3 Biotech (2017) 7:216
DOI 10.1007/s13205-017-0790-x

ORIGINAL ARTICLE

Origin and evolution of group XI secretory phospholipase A_2 from flax (*Linum usitatissimum*) based on phylogenetic analysis of conserved domains

Payal Gupta[1] · Raman Saini[2] · Prasanta K. Dash[1]

Received: 29 March 2017 / Accepted: 12 June 2017 / Published online: 1 July 2017

Abstract Phospholipase A_2 (PLA_2) belongs to class of lipolytic enzymes (EC 3.1.1.4). Lysophosphatidic acid (LPA) and free fatty acids (FFAs) are the products of PLA_2 catalyzed hydrolysis of phosphoglycerides at sn-2 position. LPA and FFA that act as second mediators involved in the development and maturation of plants and animals. Mining of flax genome identified two phospholipase A_2 encoding genes, viz., *LusPLA$_2$I* and *LusPLA$_2$II* (*Linum usitatissimum* secretory phospholipase A_2). Molecular simulation of LusPLA$_2$s with already characterized plant sPLA$_2$s revealed the presence of conserved motifs and signature domains necessary to classify them as secretory phospholipase A_2. Phylogenetic analysis of flax sPLA$_2$ with representative sPLA$_2$s from other organisms revealed that they evolved rapidly via gene duplication/deletion events and shares a common ancestor. Our study is the first report of detailed phylogenetic analysis for secretory phospholipase A_2 in flax. Comparative genomic analysis of two LusPLA$_2$s with earlier reported plant sPLA$_2$s, based on their gene architectures, sequence similarities, and domain structures are presented elucidating the uniqueness of flax sPLA$_2$.

Keywords Secretory phospholipase A_2 · Phylogenetic analysis · Gene duplication/deletion · Conserved domain · Ancestor

Introduction

Phospholipase A_2 is a large superfamily of lipolytic enzymes categorized by their ability to catalyze the hydrolysis of the 2-acyl ester bond of phosphoglycerides. Despite availability of detailed molecular information for animal phospholipases, knowledge and function of phospholipase A_2 from plants are meagre (Wang 2001). Lipids are important components of plant cell membrane and are hydrolyzed at an accelerated rate by phospholipases during stress to generate signaling molecules such as phosphatidic acid (PA), diacylglycerol (DAG), lysophospholipids (LPLs), and free fatty acids (FFA). These molecules (PA, DAG) act as second messengers to elicit various cellular processes. With increasing understanding of lipid signaling and their regulatory roles during stress, structural and evolutionary analysis of PLA_2 has gained impetus. PLA$_2$s have been classified into different classes based on function, molecular structure, and conserved domain annotation. On the basis of their structure and function, PLA$_2$s are divided into five classes: (1) secretory phospholipase A_2 (sPLA$_2$s), (2) calcium-dependent cytosolic phospholipase A_2 (cPLA$_2$s), (3) calcium-independent cytosolic phospholipase A_2 (iPLA$_2$s), (4) platelet-activating factor acetyl hydrolases (PAF-AH), and (5) lysosomal phospholipase A_2 (Six and Dennis 2000). Based on molecular structure, PLA$_2$s have further been classified into 15 groups (I–XV). Group IA, IB, IIA, IIB, IIC, IID, IIE, IIF, III, V, IX, X, XIA, XIB, XIIA, XIIB, XIII, and XIV are assigned to secretory phospholipase A_2. Group IV is assigned to

Electronic supplementary material The online version of this article (doi:10.1007/s13205-017-0790-x) contains supplementary material, which is available to authorized users.

✉ Payal Gupta
payalgupta@gmail.com

✉ Prasanta K. Dash
pdas@nrcpb.org

[1] ICAR-National Research Centre on Plant Biotechnology, Pusa Campus, New Delhi 110012, India

[2] Department of Biotechnology, Kurukshetra University, Kurukshetra 136118, India

calcium-dependent cytosolic phospholipase A_2, group VI is assigned to calcium-independent cytosolic phospholipase A_2, and groups VII and VIII are assigned to platelet-activating factor acetyl hydrolases (Schaloske and Dennis 2006).

On the basis of conserved domains, phospholipase A_2 are grouped into two major families viz, cd00618 (PLA_2_like) and pfam00068 (Phospholipase_A_2_1). Based on in silico annotated conserved domain collection the cd00618 family is categorized into five distinct sub-families (Nevalainen et al. 2012) such as "(1) the cd04704 PLA_2_bee_venom_like: phospholipase A_2, similar to bee venom; (2) the cd04705 PLA_2_group_III_like: PLA_2, similar to group III PLA_2; (3) the cd04706 PLA_2_plant: plant specific phospholipase A_2; (4) the cd04707 octoconin_90: phospholipase A_2_like domain present in octoconin_90 and octoconin_95; and (5) the cd00125 PLA_2c: secretory and cytosolic phospholipase A_2". Furthermore, the pfam00068 collection is sub-divided into two distinct sub-groups; (1) pfam06951 PLA_2G12, group XII secretory phospholipase A_2 precursor and (2) pfam09056 phospholip_A_2_3: prokaryotic phospholipase A_2 domain found in PLA_2 of bacteria and fungi (Nevalainen et al. 2012).

Among all PLA_2s, secretory phospholipase A_2 are most extensively studied. Plant $sPLA_2$s are placed in group XI and are further classified into sub-groups XIA and XIB (Six and Dennis 2000). The plant $sPLA_2$s belong to cd04706 sub-family of cd00618 collection (Marchler-Bauer et al. 2011). Based on the presence of conserved tyrosine, glycine and aspartic acid residues in the calcium binding site, and histidine residue in the active site motif, cd40706 domain has been identified in the $sPLA_2$s of Alphaproteobacteria of phylum Proteobacteria such as *Methylobacterium extorquens* and phylum Firmicutes such as *Bacillus cereus* and *Streptococcus equi*. Few patatin-like PLA_2s have also been identified (Scherer et al. 2010).

Secretory phospholipase A_2 are low-molecular-weight proteins ranging from 15 to 19 kDa (Schaloske and Dennis 2006). All plant $sPLA_2$s contain 12 Cys residues that form six intramolecular disulfide bridges (Six and Dennis 2000), a signature phospholipase A_2 (PA2c) domain containing Ca^{2+} binding loop (YGKYCGxxxxGC), a catalytic site motif (DACCxxHDxC) characterized by highly conserved His/Asp dyad (Lee et al. 2005; Mansfeld et al. 2006), and an N-terminal signal peptide required for secretion into target tissue.

In plants, $sPLA_2$s are known to be associated with numerous biological processes leading to growth (Lee et al. 2003) and development (Kim et al. 1999; Lee et al. 2010). They are also known to be involved in plant's response to wounding (Pohnert 2002), programmed cell death (Reina-Pinto et al. 2009), and abiotic stress (Scherer et al. 2010). Although all $sPLA_2$s share a common enzymatic mechanism involving the canonical histidine residue, plant $sPLA_2$s share homology with their animal counterparts only in the catalytically important Ca^{2+} binding loop and the active site motif (Lee et al. 2005). Regardless of similarity in conserved domain region, the origin and emergence of plant $sPLA_2$s compared to animal $sPLA_2$s needs elucidation.

Recently, genomic information about flax (Dash et al. 2014, 2015; Wang et al. 2012) other field and plantation crops (Dash and Rai 2016) has accumulated in the literature, but role of phospholipases and their evolution has not been investigated. Thus, our work focuses on understanding the origin and evolution of secretory phospholipase A_2 in flax and their conservation among plant PLA_2 superfamily. We bioinformatically explored on identification of conserved domains in representative collection of $sPLA_2$s belonging to annotated protein sequences from ten different plant families, bacteria, and animals. Subsequently, position of flax $sPLA_2$s within the existing model of origin and evolution of secretory phospholipase A_2 were compared to representative $sPLA_2$s of each sub-family of cd00618 collection for similarity and evolution analysis.

Methods

Sequence retrieval from database

The amino acid sequences for $sPLA_2$s were retrieved from the publicly available protein repository of the NCBI database (http://www.ncbi.nlm.nih.gov). $sPLA_2$ annotated sequences from ten different plant families, viz., Brassicaceae, Caryophyllaceae, Euphorbiaceae, Fabaceae, Malvaceae, Poaceae, Rutaceae, Saliaceae, Solanaceae including Linaceae, and venom were used for analysis. The amino acid sequence of secretory phospholipase A_2 from sub-families of cd00618 collection, viz., cd00125, cd04704, cd04705, cd04706, and cd04707 as identified earlier (Nevalainen et al. 2012) were also retrieved from NCBI.

Conserved domain search

The conserved domains of flax $sPLA_2$s were identified using the NCBI Conserved Domains Database CDDv3.14-47363 PSSMs annotation using default parameters. The conserved domain region of representative $sPLA_2$ of cd00125, cd04704, cd04705, cd04706, and cd04707 collections was identified using NCBI's Batch CD-search (Marchler-Bauer et al. 2011). Histidine and aspartic acid active site protein motifs were identified based on the PROSITE database (http://au.expasy.org/prosite) pattern annotation.

Homology search

The similarity of flax $sPLA_2$s with known plant $sPLA_2$s and bacterial $sPLA_2$ belonging to phylum Proteobacteria and Firmicutes (cd04706) and representative $sPLA_2$ of each sub-family of cd00618 collection were searched and visualized using Circoletto (Darzentas 2010) with a selected *E* value of 10^{-5} (normal) with default parameters and untangling of ribbons switched off.

Phylogenetic analysis

Multiple protein sequence alignments were carried out for flax, known plant, and snake venom $sPLA_2$s using MAFFT version 7.271 program (Katoh and Standley 2013) with L-INS-I strategy, BLOSUM62 scoring matrix for amino acid, and 1.53 gap opening penalty. Output was generated in Phylip format. A similarity score for each residue of the aligned sequences was calculated by ESPRIPT 3.0 with default parameters (Robert and Gouet 2014). Phylogenetic analysis was performed using amino acid sequences of $sPLA_2$s from 67 different plant taxa and snake venom to evaluate the evolutionary relationship among them. The amino acid alignment produced by MAFFT was submitted to modelgenerator_v_8. 51 for selecting the best model for amino acid substitution (Keane et al. 2006). Phylogenetic analysis was performed using PhyML v3.0 with an improved version of NJ (Neighbor Joining) called BioNJ. Tree topology was searched using Nearest Neighbor Interchanges (NNIs) algorithm (Guindon et al. 2010). The JTT substitution model and four gamma-distributed rate categories to account for rate heterogeneity across sites were employed. The gamma shape parameter was estimated directly from the data and analysis was performed using 1000 bootstrap replicates. The proportion of invariable sites was fixed. The tree was obtained in Newick format. Graphical representation of the phylogenetic tree was performed by i-TOL (http://itol.embl.de/). The comparison of flax $sPLA_2$s and representative $sPLA_2$s from each sub-family of cd00618 collection was performed using MAFFT-win and visualized using ESPRIPT with default parameters. The phylogenetic relationship among them was deduced using MEGA6 (Tamura et al. 2013). A maximum-likelihood (ML) tree was constructed to identify the position of flax $sPLA_2$s in the existing model of origin and evolution of $sPLA_2$ belonging to cd00618 collection as described earlier (Nevalainen et al. 2012).

Result and discussion

Flax is a multipurpose high value cash crop. It is mainly grown for its seed oil and fine fiber linen (Dash et al. 2014; Shivaraj et al. 2017). Recently, it has gained importance as a health booster and nutraceuticals. Secretory phospholipase A_2 have been isolated from a number of organisms including bacteria, fungi, and plants, but information from flax is lacking. Among plants, $sPLA_2$s have been identified from elm (*Ulmus glabra*) (Stahl et al. 1998), tobacco (*Nicotiana tabacum*) (Fujikawa et al. 2005), rice (*Oryza sativa*) (Stahl et al. 1999), arabidopsis (*Arabidopsis thaliana*) (Lee et al. 2003), carnations (*Dianthus caryophyllus*) (Kim et al. 1999), wheat (*Triticum durum*) (Verlotta et al. 2013), and soybean (*Glycine max*) (Mariani et al. 2012). A few patatin-like PLA_2s have also been isolated from plants (Scherer et al. 2010). Initially, $sPLA_2$s were classified into only two groups GI and GII, but later on, $sPLA_2$s were classified into GI/II/V/X. $sPLA_2$ superfamily has expanded over past years due to identification of conserved domains from various organisms and now $sPLA_2$s are grouped into GI/II/III/V/IX/X/XI/XII/XIII/XIV.

The Conserved Domain Database (CDD) identified the signature (PA2c) domain of secretory phospholipase A_2 in $LusPLA_2I$ and $LusPLA_2II$. The PA2c domain was characterized by calcium binding loop and catalytically active motif containing His/Asp dyad (Fig. 1). Sequence alignment using MAFFT tool and visualization with ESPRIPT 3.0 also revealed the presence of the conserved calcium binding loop and catalytic site motif. The His/Asp dyad in the catalytic site motif was also identified along with other conserved amino acids involved in calcium binding and catalysis (Fig. 1).

Amino acid sequences of plant $sPLA_2$s were submitted to Circoletto (Darzentas 2010) to deduce the similarity among them (Fig. 2a). The amino acid sequences of $LusPLA_2I$ and $LusPLA_2II$ revealed 34% identity among them. $LusPLA_2I$ showed only 29% identity with venom $sPLA_2$. $LusPLA_2I$ showed maximum similarity to *Arabidopsis lyrata* $sPLA_2$ (70%), whereas $LusPLA_2II$ showed maximum similarity to *Oryza sativa* $sPLA_2$ (65%) (Supplementary Table 1, 2). Sequence similarity of $LusPLA_2$s with nine different plant families showed more similarity to Brassicaceae (*Arabidopsis lyrata*) and Poaceae (*Oryza sativa*) (Fig. 2b).

A phylogenetic tree based on the amino acid sequence similarities between flax $sPLA_2$s and known plant $sPLA_2$s was proposed to ascertain their position in group XI (Fig. 3). Phylogenetic analysis was performed using the maximum-likelihood method and the optimal substitution model. The Akaike information criterion (AIC) and Bayesian information criterion (BIC) (Posada and Buckley 2004) evaluation for 96 protein substitution models with four gamma categories using modelgenerator_v_8. 51 revealed that JTT (Jones et al. 1992) was the best model with highest AIC and BIC value of 74,907.8221 and 75,945.8485, respectively, for phylogenetic analysis of our data set. Based on phylogeny, $sPLA_2$s belonging to 68 taxa

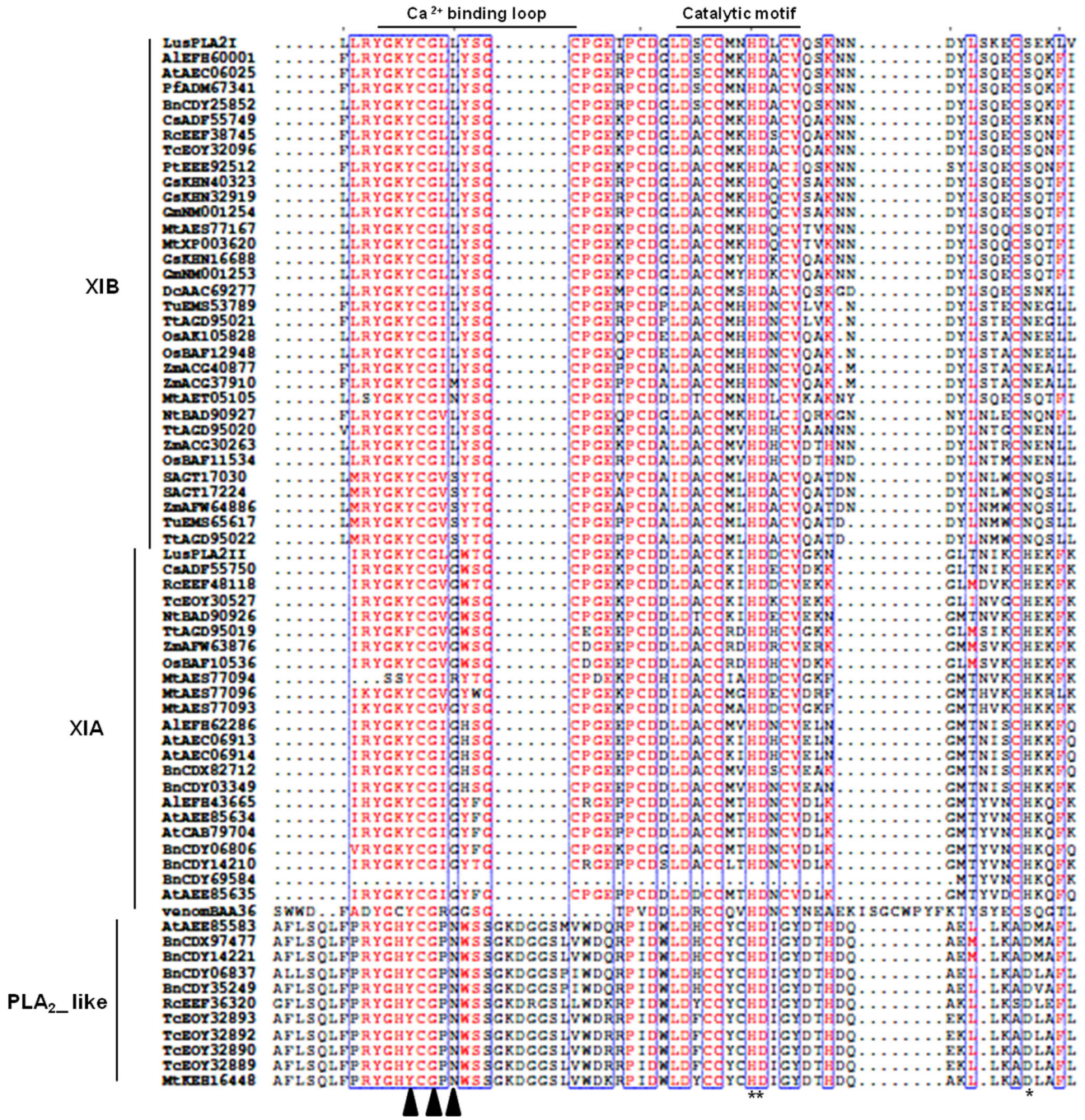
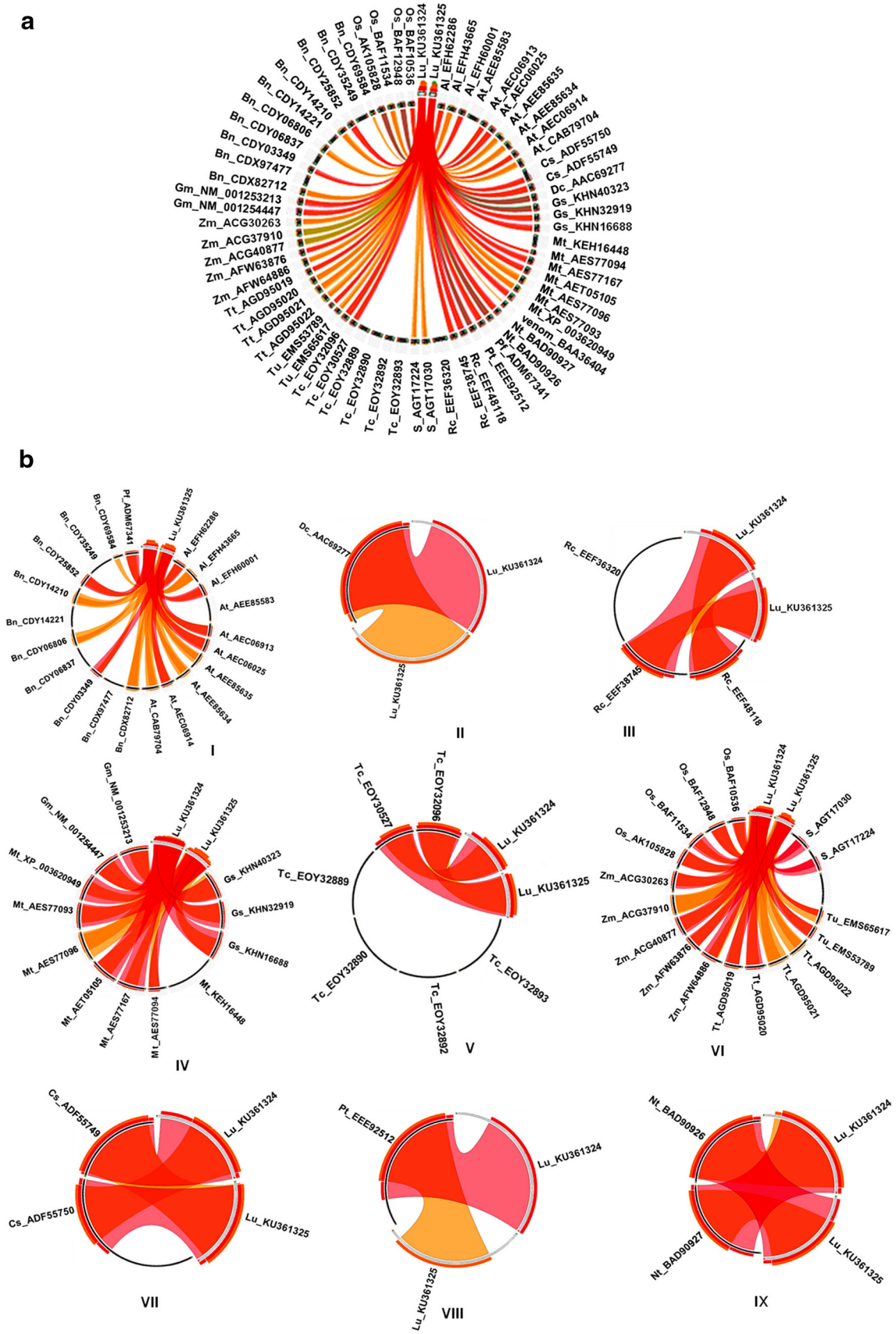


Fig. 1 Pairwise alignment of the amino acid sequences of LusPLA$_2$I and LusPLA$_2$II with multiple plant sPLA$_2$s from ten different plant families and snake venom sPLA$_2$ comprising 68 protein sequences. Amino acid sequence of snake venom was included to evaluate the similarity with the animal counterpart. Highly conserved Cys residues are highlighted in red color. Conserved Ca^{2+} binding loop and catalytically active motif are also indicated. *Triangles* denote the amino acid residues involved in calcium binding; *stars* denote the amino acid residues involved in catalysis

were classified into three groups, viz., XI, PLA$_2$_like proteins, and venom sPLA$_2$ (Fig. 3). Among these three groups, snake venom sPLA$_2$ formed a separate group sharing common blood line with group XI sPLA$_2$s. Earlier animal and plant sPLA$_2$s have been reported to show similarities in the conserved domain region and evolution (Lee et al. 2005). Group XI formed a separate class of sPLA$_2$s and due to point mutations in the conserved catalytic domain, it was further sub-divided into groups XIA and XIB. These point mutations occurring during speciation event also separated group XI from PLA$_2$_like proteins. Phylogenetic analysis revealed that cluster


a
Os_BAF10536
Os_BAF12948
Os_BAF11534
Os_AK105828
Bn_CDY69584
Bn_CDY35249
Bn_CDY25852
Bn_CDY14210
Bn_CDY14221
Bn_CDY06806
Bn_CDY06837
Bn_CDY03349
Bn_CDX97477
Bn_CDX82712
Gm_NM_001253213
Gm_NM_001254447
Zm_ACG30263
Zm_ACG37910
Zm_ACG40877
Zm_AFW63876
Zm_AFW64886
Tt_AGD95019
Tt_AGD95020
Tt_AGD95021
Tt_AGD95022
Tu_EMS53789
Tu_EMS65617
Tc_EOY30527
Tc_EOY32096
Tc_EOY32889
Tc_EOY32890
Tc_EOY32892
Tc_EOY32893
S_AGT17224
S_AGT17030
Rc_EEF36320
Rc_EEF38745
Rc_EEF48118
Pt_EEE92512
Pt_ADM67341
Nt_BAD90926
venom_BAA36404
Mt_XP_003620949
Mt_AES77093
Mt_AES77096
Mt_AET05105
Mt_AES77167
Mt_AES77094
Mt_KEH16448
Gs_KHN16688
Gs_KHN32919
Gs_KHN40323
Dc_AAC69277
Cs_ADF55749
At_CAB79704
At_AEC06914
At_AEE85634
At_AEC06025
At_AEE85635
At_AEC06913
At_AEE85583
At_EFH60001
Al_EFH43665
Al_EFH62286
Lu_KU361325
Lu_KU361324
b
I
II
III
IV
V
VI
VII
VIII
IX

◀**Fig. 2** Visualization of similarity among flax and known plant $sPLA_2$s. **a** % similarity among $LusPLA_2$s ($LusPLA_2I$ and $LusPLA_2II$) and known plant secretory PLA_2s. **b** % similarity of $LusPLA_2$s ($LusPLA_2I$ and $LusPLA_2II$) to known $sPLA_2$s from different plant families viz; *I* Brassicaceae; *II* Caryophyllaceae; *III* Euphorbiaceae; *IV* Fabaceae; *V* Malvaceae; *VI* Poaceae; *VII* Rutaceae; *VIII* Saliaceae; *IX* Solanaceae. % similarity is represented by color scheme, viz., *blue color* represents 25% similarity (worst), *green color* represents 50% similarity, *orange color* represents 75%, and *red color* represents more than 75% and similarity

PLA_2_like proteins evolved into distinct group from a common ancestor as the members of this group are less closely related to group XI of plant $sPLA_2$s.

Plant $sPLA_2$s have been placed in group XI (Six and Dennis 2000) which is divided into sub-groups XIA and XIB as per PLA_2 superfamily numbering system (Burke and Dennis 2009). The sequence analysis (Fig. 1) revealed that $LusPLA_2$s were placed in sub-group XIA and XIB. Alignment of $LusPLA_2$ protein sequences with known plant $sPLA_2$s revealed that $LusPLA_2I$ is placed in group XIB along with isozymes of $sPLA_2$ from other analyzed plant species (Supplementary Table 1).

Isoenzymes of $sPLA_2$ from multiple plant species were sorted in group XIA along with $LusPLA_2II$ as per their similarity (Supplementary Table 2). Isoenzymes from *Arabidopsis thaliana*, *Brassica napus*, *Ricinus communis*,

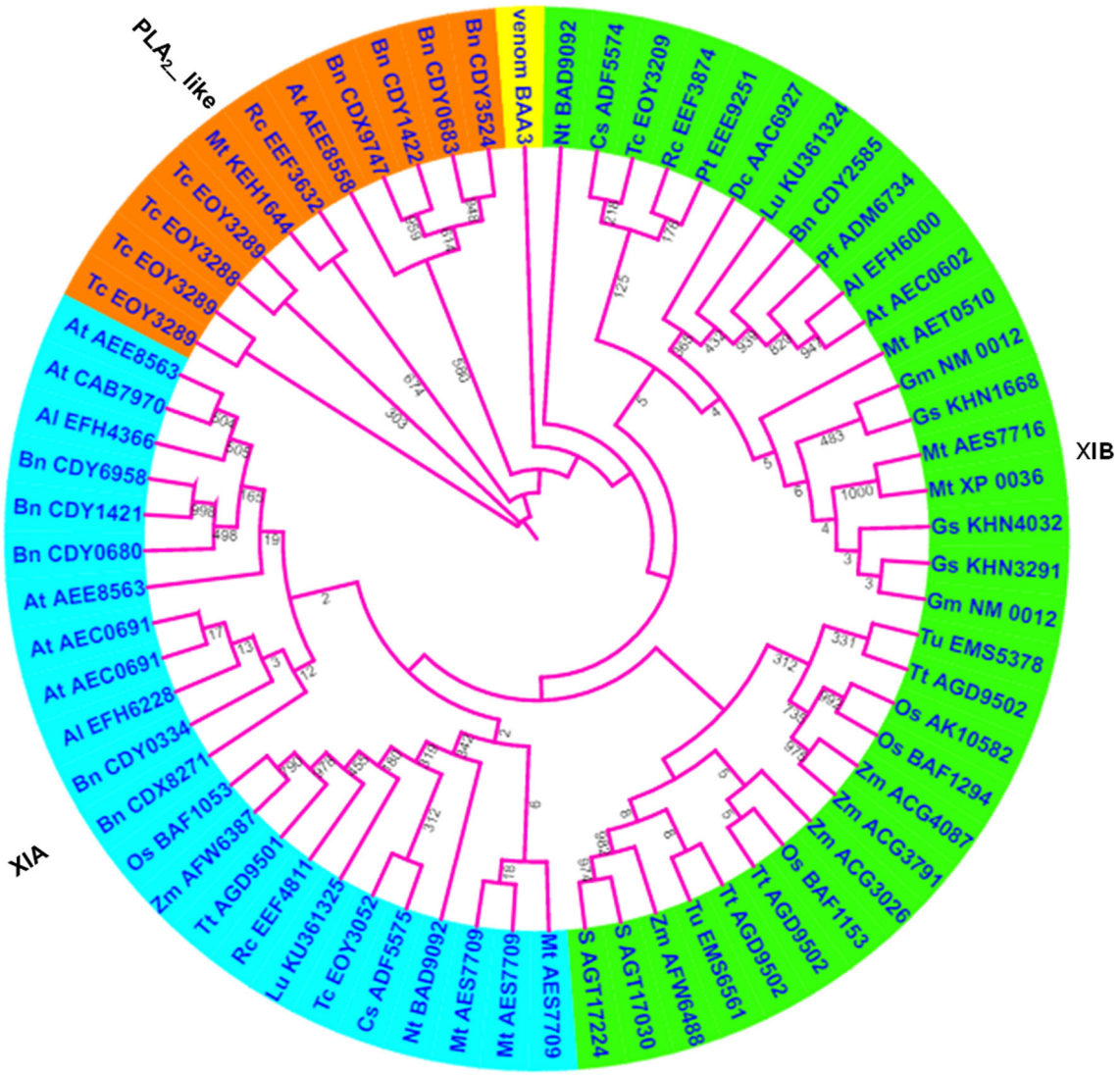


Fig. 3 Phylogenetic tree of $sPLA_2$ depicting evolutionary relationship of $sPLA_2$s from flax with orthologous plant and snake venome $sPLA_2$s. Un-rooted phylogenetic tree represents total 68 $sPLA_2$s protein sequences belonging to flax, known plants, and snake venom. $sPLA_2$s from multiple origin were divided into three groups such as XI, PLA_2_like proteins, and venom $sPLA_2$. Flax $sPLA_2$ was grouped in cluster XI, $LusPLA_2I$ in XIB, and $LusPLA_2II$ in XIA. Venom $sPLA_2$ formed a separate branch of its own and was more similar to PLA_2_like proteins. Evolutionary analysis revealed that though the $sPLA_2$s belonging to different taxa segregated into three group, they evolved from a common ancestor. *Color scheme* indicates different $sPLA_2$ groups such as *green color* indicates sub-group XIB, *blue color* indicates sub-group XIA, *orange color* indicates group PLA_2_-like proteins, and *yellow color* indicates venom $sPLA_2$

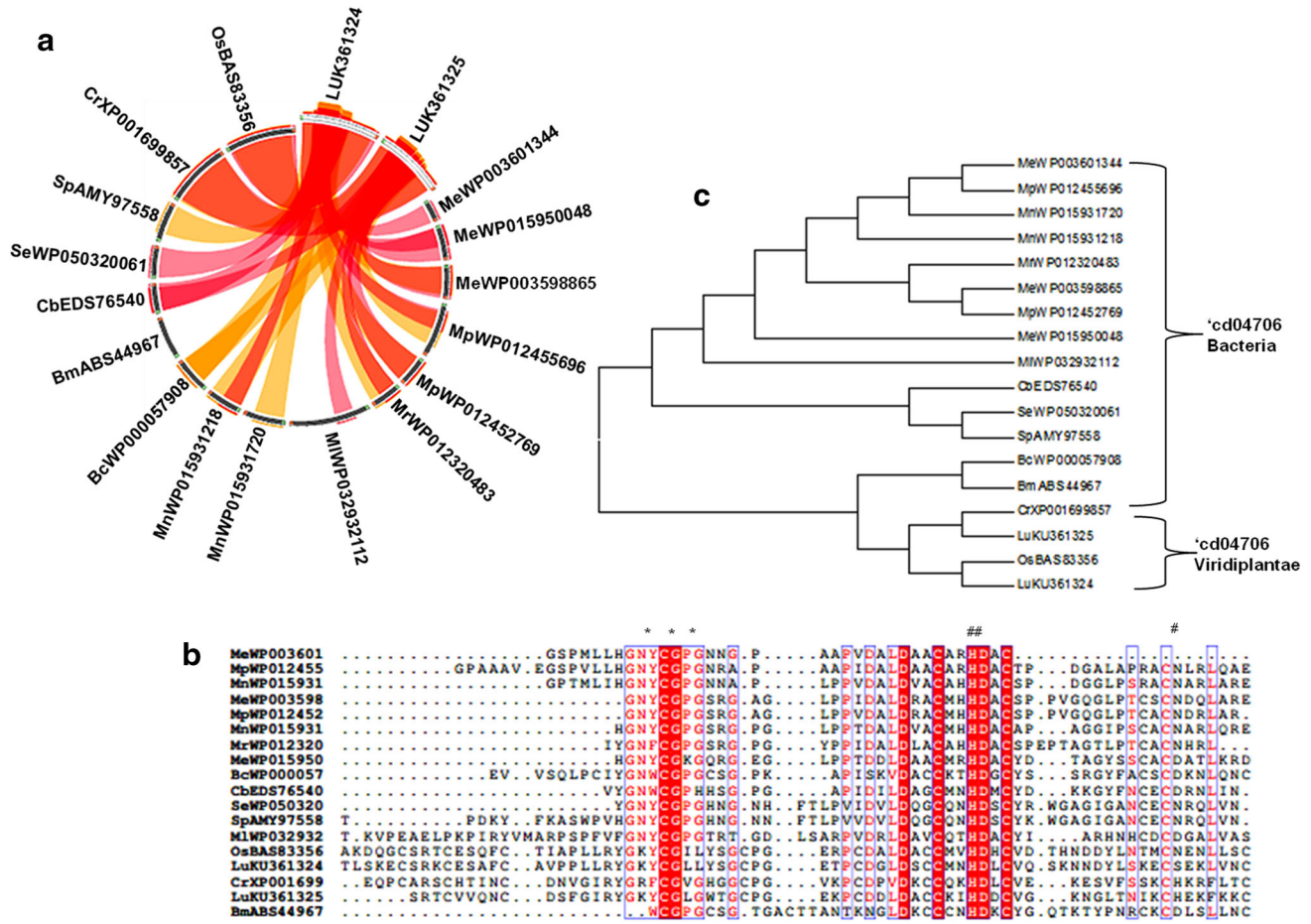


Fig. 4 Analysis of conserved domains and phylogeny of $sPLA_2$ belonging to cd04706 sub-family. **a** % similarity among the conserved domain region of flax and bacterial $sPLA_2$s of phylum Alphaproteobacteria and Firmicutes. % similarity is represented by color scheme, viz., *blue color* represents 25% similarity (worst), *green color* represents 50% similarity, *orange color* represents 75%, and *red color* represents more than 75% and similarity. **b** Pairwise alignment of the conserved domain region of flax and bacterial $sPLA_2$s. *Asterisks* denote the amino acid residues involved in Ca^{2+} binding and *hashes* denote the amino acid involved in catalysis. **c** Phylogenetic tree of flax $sPLA_2$s and bacterial $sPLA_2$s of phylum Alphaproteobacteria and Firmicutes. Evolutionary analyses reveal that the members of cd04706 collection bear common ancestor

Theobroma cacao, and *Medicago truncatula* formed a separate group of PLA_2_like proteins.

Plant $sPLA_2$s display minor differences in their Ca^{2+} binding domain and catalytic motif on the basis of which they are divided into sub-group XIA and XIB. $sPLA_2$s of sub-group XIA contain Arg or Gly and group XIB contains Asn or Ser residues as third Ca^{2+} coordinating amino acid (Fig. 1). In addition, plant $sPLA_2$s grouped within the sub-group XIA contain a His residue downstream to conserved active site His/Asp dyad, while those grouped into XIB contain Ser or Asn residue. The presence of Gly residue as third Ca^{2+} coordinating amino acid and His residue downstream to His/ASP dyad in $LusPLA_2II$ fits it into sub-group XIA. The presence of Leu residue as third Ca^{2+} coordinating amino acid and Ser residue downstream to His/ASP dyad in $LusPLA_2I$ (Fig. 1) appropriately fits it into sub-group XIB. In addition to this, the Ala residue of the conserved catalytic site motif (DACCxxHDxCV) is replaced by Ser residue (DSCCxxHDxCV) in $LusPLA_2I$. Such a replacement has also been observed in Arabidopsis (Lee et al. 2003).

Plant PLA_2-like enzymes also contain the signature phospholipase A_2 (PA2c) domain comprising the Ca^{2+} binding loop and active site motif but show minor deviation from features of a standard $sPLA_2$. They contain only six Cys residues instead of 12 Cys residues, a feature similar to animal $sPLA_2$ that contain 5–8 Cys residues. Among the catalytic active motif, the His/Asp dyad is conserved, but the amino acid downstream to the His/Asp dyad is an Asp residue instead of His/ser/Asn residue (Fig. 1). Conservation of the Asp residue in the downstream region of His/Asp dyad and presence of 6 Cys residues account for the similarity of plant PLA_2_like proteins with animal $sPLA_2$ sequence. This is in line with

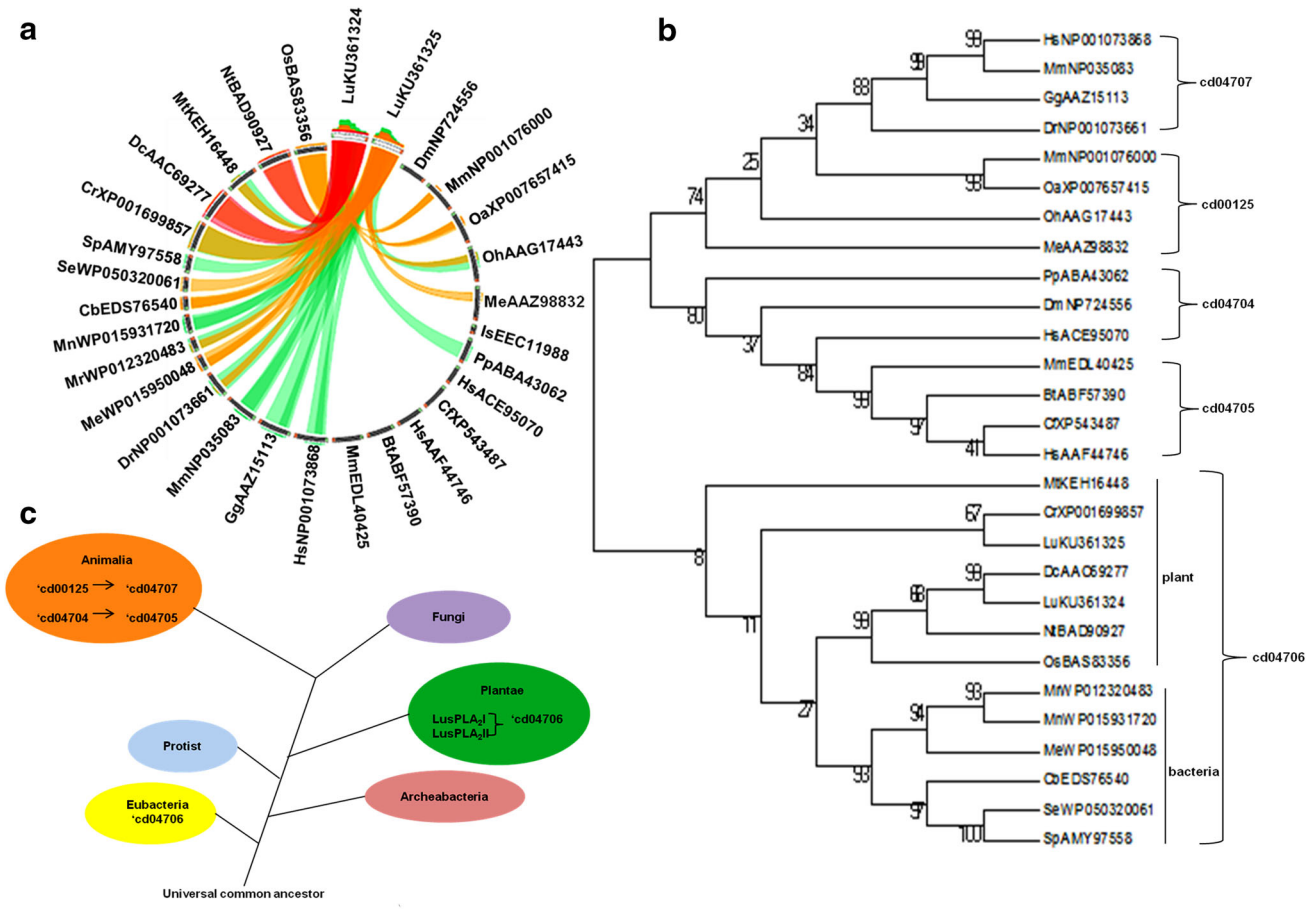


Fig. 5 Similarity and phylogenetic analysis of conserved domains of flax and representative $sPLA_2$s of each sub-family of cd00618 collection. **a** % similarity among the conserved domain region of flax and representative $sPLA_2$s of different sub-families of cd00618 collection. % similarity is represented by color scheme, viz., *blue color* represents 25% similarity (worst), *green color* represents 50% similarity, *orange color* represents 75%, and *red color* represents more than 75% and similarity. **b** Phylogenetic tree depicting the evolutionary relationship among representative $sPLA_2$ members of cd00125, cd04704, cd04705, cd04706, and cd04707 sub-families. The evolutionary analysis reveals that representative $sPLA_2$s from different cd00618 sub-families evolved from a common ancestor. Plant and bacterial $sPLA_2$s of cd04706 sub-family are grouped together. **c** Position of flax $sPLA_2$s in the existing model of evolution of $sPLA_2$s in the six kingdoms of life, viz., Animalia, Fungi, Plantae, Protista, Archeabacteria, and Eubacteria, belonging to different sub-families of cd00618 collection

the placement of PLA_2_like proteins closer to snake venom in phylogenetic tree (Fig. 3).

Secretory phospholipase A_2 of organisms from all kingdoms of life have been classified into two families based on conserved domains, viz., the cd00618 and pfam00068 collection. Since the pfam00068 collection consists of PLA_2s from bacteria, fungi, and animals with no representatives from plants, the members of this collection were not included in the study. The cd00618 collection contains $sPLA_2$s belonging to bacteria, fungi, plants, and animals. Plant $sPLA_2$s are grouped in cd04706 collection along with $sPLA_2$s of bacteria belonging to phylum Proteobacteria and Firmicutes.

Amino acid sequence comparison in the conserved domain region of flax and bacterial $sPLA_2$ revealed that despite of being assigned to same cd04706 sub-family, they showed minimal similarity among them (Fig. 4a). $LusPLA_2I$ showed high similarity (42%) to *Clostridium botulinum* of phylum firmicutes (Supplementary Table 3), whereas $LusPLA_2II$ showed similarity (42%) to *Methylobacterium nodulans* of phylum Alphaproteobacteria (Supplementary Table 4). While the $sPLA_2$ of flax contained 12 Cys residue that can form six intramolecular disulphide bonds, $sPLA_2$ of bacteria (phylum Proteobacteria and Firmicutes) contain only 4–5 Cys residues capable of forming two disulphide bonds. The conserved His/Asp dyad of catalytic active site was conserved in $sPLA_2$s of flax and bacteria. The calcium binding residues, Tyr, Gly, and Asp, are also conserved among the $sPLA_2$ of cd04706 sub-family (Fig. 4b). These features suggest that $sPLA_2$ of flax is structurally similar to $sPLA_2$ of bacteria. Phylogenetic analysis of $sPLA_2$ from plants and bacteria of

the cd04706 collection suggests that eukaryote and prokaryote sequences share a common ancestor (Fig. 4c). Prokaryotic $sPLA_2$ and plant $sPLA_2$ form distinct groups of their own suggesting that $sPLA_2$ of plant and bacteria evolved from a common ancestor through multiple speciation events. This common ancestor is suggested to be unidentified in the previous reports (Nevalainen et al. 2012).

The analysis of amino acid sequence of conserved domain of flax and representative $sPLA_2$ of cd00125, cd04704, cd04705, cd04706, and cd04707 sub-families revealed that flax $sPLA_2$s shows 70–50% similarity with known plant $sPLA_2$, 45–35% similarity to bacterial $sPLA_2$ and 40–30% similarity to $sPLA_2$ from all other taxa (Fig. 5a; Supplementary Table 5, 6). The phylogenetic analysis of representative $sPLA_2$ of cd00618 collections belonging to prokaryotic taxa, eukaryotic taxa, and flax, revealed that eukaryotic $sPLA_2$ share a common origin with their homologs in bacteria (Fig. 5b). The conservation of amino acid residues involved in catalysis and Ca^{2+} binding suggests that $sPLA_2$ of various cd00618 sub-families share a common ancestor and evolved rapidly by multiple point mutations and gene duplication.

LusPLA_2s fits well in the model of evolution for $sPLA_2$ as described earlier (Nevalainen et al. 2012) (Fig. 5c). As per this model, cd04706 collection belongs to bacteria and plants and is absent in other organisms. The cd04706 collection is a distinct group of its own. The overlap between the already described model of evolution for $sPLA_2$ and phylogenetic analysis of flax $sPLA_2$s with representative of various cd00618 collections supports the placement of flax $sPLA_2$s in the existing model of evolution.

Our investigation was focused on elucidation of the origin and evolution of flax $sPLA_2$. Phylogenetic analysis revealed that flax $sPLA_2$s share their evolutionary origin with bacteria of phylum Alphaproteobacteria and Firmicutes. Flax $sPLA_2$ fits well in the existing system of classification proposed for $sPLA_2$ belonging to different phyla. Similarity and positioning of flax $sPLA_2$ with $sPLA_2$ from other taxa relate their evolution and prevalence in the genome to their primary function of phospholipid metabolism. Occurrence of two distinct isoforms of $sPLA_2$ in flax reflects that they are involved in multitude of roles such as flowering, pollen development, auxin signaling, PIN protein trafficking, and light induced stomatal opening leading to growth and development. $sPLA_2$s are also known to be involved in wounding and pathogen attack and, besides validated genes against abiotic stress (Shivakumara et al. 2017), recently, their involvement in abiotic stress such as salt and drought stress has been enlightened. We speculate that such analysis is important to understand the existence of different isoforms of $sPLA_2$ in flax and their involvement in varied signaling pathways and their cross-talks imparting tolerance for stress survival particularly drought as it is a major constraint (Gupta and Dash 2015) in flax cultivation.

Acknowledgements The present study was supported by in-house project (RPP- Sanction No 7–5/2010-IC-IV) and ICAR-NPTC project to PKD by Indian Council of Agricultural Research (ICAR), New Delhi, India. Financial support in the form of research scholarship to PG is acknowledged.

Author contributions PG and PKD designed the experiments. PG carried out all the experiments and analyzed data. PG and RS wrote the manuscript with contribution from PKD.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest in the publication.

References

Burke JE, Dennis EA (2009) Phospholipase A_2 structure/function, mechanism, and signalling. J Lipid Res 50(Suppl.):237–242

Darzentas N (2010) Circoletto: visualizing sequence similarity with Circos. Bioinformatics 26(20):2620–2621

Dash PK, Rai R (2016) Translating the "Banana genome" to delineate stress resistance, dwarfing, parthenocarpy and mechanisms of fruit ripening. Front Plant Sci 7:1543

Dash PK, Cao Y, Jailani AK, Gupta P, Venglat P, Xiang D, Rai R, Sharma R, Thirunavukkarasu N, Abdin MZ, Yadava DK (2014) Genome-wide analysis of drought induced gene expression changes in flax (*Linum usitatissimum*). GM Crops Food 5:106–119

Dash P, Gupta P, Rai R (2015) Hydroponic method of halophobic response elicitation in flax (*Linum usitatissimum*) for precise down-stream gene expression studies. Int J Trop Agric 33:1079–1085

Fujikawa R, Fujikawa Y, Iijima N, Esaka M (2005) Molecular cloning expression and characterization of secretory phospholipase A_2 in tobacco. Lipids 40:901–908

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59:307–321

Gupta P, Dash PK (2015) Precise method of in situ drought stress induction in flax (*Linum usitatissimum*) for RNA isolation towards down-stream analysis. Anal Agric Res 36:10–17

Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci CABIOS 8:275–282

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780

Keane TM, Creevey CJ, Pentony MM, Naughton TJ, Mclnerney JO (2006) Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. BMC Evol Biol 6:29

Kim JY, Chung YS, Ok SH, Lee SG, Chung WI, Kim IY, Shin JS (1999) Characterization of the full-length sequences of phospholipase A_2 induced during flower development. Biochim Biophys Acta 1489:389–392

Lee HY, Bahn SC, Kang YM, Lee KH, Kim HJ, Noh EK, Palta JP, Shin JS, Ryu SB (2003) Secretory low molecular weight

phospholipase A_2 plays important roles in cell elongation and shoot gravitropism in *Arabidopsis*. Plant Cell 15:990–2002

Lee HY, Bahn SC, Shin JS, Hwang I, Back K (2005) Multiple forms of secretory phospholipase A_2 in plants. Prog Lipid Res 44:52–67

Lee OR, Kim SJ, Kim HJ, Hong JK, Ryu SB, Lee SH, Ganguly A, Cho HT (2010) Phospholipase A_2 is required for PIN-FORMED protein trafficking to the plasma membrane in the *Arabidopsis* root. Plant Cell 22:1812–1825

Mansfeld J, Gebauer S, Dathe K, Ulbrich-Hofmann R (2006) Secretory phospholipase A_2 from *Arabidopsis thaliana*: insights into the three-dimensional structure and the amino acids involved in catalysis. Biochemistry 45:5687–5694

Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C et al (2011) CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res 39:225–229

Mariani ME, Villarreal MA, Cheung F, Leiva EP, Madoery RR, Fidelio GD (2012) *In silico* and *in vitro* characterization of phospholipase A_2 isoforms from soybean (*Glycine max*). Biochimie 94:2608–2619

Nevalainen TJ, Cardoso JC, Riikonen PT (2012) Conserved domains and evolution of secreted phospholipases A_2. FEBS J 279:636–649

Pohnert G (2002) Phospholipase A_2 activity triggers the wound-activated chemical defense in the diatom *Thalassiosira rotula*. Plant Physiol 129:103–111

Posada D, Buckley TR (2004) Model selection and model averaging in phylogenetics: advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. Syst Biol 53:793–808

Reina-Pinto JJ, Voisin D, Kurdyukov S, Faust A, Haslam RP, Michaelson LV, Efremova N, Franke B, Schreiber L, Napier JA, Yephremov A (2009) Misexpression of FATTY ACID ELONGATION1 in the *Arabidopsis* epidermis induces cell death and suggests a critical role for phospholipase A2 in this process. Plant Cell 21:1252–1272

Robert X, Gouet P (2014) Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res 42:320–324

Schaloske RH, Dennis EA (2006) The phospholipase A_2 superfamily and its group numbering system. Biochim Biophys Acta 1761:1246–1259

Scherer GF, Ryu SB, Wang X, Matos AR, Heitz T (2010) Patatin-related phospholipase A: nomenclature, subfamilies and functions in plants. Trends Plant Sci 15:693–700

Shivakumara TN, Sreevathsa R, Dash PK, Sheshshayee MS, Papolu PK, Rao U, Tuteja N, Udayakumar M (2017) Overexpression of Pea DNA Helicase 45 (PDH45) imparts tolerance to multiple abiotic stresses in chili (*Capsicum annuum* L.). Sci Rep 7:2760

Shivaraj SM, Deshmukh RK, Rai R, Bélanger R, Agrawal PK, Dash PK (2017) Genome-wide identification, characterization, and expression profile of aquaporin gene family in flax (*Linum usitatissimum*). Sci Rep 7:46137

Six DA, Dennis EA (2000) The expanding superfamily of phospholipase A_2 enzymes: classification and characterization. Biochim Biophys Acta 1488:1–19

Stahl U, Ek B, Stymne S (1998) Purification and characterization of a low-molecular weight phospholipase A_2 from developing seeds of elm. Plant Physiol 117:197–205

Stahl U, Lee M, Sjodahl S, Archer D, Cellini F, Ek B, Iannacone R, MacKenzie D, Semeraro L, Tramontano E, Stymme S (1999) Plant low-molecular weight phospholipase A_2S (PLA_2s) are structurally related to the animal secretory PLA_2s and are present as a family of isoforms in rice (*Oryza sativa*). Plant Mol Biol 41:481–490

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725–2729

Verlotta A, Liberatore MT, Cattivelli L, Trono D (2013) Secretory phospholipases A_2 in durum wheat (*Triticum durum* Desf.) gene expression, enzymatic activity, and relation to drought stress adaptation. Int J Mol Sci 14:5146–5169

Wang X (2001) Plant phospholipases. Annu Rev Plant Physiol Plant Mol Biol 52:211–231

Wang Z, Hobson N, Galindo L, Zhu S, Shi D, McDill J, Yang L, Hawkins S, Neutelings G, Datla R, Lambert G (2012) The genome of flax (*Linum usitatissimum*) assembled de novo from short shotgun sequence reads. Plant J 72:461–473