Research article

Molecular diversity of phospholipase D in angiosperms

Marek Eliáš*[1], Martin Potocký[2], Fatima Cvrčková[1] and Viktor Žárský[1,3]

Address: [1]Department of Plant Physiology, Faculty of Science, Charles University, Viničná 5, Prague 2, Czech Republic, [2]Department of Biochemistry and Microbiology, Institute of Chemical Technology, Technická 5, Prague 6, Czech Republic and [3]Institute of Experimental Botany, The Academy of Sciences of the Czech Republic, Rozvojová 135, Prague 6, Czech Republic

E-mail: Marek Eliáš* - melias@natur.cuni.cs; Martin Potocký - potockym@vscht.cz; Fatima Cvrčková - fatima@natur.cuni.cz; Viktor Žárský - zarsky@ueb.cas.cz

*Corresponding author

Abstract

Background: The phospholipase D (PLD) family has been identified in plants by recent molecular studies, fostered by the emerging importance of plant PLDs in stress physiology and signal transduction. However, the presence of multiple isoforms limits the power of conventional biochemical and pharmacological approaches, and calls for a wider application of genetic methodology.

Results: Taking advantage of sequence data available in public databases, we attempted to provide a prerequisite for such an approach. We made a complete inventory of the *Arabidopsis thaliana* PLD family, which was found to comprise 12 distinct genes. The current nomenclature of *Arabidopsis* PLDs was refined and expanded to include five newly described genes. To asses the degree of plant PLD diversity beyond *Arabidopsis* we explored data from rice (including the genome draft by Monsanto) as well as cDNA and EST sequences from several other plants. Our analysis revealed two major PLD subfamilies in plants. The first, designated C2-PLD, is characterised by presence of the C2 domain and comprises previously known plant PLDs as well as new isoforms with possibly unusual features-catalytically inactive or independent on Ca^{2+}. The second subfamily (denoted PXPH-PLD) is novel in plants but is related to animal and fungal enzymes possessing the PX and PH domains.

Conclusions: The evolutionary dynamics, and inter-specific diversity, of plant PLDs inferred from our phylogenetic analysis, call for more plant species to be employed in PLD research. This will enable us to obtain generally valid conclusions.

Background

Phospholipase D (PLD, EC 3.1.4.4.) is a ubiquitous eukaryotic enzyme participating in various cellular processes (for a review see [1,2]). Biochemically distinct types of PLDs have been described, but only two, the mammalian glycosylphosphatidylinositol-specific PLD (GPI-PLD) and a family usually referred to as phosphatidylcholine-specific PLD (PC-PLD), have been characterised also on the molecular level. Two distinct PC-PLD genes have been identified in mammals; they seem to be involved in signal transduction and vesicular trafficking. The yeast *Saccharomyces cerevisiae* contains only one gene from the PC-PLD family and its function in sporulation has been recognised.

Plants are a traditional model for PLD research. Indeed, PLD activity was first described from a plant source [3], and the first cloned eukaryotic cDNA coding for a PLD was isolated from the castor bean, *Ricinus communis*[4]. Using mainly biochemical and pharmacological approaches, plant PLD has been implicated in many cellular processes (reviewed in [2]). Beside its roles in membrane degradation and turnover during senescence, seed germination and under stress conditions, plant PLD is emerging as an important component of signal transduction cascades, e.g. in response to wounding, abscisic acid [2] or Nod factors [5]. Earlier pharmacological evidence for the involvement of heterotrimeric G-proteins in plant PLD regulation [6] has been recently strengthened by a report on direct interaction of an alpha subunit of a G-protein with PLDα in tobacco [7]. Also products of PLD action, i.e. phosphatidic acid (PA), diacylglycerol and N-acylethanolamine, are potential signalling molecules in plants (reviewed in [8,9]).

Up to now about 20 PLDs have been cloned from plants. Multiple isoforms have been found in some species, complicating the study of plant PLD. Application of a reverse genetic approach, combining the knowledge of genomic sequences and molecular genetic techniques, holds greatest promises here. This can be documented e.g. by successful inactivation of the *AtPLDα1* gene in *Arabidopsis thaliana* by antisense strategy, which allowed identification of a novel PLD activity in plants [10,11].

Thorough characterisation of the gene family concerned is an obvious prerequisite for productive application of the reverse-genetic approach. Here we present the results of a detailed comparative analysis of the Arabidopsis and rice PLD families, combining data from the complete Arabidopsis genome sequence [12], publicly available rice (*Oryza sativa*) genomic and cDNA sequences, and the draft rice genome data made available by Monsanto [13]. Extensive EST collections from three plant species, tomato (*Lycopersicon esculentum*), *Medicago truncatula* and *Sorghum bicolor*, have been included into the analysis to provide the insight into the inter-specific variability of plant PLDs. Our results indicate that the angiosperm PLD family, comprising two major subfamilies (C2- and PXPH-PLDs), is evolutionarily very dynamic, and conclusions based on a single species (such as Arabidopsis) might not therefore be simply applicable to others.

Results and Discussion
A dozen Arabidopsis PLDs
Up to now several cDNAs representing six distinct PLD-encoding genes have been reported from Arabidopsis (Table 1). Using the cloned PLDs from Arabidopsis and other organisms we conducted exhaustive BLAST searches of the Arabidopsis sequences available from GenBank and

found 12 genes from the eukaryotic PC-PLD family, five of them not yet recorded in the literature. All genes found code for proteins containing all the conserved sequence motifs characteristic of eukaryotic PLDs, including two copies of the invariant catalytic HxKxxxxD motif [1], suggesting that all probably posses the genuine PLD enzymatic activity (thougt this must be proven experimentally). The catalytic HxKxxxxD motif is shared also by other proteins put together with the eukaryotic PLDs into the PLD superfamily [14], we however did not identify any other members of the superfamily in Arabidopsis besides the 12 PC-PLDs.

Before attempting a detailed phylogenetic analysis of the Arabidopsis PLDs, we used a combination of computational tools, comparison with cDNAs/ESTs and information from protein alignments to verify the exon-intron structures proposed by AGI annotators (see Materials and Methods). In several cases, prediction ambiguities and cloning or sequencing errors have been uncovered, and refined gene models and protein sequence predictions have been obtained and used in further analysis (see the discussion below and Additional file 1.

All Arabidopsis PLDs can be classified into two subfamilies (Fig. 1A). Since these subfamilies differ by the presence of distinct N-terminal phospholipid-binding domains (C2 vs. PX-PH), we propose denoting them as C2-PLDs and PXPH-PLDs. The C2-PLD subfamily contains 10 Arabidopsis isoforms harbouring a phospholipid/Ca^{2+}-binding fold called the C2 (or CalB) domain (Fig. 2, 5). Seven genes have been already found in the genome, including *AtPLDα1* (formerly PLDα), *AtPLDβ1* (previously PLDβ) and a tandem triplication of *AtPLDγ1*, *AtPLDγ2* and *AtPLDγ3*[2]. A gene tentatively designated as PLDδ1 [2] should in our view be more suitably labelled *AtPLDβ2* (see below), while we keep the term *AtPLDδ* for a gene recently reported with this designation [15,16]. (see Table 1).

We suggest the term *AtPLDα2* for a newly identified homolog closely related to *AtPLDα1* (88 % sequence identity at the protein level). Interestingly, *AtPLDα1* and *AtPLDα2* genes reside within one of several large-scale intragenomic duplications believed to be remnants of a tetraploidisation event dated 112 Myr ago [12], pointing toward probable evolutionary origin of these two paralogs. The remaining Arabidopsis C2-PLD genes do not correspond to any of the previously established group, so we propose terming them *AtPLDε* and *AtPLDζ*. (although a PLDε has already been mentioned in a recent review [17], it is not clear from the text to which of the Arabidopsis PLD genes it corresponds to.)

Table 1: Summary on Arabidopsis phospholipase D genes. All genes from the eukaryotic PC-PLD family identified in the complete Arabidopsis genome sequence are listed. An additional gene, PLDδ2, has been reported in the literature [2], there is, however, no corresponding ORF in the genomic clone allegedly harbouring the gene (AC004708), and the genomic clone itself has been annotated as "spurious" and excluded from the final assemblage of the complete Arabidopsis genome.

proposed gene name	original gene name[a]	genomic locus[b]	number of coding exons	gene sequence[c]	cDNA	number of ESTs	references
AtPLDα1	PLDα	At3g15730 / MSJ11.13	3	AB017071, NC_003074	U36381, AF428278[d]	40	[2], [43]
AtPLDα2		At1g52570 / F6D8.21	3	AC008016, NC_003070	-	0	-
AtPLDβ1	PLDβ	At2g42010 / T6D20.10	10[e]	U90439, NC_003071	U84568	5	[2], [11]
AtPLDβ2	PLDδ1	At4g00240 / F5I10.13	10[f]	AF195115, AL161471, AF013293, NC_003075	-	0	-
AtPLDγ1		At4g11850 / T26M18.60	10	AL161532, AL078606, NC_003075	AF027408	8	[2], 27]
AtPLDγ2		At4g11830 / T26M18.40	10 (11)[g]	AL161532, AL078606, NC_003075	AF138281[h]	1	[2], [18]
AtPLDγ3		At4g11840 / T26M18.50	10	AL161532, AL078606, NC_003075	AF138281[h]	2	[2], [18]
AtPLDδ		At4g35790 / F4B14.60	10[i]	AL161588, AL031986, NC_003075	AF322228[j], AB031047, AF306345, AF274239[k], AF424632[l]	29	[15], [16], [44], Cheuk et al., unpubl.
AtPLDε	PLDalpha	At1g55180 / F7A10.25	4	AC027034, NC_003070	-	1[m],	-
AtPLDζ		At5g25370 / F18G18.110	4	AC006258[n], NC_003076	-	0	-
AtPLDp1	PLDzeta1	At3g16790 (5' part) At3g16785 (3' part) / MGL6.24 (5' part) MGL6.27 (3' part)	20	AB022217 (5' part) AB028608 (3' part), NC_003074	AF411833	7	Qin and Wang, unpubl.
AtPLDp2		At3g05630 / F18C1.10	16[o]	AC011620, NC_003074	-	3	-

[a]For clarity, gene names used in [2] differing from our nomenclature are noted. Inappropriate designation of *AtPLDε* as "PLDalpha" has been used by AGI annotators. *AtPLDp1* was designated by Qin and Wang in the respective GenBank entry. [b]AGI gene identifier (the number following "At" refers to chromosome)/BAC-specific locus number [c]GenBank/EMBL/DDBJ accession numbers of genomic clones containing the respective PLD genes; multiple numbers noted for several genes refer to redundant clones or clone assemblages. [d]A 5' truncated cDNA sequence, the ORF incomplete. [e]Corrections introduced into the AGI annotation: the 1st exon extended with 765 bp in 5' direction (see the text), the 6th exon shortened of 16 bp at the 3' end (according to the respective cDNA and ESTs). [f]Corrections introduced into the AGI annotation: the 1st predicted intron included into the ORF and the 5th exon extended with 36 bp at the 3' end to restore conserved regions to restore conserved regions. [g]An obscure intron may be spliced off (perhaps alternatively) from the 1st predicted exon (see the text), [h]The same cDNA is noted here for both the *AtPLDγ2* and *AtPLDγ3* genes, because it is chimeric and contains portions from both the genes (see the text). [i]Corrections introduced into the AGI annotation: the last six predicted exons removed (according to the corresponding cDNAs and ESTs). [j]The independently reported *AtPLDδ* cDNAs perhaps represent two distinct splice variants (details in the text). [k]A 5' truncated cDNA sequence. [l]A one-nucleotide deletion at the very 5' end resulting in a rearranged N-terminus of the deduced protein, [m]An alleged EST (AV547754) corresponding to the PLDε gene is dubious, as it apparently represents an unspliced transcript or a contamination of the source cDNA library with a genomic fragment. [n]The gene is currently not annotated in the GenBank entry of BAC F18G18 (AC006258), but it has been included into annotation of the whole-chromosome pseudomolecule (NC_003076). [o]Corrections introduced into AGI annotation: the 8th exon extended with 21 bp at the 5' end to restore a highly conserved region.

Members of the PXPH-PLD subfamily typically bear different two phospholipid-binding domains, the PX (phox) and the PH (pleckstrin-homology) domain, in the N-terminal region (Fig. 2). PXPH-PLDs have previously been known only from animals and fungi, so identification of two Arabidopsis genes belonging to this subfamily adds an important new dimension to the picture of plant PLD. We propose terming these genes *AtPLDp1* and *AtPLDp2* (the "p" from PX and PH) to underline the principal difference between the C2-PLD branch (Greek letters) and the PXPH-PLD subfamily. Very recently, a cDNA has appeared in GenBank corresponding to *AtPLDp1* gene but unfortunately annotated as PLD zeta1 (Table 1). As we believe that our nomenclature better reflects structural and phylogenetic aspects of the plant PLD family, in the rest of this text, the gene names proposed by us are used.

Functionality of those Arabidopsis PLD genes, for which full-length cDNAs have been cloned, is undisputed. Moreover, proteins encoded by three of these genes, PLDβ1,
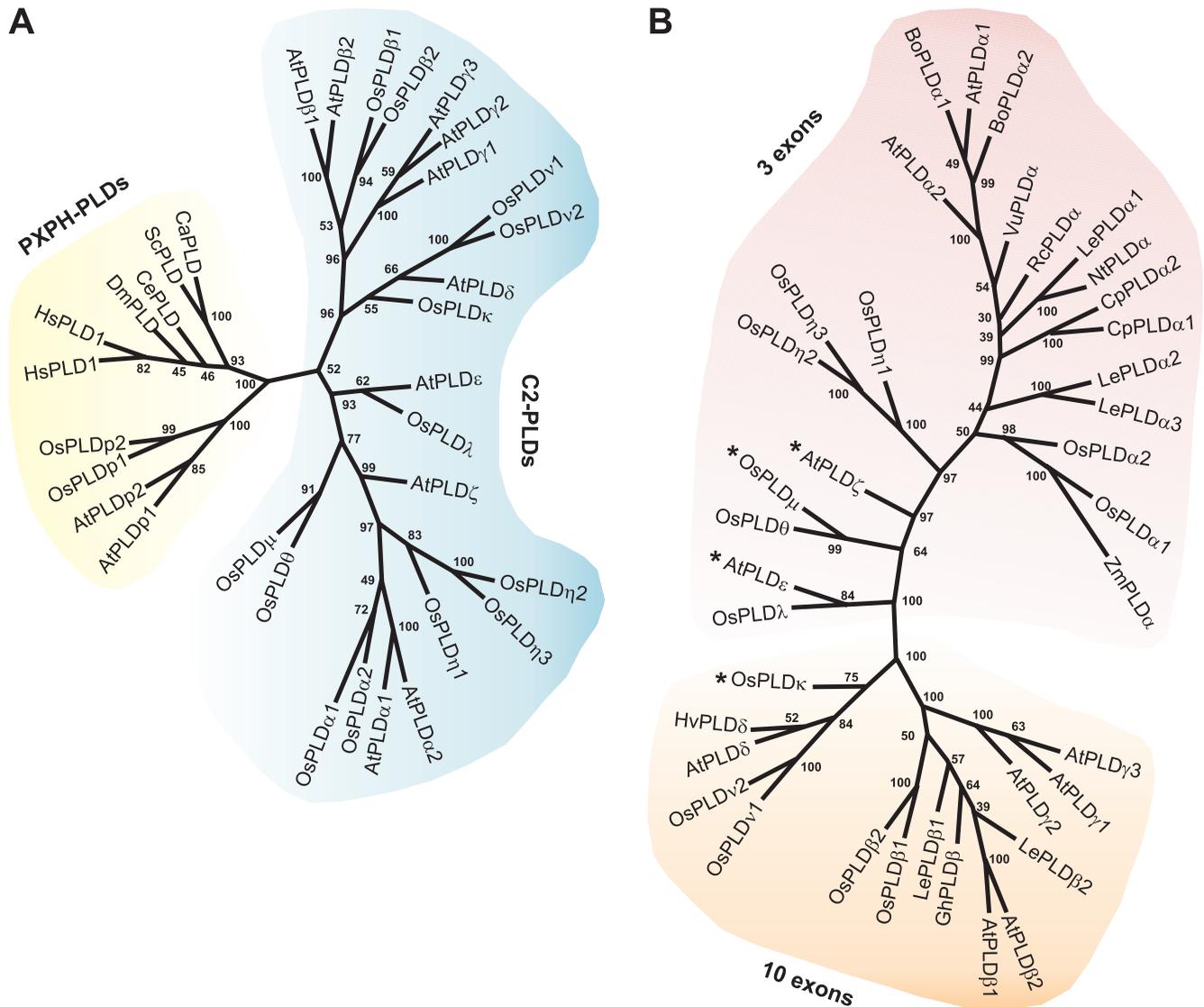
Figure 1
Phylogenetic analysis of the PC-PLD family. A, an unrooted tree of Arabidopsis, rice and selected non-plant PLDs, constructed by the neighbor-joining method on the basis of a MACAW-generated alignment (see Materials and Methods). C2-PLD and PXPH-PLD subfamilies are indicated by yellow and blue background, respectively. Rooting the tree with a bacterial PLD (not shown) revealed that the two subfamilies are monophyletic. B, more detailed phylogeny of C2-PLDs, based on a manually edited ClustalW alignment (tree constructed as above). Note that the topology corresponds to the previous tree, with the exception of the relationships within the cluster of PLDγ. Two major monophyletic subgroups indicated by different backgrounds appear to differ in the exon-intron organisation (as inferred from data from Arabidopsis, rice, cabbage and castor bean – only cDNA sequences are available for the remaining PLDs). The number of introns in the genes marked by asterisks (*) differs secondarily from the basic plans. AtPLDε, AtPLDζ and possibly OsPLDμ have independently acquired an additional exon (see Fig. 3), and OsPLDκ has lost 4 introns corresponding to the 3rd, 7th, 8th and 9th introns of beta, gamma, delta and nu PLDs. Numbers next to the nodes are percentages of bootstrap confidence levels calculated from 500 replicates. Trees with highly congruent topology were also obtained by a maximum parsimony method. Species abbreviations: At, *Arabidopsis thaliana;* Bo, *Brassica oleracea;* Ca, *Candida albicans;* Ce, *Caenorhabditis elegans;* Cp, *Craterostigma plantagineum;* Dm, *Drosophila melanogaster;* Gh, *Gossypium hirsutum;* Hs, *Homo sapiens;* Hv, *Hordeum vulgare;* Le, *Lycopersicon esculentum;* Nt, *Nicotiana tabacum;* Os, *Oryza sativa;* Sc, *Saccharomyces cerevisiae;* Rc, *Ricinus communis;* Vu, *Vigna unguiculata;* Zm, *Zea mays.* See Additional files for accession numbers and further sequence information.
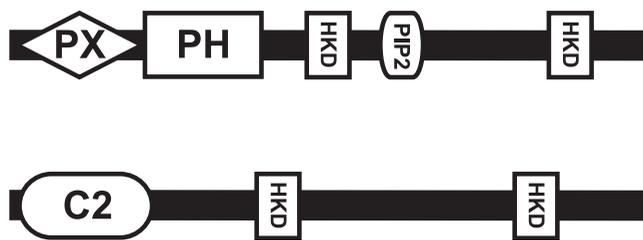
Figure 2
Schematic diagram of conserved motifs and domains in the two major subfamilies of PC-PLD family. PXPH-PLDs (at the top) and C2-PLDs (at the bottom) differ principally in their N-terminal regions by presence of distinct phospholipid-binding domains, i.e. the PX and PH domains in the case of the PXPH-PLD subfamily, and the C2 domain in the case of the C2-PLD subfamily. Note that the support for presence of the PX domain in plant PXPH-PLDs was slighly below the default cutoff when searched with the Search Pfam tool. The same pays true for the PX domain in AtPLDp1 identified by SMART (see Materials and Methods). HKD boxes refer to catalytic motifs forming together single catalytic site in each protein. The PIP2 box in the PXPH-PLDs relates to a PIP2-binding motif, see the text and Fig. 4B.

PLDγ1 and PLDδ, have been characterised biochemically (see [2,16]). Expression of several other isoforms is documented by ESTs in GenBank, but there are currently no ESTs available cognate for *AtPLDα2*, *AtPLDβ2* and *AtPLDζ* genes (Table 1). However, absence of cognate ESTs is not exceptional, since in general only about 60% of predicted Arabidopsis genes are recorded in available EST collections [12]. It is therefore very likely that expression of most of the genes without ESTs is very low or limited only to some special developmental stages or conditions.

Exon-intron organisation of Arabidopsis PLD genes
Limitations of theoretical prediction of exon-intron structures are well known and cDNA sequencing is often necessary for building accurate gene models. This proves true also for many of the Arabidopsis PLDs. Unfortunately, four reported cDNAs, i.e. *AtPLDα1*, *AtPLDβ1*, *AtPLDγ1* and *AtPLDγ2* (Table 1), contain mismatches compared to the highly accurate genomic sequences (reported to contain less than 1 error per 10^4-10^5 bp; [12]). While some of the discrepancies may represent a natural polymorphism, others, particularly those associated with frame shifts, are most likely due to sequencing errors or cloning artefacts. This suspicion is also supported by available EST sequences, which nearly always match the genomic sequences, not the cDNAs.

For example, within the coding portion of the *AtPLDα1* cDNA there are four regions with the reading frame shifted relative to the genomic sequence. As a result, the protein sequence derived from the cDNA (AAC49274.1) is

highly divergent from other PLDs in these four regions, while that predicted from the genome data (NP_188194.1) matches well the PLD consensus. We found similar discrepancies also for *AtPLDβ1*, *AtPLDγ1* and *AtPLDγ2* cDNAs. Moreover, published *AtPLDγ1* and *AtPLDγ2* cDNAs appear to be chimeric, perhaps due to cloning artefacts. The last ~180 nucleotides of *AtPLDγ1* cDNA apparently originate from a gene encoding a pseudo-response regulator (AB046955, chromosome 5). Similarly, the 3' third of the cDNA reported as *AtPLDγ2*[18] is actually derived from the *AtPLDγ3* gene. We therefore believe that the cDNA sequences have to be interpreted very cautiously, and we base our conclusions mainly on the genome project data. In several cases, however, we proposed corrections of the AGI annotation of PLD genes. Details and refined coding sequences can be found in the Additional files, most important aspects are also discussed below.

Despite sequencing errors, the *AtPLDα1* cDNA is in good agreement with the previously suggested gene structure, the gene contains three coding exons and a 5' non-coding one (Fig. 3), similarly to other characterised PLD genes from the alpha subgroup [2]. Coding portion of the *AtPLDα2* gene appears to be arranged in the same way, but whether there is also a non-coding exon in the 5' UTR that could be proven only by cloning of the respective cDNA. Two other predicted Arabidopsis genes, *AtPLDε* and *AtPLDζ*, appear also to exhibit structures related to alpha-type PLDs, but differ by the presence of an additional intron. This intron resides at different positions in both genes, suggesting that it has been acquired independently by *AtPLDε* and *AtPLDζ*.

AtPLDβ1 gene was found to consist of 10 exons [2]. Current database annotation should, however, be corrected in some points (see Table 1 and Additional file for details). Most importantly, there is a long region devoid of in-frame STOP-codons upstream from the first predicted exon, and the ORF could be thus extended in the 5' direction (Fig. 3). This would add an unusual, ~250 aa long N-terminal projection to the PLDβ1 protein. However, the potential initiation codon upstream of the predicted one is just out of the region covered by the respective cDNA, suggesting that the cDNA might be 5' truncated. There are several reasons arguing for inclusion of the N-terminal extension. First, such a long sequence would have most likely accumulated STOP-codons if not translated. Second, we found such an N-terminal extension (though loosely conserved) in other PLDs of the beta group, including also Arabidopsis PLDβ2, tomato PLDβ2 ([19]; AY013256) and two PLDβ isoforms from rice (see Table 2). Third, a cotton cDNA coding for a PLDβ (AF159139) contains a very long 5' leader sequence (more than 1300 bp) harbouring a potential ORF coding for a peptide ~300
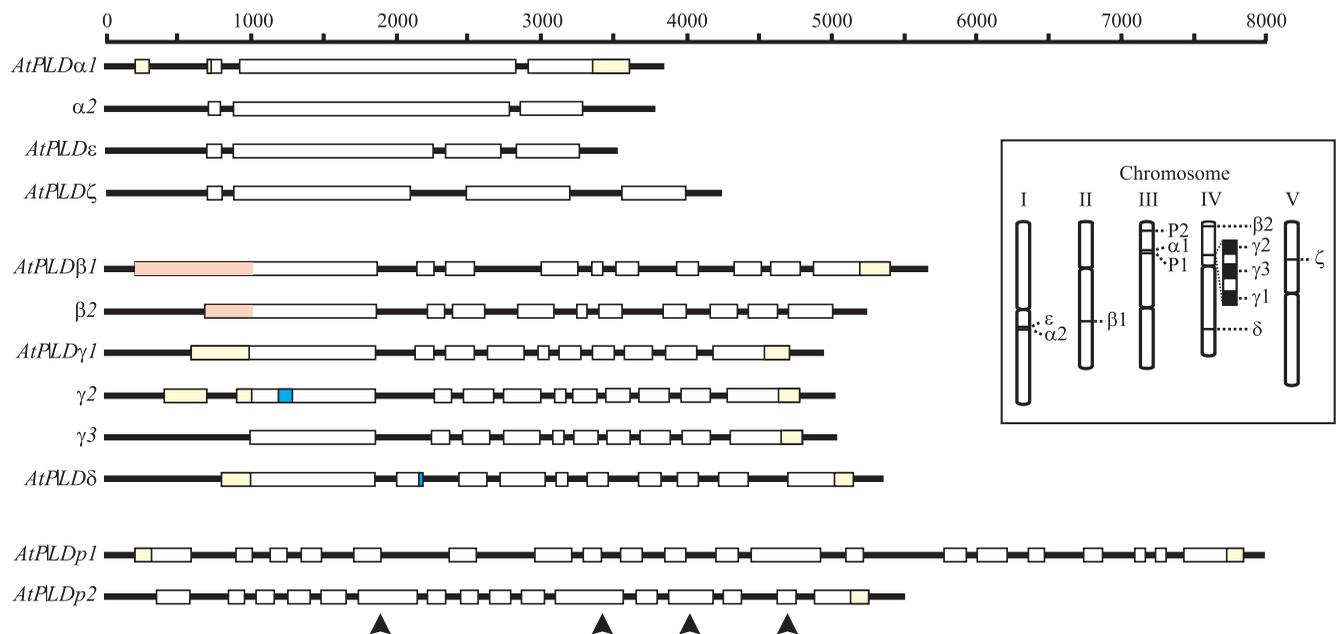
Figure 3
Exon-intron architecture and chromosomal location of PLD genes in Arabidopsis. White boxes represent coding sequences. Yellow boxes delimit un-translated regions revealed by cDNA and/or EST sequences. Blue areas in *AtPLDγ2* and *AtPLDδ* indicate potential alternative splicing (see the text). Orange regions in *AtPLDβ1* and *AtPLDβ2* indicate the portion of the first exon coding for the unusual N-terminal extension (details in the text). Note the three basic types of gene organisation. Black arrowheads indicate positions of introns presumably lost from the *AtPLDp2* gene (compare to *AtPLDp1*). The inset shows locations of the PLD genes on the Arabidopsis chromosomes.

aa long and similar to the N-terminal extensions of both beta PLDs from Arabidopsis. The 3' end of the leader ORF overlaps with the beginning of the PLD ORF but resides in a different reading frame. It is tempting to speculate that the frame shift is due to a cloning or sequencing error and that the cotton PLDβ possesses an N-terminal extension similar to other beta PLDs. The extensions in beta PLDs generally do not appear to resemble a sorting signal for any cellular compartment, so we suggest that they might e.g. mediate interaction with regulatory factors.

The *AtPLDβ2* gene was originally described as PLDδ1, and 11 exons predicted by AGI were proposed as a unique feature [2]. However, this hypothetical gene structure was not supported by the results of gene-finding programmes that we employed (see Materials and Methods). Inclusion of the 1st originally predicted intron into the ORF, which is supported by the programmes, introduces a conserved portion of the C2 domain and adjusts the splicing pattern to the 10-exonic scheme exhibited by several other PLD genes. The resulting predicted protein sequence belongs clearly to a beta PLD type (Fig. 1A,B).

The three very similar paralogs of PLDγ reside in a tandem triplication (arranged *AtPLDγ1 – AtPLDγ3 – AtPLDγ2*) on

the Arabidopsis chromosome 4 [2], indicating a relatively recent origin of the triplet. The predicted gene structure of all three genes fits the 10-exonic scheme typical for some other PLD types. However, there appears to be a probable obscure intron of 96 nucleotides in the *AtPLDγ2* gene delimited by GT-AG borders and supported by a matching cDNA sequence (see Fig. 3). Exclusion of this intron (possibly as a result of alternative splicing) deletes a region from within the C2 domain that might substantially affect function and/or regulation of AtPLDγ2. A similar potential intron may be present also in *AtPLDγ3* gene. The *AtPLDγ2* gene further contains an additional non-coding exon at the 5' end, but it is currently unknown whether the 5' UTR of the other two genes is organised in the same way (the available *AtPLDγ1* cDNA may be 5' truncated).

The *AtPLDδ* gene had been predicted by AGI annotators as consisting of 16 exons, but, as revealed by EST sequences and cDNAs, *AtPLDδ* possesses only 10 conserved exons shared with beta and gamma PLDs. Interestingly, there is an evidence for alternative splicing of the *AtPLDδ* gene, because one of the independently cloned cDNAs (AB031047, [15]) differs from the others by extension of the second exon at the 3' boundary by 33 nucleotides (Table 1, Fig. 3). Interestingly, the apparently more abundant

A

```
            xx  x xx    * *    *                      xx x   xx
AtPLDα1   YMRAQEARRFMIYVHTKMMIVDDEYIIIGSANIN 680      CIEKVNRISDKYWDFY 754
AtPLDα2   YIRAQEARRFMIYVHTKMMIVDDEYIIIGSANIN 680      CIQKVNRVADKYWDLY 754
AtPLDβ1   QALSRKSRRFMVYVHSKGMVVDDEYVVIGSANIN 953      CVRKVRTMGERNWKQF 1029
AtPLDβ2   QGLCRKSRRFMVYVHSKGMVVDDEYVVIGSANIN 797      CVRKVRTVAEENWEQF 873
AtPLDγ1   QVQALKSRRFMIYVHSKGMVVDDEFVLIGSANIN 728      CVRRVRQLSELNWRQY 804
AtPLDγ2   QVQALKSRRFMIYVHSKGMVVDDEFVLIGSANIN 726      CVRRVRQLSELNWGQY 802
AtPLDγ3   QVQALKSRRFMIYVHSKGMVVDDEFVLIGSANIN 736      CVRRVRQLSELNWRQY 812
AtPLDδ    VSDSYNFQRFMIYVHAKGMVVDDEYVLMGSANIN 726      CLKKVNTISEENWKRF 802
AtPLDε    YARAQESRRFMIYVHSKGMMIVDDEYIIIGSANIN 686     CIRMVNATADELWGLY 762
AtPLDζ    YWNAQRNRRFMVYVHSKLMIVDDTYILIGSANIN 634      CVRGLRTIGEQMWEIY 707
OsPLDα1   YSRAQEARRFMIYVHTKMMIVDDEYIIIGSANIN 682      CVQKVNRIAEKYWDMY 756
OsPLDα2   YFHAQQNRRFMIYVHTKMMIVDDEYIIVGSANIN 693      CVRKVNAMADRCWDLY 768
OsPLDβ1   REQARKHRRFMVYVHSKGMVVDDEYVIIGSANIN 916      CMRRVRQIGEQNWERF 992
OsPLDβ2   QDQAKKNRRFMVYVHSKGMVVDDEYVIIGSANIN 774      CTRQVRHIGEQNWRQF 850
OsPLDη1   YAKAQKARRFMIYVHSKMMIVDDEYIIVGSANIN 688      CVQRVNKMADKYWDLY 763
OsPLDη2   YAKAQKTRRFMIYVHSKMMIVDDEYIIVGSANIN 698      CVRRVNEMANKHWELY 774
OsPLDη3   YAKAQNARRFMIYVHSKMMIVDDEYIIVGSANIN 708      CVRRVNAMADRHWQLY 783
OsPLDθ    YMRAQQARRFKINVNAIMIVDDEYIIVGSANVN 672       CMSRVNQAARQHWDMY 755
OsPLDκ    ASSARRHRRFMIYVHSKGMVVDDEYVIVGSANIN 781      CVKRVNEIAAENWRRY 856
OsPLDλ    YWRAQVNRRFPIYVHAKLMIVDDEYVMVGSANLN 682      CVRAVRRAAEEATWDAY 762
OsPLDμ    YMRAQNARRFMIYVHAKTMIVDDEYIIVGSANIN 699      SVRLVNQAARRHWDAF 779
OsPLDν1   LRSAQKLRRFMIYVHSKGMVVDDEYVIIGSANIN 713      CVRRVREMAEENWRAY 791
OsPLDν2   QRLVQKFKRFMIYVHSKGMVVDDEYVLIGSANIN ?        CVRQVNEMAEENWARY ?
```

B

```
                                        #  #   #
HsPLD1    PRMPWHDIASAVHGKAARDVARHFIQRW 701
HsPLD2    PRMPWRDVGVVVHGLPARDLARHFIQRW 563
ScPLD     PRMPWHDVQMMTLGEPARDLARHFVQRW 895
CePLD     PRMPWHDIHSVTFGAPARDLARHFIQRW 1046
DmPLD     PRMPWHDVGLCVVGTSARDVARHFIQRW 890
AtPLDp1   PRMPWHDVHCALWGPPCRDVARHFVQRW 577
AtPLDp2   PRMPWHDVHCALWGPPCRDVARHFVQRW 572
OsPLDp1   PRMPWHDVQCALYGPACRDIARHFVQRW 701
OsPLDp2   PRMPWHDVQCALYGPPCRDVARHFVQRW 664
AtPLDα1   PREPWHDIHSRLEGPIAWDVMYNFEQRW 428
AtPLDα2   PREPWHDIHCRLEGPIAWDVLYNFEQRW 427
AtPLDβ1   PREPWHDLHSKIDGPAAYDVLTNFEERW 690
AtPLDβ2   PREPWHDLHSKIDGPAAYDVLTNFEERW 534
AtPLDγ1   PREPWHDLHSKIDGPAAYDVLANFEERW 459
AtPLDγ2   PREPWHDLHSKIDGPAAYDVLANFEERW 457
AtPLDγ3   PREPWHDLHSKIDGPAAYDVLANFEERW 466
AtPLDδ    PRQPWHDLHCRIDGPAAYDVLINFEQRW 451
AtPLDε    PREPWHDCHVSVVGGAAWDVLKNFEQRW 399
AtPLDζ    PREPWHDIHCKLDGPAAWDVLYNFEQRW 433
```

Figure 4
The second catalytic HKD motif and putative PIP2-binding sites in Arabidopsis and rice PLDs. A, multiple alignment of the second catalytic motif and adjacent regions harbouring alleged sites for PIP$_2$ binding [27]. Conserved amino acid residues are indicated by shading, asterisks denote the catalytic triad. Mutated residues of the catalytic motif in OsPLDθ are on the red background. Positions of conserved residues in the postulated PIP2-binding motifs are indicated by (x), basic residues are in bold. B, multiple alignment of a box conserved among all PLDs and known to bind PIP$_2$ in mammalian and yeast PXPH-PLDs. Three arginine residues involved in PIP$_2$ binding [31] are absolutely conserved in all PXPH-PLDs, but are not present in C2-PLDs. Motif positions are indicated by (#), arginine residues are in bold. The numbers at the right of individual sequences in both alignments refer to the position of the last residue within the whole protein, the question mark (?) indicate that complete protein sequence is not available. Species abbreviations are the same as used in the Fig. 1. See Additional files for accession numbers and other sequence information.

---

shorter alternative (as defined also by three ESTs covering the respective region) uses an unconventional 5'GC intron boundary instead of GT (see [http://www.Arabidopsis.org/splice_site_excep.html]). The longer and the shorter putative splice variants have recently been denoted PLDδa and PLDδb, respectively [16]. The PLDδa protein bears an insertion of 11 aa in an otherwise relatively conserved region, which could have profound functional consequences.

The two genes classified into the PXPH-PLD subfamily appear to exhibit the most complex exon-intron structure of all Arabidopsis PLDs. A corresponding full-length cDNA has been reported only for *AtPLDp1* (Table 1), so the prediction of the *AtPLDp2* gene remains tentative. A minor correction should perhaps be introduced into the current database prediction of the *AtPLDp2* gene to restore a highly conserved region (see Table 1 and the Additional files). Despite a difference in the number of exons (20 and 16, respectively), the structures of *AtPLDp1* and *AtPLDp2* genes are clearly related, as the difference is due to 4 in-

trons probably lost from *AtPLDp2* (20-exonic structure seems to be primordial in the plant PXPH-PLD subfamily, see the rice homologs below and Fig. 1).

PLDs in rice: an alternative view

Arabidopsis is presently the only plant for which the complete PLD set can be catalogued. Nonetheless, other species are emerging as important models for genome-wide studies. Rice genome sequencing is highly advanced, with a substantial portion (more than 230 Mbp up to now, see [http://rgp.dna.affrc.go.jp] for updates) already sequenced by the International Rice Genome Sequencing Project (IRGSP). Even greater portion of the genome (about 250 Mbp) has been sequenced by the Monsanto company, who have made their sequences publicly available [13]. With redundancy between the two data resources taken into account, we could analyse about three quarters of the whole rice genome.

We identified at least 16 complete or partial sequences of putative rice PLD genes (Table 2). Five of them have been
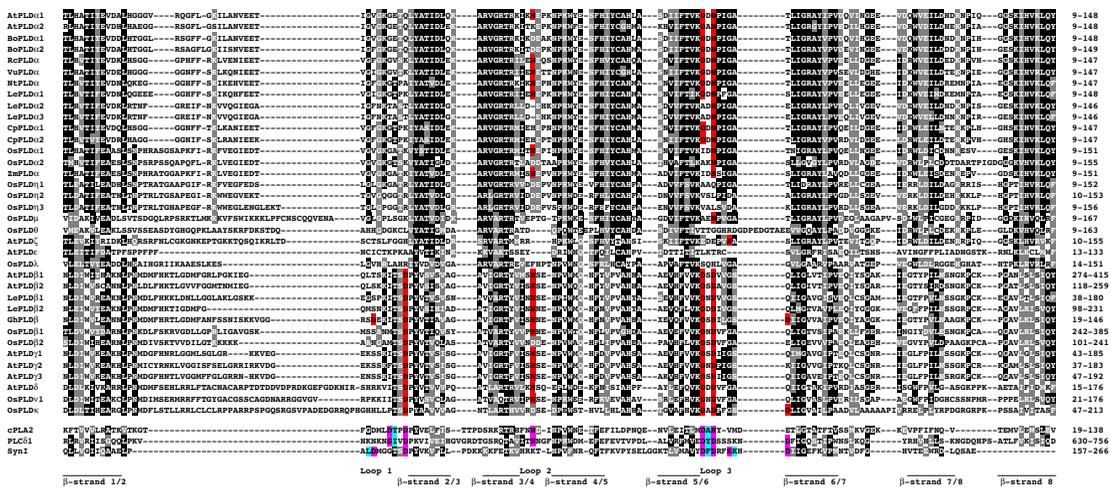
Figure 5
Multiple alignment of C2 domains from C2-PLDs. For comparison, three characterised C2 domains (from cytosolic phospholipase A_2, phospholipase Cδ and Synaptotagmin I) are included (adopted from [26]). The domain consists of eight β-strands (here indicated by lines at the bottom of the alignment) linked by loop regions. Two basic topological variants of the C2 domain have been described, resulting from a circular permutation of the β-strands (details in [26]). The Topology I is exemplified here by the C2 domain from Synaptotagmin I, whereas cPLA$_2$ and PLCδ1 exhibit the Topology II. C2 domains from PLDs are predicted to have the Topology II [45]. The first number at each β-strand refers to the Topology II, the second number to the Topology I (the first β-strand of SynI in not shown here). Three loops containing Ca^{2+}-coordinating ligands are indicated as Loop 1 through Loop 3. Black and grey background indicates more and less conserved positions, respectively. Residues, which bind Ca^{2+} by the side chains, are highlighted by a violet background. Other ligands for Ca^{2+} ions are provided by backbone carbonyls (the respective positions with a blue background). Only three Ca^{2+}-binding positions (excluding backbone carbonyls) are shared by the characterised C2 domains, the first two occupied by aspartate residues, the third either by an aspartate or an asparagine residue. Other Ca^{2+} ligands are recruited from generally non-shared positions in distinct domains. Potential Ca^{2+}-binding residues in the C2 domains from PLDs are shown on a red background (non-conserved residues contributing with the backbone carbonyls are not considered). No C2 domain from any PLD matches exactly any prototypic C2 domain with respect to the Ca^{2+}-binding sites. See the text for details.

cloned individually, 13 genes or their portions have been already sequenced by the IRGSP, sequences coming from 13 genes could be found in the Monsanto genome draft and fragments of at least one PLD gene are available only as EST or GSS sequences. Since a systematic nomenclature of rice PLD genes has not been established, we propose a terminology that would reflect phylogenetic and structural relationship with PLDs from other species (Table 2, Fig. 1A,B). Several rice PLDs appear to be orthologous to Arabidopsis genes, including two alpha-type PLDs and two beta-like PLDs. At least one presumed ortholog of *AtPLDδ* is available only as GSS and EST sequences too fragmentary to be included directly into phylogenetic analysis. Their assignment to the delta type is however supported by an analysis employing a higly similar barley PLD assembled from EST sequences and used as a placeholder (Fig. 1B). Two genes from the PXPH-PLD subfamily were

also found and denoted *OsPLDρ1* and *OsPLDρ2*. Remaining genes cannot be assigned to any of the classes established for Arabidopsis. Although *OsPLDλ*, tends to cluster together with *AtPLDε*, the two proteins share only about 42% identical amino acids and the genes appear to differ in the number of introns, supporting classification of *OsPLDλ* as a novel subtype. Similarly, according to the phylogenetic analysis *OsPLDθ* and *OsPLDμ* appear to be related, but different intron numbers and the second catalytic motif missing from OsPLDθ (see below) justify a separate classification.

Only the five individually cloned PLD genes have been annotated. Complete cDNA has been reported for *OsPLDα1*[20]; the gene has the exon-intron structure closely related to other alpha-type PLDs [2,21]. Delimitation of the coding region of the *OsPLDη1* gene has also

Table 2: Summary on PLD genes in rice. All PLD genes that could be found in public sequence databases and in the Monsanto database of the rice genome draft are listed. Sequences of PLD genes found in the Monsanto database but not yet sequenced by the IRGSP (i.e. *OsPLDβ2*, *OsPLDv2*, and *OsPLDλ*) have been deposited into GenBank. All sequences are derived from the Nipponbare cultivar, unless otherwise stated.

proposed gene name	original gene name	chromosome[a]	number of coding exons[b]	GenBank/EMBL/ DDBJ accession numbers[c]	Monsanto contigs[c]	number of ESTs	Reference
OsPLDα1	PLD1	1	3	AB001920[d], AP003215 (5' end), AP003282 (3' end), D73411[d] (cDNA)	-	18	[20,21]
OsPLDα2		5	3?[e]	AC087553	OSM15574 (5' end), OSM15576, OSM15582	2	-
OsPLDβ1		10	10	AC078894	OSM15748+ OSM15750	3	-
OsPLDβ2		?	10	AF411221	OSM12022	0	-
OsPLDδ		?	?	nbxb0077G10 (GSS sequence), BE040461 (EST), AU058024 (EST)[f]	-	2?	-
OsPLDη1	PLD2	6	3[g]	AB001919[d], AP003629,	OSM12111	2	[21]
OsPLDη2	RPLD3	6	3	AF271356[h], AP003629	OSM12115	0	McGee et al., unpubl.
OsPLDη3	RPLD4	6	3	AF271357[h], AP003629	OSM12115	0	McGee et al., unpubl.
OsPLDθ		8	3	AP004659	OSM12593	0	-
OsPLDκ		2	6	AP004078	OSM128517 +OSM128518 + OSM128519	0	-
OsPLDλ		?	3	AF411223	OSM15282	3	-
OsPLDμ		3	4	AC099323	-	0	-
OsPLDv1	RPLD5	3	10[i]	AF271358[h], AC096856	OSM1620	1	McGee et al., unpubl.
OsPLDv2		7	10?[j]	AF414565 +AF414566, AP003756(3' end)	OSM11825+ OSM11834	2	-
OsPLDρ1		5	20	AC 104272	OSM151594	1	-
OsPLDρ2		1	20	AP003934	OSM15141+ OSM15138	4	-

[a]Chromosome locations are not known for genes originally found only in the Monsanto rice genome database. [b]The exon numbers with the question mark (?) should be considered preliminary, since the corresponding gene sequences are available incomplete only. [c]In the case of several genes sequences had to be assembled from two or three contigs (contig numbers linked with "+"), the sequences are therefore incomplete and contain gaps (see Additional files). [d]Sequence derived from the Koshihikari cultivar. [e]A portion from within the 1st supposed intron has not yet been sequenced. [f]The three sequences may potentially represent two distinct PLDδ isoforms, since the last one does not overlap with the former two. [g]Erroneously reported to have two coding exons [2]. [h]Sequence derived from the Indica variety, cultivar IR54. [i]Proposed annotation corrected: the 3rd exon shortened by 12 nucleotides at the 3' end according to EST AA751500. [j]Available sequence incomplete, the very beginning of the 1st exon truncated, the 3' end of the 4th exon, a supposed intron and the 5' end of the 5th exon lacking.

been verified experimentally and found to have a similar organisation [21]. Structures of the other genes could be predicted only theoretically, but comparison with EST sequences and other PLD genes proved helpful, as exon-intron junctions appear to be highly conserved within individual subgroups of the PLD family (see below; predicted or corrected coding sequences available in Additional files). Thus, we introduced a minor correction into the previously proposed *OsPLDv1* gene structure (see Table 2 and Additional files). Annotated *OsPLDη2* and *OsPLDη3* appear to have a similar splicing pattern as *OsPLDη1*. Interestingly, we found the three *OsPLDη* genes residing in the genome adjacently in a series *OsPLDη2-OsPLDη3-OsPLDη1*, but, in contrast to the *AtPLDγ* cluster

in Arabidopsis, *OsPLDη2* is inverted with respect to the remaining two genes. Exon-intron structures proposed by us for other rice PLD genes reflect phylogenetic affinity to Arabidopsis orthologs (compare Table 1 and Table 2). The novel *OsPLDλ*, and *OsPLDθ* genes probably have 3 coding exons with introns occupying conserved positions shared with the *PLDα* and *PLDη* prototypes. The novel *OsPLDμ* gene also resembles *PLDα* and *PLDη*, although comparison with a highly similar barley EST revealed 4 coding exons. The second exon is very short and encodes a part of the first non-conserved loop of the C2 domain (Fig. 5). This exon might be completely novel or, more probably, has been insulated from the downstream coding region by acquisition of an intron. Gene structure of *OsPLDκ*, com-

prising six coding exons, appears to be derived from the 10-exonic scheme, retained by related PLDs (i.e. β, γ, δ and ν), by loss of four introns. Two identified rice PXPH-PLDs are very similar to each other including exon-intron organisation, which is obviously shared with *AtPLDp1* from Arabidopsis.

In summary, comparison of Arabidopsis and rice PLD genes revealed that they exhibit generally non-conserved exon-intron structures (Table 1, Table 2; Fig. 3), and positions of introns do not reflect boundaries between functional domains. However, three clusters of plant PLD genes can be recognised, which differ completely in exon-intron organisation from each other, but the organisation appear conserved among genes within each cluster. Independent acquisition of introns seems to be a plausible explanation (Fig. 1A,B; Fig. 3). Described plant PXPH-PLD genes appear to have primarily 20 conserved exons, with the exception of *AtPLDp2* lacking four introns. The C2-PLD subfamily comprises a clade characterised primarily with 3 coding exons, and a group of originally 10-exonic genes. Independent acquisition of additional introns (in *AtPLDε*, *AtPLDζ* and probably also *OsPLDμ*) or intron losses (suggested for *OsPLDκ* and *AtPLDp2*) may be common during the evolution of plant PLDs.

PLD diversity and expression as recorded in the EST collections

Databases of expressed sequence tags (ESTs) are available for a number of plant species and represent invaluable resource for both functional and evolutionary studies, providing information on both genetic diversity and expression profiles. To assess these aspects of the angiosperm PLD family, we identified a number of PLD-derived ESTs from Arabidopsis, rice, tomato, *Medicago truncatula* and *Sorghum bicolor* (see Table 1 and Table 2, and Additional files).

For exploration of PLD diversity beyond Arabidopsis and rice, tomato is a suitable starting point, with more than 140,000 ESTs available and five full-length PLD cDNAs cloned representing three PLDα and two PLDβ genes [19,22,23]. With the exception of *LePLDα2*, all cloned tomato PLDs are recorded among ESTs, but expression of additional isoforms is documented, too, including a PLD similar to alpha types, at least two putative delta isoforms, a PLD most similar to *AtPLDε* and at least one gene from the PXPH-PLD group. The second species analysed was *Medicago truncatula* with more than 137,000 ESTs in Gen-Bank. No PLD has yet been reported from this plant, but ESTs again indicate the presence of a complex PLD family, comprising at least two indisputable PLDα homologs highly similar to each other, at least two additional genes less similar to alpha types, potentially three PLDβ isoforms, at least one delta ortholog, a PLD most similar to *At-*

PLDε and two members of the PXPH-PLD subfamily. As a monocotyledonous model for EST analysis we chose *Sorghum bicolor*, for which more that 84,000 ESTs had been sequenced. Multiple homologs could again be found among the ESTs, including at least two obvious alpha PLDs, a gene related to the rice PLDη1, one PLDβ and a PLD most similar to the rice PLDμ.

In summary, our EST analysis revealed that PLD types identified in Arabidopsis and rice are widespread in angiosperms, but there might be additional types not yet characterised. With the help of EST clones, full-length genes/cDNAs can be easily isolated and characterised, so deeper insight into PLD diversity in plants can become soon available (the list of ESTs analysed is available in Additional files).

The relative abundance of ESTs can provide information on expression of individual genes [24]. Unfortunately, for most PLD genes there are too few ESTs for statistically significant estimation of their expression in specific tissues, developmental stages or conditions, and only general level of expression can be inferred. According to the total number of cognate ESTs in the GenBank, the most highly expressed PLD gene in Arabidopsis is *AtPLDα1* (40 EST entries) followed with *AtPLDδ* (29 entries), while other genes seem to be expressed at a considerably lower level or not recorded at all (see Table 1). In rice, expression of *OsPLDα1* predominates to a similar extent as in Arabidopsis (~40% of all ESTs from PLD genes), and expression of the other genes is markedly lower as well (Table 2). The EST collection from *Medicago* provides even more strongly substantiated evidence for an expression bias with 63% (of 74 PLD-derived ESTs in total) matching one of multiple PLDα paralogs.

Among tomato ESTs that can be assigned to the cloned cDNAs, 14 come from *LePLDα1*, 2 from *LePLDα3*, 4 from *LePLDβ1* and 1 from *LePLDβ2*, suggesting that expression of *LePLDα1* might again be prevalent. However, *LePLDα2* and *LePLDα3* have higher expression levels than *LePLDα1* when measured by Northern blots [19], so caution must be paid when few ESTs are used for conclusions on expression profiles. In *Sorghum* one of two alpha-type genes also accounts for ~40% of PLD-derived ESTs, but, in contrast to previous collections, a PLDβ-like isoform appear to be sampled to a similar extent. Interestingly, all the EST corresponding to the latter gene are derived from a cDNA library prepared from a pathogen-infected plants. It is tempting to speculate that the expression of this PLDβ gene might be induced by a pathogen-derived signal, similarly to *LePLDβ1* reported to be induced upon treatment with an elicitor xylanase [19].

Predominant expression of alpha-type PLDs inferred from our EST analysis fits with biochemical experience, since the enzymatic activity usually ascribed to alpha-type PLDs is much more abundant in plant tissues compared to the activity of the beta and gamma types [2]. Interestingly, in most plants studied (except for tomato), two PLDα genes could be found, but only one of them was highly expressed (see Table 1, Table 2 and Additional files). Similarly, two *PLDα* paralogs have been cloned from the resurrection plant *Craterostigma plantagineum*, one of them expressed constitutively and the other one induced only upon desiccation stress [25]. Differential expression mode for two very similar PLDs has been observed also in tomato, where the elicitor xylanase stimulated expression of *LePLDβ1* but not of *LePLDβ2*[19]. Henceforth, if the differences in expression did relate to differences in physiological function, it could be concluded that there is only little functional redundancy within plant PLD family, even among highly similar isoforms.

Functional aspects of the primary structure of plant PLDs
As already noted, all known eukaryotic PC-PLDs belong to two subfamilies differing in their N-terminal portion (Fig. 2), but within the core of the enzyme several highly conserved regions shared by all PLDs have been recognised [1,6]. Among these, the most important are two copies of HxKxxxxD (or HKD) motif (Fig. 3, 4A) with the three residues absolutely conserved. The two motifs form together a catalytic site and their mutation abolishes enzyme activity. Our inspection of protein sequences of both characterised and predicted plant PLDs revealed that both HKD motifs are present in all isoforms, with the exception of rice PLDθ, which, unexpectedly, has the second copy mutated (Fig. 4A). Also the tomato PLDα3 has been reported to lack the aspartate residue from the second HKD motif, and its functionality has been questioned [19]. However, we found that LePLDα3 actually does posses the second catalytic motif conserved, as the conclusion of Laxalt et al. [19] is based on a protein sequence deduced from an un-spliced transcript. Indeed, a spliced cDNA corresponding to the same gene has been cloned independently and codes for a functional enzyme [23]. On the contrary, the replacement of histidine and lysine residues in the second motif of OsPLDθ does not seem to be an artefact (the region has been sequenced independently by Monsanto and IRGSP). It is tempting to speculate that the catalytically inactive OsPLDθ protein, if expressed, may fulfil regulatory roles via competition with functional PLDs and/or other proteins in binding of phospholipids.

C2- and PXPH-PLDs are believed to differ in their dependence on Ca^{2+}. Animal and fungal PLDs are not directly dependent on Ca^{2+}[1], and the same is likely also for plant PXPH-PLDs. On the other hand, most (but perhaps not all, see below) C2-PLDs will exhibit dependence on and regulation by Ca^{2+}, as the C2 domains usually bind phospholipids in a Ca^{2+}-dependent manner [26]. Structural characterisation of several C2 domains revealed that three Ca^{2+}-coordinating sites occupied by aspartate or asparagine residues are used generally, while other ligands are specific for individual domains (see [26] and Fig. 5). Different concentrations of Ca^{2+} required for optimal stimulation of alpha, beta and gamma PLD types have been explained in terms of the number of expected Ca^{2+}-coordinating residues present within the C2 domains of individual PLDs. It was noted that beta and gamma PLDs in Arabidopsis posses all the binding residues conserved in characterised C2 domains, whereas PLDα lacks two of them because of substitution [27,28]. However, this hypothesis relies on a protein derived from the non-accurate sequence of *AtPLDα1* cDNA, with an asparagine residue accidentally substituted for by an isoleucine at the position of the third conserved Ca^{2+}-coordinating site (Fig. 1 in [28]). The asparagine residue, occurring actually in all sequenced alpha-type PLDs (Fig. 5), is a Ca^{2+}-binding ligand in the C2 domain of phospholipase A_2[26]. Nevertheless, only some PLDα isoforms posses an aspartate residue at the site corresponding to the second generally shared Ca^{2+}-binding ligand, and the first position is mutated in all alpha-type PLDs (Fig. 5), in agreement with the higher $[Ca^{2+}]$ requirement of PLDα compared to PLDβ and γ [2], and with lower affinity to Ca^{2+} of the C2 from AtPLDα1 compared to the domain from AtPLDβ1 [28]. Indeed, all PLDs belonging to the 10-exonic cluster (i.e. PLDβ, γ, δ, κ and v) do posses all the three ligands conserved and most closely resemble the C2 domain from PLCδ1. On the other hand, the remaining PLD types, i.e. PLDε, ζ, η, θ and λ, lack all the three Ca^{2+}-binding residues due to substitution or deletion, while OsPLDμ retains only one (Fig. 5). Some C2 domains have been described that do not bind Ca^{2+}, but nonetheless do target respective proteins to the membranes [26]. It is therefore tempting to speculate that the C2 domains of the latter PLD types might function similarly and that these PLDs are perhaps Ca^{2+}-independent. However, still different modes of Ca^{2+} binding employing different residues may be discovered when additional C2 domains are structurally characterised, and threfore no definite functional predictions can be made from sequence only.

All characterised PC-PLDs from both major subfamilies are stimulated by or even dependent on PIP_2 under physiological or near-physiological conditions [1,2,29]. Mammalian and yeast PLDs appear to interact with PIP_2 by the PH domain [30] and via a novel highly conserved motif located between the two copies of the catalytic HKD motif (Fig 2; [31]). Interestingly, we found that three arginine residues in this motif involved in binding of PIP_2 are shared also by plant members of the PXPH-PLD subfami-

ly (Fig. 4B). In contrast, two of these residues are replaced by non-conserved amino acids in C2-PLDs, suggesting that this subfamily might adopt different mechanism of interaction with PIP$_2$.

Two motifs rich in basic residues and allegedly similar to a polyphosphoinositide-binding motif from gelsolin or phospholipase C ([KR]X$_{3-4}$KX [KR] [KR]) have been found in plant PLDs flanking the second catalytic HKD domain and proposed to mediate PIP$_2$ binding by PLDs [27]. It was claimed that all the basic residues are conserved only in Arabidopsis PLDβ1, whereas some are replaced with non-polar or acidic residues in AtPLDα1 and AtPLDγ1. However, our inspection of revised AtPLDβ1 sequence shows that the first "motif" of AtPLDβ1 has actually also only three basic residues, since the original motif definition was based on the inaccurate cDNA sequence. A genuine gelsolin/PLC consensus motif is found only in OsPLDβ1, OsPLDv1 and OsPLDv2 (Fig. 4A). The second proposed motif does not correspond exactly to the consensus of gelsolin family and PLC, but as if it was inverted (RKXRX$_4$R). Not only that merely AtPLDβ1 out of all C2-PLDs retains all four basic residues (Fig. 4A), but also there is no experimental evidence that such a motif superficially resembling gelsolin/PLC consensus really binds phosphoinositides. Available experimental data [27–29] do not exclude the possibility that the C2 domain instead is responsible for regulatory effects of PIP$_2$ on C2-PLDs.

Conclusions
Our analysis of Arabidopsis and rice genomic data complemented by searches of EST sequences revealed that plant PLDs are unexpectedly structurally diverse in two aspects.

First, individual plant genomes harbour various PLD types from both main PLD subfamilies. This is in sharp contrast to other large eukaryotic lineages. Species with completely or almost completely sequenced genomes, i.e. *Saccharomyces cervisiae*, *Schizosaccharomyces pombe*, *Caenorhabditis elegans* and *Drosophila melanogaster*, all posses only one gene from the PC-PLD family, and mammalian diversity is perhaps limited to two thoroughly characterised isoforms (our findings and [1]). All characterised animal and fungal PLDs belong to the PXPH-PLD subfamily. The occurrence of C2-PLDs beyond plants is unsure, they have been described only from angiosperms, and mosses are the most remote group for which C2-PLD sequences can be reliably found in databases (at least four distinct genes in *Physcomitrella patens*, see Additional files). It can be inferred from our phylogenetic analysis (see Fig. 1A) that the last common ancestor of animal, fungal and plant clades did harbour at least one gene from the PXPH-PLD subfamily as well as a gene that gave rise to the plant C2-PLDs. The latter PLD did not have to posses

the C2 domain, this could be acquired later during evolution. In any case, the true C2-PLDs or their predecessors have been lost from the lineage leading to animals and fungi.

The second aspect of plant PLD diversity relates to interspecific differences in the repertoire of distinct PLD types. For instance, there are no Arabidopsis orthologs of rice *OsPLDη*, *OsPLDθ* or *OsPLDκ*, while rice may lack counterparts of *AtPLDγ* or *AtPLDζ* from Arabidopsis. Similarly, only *LePLDα1* from tomato is a true ortholog of other dicotyledonous alpha PLDs, while *LePLDα2* and *LePLDα3* form together a separate lineage within the PLDα cluster (Fig. 1B). Multiple independent losses of distinct PLD types must have occurred in individual lineages of angiosperms (e.g. PLDη or PLDκ lost in the lineage leading to Arabidopsis, PLDγ disappeared from the lineage toward rice, see Fig. 1A,B). On the other hand, independent multiplication within individual genomes seems to be common as well, exemplified by pairs of PLDα or PLDβ in Arabidopsis and rice (Fig. 1A,B; Table 1 and Table 2). Two main mechanisms for gene multiplication have apparently contributed to the diversity of plant PLDs. Origin of the two PLDα isoforms in Arabidopsis can be accounted for by polyploidisation, while the PLDγ triplet in Arabidopsis and the rice *PLDη* cluster have probably arisen by a non-reciprocal crossing-over (see above).

Diversity of plant PLDs raises the question of functional specificities of individual isoforms. Although only limited functional predictions can be made solely on the basis of sequence data, the principal difference in domain structure between C2- and PXPH-PLDs suggests that their cellular functions will also differ. PXPH-PLDs in animals and yeasts appear to be involved in regulation of vesicular and membrane trafficking (reviewed in [1]), and plant orthologs could be used in a similar context [32]. Ca^{2+}-independent PLD activity, which is probably exhibited by all PXPH-PLDs, has not been reported from plant tissues, but this is perhaps due to overabundant activity of C2-PLDs (especially PLDα) and to the notably low abundance of regulatory enzymes in general. Moreover, some stimulatory factors might be necessary for measurable activity of plant PXPH-PLDs, as is the case for mammalian PLD1 [1].

On the other hand, C2-PLDs may fulfil plant-specific tasks. Evolutionary dynamics of this subfamily in angiosperms indicates that environmental factors might exert big influence on these enzymes. Recognised role for C2-PLDs in processes such as response to wounding, pathogen attack and multiple abiotic stresses seems to fit this view, but other processes including membrane degradation during senescence also have to be considered [2]. Functioning in signalling cascades may be common to both C2- and PXPH-PLDs, although the distinction be-

tween signalling function and the previously suggested roles does not have to be unambiguous.

Directions for future research on the plant PLD are straightforward. Besides the routinely used biochemical or pharmacological approaches, methods of reverse genetics (including anti-sense silencing and screening for insertional mutants) have to be employed. Partial functional redundancy, which can be expected for some plant PLD isoforms, could be coped with by generation of multiple mutants, accompanied by monitoring of expression of individual genes upon various circumstances and by experimental analysis of promoters. For deeper understanding of PLD regulation and interconnections within cellular context, attention must be focused on possible posttranslation modifications and interacting partners. Coordination of all these approaches has the potential to answer the question why plants farm so many PLDs.

Materials and Methods

For searches of public data we used BLAST toolkit at the National Centre for Biotechnology Information ([http://www.ncbi.nlm.nih.gov/BLAST] ; [33,34]). Searches were done in parallel with low complexity filter on and off, respectively, other parameters were kept default. All sequence databases containing plant data were exploited, including non-redundant nucleotide database, HTGS, GSS and EST databases and the non-redundant protein database. The final check of these databases was done between January 24, 2002, and January 26, 2002. Rice sequence data generated by Monsanto were searched using BLASTN and TBLAST facility at the rice-research.org webpage [http://www.rice-research.org/] . Hits from all BLAST searches with E-value above 0.1 were not considered for further analysis.

Multiple alignments were constructed by CLUSTALW (version 1.8) at the BCM Search Launcher ([http://search-launcher.bcm.tmc.edu/multi-align/multi-align.html] ; [35]), with default parameters. Manual editing of the alignments was done with the assistance of GENEDOC (Free Software Foundation, Inc.). Alternatively, multiple alignments were constructed using MACAW [35], with PAM 120 matrix used for protein sequences.

Exon-intron structures of Arabidopsis PLD genes were predicted employing GENESCAN ([http://genes.mit.edu/GENSCAN.html] ; [37]), GRAIL [http://grail.lsd.ornl.gov/Grail-1.3/] , NetGene2 [http://www.cbs.dtu.dk/services/NetGene2/] , FGENEP [http://dot.imgen.bcm.tmc.edu:9331/gene-finder/gf.html] and SplicePredictor [http://bioinformatics.iastate.edu/cgi-bin/sp.cgi] . Models proposed by each programme were compared and final structures were proposed relating to the information from cognate cDNAs, ESTs and multiple align-

ments. Rice genes were predicted manually with the assistance of GENESCAN (with the options set up for maize). These predictions were again checked by comparison with ESTs, cDNAs and protein sequences of PLDs.

Specific domains in PLD protein sequences were searched by SMART ([http://smart.embl-heidelberg.de/] ; [38]) and the Search Pfam tool ([http://www.sanger.ac.uk/Software/Pfam/] ; [39]). Searches for targeting signals were performed using the TargetP programme [http://www.cbs.dtu.dk/services/TargetP/] .

Phylogenetic trees were inferred from multiple alignments of protein sequences using appropriate programmes from the PHYLIP package, version 3.57c [40]. Neighbour-joining trees were constructed as described previously [41], PROTPARS programme was employed for maximum parsimony methods and confidence of the tree topology was estimated from 500 bootstrap replications. In the case that multiple alignments generated by CLUSTALW were used for phylogenetic inference, regions that could not be aligned unambiguously or containing deletions/insertions had been removed prior. For phylogenetic inference from alignments generated by MACAW only the most conserved boxes were used.

Levels of sequence identity/similarity occasionally noted through the text refer to values calculated by the BLAST 2 Sequences programme [42] with the low complexity filter off.

Note added in proof

A reannotation of the Arabidopsis genome released into GenBank after submission of the manuscript removes some inaccuracies in predictions of exon-intron structures of PLD genes independently uncovered also by our analysis. An updated list of Arabidopsis PLD genes has been deposited into the TAIR gene families database. [http://www.arabidopsis.org/info/genefamily.html]

Additional material

Additional file 1
Corrected versions of coding sequences of previously annotated plant PLD genes
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-3-2-S1.doc]

Additional file 2
Novel rice PLD genes found in GenBank genomic sequences or Monsanto rice genome draft
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-3-2-S2.doc]

Additional file 3

Protein sequences of plant PLDs used for structural and phylogenetic analysis
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-3-2-S3.doc]

Additional file 4

A list of EST sequences used in this study
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-3-2-S4.doc]

Acknowledgements

References

1. Liscovitch M, Czarny M, Fiucci G, Tang X: **Phospholipase D: molecular and cell biology of a novel gene family.** *Biochem* 2000, **345**:401-415
2. Wang X: **Multiple forms of phospholipase D in plants: the gene family, catalytic and regulatory properties, and cellular functions.** *Prog Lipid Res* 2000, **39**:109-149
3. Hanahan DJ, Chaikoff IL: **A new phospholipide-splitting enzyme specific for the ester linkage between the nitrogenous base and the phosphoric acid grouping.** *J Biol Chem* 1947, **169**:699-705
4. Wang X, Xu L, Zheng L: **Cloning and expression of phosphatidylcholine-hydrolyzing phospholipase D from Ricinus communis L.** *J Biol Chem* 1994, **269**:20312-20317
5. den Hartog M, Musgrave A, Munnik T: **Nod factor-induced phosphatidic acid and diacylglycerol pyrophosphate formation: a role for phospholipase C and D in root hair deformation.** *Plant J* 2001, **25**:55-65
6. Munnik T, Irvine RF, Musgrave A: **Phospholipid signalling in plants.** *Biochim Biophys Acta* 1998, **1389**:222-272
7. Lein W, Saalbach G: **Cloning and direct G-protein regulation of phospholipase D from tobacco.** *Biochim Biophys Acta* 2001, **1530**:172-183
8. Chapman KD: **Emerging physiological roles for N-acylphosphatidylethanolamine metabolism in plants: signal transduction and membrane protection.** *Chem Phys Lipids* 2000, **108**:221-229
9. Munnik T: **Phosphatidic acid: an emerging plant lipid second messenger.** *Trends Plant Sci* 2001, **6**:227-233
10. Fan L, Zheng S, Wang X: **Antisense suppression of phospholipase D alpha retards abscisic acid- and ethylene-promoted senescence of postharvest Arabidopsis leaves.** *Plant Cell* 1997, **9**:2183-2196
11. Pappan K, Qin W, Dyer JH, Zheng L, Wang X: **Molecular cloning and functional analysis of polyphosphoinositide-dependent phospholipase D, PLDbeta, from Arabidopsis.** *J Biol Chem* 1997, **272**:7055-7061
12. AGI: **Analysis of the genome sequence of the flowering plant Arabidopsis thaliana.** *Nature* 2000, **408**:796-815
13. Barry GF: **The use of the Monsanto draft rice genome sequence in research.** *Plant Physiol* 2001, **125**:1164-1165
14. Ponting CP, Kerr ID: **A novel family of phospholipase D homologues that includes phospholipid synthases and putative endonucleases: Identification of duplicated repeats and potential active site residues.** *Protein Sci* 1996, **5**:914-922
15. Katagiri T, Takahashi S, Shinozaki S: **Involvement of a novel Arabidopsis phospholipase D, AtPLDδ, in dehydration-inducible**

16. Wang C, Wang X: **A Novel Phospholipase D of Arabidopsis That Is Activated by Oleic Acid and Associated with the Plasma Membrane.** *Plant Physiol* 2001, **127**:1102-12
17. Wang X: **Plant phospholipases.** *Annu Rev Plant Physiol Plant Mol Biol* 2001, **52**:211-231
18. Qin W, Dyer JH, Zheng Z, Wang X: **Isolation and nucleotide sequence of the fourth phospholipase D (accession no. Af138281), PLD-gamma 2, from Arabidopsis thaliana. (PGR99-084)** *Plant Physiol* 1999, **120**:635
19. Laxalt AM, ter Riet B, Verdonk JC, Parigi L, Tameling WIL, Vossen J, Haring M, Musgrave A, Munnik T: **Characterization of five tomato phospholipase D cDNAs: rapid and specific expression of LePLDbeta1 on elicitation with xylanase.** *Plant* 2001, **26**:237-247
20. Ueki J, Morioka S, Komari T, Kumashiro T: **Purification and characterization of phospholipase-D (PLD) from rice (Oryza-sativa L) and cloning of cDNA for PLD from rice and maize (Zea-mays L).** *Plant Cell Physiol* 1995, **36**:903-914
21. Morioka S, Ueki J, Komari T: **Characterization of two distinctive genomic clones (accession nos. AB001919 and AB001920) for phospholipase D from rice (PGR97-076).** *Plant Physiol* 1997, **114**:396
22. Almquist KC, Paliyath G: **Cloning and sequencing of a full-length cDNA coding for phospholipase D alpha (Accession Number 201661) from Lycopersicon esculentum Mill. (PGR00-013)** *Plant Physiol* 2000, **122**:292
23. Whitaker BD, Smith DL, Green KC: **Cloning, characterization and functional expression of a phospholipase D alpha cDNA from tomato fruit.** *Physiol Plant* 2001, **112**:87-94
24. Ewing RM, Ben Kahla A, Poirot O, Lopez F, Audic S, Claverie JM: **Large-scale statistical analyses of rice ESTs reveal correlated patterns of gene expression.** *Genome Res* 1999, **9**:950-959
25. Frank W, Munnik T, Kerkmann K, Salamini F, Bartels D: **Water deficit triggers phospholipase D activity in the resurrection plant Craterostigma plantagineum.** *Plant Cell* 2000, **12**:111-124
26. Rizo J, Sudhof TC: **C2-domains, structure and function of a universal Ca2+-binding domain.** *J Biol Chem* 1998, **273**:15879-15882
27. Qin W, Pappan K, Wang X: **Molecular heterogeneity of phospholipase D (PLD). Cloning of PLDgamma and regulation of plant PLDgamma, -beta, and -alpha by polyphosphoinositides and calcium.** *J Biol Chem* 1997, **272**:28267-28273
28. Zheng L, Krishnamoorthi R, Zolkiewski M, Wang X: **Distinct Ca2+ binding properties of novel C2 domains of plant phospholipase dalpha and beta.** *J Biol Chem* 2000, **275**:19700-19706
29. Pappan K, Wang X: **Plant phospholipase Dalpha is an acidic phospholipase active at near-physiological Ca(2+) concentrations.** *Arch Biochem Biophys* 1999, **368**:347-353
30. Hodgkin MN, Masson MR, Powner D, Saqib KM, Ponting CP, Wakelam MJ: **Phospholipase D regulation and localisation is dependent upon a phosphatidylinositol 4,5-biphosphate-specific PH domain.** *Curr Biol* 2000, **10**:43-46
31. Sciorra VA, Rudge SA, Prestwich GD, Frohman MA, Engebrecht J, Morris AJ: **Identification of a phosphoinositide binding motif that mediates activation of mammalian and yeast phospholipase D isoenzymes.** *EMBO J* 1999, **18**:5911-5921
32. Munnik T, Musgrave A: **Phospholipid signaling in plants: holding on to phospholipase D.** *Sci STKE* 2001, **111**:PE42
33. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**:403-410
34. Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25**:3389-3402
35. Thompson JD, Higgins DG, Gibson TJ: **Clustal-W – improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acids Res* 1994, **22**:4673-4680
36. Schuler GD, Altschul SF, Lipman DJ: **A workbench for multiple alignment construction and analysis.** *Prot Struct Fund Genet* 1991, **9**:180-190
37. Burge C, Karlin S: **Prediction of complete gene structures in human genomic DNA.** *J Mol Biol* 1997, **268**:78-94

38. Schultz J, Copley RR, Doerks T, Ponting CP, Bork P: **SMART: a web-based tool for the study of genetically mobile domains.** *Nucleic Acids Res* 2000, **28:**231-234
39. Bateman A, Birney E, Durbin R, Eddy SR, Howe KL, Sonnhammer ELL: **The Pfam protein families database.** *Nucleic Acids Res* 2000, **28:**263-266
40. Felsenstein J: **Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods.** *Methods Enzymol* 1996, **266:**418-427
41. Cvrckova F: **Are plant formins integral membrane proteins?** *Genome Biology* 2000, **1:**1-001
42. Tatusova TA, Madden TL: **Blast 2 Sequences, a new tool for comparing protein and nucleotide sequences.** *FEMS Microbiology Letters* 1999, **174:**247-250
43. Dyer JH, Zheng L, Wang X: **Cloning and nucleotide sequence of a cDNA (Accession No. U36381) encoding phospholipase D from Arabidopsis. (PGR95-096)** *Plant Physiol* 1995, **109:**1497
44. Gardiner JC, Harper JD, Weerakoon ND, Collings DA, Ritchie S, Gilroy S, Cyr RJ, Marc J: **A 90-kd phospholipase d from tobacco binds to microtubules and the plasma membrane.** *Plant Cell* 2001, **13:**2143-58
45. Nalefski AE, Falke JJ: **The C2 domain calcium-binding motif: Structural and functional diversity.** *Protein Sci* 1996, **12:**2375-2390