

bin, T. Staehelin and J. Gordon, Electrophoretic of proteins from polyacrylamide gels to nitrocellulose: procedure and some application. *Proc. Acad. Sci. USA*, 76 (1979) 4350.

Bradford, A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing a principle of protein-dye binding. *Anal. Biochem.*, 72 248-254.

Clatworthy and T.C. Hall, Histochemical analysis of 35S promoter- β -glucuronidase gene expression in transgenic rice plants. *Plant Mol. Biol.*, 15 (1990) 8.

Endo and K. Shimamoto, Expression of CaMV 35S promoter in transgenic rice plants. *Mol. Gen. Genet.*, 260 389-392.

Elroy, M. Rothenberg, K.S. Reece and R. Wu, Characterization of the rice (*Oryza sativa*) actin gene. *Plant Mol. Biol.*, 15 (1990) 257-268.



Plant Science 116 (1996) 85-96

PLANT
SCIENCE

Seed storage proteins of Cupressaceae are homologous to legumins from angiosperms: Molecular characterization of cDNAs from incense cedar (*Calocedrus decurrens* [Torr.] Florin)

Klaus-Peter Häger*, Norbert Dank

University of Bayreuth, Department of Plant Ecology and Systematics, Universitätsstr. 30, D-95440 Bayreuth, Germany

Received 25 September 1995; revision received 30 January 1996; accepted 30 January 1996

Abstract

Legumins are shown to be the predominant storage proteins present in the detergent soluble fractions of seeds from incense cedar and from all other Cupressaceae investigated. The subunits of these proteins are made up of α - and β -polypeptides with molecular weights in the range of 30-34 kDa and 23-25 kDa, respectively. Unusual for legumins, those from incense cedar are glycosylated. An incense cedar endosperm cDNA library was constructed and three different legumin cDNAs were characterized; they represent two divergent legumin subfamilies with nucleotide similarities of about 71%. Comparisons with legumins from gymnosperms and angiosperms reveal between 57% and 34% identity on the amino acid level. The deduced amino acid sequences suggest that incense cedar legumins are synthesized as precursors to be processed similarly as is known from those of angiosperms. However, none of them contains the post-translational cleavage site, Asn-Gly, which is assumed to be highly conserved in angiosperm legumin precursors. A phylogenetic analysis suggests that both legumin subfamilies described here originated relatively soon after divergence of the genes derived from the common ancestor of Pinaceae and Cupressaceae legumins. Moreover, legumins prove to be reliable molecular markers to elucidate seed plant relationships.

Keywords: Gymnosperms; Seed storage proteins; Legumin; Evolution

1. Introduction

During development of angiosperm seeds, large amounts of storage proteins are synthesized and accumulated in the storage tissues of seeds, providing a source of carbon, sulfur and nitrogen for the initial growth of the seedling. According to their solubilities, sedimentation coefficients and

structures, seed storage proteins have been divided into a series of families. The members of the legumin (11S protein) family, first described from Fabaceae [1], are hexameric proteins composed of acidic α -polypeptides and basic β -polypeptides with molecular masses of about 40 and 20 kDa, respectively. Both disulfide-linked polypeptides are derived from a single precursor by co- and post-translational cleavages. Due to their importance for human and livestock nutrition, these pro-

* Corresponding author.

teins, their biosynthesis and the genes encoding them have been intensively studied in legumes and other crop plants [2–5]. Similar proteins with significant legumin-like traits have been described from numerous species [1,6], demonstrating their widespread, if not ubiquitous, occurrence in angiosperm seeds. From comparative sequence analyses, it was suggested that these proteins have arisen from a common progenitor and that the ancestral gene that gave rise to modern forms already existed at the beginning of angiosperm evolution [7]. However, the phylogenetic origin and the molecular evolution of legumins and of the genes encoding them still remain widely unknown.

In contrast to angiosperms, only little attention has been drawn to seed proteins from gymnosperms. From electrophoretic and immunological analyses, as well as from amino acid sequencing, the occurrence of legumins in seeds from *Ginkgo biloba* [8,9] and from species of Pinaceae [10–12] has been demonstrated. To date, nucleotide sequences from legumin genes or legumin cDNAs have only been described from *Ginkgo* [13,14] and from the Pinaceae *Pseudotsuga menziesii* (Douglas fir) [15]; besides, there are unpublished sequences of Pinaceae (Loblolly pine, White spruce) and Gnetales available from the EMBL database.

Thus, the occurrence of legumins still remains uncertain for the majority of gymnosperm families. Many of the present-day conifers represent very ancient lineages [16]; therefore, the characterization of conifer legumins and of the genes encoding them should provide new insights into the fundamental organization, the evolution and the relationships of legumins from gymnosperms and angiosperms. To contribute to an improved understanding of these topics, we have conducted an electrophoretic analysis of seed storage proteins from incense cedar and from other members of the Cupressaceae family. We report the cloning and sequence analysis of three incense cedar legumin cDNAs representing two legumin subfamilies; this is the first report about the molecular structure of legumins from Cupressaceae. Finally, a phylogenetic analysis is presented to elucidate the evolutionary relationships of incense cedar legumins to those from gymnosperms and angiosperms.

2. Materials and methods

2.1. Plant material

For RNA isolation, mid-maturation stage seeds were harvested from an individual *Calocedrus decurrens* tree located at the Herrenhauser Garden, Hannover. Endosperms were isolated, immediately frozen in liquid nitrogen and stored at -70°C .

Mature seeds used for protein extractions were obtained from the Botanical Garden, Munich and from Conrad Appel GmbH, Darmstadt.

2.2. Protein extractions and gel electrophoresis

Extraction of soluble and insoluble seed proteins was performed essentially as described in [11]. Briefly, seed coats were removed and seeds were ground for 30 min in Tris-glycine-buffer (10 mM/80 mM), pH 8.2, containing 2% (w/v) NaCl, using a ratio of 0.3–0.5 g seeds to 1 ml buffer. The extracts were centrifuged (20 min, $40\,000 \times g$) and the supernatants, containing the soluble proteins, were removed. The pellets were re-extracted twice to remove all soluble proteins. Insoluble proteins were extracted by grinding the pellets in 1.0 volume of 65 mM Tris/HCl buffer, pH 6.8, containing 2% (w/v) sodium dodecyl sulfate (SDS) and 10% (v/v) glycerol, and boiling for 8 min. The slurries were centrifuged as above and the supernatants were subjected to electrophoresis. Samples of the extracts were boiled for 5 min in 62 mM Tris/HCl, pH 6.8, 3% SDS (w/v), 10% (v/v) glycerol and, if reducing conditions were required, 30 mM dithioerythritol. SDS-polyacrylamide gel electrophoresis (SDS-PAGE) was carried out in 7.5%–20% acrylamide gradient gels (Mighty Small system, Hoefer) containing 30 mM Tris, 75 mM borate, 0.1% (w/v) SDS, pH 8.8 and pH 7.9, respectively. Gels were run for 2 h at 110 V, and protein bands were visualized by staining with Coomassie brilliant blue (Serva). Molecular weight protein standards (Pharmacia) used were: phosphorylase *b* (94 kDa), bovine serum albumin (BSA) (67 kDa), ovalbumin (43 kDa), carbonic anhydrase (30 kDa), trypsin inhibitor (20.1 kDa) and α -lactalbumin (14.4 kDa). Two-dimensional electrophoresis was performed as described in [8].

2.3. Western blot analysis and gel

Seed proteins separated by electrophoretically transferred (electrophoresis unit, LKB) to nitrocellulose membranes (Schleicher & Schuell) a. The membranes were blocked with gentle shaking in 20 mM Tris, pH 7.2, containing 0.9% (w/v) NaCl and 0.5% (w/v) fat milk powder. Antiserum was diluted in blocking buffer and blots were incubated at 40°C . After washing with blocking buffer (3 min), the membranes were incubated at 40°C with goat-anti-rabbit IgG conjugated with horseradish peroxidase (HRP) (Biorad) using a 1:3000 dilution in buffer. The membranes were washed above and stained with 3,3'-diaminobenzidine tetrahydrochloride (Fluka) (0.05% (0.3 mg/ml) and H_2O_2 30% w/v (0.3 mg/ml) in Tris/HCl, pH 7.2, containing 0.05% (0.3 mg/ml) H_2O_2 . Colour development was stopped with distilled water. Glycoprotein staining was performed as described in [20].

2.4. Construction and screening of a cDNA library and sequencing of

Total RNA was extracted from incense cedar seeds collected at 10 days after pollination according to the method described by Cerff and Klotz [17]. To remove phenolic constituents, two extraction steps (20 min each) with 10% (v/v) Polyclar (Serva) were performed. Poly(A)⁺ RNA was isolated from total RNA by two passes through an oligo(dT) cellulose column (7, Pharmacia). cDNA was synthesized using a Pharmacia kit using 5 μg poly(A)⁺ RNA. Strand synthesis was monitored by the incorporation of [α - ^{32}P]dCTP (3000 Ci/mmol, Pharmacia) into double-stranded cDNA using *NotI*/*EcoRI* linkers, size fractionation on agarose gels, and cDNAs >500 bp were isolated. Recombinant phages were screened by plaque hybridization on 3 $\times$ SSPE using a random-primer synthesized fragment of the legumin gene from *Ginkgo biloba* [14]; washings were performed at 55°C in 2 $\times$ SSPE. Several hybrid

Materials and methods

Plant material

RNA isolation, mid-maturation stage seeds harvested from an individual *Calocedrus* tree located at the Herrenhauser n, Hannover. Endosperms were isolated, liately frozen in liquid nitrogen and stored at .
ture seeds used for protein extractions were ed from the Botanical Garden, Munich and Conrad Appel GmbH, Darmstadt.

Protein extractions and gel electrophoresis

raction of soluble and insoluble seed pro- was performed essentially as described in Briefly, seed coats were removed and seeds ground for 30 min in Tris-glycine-buffer (10 mM), pH 8.2, containing 2% (w/v) NaCl, a ratio of 0.3–0.5 g seeds to 1 ml buffer. The uts were centrifuged (20 min, 40 000 × g) he supernatants, containing the soluble pro- were removed. The pellets were re-extracted to remove all soluble proteins. Insoluble pro- were extracted by grinding the pellets in 1.0 ne of 65 mM Tris/HCl buffer, pH 6.8, con- ing 2% (w/v) sodium dodecyl sulfate (SDS) and (v/v) glycerol, and boiling for 8 min. The slur- were centrifuged as above and the superna- were subjected to electrophoresis. Samples of xtracts were boiled for 5 min in 62 mM HCl, pH 6.8, 3% SDS (w/v), 10% (v/v) gly- and, if reducing conditions were required, 30 lithioerythritol. SDS-polyacrylamide gel elec- roesis (SDS-PAGE) was carried out in –20% acrylamide gradient gels (Mighty Small m, Hoefer) containing 30 mM Tris, 75 mM e, 0.1% (w/v) SDS, pH 8.8 and pH 7.9, ctively. Gels were run for 2 h at 110 V, and in bands were visualized by staining with nassie brilliant blue (Serva). Molecular ht protein standards (Pharmacia) used were: phorylase *b* (94 kDa), bovine serum albumin) (67 kDa), ovalbumin (43 kDa), carbonic drase (30 kDa), trypsin inhibitor (20.1 kDa) α-lactalbumin (14.4 kDa). Two-dimensional rophoresis was performed as described in [8].

2.3. Western blot analysis and glycoprotein staining

Seed proteins separated by SDS-PAGE were electrophoretically transferred (Multiphor II electrophoresis unit, LKB) to nitrocellulose membranes (Schleicher & Schuell) at 100 mA for 1 h. The membranes were blocked overnight at 4°C with gentle shaking in 20 mM Tris/HCl buffer, pH 7.2, containing 0.9% (w/v) NaCl and 5% (w/v) non-fat milk powder. Antiserum was diluted 1:5000 in blocking buffer and blots were incubated for 2 h at 40°C. After washing with blocking buffer (3 × 10 min), the membranes were incubated for 1.5 h at 40°C with goat-anti-rabbit immunoglobulin G (IgG) conjugated with horseradish peroxidase (HRP) (Biorad) using a 1:3000 dilution in blocking buffer. The membranes were washed as described above and stained with 3,3'-diaminobenzidine tetrahydrochloride (Fluka) (0.3 mg/ml), CoCl₂ (0.3 mg/ml) and H₂O₂ 30% w/v (1 μl/ml) in 20 mM Tris/HCl, pH 7.2, containing 0.9% (w/v) NaCl. Colour development was stopped by washing with distilled water. Glycoprotein staining was performed as described in [20].

2.4. Construction and screening of an incense cedar cDNA library and sequencing of cDNAs

Total RNA was extracted from endosperm of incense cedar seeds collected approximately 140 days after pollination according to a method described by Cerff and Kloppstech [18]. To remove phenolic constituents, two additional incubation steps (20 min each) with 10% (w/v) Polyclar (Serva) were performed. Poly(A)⁺-RNA was isolated from total RNA by two successive passes through an oligo(dT) cellulose column (type 7, Pharmacia). cDNA was constructed with the Pharmacia kit using 5 μg poly(A)⁺-RNA and first strand synthesis was monitored by the incorporation of [α-³²P]dCTP (3000 Ci/mMol; NEN). Double-stranded cDNA was ligated with *NotI*/*EcoRI* linkers, size fractionated on 1% agarose and cDNAs >500 bp were cloned into λ NM1149 [19]. Recombinant phage clones were screened by plaque hybridization [20] at 55°C in 3 × SSPE using a random-prime-labeled [20] fragment of the legumin gene GbLeg1 from *Ginkgo biloba* [14]; washings were performed at 55°C in 2 × SSPE. Several hybridising cDNAs

were isolated and subcloned into the plasmid vector Bluescript pKS+ (Stratagene). Ordered sets of deletion clones were prepared using exonuclease III, and suitable clones were sequenced on both strands by the dideoxy chain termination method [20]. Sequence data were processed with the GCG software package.

2.5. Phylogenetic analysis

A neighbor-joining distance analysis was performed using the PHYLIP program package, version 3.5c (J. Felsenstein, Department of Genetics, University of Washington, 1993). Sources of sequences included: *Ginkgo* and *Magnolia*, [14,17]; *Calocedrus*, this paper; the following sequences, specified by the accession numbers given, were taken from the EMBL data library: *Pseudotsuga menziesii* L07484, *Picea glauca* X63192, *Pinus strobus* Z11486, *Oryza sativa* A X54313, *Oryza sativa* B X54314, *Avena sativa* J05485, *Raphanus sativus* S47559, *Arabidopsis thaliana* X14313, *Prunus amygdalus* X78119, *Vicia faba* A X55013, *Vicia faba* B X03677, *Pisum sativum* A X00634, *Pisum sativum* J Y00715. Computing was carried out on a VAX 6000-310 and on an Apple Macintosh IIfx.

3. Results and discussion

3.1. Legumins are major storage proteins of incense cedar seeds

According to their solubility properties, legumins from angiosperms and gymnosperms can be divided into two main classes: globulin type legumins are soluble in salt solutions and represent the majority of legumins known from angiosperms [1,5]; glutelin or crystalloid type legumins require dilute alkaline solutions or the addition of SDS, urea or other denaturing agents for complete dissolution; they have been reported for *Oryza* [21], *Quercus* [22], for Euphorbiaceae [23] and for Pinaceae [10–12].

In order to investigate the occurrence of legumins in Cupressaceae, megagametophytes and embryonic axes of incense cedar were sequentially extracted for salt-soluble and SDS-soluble proteins and both fractions were characterized by SDS-PAGE analysis. The salt-soluble fraction

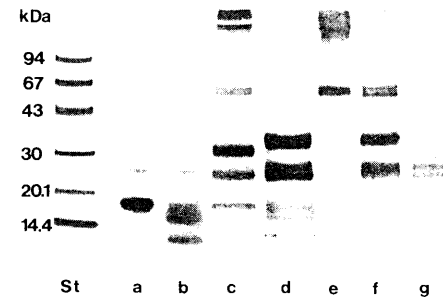


Fig. 1. Characterization of incense cedar seed storage proteins: SDS-PAGE of salt soluble (a, b) and SDS soluble (c, d) proteins under non-reducing and reducing conditions, respectively; Western blot analysis (e, f) of SDS soluble proteins separated under non-reducing and reducing conditions, probed with antibodies raised against a *Metasequoia* legumin subunit; glycoprotein staining (g) of SDS soluble proteins separated under reducing conditions. Standard proteins (St) and their molecular weights are shown on the left lane.

contains a prevalent protein with a molecular weight of about 17 kDa (Fig. 1, lane (a)). In contrast to the minor proteins, the 17 kDa protein is affected by DTE reduction, indicating the presence of disulfide bonded subunits (Fig. 1, lane (b)). Similar as in incense cedar, the 17 kDa proteins represent a major component of salt soluble seed proteins of Pinaceae [10–12]. Based on the polypeptide size and quaternary structure, it was assumed that these proteins are homologous to 2S seed storage proteins known from angiosperms [11,12]. Analysis of the SDS-soluble proteins under non-reducing conditions (Fig. 1, lane c) reveals that the most abundant proteins are in the 23 kDa and 30 kDa range and, in particular, in the high molecular weight region (>94 kDa). In addition, small amounts of proteins of 58 kDa and 17 kDa are detected, the latter obviously representing residual amounts of the prominent salt soluble protein. The reducing gel (Fig. 1, lane d) shows that the slowly migrating proteins (>94 kDa) and the 58 kDa proteins are replaced by two major sets of proteins with respective molecular weights ranging from 30 to 34 kDa and 23 to 25 kDa (the corresponding bands are partially superposed on those which are located in these regions under

nonreducing conditions). This behaviour closely resembles those described for the glutelin- or crystalloid-type legumins from the species mentioned above, in particular those from Pinaceae [10,11,12]: their legumin subunits are made up of disulfide linked α - and β -polypeptides with molecular weights of 30–35 kDa and 21–23 kDa, respectively. From the similarity in solubility, breakdown patterns upon reduction and the molecular weights of the constituent polypeptides, it is evident that the major SDS-soluble seed proteins from incense cedar should be homologous to legumins.

Nevertheless, there is a significant difference when incense cedar legumins are compared with those from other species. In general, legumins form hexamers, and their constituent subunits are held together by noncovalent forces [1]. When extracted under denaturing conditions, legumins therefore dissociate into the monomeric subunits which on reduction yield the characteristic α - and β -polypeptides. In incense cedar, however, the high molecular weight proteins (>94 kDa) are the major source of α - and β -polypeptides rather than the monomeric 58 kDa subunits. This is especially evident from two-dimensional SDS-PAGE analysis (non-reducing versus reducing conditions): the high molecular weight proteins are reduced to two polypeptide subgroups with molecular weights in the same ranges as those originating from the monomeric subunit (Fig. 2). Obviously, incense cedar high molecular weight proteins are oligomeric legumin subunits linked by intermolecular disulfide bridges or not fully accessible to SDS under nonreducing conditions. Interestingly, a similar finding has been reported for legumin subunits from Castor bean [24].

Additional support for the occurrence of legumins in incense cedar and for the existence of oligomeric legumin subunits under denaturing conditions is presented from Western blot analysis. Antibodies raised against the 60 kDa legumin subunit from the Taxodiaceae *Metasequoia glyptostroboides* — the identity of the protein has been confirmed by amino acid sequencing data (unpublished results) — were used for immunodetection of incense cedar SDS soluble seed proteins. As expected, these antibodies cross-reacted with the

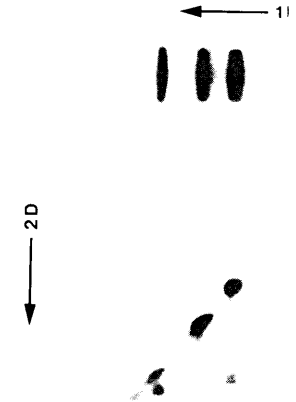


Fig. 2. Two dimensional gel electrophoresis of incense cedar seed storage proteins from incense cedar. The horizontal gel strip at the top shows the separation of the proteins in the first dimension (1D). The spots in the second dimension (2D) are affected by reduction. The spots indicated by an open arrow are located at the origin of the monomeric subunit. Note size differences in the 23–25 kDa region.

high molecular weight proteins with the 58 kDa protein, as well as with the 23–25 kDa proteins under reducing conditions (Fig. 1, lane f). Moreover, the pronounced cross-

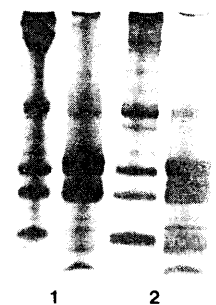
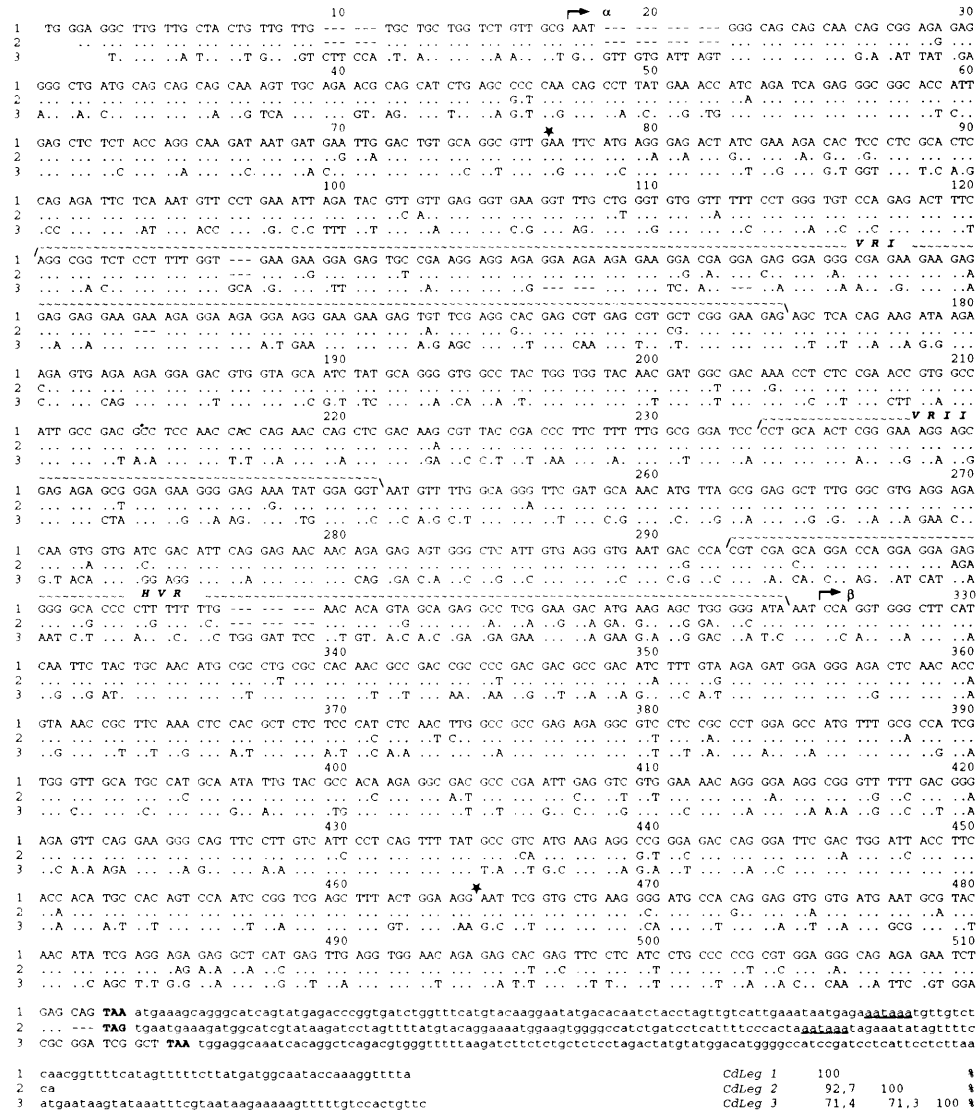


Fig. 3. Electrophoretic profiles of SDS soluble seed proteins from *Thuja orientalis*, (2) *Thuja plicata*, (3) *Thuja hodginsii*. Left lanes, non-reducing conditions; right lanes, reducing conditions.



[illegible]

icates the sequences of correspondingly
lentities relative to CdLegI are marked
ncoding regions are indicated by angled
, VR II) and to the hypervariable region
for the group I subfamily are indicated
ylation signals are underlined. Sequence

Since the species investigated represent four out of six tribes and both subfamilies (Cupressoideae and Callitroideae) of the Cupressaceae [26], our findings indicate the widespread occurrence of legumins within this family and suggest a very conservative mode of legumin evolution in the Cupressaceae.

3.3. Incense cedar legumin cDNAs can be divided into two subfamilies

A cDNA library was constructed from poly(A)⁺-RNA isolated from incense cedar endosperm at the middle stage of seed development. About 100 000 recombinant phage were screened by hybridization under moderate stringency using

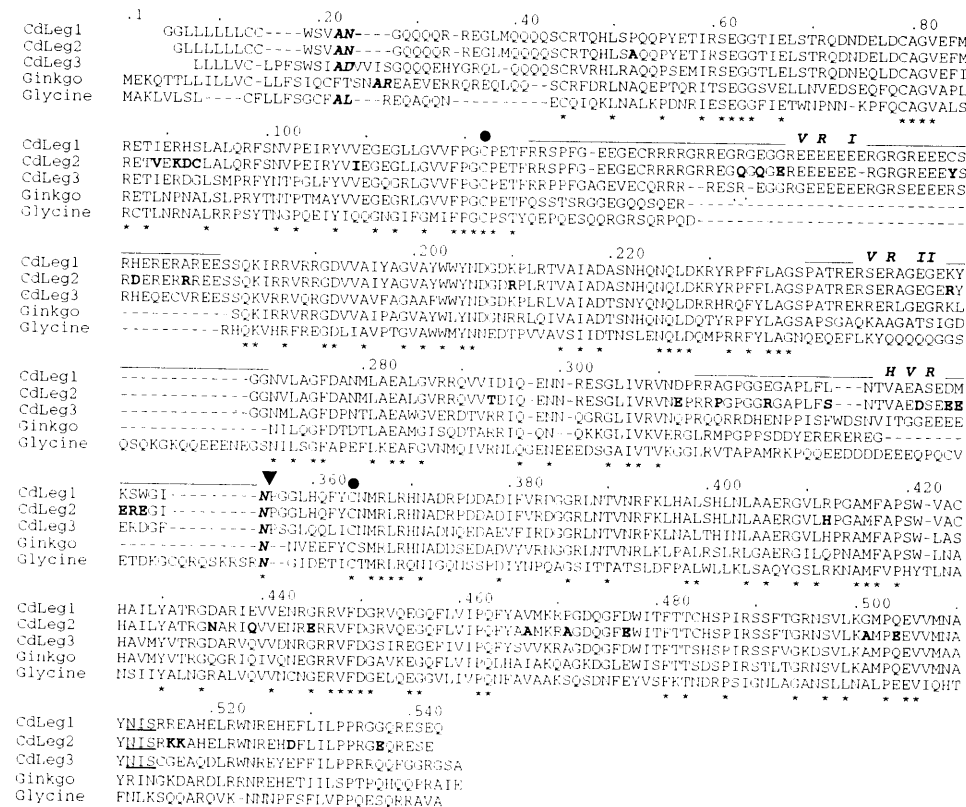


Fig. 5. Comparison of legumin precursors from incense cedar, aligned with those from *Ginkgo* and *Glycine*. Amino acid replacements within the incense cedar group I subfamily are denoted by bold type letters shown in the CdLcg2 sequence, glycosylation signals are underlined. Residues flanking the cotranslational processing sites are shown in bold italics. The arrowhead indicates post-translational processing sites, adjacent Asn residues indispensable for cleavage are in bold italics. Residues strictly conserved are marked by asterisks below the sequences, filled circles indicate cysteine residues probably involved in interpolypeptide linkage. Designations VR I, VR II and HVR are used as specified in Fig. 4. Sequences from *Ginkgo* and *Glycine* are GbLcg1 [14] and Gy2 [27].

a fragment of the *Ginkgo biloba* legumin gene GbLeg1 recently described [14]. The probe codes for the region spanning amino acid positions 370 to 511 in the primary sequence alignment shown in Fig. 5. Forty positive clones were purified, and they could be divided into two groups differing in hybridization intensities and in restriction digestion patterns. The inserts of phage clones showing weak hybridization signals contained internal *EcoRI* sites but were recovered intact when excised with *NotI* (group I clones). In contrast, the inserts of clones giving rise to strong signals remained intact when excised with *EcoRI* or *NotI*, respectively (group II clones). From each group, two of the longest cDNAs detectable, ranging in size from 1.6 to 1.7 kb, were cloned into the Bluescript plasmid vector and submitted to sequence analysis. The group I clones are slightly different from each other, whereas those from group II are identical in sequence. Open reading frames of 1505 bp (CdLeg1), 1496 bp (CdLeg2) and 1526 bp (CdLeg3) encoding legumin precursors of group I (CdLeg1, CdLeg2) and of group II (CdLeg3) were obtained. The clones are truncated in their 5' ends lacking the coding information for the *N*-terminal part of signal peptides; however, they contain the complete coding sequences of α - and β -polypeptides (codon positions 18–325 and 326–512, -511 and -514, respectively; Fig. 4) derived from precursors as discussed below. The 3' flanking regions of clones CdLeg1 and CdLeg2 contain a single potential polyadenylation signal AATAAA beginning 94 and 86 bp downstream from their stop codons, respectively. From the alignment in Fig. 4, it is obvious that CdLeg1 and CdLeg2 are more closely related than either one is to CdLeg3. The coding regions of the CdLeg1 and CdLeg2 clones show extensive sequence similarity (ca. 93% identity) on the nucleotide level. With respect to the nucleotide changes observed, 92 codons are involved and 57 of them differ exclusively at the third positions. Altogether, about 41% of the accepted mutations result in amino acid replacements. In addition, the CdLeg2 coding region shows two codon deletions at positions 154 and 512 when compared with that from CdLeg1. The similarity between both clones extends to their 3' noncoding regions with increasing divergence towards the 3' ends.

In contrast, the CdLeg3 coding sequence shows only about 71% identity with respect to those from CdLeg1 and CdLeg2. Additional changes result from insertions or deletions which are mainly located in regions corresponding to codon positions 121–174 and 293–324, respectively (Fig. 4). These regions are homologous to the variable region I and to the hypervariable region described for angiosperm legumins [27], where they display considerable heterogeneity in sequence and size. A further region of heterogeneity, designated as variable region II [27], is observed in angiosperm legumins and has been marked in Fig. 4. Interestingly, the sequences of incense cedar legumin precursors are well conserved in this region.

In angiosperms, legumins are commonly encoded by small multigene families comprising two or three types of gene subfamilies [2–5]. Likewise, this feature is also observed in gymnosperms: Southern blot analysis has revealed the presence of small legumin gene families in *Ginkgo biloba* [14] and in Douglas fir [15] and the cDNAs reported here suggest that incense cedar legumins are encoded by two divergent gene subfamilies.

3.4. The predicted structures of incense cedar legumin precursors

The primary sequences derived from incense cedar CdLeg1, CdLeg2 and CdLeg3 cDNAs have been aligned and are shown in Fig. 5. For comparison, legumin precursors from *Ginkgo biloba* [14] and soybean [27], covering a wide phylogenetic distance, have been included in the alignment. Sequence analyses using the GCG BESTFIT program revealed 70.1% identity between the CdLeg1 and CdLeg3 polypeptide sequences. About 48% to 57% identity are observed when they are compared with legumin precursors from *Ginkgo* [14] or from the Pinaceae species given in Fig. 6. When comparisons were performed with legumin precursors from angiosperms, the values for the highest and lowest similarities are 42% with *Magnolia* B14 [17] and 34% with legumins from Fabaceae, e.g., those referred to in Fig. 6. These identities clearly verify the incense cedar polypeptide sequences to be homologous with legumin precursors.

During their transport to protein bodies, legumin precursors undergo two successive cleavages: (i) *N*-terminal signal peptides of about

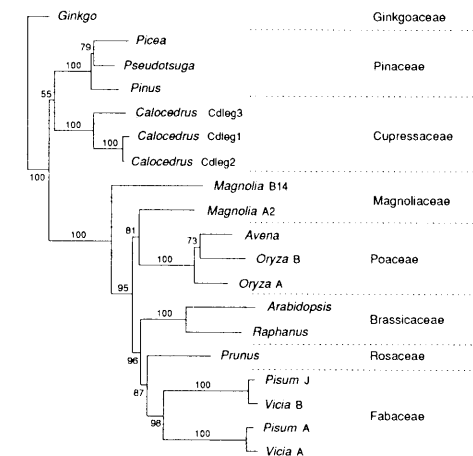


Fig. 6. Evolutionary relationships of incense cedar legumins to those from gymnosperms and angiosperms. The dendrogram results from a neighbor-joining analysis. Only first and second codon positions were included, the data set is based on 744 homologous nucleotide positions. The numbers left of the nodes are bootstrap values indicating frequencies of respective furcations found in 100 replications of subset tree calculations. For references of the sequences, see Materials and methods.

25 amino acid residues in length and directing the nascent precursor into the lumen of the endoplasmic reticulum are cotranslationally removed; (ii) upon entry into the protein storage vacuoles, the precursors are cleaved to yield α - and β -subunits typical for mature legumin subunits [2,3,5]. We assume that incense cedar legumin precursors are processed in a similar way. The leucine residues repeated in their *N*-terminal regions (around position 10, Fig. 5) are considered as hydrophobic cores typical for signal peptides. Using the algorithm of von Heijne [28], the signal peptidase cleavage sites have been determined to be located between Ala 22 and Asn 23 (CdLeg1, CdLeg2) and Ala 22 and Asp 23 (CdLeg3). The post-translational cleavage site and the adjacent sequence, representing the *N*-terminal region of the prospective β -polypeptide, are highly conserved in legumin precursors from angiosperms [7]. A similar region, which is homologous in sequence

and po
legumin
given in
legumin
ed the s
Asn 35
Appa
post-tra
sors is
resulting
polypep
had be
[7,27,30
N-termi
legumin
been re
legumin
and *Gin*
nucleoti
legumin
here, th
the prec
sequence
Asn-Gly
or addi
Ginkgo
ings and
dance v
using s
processi
the Asp
teases in
whereas
acid res
viously,
legumin
assumed
CdLeg1
mature
305 ami
polypep
kDa ea
molecul
obtained
d). The
observe
counted
values
polypep

ist, the CdLeg3 coding sequence shows 71% identity with respect to those from CdLeg2. Additional changes result in deletions or deletions which are mainly in regions corresponding to codon positions 174 and 293–324, respectively (Fig. 4). These regions are homologous to the variable region described for term legumins [27], where they display a high degree of heterogeneity in sequence and size. A region of heterogeneity, designated as region II [27], is observed in angiosperm legumins and has been marked in Fig. 4. Inter-sequence comparisons of incense cedar legumin precursors are well conserved in this region. In gymnosperms, legumins are commonly encoded by multigene families comprising two or three subfamilies [2–5]. Likewise, a similar pattern is also observed in gymnosperms: a cluster analysis has revealed the presence of three main gene families in *Ginkgo biloba* [14] and Douglas fir [15] and the cDNAs reported here suggest that incense cedar legumins are also represented by two divergent gene subfamilies.

predicted structures of incense cedar legumin precursors

Primary sequences derived from incense cedar CdLeg1, CdLeg2 and CdLeg3 cDNAs have been deduced and are shown in Fig. 5. For comparison, legumin precursors from *Ginkgo biloba* [14] and soybean [27], covering a wide phylogenetic range, have been included in the alignment. Analyses using the GCG BESTFIT program revealed 70.1% identity between the CdLeg1 and CdLeg3 polypeptide sequences. About 48% identity is observed when they are compared with legumin precursors from *Ginkgo* [14] or from soybean species given in Fig. 6. When comparisons were performed with legumin precursors from gymnosperms, the values for the highest and lowest similarities are 42% with *Magnolia B14* [17] and 21% with legumins from Fabaceae, e.g., those from *Vicia* in Fig. 6. These identities clearly verify the similarity of incense cedar polypeptide sequences to be compared with legumin precursors.

For their transport to protein bodies, legumin precursors undergo two successive steps: (i) *N*-terminal signal peptides of about

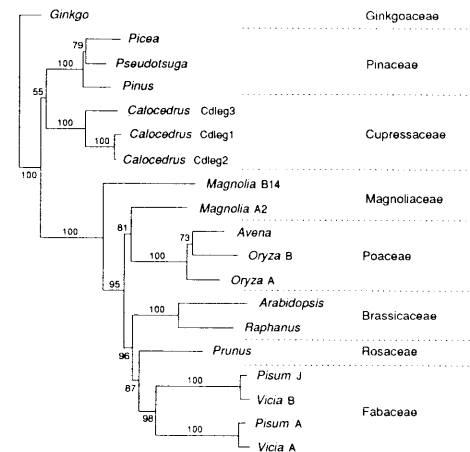


Fig. 6. Evolutionary relationships of incense cedar legumins to those from gymnosperms and angiosperms. The dendrogram results from a neighbor-joining analysis. Only first and second codon positions were included, the data set is based on 744 homologous nucleotide positions. The numbers left of the nodes are bootstrap values indicating frequencies of respective furcations found in 100 replications of subset tree calculations. For references of the sequences, see Materials and methods.

25 amino acid residues in length and directing the nascent precursor into the lumen of the endoplasmic reticulum are cotranslationally removed; (ii) upon entry into the protein storage vacuoles, the precursors are cleaved to yield α - and β -subunits typical for mature legumin subunits [2,3,5]. We assume that incense cedar legumin precursors are processed in a similar way. The leucine residues repeated in their *N*-terminal regions (around position 10, Fig. 5) are considered as hydrophobic cores typical for signal peptides. Using the algorithm of von Heijne [28], the signal peptidase cleavage sites have been determined to be located between Ala 22 and Asn 23 (CdLeg1, CdLeg2) and Ala 22 and Asp 23 (CdLeg3). The post-translational cleavage site and the adjacent sequence, representing the *N*-terminal region of the prospective β -polypeptide, are highly conserved in legumin precursors from angiosperms [7]. A similar region, which is homologous in sequence

and position with the *N*-terminal domains of legumin β -polypeptides from *Ginkgo* and *Glycine* given in Fig. 5, can be recognized in incense cedar legumin precursors too. Therefore, we have assigned the site for post-translational cleavage between Asn 355 and Pro 356, respectively.

Apparently, our assignment is unusual in that post-translational processing of legumin precursors is generally directed to an Asn-Gly junction, resulting in an *N*-terminal Gly of the β -polypeptide; it was assumed that this target site had been highly conserved during evolution [7,27,30]. However, March et al. [29] described an *N*-terminal Phe for the β -polypeptide of a minor legumin from *Pisum sativum*. Ala and Asn have been reported as unusual *N*-terminal residues of legumin β -polypeptides from *Pinus pinaster* [12] and *Ginkgo biloba* [9,13], respectively. At present, nucleotide sequences coding for these unusual legumins are only known from *Ginkgo* [13,14]; here, the Asn-Asn processing sites suggested by the precursors deduced agree with the amino acid sequencing data mentioned above. (Preceding Asn-Gly junctions which could provide alternative or additional cleavage sites are not present in *Ginkgo* nor in *Calocedrus* precursors.) These findings and our assignment stated above are in accordance with recent results from cleavage studies using synthetic oligopeptides or precursors with processing sites mutated in vitro: with respect to the Asn-Gly target site, it appears that the proteases involved show an absolute specificity for Asn whereas replacements of Gly by various amino acid residues do not hinder cleavage [13,30]. Obviously, the post-translational cleavage site of legumin precursors is less conserved than hitherto assumed. The processing predicted for the CdLeg1, CdLeg2 and CdLeg3 precursors yields mature α -polypeptides composed of 300, 299 and 305 amino acids (34.1, 34.2 and 35.2 kDa) and β -polypeptides of 187, 186 and 189 amino acids (21.4 kDa each). The close similarities of the deduced molecular weights are in agreement with the data obtained from SDS-PAGE analysis (Fig. 1, lane d). The size heterogeneity of β -polypeptides observed in 2D analysis (Fig. 2) is not fully accounted for by the cDNAs described here. The pI values calculated are 5.0 to 5.5 for the α -polypeptides and 10.3 to 10.5 for the β -

polypeptides, thus ranging similar to those from angiosperm legumins [1].

The deduced amino acid compositions of incense cedar legumin precursors present low levels of cysteine (1.4%–1.6%) and methionine (1.6%–1.8%). However, they are rich in arginine, lysine and amide amino acids (together about 25%), consistent with the role of legumins as a nitrogen source for the young seedling. The imbalance of sulfur and nitrogen observed here is also known from angiosperm legumins [5].

Most likely, the cysteine residues located at position 122 and 364 (Fig. 5) are involved in disulfide linkage of the mature α - and β -polypeptides. Cysteine residues at homologous positions are strictly conserved in angiosperm legumins [27] and it has been shown that the same cysteine residues are involved in interpolypeptide linkage in soybean legumin subunits. Legumin precursors from incense cedar and those from angiosperms share a common arrangement of conserved and variable regions. Domains of the conserved regions might be essential for interactions between subunits or for their transport to protein bodies. A notable feature unique to incense cedar legumins is the considerable size of the variable region I caused by extended runs of glutamic acid residues (Fig. 5). This is also unusual in that the variable regions of angiosperm legumins tend to be rich in glutamine [27].

3.5. Evidence for glycosylation of incense cedar legumins

In contrast to the seed storage proteins of the vicilin family, legumins are generally considered as not glycosylated [5,7]. Interestingly, all of the incense cedar legumin precursors predicted here show a potential glycosylation signal Asn-Ile-Ser in the C-terminal part of the β -polypeptides (position 512–514, Fig. 5). To determine whether the corresponding polypeptides are glycosylated, we performed glycoprotein staining using concanavalin A as the first ligand. It specifically binds to glucosyl and mannosyl residues which preferentially constitute the core oligosaccharides of glycosylated seed storage proteins. As shown in Fig. 1 (lane g), no signal can be assigned to the α -polypeptides, whereas both types of β -polypep-

tides appear to be glycosylated. When the analysis is performed with proteins separated under non-reducing conditions (not shown), corresponding signals in the 23–25 kDa region cannot be detected, whereas monomeric and oligomeric legumin subunits are stained. This staining pattern is in agreement with the disulfide-bonded structure of legumin subunits and therefore also excludes its origin in co-migrating 7S polypeptides which are devoid of S-S linkages [5].

The peculiarity of legumins containing covalently bound carbohydrates has been reported from only very few species, e.g., *Magnolia salicifolia* [17]. With respect to gymnosperms, putative glycosylation signals are present in the polypeptide sequences deduced from legumin cDNAs of *Picea glauca* and *Pinus strobus* (for references, see Fig. 6); the Asn-X-Ser/Thr motifs are located in the C-terminal region of the corresponding β -polypeptides. However, legumins from species of *Picea* and *Pinus*, as well as those from *Pseudotsuga menziesii*, do not appear to be glycosylated [10,15,31].

The physiological significance of seed storage protein glycosylation is still unclear. It has been suggested that glycosylation near the COOH-terminus of vicilins may be involved in proteolytic processing or storage of these proteins; however, the inhibition of glycosylation by tunicamycin does not affect the synthesis, assembly or transport of these proteins [5].

3.6. Evolutionary relationships of incense cedar legumins

In order to study the evolutionary relationships of incense cedar legumins to those from other seed plants, the corresponding nucleotide sequences were aligned, and the conserved regions essentially as defined in [27] were used for distance analysis. Gymnosperm legumins hitherto known and legumins from representatives of monocotyledonous as well as from dicotyledonous plants were included in this phylogenetic analysis; the resulting dendrogram is shown in Fig. 6. When *Ginkgo* legumin GbLeg1 [14] is set as an outgroup sequence, the first node in the gene tree separates the remaining gymnosperm legumins from all angiosperm legumins shown. Within gymno-

sperms, an ancestral gene of present day legumin and *Pseudotsuga* and to the incense cedar legumins CdLeg3. The Pinaceae legumins and the Cupressaceae investigate separate branches.

On condition that evolutionary genes are approximately 10% and Cupressaceae, and the compared, the gene duplication subfamilies of incense cedar considered as a very ancient distance between both subfamilies as those observed different genera of the Pinaceae suggests that intergeneric divergence on the average, took place [32]. Our assumption of origin of the legumin subfamilies incense cedar implies that they be traceable within Cupressaceae related plant families. Indexed legumin cDNAs from *tostruboides* (unpublished) similarity with CdLeg1 and less related to CdLeg3.

Within angiosperms, the *Magnolia* B14 from all other and it has been assumed that been lost during early angiosperm. Obviously, the sister gene of common ancestor of *Magnolia* angiosperm legumins known

The dendrogram not only gene evolution but it also reflects the relationship of angiosperms appear monophyletic according to the (Pinaceae, Poaceae, Brassicaceae) family relationships are consistent with concepts of plant phylogeny. The group to Cupressaceae and monocots is near the root branch [33] and the close relationship and Magnoliidae has been [34]; the representatives of appear as monophyletic too.

lated. When the analysis is separated under non-shown), corresponding Da region cannot be dimeric and oligomeric. This staining pattern sulfide-bonded structure therefore also excludes its polypeptides which are

ins containing covalent has been reported from g., *Magnolia salicifolia* inosperms, putative glycosylated in the polypeptide egumin cDNAs of *Picea* (for references, see Fig. motifs are located in the the corresponding β -egumins from species of those from *Pseudotsuga* ar to be glycosylated

ificance of seed storage till unclear. It has been tion near the COOH- involved in proteolytic these proteins; however, ylation by tunicamycin sis, assembly or transport

ns of incense cedar

volutionary relationships to those from other seed ng nucleotide sequences served regions essentially sed for distance analysis. hitherto known and natives of monocotyledonous plants were etic analysis; the resulting n Fig. 6. When *Ginkgo* is set as an outgroup in the gene tree separates perm legumins from all shown. Within gymno-

sperms, an ancestral gene gave rise to the progenitor of present day legumins from *Pinus*, *Picea* and *Pseudotsuga* and to the common ancestor of incense cedar legumins CdLeg1, CdLeg2 and CdLeg3. The Pinaceae legumins and those from the Cupressaceae investigated are clustered on two separate branches.

On condition that evolutionary rates of legumin genes are approximately homogenous in Pinaceae and Cupressaceae, and that orthologous genes are compared, the gene duplication giving rise to both subfamilies of incense cedar legumins must be considered as a very ancient event. The phylogenetic distance between both subfamilies is in the same range as those observed for legumins from different genera of the Pinaceae; the fossil record suggests that intergeneric divergence in the Pinaceae, on the average, took place 135 million years ago [32]. Our assumption of a comparably ancient origin of the legumin subfamilies present in incense cedar implies that these subfamilies should be traceable within Cupressaceae or even in closely related plant families. Indeed, we have characterized legumin cDNAs from *Metasequoia glyptostroboides* (unpublished results) which show high similarity with CdLeg1 and CdLeg2 but are clearly less related to CdLeg3.

Within angiosperms, the first node separates *Magnolia* B14 from all other angiosperm legumins and it has been assumed that B14 orthologs have been lost during early angiosperm evolution [17]. Obviously, the sister gene of *Magnolia* B14 was the common ancestor of *Magnolia* A2 and all other angiosperm legumins known.

The dendrogram not only represents legumin gene evolution but it also reflects seed plant diversification and relationships: gymnosperms and angiosperms appear monophyletic and the taxa are clustering according to their respective families (Pinaceae, Poaceae, Brassicaceae, Fabaceae). The family relationships are consistent with accepted concepts of plant phylogeny: Pinaceae are a sister group to Cupressaceae [25]; the origin of gymnosperms is near the root of the angiosperm branch [33] and the close relatedness of monocots and Magnoliidae has been previously assumed [34]; the representatives of 'higher dicots' [35] appear as monophyletic too, with reflection also of

the subclasses Dilleniidae (here, the two Brassicaceae species) and Rosidae (here, one Rosaceae and two Fabaceae members). Thus, legumin genes appear to be appropriate molecular markers to elucidate phylogenetic relationships within seed plants.

The occurrence of legumin gene families and the conservation of legumin subunit structures in the evolutionary distant lineages of gymnosperms and angiosperms suggest that the corresponding genes originated from a single ancestral gene which already existed prior to the divergence of gymnosperms and angiosperms. The early evolution of legumin genes must obviously be divorced from the origin of the seed, and it should be considered that related genes are present in non-seed plants.

Acknowledgements

We thank A. Täuber for skillful technical assistance. This study was supported by grant Ha 1925/1–2 from the Deutsche Forschungsgemeinschaft.

References

- [1] E. Derbyshire, D.J. Wright and D. Boulter, Legumin and vicilin, storage proteins of legume seeds. *Phytochemistry*, 15 (1976) 3–24.
- [2] J.A. Gatehouse, R.R.D. Croy and D. Boulter, The synthesis and structure of pea storage proteins. *Crit. Rev. Plant Sci.*, 1 (1984) 287–314.
- [3] R. Casey, C. Domoney and N. Ellis, Legume storage proteins and their genes. *Oxford Surv. Plant Mol. Cell Biol.*, 3 (1986) 1–95.
- [4] R. Casey and C. Domoney, The structure of plant storage protein genes. *Plant Mol. Biol. Rep.*, 5 (1987) 261–281.
- [5] M.A. Shotwell and B.A. Larkins, The biochemistry and molecular biology of seed storage proteins, in: A. Marcus (Ed.), *The Biochemistry of Plants, A Comprehensive Treatise*, Vol. 15, Academic Press, New York, 1989, pp. 297–345.
- [6] D.S. Luthe, Electrophoretic analysis of seed proteins in the dicotyledoneae. *Plant Mol. Biol. Rep.*, 10 (1992) 254–262.
- [7] K. Borroto and L. Dure, The globulin seed storage proteins of flowering plants are derived from two ancestral genes. *Plant Mol. Biol.*, 8 (1987) 113–131.
- [8] U. Jensen and H. Berthold, Legumin-like proteins in gymnosperms. *Phytochemistry*, 28 (1989) 1389–1394.
- [9] K.-P. Häger, U. Jensen, J. Gilroy and M. Richardson,

- The N-terminal amino acid sequence of the β -subunit of the legumin-like protein from seeds of *Ginkgo biloba*. *Phytochemistry*, 31 (1992) 523–525.
- [10] D.J. Gifford, An electrophoretic analysis of the seed proteins from *Pinus monticola* and eight other species of pine. *Can. J. Bot.*, 66 (1988) 1808–1812.
- [11] U. Jensen and C. Lixue, *Abies* seed protein profile divergent from other Pinaceae. *Taxon*, 40 (1991) 435–440.
- [12] I. Allona, R. Casado and C. Aragoncillo, Seed storage proteins from *Pinus pinaster* Ait.: homology of major components with 11S proteins from angiosperms. *Plant Sci.*, 87 (1992) 9–18.
- [13] M. Arahira and C. Fukazawa, *Ginkgo* 11S seed storage protein family mRNA: unusual Asn-Asn linkage as post-translational cleavage site. *Plant Mol. Biol.*, 25 (1994) 597–605.
- [14] K.-P. Häger, H. Braun, A. Czihal, B. Müller and H. Bäumlein, Evolution of seed storage protein genes: Legumin genes of *Ginkgo biloba*. *J. Mol. Evol.*, 41 (1995) 457–466.
- [15] I. Leal and S. Misra, Molecular cloning and characterization of legumin-like storage protein cDNA of Douglas fir seeds. *Plant Mol. Biol.*, 21 (1993) 709–715.
- [16] C.N. Miller, Mesozoic conifers. *Bot. Rev.*, 43 (1977) 217–280.
- [17] H. Fischer, V. Haake, C. Horstmann and U. Jensen, Characterization and evolutionary relationships of Magnolia legumin-encoding cDNAs representing two divergent gene subfamilies. *Eur. J. Biochem.*, 229 (1995) 645–650.
- [18] R. Cerff and K. Klopstech, Structure diversity and differential light control of mRNAs coding for angiosperm glyceraldehyde-3-phosphate dehydrogenases. *Proc. Natl. Acad. Sci. USA*, 79 (1982) 7624–7628.
- [19] N. Murray, Phage lambda and molecular cloning, in: R.W. Hendrix, J.W. Roberts, F.W. Stahl and R.A. Weisberg (Eds.), *Lambda II*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor NY, 1983, pp. 395–432.
- [20] J. Sambrook, E.F. Fritsch and T. Maniatis, *Molecular cloning: A laboratory manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor NY, 1989.
- [21] W.-M. Zhao, J.A. Gatehouse and D. Boulter, The purification and partial amino acid sequence of a polypeptide from the glutelin fraction of rice grains: homology to pea legumin. *FEBS Lett.*, 162 (1983) 96–102.
- [22] C. Collada, R.G. Caballero, R. Casado and C. Aragoncillo, Seed storage proteins in Fagaceae: similarity between Castanea globulins and Quercus glutelins. *Plant Sci.*, 75 (1991) 145–154.
- [23] L. Lalonde, D.W. Fountain, A. Kermode, F.B. Ouellette, K. Scott, J.D. Bewley and D.J. Gifford, A comparative study of the insoluble storage proteins and the lectins of seeds of the Euphorbiaceae. *Can. J. Bot.*, 62 (1984) 1671–1677.
- [24] D.J. Gifford and J.D. Bewley, Interactions between subunit polypeptides of the crystalloid protein complex of castor bean endosperm. *Plant Sci. Lett.*, 36 (1984) 37–42.
- [25] J.A. Hart, A cladistic analysis of conifers: preliminary results. *J. Arnold Arbor.*, 68 (1987) 269–307.
- [26] H. Li, A reclassification of *Libocedrus* and Cupressaceae. *J. Arnold Arbor.*, 34 (1953) 17–35.
- [27] N.C. Nielsen, C.D. Dickinson, T.-J. Cho, V.H. Thanh, B.J. Scallan, R.L. Fischer, T.L. Sims, G.N. Drews and R.B. Goldberg, Characterization of the glycinin gene family in soybean. *Plant Cell*, 1 (1989) 313–328.
- [28] G. van Heijne, Patterns of amino acids near signal sequence cleavage sites. *Eur. J. Biochem.*, 133 (1983) 17–21.
- [29] J.F. March, D.J.C. Pappin and R. Casey, Isolation and characterization of a minor legumin and its constituent polypeptides from *Pisum sativum*. *Biochem. J.*, 250 (1988) 911–915.
- [30] N.C. Nielsen, R. Jung, Y.-W. Nam, T.W. Beaman, L.O. Oliveira and R. Bassüner, Synthesis and assembly of 11S globulins. *J. Plant Physiol.*, 145 (1995) 641–647.
- [31] S. Misra and M.J. Green, Developmental gene expression in conifer embryogenesis and germination. I. Seed proteins and protein body composition of mature embryo and the megagametophyte of white spruce (*Picea glauca* (Moench) Voss.). *Plant Sci.*, 68 (1990) 163–173.
- [32] R. Florin, The distribution of conifer and taxad genera in time and space. *Acta Horti Bergiani*, 20 (1963) 121–312.
- [33] A. Cronquist, *An integrated system of classification of flowering plants*. Columbia University Press, New York, 1981.
- [34] K. Kubitzki, Probleme der Grosssystematik der Blütenpflanzen. *Ber. Dtsch. Bot. Ges.*, 85 (1972) 259–277.
- [35] M.W. Chase et al., Phylogenetics of seed plants: An analysis of nucleotide sequences from the plastid gene *rbcL*. *Ann. Missouri Bot. Gard.*, 80 (1993) 529–563.



The effect of transfer

Frans A. Krens*

*DLO-Centre for Plant Breeding

Received

Abstract

Cellular competence for regeneration (*Beta vulgaris* L.). Explants for cotyledonary petioles. De novo cotyledonary node explants. Regeneration. The mode of regeneration period on 2,4-D, gave significant mimicked the effect of 2,4-D or treatment with NAA. Using the protocol mg/l kanamycin at a frequency

Keywords: Sugarbeet explants; regeneration

1. Introduction

Sugarbeet (*Beta vulgaris* L.) is an important arable crop. The selection of existing varieties is of great importance. In the last decade the input of molecular biology has drastically increased. However, the introduction of isolated genes requires a procedure for their introduction

* Corresponding author.

