Theor Appl Genet (2003) 106:811–818
DOI 10.1007/s00122-002-1093-z

L. Natali · T. Giordani · A. Cavallini

Sequence variability of a dehydrin gene within *Helianthus annuus*

Received: 5 April 2002 / Accepted: 5 August 2002 / Published online: 31 October 2002

Abstract Dehydrins are proteins produced during the late stages of plant embryo development and following any environmental stimulus involving dehydration. In order to investigate the variability of a dehydrin-encoding gene (*Dhn1*) in cultivated and wild sunflower (*Helianthus annuus*) genotypes, near-complete alleles were isolated by the polymerase chain reaction and sequenced. All of the isolated sequences were found to contain the typical dehydrin domains, and interrupted by an intron. The number of nucleotide substitutions and indels per site was calculated. With respect to the overall sequence, variation in both the coding and noncoding [intron and 3′-UTR (untranslated region)] sequences was much larger among wild accessions than among cultivars. No variation was observed in 3′-UTRs from cultivated sunflowers. Different coding regions showed a different numbers of synonymous and nonsynonymous substitutions. The Y and K domains were the most conserved in both wild and cultivated genotypes. Sequence analysis of the deduced dehydrin proteins showed that nucleotide substitutions in wild accessions should also determine large biochemical differences at the protein level. All of the isolated alleles were however functional, at least at the transcription level. To our knowledge these are the first data on intraspecific genetic variability of such a stress response gene. The low variability of dehydrin genes from cultivated sunflower is discussed in relation to the origin of sunflower cultivars. The possibility of rescuing general genetic variability through crosses to wild accessions of *H. annuus* rather than using wild *Helianthus* species is also discussed.

Keywords Dehydrin · *Helianthus annuus* · Sunflower · DNA sequence · Nucleotide substitution · Genetic diversity

Communicated by C. Möllers

L. Natali · T. Giordani · A. Cavallini (✉)
Department of Agricultural Plant Biology, Genetics Section,
Via Matteotti 1/B, I-56124 Pisa, Italy
e-mail: acavalli@agr.unipi.it.
Fax: +39-050-576750

Introduction

Dehydrins are an immunologically distinct family of proteins, also known as the *Lea D11* subgroup of late-embryogenesis-abundant (*Lea*) proteins (Dure et al. 1989), and have been described in many angiosperm and gymnosperm species (Close 1997). They are characterised by several domains, including one or more putative amphipatic α-helix-forming consensus regions at the C-terminus (Godoy et al. 1994) and, in the majority of dehydrins, there is a region at the N-terminus with homologies to a portion of the nucleotide-binding site of chaperones of plants and bacteria (Martin et al. 1993). Many dehydrins also contain a tract of serine residues, possibly phosphorylatable (Goday et al. 1994). Dehydrins are usually produced by plants during the late stages of embryo development (Dure et al. 1989), but they are also synthesised following any environmental stimulus involving dehydration, such as drought or cold stress and salinity, as key components of dehydration tolerance (Close 1996; Zhu et al. 2000). However, their function has not yet been demonstrated, though many studies indicate that dehydrins are associated with macromolecules such as nucleoprotein complexes in the nucleus (Godoy et al. 1994) and endomembranes in the cytoplasm (Schneider et al. 1993). This suggests that these proteins are surfactants that inhibit the coagulation of a range of macromolecules and preserve their structural integrity (Close 1996).

Variability in dehydrin genes has been studied in different species. This type of analysis is complicated since dehydrins belong to a multigene family. For example, many different genes have been isolated from barley and from *Pisum sativum* and very few differences were found among sequences from within a single species (Grosselindemann et al. 1998; Lang et al. 1998; Choi et al. 1999). The recent finding of a dehydrin mutant allele demonstrated the usefulness of the normal protein in determining chilling tolerance of *Vigna unguiculata* plantlets (Ismail et al. 1999) and indicated the necessity of studying the genetic variability of dehydrins even for plant breeding purposes.

Table 1 Accession name and provenance of sunflower genotypes, GenBank accession number of the isolated *HaDhn1* sequence and putative structure of its encoded dehydrin according to Close (1997)

Accession name	Provenance[a]	Accession number	Structure
Arizona	NCRPIS, Ames, USA	AJ250224	Y_3SK_2
Colorado	NCRPIS, Ames, USA	AJ250225	Y_3SK_2
Durango	INIFAP, Durango, Mexico	AJ249273	Y_2SK_2
Iowa	NCRPIS, Ames, USA	AJ250226	Y_3SK_2
Saskatchewan	NCRPIS, Ames, USA	AJ250227	Y_3SK_2
Texas	NCRPIS, Ames, USA	AJ438979	Y_3SK_2
Washington	NCRPIS, Ames, USA	AJ250228	Y_3SK_2
HCM line	DBPA, Pisa, Italy	AJ002741	Y_3SK_2
cv. Adonis	U.S. hybrid	AJ249269	Y_3SK_2
cv. Primasol	French hybrid	AJ438980	Y_3SK_2
cv. Cernianka	Russian open pollinated cv.	AJ249270	Y_3SK_2
cv. RomsunHS53	Rumanian hybrid	AJ249271	Y_3SK_2
cv. NSH43	Yugoslavian hybrid	AJ249272	Y_3SK_2

[a] NCRPIS, North Central Regional Plant Introduction Station; INIFAP, Instituto Nacional de Investigaciones Forestales y Agropecuarias; DBPA, Dipartimento di Biologia delle Piante Agrarie

A dehydrin cDNA, *HaDhn1*, induced by drought stress, was isolated and sequenced in sunflower (Ouvrard et al. 1996), and the accumulation of its transcripts has been correlated with drought tolerance (Cellier et al. 1998). A sequence allelic to this gene, *HaDhn1a*, was found to be expressed in the latest stages of *Helianthus annuus* embryogenesis, depending on abscisic acid (ABA) accumulation; moreover, *HaDhn1a* transcripts accumulated after drought stress even in ABA-deficient sunflower mutants (Giordani et al. 1999).

Analyses of genetic polymorphism at the DNA sequence level within plant species are mostly related to alcohol dehydrogenase genes (see Lin et al. 2001). Studies on DNA sequence polymorphism of genes related to abiotic stress are even more rare despite the importance of the latter in environmental adaptation process of plants. In sunflower, genetic polymorphism has been studied by means of restriction fragment length polymorphism or by polymerase chain reaction (PCR)-based molecular markers, and no studies are presently available on DNA sequence variation in protein-encoding sequences. In order to investigate the variability of a dehydrin gene in *Helianthus annuus*, we isolated, by PCR, and sequenced a DNA fragment homologous to *HaDhn1* from different cultivated and wild sunflower genotypes. In this paper we report data on variability of this gene, at both DNA and deduced protein levels. The comparison of genetic diversity between cultivated and wild sunflowers may help in choosing breeding strategies for sunflower.

Materials and methods

Plant material and DNA isolation

The accessions used for this study and their provenance are reported in Table 1. For cultivated genotypes, one cultivar or hybrid per country, randomly chosen from different countries in which sunflower is a major crop, was analysed. Seeds were germinated and plantlets grown to maturity in the field. For each accession, one plant was collected. DNA was extracted from leaf tissues according to the method devised by Doyle and Doyle (1989) with minor modifications (Giordani et al. 1999).

For some experiments adult leaves were detached from plants and subjected to desiccation until 60% fresh weight. RNA was then isolated from the leaves according to Giordani et al. (1999).

Isolation of genomic *Dhn1* sequences by PCR

DNA sequences homologous to *HaDhn1a* were isolated by PCR on genomic DNAs obtained from the different accessions. PCR was performed using two oligonucleotides based on the published sequence of sunflower dehydrin cDNA: 5′-GCAAACTACGGAGGAGATAA-3′ (sense) and 5′-GTGAAACCACATACAAAACAAA-3′ (antisense). Sequences were amplified using 100 ng of genomic DNA as a template; thermocycling was performed at 94 °C for 30 s, 60 °C for 60 s and 72 °C for 60 s, for 30 cycles, using *Taq*-DNA polymerase (Promega, Madison, Wis.).

The amplified fragmentswere cloned into a pGEM-T Easy plasmid vector (Promega). The cloned fragments were sequenced by the dideoxy chain termination method using the PRISM dye terminator cycle sequencing kit (Perkin-Elmer, Foster City, Calif.) according to the manufacturer's instructions; sequences were analysed using the SEQUENCING ANALYSIS 2.1.2 (Perkin-Elmer) and SEQUENCHER 3.0 analysis programs (Gene Codes Corporation). For the HCM line, amplified fragments were cloned, and 50 clones were analysed on a 6% DNA sequencing gel to verify sequence length uniformity; of these clones, 15 were sequenced. For the other genotypes, one clone was analysed. Each clone was analysed twice to exclude sequencing errors.

Analysis of *Dhn1* expression by reverse transcription (RT)-PCR

For total RNA extraction, tissues were ground in liquid nitrogen with a MES-guanidine hydrochloride-containing buffer, and RNA was isolated using two phenol/chloroform extractions and an acetic acid-ethanol precipitation procedure (Giordani et al. 1999). RT-PCR was performed using the sense oligonucleotide as above, coupled with a specific antisense oligonucleotide (5′-TTCTGCATCGTTCTGGAAGT-3′). The use of these primers enabled amplification of dehydrin fragments of convenient length so that differences were apparent on the agarose gels.

Total RNA (1 µg) was heated for 3 min at 70 °C and retrotranscribed in a 25-µl volume reaction using 400 µ*M* of each deoxynucleotide triphosphate, 0.25 µ*M* antisense oligonucleotide as primer, 1× RT-Buffer (Gibco BRL, Gaithersburg, Md.,), 1 m*M* DTT, 100 U M-MLV reverse transcriptase (Gibco BRL). Retrotranscribed cDNA was used for PCR amplification; PCR was performed using 1/100 volume of retrotranscribed cDNA, 2.5 m*M* $MgCl_2$, 250 µ*M* of each deoxynucleotide triphosphate, 1 µ*M* of each primer, 1× Thermophilic buffer (Promega), 2.5 U *Taq*-DNA polymerase (Promega). Thermocycling was performed as above. PCR products were then visualised on agarose gel after EtBr staining.

Southern blotting and hybridisation

Southern blotting of DNA was performed according to standard protocols (Sambrook et al. 1989). Eight micrograms of DNA from different *Helianthus* species was digested with *Eco*RI and *Hin*dIII

restriction endonucleases in a fivefold excess according to the instructions of the suppliers (Roche, Indianapolis, In), then electrophoresed on agarose gels and blotted onto nylon membrane (Roche). Complete digestion was checked by including unmethylated bacteriophage lambda DNA which, when digested with *Eco*RI plus *Hin*dIII (DNA molecular size marker III; Roche), was also used as a fragment size marker.

Hybridisations were performed using digoxigenin-labeled *HaDhn1a* cDNA as a probe, which has been previously cloned (Giordani et al. 1999), under high stringency conditions, at 50 °C in 50% formamide, 5× SSC (1× SSC is 0.15 *M* NaCl, 0.015 *M* trisodium citrate, pH 7.0), 2% blocking reagent (Roche), 0.02% SDS, 0.1% SLS. Filters were washed twice in 2× SSC, 0.1% SDS for 15 min at room temperature, once in 1× SSC, 0.1% SDS for 30 min at 68 °C and once in 0.3× SSC, 0.1% SDS for 30 min at 68 °C. Digoxigenin haptens in DNA-DNA hybrids were detected using a Dig-DNA detection kit (Roche) by enzyme-linked immunoassay using an antidigoxigenin-alkaline phosphatase conjugate.

DNA sequence analysis

The DNA sequence was subdivided in ten subregions according to the functional characteristics of the putatively encoded protein region. Intron delimitation within genomic sequences was carried out by comparing the genomic sequences with the corresponding cDNA and confirmed using the program FEX (Baylor College of Medicine, Tex.).

Overall sequences and sequence segments were aligned using CLUSTAL W (Thompson et al. 1994). Some adjustments were made by eye.

Statistics of intraspecific polymorphism within *H. annuus* were performed using the DNASP program version 3.51 (Rozas and Rozas 1999). Nucleotide diversity (π, i.e. the average number of nucleotide differences per site) and its sampling variance were calculated according to Nei (1987), equations 8.4 and 8.12, replacing 2n by n. Numbers of synonymous and nonsynonymous substitutions per site were estimated for coding nucleotide sequences using the DNASP program as above, according to the method of Nei and Gojobori (1986). Alignment gaps were excluded from comparisons.

Relationships among dehydrin DNA sequences from different genotypes were investigated by the neighbor-joining (NJ) method (distance algorithm after Kimura), using the PHYLIP program package version 3.572 (Felsenstein 1989): with the SEQBOOT program, 1,000 versions of the original alignment were generated; then, trees were generated using the DNADIST and NEIGHBOR programs. A strict consensus tree was obtained from the available trees using the CONSENSE program.

Isoelectric points of the deduced proteins were calculated using the program COMPUTE PI/MW at the Expasy server (Switzerland), according to Wilkins et al. (1998). Hydrophobicity profiles were constructed by the program PROTSCALE, at the Expasy server, according to amino acid scale values by Kyte and Doolittle (1982), using a window size of nine amino acids, with a 100% relative weight of the window edges compared to the window centre. The predicted secondary structure of deduced proteins (percentage of α-helix, extended strand and random coil) was analysed using the program HNN at the Pole Bio-Informatique Lyonnais server (France).

Results and discussion

Sequences homologous to *HaDhn1a* were isolated by PCR from genomic DNA of 13 *H. annuus* genotypes. The primers used to isolate the *HaDhn1a* sequences in the present investigation correspond to sequence portions specific to this gene and do not belong to shared typical dehydrin domains. Primers were designed to obtain one specific DNA fragment by PCR: after amplification, 50 clones were first analysed for inserted fragment length in high-resolution gel; no length variation was found (data not shown). Then, 15 clones were sequenced and all inserted fragments corresponded to the *HaDhn1a* sequence, showing no sequence variation. Therefore, we deduced that, under the experimental conditions used (primers, stringency, temperature), only one product could be obtained using that primer pair. Moreover, after digestion with *Eco*RI or *Hin*dIII and Southern blot hybridisations with digoxigenin-labeled *HaDhn1a* only single bands were observed (Fig. 1), further supporting the hypothesis that *HaDhn1a*, though belonging to a gene family, identifies a single-copy gene.

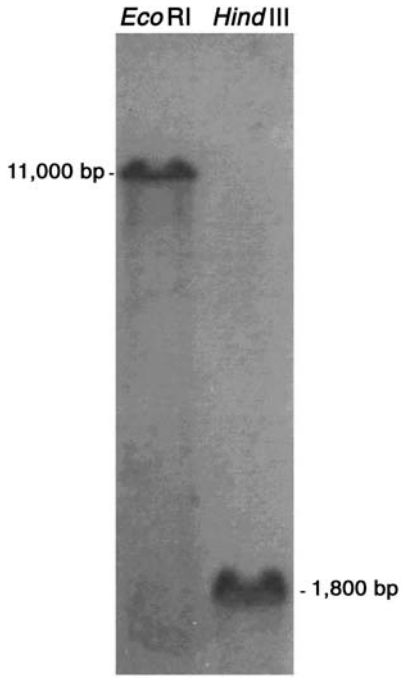

Fig. 1 Southern blot of genomic DNA of *Helianthus annuus* (HCM line), digested with *Hin*dIII or *Eco*RI, hybridised to digoxigenin-labeled *Hadhn1a*. Molecular weights of bands (in basepairs) are reported

The GenBank accession number of the sequences isolated are reported in Table 1 together with their structure, according to the nomenclature proposed by Close (1997).

The nucleotide diversity (π), i.e. the average number of nucleotide differences per site (Nei 1987) in coding and noncoding regions of the dehydrin gene are reported in Table 2, calculated excluding sites subjected to insertions or deletions. With respect to the overall sequence and coding region, a two-three fold higher uniformity was observed among cultivars than among wild accessions. As far as noncoding regions, the introns showed generally a larger variability than the 3′-UTR (untranslated region) (Table 2). Quite surprisingly, substitutions in 3′-UTRs of sunflower cultivars are extremely rare, indicating the importance, in the tested cultivated geno-

Table 2 Nucleotide diversity (π, the average number of nucleotide differences per site) and its sampling variance (in parenthesis) of different regions of PCR-amplified *HaDhn1* sequences from *Helianthus annuus* (7 wild accessions and 6 cultivars)

Gene portion	Wild	Cultivars
Coding	0.01494 (0.00179)	0.00649 (0.00102)
Intron	0.05390 (0.00988)	0.02358 (0.00549)
3′-UTR	0.03121 (0.00925)	0.00262 (0.00169)
Overall	0.02336 (0.00243)	0.00841 (0.00138)

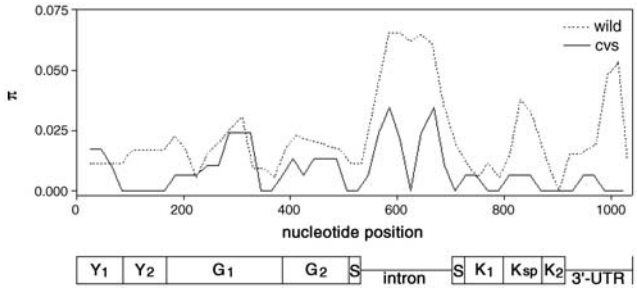

Fig. 2 Graphic representation of the pattern of change of nucleotide diversity along *Hadhn1* sequences from wild and cultivated genotypes and the corresponding schematic representation of the sunflower dehydrin. *Boxes* indicate different domains of the gene according to their encoded protein sequence: *Y* TDEYGNP-containing domains, *G* glycine-rich domains, *S* serine-rich domain, *K* lysine-rich domains, K_{sp} spacer between K regions

types, of conserving the 3′-UTR–i.e. beyond specific elements as FUEs and NUEs–involved in transcript processing (Rothnie 1996).

Data on sequence polymorphism in plant genes are rare. Values of *HaDhn1a* genes from wild genotypes are, for example, comparable to those reported for the *Adh3* locus in wild barley (0.0219 for overall sequence; Lin et al. 2001) and higher than those of other *Adh* loci of allogamous species (from 0.00204 to 0.01742 for the overall sequence; Cummings and Clegg 1998) and of a chitinase-encoding gene (i.e., involved in fungal response) of *Arabidopsis thaliana* (0.0104 for the entire region; Kawabe et al. 1997).

Nucleotide diversity was also calculated along the dehydrin DNA sequence, keeping separated sequences isolated from cultivated genotypes from those isolated from wild ones. Following alignment using the program CLUSTAL W, a 50-bp window was moved along the sequences in steps of 20 nucleotides. π was calculated in each window, and the value was assigned to the nucleotide at the midpoint of the window (Fig. 2); alignment gaps were not considered in the length of the windows. In Fig. 2, a schematic representation of the dehydrin gene, subdivided in ten nucleotide regions (see below) according to their encoded amino acid sequence, is also reported.

All of the sequences show the typical dehydrin-conserved domains and interrupted by an intron. At the N terminus, the Y domain contained the TDEYGNP motifs, which are similar to portions of plant and bacterial chaperones with an affinity to nucleic acids (Martin et al. 1993). All analysed genotypes, except one, show three TDEYGNP repetitions. Only two TDEYGNPs are found in the Durango wild accession. It is to be noted that two TDEYGNPs are generally found in dehydrin genes from wild *Helianthus* species (Giordani et al. in preparation). In cultivated genotypes the Y domain may be distinguished into two regions characterised by different genetic diversity (Fig. 2): Y_1 (that does not contain TDEYGNPs) and Y_2 (containing two, or three, TDEYGNPs). No diversity is observed in the Y_2 region in cultivated genotypes, while in wild accessions π is nearly constant along Y_1 and Y_2.

After the Y domain, a hydrophilic, glycine-rich domain may be subdivided in two parts: the first part (G_1) containing eight TGG-XG motifs and the second (G_2) in which these motifs are not apparently observed, though it is characterised by a high glycine content. Glycine-rich domains are probably involved in general dehydrin hydrophilicity (Close et al. 1989), though its function is far from clear. It is notable that just a deletion in one G-rich repetition determines a strong reduction of chilling tolerance in *Vigna unguiculata* plantlets (Ismail et al. 1999), indicating the importance of these segments in dehydrin function. The G domain is the most variable (excluding the intron) in cultivated genotypes (Fig. 2).

After the G regions, a serine-rich domain (domain S) was observed, with eight serine residues, whose nucleotide sequence is divided into two parts by an intron (Fig. 2). It has been hypothesised that the polyserine trait is a phosphorylatable site, possibly related to the translocation of dehydrins from the cytoplasm to the nucleus where they could interact with nucleic acid (Vilardell et al. 1990; Godoy et al. 1994). Probably because of its importance in the protein and due to the presence of the splicing site, no variations were observed in this domain. The intron is highly variable, and different profiles were observed between wild and cultivated genotypes.

At the C-terminus, there is a domain showing alternate hydrophilic and hydrophobic portions that should form an amphipatic α-helix (Segrest et al. 1990), probably involved in the stabilisation of partially denatured proteins (Hartl et al. 1994). This domain (K) contains two typical KIKEKLPG motifs in all of the analysed sequences and can be subdivided, on the basis of nucleotide diversity (Fig. 2), into K_1, containing the first KIKEKLPG, K_{sp}, a spacer between the two KIKEKLPG motifs, and K_2 region, with the second KIKEKLPG. Within the K domain, the spacer is the most variable in both wild and cultivated genotypes.

Nucleotide diversity per synonymous and nonsynonymous sites (π_s and π_a) was calculated for each subregion of the dehydrin gene (Table 3). With respect to the overall sequence, nonsynonymous substitutions in sunflower wild accessions seem to be relatively high, if we consider that we are studying intraspecific variation. Within sunflower cultivars, a few nonsynonymous substitutions

Table 3 Number of nucleotides, number of sites (excluding gaps and including stop codon), number of mutations, nucleotide diversity per site (π) from the total number of mutations, for synonymous and nonsynonymous sites, of different coding portions of PCR-amplified *HaDhn1* sequences from seven wild accessions and six cultivated genotypes of *H. annuus*

Gene portion	Number of nucleotides	Wild accessions						Cultivated genotypes					
		Synonymous sites			Nonsynonymous sites			Synonymous sites			Nonsynonymous sites		
		Number of sites	Number of mutations	π_s	Number of sites	Number of mutations	π_a	Number of sites	Number of mutations	π_s	Number of sites	Number of mutations	π_a
Overall	760	158.24	14	0.02949	471.76	16	0.01050	194.42	3	0.00514	564.58	10	0.00697
Y_1	88	20.90	1	0.01346	66.10	3	0.01299	20.56	0	0.00000	66.44	2	0.01303
Y_2	90	11.83	1	0.04829	36.17	0	0.00000	22.50	0	0.00000	67.50	0	0.00000
G_1	204	57.71	3	0.01813	140.29	6	0.01493	59.89	1	0.00556	144.11	4	0.01202
G_2	144	23.43	1	0.02443	57.57	4	0.01986	40.08	1	0.00832	103.92	3	0.00963
S	54	12.19	0	0.00000	41.81	2	0.01373	12.00	0	0.00000	42.00	0	0.00000
K_1	45	8.33	2	0.06857	36.67	0	0.00000	8.33	1	0.04000	36.67	0	0.00000
K_{sp}	81	15.83	5	0.10226	50.17	1	0.00570	20.06	0	0.00000	60.94	1	0.00545
K_2	54	11.00	0	0.00000	43.00	0	0.00000	11.00	0	0.00000	43.00	0	0.00000

occur only in Y_1, G_1, G_2, and K_{sp} regions. Differences among subregions indicate a tendency for differences in sequence conservation. It is to be noted that π_a is null for Y_2, K_1 and K_2, i.e. subregions corresponding to the most typical dehydrin domains.

On the whole, the data reported in Tables 2 and 3 and Fig. 2 suggest that different nucleotide regions have evolved differently. This phenomenon has been observed for members of other plant multigene families, for example that encoding phytochromes (Mathews et al. 1995; Alba et al. 2000).

Concerning insertions or deletions, single nucleotide gaps in the coding regions were absent, as already observed in barley dehydrins by Choi et al. (1999).

A three-nucleotide gap was found only in the intron of HCM, NSH43 and RomsunHS53. Among wild *H. annuus* accessions, one to nine nucleotide insertions and one to four nucleotide deletions were found in the intron; one to seven nucleotide indels were observed in the 3′-UTR. In the coding regions, a 39-nucleotide deletion in the Y_2 region of Durango accession, a six-nucleotide insertion in the G1 region of Durango and Arizona accessions, a 63-nucleotide deletion in the G_2 region of Colorado accession and a 12-nucleotide deletion in the K_{sp} region of the Washington accession were observed. In all cases, insertions or deletions did not determine frame shifts.

To study if the observed nucleotide sequence variations also determine variations in biochemical and biophysical characteristics of deduced proteins, different parameters were studied. The consensus protein sequence of sunflower deduced dehydrin is reported in Fig. 3, together with variant amino acids found among genotypes. It can be observed that 10 out of 14 (71.8%) amino acid substitutions in wild accessions are nonconservative with regard to simple biochemical classification of amino acids (i.e. polar, non-polar, acidic, basic, aromatic and cystein); four out of ten (40.0%) amino acid substitutions are nonconservative in cultivated genotypes.

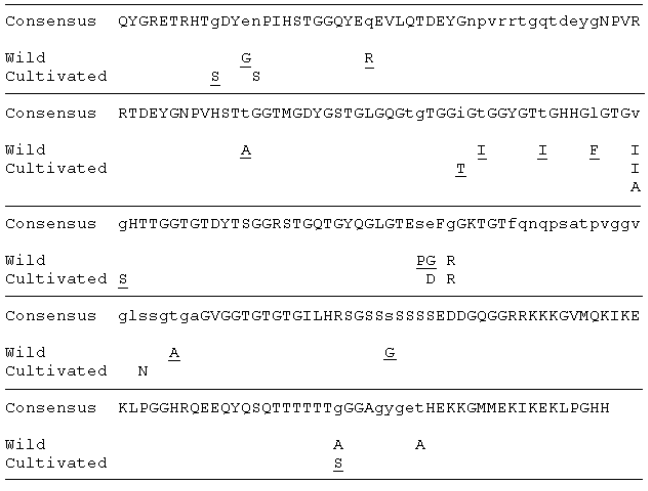

Fig. 3 Consensus sequence of *Hadhn1* deduced protein. Changes relative to consensus sequence are reported *below* each sequence. Non-conservative amino acid substitutions are *underlined*

The hydrophobicity profiles are quite similar between wild and cultivated genotypes (Fig. 4a). However, if the standard error for each amino acid position is considered (Fig. 4b), a much higher variability can be deduced for wild accessions than for cultivated accessions.

Deduced proteins were also analysed for their molecular weight, calculated isoelectric point, and predicted secondary structure (percentage of α-helix, extended strand and random coil) (Table 4). Again, variability is much larger in wild than in cultivated sunflowers, as evidenced by the ratio between standard errors.

The functionality of isolated alleles was studied at the RNA expression level, by RT-PCR. Alleles from each genotype were expressed after leaf desiccation until 60% fresh weight. In Fig. 5, RT-PCR products from HCM line, Colorado and Durango accessions (that present large deletions in the isolated alleles) are reported.

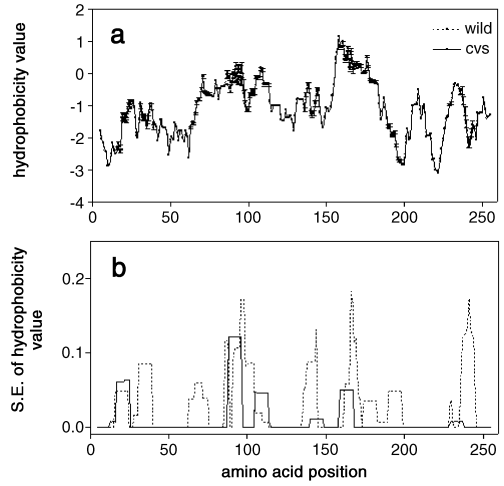

Fig. 4 Mean hydrophobicity profiles (±SE) of deduced dehydrins from wild and cultivated genotypes of *H. annuus* (**a**) and SEs of hydrophobicity value at different amino acid position (**b**)

Table 4 Mean (±standard error, SE) molecular weights (MW), isoelectric points (I_p), percentages of α-helix, extended strand and random coil in dehydrin proteins deduced from DNA sequences isolated from wild and cultivated sunflower genotypes. For each parameter, the ratio between SEs in cultivated and wild genotypes is reported

	Wild	Cultivated	SE cv./ SE wild
MW	25,160.00 ± 299.18	25,811.31 ± 7.88	0.026
Ip	6.893 ± 0.091	6.630 ± 0.000	0.000
α-helix	9.060 ± 0.241	8.620 ± 0.130	0.539
Extended strand	5.256 ± 0.524	5.085 ± 0.065	0.124
Random coil	85.684 ± 0.631	86.295 ± 0.133	0.211

Finally, a neighbor-joining analysis of the 13 nucleotide sequences is reported in Fig. 6. Apart from the Washington accession (whose position is not supported by the bootstrap value, however), cultivated and wild genotypes are separated and in different positions in the unrooted tree, possibly indicating a unique origin for cultivated genotypes.

On the whole, both nucleotide and amino acid sequence data showed that variability in the dehydrin sequence is much larger in wild sunflowers than in cultivated ones. This uniformity of domesticated sunflower had been observed by Cronn et al. (1997) using allozyme analysis but our data, to date the only available at the DNA sequence level, indicate strong genetic uniformity for sunflower cultivars, even though only one gene and relatively few genotypes were analysed. The cultivated genotypes tested were randomly chosen according to their different provenance (one for each country in which sunflower is a major crop), hence it is plausible that they do not derive from the same inbred lines.

The different genetic variability between cultivars and wild accessions is probably related to the time course of

Fig. 5 RT-PCR products of *Dhn1* sequences from leaves of HCM line, Colorado and Durango accessions. Molecular weights of bands (in basepairs) are reported

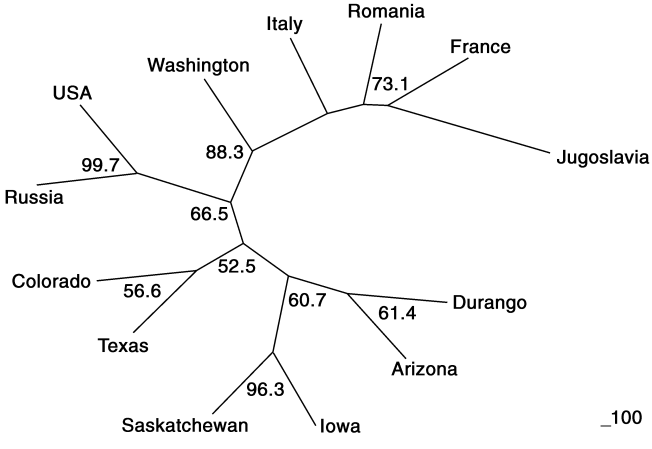

Fig. 6 Strict consensus tree, based on nucleotide sequences and obtained by the neighbor-joining analysis (number of replicates = 1,000). *Numbers* at internal branches indicate bootstrap values higher than 50%

sunflower breeding. It began in Russia in the 19th century using relatively few American genotypes introduced into Europe by early Spanish explorers (Putt 1978) and in Russia by Peter the Great in the 18th century (Zukovsky 1950). This small number of genotypes was the starting material for subsequent breeding, as has been shown by pedigree analysis of modern cultivars (Cheres and Knapp 1998), even in North America (Semelczi-Kovacs 1975). This may have determined uniformity, at least for some genes, even in cultivars of very different origin.

Such a uniformity of the dehydrin gene in cultivars could also indicate that these allelic forms were the most suitable for the environments in which sunflower is cultivated and they were unconsciously selected by breeders.

The relatively large genetic variability of the dehydrin gene could be related to the geographic distribution of wild *H. annuus* in North America, which is the largest among *Helianthus* species (Rogers et al. 1982). Since dehydrins are involved in environmental adaptation, their variability in wild *H. annuus* could be related to the different environments in which this species lives and useful for breeding purposes.

If variability of the dehydrin gene is an indicator of the general genetic variability of wild *H. annuus*, then our experiments suggest that this should be taken into consideration in sunflower breeding programs. The narrow genetic background of cultivated sunflower has been a concern for its potential for improvement, and efforts to widen its genetic base are underway. Wild *Helianthus* species often constitute the basic genetic stock from which cultivated sunflower originated, and they are used for variability rescue and the introgression of important traits in *H. annuus*. The 49 North American wild *Helianthus* species have long survived extreme environments and possess resistance or tolerance genes to salt, drought, insects, diseases, as well as cytoplasmic male-sterility (Seiler 1992; Jan 2001). Gene transfer from wild species into the cultivated background largely depends on the success of interspecific hybridisation, F_1 fertility, chromosome pairing for genetic recombination, efficient screening methods and a sufficient number of progenies for selection. When used in crosses with *H. annuus*, wild annual *Helianthus* species generally produce F_1 seeds, perennial species do not.

Such difficulties are obviously absent in crosses involving wild *H. annuus*. If wild populations of *H. annuus* indeed show large genetic variability, they could be the best genetic resources for sunflower improvement, together with or alternative to interspecific crosses.

Acknowledgements This research work was partially supported by CNR-Target Project on Biotechnology.

References

Alba R, Kelmenson PM, Cordonnier-Pratt MM, Pratt LH (2000) The phytochrome gene family in tomato and the rapid differential evolution of this family in Angiosperms. Mol Biol Evol 17:362–373

Cellier F, Conejero G, Breitler JC, Casse F (1998) Molecular and physiological responses to water deficit in drought-tolerant and drought-sensitive lines of sunflower. Accumulation of dehydrin transcripts correlates with tolerance. Plant Physiol 116:319–328

Cheres MT, Knapp SJ (1998) Ancestral origins and genetic diversity of cultivated sunflower: coancestry analysis of public germplasm. Crop Sci 38:1476–1482

Choi DW, Zhu B, Close TJ (1999) The barley (*Hordeum vulgare* L.) dehydrin family: sequences, allele types, chromosome assignments, and expression characteristics of 11 Dhn genes of cv. Dicktoo. Theor Appl Genet 98:1234–1247

Close TJ (1996) Dehydrins: emergence of a biochemical role of a family of plant dehydration proteins. Physiol Plant 97:795–803

Close TJ (1997) Dehydrins: a commonalty in the response of plants to dehydration and low temperature. Physiol Plant 100:291–296

Close TJ, Kortt AA, Chandler PM (1989) A cDNA-based comparison of dehydration-induced proteins (dehydrins) in barley and corn. Plant Mol Biol 13:95–108

Cronn R, Brothers M, Klier K, Bretting PK, Wendel JF (1997) Allozyme variation in domesticated annual sunflower and its wild relatives. Theor Appl Genet 95:532–545

Cummings MP, Clegg MT (1998) Nucleotide sequence diversity at the alcohol dehydrogenase 1 locus in wild barley (*Hordeum vulgare* ssp. *spontaneum*): an evaluation of the background selection hypothesis. Proc Natl Acad Sci USA 95:5637–5642

Doyle JJ, Doyle JL (1989) Isolation of plant DNA from fresh tissue. Focus 12:13–15

Dure III L, Crouch M, Harada J, Ho THD, Mundy J, Quatrano R, Thomas T, Sung ZR (1989) Common amino acid sequence domains among the LEA proteins of higher plants. Plant Mol Biol 12:475–486

Felsenstein J (1989) PHYLIP: phylogeny inference package. Cladistics 5:164–166

Giordani T, Natali L, D'Ercole A, Pugliesi C, Fambrini M, Vernieri P, Vitagliano C, Cavallini A (1999) Expression of a dehydrin gene during embryo development and drought stress in ABA-deficient mutants of sunflower (*Helianthus annuus* L.). Plant Mol Biol 39:739–748

Goday A, Jensen AB, Culianez-Macia FA, Alba MM, Figueras M, Serratosa J, Torrent M, Pagès M (1994) The maize abscisic acid-responsive protein *Rab17* is located in the nucleus and interacts with nuclear localization signals. Plant Cell 6:351–360

Godoy JA, Lunar R, Torres-Schumann S, Moreno J, Rodrigo RM, Pintor-Toro JA (1994) Expression, tissue distribution and subcellular localization of dehydrin TAS14 in salt-stressed tomato plants. Plant Mol Biol 26:1921–1934

Grosselindemann E, Robertson M, Wilmer JA, Chandler PM (1998) Genetic variation in pea (*Pisum*) dehydrins: sequence elements responsible for length differences between dehydrin alleles and between dehydrin *loci* in *Pisum sativum* L. Theor Appl Genet 96:1186–1192

Hartl FU, Hlodan R, Langer T (1994) Molecular chaperones in protein folding: the art of avoiding sticky situations. Trends Biochem Sci 19:20–25

Ismail AM, Hall AE, Close TJ (1999) Allelic variation of a dehydrin gene cosegregates with chilling tolerance during seedling emergence. Proc Natl Acad Sci USA 96:13566–13570

Jan CC (2001) Interspecific hybridization, gene transfer, and the development of resistant sunflower to the new broomrape race F in Spain. In: Durante M, Bernardi R (eds) Fifth Eur Conf Sunflower Biotechnol. Abstract Book, Università di Pisa, Pisa, p 37

Kawabe A, Innan H, Terauchi R, Miyashita NT (1997) Nucleotide polymorphism in the acidic chitinase locus (*ChiA*) region of the wild plant *Arabidopsis thaliana*. Mol Biol Evol 14:1303–1315

Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol 157:105–132

Lang V, Robertson M, Chandler PM (1998) Allelic variation in the dehydrin gene family of "Himalaya" barley (*Hordeum vulgare* L.). Theor Appl Genet 96:1193–1199

Lin JZ, Brown AHD, Clegg MT (2001) Heterogeneous geographic patterns of nucleotide sequence diversity between two alcohol dehydrogenase genes in wild barley (*Hordeum vulgare* ssp. *spontaneum*). Proc Natl Acad Sci USA 98:531–536

Martin J, Geromanos S, Tempst P, Hartl FU (1993) Identification of nucleotide-binding regions in the chaperonin proteins GroEL and GroES. Nature 366:279–282

Mathews S, Lavin M, Sharrock RA (1995) Evolution of the phytochrome gene family and its utility for phylogenetic analyses of Angiosperms. Ann Mo Bot Gard 82:296–321

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418–426

Ouvrard O, Cellier F, Ferrare K, Tousch D, Lamaze T, Dupuis JM, Casse-Delbart F (1996) Identification and expression of water stress- and abscisic acid-regulated genes in a drought-tolerant sunflower genotype. Plant Mol Biol 31:819–829

Putt ED (1978) History and present world status. In: Carter J (ed) Sunflower science and technology. Am Soc Agron, Madison, Wis. pp 1–30

Rogers CE, Thompson TE, Seiler GJ (1982) Sunflower species of the United States. Nat Sunflower Assoc, Fargo, N.D.

Rothnie HM (1996) Plant mRNA 3′-end formation. Plant Mol Biol 32:43–61

Rozas J, Rozas R (1999) DNASP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics 15:174–175

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory, Cold Spring Harbor

Schneider K, Wells B, Schmelzer E, Salamini F, Bartels D (1993) Desiccation leads to the rapid accumulation of both cytosolic and chloroplastic proteins in the resurrection plant *Craterostigma plantagineum* Hochst. Planta 189:120–131

Segrest JP, De Loof H, Dohlman JG, Brouillette CG, Anantharamaiah GM (1990) Amphipatic helix motif: classes and properties. Prot Struct Funct Genet 8:103–117

Seiler GJ (1992) Utilization of wild sunflower species for the improvement of cultivated sunflower. Field Crops Res 30:231–270

Semelczi-Kovacs A (1975) Acclimatization and dissemination of the sunflower in Europe. Acta Ethnogr Acad Sci Hung 24:47–88

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Vilardell J, Godoy A, Freire MA, Torrent M, Martinez MC, Torne JM, Pagès M (1990) Gene sequence, developmental expression and protein phosphorylation of RAB-17 in maize. Plant Mol Biol 14:423–432

Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, Hochstrasser DF (1998) Protein identification and analysis tools in the ExPASy server. In: Link AJ (ed) 2-D Proteome Analysis Protocols. Humana Press, Totowa

Zhu B, Choi DW, Fenton R, Close TJ (2000) Expression of the barley dehydrin multigene family and the development of freezing tolerance. Mol Gen Genet 264:145–153

Zukovsky PM (1950) Cultivated plants and their wild relatives. Commonw Agric Bur, Farnham Royal, UK