Theor Appl Genet (2003) 107:316–325
DOI 10.1007/s00122-003-1249-5

T. Giordani · L. Natali · A. Cavallini

Analysis of a dehydrin encoding gene and its phylogenetic utility in *Helianthus*

Received: 16 October 2002 / Accepted: 3 February 2003 / Published online: 23 April 2003

Abstract Dehydrins are ubiquitous plant proteins, synthesized in late stages of plant embryo development and following any environmental stress involving dehydration. With the aim to study the evolution of such a stress-responsive gene within *Helianthus* and to test the possibility of using this gene for phylogenetic studies, fragments of the same dehydrin gene were isolated by PCR and sequenced in 16 wild *Helianthus* species or subspecies. All isolated sequences included the typical dehydrin domains (Y, S and K), a portion of 3′-UTR and an intron, inserted in the same position within the S domain-encoding region. The number of nucleotide substitutions (both synonymous and nonsynonymous) was calculated keeping separate the different gene regions, and differences occur even among coding domains, indicating that evolutionary constraints act differently on each region. The occurrence of indels and/or insertions was also observed. At the deduced protein level, the calculation of isoelectric point, molecular weight and the percentage of α-helix showed a diversification of biochemical properties of this protein between annual and perennial *Helianthus* species. Phylogenetic trees were built by the maximum-likelihood, maximum-parsimony, and neighbor-joining methods. In all cases the same topology was observed; perennial and annual species form a supported clade, and *H. annuus* was separated from the other annuals and from perennials. These data support the use of this stress-responsive gene to study the phylogeny of *Helianthus*.

Communicated by C. Möllers

Electronic Supplementary Material is available if you access this article at http://dx.doi.org/10.1007/s00122-003-1249-5. On that page (frame on the left side), a link takes you directly to the supplementary material.

T. Giordani · L. Natali · A. Cavallini (✉)
Department of Agricultural Plant Biology, Genetics Section,
Via Matteotti 1/B, I-56124 Pisa, Italy
e-mail: acavalli@agr.unipi.it
Tel.: +39-050571567
Fax: +39-050576750

Keywords Dehydrin · *Helianthus* · Sunflower · Indel · Substitution · Phylogenetic trees

Introduction

Dehydrin-encoding genes have been isolated from many angiosperm and gymnosperm species (Close 1997). Dehydrins are characterized by one or more putative amphipatic α-helix-forming consensus regions at the C-terminus and, often, 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). These proteins are synthesized by plants in the late stages of embryo development [they are also known as the *Lea D11* subgroup of late-embryogenesis-abundant (*Lea*) proteins, Dure et al. 1989], but also following any environmental stress involving dehydration (drought, cold stress, salinity). Dehydrins are key components of dehydration tolerance (Close 1996; Zhu et al. 2000) through their association with macromolecules such as nucleoprotein complexes in the nucleus (Godoy et al. 1994) and endomembranes in the cytoplasm (Schneider et al. 1993). It was proposed that these proteins are surfactants inhibiting the coagulation of a number of macromolecules and preserving their structural integrity (Close 1996). It is worth noting that in *Vigna unguiculata* plantlets, the occurrence of a mutation in a dehydrin gene determines chilling sensitivity (Ismail et al. 1999). This finding suggests the importance of studying the genetic variability of dehydrins even for plant breeding purposes.

Analysis of variability of dehydrin genes has been studied in different species. Such 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 there are few differences among sequences from within a single species (Grosselindemann et al. 1998; Lang et al. 1998; Choi et al. 1999).

A dehydrin-encoding cDNA, *HaDhn1*, has been isolated and sequenced in sunflower (Ouvrard et al. 1996),

whose transcript accumulation is related to drought tolerance (Cellier et al. 1998). A sequence allelic to this gene, *HaDhn1a* (Giordani et al. 1999), is expressed in the latest stages of *Helianthus annuus* embryogenesis, depending on abscisic acid accumulation, and after drought stress even in ABA-deficient sunflower mutants.

The *HaDhn1a* sequence from the *H. annuus* HCM homozygous line was isolated by PCR using primers that do not belong to shared typical dehydrin domains and were designed to obtain one specific DNA fragment by PCR (Natali et al. 2003). This was deduced since a homozygous line was used to determine how many genes could be isolated with these primers: with a homozygous line, the occurrence of different PCR-amplified fragments should have to correspond to more than one gene and not to different alleles of the same gene. After amplification and cloning, 50 clones were first analyzed for inserted fragment length in a high-resolution gel and no length variation was found; then, 15 out of 50 clones were sequenced and all inserted fragments corresponded to the same sequence, showing no sequence variation. Therefore, in the experimental conditions used (primers, stringency, temperature) only one product could be obtained using that primer pair and, in *H. annuus, HaDhn1a*, though belonging to a gene family, identified a single-copy gene (Natali et al. 2003).

Helianthus belongs to the Compositae family and provides two food plants, *H. annuus*, and *Helianthus tuberosus*, the Jerusalem artichoke. In their classification, based on morphological and crossability analyses, Heiser et al. (1969) included 67 species, annual or perennial, 50 native to North America and 17 to South America. Then, Schilling and Heiser (1981) excluded the South American species and subdivided *Helianthus* into four sections, section *Helianthus* (comprising only diploid and annual species), section *Agrestes* (comprising only the diploid and annual *H. agrestis*), section *Ciliares* (with six perennial species) and section *Atrorubentes* (with 30 perennial and one annual species, *Helianthus porteri*). Further analyses, based on commonly used molecular markers for species-level phylogenies in plants (i.e. ITS and cpDNA RFLPs), showed that all perennial species are comprised in one section and that *H. porteri* forms a section *per se* (Schilling 1997). Phylogenetic analyses suggested a close relationship between *Helianthus* and *Phoebanthus* (Soltis and Soltis 2000; Schilling 2001); other close relatives of *Helianthus* are *Viguiera* and *Tithonia*. Based on the geographic distributions of its closest relatives, *Helianthus* most likely originated in Mexico, with subsequent migration through North America (Schilling et al. 1998). Several species of *Helianthus* are known to be of hybrid origin (Rieseberg 1995).

Phylogenetic reconstruction in *Helianthus* has been unusually difficult because of the recent and rapid diversification of the group and because of widespread hybridization. Hence, the use of different gene sequences to resolve phylogenetic relationships in this group can be necessary. In this paper we report on the analysis of DNA fragments homologous to *HaDhn1* from 16 available wild *Helianthus* species or subspecies. The main objectives of this work were to analyze the variability of this dehydrin gene and its evolution in *Helianthus*, and to test the possibility to use a plant stress-responsive gene to investigate the phylogeny of this genus.

Materials and methods

Plant material and DNA isolation

The species and subspecies from which DNA sequences putatively orthologous to *HaDhn1a* were isolated and their provenances are reported in Table 1. Beyond these species, DNA was isolated also from *Helianthus giganteus, Helianthus argophyllus, Helianthus bolanderi, Helianthus pumilus, Helianthus simulans, Helianthus atrorubens, Helianthus decapetalus, Helianthus divaricatus, Helianthus strumosus, Helianthus grosseserratus, Helianthus nuttallii, Helianthus occidentalis*. Seeds were germinated and plantlets grown to maturity in the field. For some analyses, plants were hand-forced for self-pollination, that ensured a small number of seeds. For each accession, DNA was extracted from leaf tissues of one plant according to the method devised by Doyle and Doyle (1989) with minor modifications. Leaves from single plants were ground in a pre-heated mortar in CTAB isolation buffer [2% (w/v) CTAB (Sigma), 1.4 M of NaCl, 0.2% (v/v) 2-mercaptoethanol, 20 mM of EDTA, 100 mM of Tris-HCl, (pH 8.0)] at 60 °C. The samples were incubated at 60 °C for 30 min with occasional gentle swirling and then extracted once with chloroform/isoamyl alcohol (24:1, v/v). After centrifugation (5,000 rpm) at room temperature, nucleic acids were precipitated from the aqueous phase by adding 2/3 volumes of cold isopropanol and then washed in 76% (v/v) ethanol, 10 mM ammonium acetate for 1–2 h, allowed to dry briefly and resuspended in water. RNA was digested by RNAse A (100 μg/ml at 37 °C for 1 h); RNAse was then extracted with chloroform/isoamyl alcohol (24:1, v/v), and DNA was ethanol precipitated and re-solubilized in water. Seeds were collected from plants used for analyses and are stored in our Department.

Isolation of genomic Dhn1 sequences by PCR

DNA sequences putatively orthologous to *HaDhn1a* were isolated by PCR on genomic DNAs from the species listed in Table 1. 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 30 s and 72 °C for 60 s, for 30 cycles, using *Taq*-DNA polymerase (Promega).

The amplified fragments were 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) using SP6 and T7 primers according to the manufacturer's instructions; sequences were analyzed using Sequencing Analysis 2.1.2 (Perkin-Elmer) and the Sequencher 3.0 analysis program (Gene Codes Corporation); forward and reverse sequencing runs were compared.

Southern blotting and hybridization

Southern blotting of DNA was performed according to standard protocols (Sambrook et al. 1989). Eight micrograms of DNA from 16 *Helianthus* species or subspecies (Table 1) were digested with *Eco*RI and *Bam*HI restriction endonucleases (not cutting within the *HaDhn1a*) in a five-fold excess according to the instructions of the suppliers (Roche), then electrophoresed on agarose gels and blotted onto the nylon membrane (Roche). Complete digestion was

Table 1 Species or subspecies name and source, Genbank accession number of the isolated *Dhn1* sequence and putative structure of dehydrin according to Close (1997)

Species	Source[a]	GenBank accession number	Structure
H. annuus (HCM line)	DBPA, Pisa, Italy	AJ002741	Y_3SK_2
H. annuus (acc. Colorado)	NCRPIS, Ames, USA	AJ250225	Y_3SK_2
H. annuus (acc. Durango)	INIFAP, Durango, Mexico	AJ249273	Y_2SK_2
H. annuus (acc. Iowa)	NCRPIS, Ames, USA	AJ250226	Y_3SK_2
H. annuus (acc. Saskatchewan)	NCRPIS, Ames, USA	AJ250227	Y_3SK_2
H. annuus (acc. Texas)	NCRPIS, Ames, USA	AJ438979	Y_3SK_2
H. annuus (acc. Washington)	NCRPIS, Ames, USA	AJ250228	Y_3SK_2
H. annuus (acc. Arizona)	NCRPIS, Ames, USA	AJ250224	Y_3SK_2
H. debilis cucumerifolius	NCRPIS, Ames, USA	AJ249708	Y_2SK_2
H. debilis debilis	NCRPIS, Ames, USA	AJ249709	Y_2SK_2
H. debilis silvestris	NCRPIS, Ames, USA	AJ249710	Y_2SK_2
H. neglectus	NCRPIS, Ames, USA	AJ250150	Y_2SK_2
H. niveus canescens	DBG, Phoenix, USA	AJ250147	Y_2SK_2
H. petiolaris petiolaris	NCRPIS, Ames, USA	AJ250152	Y_3SK_2
H. petiolaris fallax	NCRPIS, Ames, USA	AJ250151	Y_3SK_2
H. praecox hirtus	NCRPIS, Ames, USA	AJ250125	Y_2SK_2
H. praecox praecox	NCRPIS, Ames, USA	AJ250153	Y_2SK_2
H. praecox runyoni	NCRPIS, Ames, USA	AJ250126	Y_2SK_2
H. hirsutus	NCRPIS, Ames, USA	AJ250145	Y_2SK_2
H. mollis	BG, Poznan, Poland	AJ250146	Y_2SK_2
H. maximiliani	NCRPIS, Ames, USA	AJ250149	Y_2SK_2
H. tuberosus	DBPA, Pisa, Italy	AJ250148	Y_2SK_2
H. ciliaris	NCRPIS, Ames, USA	AJ297737	Y_2SK_2
T. rotundifolia	NCRPIS, Ames, USA	AJ250127	Y_2SK_2

[a] NCRPIS: North Central Regional Plant Introduction Station; INIFAP: Instituto Nacional de Investigaciones Forestales y Agropecuarias; DBPA: Dipartimento di Biologia delle Piante Agrarie; DBG: Desert Botanical Garden; BG: Botanical Garden

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-labelled *HaDhn1a* cDNA as a probe, that was previously cloned (Giordani et al. 1999), under high stringency conditions, at 50 °C in 50% formamide, 5 × SSC (1 × SSC in 0.15 M NaCl, 0.015 M trisodium citrate, pH 7.0), 2% blocking reagent (Roche), 0.02% SDS and 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. Detection of hybridization signals was performed using the DIG-Nucleic Acid Detection Kit (Roche) according to manufacturer's instructions.

DNA sequence analysis

The coding portion of the DNA sequence was subdivided into four regions according to the functional characteristics of the putatively encoded protein domain; an intron and the 3′-UTR fragment were also analyzed. Intron delimitation within genomic sequences was made by comparing the genomic sequence of *H. annuus* with the corresponding cDNA (Giordani et al. 1999) and confirmed by the use of the program FEX (Baylor College of Medicine, Texas, USA).

Overall sequences and sequence segments were aligned using CLUSTALW (Thompson et al. 1994). Some adjustments were made by eye. Alignments are available by the authors upon request.

Overall numbers of nucleotide substitutions per site were calculated for different gene regions from pairwise sequence comparisons, according to the two-parameter model of Kimura (1980) using the program DNADIST of the PHYLIP program package version 3.572 (Felsenstein 1989). Numbers of synonymous and nonsynonymous substitutions per site (K_s and K_a) were estimated for coding nucleotide sequences using the DnaSP program (Rozas and Rozas 1999), according to the method of Nei and Gojobori (1986). Alignment gaps were excluded from comparisons.

The number of indels per site in the coding and noncoding *Dhn1* regions was calculated according to Laroche and Bousquet (1999). Their equation is based on the assumption that each indel is the result of a single mutational event (Laroche et al. 1997), but it allows for a more realistic estimation of number of indels per site. The number of indels per site between two nucleotide sequences was obtained by the formula

$$I = N/(L - D + N),$$

where I = indel rate, N = total number of indels, L = total number of sites and D = number of sites involved in indels (Laroche and Bousquet 1999).

Relationships among dehydrin DNA sequences from different species were investigated by three commonly used methods of phylogenetic inference, maximum parsimony (MP), neighbor-joining (NJ) and maximum likelihood (ML), employing the PHYLIP program package version 3.572 (Felsenstein 1989). Using the SEQBOOT program, 100 versions of the original alignment were generated; then, trees were generated using the DNAPARS, DNADIST or DNAML programs, for MP, NJ and ML analyses, respectively, using default options. The CONSENSE program was used to obtain consensus trees.

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 100% relative weight of the window edges compared to the window center. The predicted secondary structure of deduced proteins (percentage of α-helix, extended strand and random coil) was analyzed using the program HNN at the Pole Bio-Informatique Lyonnais server (France).

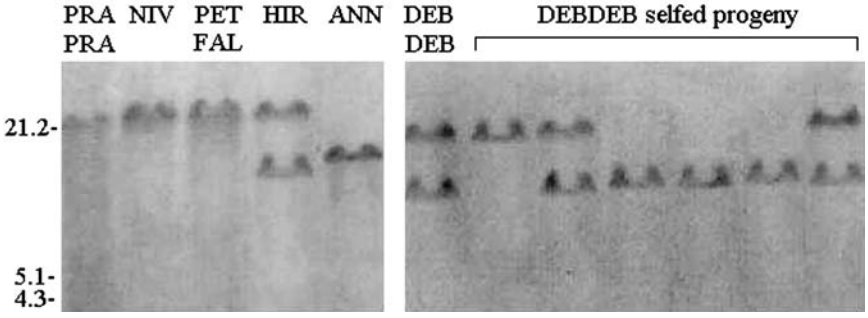

Fig. 1 Southern blot of genomic DNA from single plants of five *Helianthus* species (left; *PRAPRA*: *praecox praecox*; *NIV*: *niveus*; *PETFAL*: *petiolaris fallax*; *HIR*: *hirsutus*; *ANN*: *annuus*) and of a single plant of *H. debilis debilis* (*DEBDEB*) and its selfed progeny (right), digested with *Bam*HI and hybridized with digoxigenin-labeled *HaDhn1a*. Molecular-weight marker is reported on the left (in kbp)

Results and discussion

Structure and genetic variability of the *Dhn1* gene in the genus *Helianthus*

In *H. annuus, HaDhn1a*, though belonging to a gene family, identifies a single-copy gene (Natali et al. 2003, see Introduction). The same primers and experimental conditions used to amplify *HaDhn1a* by PCR were used on wild genotypes (presumably heterozygous).

Sequences orthologous to *HaDhn1a* were isolated from genomic DNA of 15 *Helianthus* species or subspecies (beyond *H. annuus*) and of *Tithonia rotundifolia*. The orthologous nature of the isolated sequences was assumed because: (1) the same non-degenerated oligonucleotides were used as PCR-primers; (2) only one electrophoretic band was observed after amplification for each genotype; and (3) after digestion with *Eco*RI or *Bam*HI and Southern-blot hybridisations, with labeled *HaDhn1a*, only single or double hybridization bands, were observed. In Fig. 1a, Southern hybridization in five species is reported; single or double hybridization bands were found also in the other *Helianthus* species analyzed (data not shown). When two bands were observed, as in *Helianthus debilis* ssp. *debilis*, this was due to heterozygosity in the sites recognized by the restriction enzymes: in fact, in these cases, after selfing plants, some individuals belonging to the selfed progeny displayed only one hybridization band, indicating that the segregation of bands occurred (Fig. 1b), i.e. the two bands represent two *Dhn1* alleles. This suggests that also in the other *Helianthus* species, *Dhn1* is a single-copy gene. Only in *Helianthus hirsutus* (tetraploid, Fig. 1a) and *H. tuberosus* (hexaploid) multiple bands occurred; in these species only one allele was isolated. The possibility that isolated *Dhn1* sequences of these species are homoeologous can not be ruled out. *HaDhn1a* sequences were also isolated from seven *H. annuus* wild accessions (Natali et al. 2003), and the sequence from accession Arizona, the most similar to consensus *H. annuus Dhn1*, was used for comparisons to the other *Helianthus* species.

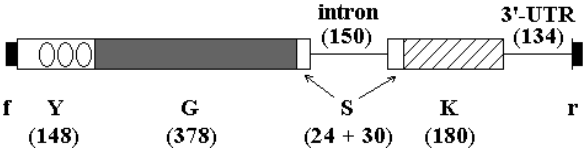

Fig. 2 Schematic representation of dehydrin-encoding *Dhn1* sequence of *H. annuus*. *Boxes* indicate different domains of the gene according to their encoded protein sequence. *Circles* within the Y domain indicate the TDEYGNP repeated motifs. Number of nucleotides for each domain is indicated in parentheses. *f* and *r* are the forward and reverse primers, respectively

The GenBank accession numbers of the sequences isolated are reported in Table 1 together with their structure, according to the nomenclature proposed by Close (1997). It is to be considered that no DNA sequence was amplified from the genomic DNA of the following species: *H. giganteus, H. argophyllus, H. bolanderi, H. pumilus, H. simulans, H. atrorubens, H. decapetalus, H. divaricatus, H. strumosus, H. grosseserratus, H. nuttallii* and *H. occidentalis*. This could be due to differences between the selected oligonucleotides and the actual nucleotide sequence of *Dhn1* in these species. With the exceptions of *H. argophyllus* and *H. bolanderi*, all these species are perennial.

Concerning the gene structure, all the isolated sequences present the typical dehydrin domain encoding regions. After alignment using the program CLUSTALW, they were subdivided in four nucleotide coding regions according to their encoded amino-acid sequence (regions Y, G, S and K) and two non-coding regions, intron and 3′-UTR (Fig. 2).

At the N terminus, the Y domain contained the TDEYGNP motifs which are similar to portions of plant and bacterial chaperones with affinity to nucleic acids (Martin et al. 1993). TDEYGNPs are arranged differently in *Helianthus* species, as shown in Fig. 3. All analyzed species, except *H. annuus, Helianthus petiolaris* ssp. *petiolaris*, and *H. petiolaris* ssp. *fallax*, showed two TDEYGNP repetitions. A third TDEYGNP is found in *H. annuus*. Also *H. petiolaris* ssp. *petiolaris* and *H. petiolaris* ssp. *fallax* showed a third TDEYGNP, but in a different

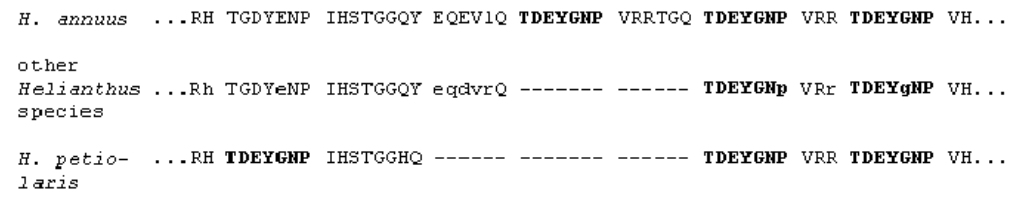

Fig. 3 Alignment of the deduced amino-acid sequences of the Y domain of *H. annuus, Helianthus* spp. consensus and *H. petiolaris* dehydrin. Conserved amino acids are in *capitals*, consensus amino acids in *lower case*, the TDEYGNP motifs in *boldface*

Table 2 Number of sites and number of nucleotide substitutions per site (K), obtained by pairwise comparison of sequences of different portions of PCR-amplified *Dhn1* sequences from *Helianthus* species (HIR: *hirsutus*, MAX: *maximiliani*, DEBDEB: *debilis debilis*, ANN: *annuus*, PRAPRA: *praecox praecox*, PRARUN: *praecox runyoni*, PETFAL: *petiolaris fallax*, PETPET: *petiolaris petiolaris*)

Compared species	Overall		Coding region		Intron		3′-UTR	
	No. of sites	K	No. of sites	K	No. of sites	K	No. of sites	K
HIR/MAX	917	0.0266	666	0.0260	129	0.0078	122	0.0509
HIR/DEBDEB	947	0.0513	703	0.0409	121	0.1160	123	0.0504
HIR/ANN	942	0.0654	697	0.0566	121	0.1556	124	0.0330
MAX/DEBDEB	919	0.0530	667	0.0416	121	0.1064	131	0.0637
MAX/ANN	914	0.0699	661	0.0615	121	0.1455	132	0.0469
DEBDEB/ANN	973	0.0621	715	0.0566	125	0.1308	133	0.0307
PRAPRA/PRARUN	985	0.0237	721	0.0240	127	0.0159	137	0.0298
PETFAL/PETPET	951	0.0117	688	0.0102	129	0.0317	134	0.0000

position from *H. annuus*; this sequence appeared to have degenerated in all the other sequences (Fig. 3).

After the Y domain, a hydrophilic, glycine-rich domain (G region) may be found. The presence of glycine-rich domains was hypothesized to be involved in general dehydrin hydrophilicity (Close et al. 1989), though its function is far from clear. Functionally, it should correspond to the ϕ regions described in dehydrins of other species (see for example Choi et al. 1999). It is to be considered that just a deletion in one ϕ repetition determines a strong reduction of chilling tolerance in *V. unguiculata* plantlets (Ismail et al. 1999), indicating the importance of these segments in dehydrin function.

After the G domain, a serine-rich domain (domain S) is observed, with eight serine residues, whose nucleotide sequence is divided into two parts by an intron, inserted at the same position in all analyzed sequences (Fig. 2). It has been hypothesized 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).

At the C-terminus, there is a K domain, showing alternate hydrophilic and hydrophobic portions that should form an amphipatic α-helix (Segrest et al. 1990), probably involved in the stabilization of partially denatured proteins (Hartl et al. 1994). This domain contains two typical KIKEKLPG motifs in all the analyzed sequences. After the K domain, the 3′-UTR follows and is incomplete in our clones.

To study the evolution of different regions in *Helianthus* species, we analyzed the number of nucleotide substitutions per site rather than substitution rates. In fact, the age of the genus *Helianthus* ranges between 4.75 and 22.7 million years, measured after analysis of cpDNA, and, within the genus, the extant lineages arose between 1.7 and 8.2 million years ago (Schilling 1997). Considering that the use of cpDNA data as a molecular clock is not without difficulty (Wendel and Albert 1992; Schilling 1997), particularly for closely related species (Li 1997), we preferred to analyze absolute numbers of nucleotide substitutions per site.

The number of nucleotide substitutions per site (K) was calculated for the overall sequence or keeping separate the coding portion, intron and 3′-UTR, in pairwise comparisons between species or subspecies. In Table 2 K values, obtained by pairwise comparison of sequences from selected species, two annuals (*H. annuus* and *H. debilis* ssp. *debilis*) and two perennials (*H. hirsutus* and *Helianthus maximiliani*), and from selected subspecies (*Helianthus praecox* ssp. *praecox* and ssp. *runyoni*, and *H. petiolaris* ssp. *fallax* and ssp. *petiolaris*) are reported, to show differences between perennials and annuals, and between subspecies. The use of the mean of pairwise comparisons of all species was considered inappropriate because it would ignore the phylogenetic correlation (and thus non-independence) of substitutions. The same trend comparing selected species or subspecies was observed comparing other annuals or other perennials between them and other annuals to other perennials (data not shown).

As for the overall sequence and the coding regions, subspecies show the lowest K values. However, analogous low values are observed between the two perennials. This pattern is observed also concerning the intron sequence, with a surprisingly low K value between

perennials. Considering the 3′-UTR, K values are analogous between each species and between the two subspecies of *H. praecox*, while no difference is observed between subspecies of *H. petiolaris*. As a general rule, the 3′-UTR shows nucleotide substitution numbers analogous to the coding region and much lower than the intron, indicating the presence of evolutionary constraints.

Synonymous and nonsynonymous substitutions numbers, obtained from pairwise comparisons of the same species, are reported in Table 3, with regard to the single regions of the *Dhn1* gene. As far as the overall sequence is concerned, the highest K_a values are observed when *H. annuus* is involved in comparison; in the other cases (excluding subspecies comparisons), they range from 0.0216 to 0.0250, i.e. are nearly the same. K_s values are much higher than the K_a ones, indicating the occurrence of evolutionary constraints on this gene. It is to be noted that between the two perennial species (*H. hirsutus* and *H. maximiliani*) K_s is very low. Concerning the different coding regions, in some cases the number of sites is too low to obtain reliable data. For the Y region, K_s values are the highest, but the tendency for conservation of this protein domain is demonstrated by the very low K_a values. S and K regions appear very conserved, at both synonymous and nonsynonymous sites; the relatively high K_a values for the S region are actually largely related to a single amino-acid substitution. When nonsynonymous substitutions are taken into account, the G region is largely the most variable. This large variability is also found comparing the other species from which dehydrin was isolated, that are not reported in Table 3. On the whole, the occurrence of differences in the number of nucleotide substitutions within the *Dhn1* gene suggests that evolutionary constraints act differently on different dehydrin regions, as already reported for other genes (see for example Gaut et al. 1999).

Concerning the subspecies, Table 3 shows very different patterns of nucleotide substitutions within *H. praecox* and within *H. petiolaris*: very low K_s and K_a values are found comparing ssp. *fallax* and ssp. *petiolaris* of *H. petiolaris*. On the contrary, K_s and K_a between ssp. *praecox* and ssp. *runyoni* of *H. praecox* are in some cases higher than between perennial species (Table 3). Subspecies are usually defined by geographic and morphological differences; our results indicate that the amount of sequence variation might be considered a critical feature for the appropriateness of the subspecific rank.

Concerning insertions or deletions, single nucleotide gaps in the coding regions were absent and double ones very rare, as already observed in barley dehydrins by Choi et al. (1999). Table 4 reports values for comparisons between the chosen *Helianthus* species and subspecies.

Intron indels per site in *Helianthus* species are generally higher (except for the comparison between *H. hirsutus* and *H. maximiliani*) than those found in the 3′-UTR (Table 4). As for the coding regions at the interspecific level, the highest numbers of indels per site are found in those comparisons involving *H. annuus* (Table 4), indicating the peculiarity of this species within the genus *Helianthus*. No indels were found between subspecies of *H. praecox*, while some indels were observed between subspecies of *H. petiolaris*: this variability trend is opposite to that reported for nucleotide substitutions (Tables 2, 3).

Concerning coding regions, gaps appeared only in the alignment of the K regions of perennial species. A 15-nucleotide deletion in *H. hirsutus, H. maximiliani, Helianthus mollis, H. tuberosus* and a 18 nucleotide deletion in *Helianthus ciliaris*, were observed. In the K domain of annual species, we found a six-nucleotide deletion and a 15-nucleotide deletion in *Helianthus debilis cucumerifolius* and *H. petiolaris fallax*, respectively. Deletions do not produce frame shifts.

The observed nucleotide sequence variations were studied also in terms of the biochemical and biophysical characteristics of the deduced proteins. Concerning the positions of primers in the *Dhn1* gene, the deduced proteins from amplified sequences are presumably complete, excluding the first methionin, that was added in analyses. Deduced proteins were analyzed comparing their molecular weight, their calculated isoelectric point and their predicted secondary structure (hydrophobicity profiles, percentage of α-helix, extended strand and random coil) between annuals and perennials. No special trend was observed between annuals and perennials concerning hydrophobicity. A tendency for higher isoelectric points (6.86 ± 0.09 vs 6.46 ± 0.11, $P = 0.01$) and molecular weights ($24{,}150 \pm 195.1$ vs $23{,}380 \pm 295.2$, $P < 0.05$) and for minor extensions of the α-helix (9.33 ± 0.51 vs 10.53 ± 0.15, $P < 0.05$) in annuals than in perennials was found. These differences between annuals and perennials indicate functional differentiation of the *Dhn1* gene between these two groups. Biochemical analyses of the different dehydrins purified from the different species will help to confirm this hypothesis.

In conclusion, our data indicate that, within *Helianthus*, the evolution of the dehydrin gene, *Dhn1*, is characterized by a different conservation of protein regions, with consequences on protein structure and, possibly, on function. The largest variability is observed in the glycine-rich domain (G domain) encoding region. Glycine-rich domains are important in determining environmental adaptation, as indicated by the data of Ismail et al. (1999) on a dehydrin of *V. unguiculata*. Though sequence similarity between glycin-rich domain coding sequences of *V. unguiculata* and *Helianthus* is very limited, indicating that these dehydrins are not related, the large variability observed between the compared *Helianthus* species might reflect variability in adaptation to the different environments in which *Helianthus* species live, especially in relation to chilling tolerance which is presumably different between annuals and perennials.

Phylogenetic analyses

Prior to the 1990s, phylogenetic analyses of *Helianthus* have been largely based on morphological characteristics

Table 3 Number of synonymous (n_s) or nonsynonymous (n_a) sites, synonymous (K_s) or nonsynonymous (K_a) nucleotide substitutions per site, obtained by pairwise comparison of sequences of different coding portions of PCR-amplified *Dhn1* sequences from *Helianthus* species (names of the species as in Table 2)

Gene region		Compared species							
		HIR/MAX	HIR/DEBDEB	HIR/ANN	MAX/DEBDEB	MAX/ANN	DEBDEB/ANN	PRAPRA/PRARUN	PETFAL/PETPET
Overall	n_s	164.67	176.42	174.17	165.58	163.33	182.08	183.67	173.17
	K_s	**0.0310**	**0.0967**	**0.1229**	**0.0965**	**0.1282**	**0.1317**	**0.0684**	**0.0175**
	n_a	489.33	516.58	512.83	488.42	484.67	525.92	530.33	504.83
	K_a	**0.0228**	**0.0216**	**0.0344**	**0.0250**	**0.0419**	**0.0330**	**0.0095**	**0.0080**
Y	n_s	32.16	32.16	32.75	32.16	32.75	32.75	32.16	28.83
	K_s	**0.1001**	**0.1557**	**0.1557**	**0.1557**	**0.1557**	**0.2158**	**0.2158**	**0.0000**
	n_a	102.84	102.84	102.25	102.84	102.25	102.25	102.84	88.17
	K_a	**0.0000**	**0.0154**	**0.0000**	**0.0154**	**0.0000**	**0.0154**	**0.0154**	**0.0000**
G	n_s	86.67	98.58	95.83	87.58	84.83	97.75	99.83	97.83
	K_s	**0.0350**	**0.1085**	**0.1735**	**0.0960**	**0.1672**	**0.1611**	**0.0581**	**0.0233**
	n_a	216.33	243.42	240.17	215.42	212.17	241.25	245.17	247.17
	K_a	**0.0319**	**0.0321**	**0.0474**	**0.0274**	**0.0545**	**0.0416**	**0.0136**	**0.0045**
S	n_s	12.50	12.33	12.33	12.17	12.17	12.00	12.00	12.00
	K_s	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**
	n_a	41.50	41.67	41.67	41.83	41.83	42.00	42.00	42.00
	K_a	**0.0498**	**0.0244**	**0.0244**	**0.0754**	**0.0754**	**0.0000**	**0.0000**	**0.0000**
K	n_s	33.33	33.33	33.25	33.66	33.58	39.58	39.66	34.33
	K_s	**0.0000**	**0.0353**	**0.0000**	**0.0349**	**0.0000**	**0.0350**	**0.0000**	**0.0000**
	n_a	128.67	128.67	128.75	128.34	128.42	140.42	140.34	127.50
	K_a	**0.0085**	**0.0085**	**0.0085**	**0.0172**	**0.0172**	**0.0172**	**0.0000**	**0.0086**

Table 4 Number of indels per site, obtained by pairwise comparison of sequences, of different portions of PCR-amplified *Dhn1* sequences from *Helianthus* species (names of the species as in Table 2)

Compared species	Gene portion			
	Overall	Coding	Intron	3′-UTR
HIR/MAX	0.0065	0.0030	0.0077	0.0240
HIR/DEBDEB	0.0117	0.0028	0.0454	0.0315
HIR/ANN	0.0171	0.0097	0.0538	0.0159
MAX/DEBDEB	0.0130	0.0030	0.0500	0.0296
MAX/ANN	0.0189	0.0103	0.0630	0.0148
DEBDEB/ANN	0.0132	0.0062	0.0469	0.0148
PRAPRA/PRARUN	0.0000	0.0000	0.0000	0.0000
PETFAL/PETPET	0.0053	0.0044	0.0077	0.0074

(see Heiser et al. 1969; Rogers et al. 1982). In the last decade, molecular studies have been made to clarify the relationships among *Helianthus* species. However, these studies have been mostly concerned on molecular markers such as RFLP of chloroplast DNA and RAPD (Schilling 1997; Sossey-Alaoui et al. 1998). Sossey-Alaoui et al. (1998) supported the hypothesis that three types of basic genomes exist and are differently combined in *Helianthus* species. RFLP analysis has shown that four sections are present, one including annual *H. agrestis*, another including annual *H. porteri*, a third (sect. *Helianthus*) with all other annuals and a fourth including all perennials (Schilling 1997); section *Helianthus* needs however to be further evaluated, probably due to its recent species differentiation (Schilling 2001). The first sequence data available for systematic and phylogenetic studies are relatively recent (Schilling et al. 1998) and involved comparisons of the rDNA ITS sequence, that proved not to be useful for effective separation of species, due to its lack of variability. Phylogenetic analyses on six *Helianthus* species showed that the rDNA ETS sequence is more informative and promising than ITS for *Helianthus* phylogeny (Linder et al. 2000).

No data were available until now using DNA sequences of coding genes. For our analyses, dendrograms were constructed from isolated nucleotide sequences of *Helianthus* species, using *T. rotundifolia Dhn1* (obtained by PCR as described for *Helianthus* species) as an outgroup, to visualize possible phylogenetic relationships between the isoforms, and including seven *Dhn1* sequences from wild accessions of *H. annuus*. The dehydrin data set was subjected to MP, NJ and ML analyses (Fig. 4).

Phylogenetically, *Dhn1* sequences are more informative in separating the perennial species than in separating annuals. Dehydrins from perennial species *H. tuberosus, H. hirsutus, H. ciliaris, H. maximiliani, H. mollis*) form distinct, highly supported clades, placed in the lowest diverging branches. The apparent separation of perennial species from annuals is in agreement with molecular data on the evolution of the *Helianthus* ITS sequence (Schilling et al. 1998).

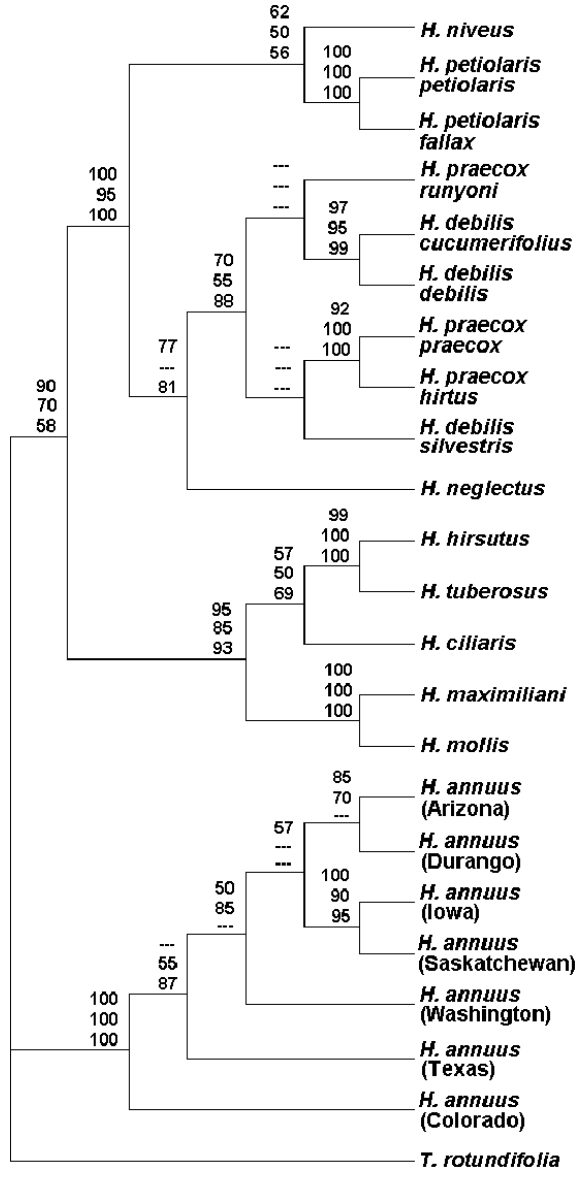

Fig. 4 Fifty percent majority rule bootstrap consensus tree, based on nucleotide sequences corresponding to the *Dhn1* coding portion subjected to MP, ML and NJ analyses (percent bootstrap values are listed as upper, middle and lower numbers, respectively). The tree was rooted using a *T. rotundifolia Dhn* sequence as an outgroup

The separation in distinct clades is less supported, when considering the dehydrins from the annual species. Schilling et al. (1998), studying the ITS sequence, described three clades within annual *Helianthus* species: one including *H. annuus*, a second with *H. petiolaris, Helianthus neglectus* and *Helianthus niveus*, a third with *H. praecox* and *H. debilis*. It is also worth noting that *H. annuus* occupies an intermediate position between annuals and perennials, closer to the outgroup. This might be due to possible paralogy of the isolated *H. annuus* dehydrin gene; however, seven alleles were isolated from different accessions of wild sunflowers and all placed in the same clade. On the other hand, phylogenetic analyses

based on Southern blotting and hybridization with a *copia*-like retrotransposon sequence also indicate that *H. annuus* is placed at the lowest diverging branches of the phylogenetic tree (Cavallini et al. 2002).

Finally, it is to be noted that the subspecies separation within *H. debilis* and *H. praecox* is not completely supported by *Dhn1* phylogenetic data: further analysis, using other DNA sequences are necessary to verify genetic relationships between subspecies.

On the whole, the results from these phylogenetic analyses are encouraging. The dehydrin sequences are evolving quickly enough to provide phylogenetic resolution. The phylogenies presented are largely consistent with results from earlier studies, showing an apparent separation between perennial and annual species, and confirming the data by Schilling (1997). Our data, however, indicate that *H. annuus* is placed closest to *Tithonia*. Chloroplast DNA, nuclear ribosomal genes and isozymes all place *H. annuus* well within the annual sunflowers, while *copia*-like retrotransposon sequences confirm results obtained using *Dhn1*. The data presented in this paper encourage us to design internal primers that can be conveniently used to amplify the *Dhn1* gene from other *Helianthus* species and establish *Helianthus* phylogeny. Analyses of sequence data of other genes will contribute to further clarify the phylogenetic placement of *H. annuus*.

References

Cavallini A, Santini S, Giordani T, Natali L (2002) Use of repeated and coding DNA sequences to study genetic relationship in the genus *Helianthus*. In: Sharma AK, Sharma A (eds) Plant genome – biodiversity and evolution, vol 1. Science Publishers, Inc, Enfield, New Hampshire (in press)

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

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

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

Dure L III, Crouch M, Harada J, Ho T-HD, 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

Gaut BS, Peek AS, Morton BR, Clegg MT (1999) Patterns of genetic diversification within the *Adh* gene family in the grasses (Poaceae). Mol Biol Evol 16:1086–1097

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

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

Heiser CB, Smith DM, Clevenger S, Martin WC (1969) The North American sunflowers. Memoirs Torrey Bot Club 22:1–218

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 96:13,566–13,570

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120

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

Laroche J, Bousquet J (1999) Evolution of the mitochondrial *rps3* intron in perennial and annual Angiosperms and homology to *nad5* intron 1. Mol Biol Evol 16:441–452

Laroche J, Li P, Maggia L, Bousquet J (1997) Molecular evolution of angiosperm mitochondrial introns and exons. Proc Natl Acad Sci USA 94:5722–5727

Li WH (1997) Molecular evolution. Sinauer, Sunderland, Massachusetts

Linder CR, Goertzen LR, Van den Heuvel B, Francisco-Ortega J, Jansen RK (2000) The complete external transcribed spacer of 18S–26S rDNA: amplification and phylogenetic utility at low taxonomic levels in Asteraceae and closely allied families. Mol Phylog Evol 14:285–303

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

Natali L, Giordani T, Cavallini A (2003) Sequence variability of a dehydrin gene within *Helianthus annuus*. Theor Appl Genet 106:811–818

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 J-M, 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

Rieseberg LH (1995) The role of hybridization in evolution: old wine in new skins. Am J Bot 82:944–953

Rogers CE, Thompson TE, Seiler GJ (1982) Sunflower species of the United States. National Sunflower Assoc, Fargo, USA

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, New York, USA

Schilling EE (1997) Phylogenetic analysis of *Helianthus* (Asteraceae) based on chloroplast DNA restriction-site data. Theor Appl Genet 94:925–933

Schilling EE (2001) Phylogeny of *Helianthus* and related genera. Oleagineaux Corps Gras Lipides 8:22–25

Schilling EE, Heiser CB (1981) Infrageneric classification of *Helianthus* (*Compositae*). Taxonomy 30:393–403

Schilling EE, Linder CR, Noyes RD, Rieseberg LH (1998) Phylogenetic relationships in *Helianthus* (Asteraceae) based on nuclear ribosomal DNA internal transcribed spacer region sequence data. Syst Bot 23:177–187

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. Proteins Struct Funct Genet 8:103–117

Soltis ED, Soltis PS (2000) Contributions of plant molecular systematics to studies of molecular evolution. Plant Mol Biol 42:45–75

Sossey-Alaoui K, Serieys H, Tersac M, Lambert P, Schilling EE, Griveau Y, Kaan F, Bervillé A (1998) Evidence for several genomes in *Helianthus*. Theor Appl Genet 97:422–430

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

Wendel JF, Albert VA (1992) Phylogenetics of the cotton genus (*Gossypium*): character-state weighted parsimony analysis of chloroplast-DNA restriction site data and its systematic and biogeographic implications. Syst Bot 17:115–143

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, New Jersy, pp 131–141

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