ΦYTON
REVISTA INTERNACIONAL DE BOTÁNICA EXPERIMENTAL
INTERNATIONAL JOURNAL OF EXPERIMENTAL BOTANY
FUNDACION ROMULO RAGGIO
Gaspar Campos 861, 1638 Vicente López (BA), Argentina
www.revistaphyton.fund-romuloraggio.org.ar

Identification and evolutionary relationships of partial gene sequences from dehydrin group in three species of cacti

Identificación y relaciones evolutivas de secuencias parciales de genes del grupo dehidrina en tres especies de cactáceas

Hernández-Camacho S[1], E Pérez-Molphe-Balch[1], AG Alpuche-Solís[2], JF Morales-Domínguez[1]

Abstract. Dehydrins or Group 2 Late Embryogenesis Abundant (LEA) proteins play an important role in the response and adaptation to different types of abiotic stresses such as droughts, high salinity and low temperatures. Using PCR techniques, we identified three gene fragments that encoded dehydrin-like proteins in three cacti species *Opuntia ficus-indica* (OpfiDHN-like), *Leuchtenbergia principis* (LepDHN-like) and *Mammillaria bombycina* (MabDHN-like). Bioinformatic sequence analysis showed an identity between 96 and 97% with the *Opuntia streptacantha* dehydrin 1 (OpsDHN1) gene, demonstrating that the amplified fragments corresponded to dehydrin-like gene sequences, and that the designed oligonucleotides were effective for similar gene amplification in different cacti genera. Multiple OpfiDHN-like, LepDHN-like and MabDHN-like alignments showed that they possessed three repetitions of the conserved K segment. Also, a histidine rich motif was found, which is believed to facilitate the binding of these proteins with metal ions that probably evolved differently in the Opuntioidea and Cactoidea subfamilies of the Cactaceae family. Bioinformatic tools demonstrated that each of the three partial amino acid sequences corresponded to acidic, highly hydrophilic, and disordered protein fragments, which are characteristics of dehydrin proteins. Phylogenetic analysis using maximum parsimony, indicated that cacti dehydrins-like proteins were monophyletic, as well as those of other families.

Keywords: LEA; Hydrophilins; Alignment; Clade; Disordered proteins.

Resumen. Las dehidrinas o proteínas abundantes de la embriogénesis tardía (LEA) del grupo 2 juegan un rol importante en la respuesta y adaptación a diferentes tipos de estrés abiótico como deshidratación, alta salinidad y bajas temperaturas. Usando técnicas de PCR, se identificaron tres fragmentos de genes que codifican para proteínas tipo dehidrina de tres especies de cactus: *Opuntia ficus-indica* (OpfiDHN-like), *Leuchtenbergia principis* (LepDHN-like) y *Mammillaria bombycina* (MabDHN-like). El análisis bioinformático de las secuencias mostró que poseen una identidad entre el 96 y 97% con la secuencia del gen dehidrina 1 (OpsDHN1) de *Opuntia streptacantha*, demostrando que los fragmentos amplificados corresponden a secuencias de genes tipo dehidrina, y que los oligonucleótidos diseñados fueron efectivos para la amplificación de genes similares en diferentes géneros de cactáceas. El alineamiento múltiple de OpfiDHN-like, LepDHN-like y MabDHN-like mostró que poseen tres repeticiones del segmento K conservado. También se encontró un motivo rico en histidinas, el cual se cree que facilita la unión de estas proteínas con iones metálicos que probablemente evolucionaron de manera diferente en las subfamilias Opuntioidea y Cactoidea de la familia Cactaceae. Se demostró mediante técnicas bioinformáticas que cada una de las tres secuencias de aminoácidos parciales son ácidas, altamente hidrofílicas y desordenadas, las cuales son características de las proteínas tipo dehidrina. El análisis filogenético usando máxima parsimonia demuestra que las proteínas tipo dehidrina de las cactáceas son monofiléticas, así como las de otras familias de plantas.

Palabras clave: LEA; Hidrofilinas; Alineamiento; Clado; Proteínas desordenadas.

[1] Departamento de Química. Universidad Autónoma de Aguascalientes. Av. Universidad No. 940, Ciudad Universitaria, C.P. 20131; Aguascalientes, Ags., México. Tel. 052 449 9107400 ext. 8420 y 365.
[2] Departamento de Biología Molecular de Plantas. Instituto Potosino de Investigación Científica y Tecnológica A.C. Camino a la presa San José No. 2055, Col. Lomas 4a Sección, C.P 78216. San Luis Potosí, San Luis Potosí, México.
Address correspondence to: José Francisco Morales Domínguez, *e-mail:* jfmoral@correo.uaa.mx
Received 5.VI.2015. Accepted 24.I.2016.

INTRODUCTION

Hydrophilins are a very large and abundant group of plant proteins, also known as Late Embryogenesis Abundant (LEA) proteins (Reyes et al., 2008). These proteins were first isolated from cotton seeds accumulated during the late embryogenesis phase; their genes are expressed during late embryogenesis, as well as in vegetative tissues subjected to drought (Yang et al., 2012). Many LEA genes have been cloned from different plant species, and at least five different LEA groups have been classified on the basis of specific domains, amino acids sequences and similarity, distinctive motifs and gene expression (Bies-Ethève et al., 2008; George et al., 2009; Amara et al., 2012). Group 2 LEA proteins, also called Dehydrins (DHNs), is the most important and widely studied among LEA; they share a common K segment (EKKGIMDKIKEKLPG) present in one or several copies, also contain an S segment (polyserine) that can undergo extensive phosphorylation and a Y-domain (DEYGNP) similar to the nucleotide-binding site of plant and bacterial chaperones (Amara et al., 2012). LEA 2 proteins have been divided into five subclasses depending on the presence or absence of these motifs, including Y_nSK_n, Y_nK_n, SK_n, K_n and K_nS segments (where "n" corresponds to the segments number) (Shih et al., 2008). Amara et al. (2012) mentioned that all dehydrins have at least one K segment, which is usually located near the C-terminus and has the ability to form an amphipathic helix structure that may play a role in membrane and protein interactions. The S segment consists of a number of serine residues, which are possibly involved in the regulation of protein conformation and can be modified by phosphorylation. Finally, the Y segment (DEYGNP) is located near the N-terminus and is homologous to the nucleotide binding site, found in chaperone-like activity proteins of several organisms. A lysine rich segment, probably involved in the DNA or RNA dehydrin-binding process, has been identified between S and K segments (Yang et al., 2012). LEA proteins are part of an evolutionarily conserved group of hydrophilic proteins termed "hydrophilins" involved in various adaptive responses to hyperosmotic conditions (Garay-Arroyo et al., 2000). These proteins are highly hydrophilic and contained charged amino acid residues, and their expression appears to be abscisic acid-dependent. Both the pattern of expression and the structural features of LEA proteins suggest a general protective role in desiccation tolerance (Yang et al., 2012).

Dehydrins play an important role in the response and adaptation to different abiotic stresses such as drought, high salinity and low temperatures (Rodziewicz et al., 2014). Lopez et al. (2003) found a correlation between dehydrin accumulations and water deficit resistance in seven wheat crops during winter time, where three of these cultivars showed a significant overexpression of a 24 kD dehydrin. Furthermore, it has also been shown that some dehydrins can be overexpressed due to the action of phytohormones that are synthesized under biotic stress. Some of these dehydrins are: TaDHN (*Triticum aestivum*) (Shakirova et al., 2003) BcDh2 (*Boea crassifolia*) (Shen et al., 2004), Dc3 (*Daucus carota*) (Siddiqui et al., 1998), DHN-5 (*Triticum* sp) (Drira et al., 2015), among others. Ochoa-Alfaro et al. (2012) reported the overexpression of the *OpsDHN1* gene that encodes a SK_3 dehydrin protein in *Opuntia streptacantha* pads because of high salinity, cold and the presence of exogenous abscisic acid (ABA).

Li and Cao (2015) performed a comparative genome analysis including phylogenetic relationship, chromosomal localization, gene duplication, gene structure, and expression profile of the LEA gene family in maize, and they demonstrated that gene organization and motif composition of the LEA members are highly conserved in each of the groups, indicating their functional conservation. Bassett et al. (2015) made a phylogenetic analysis of the complete peach dehydrin family and in their results corroborate a close relationship between PpDhn1, 2 and 6 which are all or in part homologous with the *Arabidopsis* paralogs Xero1 and 2 and Rab18. PpDhn3 is located on scaffold 1 and shares sequence similarity and phylogenetic affinity with the *Arabidopsis* paralogs Cor47, Erd10 and Erd14, which were most likely duplicated after the peach and *Arabidopsis* lineages diverged from their most recent common ancestor.

Cacti are widely adapted plants that live under arid conditions and have the ability to reduce water loss associated with: (1) morphological adaptations (Ochoa-Alfaro et al., 2012), (2) a crassulacean acid metabolism (CAM), and (3) changes in gene expression (Tunnacliffe & Wise, 2007). Cactaceae family comprises 127 genera and 1438 species divided in four subfamilies: Cactoideae, Pereskiodeae, Maihuenioideae and Opuntioideae. Cacti are distributed exclusively in the New World except for *Rhipsalis baccifera*, which also occurs in Africa and Asia. The taxonomy of cacti has been traditionally based on comparative observation of morphological, molecular and biogeographic data (Calvente et al., 2011).

Little work has been reported on the presence of LEA genes in cacti. Here, we describe in detail three new partial sequences of dehydrins from the three cacti species: *Mammillaria bombycina, Leuchtenbergia principis* (belonging to Cactoideae subfamily), and *Opuntia ficus-indica* (belonging to Opuntioideae subfamily). These dehydrin protein genes are related to OpsDHN1 gene in *Opuntia streptacantha* showing putative amino acid sequence similarity with group LEA 2 proteins and a hydropathy index lower than one, indicating a higher abundance of hydrophilic amino acids. The phylogenetic study of these partial genes showed that these three sequences are monophyletic as well as other plant families.

MATERIALS AND METHODS

Plant material. *Leuchtenbergia principis* (W. J. Hooker), *Mammillaria bombycina* (Quehl) and *Opuntia ficus-indica* (Linnaeus P. Miller) were selected from the *in vitro* Germplasm bank of the Plant Biotechnology Unit located at the

Universidad Autónoma de Aguascalientes, Mexico. *M. bombycina*, *O. ficus-indica* and *L. principis* were propagated and maintained in a Murashige and Skoog (MS) medium (Murashige and Skoog, 1962). This medium was supplemented with 0.5 mg/L benzyl adenine (BA) for *M. bombycina*, and with 1 mg/L BA for *O. ficus-indica* and *L. principis*.

DNA isolation. A modified Murray and Thompson (1980) method was implemented for the cacti DNA extraction, which is described as follows: For each species, 300 mg of fresh tissue were weighed, frozen in liquid nitrogen and pulverized with a mortar. Samples were transferred into 1.5 mL microcentrifuge tubes and 1 volume of lysis buffer (2% CTAB, 1.4 M NaCl, 20 mM EDTA, 100 mM Tris-HCl pH 8, 0.2% β-mercaptoethanol) was added; all tubes were maintained at 65 ˚C for 10 min and then were placed in ice for 5 min; 300 µL of 1.4 M NaCl were added into the tubes and then incubated for 10 min at room temperature. Extractions were performed with 1 volume of phenol-chloroform (1:1), and 1 volume of chloroform-isoamyl alcohol (24:1). The aqueous phase was recovered and the nucleic acids were precipitated with 1 volume of isopropanol. This mix was incubated at room temperature for 10 min. DNA pellet was washed twice with 100 µL of 70% ethanol, and finally resuspended in 50 µL of sterile distilled water. The concentration and purity were measured using UV spectrometry at a wavelength of 260 and 280 nm; DNA was analyzed by electrophoresis using a 0.8% agarose gel, stained with ethidium bromide and visualized under ultraviolet light.

PCR primer design and amplification of dehydrin-like genes. Specific oligonucleotides were designed from the gene sequence OpsDHN1 (*Opuntia streptacantha*; gene bank access HO058650) using the Primer Select 5.0 DNAStar (DNASTAR, Inc.) program. The oligonucleotides were FOpsDHN1: 5'GAGGAGGAGGGAGATGACGAAGAC3' and ROpsDHN1: 5'GAAGGGGGTTGATCACACTCCACA3'. A fragment amplification of 497 bp size was expected. PCR techniques were performed using the Phusion Flash High-Fidelity PCR Master Mix (Thermo Scientific) commercial kit with 0.5 µM of each primer and 50 ng of DNA, under the following conditions: 1 cycle of 98 ˚C for 10 s, 30 cycles of 98 ˚C for 1 s, 62 ˚C for 5 s and 72 ˚C for 15 s, and finally one cycle of 72 ˚C for 1 min. A thermocycler TC-312 (Techne) was used. Amplifications were verified by electrophoresis in a 1.2% agarose gel, stained with ethidium bromide and visualized under ultraviolet light. Purification of the amplified fragments from the agarose gels was performed using the Wizard SV Gel and PCR Clean-Up System kit (Promega). All fragments were cloned into the pGEM® T-Easy Vector (Promega), and were sequenced at the National Genomic Laboratory for Biodiversity (LANGEBIO), Irapuato, Guanajuato, Mexico.

***In silico* analysis.** PCR product sequences were analyzed in order to search for homologous genes, using the Blastp program in the GenBank database on the NCBI web-server (http://www.ncbi.nlm.nih.gov). Virtual sequence translations were performed on the Bioinformatic Swiss Institute ExPASy platform (http://web.expasy.org/translate/). The NCBI BLASTx program was employed for the search of homologous proteins using the nucleotide sequences. Multiple alignment was performed with the ClustalW 3.0 method of the DNAstar MegAlign program (DNASTAR, Inc.). The conserved motifs were analyzed with the MEME platform [Motif-based sequence analysis tools; http://meme.nbcr.net/meme/] (Bailey et al., 2009), and the domains with PROSITE tool (http://prosite.expasy.org/prosite.html) platform. Also, we analyzed the amino acid sequence in order to determine: a) protein mass analysis; b) number of acid and basic amino acids with the GPMAW lite (General Protein Mass Analysis for Windows) of Alphalyse (http://www.alphalyse.com/gpmaw_lite.html); c) total number of amino acids; d) the isoelectric point; and e) GRAVY (Grand Average of Hydropathicity Index) using the ProtParam platform [http://web.expasy.org/protparam/] (Gasteiger et al., 2005). Amino acid hydropathicity plots were performed with the ProtScale tool [http://web.expasy.org/protscale/] (Gasteiger et al. 2005), based on the Kyte and Doolittle (1982) scale. The analysis of disordered protein degree was performed using the DisProt platform [http://www.disprot.org/pondr-fit.php] (Sickmeier et al., 2007). All the graphics were constructed with the GraphPad Prism 6.01 program.

Phylogenetic analysis. Multiple amino acid sequence alignment was performed using the ClustalW2 program of the European Bioinformatic Institute (http://www.ebi.ac.uk/Tools/msa/clustalw2/). We compared dehydrin-like proteins of *O. ficus-indica, M. bombycina* and *L. principis* with homologous dehydrins (full and partial sequences) of 18 different plant species (Table 1). Homologous proteins were obtained from the UniProtKB platform (UniProt Knowledgebase, http://www.uniprot.org/uniprot/) of the Universal Protein Resource server (UniProt server; http://www.uniprot.org/help/about) with the help of the FASTA program (http://www.ebi.ac.uk/Tools/sss/fasta/). Phylogenetic tree analysis was performed using a Maximum Parsimony (MP) method in PAUP * 4.0b10 software (Swofford et al., 2003); *Arabidopsis thaliana* DHN ERD10 was selected as outgroup. Characters were handled as disordered and with equal weight; all gaps were treated as missing data (uninformative). MP analysis trees were constructed using a heuristic search with 1 gradual random replica addition; the branch-swapping algorithm corresponded to tree-bisection-reconnection (TBR). Character state, distance matrix, strict consensus tree, and consistency index (CI) were analyzed. Clade strength was estimated using a

bootstrap analysis (Felsenstein, 1985) with 100 replicates. Through this analysis a tree was obtained for each step, using TBR branch swapping algorithm and collapsing the branches with a maximum length of zero; also all shorter trees (Multrees) were maintained. Only groups with >50% frequencies were retained.

RESULTS

PCR amplification showed an approximately 500 bp band in the three cacti species analyzed (Fig. 1). Partial nucleotide sequences were deposited in the GenBank database on the NCBI web-server (Table 1): *OpfiDHN-like* (445 nt) for *O. ficus-indica* (accession number: KP720561), *LepDHN-like* (493 nt) for *L. principis* (accession number: KP720562) and *MabDHN-like* (432 nt) for *M. bombycina* (accession number: KP720560). Comparison of the three nucleotide sequences and the virtual translation with other homologous sequences of the NCBI database showed that the three fragments have an identity between 96 and 97% with OpsDHN1 (accession number: HM581971.1 for the gene, and AEI52546.1 for the protein). Translated fragments of the OpfiDHN-like protein has 148 amino acids (accession number: AKC92526.1), LepDHN- like had 164 amino acids (accession number: AKC92527.1), and MabDHN-like had 145 amino acids (accession number: AKC92525.1); all of them had a +1 reading frame (Table 1). Multiple sequence alignment of the amino acids showed a similarity shared in OpfiDHN-like (AKC92526.1), LepDHN-like (AKC92527.1) and MabDHN-like (AKC92525.1) with OpsDHN1 (AEI52546.1); this similarity is higher at the K conserved segment of dehydrins (Fig. 2). Furthermore was observed in OpfiDHN-like and OpsDHN1 a histidine (H) rich segment, formed by a serie of six H; meanwhile in LepDHN-like and MabDHN-like this motif only have three H (Fig. 2). The searching for conserved motifs within the 18 dehydrins retrieved from UniProtKB platform (Table 2) showed that all sequences have at least 2 motifs or K segments, represented as motif 1 (PEAAVEHEAEAKEKKGFLDKIKEKLPGYH) and motif 3 (YEETEEKKGFLDKIKEKLPGH) (Fig. 3). Additionaly a motif 2 (EESGNVESTDRGLFDFLGKKEEEKPQHAH) is present in almost all total and partial sequences at the N-terminus of the peptides, but not in OpfiDHN-like, LepDHN-like, MabDHN-like and *Pinus mugo* subsp *rotundata* DHN1-like (Fig. 3). The graphical representation of the domains of the three dehydrin-like and OpsDHN1 proteins (Fig. 4) showed that OpsDHN1 contained two distinct domains: a Dehydrin 1 domain, which was formed by the S segment, and a conserved Dehydrin 2 domain formed by a K segment. We observed that OpfiDHN-like, LepDHN-like and MabDHN-like lack the Dehydrin 1 domain, although they have three repetitions of the Dehydrin 2 domain (Fig. 4). The amino acid characteristic study of putative proteins and OpsDHN1 showed that all sequences had a higher number of acid than basic amino acids (Table 2). The GRAVY values were negative between -1.331 to -1.624, indicating that these proteins are highly hydrophilic (Table 3). This was confirmed by the hydropathy values obtained of amino acids where most of them fall into a score lower than 0 (Fig. 5). Analysis of the disordering degree of amino acid sequences showed that most of the amino acids were arranged in values above 0.5 (Fig. 6).

Phylogenetic analysis using the Maximum Parsimony method (MP), showed a total *ingroup* of 338 characters, where 82 of them were variable and 159 were parsimony-informative. The distance matrix (divergence proportions of total amino acids, excluding gaps and polymorphic sites) had a variation from 0.00444 to 0.76744, and the heuristic search identified 32 most parsimonious trees out of 765 steps (CI = 0.852). Strict consensus tree (Fig. 7) showed the C clade conformed by two subclades with a 91% bootstrap support (PBS); the first subclade contained the Cactaceae family dehydrins with 100% of PBS; sister subclade corresponded to the Amaranthaceae family dehydrins and a *Tamarix hispida* dehydrin with 66% of PBS. Closely related to the clade C is the *Eriobotrya japonica* dehydrin and the sister B clade composed by dehydrins of the Salicaceae family with a 100% SBP. The A clade is composed by dehydrins of the Solanaceae family with a 100% SBP.

Table 1. Nucleotide and amino acid sequences length of *O. ficus-indica*, *L. principis* and *M. bombycina* dehydrins.
Tabla 1. Tamaño de las secuencias de nucleótidos y aminoácidos de las dehidrinas de *O. ficus-* indica, *L. principis* y *M. bombycina*.

Specie	Nucleotide sequence length and accession number	Amino acid sequence length and accession number	Amino acid Identities (%)	Alignment with the OpsDHN1 protein in the amino acids	Expected value
Opuntia ficus-indica	445 (KP720561)	148 (AKC92526.1)	97	81- 228	1e-89
Leuchtenbergia principis	493 (KP720562)	164 (AKC92527.1)	96	81- 246	2e-99
Mammillaria bombycina	435 (KP720560).	145 (AKC92525.1)	96	99- 245	2e-86

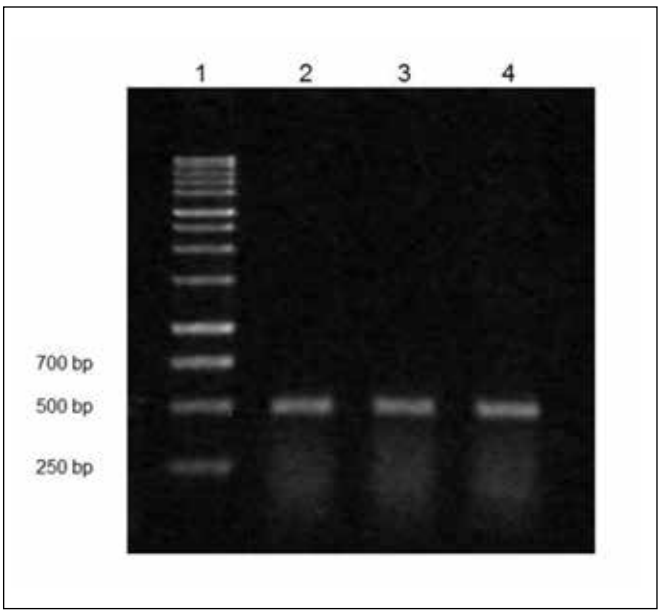

Fig. 1. PCR amplification of dehydrin-like genes in *Opuntia ficus-indica, Leuchtenbergia principis* and *Mammillaria bombycina*. Lanes: (1) Molecular weight marker 1kb (Promega), (2) *O. ficus-indica*, (3) *L. principis*, (4) *M. bombycina.*

Fig. 1. Amplificación por PCR de genes tipo dehidrina en *Opuntia ficus-indica, Leuchtenbergia principis* y *Mammillaria bombycina*. Carriles: (1) Marcador de peso molecular 1kb (Promega), (2) *O. ficus indica*, (3) *L. principis*, (4) *M. bombycina*.

DISCUSSION

For the first time, dehydrin-like genes of *Opuntia ficus-indica*, *Mammillaria bombycina* and *Leuchtenbergia principis* were described. In the search of homologous sequences, we found that they share an identity between 96 and 97% with the *Opuntia streptacantha* dehydrin 1 (OpsDHN1) (Table 1), that have a full length cDNA sequence (accession no. HO058650) of 905 bp, including 76 bp of 5'-UTR and 82 bp of 3'-UTR, the genomic fragment of 1,139 bp in length includes part of the 5' and 3' -UTR regions, as well as the entire coding region, which was interrupted by an intron of 234 bp inserted within the sequence that encodes the S-segment (Ochoa-Alfaro et al., 2012). This confirmed that the amplified fragments corresponded to LEA 2 gene sequences, and that the designed oligonucleotides were effective for homologous gene amplification in different cacti genera (Fig. 1).

Multiple alignment of putative amino acid sequences (Fig. 2) evidenced the presence of three K segment repetitions in agreement with Sun et al. (2013) who perform a comparison of amino acid sequences of AmDHN from *Ammopiptanthus mongolicus* with other reported dehydrin proteins in NCBI database and found a conserved sequence corresponding to K segment, whose sequence was EKKGIMNKIKEKLPG (Fig.1) similar to that found in cacti. This segment is extremely important when the protein is exposed to dehydrated

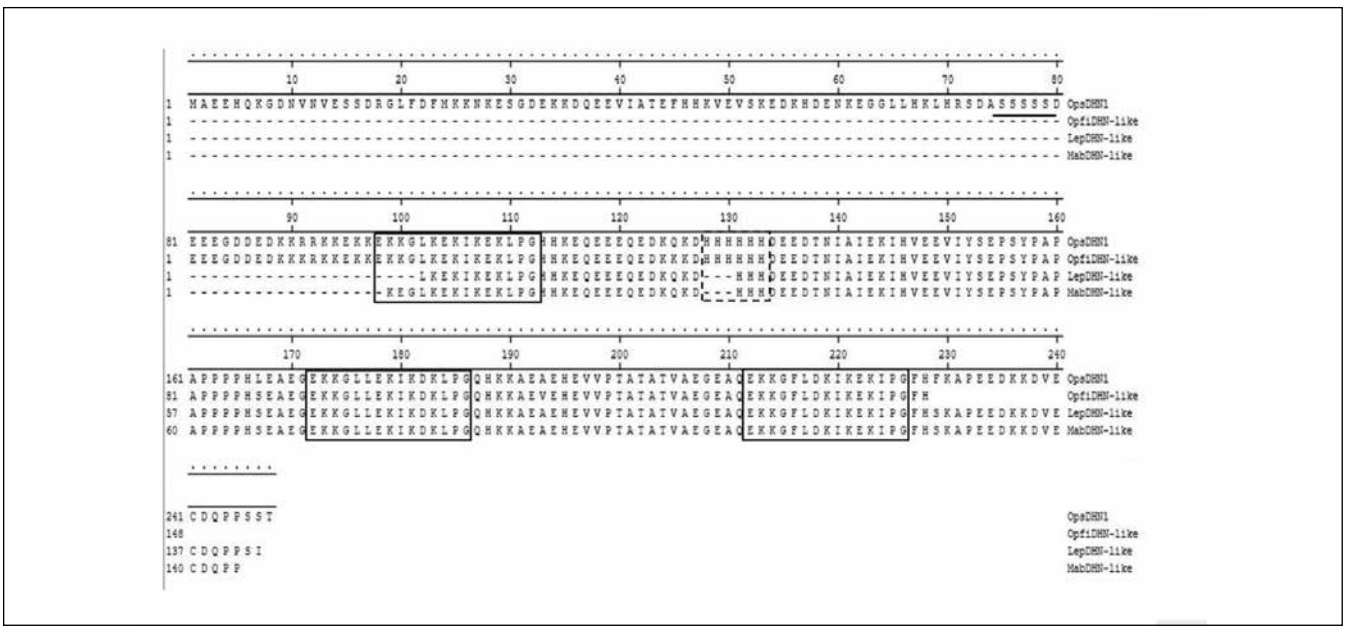

Fig. 2. Multiple alignment of the amino acid sequences of OpfiDHN-like, LepDHN-like, MabDHN-like and OpsDHN1. The sequences corresponding to the S segment are underlined; K segments are in the black squares and a Histidine rich motif is highlighted in the dashed box.

Fig. 2. Alineamiento múltiple de las secuencias de aminoácidos de OpfiDHN-like, LepDHN-like, MabDHN-like y OpsDHN1. Las secuencias correspondientes al segmento S se encuentran subrrayadas; los segmentos K se encuentran en los recuadros negros y un motivo rico en histidinas se encuentra en el recuadro con líneas punteadas.

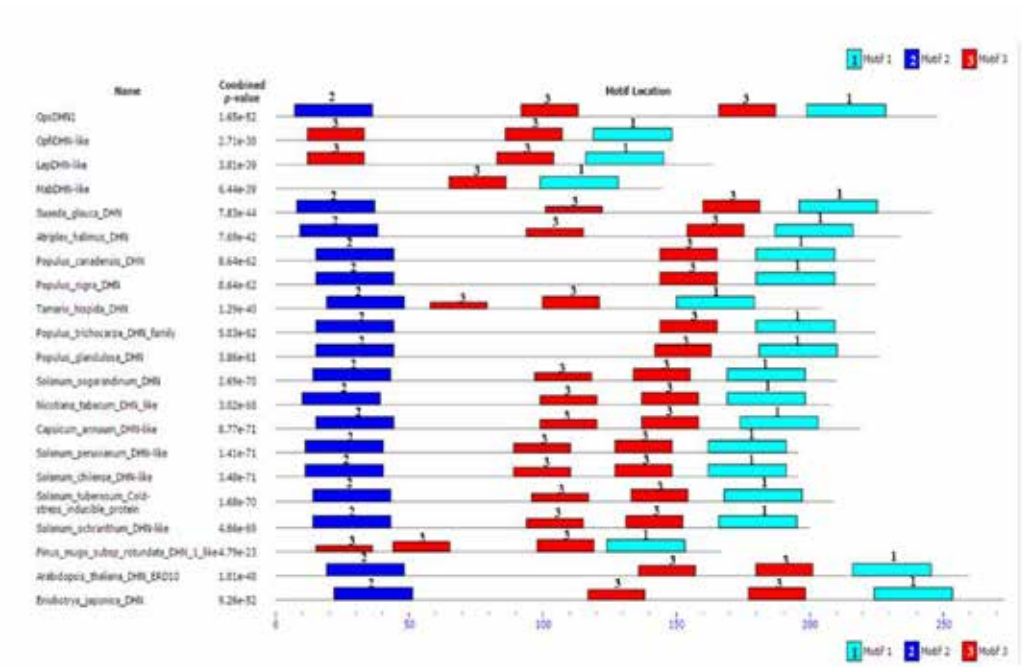

Fig. 3. Analysis of motifs of 21 dehydrin proteins. Numbers correspond to motifs 1 (PEAAVEHEAEAKEKKGFLDKIKEKLPGYH), 2 (EESGNVESTDRGLFDFLGKKEEEKPQHAH), and 3 (YEETEEKKGFLDKIKEKLPGH), respectively. Repetitions indicate the same motif in a given sequence.

Fig. 3. Análisis de motivos de 21 proteínas tipo dehidrina. Los números corresponden a los motivos 1 (PEAAVEHEAEAKEKKGFLDKIKEKLPGYH), 2 (EESGNVESTDRGLFDFLGKKEEEKPQHAH) y 3 (YEETEEKKGFLDKIKEKLPGH), respectivamente. Las repeticiones indican la repetición del mismo motivo en una secuencia dada.

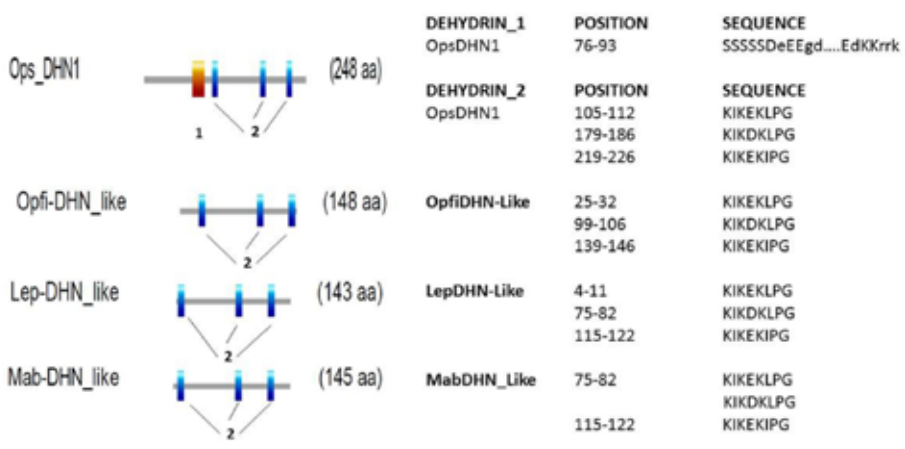

Fig. 4. Graphical representation of the functional domains of OpfiDHN-like, LepDHN-like, MabDHN-like and OpsDHN1. The domains dehydrin 1 and dehydrin 2 are indicated with numbers 1 and 2, respectively.

Fig. 4. Representación gráfica de los dominios funcionales de OpfiDHN-like, LepDHN-like, MabDHN-like and OpsDHN1. El dominio dehidrina 1 y dehidrina 2 se representan con los números 1 y 2, respectivamente.

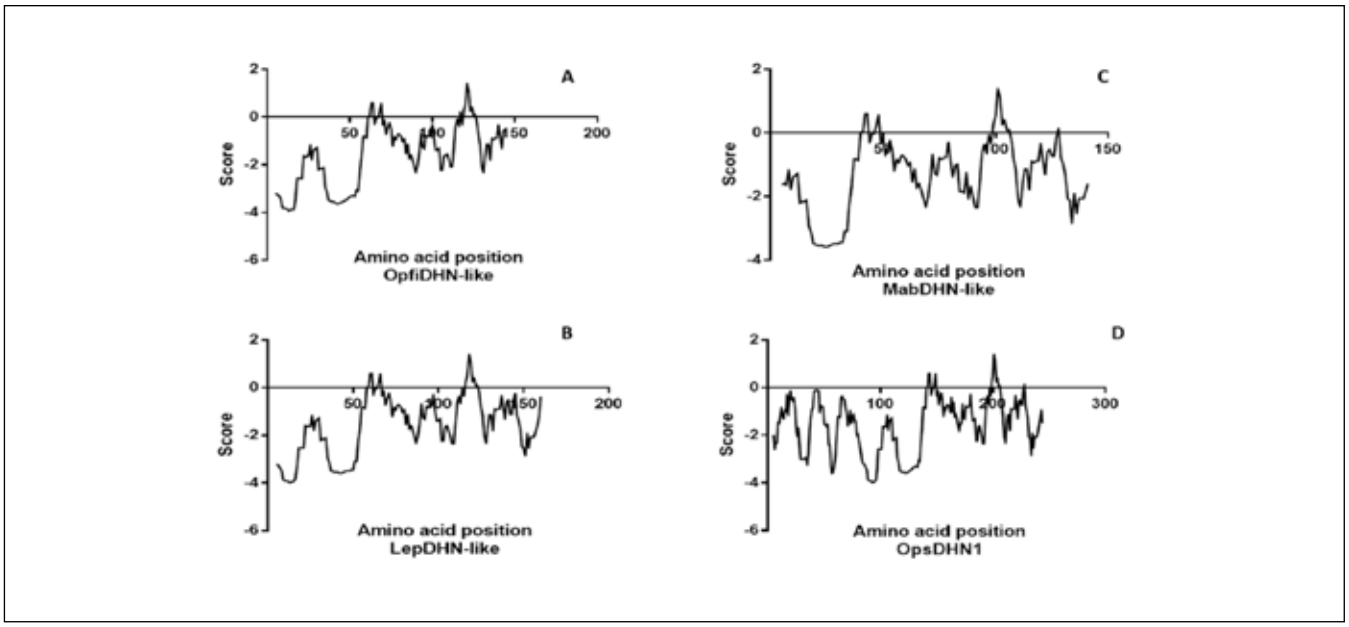

Fig. 5. Hydropathicity values of OpfiDHN-like, LepDHN-like and MabDHN-like amino acids. The data plotted were based on the Kyte and Doolittle (1982) scale, where the smaller the value the more hydrophilic is the amino acid. Hydropathicity of: (A) OpfiDHN-like; (B) LepDHN-like; (C) MabDHN-like; (D) OpsDHN1.

Fig. 5. Hidropaticidad de los aminoácidos de OpfiDHN-like, LepDHN-like y MabDHN-like. Los datos graficados se obtuvieron en base a la escala de Kyte y Doolittle (1982) donde cuanto menor es el valor más hidrofílico es el aminoácido. Hidropatía de: (A) OpfiDHN-like; (B) LepDHN-like; (C) MabDHN-like; (D) OpsDHN1.

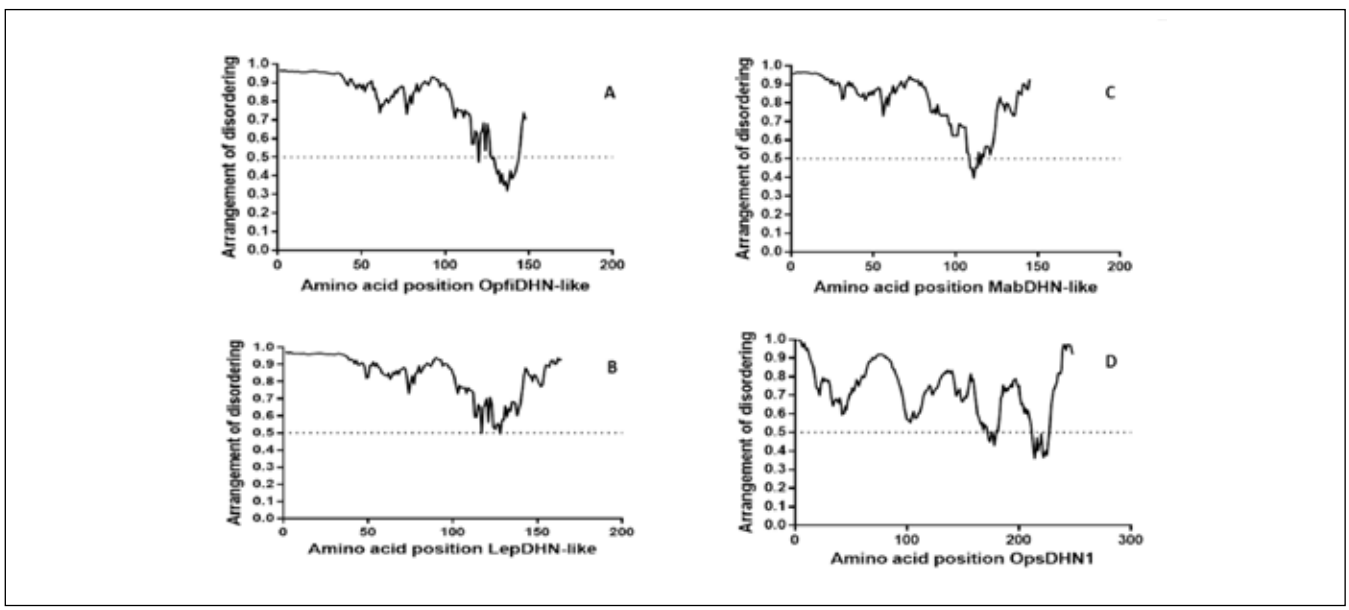

Fig. 6. Degree of disorder of OpfiDHN-like, LepDHN-like and MabDHN-like proteins according to DisProt platform (Sickmeier et al., 2007). (A) OpfiDHN-like; (B) LepDHN-like; (C) MabDHN-like; (D) OpsDHN1.Values above 0.5 are disordered regions of the protein.

Fig. 6. Grado de desordenamiento de las proteínas OpfiDHN-like, LepDHN-like y MabDHN-like de acuerdo con la plataforma Disprot (Sickmeier et al., 2007). (A) OpfiDHN-like; (B) LepDHN-like; (C) MabDHN-like; (D) OpsDHN1. Los valores mayores a 0,5 son regiones desordenadas de las proteínas.

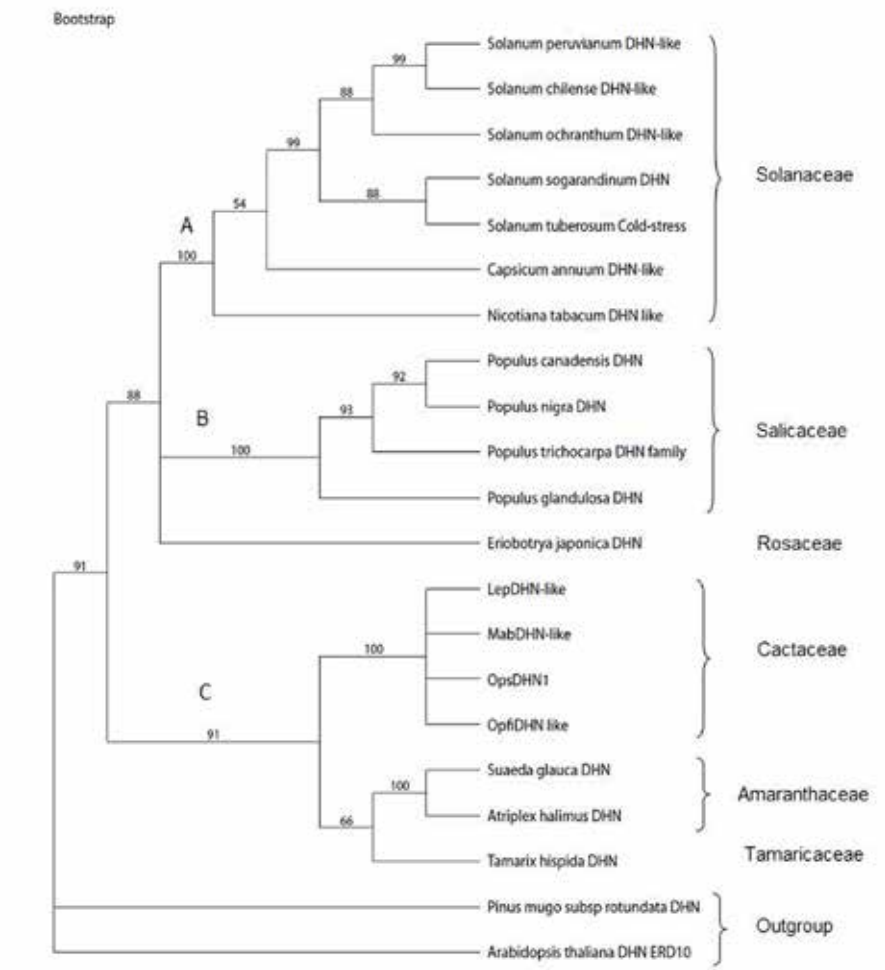

Fig. 7. Strict consensus tree of 32 equally parsimonious trees of 21 dehydrin-like proteins including the three cacti species of this study. Bootstrap values are given on the basis of the most parsimonious groups retained under the majority rule of 50%; the values are presented above branches.

Fig. 7. Árbol consenso estricto de 32 árboles igualmente parsimoniosos de 21 proteínas tipo dehidrina incluyendo las cactáceas analizadas. Los valores del Bootstrap están dados en base a los grupos más parsimoniosos retenidos bajo la regla de la mayoría del 50%, los valores se presentan por arriba de las ramas.

environments, since it can form an amphiphilic α-helix which allows binding to other partially denatured proteins (Qiu et al., 2014).

Furthermore, we identified a histidine (H)-rich motif that has also been found in dehydrins from other species, such as *Saussurea involucrata* (Qiu et al., 2014), *Ricinus communis* (Krüger et al. 2002) and *Opuntia streptacantha* (Hernández-Sánchez et al., 2015). According to Hara et al. (2005) this H-segment has shown binding with metal ions in their experiments with the CuCoR15 dehydrin protein of *Citrus unshiu,* which has two H-together in their sequence that contributed to a stronger metal ion binding than those proteins that have separated H in the same domain. Our results showed that the virtual translation of OpfiDHN-like, MabDHN-like and LepDHN-like have from 3 to 6 H-together (Fig. 2), suggesting they could play a role in the protection against high metal ion concentrations or also as metal ion stabilizers. Biochemical approaches are needed to confirm this assumption. Hernández-Sánchez et al. (2015) demonstrated that deletion of the regions that contain the histidine- and serine-rich motifs in OpsDHN1 protein affect its nuclear localization; the histidine-rich motif of OpfiDHN-like, LepDHN-like and MabDHN-like may be involved in the same function as they are very similar to the OpsDHN1.

After the conserved motif analysis search (Fig. 3) we identified at least two K segment repetitions (represented as motifs 1 and 3) in our cacti and homologous dehydrin sequences (Table 2). This was confirmed during the conserved domain analysis (Fig. 4), in which all the analyzed cacti proteins displayed three Dehydrin 2 domain repetitions, corresponding to K segments that have been found also in *Arabidopsis thaliana*, *Gossypium* sp, and *Pisum sativum* (Roberton & Chandler, 1992).

The hydropathicity scale proposed by Kyte and Doolittle (1982) have been used by many bioinformatic platforms to determine the charge-hydropathy (C-H) plot of amino acid sequences. This simple C-H plot largely separates Intrinsically Disordered Proteins (IDPs) from structured proteins (Huang et al., 2014). Figure 5 shows that OpfiDHN- like, MabDHN-like and LepDHN- like are hydrophilic proteins because according to the Kyte and Doolittle (1982) scale, the smaller the hydropathicity and GRAVY values (Table 2), the more hydrophilic the protein is. The hydrophilic property of the LEA 2 proteins is due to a high number of charged and polar amino acids, a low number of non-polar amino acids, which are mostly hydrophobic and also to the absence of tryptophan (W) and cysteine (C), which is the reason that they are also called hydrophilins (Reyes et al., 2005; Battaglia et al., 2008). Hydrophilins are able to interact with water or other polar/charged small molecules, and thus could share a common physiological role in dehydration tolerance (Jaspard & Hunault, 2014). In an aqueous solution, hydrophilic LEA proteins have a high degree of structure disorder, so they are considered intrinsically disordered proteins (IDP) (Mouillon et al., 2006). Most of LEA proteins are IDP and little is known about their molecular mechanism of action, although *in vitro* assays with various LEA proteins suggested roles in desiccation and/or freezing aggregation or membrane protection (Jaspard & Hunault, 2014). Concerning this parameter, the putative amino acid sequences of our dehydrins (Fig 6) had a high disorder degree (above 0.5) in agreement with Radivojac et al. (2007).

Dehydrins can be classified as neutral, acidic or basic, depending on their amino acid sequences (Alsheikh et al., 2005). When our sequences were compared with OpsDHN1, which is acid by nature (Ochoa-Alfaro et al., 2012), it could be observed that these proteins correspond to acid dehydrins based on their higher content of E and D residues (between 37 and 39 out of 143 to 148 total amino acid residues) (Table 2). The motif analysis (Fig. 3) revealed the presence of a rich K and E motif (represented as motif 2 in Fig. 3) in most of the homologous proteins, which allows dehydrins to interact with DNA molecules according to Tunnacliffe and Wise (2007). Similar motifs have been recently reported in proteins such as those of *Solanum habrochaites* ShDHN (Liu et al., 2015). However, OpfiDHN-like, LepDHN-like and MabDHN-like did not have an S-segment or a motif 2, possibly because they were partial sequences since they are between 96 and 97% similar to OpsDHN1 in their peptide sequence (Table 1) and is very probable that this motifs are absent because the sequence is not complete.

The strict consensus tree, constructed from 32 most parsimonious trees, showed the presence of 3 clades (A, B and C), in which OpfiDHN- like, OpsDHN1, LepDHN- like and MabDHN- like formed a monophyletic subclade within clade C; all the other analyzed families presented this monophyletic arrangement (Fig. 7). In the multiple alignment of amino acids (Fig. 2) some differences in the H-rich motif were observed. In the Opuntioideae subfamily (OpfiDHN-like and OpsDHN1) 6 H were present in this motif, whereas only 3 H were in the Cactoideae subfamily (LepDHN-like and MabDHN-like). In the multiple nucleotide alignment we can observe a gap of 9 nucleotides in the 699-707 positions corresponding to the absence of these histidines (data not shown).

The loss of these H might be due to some adaptive mechanism induced by selective pressure of the specific ecological niche from where diversification to different subfamilies and species occurred; according to Hernández-Hernández et al. (2014) the high diversification rates that characterize genera such as *Opuntia, Mammillaria* and *Echinocereus* might be associated with their geographic expansion during the recent aridification of North America (particularly the expansion of the Chihuahuan Desert) during the Miocene. It has been shown that aridification can shape the evolution in functional traits (Livshultz et al., 2011). Subfamilies Opuntioideae and Cactoideae have long been recognized as monophyletic on the basis of morphological and molecular data. Studies based on molecular data show that Opuntioideae subfamily is characterized by a deletion in the *accD* region of the chloroplast genome (Wallace & Dickie, 2002; Griffith & Porter, 2009). Subfamily Cactoideae has been also strongly supported as monophyletic in molecular phylogenies (Nyffeler, 2002) and the members of this subfamily are characterized by the loss of an intron in the *rpoC1* chloroplast gene (Wallace & Cota, 1994, Hernández-Hernández et al., 2011), while members of the Opuntioideae and Pereskioideae subfamilies have it (Wallace & Cota, 1994). However in our consensus tree (Fig 7) this separation of these subfamilies can not be observed probably because a more complete sequence is needed or the number of changes is not sufficient to assert a possible evolution using these dehydrins as model.

Cacti dehydrins were closely related to the Amaranthaceae (*Sueda glauca* DHN and *Atriplex halimus* DHN) and Tamaricaceae (*Tamarix hispida* DHN) dehydrins families (Fig. 7); these two plant families are evolutionarily close because all belong to the Caryophyllales order (Cronquist, 1981, The Angiosperm Phylogeny Group, 2009).

Using PCR techniques, we identified three dehydrin-like gene fragments in *Opuntia ficus-indica*, *Leuchtenbergia prin-*

Table 2. Total and partial sequences of dehydrin-like proteins homologous to OpfiDHN- like, LepDHN-like and MabDHN-like.
Tabla 2. Secuencias totales y parciales de proteínas tipo dehidrina homólogas a OpfiDHN- like, LepDHN-like y MabDHN-like.

Taxa	UNIPROT KB Accesion Number	Name	Sequence type	Family
Suaeda glauca	H8YHV1	Dehydrin	Total	Amaranthaceae
Atriplex halimus	U5Y959	Dehydrin	Total	Amaranthaceae
Arabidopsis thaliana	F4HST2	Dehydrin ERD 10	Total	Brassicaceae
Opuntia streptacantha	G9B6J3	Dehydrin 1	Total	Cactaceae
Pinus mugo subsp *rotundata*	F1BQ11	Dehydrin 1 protein (Fragment)	Parcial	Pinaceae
Eriobotrya japonica	B8Y3W6	Dehydrin 2	Total	Rosaceae
Populus canadensis	A7L2U5	Dehydrin	Total	Salicaceae
Populus nigra	A7L2U4	Dehydrin	Total	Salicaceae
Populus trichocarpa	A9PA80	Dehydrin family protein	Total	Salicaceae
Populus glandulosa	A7L2U1	Dehydrin	Total	Salicaceae
Solanum sogarandinum	Q7Y1A0	25 kDa protein dehydrin	Total	Solanaceae
Nicotiana tabacum	Q76MG1	Dehydrin (Fragment)	Parcial	Solanaceae
Capsicum annuum	Q6XLQ1	Dehydrin-like protein	Parcial	Solanaceae
Solanum peruvianum	E5F397	Dehydrin (Fragment)	Parcial	Solanaceae
Solanum chilense	E5F352	Dehydrin (Fragment)	Parcial	Solanaceae
Solanum tuberosum	O04232	Cold-stress inducible protein	Total	Solanaceae
Solanum ochranthum	E5F3B1	Dehydrin (Fragment)	Parcial	Solanaceae
Tamarix hispida	C0KTL3	Dehydrin	Total	Tamaricaceae

Table 3. Parameters of dehydrin-like proteins.
Tabla 3. Parámetros de las proteínas tipo dehidrina.

Parameters	OpfiDHN- Like	LepDHN-Like	MabDHN-Like	OpsDHN1
aa[1] total number	148	143	145	248
Acid aa number (E[4], D[5])	39	37	37	66
Basic aa number (K[2], R[3])	31	23	24	46
GRAVY[6]	-1.624	-1.331	-1.37	-1.53
Molecular weight	17025.0	16138.9	16370.1	28362.1
kD	17.02	16.61	16.36	28.6
Isoelectric Point	5.92	5.22	5.23	5.43

aa[1]= amino acid; K[2]= Lysine; R[3]= Arginine; E[4]= Glutamic acid; D[5]= Aspartic acid; GRAVY[6]= grand average of hydropathicity.
aa[1]=aminoácidos; K[2]= Lisina; R[3]= Arginina; E[4]= Ácido Glutámico; D[5]= Ácido Aspártico; GRAVY[6]= promedio general de hidropatía.

cipis and *Mammillaria bombycina* with the sufficient base pairs to identify them by *in silico*. The three putative amino acid sequences showed the presence of 3 K segments, typical of LEA 2 proteins and a histidine-rich segment. Parameter analysis showed that these peptides were acidic, highly hydrophilic and intrinsically disordered. The gene tree showed that dehydrins of different cacti genera have a monophyletic origin, as well as in other plant families. Also, we found a probable evolutionary relationship between different members of analyzed subfamilies, because the representatives of the Cactoideae subfamily exhibited a 3 histidine deletion in the H motif of their partial amino acid sequence; this deletion is not present in the representatives of the Opuntioideae subfamily. However, in the consensus tree this subfamilies separation cannot be observed probably because the number of changes is not sufficient to assert a possible evolution using these dehydrins as a model.

ACKNOWLEDGMENTS

This work was supported by the Ministry of Education (SEP; Programa de Fortalecimiento de Cuerpos Académicos). We want to thank to the National Science and Technology Council of Mexico (CONACYT) for the PhD grant given to SHC.

REFERENCES

Alsheikh, M.K., J.T. Svensson & S.K. Randall (2005). Phosphorylation regulated ion-binding is a property shared by the acidic subclass dehydrins. *Plant Cell and Environment* 28: 1114-1122.

Amara, I., A. Odena, E. Oliveira, A. Moreno, K. Masmoudi, M. Pagès & A. Goday (2012). Insights into Maize LEA Proteins: From Proteomics to Functional Approaches. *Plant Cell Physiology* 53: 312-329.

Bailey, T.L., M. Bodén, F.A. Buske, M. Frith, C.E. Grant, L. Clementi, J. Ren, W.W. Li & W.S. Noble (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research* 37: 202-208.

Bassett, C.L., K.M. Fisher & R.E. Farrell Jr (2015). The complete peach dehydrin family: characterization of three recently recognized genes. *Tree Genetics & Genomes* 11: 1-14.

Battaglia, M., Y. Olvera-Carrillo, A. Garciarrubio, F. Campos & A.A. Covarrubias (2008). The enigmatic LEA proteins and other hydrophilins. *Plant Physiology* 148: 6-24.

Bies-Ethève, N., P.G. Comella, A. Debures, E. Lasserre, E. Jobet, M. Raynal, R. Cooke & M. Delseny (2008). Inventory, evolution and expression profiling diversity of the LEA (Late Embryogenesis Abundant) protein gene family in *Arabidopsis thaliana*. *Plant Molecular Biology* 67: 107-124.

Calvente, A., D.C. Zappi, F. Forest & L.G. Lohmann (2011). Molecular phylogeny of tribe Rhipsalideae (Cactaceae) and taxonomic implications for Schlumbergera and Hatiora. *Molecular Phylogenetics and Evolution* 58: 456-468.

Cronquist, A. (1981). An integrated system of classification of flowering plants. Columbia University Press. New York. Connaissance Sur La Flore Du Maroc 91, pp. 1984-1989.

Drira, M., W. Saibi, I. Amara, K. Masmoudi, M. Hanin & F. Brini (2015). Wheat Dehydrin K-Segments Ensure Bacterial Stress Tolerance, Antiaggregation and Antimicrobial Effects. *Applied Biochemistry and Biotechnology* 175: 3310-3321.

Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39: 783-791.

Garay-Arroyo, A., J. M. Colmenero-Flores, A. Garciarrubio & A. A. Covarrubias (2000). Highly hydrophilic proteins in prokaryotes and eukaryotes are common during conditions of water deficit. *Journal of Biological Chemistry* 275: 5668-5674.

Gasteiger, E., C. Hoogland, A. Gattiker, S. Duvaud, M.R. Wilkins, R.D. Appel & A. Bairoch (2005). Protein Identification and Analysis Tools on the ExPASy Server. In: John M. Walker (ed), pp. 571-607. The Proteomics Protocols Handbook, Humana Press.

George, S., B. Usha & A. Parida (2009). Isolation and characterization of an atypical LEA protein coding cDNA and its promoter from drought-tolerant plant *Prosopis juliflora*. *Applied Biochemistry and Biotechnology* 157: 244-253.

Griffith, M.P. & J.M. Porter (2009). Phylogeny of Opuntioideae (Cactaceae). *International Journal of Plant Sciences* 170: 107-116.

Hara, M., M. Fujinaga & T. Kuboi (2005). Metal binding by citrus dehydrin with histidine-rich domains. *Journal of Experimental Botany* 56: 2695-2703.

Hernández-Hernández, T., H.M. Hernández, J.A. De-Nova, R. Puente, L.E. Eguiarte & S. Magallón (2011). Phylogenetic relationships and evolution of growth form in Cactaceae (Caryophyllales, Eudicotyledoneae). *American Journal of Botany* 98: 44-61.

Hernández-Hernández, T., J.W. Brown, B. O. Schlumpberger, L.E. Eguiarte & S. Magallón (2014). Beyond aridification: multiple explanations for the elevated diversification of cacti in the New World Succulent Biome. *New Phytologist* 202: 1382-1397.

Hernández-Sánchez, I.E., I. Maruri-López, A. Ferrando, J. Carbonell, S.P. Graether & J.F. Jiménez-Bremont (2015). Nuclear localization of the dehydrin OpsDHN1 is determined by histidine-rich motif. *Frontiers in Plant Science* 6: 1-8

Huang, F., C.J. Oldfield, B. Xue, W.L. Hsu, J. Meng, X. Liu & A.K. Dunker (2014). Improving protein order-disorder classification using charge-hydropathy plots. *BMC bioinformatics* 15: S4.

Jaspard E. & G. Hunault (2014). Comparison of Amino Acids PhysicoChemical Properties and Usage of Late Embryogenesis Abundant Proteins, Hydrophilins and Why Domain. *PLoS ONE* 9.

Krüger, C., O. Berkowitz, U.W. Stephan & R. Hell (2002). A metal-binding member of the late embryogenesis abundant protein family transports iron in the phloem of *Ricinus communis* L. *Journal of Biological Chemistry* 277: 25062-25069.

Kyte, J. & R.F. Doolittle (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology* 157: 105-132.

Li, X. & J. Cao (2015). Late Embryogenesis Abundant (LEA) Gene Family in Maize: Identification, Evolution, and Expression Profiles. *Plant Molecular Biology Reporter*: 1-14.

Liu, S., H. Zhuanfang, W. Jianfeng, L. Mingshun, Z. Degui, P. Guangtang, Z. Shihuang & L. Xinhai (2015). Identification of two functional markers associated with drought resistance in maize. *Molecular Breeding* 35: 1-10.

Livshultz, T., J.V Mead, D.J. Goyder & M. Brannin (2011). Climate niches of milkweeds with plesiomorphic traits (Secamonoideae; Apocynaceae) and the milkweed sister group link ancient African climates and floral evolution. *American Journal of Botany* 98: 1966-1977.

Lopez, C.G., G.M. Banowetz, C.J. Peterson & W. E. Kronstad (2003). Dehydrin expression and drought tolerance in seven wheat cultivars. *Crop Science* 43: 577-582.

Mouillon, J.M., P. Gustafsson & P. Harryson (2006). Structural investigation of disordered stress proteins. Comparison of full-length dehydrins with isolated peptides of their conserved segments. *Plant Physiology* 141: 638-650.

Murashige, T. & F. Skoog (1962). A revised medium for rapid growth and bioassays with tabacco cultures. *Physiologia Plantarum* 15: 473-497.

Murray, M.G. and W.F. Thompson (1980). Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Research* 8: 4321-4326.

Nyffeler R. (2002). Phylogenetic relationships in the cactus family (Cactaceae) based on evidence from trnK/matK and trnL- trnF sequences. *American Journal of Botany* 89: 312-326.

Ochoa-Alfaro, A.E., M. Rodríguez-Kessler, M.B. Pérez-Morales, P. Delgado-Sánchez, C.L. Cuevas-Velazquez, G. Gómez-Anduro & J.F. Jiménez-Bremont (2012). Functional characterization of an acidic SK3 dehydrin isolated from an *Opuntia streptacantha* cDNA library. *Planta* 235: 565-578.

Qiu, H., L. Zhang, C. Liu, L. He, A. Wang, H.L. Liu & J.B. Zhu (2014). Cloning and characterization of a novel dehydrin gene, SiDhn2, from *Saussurea involucrata* Kar. et Kir. *Plant Molecular Biology* 84: 707-718.

Radivojac, P., L.M. Iakoucheva, C.J. Oldfield, Z. Obradovic, V.K. Uversky & A.K. Dunker (2007). Intrinsic disorder and functional proteomics. *Biophysical Journal* 92: 1439-1456.

Reyes, J.L., F. Campos, H. Wei, R. Arora, Y. Yang, D.T. Karlson & A.A. Covarrubias (2008). Functional dissection of hydrophilins during *in vitro* freeze protection. *Plant Cell and Environment* 31: 1781-1790.

Reyes, J.L., M.J. Rodrigo, J.M. Colmenero-Flores, J.V. Gil, A. Garay-Arroyo, F. Campos, F. Salamini, D. Bartels & A.A. Covarrubias (2005). Hydrophilins from distant organisms can protect enzymatic activities from water limitation effects in vitro. *Plant, Cell & Environment* 28: 1365-3040.

Roberton, M. & P.M. Chandler (1992). Pea dehydrins: identification, characterization and expression. *Plant Molecular Biology* 19: 1031-1044

Rodziewicz, P., B. Swarcewicz, K. Chmielewska, A. Wojakowska & M. Stobiecki (2014). Influence of abiotic stresses on plant proteome and metabolome changes. *Acta Physiologiae Plantarum* 36: 1-19.

Shakirova, F.M., A.R. Sakhabutdinova, M.V. Bezrukova, R.A. Fatkhutdinova & D.R. Fatkhutdinova (2003). Changes in the hormonal status of wheat seedlings induced by salicylic acid and salinity. *Plant Science* 164: 317-322.

Shen, Y., M.J. Tang, Y.L. Hu & Z.P. Lin (2004). Isolation and characterization of a dehydrin-like gene from drought-tolerant *Boea crassifolia*. *Plant Science* 166: 1167-1175.

Shih, M.D., F.A. Hoekstra & Y.I.C. Hsing (2008). Late Embryogenesis Abundant Proteins. *Advances in Botanical Research* 48: 211-255. doi:10.1016/S0065-2296(08)00404-7

Sickmeier, M., J.A. Hamilton, T. LeGall, V. Vacic, M.S. Cortese et al. (2007). DisProt: the database of disordered proteins. *Nucleic Acids Research* 35: D786-D793.

Siddiqui, N.U., H.J. Chung, T.L. Thomas & M.C. Drew (1998). Abscisic acid dependent and independent expression of the carrot late embryogenesis abundant class gene Dc3 in transgenic tobacco seedlings. *Plant Physiology* 118: 1181-1190.

Sun, J., L. Nie, G. Sun, J. Guo & Y. Liu (2013). Cloning and characterization of dehydrin gene from *Ammopiptanthus mongolicus*. *Molecular Biology Reports* 40: 2281-2291.

Swofford, D.L. (2003). PAUP* Phylogenetic Analysis Using Parsimony (*and Other Methods), v. 4.0 beta 10. Sunderland: Sinauer Associates.

THE ANGIOSPERM PHYLOGENY GROUP (2009). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society* 161: 105-121.

Tunnacliffe, A. & M.J. Wise (2007). The continuing conundrum of the LEA proteins. *Naturwissenschaften* 94: 791-812.

Wallace R.S. & S.L. Dickie (2002). Systematic implications of chloroplast DNA sequence variation in subfam. Opuntioideae (Cactaceae). *Succulent Plant Research* 6: 9-24.

Wallace, R.S. & J.H. Cota (1996). An intron loss in the chloroplast generpoC1 supports a monophyletic origin for the subfamily Cactoideae of the Cactaceae. *Current genetics* 29: 275-281.

Yang, Y., M. He, Z. Zhu, S. Li, Y. Xu, C. Zhang, S.D. Singer & Y. Wang (2012). Identification of the dehydrin gene family from grapevine species and analysis of their responsiveness to various forms of abiotic and biotic stress. *BMC Plant Biology* 12: 140.