypeptides representing 46.7%. Here we can find the adenosyl homocysteinase (C0PHR4), farnesyl pyrophosphate synthase (P49353) and glucose-6-phosphate isomerase (P49105). The metabolic process (GO:0008152) is the second most representative biological process, with 19 polypeptides representing 34.5% (Fig. 3b). Some members of this biological process are triosephosphate isomerase, cytosolic (P12863), ATP synthase subunit β, mitochondrial (P19023) and phosphoglucomutase (P93804).

A total of 15 proteins were submitted to PANTHER to analyse the protein intersection of the hybrid MR2008 and LUG03 (female parent line). In this way, six class hits related to molecular function were detected, where only four proteins belong to the binding category (GO:0005488), representing 40% with members such as ferritin-1 (P29036). Two polypeptides represented the catalytic activity category (GO:0003824). The proteasome subunit α type (A3FMB9) is a member of this category. In the biological processes, we got four categories with a total of 12 hits. The cellular process was the category with the greatest number of proteins (5) followed by the biological regulation categories (GO:0065007) and metabolic process (GO:0008152), with three and two members respectively (data not shown).

Regarding the proteins identified in both, the hybrid MR2008 and line CML491, a total of 26 proteins were submitted for classification. With molecular function, 11 class hits were detected. The highest number of genes identified belongs to catalytic activity (GO:0003824), a class with five genes, representing 45.5%. This

https://doi.org/10.1017/S0021859621000812 Published online by Cambridge University Press

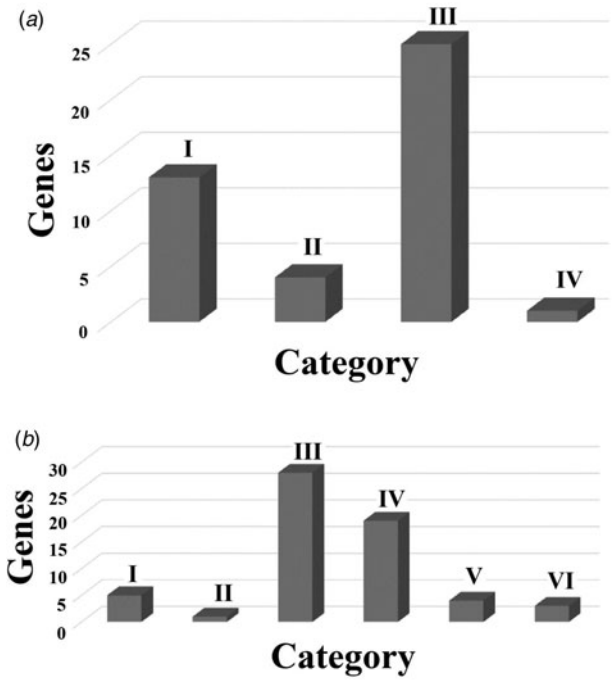

Fig. 3. Intersection proteins for MR2008, LUG03 and CML491, with 75 polypeptides and 43 protein class hits. (a) Molecular function: catalytic activity with more hits and 25 polypeptides representing 58.1%; for the group of binding proteins with 13 members representing 30.2%; total number of genes: 55. (I) Binding (GO:0005488); (II) structural molecule activity (GO:0005198); (III) catalytic activity (GO:0003824); (IV) transporter activity (GO:0005215). (b) Biological process; with six categories with 60 hits. The cellular process with 28 polypeptides and 46.7% and metabolic process with 19 polypeptides and 31.7% were the most representative categories. Total number of genes: 55. (I) Response to stimulus (GO:0050896); (II) signalling (GO:0023052); (III) cellular process (GO:0009987); (IV) metabolic process (GO:0008152); (V) biological regulation (GO:0065007); (VI) localization (GO:0051179).

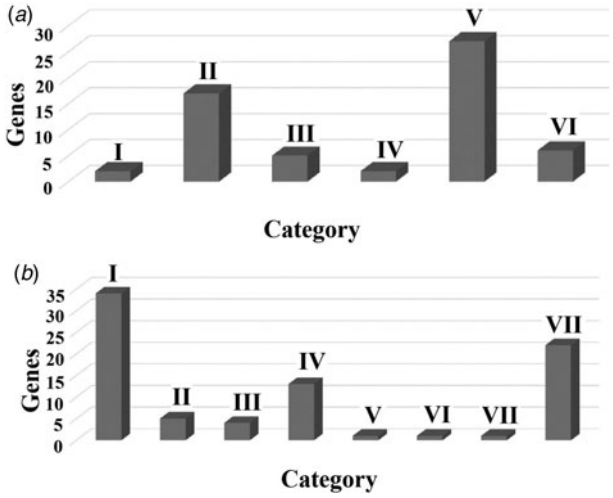

Fig. 4. Exclusive proteins in MR2008. (a) Molecular function; were identifying 85 genes and classes with 58 hits grouped into seven protein classes. The catalytic activity was the more significant group with 27 proteins members representing approximately 46.6%, the second was binding proteins (GO:0005488) with 27 genes representing 29.3%. (I) Translation regulator activity (GO:0045182); (II) binding (GO:0005488); (III) structural molecule activity (GO:0005198); (IV) molecular function regulator (GO:0098772); (V) catalytic activity (GO:0003824); (VI) transporter activity (GO:0005215). (b) Biological process; the classes with the more genes were the cellular processes (GO:0009987) with 34 genes representing 42% followed by the metabolic processes (GO:0008152) with 22 genes representing 27.2%. Total number of genes: 85. (I) Cellular process (GO:0009987); (II) localization (GO:0051179); (III) biological regulation (GO:0065007); (IV) response to stimulus (GO:0050896); (V) signalling (GO:0023052); (VI) developmental process (GO:0032502); (VII) multicellular organismal process (GO:0032501); (VIII) metabolic process (GO:0008152).

class has members such as peroxidase 1 (A5H8G4), acetolactate synthase 1, chloroplastic (Q41768) and alcohol dehydrogenase class-3 (P93629). The second class with more members was the binding class (GO:0005488). This class has the translationally controlled tumour protein (Q8H6A5), alcohol dehydrogenase class-3 (P93629) and histone H2B.1 (P30755) gene, representing 36.4%. In the biological process classification (data not shown), 17 genes and ten class hits were detected. Here, the most remarkable classes were the cellular process (GO:0009987) and the metabolic process (GO:0008152). The class with four and three genes, representing 40% and 30%, respectively, grouping proteins such as histone H2B.1 (P30755) in cellular process, alcohol dehydrogenase class-3 (P93629), in both classes. Because different proteins such as ferritin 1, alcohol dehydrogenase class 3 and peroxidase 1 were repeated, and the low number of polypeptides identified, we decided not to show both analyses.

Regarding the exclusive proteins of hybrid MR2008, a total of 174 entries were submitted for clustering. At the molecular function level, 85 genes were identified and classified into six protein classes with 58 class hits (Fig. 4a). The bigger group was the catalytic activity (GO:0003824) category, containing 27 members that represent 46.6%. Here we found the glutathione S-transferase (GST) 30 (Q9FQA9), glutathione peroxidase (B6T5N2), catalase (B6UHU1) and β-galactosidase (B4F9J1). The second major group was the binding proteins (GO:0005488) category conformed by 17 genes that represent 29.3%. As examples of this category, there is the heat shock 70 kDa protein (P11143), calnexin (B6TNF1), superoxide dismutase [Cu-Zn] 2 (P11428) and amine oxidase (A0A096SGQ3). Finally, in the biological processes (Fig. 4b), the classes with the highest number of genes were the cellular processes (GO:0009987), with 34 genes representing 42%; this class has members such as calnexin (B6TNF1); heat shock cognate 70 kDa protein 2 (B6SZ69) and glutaredoxin-dependent peroxiredoxin (B4FN24). The following are the metabolic processes (GO:0008152) with 22 genes that represent 27.2%. Here we found proteins related to the carbohydrate metabolic process like citrate synthase (B4FIC0).

Metabolic pathways

To identify the metabolic pathways, we submitted the UniProtKB/Swiss-Prot access codes to the KEGG database. We found that the exclusive proteome of the MR2008 hybrid is implicated in 53 different pathways (Table 1), while the CML491 and LUG03 exclusive proteome are involved in 16 and 28, respectively. In the same way, the number of pathways to the biosynthesis of secondary metabolites was higher in MR2008 than the male and female parents (20 *v.* 6 and 8, respectively).

Among the protein found in all genotypes, there is the linoleate 13S-lipoxygenase10 (A1XCI5; zma:100037809), which participates in linoleic acid metabolism (zma00591), as well as in the α-linolenic acid metabolism (zma00592). Moreover, there is a DIMBOA UDP-glucosyltransferase (B4G072; zma:100274317), which participates in the benzoxazinoid biosynthesis (zma00402).

In the proteins shared between MR2008 and LUG03, there is the protein HSP90-2 (C3UZ63; zma:100384478). This enzyme has been reported in the plant–pathogen interaction (zma04626),

https://doi.org/10.1017/S0021859621000812 Published online by Cambridge University Press

Table 1. Number of pathways in the exclusive proteomes of genotypes; in the hybrid MR2008 proteome, there were 53 pathways identified in contrast to the CML491 and LUG03, which presented 28 and 16 pathways, respectively

Pathways KEGG			
Genotype	MR2008	LUG03	CML491
Pathways	53	28	16
Metabolic pathways	31	11	8
Biosynthesis of secondary metabolites	20	8	6

as well as protein processing in the endoplasmic reticulum (zma04141).

Within proteins shared between CML491 and MR2008, we found the anthocyanin 3-O-glucosyltransferase (P16166), which is involved in the biosynthesis of secondary metabolites (zma01110) and anthocyanin biosynthesis (zma:732800). As mentioned in the morphological description, the anthocyanin colourations are present in several parts in both male parent and hybrid plants. The anthocyanin biosynthesis pathway requires a previous acetyl CoA, phenylpropanoids and phenylalanine synthesis (Ortíz *et al*., 2011; Zhang *et al*., 2014; Matin, 2018), in addition to proteins such as 3-hydroxybutyryl-CoA dehydrogenase (B4F890; zma:100191282), peroxidase 1 (A5H8G4) (zma:542029) and acetolactate synthase 1(Q41768) (zma:100285396). All these proteins additionally participate in phenylalanine metabolism, pantothenate and CoA biosynthesis (zma00770).

In the hybrid MR2008 proteome, we found that 31 proteins contribute to the metabolic pathways, and 13 proteins are involved with ribosomes. Proteins such as citrate synthase (B4FIC0) has been involved in the following seven pathways of *Z. mays L*: citrate cycle (zma00020), glyoxylate and dicarboxylate metabolism (zma00630), 2-oxocarboxylic acid metabolism (zma01210) metabolic pathways (zma01100), biosynthesis of secondary metabolites (zma01110), carbon metabolism (zma01200) and biosynthesis of amino acids (zma01230).

Discussion

Genomics offers knowledge and new possibilities to predict protein functions. Providing large-scale functional and molecular interaction networks allows the inference of functional annotations for genes and proteins. Moreover, it provides a better understanding of biological systems with a higher possibility of genetic improvement, particularly crops such as maize. Furthermore, the design of strategies based on omics can improve the production and selection of better seeds to achieve higher food quality control, as well as obtaining products such as enzymes. The plant analysis through LC-MS/MS and database searching allows the identification of large numbers of peptides from a generated spectrum. In this study, the main objective was to create a morphological and proteomic description through mass spectrometry and bioinformatics analysis of hybrid maize and its parental lines. A total of 500 proteins were identified. Subsequently, they were classified by genotype expression and functional annotation. The Supplementary Table 1 data contain descriptive characteristics for each protein, gene, Ip, m/z, as well as the access codes of each identified peptide.

Although proteomic analyses are subject to modifications as any experiment that involves plant processes, our results showed some components of kernel proteome that can be associated to plant survival function or a quality industrial, food and selection process improvement. The proteomic, metabolomic and enzymatic activity profiles have been used to develop biological systems and identify genetic traits.

In Mexico, the use of white maize kernel for human consumption is preferred, and the industrial use of hybrid maize is extensive. For this reason, our study was developed with the hybrid MR2008 from the Universidad de Guadalajara and its parents, LUG03 from Universidad de Guadalajara (U de G) and CML491 from CIMMYT. CUCBA has the breeding title of MR2008.

Although the proteomic analysis was carried out from the maize kernel, the seed determinate several processes such as respiration, synthesis, and degradation of compounds, growth, development and specific differentiation, intracellular controls as well as the response to external stimuli to ensure the embryo growth. Even though this type of study is usually influenced by environmental factors (pH, temperature and light), the data obtained show the proteome, as well as the molecular and biological processes for each genotype.

The morphology similarity between the hybrid MR2008 and CML491 has been exhaustively evaluated for their behaviour under biotic and abiotic stress. Therefore, the released lines showed good potential to be components for the improvement of pedigrees (Sierra-Macías, 2007; Andrés-Meza *et al*., 2011; Mageto *et al*., 2020). They generally present a good combining ability. CML491 is a QPM (Quality Protein Maize) line, exhibiting the aforementioned characteristics highlighted in the morphological description of this work. Moreover, CML491 stands out as an important component from other hybrids or pedigrees to incorporate tolerance to stressful conditions such as drought or little water.

On the other hand, LUG03 (Universidad de Guadalajara) generated at CUCBA stands out for being a line that presents a healthy aspect in the foliage development, as well as a floral synchronization with CML491, which promoted the hybrid MR2008 development with a good production of white kernel, adapted to subtropical zones.

The molecular function and biological process of the proteins found in all genotypes show a correlation with physiological activities of the seed, such as germination, respiration, mobilization of food reserves and nucleic acid synthesis (Cao *et al*., 2006). The metabolic processes during the germination of the seed for plant development are decisive for survival. Therefore, once the germination factors are conjugated, physiological, biochemical and morphological changes are activated during this stage, ensuring embryo development. Many of these processes are aimed at the degradation of reserve biomolecules in simpler components to be used efficiently, but also in the synthesis of amino acids, proteins and nucleic acids to ensure the growth of the new plant. All these processes are triggered when a series of factors such as humidity, temperature, light, oxygen and pH are appropriate for the species in question (Ali and Elozeiri, 2017).

Due to the above exposed in the ontological studies, it seems logical that the classification of genes based on the proteins detected in the corn kernel is aimed at ensuring metabolic processes, the division, growth and function of cellular components, the identification of responses to internal and external stimuli, as well as the degradation and synthesis of the necessary compounds in the germination process.

https://doi.org/10.1017/S0021859621000812 Published online by Cambridge University Press

The classification of the functions of the genes identified through the proteins detected by the MS/MS analysis at the levels of molecular function for the three genotypes. In the intersection proteins such as the exclusive proteins of the hybrid MR2008, the classes with the highest number of genes were those with catalytic activity (GO: 0003824), which indicates the importance of carbohydrate and protein metabolism from the seed to begin the germination process. The second class with the highest number of genes identified belongs to the binding class (GO: 0005488). Genetic regulations are vital processes in all organisms, which allow an organism to respond to its environment and promote its complexity. This can be achieved with the help of nucleic acid-binding proteins (Tak Leung *et al*., 2019).

Simultaneously at the level of biological processes, the gene ontology identified the classes with the highest number of genes in these proteomes, the intersection proteins in the three genotypes, the exclusive ones of the hybrid and the intersection of the hybrid with the parents, then the class with the highest number of genes identified were cellular processes (GO: 0009987) with 34 genes and 81 hits and metabolic processes (GO: 0008152) with 22 genes in hybrid's exclusive proteins.

In corn kernels, approximately 70% of the seed is made up of storage compounds (Ali and Elozeiri, 2017); therefore, the classification of proteins must show a breakdown of the processes that have to be carried out during the germination process after inhibition, the activation of metabolic processes, such as the synthesis of hydrolytic enzymes for the degradation of starches (Ali and Elozeiri, 2017) by hydrolases or oxidoreductases, but also in the genetic regulation of the vital processes of its environment (Tak Leung *et al*., 2019) such as transcription, translation, gene silencing (Hudson and Ortlud, 2014) in the case of the classification of genes that encode nucleic acid-binding proteins.

Within the regulation of biological processes by proteins shared between the hybrid MR2008 and LUG03, we detected the ferritin-1 chloroplastic protein; an iron-binding protein with an oxidoreductase function linked to the flowering process, leaf development, photosynthesis and reactive oxygen species regulation. This protein is synthesized from the *FER1* gene (P29036 in Uniprot KB). Moreover, this protein can be classified within the cellular process category as well as the heat shock protein 90 (HSP90 code UniProtKB: C3UZ63).

In the male parental proteome, proteins such as alcohol dehydrogenase class-3 (UniprotKB:P93629) have an essential role in response to a toxic substance and cellular catabolic process. Another example of proteins involved in the stimuli response process (GO:0050896) is the peroxidase 1 A5H8G4. This enzyme that oxidizes substrates plays a catalytic role in the processes of xylem lignification (Coleman *et al*., 2008). Regarding metabolic processes, the anthocyanidin 3-O-glucosyltransferase (Uniprot code: P16166) is included in the process of synthesis of secondary metabolites involved in pigmentation. This glucosyltransferase is necessary for pigment production through the flavonoid-anthocyanin pathway in maize (KEGG). These pigments are present as higher amounts of peonidin, cyanidin and pelargonidin in maize kernels (Salinas *et al*., 2013). Also, they can be found as dyes mainly in spikes, brace roots, and stigmas. This pigmentation was identified both in the male line and the hybrid MR2008 (Cocciolone and Cone, 1993; Quattrocchio *et al*., 1993; Lea and Leegood, 2001).

Within the MR2008 and CML491 proteome, there is the GST, an enzyme that facilitates the conjugate's entry to vacuoles. We also detected a protein encoded by the *Bronze-2* gene in maize, which performs the last genetically defined step of the anthocyanins biosynthesis, resulting in the deposition of red and purple pigments in the vacuoles of maize tissues (Marrs *et al*., 1995; Alfenito *et al*., 1998). In the MR2008 proteome, we identified the GST U6 (P12653), glutathione transferase 30 (Q9FQA9) and glutathione peroxidase (B6T5N2), while in CML491, we found the GST F2 (B6U1A7), glutathione transferase 5 (O24595), GST 22 (Q9FQB7). In contrast, in the LUG03 proteome analysis, there was no detection of the glutathione transferase enzyme, possibly for the characteristic colourless stigmas adventitious roots, stem glumes and leaf sheaths of this genotype.

The previous description is an example of the proteomic applicability in the maize kernel selection with the best phenotypic characters. In this study, we can observe that in the female parent, there are more proteins present associated with the functions of cellular regulation such as biological process (12 hits) than molecular functions (six hits). The male parent had 11 hits associated with the molecular function and ten hits for biological processes.

Also, we found more variation in the biosynthesis of the secondary metabolites of the hybrid (20) than the parents (Table 1). Metabolites are essential to plant survival in stress conditions (Labarrere *et al*., 2019). The variability of the proteome in the hybrid is also shown in ribosomal proteins. In this analysis, 18 ribosomal proteins were detected, which were expressed exclusively in the hybrid proteome. In contrast, CML491 exhibited six, and LUG03 eight ribosomal proteins (Supplementary Table 1; Marcon *et al*., 2013).

In the Venn diagram (Fig. 2), we show that 174 proteins are exclusively expressed in the hybrid MR2008. It has been reported that hybrid proteomes are more variable than parental (Xing *et al*., 2016; Cañas *et al*., 2017; Holá *et al*., 2017). In this study, about 34.8% of the proteome is unique in the hybrid proteome (174/500). Hu *et al*. (2017) analysed the proteome of a hybrid in immature maize cobs using tandem mass analysis, finding that 38.6% of the hybrid proteome was differentially expressed compared with its parents.

It could be hard to associate all the proteins exclusively identified in the hybrid to a morphological characteristic. According to Mendelian genetics, qualitative characters are conceptualized by the presence of a gene in plants with the expression of a character such as the colour of a flower. However, it is necessary to point out that quantitative or polygenic characteristics, mediated by the presence of several genes and their interaction, are assigned to many morphological characters in hybrids. Although we were able to identify in the proteome the enzyme anthocyanidin 3-O glucosyltransferase (P16166), which participates in the synthesis of anthocyanins that gives colouration in parts of the plant, several genes can synthesize different types of proteins; these proteins can present post-translational modifications, which can originate different functions depending on the state of the plant. On the other hand, concerning the specific proteins of the hybrid, if we remember the outstanding characteristics in the hybrid, the length of the upper internodes, the ear diameter, they are quantitative characteristics that were expressed by several genes, with complex interactions, which cannot be associated simply to a protein or the function of a gene. Therefore, expressing a phenotypic relationship for each protein would be complex. However, with the information provided in databases (Schmidt *et al*., 2014), we can identify the function of the proteins and the metabolic pathway where they belong, granting greater knowledge for each protein and their effects in the hybrids.

https://doi.org/10.1017/S0021859621000812 Published online by Cambridge University Press

This proteomic variation is most likely related to the diversity of proteins, which can participate in a diversity of metabolic pathways. The differential proteins in the products of the breeding process of two inbred lines are documented in the literature (Guo *et al*., 2013; Marcon *et al*., 2013); but the proteome of the corn kernel and the way to analyse that data can give us an idea of the specific changes that are shown in a hybrid.

Currently, without proteomics, the understanding of the response of biological systems is only partial because not all genes are transcribed or translated to functional products (Feussner and Polle, 2016).

Here, the proteome analysis of three genotypes shows us their diversity and the contributions of the parental lines to the F1. Additionally, we show the essential proteins implied in the vital process, such as binding proteins, catalytic activity and response to stimuli that are required for embryo development. Moreover, it is shown that some proteins are critical in metabolic pathways, such as the synthesis of secondary metabolites and homeostatic processes.

Regarding the low percentage of peptides shared between the parents and the hybrid, the process of synthesis of a protein involves several steps at the molecular level; such as the transcription and translation of mRNAs to synthesize proteins; also, various transcription factors may be linked to transcription mechanisms and regulate the protein synthesis. In addition, once synthesized, these proteins can be objects of post-translational modifications such as phosphorylation and methylation and that epigenetic states respond to environmental signals (Zhu, 2008; Chinnusamy and Zhu, 2009). The breeding of inbred lines produces seeds and tissues in F1 with a great variation in differential proteins, with fewer proteins shared between F1 and its parental lines. Other studies show that genome introgression has an average of 13.4% with ranges from 3.7 to 26.6% in *Hordeum vulgare* using genotyping and SNP markers (De la Fuente *et al*., 2018); for the case of corn through the hybridization of short gene probes with BAC libraries with fingerprints of two lines B73 and Mo17, Springer and Stupar (2007) found that both lines shared 80% of the genes, but B73 had an 11% that was specific and Mo 17 9%. Moreover, Meena *et al*. (2018) with several proteomic studies on heterosis revealed that the fraction of non-additive proteins in the hybrid varied depending on the genotype, tissue and stage of development, and they consider that multiple molecular mechanisms could lead to an expression of differential protein profile, which in turn could be contributing to heterosis. A constant in these breeding is that F1 encodes a great variety of proteins with respect to the parental inbred lines (Wang *et al*., 2020). Therefore, identifying the participation of proteins in metabolic pathways may be important to clarify the mechanisms of heterosis.

Regarding the LC-MS/MS, this technique shows those proteins that are in sufficient quantity to be detected, the fact that a protein is not present in the qualitative analysis does not indicate that it is not present in the proteome of the genotype analysed.

Finally, the maize kernel characterization provides a broad view of the protein diversity, as well as the phenotypic impact of their location and activity during the embryo development and the germination process. As well as the proportion of proteins present in each genotype and the contributions to the hybrid.

Although this is a 'shotgun' qualitative analysis, that does not require experimental replicates for the proteins found in the seeds of a simple hybrid and its two parents; it also shows that the proteins present in this kernel can give guidelines for the study of quality of hybrid seed, either through the test with several hybrids, their proteomes and their specific characteristics in their contained proteins, the characteristics between elite lines, their proteomes and their morphological characteristics, their differential proteomes, and the outstanding properties that they may have. In other areas such as pharmacological ones, proteomic analysis can show us those proteins that are more allergenic than are present in the various genotypes. For nutritional questions, they can give us an idea of which proteins are present in a genotype and which are not.

Other perspectives that can be taken from these qualitative analyses are that they can be carried out by quantifying proteins of interest, which were detected in the presented proteomes, or carrying out the study for a specific gene of the proteins present in the maize kernel. In other areas, it is also possible to continue replicating the studies and analysing the metabolic pathways of the enzymes detected, in order to clarify the processes carried out by the seed for its conservation and germination. On the other hand, secondary metabolites, which largely obey the stress response, can be studied in more detail to clarify the processes of homeostasis in the conservation of the viability of the seed. The ontology of genes appears in each genotype and the hybrids of other crosses, when they are in different development conditions to identify the response of genes to these conditions.

The proteins identified in MS supports by bioinformatic analysis could give us information related to different functions that might play into both germination and development of the plant to clarify those proteins involved these or in other physiological function like external stimuli, germination. Identify not only proteins, but also their sateges of development, the genes involved and compare the functions performed in the see on other organelles and in other stages of development.

Conclusions

In recent years, there has been a rapid growth in genotypic and bioinformatics characterization, through proteomic analysis using the LC-MS/MS technique. This work allows us to carry out proteomic analysis on maize seeds, standardize the technique and perform a bioinformatic analysis.

Hence, the proteomes of the three genotypes allow us to identify their diversity (greater in the case of the hybrid), as well as the contributions of the parents for F1; also enzymatic proteins fundamental for the development of important metabolic pathways in plants, a broad vision of the diversity of proteins, their classification based on its functions, given the various phenomena that occur in the germination and survival of the embryo. In the same way, we also found the proportion of proteins present in the hybrid and the lines, the contributions to the hybrid of their parents, as well as the expression of some of them in the phenotype of the F1 studied.

It is important to continue performing analysis within the seeds of the hybrids and the lines that form them; to observe and identify protein characterization patterns, based on an ontological classification; with the identification of the metabolic pathways in which their enzymes participate and clearly trace the processes carried out by both seeds and plants, in order to have a clearer and more precise understanding of the processes carried out by these organisms in their adaptation and survival; although this depends on future work.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0021859621000812.

https://doi.org/10.1017/S0021859621000812 Published online by Cambridge University Press

Acknowledgements. The authors would like to thank CONACYT Mexico for the support of EEOM (724639/634015 CONACYT).

Financial support. The authors would like to thank the University of Guadalajara, Mexico for the support of project 251987.

Conflict of interest. The authors declare no competing financial interest and took part in and approved the final manuscript.

Ethical standards. Not applicable.

References

Alfenito MR, Souer E, Goodman CD, Buell R, Mol J, Koes R and Walbot V (1998) Functional complementation of anthocyanin sequestration in the vacuole by widely divergent glutathione S-transferases. *The Plant Cell* **7**, 1135–1149.

Ali AS and Elozeiri AA (2017) Metabolic processes during seed germination Intech open. doi: 10.5772/intechopen.70653. Available at https://www.intechopen.com/books/advances-in-seed-biology/metabolic-processes-during-seed-germination.

Anderson MJ, Gorley RN and Clarke KR (2008) *PERMANOVA + for PRIMER, Guide to Software and Statistical Methods*. Plymouth, UK: PRIMER.

Andrés-Meza P, López Collado C, Sierra-Macías M, López-Romero G, Leyva-Ovalle O, Palafox-Caballero A and Rodríguez-Montalvo F (2011) Combining ability in maize lines using a diallelic cross. *Tropical and Subtropical Agroecosystems* **13**, 525–532.

Bradford MM (1976) A rapid and sensitive for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, *Analytical Biochemistry* **72**, 248–254.

Cañas RA, Yesbergenova-Cuny Z, Simons M, Chardon F, Armengaud P, Quilleré A, Cukier C, Yves Gibon Y, Limami A, Nicolas S, Brulé L, Lea P, Maranas C and Irel B (2017) Exploiting the genetic diversity of maize using a combined metabolomic, enzyme activity profiling, and metabolic modeling approach to link leaf physiology to kernel yield. *The Plant Cell* **29**, 919–943.

Cao D, Cheng H, Wu W, Soo HM and Peng J (2006) Gibberellin mobilizes distinct DELLA-dependent transcriptomes to regulate seed germination and floral development in Arabidopsis. *Plant Physiology* **2**, 509–525.

Chen JZ (2013) Genomic and epigenetic insights into the molecular bases of heterosis. Nature *Reviews Genetic* **14**, 471–482.

Chinnusamy V and Zhu J-K (2009) Epigenetic regulation of stress responses in plants. *Current Opinion in Plant Biology* **12**, 133–139.

CIMMYT (2013) El CIMMYT lanza 22 líneas endogámicas o puras de maíz para el trópico y el subtrópico, https://www.cimmyt.org/es/uncategorized/el-cimmyt-lanza-22-lineas-endogamicas-o-puras-de-maiz-para-el-tropico-y-el-subtropico/. (Accessed 14 August 2021).

CIMMYT (2019) Blogs: maize: from Mexico to the world, 20 May 2016. https://www.cimmyt.org/blogs/maize-from-mexico-to-the-world/ (Accessed 13 June 2021).

Cocciolone SM and Cone KC (1993) Pl-Bh, an anthocyanin regulatory gene of maize that leads to variegated pigmentation. *Genetics* **2**, 575–588.

Coleman HD, Samuels AL, Guy RD and Mansfield SD (2008) Perturbed lignification impacts tree growth in hybrid poplar – a function of sink strength, vascular integrity, and photosynthetic assimilation. *Plant Physiology* **3**, 1229–1237.

Cummans J (2015) What is corn used for? Capitule 13 surprising uses of corn, how to invest in commodities. https://commodityhq.com/investing-ideas/13-ways-corn-is-used-in-our-everyday-lives/ (Accessed 18 January 2021).

De la Fuente Cantó C, Russell J, Hackett CA, Booth A, Siobhan D, George T and Waugh R (2018) Genetic dissection of quantitative and qualitative traits using a minimum set of barley recombinant chromosome substitution lines. *BMC Plant Biology* **18**, 340.

Duvick DN (2005) The contribution of breeding to yield advances in maize (*Zea mays L.*). *Advance in Agronomy* **86**, 83–145, https://doi.org/10.1016/S0065-2113(05)86002-X.

Edgerton MD (2009) Increasing crop productivity to meet global needs for feed, food, and fuel plant physiology. doi: https://doi.org/10.1104/pp.108.130195.

FAOSTAT (2014) 2014/para proteínas g/persona por día suministro alimentario/cultivos equivalentes. Available at http://faostat3.fao.org/browse/FB/CC/S (Accessed June 20 2021).

FAOSTAT (2017) http://www.fao.org/worldfoodsituation/csdb/es/ (Accessed 18 January 2021).

Feussner I and Polle A (2016) What the transcriptome does not tell – proteomics and metabolomics are closer to the plants' patho-phenotype. *Current Opinion in Plant Biology* **26**, 26–31, ISSN 1369-5266.

Foley J (2013) It's time to rethink America's corn system; Scientific American. Available at https://www.scientificamerican.com/article/time-to-rethink-corn/ (Accessed 18 April 2021).

Guo B, Chen Y, Zhang G, Xing J, Hu Z, Feng W, Yao Y, Peng H, Du J, Zhang Y, Ni Z and Sun Q (2013) Comparative proteomic analysis of embryos between a maize hybrid and its parental lines during early stages of seed germination. *PLoS ONE*, **8**, e65867.

Holá D, Benešová M, Fischer L, Haisel D, Hnilička F, Hniličková H, Jedelský PL, Kočová M, Procházková D, Rothová O, Tůmová L and Wilhelmová N (2017) The disadvantages of being a hybrid during drought: a combined analysis of plant morphology, physiology and leaf proteome in maize. *PLoS ONE* **12**, e0176121.

Hu X, Wang H, Li K, Wu Y, Liu Z and Huang C (2017) Genome-wide proteomic profiling reveals the role of dominance protein expression in heterosis in immature maize ears. *Scientific Report* **7**, 1–12.

Hudson WH and Ortlund EA (2014) The structure, function and evolution of proteins that bind DNA and RNA. *Nature Reviews Molecular Cell Biology* **15**, 749–760.

Indexmundi.com. Available at https://www.indexmundi.com/agriculture/?country=mx&commodity=corn&graph=domestic-consumption. (Accessed February 20 2021).

KEGG (Kyoto Encyclopedia of Genes and Genomes). Available at http://www.genome.jp/kegg/ (Accessed February 20 2021).

Labarrere B, Prinzing A, Dorey T, Chesneau E and Hennion F (2019) Variations of secondary metabolites among natural populations of Sub-Antarctic *Ranunculus species* suggest functional redundancy and versatility. *Plants* **8**, 234.

Lea P and Leegood R (2001) *Plant Pigments GA Hendry, Plant Biochemistry and Molecular Biology Ed*. Wiley second edition. UK: Lancaster University.

Mageto E, Lee M, Dhliwayo T, Palacios-Rojas N, San Vicente F, Burgueño J and Hallauer AR (2020) An evaluation of kernel zinc in hybrids of elite quality protein maize (QPM) and non-QPM inbred lines adapted to the tropics based on a mating design. *Agronomy* **691**, 2–18.

Malik KA, Blasco C and Pico Y (2010) Liquid chromatography–mass spectrometry in food safety. *Journal of Chromatography A* **25**,4018–4040.

Marcon C, Lamkemeyer T, Malik W, Ungrue D, Piepho HP and Hochholdinger F (2013) Heterosis-associated proteome analyses of maize (*Zea mays L.*) seminal roots by quantitative label-free LC-MS. *Journal of Proteomics* **93**, 295–302.

Marrs K, Alfenito M and Lloyd A (1995) A glutathione S-transferase involved in vacuolar transfer encoded by the maize gene *Bronze-2*. *Nature* **375**, 397–400.

Martin GDA (2018) Los compuestos fenólicos: un acercamiento a su biosíntesis, síntesis y actividad biológica, *Revista de Investigación Agraria y Ambiental* **9**, 82–106.

Meena RK, Pullaiahgari D and Gudipalli P (2018) Proteomic analysis of heterotic seed germination in maize using F1 hybrid DHM 117 and its parental inbreds. *Turkish Journal of Biology* = Turk biyoloji dergisi **42**, 345–363.

Mi H, Muruganujan A and Thomas PD (2012) PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Research* **41**, D377–D386.

Minjarez B, Valero RML, del Sanchez PMM, González-Robles A, Sosa-Melgarejo JA, Luna-Muñoz J, Mena R and Luna-Arias JP (2013) Identification of polypeptides in neurofibrillary tangles and total homogenates of brains with Alzheimer's disease by tandem mass spectrometry. *Journal of Alzheimer's Disease: JAD* **34**, 239–262.

Moose SP and Mumm RH (2008) Molecular plant breeding as the foundation for 21st century crop improvement. *Plant Physiology* **3**, 969–977.

Ortíz M, Reza Vargas MC, Chew MR and Meza VJ (2011) Propiedades funcionales de las antocianinas. *Biotecnia* **13**, 16–22.

PANTHER DB. Available at http://pantherdb.org/ (Accessed 14 June 2021).

Prassad MNV (2008) *Trace elements as Contaminants and Nutrients: Consequences in Ecosystems and Human Health*. USA: Wiley and Sons, Inc Publication.

Quattrocchio F, Wing JF, Leppen H, Mol J and Koes RE (1993) Regulatory genes controlling anthocyanin pigmentation are functionally conserved among plant species and have distinct sets of target genes. *The Plant Cell* **11**, 1497–1512.

SAGARPA (2016) Panorama agroalimentario maíz 2016, Dirección de investigación y evaluación económica sectorial. Available at https://www.gob.mx/cms/uploads/attachment/file/200637/Panorama_Agroalimentario_Ma_z_2016.pdf (Accessed 10 June 2021).

Salinas Moreno Y, García Salinas C, Coutiño Estrada B and Vidal Martínez VA (2013) Variabilidad en contenido y tipos de antocianinas en granos de color azul/morado de poblaciones mexicanas de maíz. *Revista Fitotecnia Mexicana* **36**(Supl. 3-a), 285–294. (Accessed 07 June 2021). Available at http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0187-73802013000500005&lng=es&tlng=es.

Sauer S and Luge T (2014) Nutriproteomics: facts, concepts, and perspectives. *Proteomics* **15**, 5–6.

Schmidt A, Forne I and Imhof A (2014) Bioinformatic analysis of proteomics data. *BMC Systems Biology* **8**, S3.

Shevchenko A, Jensen ON, Podtelejnikov AV, Sagliocco F, Wilm M, Vorm O, Mortensen P, Boucherie H and Mann M (1996) Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. *Proceedings of the National Academy of Sciences of the USA* **93**, 14440–14445.

Shilov I, Seymour S, Patel A, Loboda A, Tang W, Keating S, Hunter C, Nuwaysir L and Schaeffer D (2007) The paragon algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass Spectra. *Molecular & Cellular Proteomics* **9**, 1638–1655.

Sierra-Macías M, Palafox-Caballero A, Becerra-Leor E, Córdova-Orellana E, Espinosa Calderón A and Rodríguez-Montalvo F (2007) Comportamiento de híbridos de maíz con alta calidad de proteína, por su buen rendimiento y tolerancia al 'Achaparramiento'. *Agronomía Mesoamericana* **18**, 93–101.

SNICS (2010) Manual gráfico para la descripción varietal en maíz, Segunda edición, México. https://www.gob.mx/cms/uploads/attachment/file/390820/MGDVMaiz.pdf. (Accessed 04 June 2021).

Springer NM and Stupar R (2007) Allelic variation and heterosis in maize: how do two halves make more than a whole?. *Genome Research* **17**, 264–275.

Tak Leung RW, Jiang X, Chu KH and Qin J (2019) ENPD – a database of eukaryotic nucleic acid binding proteins: linking gene regulations to proteins. *Nucleic Acids Research* **47**, D322–D329.

Turrent FA, Wise TA and Garvey E (2012) Achieving Mexico's maize potential, GDAE Working Paper 12-03. http://www.ase.tufts.edu/gdae/Pubs/wp/12---03TurrentMexMaize.pdf.

UNITPROT http://www.uniprot.org/ (Accessed 05 June 2021).

Venny 2.1 http://bioinfogp.cnb.csic.es/tools/venny/ (Accessed 12 June 2021).

Wang D, Mu Y, Hu X, Ma B, Wang Z, Zhu L, Xu J and Huang C (2020) Comparative proteomic analysis reveals that the heterosis of two maize hybrids is related to enhancement of stress response and photosynthesis respectively. *Research Square*. https://doi.org/10.21203/rs.3.rs-29700/v2 (Accessed 16 July 2021).

Xing J, Sun Q and Ni Z (2016) Proteomic patterns associated with heterosis. *Biochimica et Biophysica Acta* **8**, 908–915.

Zhang Y, Butelli E and Matin C (2014) Engineering anthocyanin biosynthesis in plants. *Plant Biology* **19**, 81–90. (Accessed 15 June 2021).

Zhu JK (2008) Epigenome sequencing comes of age. *Cell* **133**, 395–397. (Accessed June 5 2021).