

Evolutionary History of the Non-Specific Lipid Transfer Proteins

Monika M. Edstam^a, Lenita Viitanen^b, Tiina A. Salminen^b and Johan Edqvist^{a,1}

^a IFM Molecular Genetics, Linköping University, 581 83 Linköping, Sweden

^b Structural Bioinformatics Laboratory, Department of Biosciences, Åbo Akademi University, Artillerigatan 6 A III, FIN-20520 Turku, Finland

ABSTRACT The non-specific lipid transfer proteins (nsLTPs) are small, basic proteins characterized by a tunnel-like hydrophobic cavity, capable of transferring various lipid molecules between lipid bilayers. Most nsLTPs are synthesized with an N-terminal signal peptide that localizes the protein to the apoplastic space. The nsLTPs have only been identified in seed plants, where they are encoded by large gene families. We have initiated an analysis of the evolutionary history of the nsLTP family using genomic and EST information from non-seed land plants and green algae to determine: (1) when the nsLTP family arose, (2) how often new nsLTP subfamilies have been created, and (3) how subfamilies differ in their patterns of expansion and loss in different plant lineages. In this study, we searched sequence databases and found that genes and transcripts encoding nsLTPs are abundant in liverworts, mosses, and all other investigated land plants, but not present in any algae. The tertiary structures of representative liverwort and moss nsLTPs were further studied with homology modeling. The results indicate that the nsLTP family has evolved after plants conquered land. Only two of the four major subfamilies of nsLTPs found in flowering plants are present in mosses and liverworts. The additional subfamilies have arisen later, during land plant evolution. In this report, we also introduce a modified nsLTP classification system.

Key words: Lipids; evolutionary genetics; molecular biology; molecular evolution; bryophytes; ferns.

INTRODUCTION

The non-specific lipid transfer proteins (nsLTPs) are a part of the prolamin superfamily—a plant-specific protein superfamily characterized by conserved cysteine residues, low molecular mass, and a high content of α -helices (Kreis et al., 1985; Shewry et al., 2002). The prolamin superfamily also includes the α -amylase inhibitors, the prolamins, the 2S albumins, the indolines, and the hybrid proline-rich proteins (Radauer and Breiteneder, 2007). The nsLTPs possess four or five α -helices, which are stabilized by four conserved disulfide bridges formed by an eight-Cys motif (8CM) with the general form C-X_n-C-X_n-CC-X_n-CXC-X_n-C-X_n-C. The nsLTPs are further characterized by the presence of a tunnel-like hydrophobic cavity, capable of transferring various lipid molecules between lipid bilayers *in vitro* (Kader et al., 1984; Shin et al., 1995; Lee et al., 1998; Lascombe et al., 2008).

Almost all nsLTPs carry an N-terminal signal peptide directing the protein to the apoplastic space. Many nsLTPs also carry a sequence motif for the post-translational addition of a glycosylphosphatidylinositol (GPI)-anchor (DeBono et al., 2009; Lee et al., 2009). The nsLTPs have been divided into two families, based on the length of the polypeptide of the mature protein: LTP1s have approximately 90 amino acids, while LTP2s are about 70 amino acids long (Douliez et al., 2000). The two

families have a different pairing of their disulfide bridges: in the LTP1s, the Cys residues 1–6, 2–3, 4–7, and 5–8 are paired (Pasquato et al., 2006), while, in the LTP2s, the fifth and the sixth Cys residues have switched partners and, thus, Cys 1–5, 2–3, 4–7, and 6–8 are paired (Hoh et al., 2005).

The nsLTPs are only found in plants where they are encoded by large gene families with more than 40 members in the genomes of many flowering plants (Boutrot et al., 2008). nsLTPs have been thoroughly studied in flowering plants like rice (*Oryza sativa*), *Arabidopsis thaliana*, wheat (*Triticum aestivum*), barley (*Hordeum vulgare*), and maize (*Zea mays*) (Molina and Garcia-Olmedo, 1997; Boutrot et al., 2008; Li et al., 2008). Their exact biological role *in vivo* remains unclear, but several potential functions have been proposed. DIR1 in *A. thaliana* is an nsLTP with a role in long-distance signaling (Maldonado et al., 2002; Lascombe et al., 2008) that possibly is involved in plant defense against pathogens. A role for

¹ To whom correspondence should be addressed. E-mail johan.edqvist@liu.se, fax +46-13-281399, tel. +46-13-281288.

© The Author 2011. Published by the Molecular Plant Shanghai Editorial Office in association with Oxford University Press on behalf of CSPB and IPPE, SIBS, CAS.

doi: 10.1093/mp/ssr019, Advance Access publication 12 April 2011

Received 2 February 2011; accepted 14 February 2011

nsLTPs in pathogen defense is also suggested from experiments showing that heterologously expressed and purified nsLTPs in some cases are toxic, *in vitro*, to a wide range of fungal plant pathogens (Nielsen et al., 1996; Kirubakaran et al., 2008). The *Medicago truncatula* gene **N5** encodes an nsLTP, which is required for the symbiotic association between *M. truncatula* and the nitrogen-fixing rhizobial bacteria *Sinorhizobium meliloti* (Pii et al., 2009).

There are also several indications that some nsLTPs are involved in the formation of protective hydrophobic layers on the surfaces of plant aerial organs (Sterk et al., 1991; Thoma et al., 1994; Cameron et al., 2006; DeBono et al., 2009; Lee et al., 2009; Panikashvili et al., 2010) and also in pollen wall development (Zhang et al., 2008). Other nsLTPs are highly expressed in the endosperm of germinating seeds and may have a role in the recycling of endosperm lipids or as protease inhibitors protecting the cotyledons from proteases released from the endosperm during programmed cell death (Edqvist and Farbos, 2002; Eklund and Edqvist, 2003). One nsLTP is reported to enhance cell wall extension in tobacco (Nieuwland et al., 2005). Furthermore, several members of the nsLTP family have been identified as major allergens in foods and pollens (Sánchez-Monge et al., 1999; Salcedo et al., 2004; Lauer et al., 2009; Zoccatelli et al., 2010).

Clearly, it is rather difficult to get a comprehensive understanding of the biological function of nsLTPs, with reports pointing in many various directions. In this study, we attempt to provide clues to the function of the nsLTPs by investigating the evolutionary history of these enigmatic plant proteins. We show that the nsLTP genes are abundant in all investigated land plants, but absent from chlorophyte and charophyte green algae. Seemingly, the nsLTP genes evolved early after the establishment of land plants. The non-seed plants have a more limited set of nsLTP subfamilies or types compared to seed plants, which indicates that novel nsLTP types have evolved during land plant evolution. To accommodate the nsLTP subfamilies from early diverging land plants in an nsLTP classification system, we introduce an expanded categorization system, which considers intron positions, presence of GPI modification sites, Cys spacing in the 8CM, and sequence similarity. Furthermore, we have studied the three-dimensional structure of representative sequences from the two earliest nsLTP subtypes, using homology modeling.

RESULTS

Classification of nsLTP Genes

Genome sequences and ESTs from green and red algae, the liverwort *Marchantia polymorpha*, the moss *Physcomitrella patens*, the lycopod *Selaginella moellendorffii*, the fern *Adiantum capillus-veneris*, and the conifer *Pinus taeda* were analyzed for nsLTP genes. A total of 14 unique nsLTPs were found in *M. polymorpha*, 40 in *P. patens*, 43 in *S. moellendorffii*, six in *A. capillus-veneris*, and 43 in *P. taeda*.

InterProScan searches (www.ebi.ac.uk/Tools/InterProScan/) using plant nsLTP as bait also identified a few proteins from fungi and metazoans that possibly contained the C-X_n-C-X_n-CC-X_n-CXC-X_n-C-X_n-C motif. Two of these proteins with accession numbers AAQ54611 and ACG76228 are from insects. Both show very high similarity, 100 and 89%, respectively, to plant nsLTP amino acid sequences available in the database at NCBI. Since no other nsLTP genes have been identified in insect genomes, we believe that these genes or proteins are plant contaminations of projects on insect proteomics or insect gene cloning. We also obtained hits to some apparently true non-plant proteins such as the major surface glycoprotein Rucl4 (accession number ABQ50993) from the fungi *Pneumocystis jirovecii*, the laminin-related Imb-family member (lam-1) (accession number NP_500734) from *Caenorhabditis elegans*, and the F-box/LRR-repeat protein 18 (accession number NP_001025416) from *Danio rerio*. However, a closer examination of these sequences showed that, although they all are rich in Cys residues, none contains the 8CM motif C-X_n-C-X_n-CC-X_n-CXC-X_n-C-X_n-C found in plant nsLTPs. Thus, our investigation does not indicate that the 8CM motif is present in proteins from species outside the plant kingdom.

We systematically classified the identified plant nsLTP genes in order to analyze their evolutionary history. Previous nsLTP classification systems have been developed for genes from flowering plants and are, in general, based on sequence similarity and the putative molecular mass of the mature protein. For instance, the nsLTPs from rice, wheat, and *A. thaliana* were on the basis of sequence similarity recently classified in nine different types (Boutrot et al., 2008). Many of the nsLTP genes we have identified in *M. polymorpha*, *P. patens*, and *S. moellendorffii* show outside the 8CM rather low sequence similarity to the genes in flowering plants. Therefore, we found that sequence similarity was not a sufficient tool for a classification system that should also include genes from non-flowering plants. In this study, the two most established subfamilies, Types 1 and 2, are retained, while the other nsLTP genes are classified in the subfamilies Types C–K (Table 1). Some of these identified subfamilies, such as D and G, are expressed in liverworts, mosses, and vascular plants, whereas other types may in this investigation be restricted to a single species (Figure 1).

The classification presented in this work was initiated by first sorting the identified nsLTP sequences into Type G, based on the presence of a GPI modification site. In the second round of classification, we grouped the remaining sequences on the basis of the presence and position of introns. This sorting step was based on the observation that Type 1 may be identified by a conserved intron at nucleotide 5 after the last Cys in the 8CM and that Type 2 lack introns (Figure 2; Boutrot et al., 2008). The intron-based sorting identified the abundant Type D, as well as Type C (Figure 2). In the next step of the sorting procedure, hitherto unclassified sequences were sorted based on amino acid identity above 30% to nsLTP sequences already classified in the two previous steps. In this third step of the sorting procedure, we mainly identified Type D sequences lacking introns

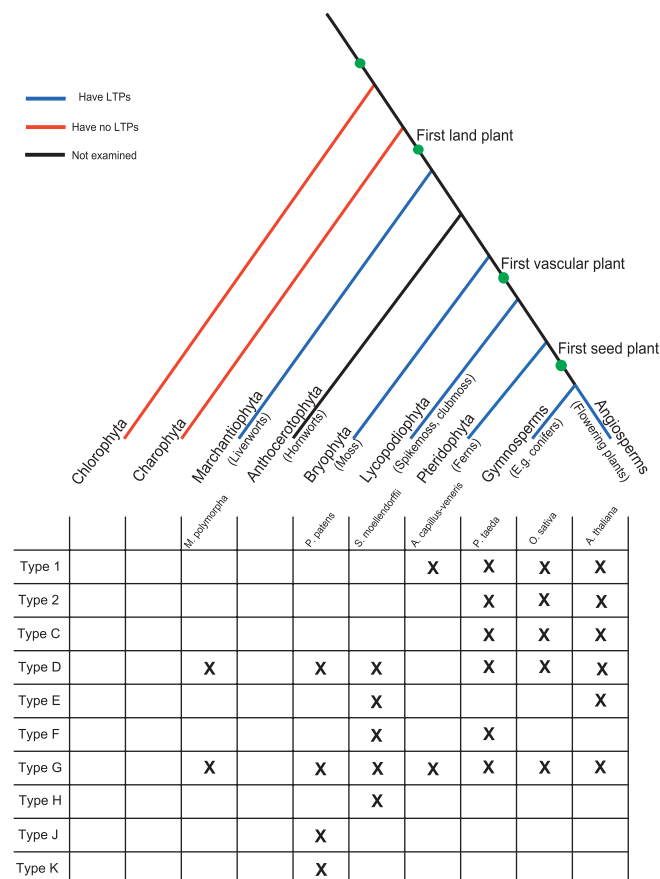
Table 1. Some characteristics for the different types of non-specific lipid transfer proteins found in *Marchantia polymorpha*, *Physcomitrella patens*, *Selaginella moellendorffii*, *Adiantum capillus-veneris*, *Pinus taeda*, *Oryza sativa* and *Arabidopsis thaliana*.

Type	Intron position ^a	GPI-anchored	Spacing pattern										
1	5	No	C	8, 9	C	13-17	CC	18-20 ^b	C-1-C	19-24	C	7-14	C
2	none	No	C	7 ^b	C	13, 15	CC	8-10	C-1-C	16, 21, 23	C	5-6 ^b	C
C	1	No	C	9	C	14, 16, 19	CC	9	C-1-C	12 ^b	C	6	C
D	4	No	C	6-14	C	8-18	CC	9-17	C-1-C	19-28	C	6-14	C
E	none	No	C	11, 13	C	15-16	CC	9	C-1-C	21-26	C	6-7	C
F	10 ^c	No	C	7	C	11 ^b	CC	10	C-1-C	19	C	9	C
G	various	Yes	C	6-11	C	14-19	CC	12-13	C-1-C	20-34	C	5-12	C
H	none	No	C	8, 11	C	13, 16	CC	9	C-1-C	21	C	9	C
J	25-58	No	C	11, 14	C	14-15	CC	12	C-1-C	20-22	C	6	C
K	various	No	C	11	C	14	CC	13	C-1-C	24	C	7	C

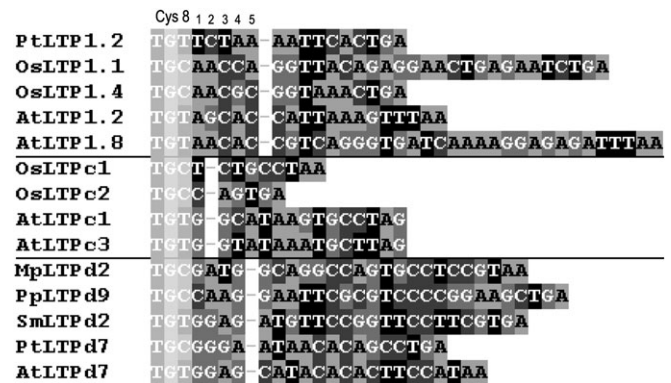
a Position of the intron is given as the number of bases from the codon encoding the eighth cysteine in the 8CM.

b Grey shaded numbers indicate spacings specific for a type. Multiple shaded numbers in same type must be combined to achieve specificity.

c This is based on the intron position in one of three sequences. The other two sequences lack either EST or genomic DNA sequence.

**Figure 1.** Occurrence and Distribution of Non-Specific Lipid Transfer Proteins (nsLTPs) in Different Plant Species.

Red branches in the cladogram indicate that no nsLTPs were found, while blue branches indicate presence of nsLTPs. Green dots mark important events in plant evolution. The table shows in which species each type of nsLTPs can be found.

**Figure 2.** The Position of the Intron in the DNA Sequence Differs between the nsLTP Types.

It is always located behind the last Cys in the 8CM, but the exact position varies. In Type 1, it is always behind the fifth nucleotide, in Type C behind the first, and in Type D behind the fourth.

and Type G sequences lacking an evident GPI modification site. In the fourth sorting step, the remaining unclassified nsLTP sequences showing at least 30% identity were grouped together. This step separated the abundant Type 2 from the minor Types E, F, H, J, and K. Finally, in an attempt to establish a specific 8CM consensus for each obtained nsLTP type, we investigated the spacing of the Cys residues in all of the classified nsLTPs (Table 1).

The obtained groups were tested in phylogenetic analyses with trees constructed with both maximum likelihood and neighbor-joining methods (Figure 3). Types 1 and 2 sequences form specific clades in the trees, indicating that the genes in both these major nsLTP types share a common ancestor. The sequences from the minor types E, H, and J also form separated clades in both trees. Type D and G sequences locate to specific

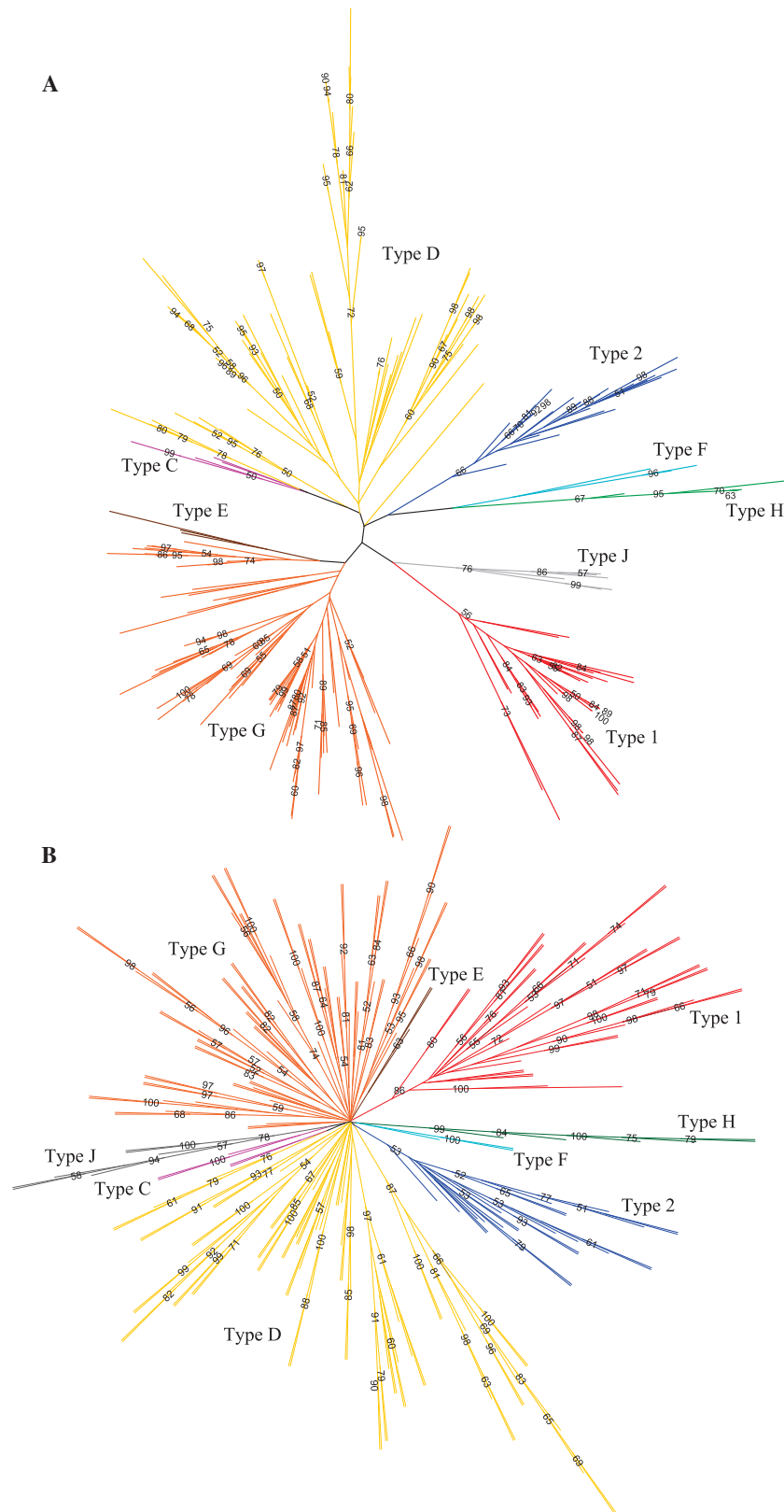


Figure 3. Unrooted Phylogenetic Trees over Different Subfamilies of nsLTP Sequences from the Studied Species, plus *A. thaliana* and Rice.
(A) The maximum likelihood phylogenetic tree.
(B) The tree based on the neighbor-joining method.

clusters. However, none of these families forms distinct monophyletic clades supported with high boot-strap values in any of the phylogenetic trees.

nsLTPs Were Not Identified in Algae

The four green algae that belong to Chlorophyta (*Volvox carteri*, *Chlamydomonas reinhardtii*, *Ostreococcus lucimarinus*, and *Ostreococcus tauri*) and the red algae (*Cyanidioschyzon merolae*) have completely sequenced genomes; still, no genes coding for nsLTPs or nsLTP-like proteins were found. Anyhow, InterProScan sequence searches using plant nsLTPs as baits revealed that the 8CM could be detected in some proteins from *O. lucimarinus*, *O. tauri*, and *Micromonas* spp. (accession numbers A456G3, C1E7Q8, C1E8G6, XP_002505095, C1MZS3, XP_003059902, Q00X46, and Q019T7). These proteins have molecular sizes between 50 and 340 kDa and thus have much larger molecular masses compared to the plant nsLTPs. In addition, our investigations showed that five of them have transmembrane regions, while the remaining three are not secretory proteins. When the algal proteins were used for BLAST searches against land plant proteins no nsLTPs were found. We therefore did not classify these algal proteins as plant nsLTPs.

For the green alga from Charophyta (*Mesostigma viride*) and the red alga *Porphyra yezoensis*, EST databases were available and searched, but no nsLTP genes were identified among the numerous ESTs. The results indicate that nsLTPs are not encoded in algal genomes; rather, the nsLTP genes have been acquired by land plants. It should be pointed out, though, that until we have searched fully sequenced genomes, we cannot completely rule out that some nsLTP genes may exist in charophytes.

nsLTPs in *Marchantia polymorpha*

M. polymorpha belongs to the liverworts, which, according to morphology and sequence data, resemble the first plants that conquered land. The genome for *M. polymorpha* is not sequenced yet; hence, only EST databases were searched in this study. Among these 33 692 ESTs, 14 unique sequences encoding nsLTPs were found. Eight of these sequences belong to Type D (Table 2). Amplification and sequencing of six of the nsLTP Type D genes revealed that five genes carry introns whereas one gene is without intron. The introns in two of the nsLTP genes are found four bases downstream of the 8CM. The exact intron position has not been determined in the remaining sequences.

Four Type G sequences were found in *M. polymorpha*. MplPTg1 does not have a GPI-anchor site and is lacking the eighth Cys as well. It is still regarded as a Type G sequence due to 37% amino acid identity to another Type G sequence. An intron is found in one of the Type G sequences, but the position has not been determined (Table 2). The presence of introns in the other three sequences is unknown. The two remaining nsLTP sequences do not fit into any of the proposed types. Our results show that nsLTPs were present and abundant even in the earliest land plant ancestors. Types D and G were possibly the first that evolved in land plants, while Types 1 and 2 that seem to be the dominant, or most abundant, nsLTP types in flowering plants are not found in *M. polymorpha*.

nsLTPs in *Physcomitrella patens*

The moss *P. patens* is the earliest diverging land plant with a completely sequenced genome (Rensing et al., 2008). A genome-wide search revealed that a total of 40 nsLTPs

In both trees, bootstrap values above 50 are presented. Complete trees with designations of the included sequences are available at TreeBASE (www.treebase.org/; submission id: 11145; study access <http://purl.org/phylo/treebase/phylows/study/TB2:S11155>). Each nsLTP type is shown in a specific color: Type 1 in red, Type 2 in blue, Type C in pink, Type D in yellow, Type E in brown, Type F in cyan, Type G in orange, Type H in green, and Type J in gray. Type 1 sequences included in the trees: AcvLTP1.1, AcvLTP1.2, AcvLTP1.3, PtLTP1.1, PtLTP1.2, PtLTP1.3, PtLTP1.4, PtLTP1.5, PtLTP1.6, PtLTP1.7, PtLTP1.8, PtLTP1.9, OsLTP1.1, OsLTP1.2, OsLTP1.3, OsLTP1.4, OsLTP1.6, OsLTP1.7, OsLTP1.8, OsLTP1.9, OsLTP1.10, OsLTP1.12, OsLTP1.13, OsLTP1.14, OsLTP1.15, OsLTP1.16, OsLTP1.17, OsLTP1.18, OsLTP1.19, OsLTP1.20, AtLTP1.1, AtLTP1.3, AtLTP1.4, AtLTP1.5, AtLTP1.6, AtLTP1.7, AtLTP1.8, AtLTP1.9, AtLTP1.10, AtLTP1.11, AtLTP1.12. Type 2 sequences included in the trees: PtLTP2.1, OsLTP2.1, OsLTP2.2, OsLTP2.3, OsLTP2.4, OsLTP2.5, OsLTP2.6, OsLTP2.7, OsLTP2.8, OsLTP2.9, OsLTP2.10, OsLTP2.11, OsLTP2.12, OsLTP2.13, AtLTP2.1, AtLTP2.2, AtLTP2.3, AtLTP2.4, AtLTP2.5, AtLTP2.6, AtLTP2.7, AtLTP2.9, AtLTP2.10, AtLTP2.11, AtLTP2.12, AtLTP2.13, AtLTP2.14. Type C sequences included in the trees: PtLTPc1, OsLTPc1, OsLTPc2, AtLTPc1, AtLTPc3. Type D sequences included in the trees: MplTPd1, MplTPd2, MplTPd3, MplTPd4, MplTPd5, MplTPd6, MplTPd7, PpLTPd1, PpLTPd2, PpLTPd3, PpLTPd4, PpLTPd5, PpLTPd6, PpLTPd7, PpLTPd8, PpLTPd9, PpLTPd10, PpLTPd11, PpLTPd12, PpLTPd13, PpLTPd14, PpLTPd15, PpLTPd16, PpLTPd17, PpLTPd18, PpLTPd19, PpLTPd20, SmlTPd1, SmlTPd2, SmlTPd3, SmlTPd4, SmlTPd5, SmlTPd6, SmlTPd7, SmlTPd8, SmlTPd9, SmlTPd11, SmlTPd12, SmlTPd13, SmlTPd14, SmlTPd15, SmlTPd16, SmlTPd17, SmlTPd18, SmlTPd19, PtLTPd1, PtLTPd2, PtLTPd3, PtLTPd4, PtLTPd5, PtLTPd6, PtLTPd7, PtLTPd8, PtLTPd9, PtLTPd10, PtLTPd11, PtLTPd12, OsLTPd1, OsLTPd2, OsLTPd3, OsLTPd4, OsLTPd5, OsLTPd6, OsLTPd7, OsLTPd8, OsLTPd9, OsLTPd10, OsLTPd12, AtLTPd1 (DIR1), AtLTPd2, AtLTPd3, AtLTPd4, AtLTPd5, AtLTPd6, AtLTPd7, AtLTPd8, AtLTPd9, AtLTPd10, AtLTPd11, AtLTPd12. Type E sequences included in the trees: SmlLTPe1, SmlLTPe2, SmlLTPe3. Type F sequences included in the trees: SmlLTPf1, SmlLTPf2, PtLTPf1. Type G sequences included in the trees: MplTPg1, MplTPg2, MplTPg3, PpLTPg1, PpLTPg2, PpLTPg3, PpLTPg4, PpLTPg5, PpLTPg6, PpLTPg7, PpLTPg8, PpLTPg9, PpLTPg10, SmlTPg1, SmlTPg2, SmlTPg3, SmlTPg4, SmlTPg5, SmlTPg6, SmlTPg7, SmlTPg8, SmlTPg9, SmlTPg11, SmlTPg12, AcvLTPg1, AcvLTPg2, PtLTPg1, PtLTPg2, PtLTPg3, PtLTPg4, PtLTPg5, PtLTPg6, PtLTPg7, PtLTPg8, PtLTPg9, PtLTPg10, PtLTPg11, PtLTPg12, PtLTPg13, PtLTPg14, PtLTPg15, PtLTPg16, PtLTPg17, OsLTPg1, OsLTPg2, OsLTPg3, OsLTPg4, OsLTPg5, OsLTPg6, OsLTPg7, OsLTPg8, OsLTPg9, OsLTPg10, OsLTPg11, OsLTPg12, OsLTPg13, OsLTPg14, OsLTPg15, OsLTPg16, OsLTPg17, OsLTPg18, OsLTPg19, OsLTPg20, OsLTPg21, OsLTPg24, OsLTPg25, AtLTPg1, AtLTPg2, AtLTPg3, AtLTPg5, AtLTPg6, AtLTPg7, AtLTPg8, AtLTPg9, AtLTPg10, AtLTPg11, AtLTPg12, AtLTPg13, AtLTPg14, AtLTPg15, AtLTPg16, AtLTPg19, AtLTPg20, AtLTPg21, AtLTPg22, AtLTPg23, AtLTPg26, AtLTPg29, AtLTPg30, AtLTPg31, AtLTPg32. Type H sequences included in the trees: SmlLTPh1, SmlLTPh2, SmlLTPh3, SmlLTPh4, SmlLTPh5, SmlLTPh6. Type J sequences included in the trees: PpLTPj1, PpLTPj2, PpLTPj3, PpLTPj4, PpLTPj5, PpLTPj6, PpLTPj7. Type K is not included.

Table 2. The occurrence of non-specific lipid transfer proteins in *Marchantia polymorpha*, and some of their features.

Name ^a	Signal peptide		Mature protein			Intron Amount / position ^c	Amino acid identity ^d
	Amino acids	Target ^b	Amino acids	Mass (Da)	pI		
Type D							
MpLTPd1	23	s	179	19229.06	4.96	1 ^e	41% (SmLTPd2)
MpLTPd2	22	s	83	8963.22	8.35	1 (4)	
MpLTPd3	22	s	82	8829.13	10.07	n/a ^f	36% (MpLTPd2)
MpLTPd4	23	s	80	8412.93	11.26	n/a ^f	39% (MpLTPd2)
MpLTPd5	25	s	88	9200.14	10.44	1 ^e	38% (PpLTPd15)
MpLTPd6	20	s	101	10430.82	4.73	1 ^e	48% (PpLTPd15)
MpLTPd7	24	s	77	8053.33	5.00	0	34% (MpLTPd6) ^g
MpLTPd8	26	s	100	10852.21	3.78	1 (4)	
Type G							
MpLTPg1	23	s	88	9065.49	5.98	n/a ^f	37% (SmLTPg4)
MpLTPg2	24	s	259	25641.25	4.05	n/a ^f	
MpLTPg3	23	s	141	13693.65	3.96	1 ^e	
MpLTPg4	25	s	143	14973.05	4.38	n/a ^f	
Singles							
MpLTPx1	24	s	86	9688.20	9.52	n/a ^f	
MpLTPx2	26	s	80	8554.69	4.23	0	

^a For accession numbers of corresponding ESTs, see Supporting Information Table S2

^b Subcellular target for the protein. S = secretory pathway, m = mitochondria

^c Position of the intron is given as the number of bases from the codon encoding the eighth cysteine in the 8CM

^d The highest amino acid identity the sequence shows to a protein with specific characteristics for the type.

^e The position of the intron has not yet been determined

^f The presence or absence of an intron has not yet been determined

^g The amino acid identity to proteins with specific characteristics is below 30%. Instead, identity to a protein already included through the identity criteria is used.

are found in *P. patens* (Table 3). The nsLTP content in *P. patens* resembles that in the liverwort *M. polymorpha*; Types D and G are the dominant types and constitute the main part of the nsLTPs in both species. Type D has 21 members, of which 16 have an intron four bases downstream of the 8CM. Thus, the position of this nsLTP intron is conserved in liverworts and mosses. The remaining five genes do not have any introns, but are classified as Type D due to amino acid identities between 36 and 58% to other Type D sequences (Table 3).

Ten sequences belong to Type G and all of them have a GPI-anchor site. Curiously, two of the Type G sequences in *P. patens* have two 8CMs and one has three 8CMs (Figure 4). Two introns each are found in three of the Type G sequences. In two of them, the first intron is located four nucleotides downstream of the last Cys codon, as in Type D. The second intron does not show any obvious positional conservation. The remaining seven Type G sequences are without introns. Two additional types (Types J and K) were identified in *P. patens*. These types were not found in other species, which suggests that they have appeared after the separation of mosses. Types 1, 2, C, E, F, and H are not encoded in *P. patens* and seemingly have evolved in vascular plants.

Only a few of the *P. patens* nsLTP genes are organized in tandem repeats. The Type D protein PpLTPd6 is encoded by two duplicated genes organized in a tandem repeat and

the same is found for PpLTPd9. Further, another gene is repeated four times in the genome, but the copies are not adjacent to each other. The reading frames encoded from two of these four repeats are identical and encode PpLTPd19, whereas the other two copies that encode PpLTPd18 and PpLTPd20 show, at the DNA level, 94% identity to each other, and 94 and 89% identity, respectively, to PpLTPd19. None of these repeated genes carries any intron.

nsLTPs in *Selaginella moellendorffii*

S. moellendorffii is a member of the phylum Lycopodiophyta and is one of the earliest diverging vascular land plants. In the completely sequenced *S. moellendorffii* genome, we identified 43 nsLTP genes, of which 19 belong to Type D (Table 4). A comparison between DNA and mRNA sequences revealed that 11 of the Type D genes have an intron located four nucleotides downstream of the last Cys codon. Two other genes (SmLTPd14 and SmLTPd17) have, at the same position, a sequence element with the conserved features of an intron. However, mRNA sequence data are lacking and the sequence elements have therefore not been experimentally confirmed as introns. Six Type D genes in *S. moellendorffii* are without introns. As described above, there are no ESTs that confirm the expression of *SmLTPd14* and *SmLTPd17*. This is also the case

Table 3. The occurrence of non-specific lipid transfer proteins in *Physcomitrella patens*, and some of their features.

Name ^a	Signal peptide		Mature protein			Intron Amount / position ^c	Amino acid identity ^d
	Amino acids	Target ^b	Amino acids	Mass (Da)	pI		
Type D							
PpLTPd1	26	s	86	9047.1	4.14	1 (4)	
PpLTPd2	-		98	10039.38	4.37	1 (4)	
PpLTPd3	28	s	81	8258.44	9.82	1 (4)	
PpLTPd4	25	s	88	8948.29	9.87	1 (4)	
PpLTPd5	23	s	87	9141.45	8.4	1 (4)	
PpLTPd6 ^f	-		100	10476.29	12.6	1 (4)	
PpLTPd7	24	s	87	8329.62	10.58	1 (4)	
PpLTPd8	25	s	90	9916.76	9.33	1 (4)	
PpLTPd9 ^f	25	s	87	8927.75	12.4	1 (4)	
PpLTPd10	27	s	85	8355.9	9.88	1 (4)	
PpLTPd11	27	s	79	7817.93	8.34	1 (4)	
PpLTPd12	28	s	82	8507.65	9.46	0	58% (PpLTPd11)
PpLTPd13	28	s	82	8890.11	8.54	1 (4)	
PpLTPd14	27	s	81	8276.41	9.52	1 (4)	
PpLTPd15	30	s	98	10102.66	9.3	1 (4)	
PpLTPd16 ^e	31	s	99	10873.64	8.3	1 (4)	
PpLTPd17	26	s	77	7887.26	8.38	0	40% (PpLTPd15)
PpLTPd18	32	s	80	8762.38	11.8	0	38% (SmLTPd2)
PpLTPd19	34	s	80	8746.32	11.92	0	36% (SmLTPd2)
PpLTPd20 ^e	32	s	80	8691.20	11.8	0	36% (SmLTPd2)
PpLTPd21 ^e	21	s	135	14497.69	5.79	2 (4)	
Type G							
PpLTPg1	27	s	273	27693.19	4.66	0	
PpLTPg2	21	s	307	30008.89	3.59	0	
PpLTPg3	22	s	139	14001.99	5.58	2 (4, 45)	
PpLTPg4 ^e	21	s	270	25480.69	3.93	0	
PpLTPg5	27	s	296	29564.36	4.25	0	
PpLTPg6	21	s	178	17274.89	8.15	0	
PpLTPg7 ^e	20	s	419	42551.16	4.46	0	
PpLTPg8	20	s	405	38831.32	4.28	0	
PpLTPg9	23	s	142	13953.87	5.52	2 (4, 73)	
PpLTPg10 ^e	24	s	237	25161.82	9.32	2 (117, 218)	
Type J							
PpLTPj1	28	s	134	14634.53	5.59	0	38% (PpLTPj2)
PpLTPj2	21	s	184	19726.69	5.03	1 (49)	50% (PpLTPj4)
PpLTPj3	22	s	165	18119.05	9.82	1 (25)	59% (PpLTPj5)
PpLTPj4	35	s	188	19693.6	9.51	1 (58)	73% (PpLTPj7)
PpLTPj5 ^e	19	s	241	25724.42	8.41	0	59% (PpLTPj3)
PpLTPj6	23	s	195	20529.75	9.1	1 (58)	69% (PpLTPj7)
PpLTPj7	35	s	188	19938.96	9.26	1 (58)	73% (PpLTPj4)
Type K							
PpLTPk1 ^e	26	s	271	28612.88	5.82	3 (118, 439, 535)	54% (PpLTPk2)
PpLTPk2 ^e	26	s	260	27098.97	4.92	1 (130)	54% (PpLTPk1)

^a For accession numbers of corresponding proteins, see Supporting Information Table S2^b Subcellular target for the protein. S = secretory pathway, m = mitochondria^c Position of the intron is given as the number of bases from the codon encoding the eighth cysteine in the 8CM^d The highest amino acid identity the sequence shows to a protein with specific characteristics for the type.^e The putative amino acid sequence is not, partly or completely, confirmed by ESTs^f The sequence is encoded by two identical genes

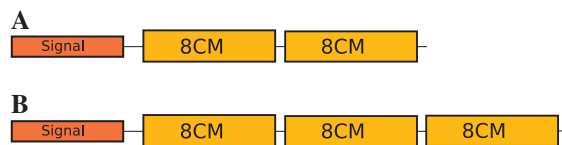


Figure 4. The Structure of Polydomain nsLTPs.

(A) One additional 8CM follows directly behind the first. This polydomain structure is found in MpLTPg2, PpLTPg1, PpLTPg5, and PpLTPj5.

(B) Two additional 8CMs follow directly behind the first. This polydomain structure is found in PpLTPg7.

for three of the genes that are without intron (*SmLTPd1*, *SmLTPd3*, and *SmLTPd4*). We can therefore not exclude that at least some of these sequences are pseudogenes.

Twelve of the *S. moellendorffii* nsLTP genes encode Type G proteins. Six of them (*SmLTPg1*, *SmLTPg4*, *SmLTPg8*, *SmLTPg9*, *SmLTPg11*, *SmLTPg12*) do not have a predicted GPI-anchor site, but are still considered as members of Type G due to above 30% amino acid identity to other Type G sequences (Table 4). Eight of the Type G genes contain one to three introns and six of these genes have an intron localized to the conserved position four bases downstream of the

Table 4. The occurrence of non-specific lipid transfer proteins in *Selaginella moellendorffii*, and some of their features.

Name ^a	Signal peptide		Mature protein			Intron Amount/Position ^c	Amino acid identity ^d
	Amino acids	Target ^b	Amino acids	Mass (Da)	pI		
Type D							
SmLTPd1 ^e	22	s	154	16105.77	9.69	0	37% (PpLTPd15)
SmLTPd2	22	s	81	8464.80	5.1	1 (4)	
SmLTPd3 ^e	24	s	80	8233.37	6.36	0	36% (SmLTPd12)
SmLTPd4 ^e	22	s	81	8271.42	6.38	0	40% (SmLTPd12)
SmLTPd5 ^e	25	s	102	10546.2	6.7	1 (4)	
SmLTPd6	25	s	86	9211.45	9.52	1 (4)	
SmLTPd7	21	s	80	8211.34	9.99	0	37% (SmLTPd12)
SmLTPd8	27	s	86	9062.86	10.3	1 (4)	
SmLTPd9	27	s	87	9001.29	9.87	1 (4)	
SmLTPd10	23	s	91	9585.78	5.47	0	47% (SmLTPd5)
SmLTPd11	22	s	107	11439.29	9.8	0	38% (SmLTPd8)
SmLTPd12	21	s	86	8963.33	10.52	1 (4)	
SmLTPd13	24	s	92	9477.02	9.46	1 (4)	
SmLTPd14 ^{e, f}	21	s	95	10397.15	9.78	0	49% (SmLTPd16)
SmLTPd15	30	s	88	9660.13	6.93	1 (4)	
SmLTPd16	28	s	85	9089.56	9.4	1 (4)	
SmLTPd17 ^e	30	s	84	9164.64	8.25	0	93% (SmLTPd15)
SmLTPd18	21	s	81	8284.49	10.95	1 (4)	
SmLTPd19 ^{e, f}	21	s	77	7842.03	11.84	1 (4)	
Type E							
SmLTPe1	26	s	82	8227.25	4.27	0	37% (SmLTPe2)
SmLTPe2	28	s	78	8136.12	4.52	0	33% (AtLTPe1)
SmLTPe3 ^e	24	s	82	8533.97	8.24	0	36% (SmLTPe2)
Type F							
SmLTPf1 ^e	35	m	84	8903.11	5.42	0	
SmLTPf2	22	s	108	11402.02	5.52	1 (10)	
Type G							
SmLTPg1 ^e	23	s	91	9466.17	10	0	43% (PtLTPg2)
SmLTPg2	22	s	115	11993.98	5.09	1 (4)	
SmLTPg3	22	s	137	13810.68	5.08	2 (4, 70)	
SmLTPg4 ^e	24	s	237	25731.29	5.99	2 (-6, 36)	39% (OsLTPg2)
SmLTPg5 ^e	20	s	162	16683.61	4.74	1 (85)	
SmLTPg6	19	s	136	14370.56	5.84	3 (4, 37, 58)	
SmLTPg7	23	s	153	14764.01	9.3	2 (4, 37)	

Table 4. Continued

Name ^a	Signal peptide		Mature protein			Intron Amount/Position ^c	Amino acid identity ^d
	Amino acids	Target ^b	Amino acids	Mass (Da)	pI		
SmLTPg8	27	s	129	14434.95	9.69	1 (4)	42% (SmLTPg2)
SmLTPg9	19	s	88	8952.16	6.26	0	38% (PtLTPg9)
SmLTPg10 ^e	23	s	132	13942.29	9.19	2 (4, 37)	
SmLTPg11 ^e	35	s	87	8790.08	9.53	0	44% (PtLTPg3)
SmLTPg12 ^e	21	s	132	14115.15	8.84	0	39% (PpLTPg7)
Type H							
SmLTPh1	30	s	82	8963.58	10.51	0	38% (SmLTPh5)
SmLTPh2 ^e	25	s	83	9069.43	5.42	0	92% (SmLTPh6)
SmLTPh3 ^e	25	s	81	9086.54	5.83	0	87% (SmLTPh6)
SmLTPh4	26	s	81	8509.98	9.7	0	39% (SmLTPh2)
SmLTPh5 ^e	25	s	83	9112.54	6.04	0	97% (SmLTPh6)
SmLTPh6 ^e	25	s	83	9147.59	6.04	0	97% (SmLTPh5)
Single							
SmLTPx1	25	s	77	7674.89	6.17	0	

a For accession numbers of corresponding proteins, see Supporting Information Table S2

b Subcellular target for the protein. S = secretory pathway, m = mitochondria

c Position of the intron is given as the number of bases from the codon encoding the eighth cysteine in the 8CM

d The highest amino acid identity the sequence shows to a protein with specific characteristics for the type

e The putative amino acid sequence is not, partly or completely, confirmed by ESTs

f The sequence is encoded by two identical genes

last Cys codon. Three of the five Type G genes, which contain more than one intron, have an additional intron inserted 37 nucleotides downstream of the last Cys codon. One of the Type G genes in *S. moellendorffii* has a predicted intron within the 8CM, but the expression of this gene has not been confirmed.

There are three genes encoding Type E and two genes encoding Type F nsLTPs in the *S. moellendorffii* genome. Interestingly, one of the *S. moellendorffii* Type F sequences is predicted to have a mitochondrial signal peptide with the highest possible score in TargetP. The expression of this gene has, however, not been confirmed by an EST. Type H, which is found only in *S. moellendorffii*, consists of six genes, all without introns. The four highly similar (87–97% identity) Type H genes that encode SmLTPh2, SmLTPh3, SmLTPh4, and SmLTPh5 are located in a repeat in the *S. moellendorffii* genome. None of these genes is known to be expressed. Otherwise, only a few nsLTP genes are organized in tandem repeats in *S. moellendorffii*. There are two tandem repeats with Type D genes. In one of these repeats, there are two identical genes encoding SmLTPd14. In the other Type D repeat, translation of the two tandem genes results in two different proteins: SmLTPd16 and SmLTPd17. Furthermore, in *S. moellendorffii*, there are two genes with identical open reading frames that encode SmLTPd19. These genes are not, however, organized in a tandem repeat but found on separate locations in the genome.

Types D and G are dominating the nsLTP content in *S. moellendorffii*, *P. patens*, and *M. polymorpha*. Similarly to *P. patens*, *S. moellendorffii* has some additional nsLTP genes, which could not be classified into the major types. These

genes rather belong to minor types, which, on the basis of the current information, appear to be species-specific, such as Type H, or are shared by only a few other species, such as Type E.

nsLTPs in *Adiantum capillus-veneris*

The fern *A. capillus-veneris* belongs to the family Pteridaceae and has diverged between *S. moellendorffii* and the seed plants. *A. capillus-veneris* does not have a sequenced genome, but 30 540 ESTs are available. Among these ESTs, only six unique nsLTPs were found (Table 5). Three unique ESTs encode Type 1 nsLTPs, which shows that Type 1 has evolved after the divergence of Lycopodiophyta and Euphyllophyta. Only one of the two Type G sequences has a predicted GPI-anchor modification site, but both are considered as members of Type G due to high sequence similarity to other Type G sequences (Table 5). The remaining one of the identified *A. capillus-veneris* nsLTPs does not match any of the proposed types.

This investigation has revealed that the Type D genes are the dominant, most highly expressed nsLTP type in the early diverging land plants. Still, no such genes were identified in *A. capillus-veneris*. It is important to point out, though, that we identified an unexpectedly low number of nsLTPs among the *A. capillus-veneris* ESTs and it seems most likely that there must be a significant number of nsLTP genes remaining to be identified. Therefore, we do not exclude that Type D, or any of the other missing nsLTP types, is encoded in *A. capillus-veneris*. Nevertheless, it seems that a shift from Type D to Type 1 as the dominant nsLTP occurred after the separation of Lycopodiophyta and Euphyllophyta.

Table 5. The occurrence of non-specific lipid transfer proteins in *Adiantum capillus-veneris*, and some of their features.

Name ^a	Signal peptide		Mature protein			Amino acid identity ^c
	Amino acids	Target ^b	Amino acids	Mass (Da)	pI	
Type 1						
AcvLTP1.1	28	s	91	9633.34	10.51	
AcvLTP1.2	32	s	108	11685.62	10.19	
AcvLTP1.3	26	s	102	10041.44	10.36	
Type G						
AcvLTPg1	33	s	117	11994.48	4.2	36% (PtLTPg9)
AcvLTPg2	29	s	174	17179.47	3.97	
Single						
AcvLTPx1	28	s	79	7949.01	10.00	

a For accession numbers of corresponding ESTs, see Supporting Information Table S2

b Subcellular target for the protein. S = secretory pathway, m = mitochondria

c The highest amino acid identity the sequence shows to a protein with specific characteristics for the type

nsLTPs in *Pinus taeda*

There is no completely sequenced gymnosperm genome, but there are 328 628 ESTs available from *P. taeda*. Among these ESTs, 43 unique nsLTP sequences were identified and all except one could be placed into Types 1, 2, or C–G (Table 6). The remaining sequence did not match the criteria for any of the specified nsLTP types. The nsLTP content in *P. taeda* generally resembles much of that in angiosperms. Types 2 and C are present and both types probably evolved after the separation of ferns and seed plants. Type E, previously found in *A. thaliana* and *S. moellendorffii*, is not found in *P. taeda*, while Type F is only found in *P. taeda* and *S. moellendorffii*. Four of the 17 Type G sequences found in *P. taeda* were predicted not to have a GPI-anchor site, but are considered as members of Type G due to high sequence similarity (50–64% identity) to other Type G sequences.

Among 5393 genomic sequences for *P. taeda*, five were found to encode nsLTPs. From these genomic sequences, it was possible to examine the intron positions for one Type 1 sequence, three Type D sequences, and one Type G sequence. The Type 1 sequence has an intron in position five, two of the Type D sequences and the Type G sequence have an intron in position four, and the third Type D sequence is without any intron. In addition to the nsLTPs, 14 other sequences with 8CM were found in *P. taeda*. Homology studies showed that they were hybrid proline-rich proteins and 2S albumin seed storage proteins.

Structural Modeling of nsLTPs from *Marchantia polymorpha* and *Physcomitrella patens*

In order to investigate the folding and the pairing of the cysteines in nsLTPs from early diverged land plants, we modeled the 3-D structure of Type D and G sequences from *M. polymorpha* and *P. patens*. MpLTPd5 and MpLTPg3 from *M. polymorpha* (Table 2) and PpLTPd15 and PpLTPg5 from *P. patens* (Table 3) were selected as representative sequences for structural modeling. The crystal structure of *A. thaliana*

DIR1 (pdb code 2RKN; Lascombe et al., 2008) was selected as a template for modeling based on a search against the PDB using the Basic Local Alignment Search Tool (BLAST) at NCBI (<http://blast.ncbi.nlm.nih.gov/>) with the target Type D and G proteins as bait. DIR1 has 77 amino acids, all of which are visible in the crystal structure (Figure 5A and 5B). Based on the DIR1 structure, the amino acids 2–82 of MpLTPd5, 1–80 of PpLTPd15, 1–84 of MpLTPg3, and 1–81 of PpLTPg5 could be modeled. The sequence identity between the *A. thaliana* DIR1 and the modeled proteins is 27% for MpLTPd5, 34% for PpLTPd15, 31% for MpLTPg3, and 24% for PpLTPg5.

The structural models of MpLTPd5, PpLTPd15, MpLTPg3, and PpLTPg5 all have the typical nsLTP fold: five α -helices stabilized by the four disulfide bridges of the conserved 8CM (Figure 5C and 5D). All of the models also have an inner cavity lined by hydrophobic amino acids (Figure 5A). In the template protein *A. thaliana* DIR1, the Cys residues 1–5, 2–3, 4–7, and 6–8 are paired, similarly to the family 2 nsLTPs. Since the structural models of MpLTPd5, PpLTPd15, MpLTPg3, and PpLTPg5 were constructed based on the *A. thaliana* DIR1 structure, they have the same disulfide pairing as DIR1. This disulfide pattern creates good-quality 3-D models, which suggests that the MpLTPd5, PpLTPd15, MpLTPg3, and PpLTPg5 proteins indeed have the same type of 8CM as the *A. thaliana* DIR1 and the LTP2s.

DISCUSSION

Previously, it has been shown that flowering plants have large gene families encoding nsLTPs (Jang et al., 2007; Boutrot et al., 2008). Here, we report that non-flowering plants such as gymnosperms, ferns, lycophytes, mosses, and liverworts also carry numerous nsLTP genes. On the other hand, the nsLTP genes were not identified in algae. The nsLTP gene content in non-seed plants is rather different compared to seed plants, since novel subfamilies have evolved during plant evolution (Figure 6). The nsLTP gene family Type D was established in

Table 6. The occurrence of non-specific lipid transfer proteins in *Pinus taeda*, and some of their features.

Name ^a	Signal peptide		Mature protein			Intron Amount / position ^c	Amino acid identity ^d
	Amino acids	Target ^b	Amino acids	Mass (Da)	pI		
Type 1							
PtLTP1.1	24	s	99	9982.51	9.23	n/a	
PtLTP1.2	24	s	99	10008.63	9.52	1 (5)	
PtLTP1.3	27	s	93	9615.06	9.9	n/a	
PtLTP1.4	24	s	99	10133.73	9.52	n/a	
PtLTP1.5	31	s	96	9527.98	10.22	n/a	
PtLTP1.6	27	s	92	9130.46	9.9	n/a	
PtLTP1.7	33	s	102	10996.18	4.02	n/a	
PtLTP1.8	28	s	96	9832.39	9.52	n/a	
PtLTP1.9	28	s	90	9754.14	6.91	n/a	
Type 2							
PtLTP2.1	26	s	71	7700.15	9.85	0	
Type C							
PtLTPc1	25	s	67	6693.72	9.82	n/a	
PtLTPc2	32	s	66	6686.42	3.34	n/a	
Type D							
PtLTPd1	21	s	85	9098.64	9.06	n/a	37% (AtLTPd5)
PtLTPd2	17	s	82	9081.77	9.11	0	35% (AtLTPd3)
PtLTPd3	26	s	78	8245.59	4.46	n/a	34% (AtLTPd3)
PtLTPd4	25	s	91	9287.16	10.86	n/a	62% (AtLTPd6)
PtLTPd5	23	s	88	8995.33	6.82	n/a	59% (AtLTPd7)
PtLTPd6	27	s	89	9119.52	9.17	n/a	56% (AtLTPd7)
PtLTPd7	25	s	98	10286.87	8.3	1 (4)	
PtLTPd8	25	s	88	9074.50	8.32	n/a	59% (AtLTPd7)
PtLTPd9	23	s	88	9032.46	9.17	n/a	59% (AtLTPd7)
PtLTPd10	25	s	91	9098.75	11.84	n/a	59% (OsLTPd5)
PtLTPd11	24	s	79	8211.56	9.99	1 (4)	
PtLTPd12	26	s	89	8178.67	10.71	n/a	43% (AtLTPd11)
Type F							
PtLTPf1	23	s	96	10134.83	10.22	n/a	
Type G							
PtLTPg1	19	s	228	21055.88	10.37	n/a	
PtLTPg2	25	s	176	17087.56	9.83	n/a	
PtLTPg3	27	s	152	14762.91	9.88	n/a	
PtLTPg4	23	s	178	18966.01	9.42	n/a	58% (OsLTPg2)
PtLTPg5	35	s	148	14801.23	10.12	1 (4)	
PtLTPg6	24	s	171	16110.11	4.37	n/a	50% (PtLTPg11)
PtLTPg7	24	s	170	17553.05	9.36	n/a	51% (OsLTPg2)
PtLTPg8	25	s	192	18839.59	11.17	n/a	
PtLTPg9	26	s	201	19568.32	10.6	n/a	
PtLTPg10	26	s	165	16279.59	8.31	n/a	
PtLTPg11	29	s	228	22066.11	10.9	n/a	
PtLTPg12	21	s	237	22889.84	10	n/a	
PtLTPg13	25	s	100	10252.62	5.04	n/a	
PtLTPg14	27	s	124	12940.80	4.39	n/a	
PtLTPg15	24	s	164	16691.87	4.57	n/a	64% (PtLTPg14)
PtLTPg16	25	s	164	17037.50	6.82	n/a	
PtLTPg17	22	s	199	19523.01	4.69	n/a	

Table 6. Continued

Name ^a	Signal peptide		Mature protein			Intron Amount / position ^c	Amino acid identity ^d
	Amino acids	Target ^b	Amino acids	Mass (Da)	pI		
Single							
PtLTPx1	28	s	103	11799.50	4.91	n/a	

a For accession numbers of corresponding ESTs, see Supporting Information Table S2
b Subcellular target for the protein. S = secretory pathway, m = mitochondria
c Position of the intron is given as the number of bases from the codon encoding the eighth cysteine in the 8CM. Due to lack of genomic sequences, information of introns is not available for most *P. taeda* nsLTPs
d The highest amino acid identity the sequence shows to a protein with specific characteristics for the type

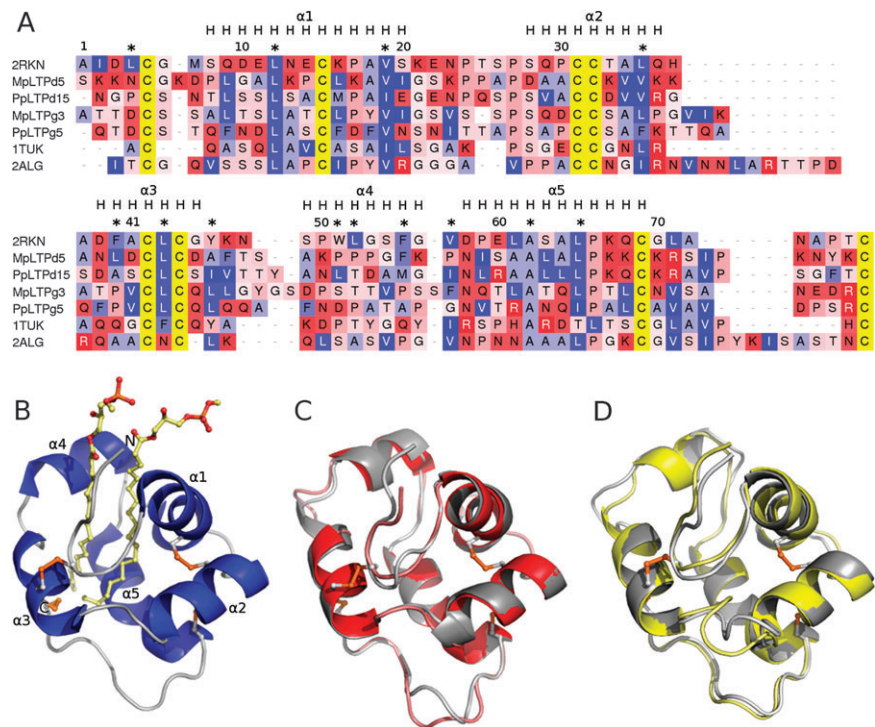


Figure 5. Sequence Alignment and Homology Models Based on the *A. thaliana* DIR1 Structure.

(A) Sequence alignment of *A. thaliana* DIR1 (2RKN; amino acids (aa) 1–77), MplTPd5 (aa 2–82), PpLTPd15 (aa 1–80), MplTPg3 (aa 1–84), PpLTPg5 (aa 1–81), wheat nsLTP (1TUK; aa 1–67), and peach nsLTP (2ALG; aa 1–87). The numbering of the MplTP and PpLTP sequences corresponds to the predicted mature protein after the removal of the signal peptide. The sequence numbering and the secondary structure of DIR1 are shown above the alignment. The wheat and peach nsLTPs were chosen as examples of a Type 2 and Type 1 nsLTP, respectively, since their crystal structure is known (Hoh et al., 2005; Pasquato et al., 2006). Amino acids that line the inner cavity of DIR1 (Lascombe et al., 2008) are marked with an asterisk and the conserved Cys residues of the 8CM are in yellow.
(B) The crystal structure of DIR1 (blue) in complex with two lysophosphatidyl choline molecules (shown as ball-and-stick; carbons in yellow). The five α -helices are marked similarly as in the sequence alignment in (A).
(C) The homology models of PpLTPd15 (red) and MplTPd5 (gray).
(D) The homology models of PpLTPg5 (yellow) and MplTPg3 (gray). The 8CMs of DIR1 (B), PpLTPd15 (C) and PpLTPg5 (D) are shown as ball-and-stick.

early diverged land plants and has since been retained in at least three of the four major lineages of land plants, such as liverworts, mosses, and vascular plants. We have not yet investigated whether the fourth major lineage, the hornworts, also carries this gene family.

Seemingly, the Type D genes represent the earliest nsLTP genes (Figure 1). The results from the phylogenetic analyses do not unambiguously indicate a monophyletic origin for Type D. Nonetheless, a monophyletic origin is strongly supported by the conserved intron positioned at the fourth base after the

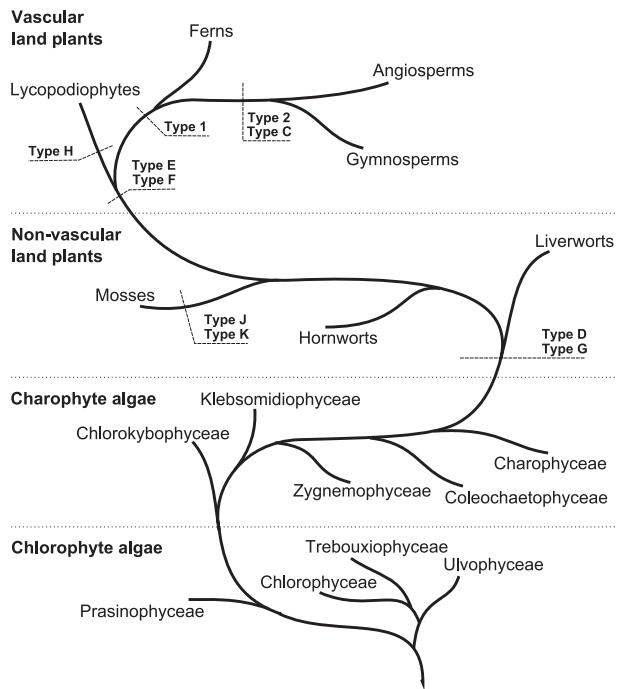


Figure 6. A Model for the Expansion of the nsLTP Family during Plant Evolution.

last Cys in the 8CM. The position of the intron has been conserved throughout the evolution of land plants, since it is present from liverworts and mosses up to flowering plants, such as *A. thaliana* (Figure 2). According to our expanded nsLTP classification system, the DIR1 protein from *A. thaliana* is a Type D nsLTP. The results of the structural modeling also support this classification of DIR1, since the sequence lengths and the structural features of the Type D proteins MpLTPd5 and PpLTPd15 are highly similar to the crystal structure of *A. thaliana* DIR1 (Figure 5). Furthermore, using the disulfide pattern of *A. thaliana* DIR, good-quality 3-D models of MpLTPd5 and PpLTPd15 could be generated, which additionally indicates that these proteins are of the same nsLTP type. Interestingly, the *A. thaliana* mutant *Defective in induced resistance 1-1* (*dir1-1*) is compromised in the production or transmission of an essential mobile signal to promote long-distance signaling. This leads to a deficient systemic acquired resistance in *dir1-1* (Maldonado et al., 2002). Whether other Type D nsLTPs have functions related to signaling remains to be investigated.

The Type G genes that carry a GPI-anchoring motif are also present in liverworts and all other investigated land plant species. Several of the Type G genes in *P. patens* and *S. moellendorffii* carry an intron at the same position as the Type D genes (Tables 2–4 and 6). It therefore seems likely that the Type G genes evolved after a fusion of a Type D gene with a GPI-anchoring motif. Since the Type G genes are present also in liverworts, this fusion occurred before the split between liverworts and mosses. However, we cannot completely

exclude that such a gene fusion event has occurred more than once during land plant evolution. The results of the structural modeling also support a close evolutionary relationship for the Type D and G proteins, since they indicate a similar, Type 2-like, disulfide pairing for the proteins (Figure 5). GPI-anchors are involved in attaching proteins to plasma membrane, although some proteins with GPI-anchors are further processed at the plasma membrane and covalently linked to components of the cell wall (Orlean and Menon, 2007). Clearly, the addition of the GPI-motif added novel functional features to the nsLTPs and the Type G genes were therefore retained in the plants.

A recent study by Lee et al. (2009) showed that the Type G nsLTP, LTPg1, from *A. thaliana* contributes either directly or indirectly to cuticular lipid accumulation. The *ltpg1* mutant contains a decreased amount of C29 alkane in the cuticle, compensated by an increased amount of C29 secondary alcohol and C29 ketone (Lee et al., 2009). DeBono et al. (2009) showed that LTPg1 is required for normal export of wax to the cuticle. The cuticle is a barrier not only to water, but also towards many other molecules and pathogens. It is present in all land plants but not in green algae.

The main nsLTP types in early diverged land plants, Types G and D, may be related to the adaptation to stressful conditions that followed with the establishment of plants on land. The other nsLTP types, such as Types 1 and 2, might then have evolved following the development towards more complex structures, such as roots, flowers, and seeds. This is supported by Park et al. (2000), who showed that a Type 1 nsLTP from lily (*Lilium longiflorum* cv. Nellie White) is necessary for adhesion of the pollen tube to the stigma/style during pollination, and further by Tomassen et al. (2007), who identified a Type 1 nsLTP from tomato (*Lycopersicon esculentum* cv. Counter) as an activator of polygalacturonase, a pectin-degrading enzyme involved in fruit ripening.

The Type 1 genes are expressed in the fern *A. capillus-veneris*, which suggests that this nsLTP type evolved after the separation of Lycopods and ferns. The Type 1 also shows resemblance to the Type D genes where the most obvious difference is the position of the intron. In Type D genes, the intron is located four bases downstream of the last Cys codon in the 8CM, while, in Type 1 genes, the intron boundary has slid one base and is now located five bases from the last Cys codon of the 8CM. Furthermore, the pairing of the Cys residues in the 8CM is different in Type 1. The comparison between Type 1 and 2 sequences and angiosperms showed that Type1 nsLTPs have either a hydrophilic or a small amino acid between Cys5 and Cys6 in the 8CM, whereas Type2 nsLTPs have a conserved large hydrophobic residue in this position. It has been speculated that the properties of the amino acid between Cys5 and Cys6 influence the selection of partners for the formation of the disulfide bridges in the 8CM (Douliez et al., 2001). The Type 1 sequences we identified in the fern *A. capillus-veneris* and in the gymnosperm *P. taeda* follow this pattern with glycine or a hydrophilic amino acid separating Cys5 and Cys6. On the other hand, all nsLTPs identified in *M. polymorpha*, *P. patens*,

and *S. moellendorffii* have a hydrophobic residue, such as leucine, phenylalanine, or valine, in this position (Figure 5). This further supports the results from our structural modeling, indicating that the disulfide bridges are formed by pairings of the Cys residues 1–5, 2–3, 4–7, and 6–8 in the nsLTPs expressed in liverworts, mosses, and lycopods.

In this investigation of nsLTP genes in non-flowering plants, the Type 2 genes were only identified in *P. taeda*, which suggests that this nsLTP type evolved after the separation of ferns and seed plants. The Type 2 genes always lack introns, but resemble the Type D genes in the spacing and disulfide bridge patterns between Cys residues in the 8CM (Table 1 and Figure 5). It seems possible that the Type 2 genes evolved from an intron loss event involving a Type D gene. Introns may be lost when the genomic copy of a gene undergoes a gene conversion with a reverse-transcribed copy of a spliced mRNA transcript (RT-mRNA), deleting one or more adjacent introns. Alternatively, introns could be lost by integration of a reverse transcribed mRNA into a novel locus or through exact genomic deletion (Roy and Gilbert, 2006). Events involving the loss of introns have not only been important for the formation of novel nsLTP types. They have also been important for gene formation within the specific gene types. For example, in *P. patens*, four of 21 and, in *S. moellendorffii*, six of 19 Type D genes lack introns.

Other members of the prolamine superfamily could not be found in green algae or in early diverged land plants. In *P. taeda*, we identified ESTs encoding proteins belonging to three other prolamine families: 2S Albumins, Hybrid proline-rich proteins and Prolamines. The two remaining families (α -amylase/trypsin inhibitors and puroindolines/grain softness proteins) only seem to exist in angiosperms. These protein families are probably also more specialized towards functions in complex structures and systems not found in the early diverging land plants. For example, the absence of puroindolines in all plants but angiosperms is explained by their function as regulators of hardness and texture of the endosperm (Morris, 2002). However, the precise function is not yet revealed for any of the families constituting the prolamine superfamily, so further investigations of all of them are of great importance.

In this work, we have explored the evolutionary history of plant nsLTPs. The results suggest that nsLTPs evolved very early during land plant evolution. The first land plants faced numerous challenges that included increased exposure to UV radiation, desiccation, and temperature stress when they adapted to a life on land approximately 470 million years ago (Wellman et al., 2003). Thus, it was necessary to develop many tools to cope with the numerous types of stresses on land. We speculate that the nsLTPs have properties that are useful for the plants during stressful conditions and therefore this gene family was selected and expanded in the first land plants. Further investigations, using, for instance, gene disruptions, ectopic expression, and gene expression analysis, will be required to let us learn more about the function of nsLTPs in non-seed plants. We hope that this work will provide a novel

perspective and additional tools in the search for the biological functions of the enigmatic nsLTP family.

METHODS

Databases

Databases with sequences from *P. taeda* (loblolly pine), *A. capillus-veneris* (maidenhair fern), *S. moellendorffii* (spikemoss), *P. patens* (moss), *M. polymorpha* (liverwort), *M. viride* (charophyte green alga), *V. carteri* (chlorophyte green alga), *C. reinhardtii* (chlorophyte green alga), *O. lucimarinus* (chlorophyte green alga), *O. tauri* (chlorophyte green alga), *C. merolae* (red alga), and *P. yezoensis* (red alga) were used. For *S. moellendorffii* (Weng et al., 2005), *P. patens* (Rensing et al., 2008), *V. carteri* (Prochnik et al., 2010), *C. reinhardtii* (Asamizu et al., 1999; Merchant et al., 2007), *O. lucimarinus* (Palenik et al., 2007; Lanier et al., 2008), and *C. merolae* (Misumi et al., 2005), both complete genome and ESTs were available, while there were only ESTs for *P. taeda* (Cairney et al., 2006), *A. capillus-veneris* (Yamauchi et al., 2005), *M. polymorpha*, *M. viride* (Simon et al., 2006), and *P. yezoensis* (Lee et al., 2000; Nikaido et al., 2000) and only genomic sequences for *O. tauri* (Derelle et al., 2006). When available, both databases with scaffold sequences and databases with gene models were used in the data mining. In searches for *S. moellendorffii* genes, genome version 1.0 and filtered gene models 3 were used. In *P. patens*, we used genome version 1.1 and filtered gene models 3. For *C. reinhardtii*, the gene models in version 3.0 of the genome sequence was used for the analysis. Only one version of genome sequence and gene models were available for the other organisms.

328 628 ESTs from *P. taeda*, 30 540 ESTs from *A. capillus-veneris*, 382 584 ESTs from *P. patens*, 33 692 ESTs from *M. polymorpha*, 15 972 ESTs from *M. viride*, and 22 069 ESTs from *P. yezoensis* were downloaded from the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov/). The EST databases were assembled into contigs by CAP EST Assembler (<http://host9.bioinfo3.ifom-iec-campus.it/cap/>) (Huang, 1992, 1996).

All protein and genomic sequences are from databases available at the US Department of Energy Joint Genome Institute's homepage (<http://genome.jgi-psf.org/>), except for the *C. merolae* sequences that were downloaded from the *C. merolae* Genome Project's homepages (<http://merolae.biol.s.u-tokyo.ac.jp/>).

Data Mining

BLAST was used for searches in sequences databases (Altschul et al., 1990). The searches were performed with blastp and tblastn, with default settings, except for the e-value, which was set to 1. All known nsLTP amino acid sequences from *A. thaliana* were used as bait. BLAST searches in genome, protein, and EST databases from each selected organism were performed, to increase the probability of detecting putative

nsLTP sequences. In addition, a search with the Hidden Markov model software HMMer was performed (<http://hmmer.janelia.org/>). The HMM model was built with HMMbuild, using an alignment with all sequences in the Pfam family 'Tryp_Alpha_Amyl' (Pfam:PF00234) as seed. The model was used as bait in a search with HMMsearch in the different databases. All settings were left as default. Searches with HMMer generated fewer hits than searches with BLAST, and no additional sequences were found with this method. To identify proteins with the 8CM, we also applied InterProScan searches (www.ebi.ac.uk/Tools/InterProScan/) using plant nsLTPs as baits.

Each identified protein sequence was BLAST searched back against the genome and the EST database, to check for introns and possible errors. The predicted amino acid sequences in the various databases were edited according to the ESTs, when such were available. Comparison of genomic sequences and ESTs were also used for identification of introns. Sixteen of the putative nsLTPs in *P. patens* and 30 in *S. moellendorffii* were not confirmed by ESTs. Since the genomes of *P. taeda*, *A. capillus-veneris*, *M. polymorpha*, and *M. viride* have not been completely sequenced, the use of genome databases to identify introns in the nsLTP genes was not possible in these species. The designations used for nsLTP sequences in this report and their corresponding database accession numbers are shown in Supplemental Table 1.

Phylogenetic Analyses

To predict sub-cellular targeting and the length of possible signal peptides, the programs SignalP and TargetP were used, with default settings (Emanuelsson et al., 2007). Molecular weight (Mw) and isoelectric point (pI) were computed for mature proteins with ExPasy's pI/Mw tool (Gasteiger et al., 2005). The eight Cys residues of the 8CM domain were excluded during pI calculation, since Cys residues in a disulfide bond have no charge. The prediction method PredGPI (<http://gpcr.biocomp.unibo.it/predgpi/>) was used to determine the presence of GPI-anchor sites (Pierleoni et al., 2008).

Multiple alignments were performed with the ClustalW method (Thompson et al., 1994), using the program ClustalW2 v2.0.7 (Larkin et al., 2007). All settings were left as default, except for the 'Gap extension penalty' and the 'Gap opening penalty', which were set to 0.01 and 10, respectively. Due to unconfirmed C-terminal ends of proteins lacking corresponding ESTs, only the 8CM part of the mature proteins were used in alignments.

The alignments were manually refined before input in ProtTest v2.4 (Abascal et al., 2005), which was set to all candidate models, a BIONJ tree, and the slow optimization strategy. LG +I +G was predicted as the best model and thus used to construct Maximum Likelihood phylogenetic trees using the program Phym1 v3.0 (Guindon and Gascuel, 2003). All other settings were left as default, but with 100 replicates for bootstrapping. In addition, a neighbor-joining tree was constructed using the programs Seqboot, ProtDist, Neighbor, and Consense from the PHYLIP package v3.67 (Felsenstein,

1989). The model PAM was used in ProtDist and 100 replicates for bootstrapping were constructed in Seqboot and used in both ProtDist and Neighbor, but all other settings were left as default. Some sequences were excluded from the final trees. The trees were viewed in the program FigTree v1.3.1 (<http://tree.bio.ed.ac.uk/software/figtree/>). BioEdit v7.0.9.0 (Hall, 1999) was used to calculate amino acid identities and to visualize the alignment. Both the trees and the multiple sequence alignment are available online at TreeBASE (www.treebase.org/; submission id: 11145; study access <http://purl.org/phylo/treebase/phyloids/study/TB2:S11155>).

mRNA Sequencing

Since many of the predicted proteins from *P. patens* and *S. moellendorffii* lacked corresponding ESTs in the databases, we isolated RNA from the plants in order to detect and sequence the missing mRNAs. *S. moellendorffii* was obtained from World of Ferns (Bangor, UK) and grown in soil in a greenhouse, while *P. patens* was grown on BCD agar supplemented with 1 mM CaCl₂, 5 mM Ammonium tartrate, and 0.8% agar (Nishiyama et al., 2000). RNA was isolated with RNeasy Plant Mini Kit (Qiagen, Hilden, Germany), eluted in RNase-free water, and stored at -80°C. Superscript III Reverse Transcriptase (Invitrogen, Carlsbad, CA, USA) was used to generate cDNA from the isolated RNA, according to the manufacturer's protocol. The primer used was an oligo(dT)₁₇ with an adaptor attached for 'rapid amplification of cDNA ends' (Supplemental Table 2). Polymerase chain reaction (PCR) was performed in order to amplify the gene-specific cDNAs. In PCR, the 3' primer was complementary towards the adaptor of the oligo(dT)₁₇ and the 5' primer was specific for the gene of interest. All primers for cDNA synthesis and PCR were ordered from Invitrogen (Supplemental Table 2). PCR products with a molecular size within the expected range were extracted from gel pieces (QIAEX II Gel Extraction Kit (Qiagen)) and sent to Eurofins MWG Operon (Ebersberg, Germany) for sequencing.

Genomic DNA Sequencing

In order to determine the intron positions in *M. polymorpha*, we isolated and sequenced DNA from plants growing in a greenhouse. For isolation of DNA, a DNeasy Plant Mini Kit (Qiagen) was used according to the manufacturer's protocol. The isolated DNA was then used in PCR with gene-specific primers (Supplemental Table 2). PCR products longer than estimated from the EST sequence were assumed to contain introns and were therefore extracted from agarose gels with the QIAEX II Gel Extraction Kit (Qiagen) and sequenced by Eurofins MWG Operon (Germany).

Structural Modeling

Homology models of the nsLTPs MpLTPd5, MpLTPg3 (Table 2), PpLTPd15, and PpLTPg5 (Table 3) were constructed based on the crystal structure of DIR1 from *A. thaliana* (pdb code 2RKN; Lascombe et al., 2008). A sequence alignment of MpLTPd5, PpLTPd15, MpLTPg3, PpLTPg5, and DIR1 was made using the

program MALIGN within the BODIL modeling environment (Lehtonen et al., 2004). For each of the target proteins, a set of 15 models was made based on the sequence alignment and the template structure and using the program MODELLER (Sali and Blundell, 1993). From each set of 15 models, the one with the lowest value of the MODELLER objective function was selected for further analysis. The models were superimposed on the DIR1 structure using the program VERTAA in BODIL.

Accession Numbers

Sequence data from this article can be found in the EMBL/GenBank data libraries under accession numbers GU330199, GU330200, GT889122, GT889123, GT889124, GT889125, GT889126, GT889127, GT889128, and GT889129.

SUPPLEMENTARY DATA

Supplementary Data are available at *Molecular Plant Online*.

FUNDING

This work was supported by the Carl Trygger Foundation, the Sigrid Jusélius Foundation, and the Joe, Pentti and Tor Borg Memorial Fund.

ACKNOWLEDGMENTS

We would like to thank Joel Hinz for assistance with the databases and Professor Mark Johnson for the excellent facilities at the Structural Bioinformatics Laboratory at Åbo Akademi University. No conflict of interest declared.

REFERENCES

- Abascal, F., Zardoya, R., and Posada, D. (2005). ProtTest: selection of best-fit models of protein evolution. *Bioinformatics*. **21**, 2104–2105.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.
- Asamizu, E., Nakamura, Y., Sato, S., Fukuzawa, H., and Tabata, S. (1999). A large scale structural analysis of cDNAs in a unicellular green alga, *Chlamydomonas reinhardtii*. I. Generation of 3433 non-redundant expressed sequence tags. *DNA Res.* **6**, 369–373.
- Boutrot, F., Chantret, N., and Gautier, M.F. (2008). Genome-wide analysis of the rice and *Arabidopsis non-specific lipid transfer protein (nsLtp)* gene families and identification of wheat *nsLtp* genes by EST data mining. *BMC Genomics*. **9**, 86.
- Cairney, J., et al. (2006). Expressed sequence tags from loblolly pine embryos reveal similarities with angiosperm embryogenesis. *Plant Mol. Biol.* **62**, 485–501.
- Cameron, K.D., Teece, M.A., and Smart, L.B. (2006). Increased accumulation of cuticular wax and expression of lipid transfer protein in response to periodic drying events in leaves of tree tobacco. *Plant Physiol.* **140**, 176–183.
- DeBono, A., Yeats, T.H., Rose, J.K., Bird, D., Jetter, R., Kunst, L., and Samuels, L. (2009). *Arabidopsis* LTPG is a glycosylphosphatidylinositol-anchored lipid transfer protein required for export of lipids to the plant surface. *Plant Cell*. **21**, 1230–1238.
- Derelle, E., et al. (2006). Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. *Proc. Natl Acad. Sci. U S A*. **103**, 11647–11652.
- Douliez, J.P., Michon, T., Elmorjani, K., and Marion, D. (2000). Structure, biological and technological functions of lipid transfer proteins and indolines, the major lipid binding proteins from cereal kernels. *J. Cer. Sci.* **32**, 1–20.
- Douliez, J.P., Pato, C., Rabesona, H., Molle, D., and Marion, D. (2001). Disulfide bond assignment, lipid transfer activity and secondary structure of a 7-kDa plant lipid transfer protein, LTP2. *Eur. J. Biochem.* **268**, 1400–1403.
- Edqvist, J., and Farbos, I. (2002). Characterization of germination-specific lipid transfer proteins from *Euphorbia lagascae*. *Planta*. **215**, 41–50.
- Eklund, D.M., and Edqvist, J. (2003). Localization of nonspecific lipid transfer proteins correlate with programmed cell death responses during endosperm degradation in *Euphorbia lagascae* seedlings. *Plant Physiol.* **132**, 1249–1259.
- Emanuelsson, O., Brunak, S., von Heijne, G., and Nielsen, H. (2007). Locating proteins in the cell using TargetP, SignalP, and related tools. *Nat. Protoc.* **2**, 953–971.
- Felsenstein, J. (1989). PHYLIP—Phylogeny Inference Package (Version3.2). *Cladistics*. **5**, 164–166.
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., and Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy Server. In *The Proteomics Protocols Handbook*, Walker J.M., ed. (Totowa, NJ: Humana Press Inc.), pp. 571–607.
- Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696–704.
- Hall, T.A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41**, 95–98.
- Hoh, F., Pons, J.L., Gautier, M.F., de Lamotte, F., and Dumas, C. (2005). Structure of a liganded type 2 non-specific lipid-transfer protein from wheat and the molecular basis of lipid binding. *Acta Crystallogr. D Biol.* **61**, 397–406.
- Huang, X. (1992). A contig assembly program based on sensitive detection of fragment overlaps. *Genomics*. **14**, 18–25.
- Huang, X. (1996). An improved sequence assembly program. *Genomics*. **33**, 21–31.
- Jang, C.S., Jung, J.H., Yim, W.C., Lee, B.M., Seo, Y.W., and Kim, W. (2007). Divergence of genes encoding non-specific lipid transfer proteins in the poaceae family. *Mol. Cells*. **24**, 215–223.
- Kader, J.C., Julienne, M., and Vergnolle, C. (1984). Purification and characterization of a spinach-leaf protein capable of transferring phospholipids from liposomes to mitochondria or chloroplasts. *Eur. J. Biochem.* **139**, 411–416.
- Kirubakaran, S.I., Begum, S.M., Ulaganathan, K., and Sakthivel, N. (2008). Characterization of a new antifungal lipid transfer protein from wheat. *Plant Physiol. Biochem.* **46**, 918–927.

- Kreis, M., Forde, B.G., Rahman, S., Mifflin, B.J., and Shewry, P.R. (1985). Molecular evolution of the seed storage proteins of barley, rye and wheat. *J. Mol. Biol.* **183**, 499–502.
- Lanier, W., Moustafa, A., Bhattacharya, D., and Comeron, J.M. (2008). EST analysis of *Ostreococcus lucimarinus*, the most compact eukaryotic genome, shows an excess of introns in highly expressed genes. *PLoS One*. **3**, e2171.
- Larkin, M.A., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*. **23**, 2947–2948.
- Lascombe, M.B., Bakan, B., Buhot, N., Marion, D., Blein, J.P., Larue, V., Lamb, C., and Prangé, T. (2008). The structure of 'defective in induced resistance' protein of *Arabidopsis thaliana*, DIR1, reveals a new type of lipid transfer protein. *Prot. Sci.* **17**, 1522–1530.
- Lauer, I., et al. (2009). The non-specific lipid transfer protein, Ara h 9, is an important allergen in peanut. *Clin. Exp. Allergy*. **39**, 1427–1437.
- Lee, E.K., Seo, S.B., Kim, T.H., Sung, S.K., An, G., Lee, C.H., and Kim, Y.J. (2000). Analysis of expressed sequence tags of *Porphyra yezoensis*. *Mol. Cells*. **10**, 338–342.
- Lee, J.Y., Min, K., Cha, H., Shin, D.H., Hwang, K.Y., and Suh, S.W. (1998). Rice non-specific lipid transfer protein: the 1.6 Å crystal structure in the unliganded state reveals a small hydrophobic cavity. *J. Mol. Biol.* **276**, 437–448.
- Lee, S.B., Go, Y.S., Bae, H.J., Park, J.H., Cho, S.H., Cho, H.J., Lee, D.S., Park, O.K., Hwang, I., and Suh, M.C. (2009). Disruption of glycosylphosphatidylinositol-anchored lipid transfer protein gene altered cuticular lipid composition, increased plastoglobules and enhanced susceptibility to infection by the fungal pathogen, *Alternaria brassicicola*. *Plant Physiol.* **150**, 42–54.
- Lehtonen, J.V., et al. (2004). BODIL: a molecular modelling environment for structure–function analysis and drug design J.. *Comput. Aided Mol. Des.* **18**, 401–419.
- Li, C., Xie, W., Bai, W., Li, Z., Zhao, Y., and Liu, H. (2008). Calmodulin binds to maize lipid transfer protein and modulates its lipids binding ability. *FEBS J.* **275**, 5298–5308.
- Maldonado, A.M., Doerner, P., Dixon, R.A., Lamb, C.J., and Cameron, R.K. (2002). A putative lipid transfer protein involved in systemic resistance signalling in *Arabidopsis*. *Nature*. **419**, 399–403.
- Merchant, S.S., et al. (2007). The Chlamydomonas genome reveals the evolution of key animal and plant functions. *Science*. **318**, 245–250.
- Misumi, O., Matsuzaki, M., Nozaki, H., Miyagishima, S.Y., Mori, T., Nishida, K., Yagisawa, F., Yoshida, Y., Kuroiwa, H., and Kuroiwa, T. (2005). *Cyanidioschyzon merolae* genome. A tool for facilitating comparable studies on organelle biogenesis in photosynthetic eukaryotes. *Plant Physiol.* **137**, 567–585.
- Molina, A., and Garcia-Olmedo, F. (1997). Enhanced tolerance to bacterial pathogens caused by the transgenic expression of barley lipid transfer protein LTP2. *Plant J.* **3**, 669–675.
- Morris, C.F. (2002). Puroindolines: the molecular genetic basis of wheat grain hardness. *Plant Mol. Biol.* **48**, 633–647.
- Nielsen, K.K., Nielsen, J.E., Madrid, S.M., and Mikkelsen, J.D. (1996). New antifungal proteins from sugar beet (*Beta vulgaris* L.) showing homology to non-specific lipid transfer proteins. *Plant Mol. Biol.* **31**, 539–552.
- Nieuwland, J., Feron, R., Huisman, B.A., Fasolino, A., Hilbers, C.W., Derksen, J., and Mariani, C. (2005). Lipid transfer proteins enhance cell wall extension in tobacco. *Plant Cell*. **17**, 2009–2019.
- Nikaido, I., Asamizu, E., Nakajima, M., Nakamura, Y., Saga, N., and Tabata, S. (2000). Generation of 10,154 expressed sequence tags from a leafy gametophyte of a marine red alga, *Porphyra yezoensis*. *DNA Res.* **7**, 223–227.
- Nishiyama, T., Hiwatashi, Y., Sakakibara, I., Kato, M., and Hasebe, M. (2000). Tagged mutagenesis and gene-trap in the moss, *Physcomitrella patens* by shuttle mutagenesis. *DNA Res.* **7**, 9–17.
- Orlean, P., and Menon, A.K. (2007). Thematic review series: lipid posttranslational modifications. GPI anchoring of protein in yeast and mammalian cells, or: how we learned to stop worrying and love glycosphospholipids. *J. Lipid Res.* **48**, 993–1011.
- Palenik, B., et al. (2007). The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation. *Proc. Natl Acad. Sci. U S A.* **104**, 7705–7710.
- Panikashvili, D., Shi, J.X., Bocobza, S., Franke, R.B., Schreiber, L., and Aharoni, A. (2010). The *Arabidopsis* DSO/ABCG11 transporter affects cutin metabolism in reproductive organs and suberin in roots. *Mol. Plant*. **3**, 563–575.
- Park, S.Y., Jauh, G.Y., Mollet, J.C., Eckard, K.J., Nothnagel, E.A., Walling, L.L., and Lord, E.M. (2000). A lipid transfer-like protein is necessary for lily pollen tube adhesion to an in vitro matrix. *Plant Cell*. **12**, 151–163.
- Pasquato, N., Berni, R., Folli, C., Folloni, S., Cianci, M., Pantano, S., Helliwell, J.R., and Zanotti, G. (2006). Crystal structure of peach Pru p 3, the prototypic member of the family of plant non-specific lipid transfer protein pan-allergens. *J. Mol. Biol.* **356**, 684–694.
- Pierleoni, A., Martelli, P.L., and Casadio, R. (2008). PredGPI: a GPI-anchor predictor. *BMC Bioinformatics*. **9**, 392.
- Pii, Y., Astegno, A., Peroni, E., Zaccardelli, M., Pandolfini, T., and Crimi, M. (2009). The *Medicago truncatula* N5 gene encoding a root-specific lipid transfer protein is required for the symbiotic interaction with *Sinorhizobium meliloti*. *Mol. Plant–Microbe Interact.* **22**, 1577–1587.
- Prochnik, S.E., et al. (2010). Genomic analysis of organismal complexity in the multicellular green alga *Volvox carteri*. *Science*. **329**, 223–226.
- Radauer, C., and Breiteneder, H. (2007). Evolutionary biology of plant food allergens. *J. Allergy Clin. Immunol.* **120**, 518–525.
- Rensing, S.A., et al. (2008). The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science*. **319**, 64–69.
- Roy, S.W., and Gilbert, W. (2006). The evolution of spliceosomal introns: patterns, puzzles and progress. *Nat. Rev. Genet.* **3**, 211–221.
- Salcedo, G., Sanchez-Monge, R., Diaz-Perales, A., Garcia-Casado, G., and Barber, D. (2004). Plant non-specific lipid transfer proteins as food and pollen allergens. *Clin. Exp. Allergy*. **34**, 1336–1341.
- Sali, A., and Blundell, T.L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815.
- Sánchez-Monge, R., Lombardero, M., García-Sellés, F.J., Barber, D., and Salcedo, G. (1999). Lipid-transfer proteins are relevant allergens in fruit allergy. *J. Allergy Clin. Immunol.* **103**, 514–519.

- Shewry, P.R., Beaudoin, F., Jenkins, J., Griffiths-Jones, S., and Mills, E.N.C. (2002). Plant protein families and their relationships to food allergy. *Biochem. Soc. Transact.* **30**, 906–910.
- Shin, D.H., Lee, J.Y., Hwang, K.Y., Kim, K.K., and Suh, S.W. (1995). High-resolution crystal structure of the non-specific lipid-transfer protein from maize seedlings. *Structure*. **3**, 189–199.
- Simon, A., Glöckner, G., Felder, M., Melkonian, M., and Becker, B. (2006). EST analysis of the scaly green flagellate *Mesostigma viride* (Streptophyta): implications for the evolution of green plants (Viridiplantae). *BMC Plant Biol.* **6**, 2.
- Sterk, P., Booij, H., Schellekens, G.A., Van Kammen, A., and De Vries, S.C. (1991). Cell-specific expression of the carrot EP2 lipid transfer protein gene. *Plant Cell*. **3**, 907–921.
- Thoma, S., Hecht, U., Kippers, A., Botella, J., De Vries, S., and Somerville, C. (1994). Tissue-specific expression of a gene encoding a cell wall-localized lipid transfer protein from *Arabidopsis*. *Plant Physiol.* **105**, 35–45.
- Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucl. Acids Res.* **22**, 4673–4680.
- Tomassen, M.M., Barrett, D.M., van der Valk, H.C., and Woltering, E.J. (2007). Isolation and characterization of a tomato non-specific lipid transfer protein involved in polygalacturonase-mediated pectin degradation. *J. Exp. Bot.* **58**, 1151–1160.
- Wellman, C.H., Osterloff, P.L., and Mohiuddin, U. (2003). Fragments of the earliest land plants. *Nature*. **425**, 282–285.
- Weng, J.K., Tanurdzic, M., and Chapple, C. (2005). Functional analysis and comparative genomics of expressed sequence tags from the lycophyte *Selaginella moellendorffii*. *BMC Genomics*. **6**, 85.
- Yamauchi, D., Sutoh, K., Kanegae, H., Horiguchi, T., Matsuoka, K., Fukuda, H., and Wada, M. (2005). Analysis of expressed sequence tags in prothallia of *Adiantum capillus-veneris*. *J. Plant Res.* **118**, 223–227.
- Zhang, D.S., Liang, W.Q., Yuan, Z., Li, N., Shi, J., Wang, J., Liu, Y.M., Yu, W.J., and Zhang, D.B. (2008). Tapetum degeneration retardation is critical for aliphatic metabolism and gene regulation during rice pollen development. *Mol. Plant*. **1**, 599–610.
- Zoccatelli, G., Pokoj, S., Foetisch, K., Bartra, J., Valero, A., Del Mar San Miguel-Moncin, M., Vieths, S., and Scheurer, S. (2010). Identification and characterization of the major allergen of green bean (*Phaseolus vulgaris*) as a non-specific lipid transfer protein (Pha v 3). *Mol. Immunol.* **47**, 1561–1568.

Supplemental Table 1. Designations Used for nsLTP Sequences in this Report and their Corresponding Database Accession Numbers.

Supplemental Table 2. Primers Used for cDNA Synthesis and Polymerase Chain Reactions (PCR).

Supplementary Table S1. Primers used for cDNA synthesis and Polymerase Chain Reactions (PCR).

Name		
Oligo(dT) with adaptor	GGCCACGCGTCGACTAGTACTTTTTTTTTTTTTTTTTT	
Primer specific for adaptor	GGCCACGCGTCGACTAGTAC	
	Forward primer	Reverse primer
MpLTPd1	TGTTGCAGATGTTTGTGACAC	CTCAACAGAATTCTGAACTTGA
MpLTPd2	GAGAGATTTTCCTCATGGCATT	AGAGAAAACCCAGCGAATAACTC
MpLTPd3	CGAGTCTCTCTGTCTCACAATGG	TTGTATACCTCATGCTCATCCAAG
MpLTPd4	TCGAGCTATGAGAATTGTAGCA	TTGACGTTTCATGGCGACA
MpLTPd5	CAATGGCGACTACTCGTGC	GAAGACCAGGACGAATTCG
MpLTPd6	GGGCCCTTCATTTAGCTCTG	TATGGCAAGATCTTCCAATCTG
MpLTPd7	CGATCATAATTCTCTGTTATTCG	TGTTGCATGATAAACACGTT
MpLTPd8	CTGCCCTGCAGATGAGAAG	CATGGCCAAAACAAAGAACG
MpLTPg1	AACATGGCAGCGGTGTCT	CTACTGACAATGACAGATCTGAGC
MpLTPg2	TGCACCAACAGTTGCACC	AAGCAAGTCTTTCAGCCCTTC
MpLTPg3	TCTTGGGCTGAGATAGAAGAATG	AGGGGATCGGGATGGAGA
MpLTPg4	GGAGGCCTCATTTATTGTC	AGTCTAAGCGGATTGAGCTTCTC
MpLTPx1	GCTGTTTTTCATAACAGCAGGTC	CTAAAAGCAGGCAAATCAAGAT
MpLTPx2	GCAAATTCAGTGGACCTTTGT	CGAGCTTGCACGGAAAAG
	Outer primer	Inner primer
PpLTPd1	TTGTGTCCTCTTCACAACGA	GGCGAATAGCTCTGCTGTG
PpLTPd2	CAACAGCTCTGGGGTAACAG	AGCTCTGGTGGTGGTGTCT
PpLTPd6	GTATTGCAGCCAGCAAGAAT	AGGCAATGCTGCTCCAAC
PpLTPd7	TCAGCCTCTTGTCATAGTTGC	AAAGTTTTTGAAGCGTTGC
PpLTPd12	GAAGTTGTGGGCTTGTGACT	GTCTCAGTGTGAGCCGTTTG
PpLTPd15	CAAGGGAAAATATGGAAGCA	CATCGAAAATCCACGCTGT
PpLTPd16	GAAGCCTGTGCAATATCCTC	TCTCCTCGTTGTGACCTCT
PpLTPd17	GCCTCATCCACCTAACAAGA	ATGAAGAGCTTTGCGGTGAT
PpLTPd21	GGAAGCACATTTGGGACTCT	CCTCGTGCTCCACTTTCTA
PpLTPg10	TCTAGAGGATCCCCAGGTTT	GACGTGGATCCGTTTCTGAT
PpLTPi1	TGGACTCCCTGAACAACCTA	TGAAATCTGCACCAAAGCAC
PpLTPi2	CCGAGTGTGAGGAGCTAAGA	AATGTCTGCGACGAGAATCC
PpLTPi5	CTGTATGAGGCGTCGGTAAC	AATGGGGACTTGGAATGTGA
PpLTPk1	GGGTGCACTGAAACAAGTCT	GACCCTCGACATGACGAAAC
SmLTPd1	TCCAAGCCAGGGATCATCTA	TTGGGAGACTTCCGAGCTT
SmLTPd2	AGAAGCACGGAGATCAGCAT	AGGCGAGCGAGTAGGAAGAT
SmLTPd3	ACAGTTCAGGCGATCACTCC	ATTCGCGTTCTTCCAATCAG
SmLTPd4	TTGCTTCTCAGCTCCTCACA	GTATCCATGGCTCTCGCAA
SmLTPd5	GTGCGTAGATCTCGGCTTTC	CAATGGATTCAAGACGCTCT
SmLTPd6	GCTGGATCACCGCTCTTTAC	GTTTGGATCCATCCATCGTC
SmLTPd14	AGTGTTTCATCAAACAGATCTCCA	GCATTGATGAATGGCAGTGA

SmLTPd15	TCTCGAGTGTCTCTCCTTCG	GCACAAATGGAGACGATCAA
SmLTPe1	TTGCTCGCAGTTGATCTTTG	AAGGATGGAGTCAGGTGCAG
SmLTPe3	ATCGCGATCACGAGTAACTG	TTTCGCTGAATCGTCTTTGA
SmLTPf1	CCAGAGGGTATCACCACCAC	GCATTCAAAGGAGGGAGAGA
SmLTPf2	TCGATCGAAATCTCCCAGTC	CGATCACCGCGAAATTATG
SmLTPg1	TCTTTGGCTCACCAAATTCC	TTCTTCCAATTGCCTCGATT
SmLTPg2	CTCGATCACGAAAAGCCTTC	ACCAGGCTGTGAAACACCTC
SmLTPg4	CAGGCAAAGGGATCGAGATA	GGATCGATCGGTAGGATGAA
SmLTPg5	TTTGTACAGGGGATTTGTTAGC	AGACAATGCTATGTTGAATGACA
SmLTPg6	GGGCATCACGAACTCCATAG	CTCTACTGCAATGGGCTCGT
SmLTPg10	AACTTGTAAGGATGCCAAGGAG	CTCTTGCTGGCGTTGGAT
SmLTPg11	CGCACTTTCCCTCACTCAAT	ATCACTCCACCGCATTCTCT
SmLTPg12	CGCGGTAATAAGATCCGAAG	TTTGAGACCGACCATCTTCA
SmLTPh1	CGAAGCAAACATCCATCG	TCGAGATTGCAGCAGTAGTGA
SmLTPh2	TCGTTTTGTCAAACCAACTCA	ACTACCTCTGGAGTTGCCTCT
SmLTPh3	GAAAGGCTTCCTTGCTGCTA	CCCTAATGGATTCCAGAGCA
SmLTPx1	GGACGAGCTCTTGAGGAACA	GGTTGCTTGCTTGCTTCTCT

Supplementary Table S2. Designations used for nsLTP sequences in this report, and their corresponding database accession numbers. The EST names are GenBank accession numbers, while the names of protein and genome sequences are DOE Joint Genome Institute's identification numbers.

M. polymorpha

Name	EST
MpLTPd1	BJ869728
MpLTPd2	BJ842409
MpLTPd3	BJ842727
MpLTPd4	BJ845694
MpLTPd5	BJ855616
MpLTPd6	BJ858639
MpLTPd7	BJ862476
MpLTPd8	BJ872793
MpLTPg1	BJ842922
MpLTPg2	BJ843944, BJ855633
MpLTPg3	BJ846972
MpLTPg4	BJ850882
MpLTPx1	BJ848978
MpLTPx2	BJ863805

P. patens

Name	Protein	Adjusted to
PpLTPd1	130943	EST DC904088
PpLTPd2	135799	EST GT889122
PpLTPd3	161674	EST BJ976131
PpLTPd4	168323	EST BJ977829
PpLTPd5	233050	-
PpLTPd6	122287	EST GT889123
PpLTPd7	134241	EST GT889124
PpLTPd8	162185	EST BJ977311
PpLTPd9	145701	EST AW561377
PpLTPd10	202862	-
PpLTPd11	77295	EST DC904521
PpLTPd12	159734	EST DC941596
PpLTPd13	171216	EST FC430250
PpLTPd14	171703	EST DC908802
PpLTPd15	81072	-
PpLTPd16	93101	-
PpLTPd17	123360	-

PpLTPd18	Not predicted	EST BJ590127
PpLTPd19	Not predicted	EST BJ611281
PpLTPd20	Not predicted	Genome. Scaffold 343, 348846-349181
PpLTPd21	97321	-
PpLTPg1	65746	Start adjusted to scaffold 7, pos 2457136
PpLTPg2	111664	-
PpLTPg3	163050	EST BJ975000
PpLTPg4	163905	Start adjusted to scaffold 61, pos 1679106
PpLTPg5	168261	Start adjusted to scaffold 171, pos 142928
PpLTPg6	170287	-
PpLTPg7	171728	ESTs BJ966007, DC957097
PpLTPg8	217301	-
PpLTPg9	234183	-
PpLTPg10	172621	-
PpLTPj1	72720	Genome. Scaffold 38, 1259147-1259632
PpLTPj2	91074	-
PpLTPj3	168243	-
PpLTPj4	168466	EST DC924087
PpLTPj5	170209	Genome. Scaffold 235. 511536-512315
PpLTPj6	172940	ESTs BJ947435 and PP004038338R
PpLTPj7	168223	ESTs DC945461 and BJ176916.1
LTPk1	82551	-
LTPk2	166169	-

S. moellendorffii

Name	Protein	Adjusted to
SmLTPd1	412994	-
SmLTPd2	427211	Start adjusted to scaffold 89, pos.194054.
SmLTPd3	402749	-
SmLTPd4	403588	-
SmLTPd5	403691	-
SmLTPd6	405720	-
SmLTPd7	413193	-
SmLTPd8	428963	EST FE461209
SmLTPd9	441550	-
SmLTPd10	445624	-
SmLTPd11	446003	EST FE462950
SmLTPd12	446685	-
SmLTPd13	448612	-

SmLTPd14	135066	-
SmLTPd15	412273	EST GT889125
SmLTPd16	412899	EST FE457381
SmLTPd17	Not predicted	Genome. Scaffold 61. 1095996-1096337
SmLTPd18	445118	EST DN838270
SmLTPd19	431493	Genome. Scaffold 155. 55380-55785
SmLTPe1	425055	-
SmLTPe2	425085	-
SmLTPe3	437676	Genome. Scaffold 0. 6116200-6116517
SmLTPf1	403057	Genome. Scaffold 0. 6392788-6393144
SmLTPf2	408482	EST GT889126
SmLTPg1	79697	-
SmLTPg2	87601	EST GT889127
SmLTPg3	140578	-
SmLTPg4	403297	-
SmLTPg5	403556	-
SmLTPg6	403686	EST GT889128
SmLTPg7	409739	-
SmLTPg8	443870	-
SmLTPg9	Not predicted	EST FE426958
SmLTPg10	413655	-
SmLTPg11	99365	-
SmLTPg12	124017	-
SmLTPh1	413644	EST GT889129
SmLTPh2	413648	Genome. Scaffold 31. 882194-882517
SmLTPh3	416643	Genome. Scaffold 31. 1110485-1110802
SmLTPh4	430505	EST FE450991
SmLTPh5	416608.1	Genome. Scaffold 31. 888446-888769
SmLTPh6	416608.2	Genome. Scaffold 31. 884697-885020
SmLTPx1	427417	-

A. capillus-veneris

Name	EST
AcvLTP1.1	DK952812
AcvLTP1.2	BP917429
AcvLTP1.3	DK947840
AcvLTPg1	DK956455
AcvLTPg2	DK949778
AcvLTPx1	DK958277

P. taeda

Name	EST
PtLTP1.1	DR011430
PtLTP1.2	DR055396
PtLTP1.3	DN449630
PtLTP1.4	CO361240
PtLTP1.5	AW043179
PtLTP1.6	DN465131
PtLTP1.7	AW982101, DT625695
PtLTP1.8	AW011402
PtLTP1.9	DT637000, DR691856
PtLTP2.1	DR743561
PtLTPc1	AW982009
PtLTPc2	AW981528
PtLTPd1	DR168700
PtLTPd2	DR091150
PtLTPd3	DR052695
PtLTPd4	DR052578
PtLTPd5	DR745097
PtLTPd6	DR072724
PtLTPd7	CF403062
PtLTPd8	DR090808
PtLTPd9	DR119451
PtLTPd10	CX648438
PtLTPd11	DR389023
PtLTPd12	DR743808
PtLTPf1	DR022257
PtLTPg1	CF393843, DR055946
PtLTPg2	CX652807
PtLTPg3	CO363298
PtLTPg4	DR014439
PtLTPg5	DR071209
PtLTPg6	CF672424
PtLTPg7	DT625889
PtLTPg8	CX649189
PtLTPg9	CF473417, CF479647
PtLTPg10	CV135072
PtLTPg11	DT626206

PtLTPg12	DR057416
PtLTPg13	CF399301
PtLTPg14	DR097096
PtLTPg15	DR684739
PtLTPg16	DR079008
PtLTPg17	DR094242
PtLTPx1	DR079057

A. thaliana

Name	Locus
AtLTP1.1	At2g15050.2
AtLTP1.2	At2g15325.1
AtLTP1.3	At2g18370.1
AtLTP1.4	At2g38530.1
AtLTP1.5	At2g38540.1
AtLTP1.6	At3g08770.1
AtLTP1.7	At3g51590.1
AtLTP1.8	At3g51600.1
AtLTP1.9	At4g33355.1
AtLTP1.10	At5g01870.1
AtLTP1.11	At5g59310.1
AtLTP1.12	At5g59320.1
AtLTP2.1	At1g43665
AtLTP2.2	At1g43666.1
AtLTP2.3	At1g43667.1
AtLTP2.4	At1g48750.1
AtLTP2.5	At1g66850.1
AtLTP2.6	At1g73780.1
AtLTP2.7	At2g14846.1
AtLTP2.8	At3g12545
AtLTP2.9	At3g18280.1
AtLTP2.10	At3g57310.1
AtLTP2.11	At5g38160.1
AtLTP2.12	At5g38170.1
AtLTP2.13	At5g38180.1
AtLTP2.14	At5g38195.1
AtLTPc1	At5g07230.1
AtLTPc2	At5g52160.1
AtLTPc3	At5g62080.1

O. sativa

Name	Locus
OsLTP1.1	Os01g12020.1
OsLTP1.2	Os01g60740
OsLTP1.3	Os03g59380.1
OsLTP1.4	Os05g40010.1
OsLTP1.5	Os06g06340.1
OsLTP1.6	Os06g34840.1
OsLTP1.7	Os08g03690.1
OsLTP1.8	Os11g02330
OsLTP1.9	Os11g02350.1
OsLTP1.10	Os11g02379.1
OsLTP1.11	Os11g02379.2
OsLTP1.12	Os11g02400.1
OsLTP1.13	Os11g02424.2
OsLTP1.14	Os11g24070.1
OsLTP1.15	Os12g02290.1
OsLTP1.16	Os12g02300.1
OsLTP1.17	Os12g02310.1
OsLTP1.18	Os12g02320.1
OsLTP1.19	Os12g02330.1
OsLTP1.20	Os12g02340.1
OsLTP2.1	Os01g49640.1
OsLTP2.2	Os01g49650.1
OsLTP2.3	Os03g02050.1
OsLTP2.4	Os05g47700.1
OsLTP2.5	Os05g47730.1
OsLTP2.6	Os06g49190.1
OsLTP2.7	Os10g36070.1
OsLTP2.8	Os10g36090.1
OsLTP2.9	Os10g36100

AtLTPd1	At5g48485.1	OsLTP2.10	Os10g36110.1
AtLTPd2	At5g48490.1	OsLTP2.11	Os10g36160.1
AtLTPd3	At5g55410.1	OsLTP2.12	Os10g36170.1
AtLTPd4	At5g55450.1	OsLTP2.13	Os11g40530.1
AtLTPd5	At5g55460.1	OsLTPc1	Os08g43290.1
AtLTPd6	At2g37870.1	OsLTPc2	Os09g35700.1
AtLTPd7	At3g53980.1	OsLTPd1	Os01g68580.1
AtLTPd8	At5g05960.1	OsLTPd2	Os01g68589.1
AtLTPd9	At1g32280.1	OsLTPd3	Os07g18750.1
AtLTPd10	At4g30880.1	OsLTPd4	Os07g18990.1
AtLTPd11	At4g33550	OsLTPd5	Os01g62980.1
AtLTPd12	At5g56480.1	OsLTPd6	Os04g33920.1
AtLTPe1	At3g07450.1	OsLTPd7	Os04g33930.2
AtLTPe2	At3g52130.1	OsLTPd8	Os05g06780.1
AtLTPg1	At1g03103	OsLTPd9	Os01g58650.1
AtLTPg2	At1g05450	OsLTPd10	Os01g58660.1
AtLTPg3	At1g18280	OsLTPd11	Os10g05720.2
AtLTPg4	At1g27950	OsLTPd12	Os11g29420.1
AtLTPg5	At1g36150	OsLTPg1	Os01g59870
AtLTPg6	At1g55260	OsLTPg2	Os03g07100
AtLTPg7	At1g62790	OsLTPg3	Os03g09230
AtLTPg8	At1g73550	OsLTPg4	Os03g20760
AtLTPg9	At1g73560	OsLTPg5	Os03g26800
AtLTPg10	At1g73890	OsLTPg6	Os03g26820
AtLTPg11	At2g13820	OsLTPg7	Os03g57970
AtLTPg12	At2g27130	OsLTPg8	Os03g57980
AtLTPg13	At2g44290	OsLTPg9	Os03g57990
AtLTPg14	At2g44300	OsLTPg10	Os03g58940
AtLTPg15	At2g48130	OsLTPg11	Os04g38840
AtLTPg16	At2g48140	OsLTPg12	Os05g41030
AtLTPg17	At3g22570	OsLTPg13	Os06g47200
AtLTPg18	At3g22580	OsLTPg14	Os07g07790
AtLTPg19	At3g22600	OsLTPg15	Os07g07870
AtLTPg20	At3g22620	OsLTPg16	Os07g07860
AtLTPg21	At3g43720	OsLTPg17	Os07g07920
AtLTPg22	At3g58550	OsLTPg18	Os07g07930
AtLTPg23	At4g08670	OsLTPg19	Os07g09970
AtLTPg24	At4g12360	OsLTPg20	Os07g30590
AtLTPg25	At4g14805	OsLTPg21	Os07g43290

AtLTPg26	At4g14815	OsLTPg22	Os08g42040
AtLTPg27	At4g22630	OsLTPg23	Os11g37320
AtLTPg28	At4g22666	OsLTPg24	Os06g49770
AtLTPg29	At5g09370	OsLTPg25	Os11g37280.1
AtLTPg30	At5g13900	OsLTPx1	Os03g44000.1
AtLTPg31	At5g64080	OsLTPx2	Os07g27940.1
AtLTPg32	At1g70250	OsLTPx3	Os11g34660
AtLTPx1	At1g52415		
AtLTPx2	At1g64235		
AtLTPx3	At4g08530		
AtLTPx4	At4g28395		