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<sup>1</sup> · Raman Saini<sup>2</sup> · Prasanta K. Dash<sup>1</sup>

Received: 29 March 2017/Accepted: 12 June 2017/Published online: 1 July 2017 © Springer-Verlag GmbH Germany 2017

Abstract Phospholipase A<sub>2</sub> (PLA<sub>2</sub>) belongs to class of lipolytic enzymes (EC 3.1.1.4). Lysophosphatidic acid (LPA) and free fatty acids (FFAs) are the products of PLA<sub>2</sub> 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<sub>2</sub> encoding genes, viz., LusPLA2I and LusPLA2II (Linum usitatissimum secretory phospholipase A<sub>2</sub>). Molecular simulation of LusPLA<sub>2</sub>s with already characterized plant sPLA<sub>2</sub>s revealed the presence of conserved motifs and signature domains necessary to classify them as secretory phospholipase A2. Phylogenetic analysis of flax sPLA2 with representative sPLA<sub>2</sub>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<sub>2</sub> in flax. Comparative genomic analysis of two LusPLA<sub>2</sub>s with earlier reported plant sPLA<sub>2</sub>s, based on their gene architectures, sequence similarities, and domain structures are presented elucidating the uniqueness of flax sPLA<sub>2</sub>.

**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

<sup>1</sup> ICAR-National Research Centre on Plant Biotechnology, Pusa Campus, New Delhi 110012, India

<sup>2</sup> Department of Biotechnology, Kurukshetra University, Kurukshetra 136118, India Keywords Secretory phospholipase  $A_2 \cdot$  Phylogenetic analysis  $\cdot$  Gene duplication/deletion  $\cdot$  Conserved domain  $\cdot$  Ancestor

# Introduction

Phospholipase A<sub>2</sub> 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<sub>2</sub> 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 PLA2 has gained impetus. PLA2s have been classified into different classes based on function, molecular structure, and conserved domain annotation. On the basis of their structure and function, PLA<sub>2</sub>s are divided into five classes: (1) secretory phospholipase  $A_2$ (sPLA<sub>2</sub>s), (2) calcium-dependent cytosolic phospholipase A<sub>2</sub> (cPLA<sub>2</sub>s), (3) calcium-independent cytosolic phospholipase A<sub>2</sub> (iPLA<sub>2</sub>s), (4) platelet-activating factor acetyl hydrolases (PAF-AH), and (5) lysosomal phospholipase A<sub>2</sub> (Six and Dennis 2000). Based on molecular structure, PLA<sub>2</sub>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 A2. Group IV is assigned to



calcium-dependent cytosolic phospholipase A<sub>2</sub>, group VI is assigned to calcium-independent cytosolic phospholipase A<sub>2</sub>, 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<sub>2</sub> are grouped into two major families viz, cd00618 (PLA<sub>2</sub> like) and pfam00068 (Phospholipase A<sub>2</sub> 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<sub>2</sub>\_bee\_venom\_like: phospholipase A<sub>2</sub>, similar to bee venom; (2) the cd04705 PLA2\_group\_III\_like: PLA2, similar to group III PLA<sub>2</sub>; (3) the cd04706 PLA<sub>2</sub>\_plant: plant specific phospholipase A2; (4) the cd04707 octoconin\_90: phospholipase A2\_like domain present in octoconin\_90 and octoconin\_95; and (5) the cd00125 PLA<sub>2</sub>c: secretory and cytosolic phospholipase A2". Furthermore, the pfam00068 collection is sub-divided into two distinct sub-groups; (1) pfam06951 PLA<sub>2</sub>G12, group XII secretory phospholipase A<sub>2</sub> precursor and (2) pfam09056 phospholip\_A2\_3: prokaryotic phospholipase A2 domain found in PLA<sub>2</sub> of bacteria and fungi (Nevalainen et al. 2012).

Among all PLA<sub>2</sub>s, secretory phospholipase A<sub>2</sub> are most extensively studied. Plant sPLA<sub>2</sub>s are placed in group XI and are further classified into sub-groups XIA and XIB (Six and Dennis 2000). The plant sPLA<sub>2</sub>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<sub>2</sub>s of Alphaproteobacteria of phylum Proteobacteria such as *Methylobacterium extorquens* and phylum Firmicutes such as *Bacillus cereus* and *Streptococcus equi*. Few patatin-like PLA<sub>2</sub>s 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<sub>2</sub>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 (YGKYCGxxxGC), 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<sub>2</sub>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<sub>2</sub>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<sup>2+</sup> 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<sub>2</sub> in flax and their conservation among plant PLA<sub>2</sub> superfamily. We bioinformatically explored on identification of conserved domains in representative collection of sPLA<sub>2</sub>s belonging to annotated protein sequences from ten different plant families, bacteria, and animals. Subsequently, position of flax sPLA<sub>2</sub>s within the existing model of origin and evolution of secretory phospholipase A<sub>2</sub> were compared to representative sPLA<sub>2</sub>s of each sub-family of cd00618 collection for similarity and evolution analysis.

### Methods

#### Sequence retrieval from database

The amino acid sequences for sPLA<sub>2</sub>s were retrieved from the publicly available protein repository of the NCBI database (http://www.ncbi.nlm.nih.gov). sPLA2 annotated sequences from ten different plant families, viz., Brassicaceae, Caryophyllaceae, Euphorbiaceae, Fabaceae, Mal-Poaceae, Rutaceae, Saliaceae, Solanaceae vaceae, including Linaceae, and venom were used for analysis. The amino acid sequence of secretory phospholipase A2 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<sub>2</sub>s were identified using the NCBI Conserved Domains Database CDDv3.14-47363 PSSMs annotation using default parameters. The conserved domain region of representative sPLA<sub>2</sub> 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<sub>2</sub>s with known plant sPLA<sub>2</sub>s and bacterial sPLA<sub>2</sub> belonging to phylum Proteobacteria and Firmicutes (cd04706) and representative sPLA<sub>2</sub> 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<sub>2</sub>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<sub>2</sub>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 gammadistributed 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<sub>2</sub>s and representative sPLA<sub>2</sub>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 sPLA2s in the existing model of origin and evolution of sPLA<sub>2</sub> 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<sub>2</sub> have been isolated from a number of organisms including bacteria, fungi, and plants, but information from flax is lacking. Among plants, sPLA<sub>2</sub>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<sub>2</sub>s have also been isolated from plants (Scherer et al. 2010). Initially, sPLA<sub>2</sub>s were classified into only two groups GI and GII, but later on, sPLA<sub>2</sub>s were classified into GI/II/V/X. sPLA2 superfamily has expanded over past years due to identification of conserved domains from various organisms and now sPLA<sub>2</sub>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<sub>2</sub> in LusPLA<sub>2</sub>I and LusPLA<sub>2</sub>II. 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<sub>2</sub>s were submitted to Circoletto (Darzentas 2010) to deduce the similarity among them (Fig. 2a). The amino acid sequences of LusPLA<sub>2</sub>I and LusPLA<sub>2</sub>II revealed 34% identity among them. LusPLA<sub>2</sub>I showed only 29% identity with venom sPLA<sub>2</sub>. LusPLA<sub>2</sub>I showed maximum similarity to *Arabidopsis lyrata* sPLA<sub>2</sub> (70%), whereas LusPLA<sub>2</sub>II showed maximum similarity to *Oryza sativa* sPLA<sub>2</sub> (65%) (Supplementary Table 1, 2). Sequence similarity of LusPLA<sub>2</sub>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<sub>2</sub>s and known plant sPLA<sub>2</sub>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<sub>2</sub>s belonging to 68 taxa



XIB         Important international internatinternatintereacondular internate international international inte			Ca <sup>2+</sup> binding	loop Catalytic motif	
XIB         Image: Construction of the consthe construction of the construction of the constru		LusPLA2T	LEYCKYCCULYSC	CRORTPODOLDSCONNEDLCV	
XIB         Rescuence         PLANKY COLL 12 GC         COUNT OF COLL 22 GC <thcount 22="" coll="" gc<="" of="" th="">         COUNT OF COUNT OF</thcount>		Alefe60001	FLRYGKYCGLLYSG.	CPGERPCDGLDSCOMKHDACV	OSKNNDYLSOECSONFI
YIB         PEAMOR 74.1 BACCY2525         FLATCHY CCLL/12 GC         CPC BILL DC CMAIN 10.0 (0.0 KMM)         DY 16 DIES OF CCL MARK 10.0 (0.0 KMM) <thdy 16="" dies="" of<br="">CCL MARK 10.0 (0.0 KMM)         DY 16 DI</thdy>		AtAEC06025	FLRYGKYCGLLYSG.	CPGERPCDCLDSCCMKHDACV	OSKNNDYLSOECSOKFI
XIB         Image: State of the state		P£ADM67341	FLRYCKYCGLLYSC.	CPGERPCDGLDSCGMNHDACV	QSKNNDYLSQECSQKFI
XIB         CHEMP 24:12 TELOT 23:05 CHEMP 23:12 CHEMP 23:1		BnCDY25852	LLRYCKYCGLLYSC.	CPGERPCDGLDSCGMKHDACV	QSKNNDYLSQECSQKFI
XIB         TELEDY 32094 CREMEND323 CREMEND32		CSADF55749	FLRYCKYCGLLYSC.	CPGERPCDCLDACOMKHDACV	OAKNNDYLSOECSKNFI
XIB         PEREPS2512 GARDING 233 GARDING 234 GARDING 234 HEADSTITUT Common Commo		TOEOV32096	PLAYCKYCCI I YSC	CPCZEPCDCLDACCMERDACY	ONENN DYLSOPCSONET
XIB         Carbon 0323 Carbon 02254 Carbon 02257 Carbon 02257 C		Pt EEE 92512	FLBYCKYCCLLYSC	CPCENPCDCLDACCMEHDACI	OSKNN SYLSOFCSONFL
XIB         CREMIN 229.19 CREMIN 2254 MEASO 71.67 MEASO 71		GSKEN40323	LLRYGKYCGLLYSG.	CPGERPCDGLDACOMKHDOCV	SAKNN DYLSOECSOTFI
XIB         Comm00:254 (MALSS7):167 (M		GsKEN32919	LLRYGKYCGLLYSG.	CPGERPCDCLDACCMKHDOCV	SAKNNDYLSOECSOTFI
XIB         NUMBER 17167         LLAYGYYCC/LLYSG         CPCARP CIGL/ACCM HED/CVTVINN         DY SOUGH OF SOUGH OF COMMUNICATION OF SUPERIOR OF SUPERIN		CmNM001254	LLRYGKYCGLLYSG.	CPGERPCDGLDACGMKHDOCV	SAKNNDYLSQECSQTFI
XIB         Interview         Inte		MEAES77167	LLRYGKYCGLLYSG.	CPGENPCDGLDACGMKHDOCV	TVKNNDYLSOOCSOTFI
XIB         Common 253         LLNYEYYCLIYYS         CP CONNECTICUDACCM NUMERCYCLARK         DY 180 DESC DY 180 DES		CarEN16699	LEXCENCELLING	CROENPEDGLOACONNHOUSY	
XIB         DeALC 5927         LL RYCKYCCULYSG         PPCIM PC CLUAC (MSITIAL CV) SINCE         DYL S DYL SINCE           TLACD 55021         FL RYCKYCCULYSG         PPC SINC CP FL ALC (MSITIAL CV) SINCE         DYL S DYL SINCE           OBALT 2543         LL RYCKYCCULYSG         CPC SINC CP FL ALC (MSITIAL CV) SINCE         DYL S TENNE           OBALT 2543         LL RYCKYCCULYSG         CPC SINC CP FL ALC (MSITIAL CV) SINCE         DYL S TENNE           MART 2543         LL RYCKYCCULYSG         CPC SINC CP FL ALC (MSITIAL CV) SINCE         DYL S TENNE           MART 2543         LL RYCKYCCULYSG         CPC SINC CONTRACT (MSITIAL CV) SINCE         DYL S TENNE           MART 25505         LL SYCKYCCULYSG         CPC SINC CONTRACT (MSITIAL CV) SINCE         DYL S TENNE           MART 255020         VIR NYCKYCCULYSG         CPC SINC CONTRACT (MSITIAL CV) SINCE         DYL S TENNE           MART 255020         VIR NYCKYCCULYSG         CPC SINC CONTRACT (MSITIAL CV) SINCE         DYL S TENNE           MART 25502020         VIR NYCKYCCULYSG         CPC SINC CONTRACT (MSITIAL CV) SINCE         DYL S TENNE           MART 255022         LANYCKYCCULYSG         CPC SINC CONTRACT (MSITIAL CV) SINCE         DYL S TENNE           MART 255131         LANYCKYCCULYSG         CPC SINC CONTRACT (MSITIAL CV) SINCE         DYL S TENNE           MART 255131         L		CmNM001253	LLRYCKYCGLLYSG.	CPGENPCDGLDACGNYHDKCV	OAKNN
PLACE         PLACE <th< th=""><th>XIB</th><th>DcAAC69277</th><th>LLRYCKYCGLLYSC.</th><th>CPGEMPCDGLDACGMSHDACV</th><th>OSKCDDYLSOECSNELI</th></th<>	XIB	DcAAC69277	LLRYCKYCGLLYSC.	CPGEMPCDGLDACGMSHDACV	OSKCDDYLSOECSNELI
TAGD 95021         FLAC 95012         FLAC 95012         FLAC 95		TuEMS53789	FLRYCKYCGILYSC.	CPGERPCDPLDACOMHHDNCV	LVK.NDYLSTECNEGLL
VIANCIOSE28		TLACD95021	FLRYCKYCGILYSC.	CPGERPCDPLDACOMHHDNCV	LVK.NDYLSTECNECIL
VIA         VIA <th></th> <th>OsAK105828</th> <th>LLRYGKYCGILYSG.</th> <th>CPGEQPCDELDACOMHHDNCV</th> <th>QAK.NDYUSTACNEEUL</th>		OsAK105828	LLRYGKYCGILYSG.	CPGEQPCDELDACOMHHDNCV	QAK.NDYUSTACNEEUL
XIA         TENDED3105         TENTER TO DIVER         EPERAPPECTAL CHARGE CONTROL CHARGE CONTROL CHARMED CO		OSBAF12948	ILRIGRICGILISC.	CPGEOPCDELDACOMHHDNCV	OAK NDTUSTACNEEUL
NLARTÓSIOS         LLSYGXYCCUNYEG         PORTPORTUDATCONNECT/CANADY         PORTPORTUDATCONNECT/CANADY         PORTPORTUDATCONNECT/CANADY           THAG095020         YLAYGYYCCULYEG         PORTPORTUDATCONNECT/CANADY         PYTACONEC           Damacosoci         YLAYGYYCCULYEG         PORTPORTUDATCONNECT/CANADY         PYTACONEC           Damacosoci         YLAYGYYCCULYEG         PORTPORTUDATCONNECT/CANADY         PYTACONE           Damacosoci         YLAYGYYCCULYEG         PORTPORTUDATCONNECT/ANDAC         PYTACONE           DAMATISSA         LLAYGYYCCULYEG         PORTPORTUDATCONNECT/ANDAC         PYTALMENCO           DAMACOSOCI         MARTORYCCULYEG         PORTPORTUDATCONNECT/ANDAC         PYTALMENCO           DAATTISSA         LLAYGYYCCULYEG         PORTPORTUDATCONNECTOR WINDECOTTEN         YTATENNE           DAATTISSA         MARTORYCOVYTT         PORTPORTUDATCONNECTOR WINDECOTTEN         YTATENNE           DAATTISSA         MARTORYCOVYTT         PORTPORTUDATCONNECTOR WINDECOTTEN         YTATENNE           TALGOPSO22         MARTORYCOVYTT         PORTPORTUDATCONNECTOR WINDECOTTEN         YTATENNE           TALGOPSO22         MARTORYCOVYTT         PORTPORTUDATCONNECTOR WINDECOTTEN         YTATENNE           TALGOPSO22         MARTORYCOVYTT         PORTPORTUDATCONNECTOR         YTATENNE		ZmACG37910	FLRYCKYCGIMYSC	CPCEPPCDALDACOMHHDNCV	OAK M. DYLSTACNEALL
NEBAD90927		MtAET05105	LSYCKYCGINYSC.	CPGETPCDDLDTCCMNHDLCV	KAKNYDYLSOECSOTFI
TLAG095020        , LARGKYCCILLYSG         CPGIKPCCALDACCMVIDBEVCHANNA		NtBAD90927	FLRYGKYCGVLYSG.	CPGEOPCDGLDACGMKHDLCI	QRKGNNYLNLECNONFL
PLA2_like       PLASTOS 2023       LLPYCKYCCG LLYSG       PCCMPCTALLACCMVEDHC/D TMMN       DYNTROME         SACT17330       LLPYCKYCCG LLYSG       PCCMPCTALLACCMVEDHC/D TMMN       DYNTROME         SACT17224       LPYCKYCCG LYSG       PCCMPCTALLACCMVEDHC/D TMMN       DYNTROME         SACT17230       LPYCKYCCG LYSG       PCCMPCTALLACCMVEDHC/D TMMN       DYNTROME         SACT1724       LPYCKYCCG LYSG       PCCMPCTALLACCMVEDHC/D TMMN       DYNTROME         TCAC05022       LPYCKYCCG LYSG       PCCMPCTALLACCMVEDHC/D TMMN       DYNTROME         TCAC05022       LPYCKYCCG LYSG       PCCMPCTALLACCMVEDHC/D TMMN       DYNTROME         CCADF55750       RYCKYCCG LYSG       PCCMPCTALLACCMVEDACCMERA       CLINIX CHEMANON         RCEEFF4118       RYCKYCCG LYSG       PCCMPCTDLLACCK HEADCOCK       CLINIX CHEMANON         TCEO730527       RYCKYCCG LYSG       PCCMPCTDLLACCK HEADCOCK       CLINIX CHEMANON         WEABS77094       RYCKYCCG LYSG       PCCMPCTDLLACCK HEADCOCK       CLINIX CHEMANON         WEABS77095       RYCKYCCG LYSG       PCCMPCTDLLACCK HEADCOCK       CLINIX CHEMANON         MARKENSTROM       RYCKYCCG LYSG       PCCMPCTDLLACCK HEADCOCK       CLINIX CHEMANON         MYLBSD7709       RYCKYCCG LYSG       PCCMPCTDLLACCK HEADCOCK       CLINIX CHEMANON		TLACD95020	VLRYGKYCGILYSC.	CPGENPCDALDACONVHDHCV	AANNNDYLNTCONENLL
XIA       Signature		ZmACC30263	LRYCKYCGILYSG.	CPGENPCDALDACOMVHDHCV	DTHNNDYUNTRONENUL
XIA       Sharti 7224 EAAW 6486 Comparison TCACD95022 CAMPS 6561 CADP55750 CADP5586 CADP5586 CADP5586 CADP5586 CADP5586 CADP5586 CAD5587 CADP5586 CA		GACT17030	IMPYCKYCCURYTC	CPCEVPCDALDACONVRDACY	
XIA       Zmarfefe86       LNRYCRYCCVSTG       CPCEPPCCALDACCM_HDACV0ATD       DY1MMM/NO         Ttacg09502       LNRYCRYCCVSTG       CPCEPPCCALDACCM_HDACV0ATD       DY1MMM/NO         Ttacg09502       LNRYCRYCCVGVSTG       CPCEPPCCALDACCM_HDACV0ATD       DY1MMM/NO         CaADP55750       IRYCRYCCVGVSG       CPCEPPCCALDACCM_HDACV0ATD       DY1MMM/NO         CaADP55750       IRYCRYCCVGVSG       CPCEPPCCALDACCM_HDACV0ATD       DY1MMM/NO         CaADP55750       IRYCRYCCVGVSG       CPCEPPCCALDACCM_HDACV0ATD       DY1MMM/NO         TcD0730527       IRYCRYCCVGVSG       CPCEPPCCDLDACCM_HDACV6K_HDFCVEKK       GLMVXCHE         TcD0730527       IRYCRYCCVGVSG       CPCEPPCCDLDACCM_HDACVEKK       GLMVXCHE         TcD0730529       IRYCRYCCVGVSG       CPCEPPCCDLDACCM_HDACVEKK       GLMVXCHE         TcAD95019       IRYCRYCCVGVSG       CDCEPPCCDLDACCM_HDACVEKK       GLMVXCHE         MEAE977094       SYCRYCCVGVSG       CDCEPPCCDLDACCM_HDACVEKK       GLMVXCHE         MEAE977093       IRYCRYCCVGVSG       CDCEPPCCDLDACCM_HDACVEKK       GLMVXCHE         MEAE977093       IRYCRYCCVGVSG       CDCEPPCCDLDACCM_HDACVEKK       GLMVXCHE         MEAE977093       IRYCRYCCVGVSG       CDCEPPCCDLDACCM_HDACVEKK       GLMVXCHE         ALEE57039       IRYCRYCCGGVGSG <th></th> <th>SACT17224</th> <th>LMRYCKYCCVSYTC.</th> <th>CPGEAPCDAIDACONLHDACV</th> <th>OATDN DYLNLWCNOSIL</th>		SACT17224	LMRYCKYCCVSYTC.	CPGEAPCDAIDACONLHDACV	OATDN DYLNLWCNOSIL
Tubes 55:17      LMRTCKYCCV/SYTG      CPCEPPCDALDACCMLPDACVQATD      DYIINMMCNO;         LusPLA2II       IRTCKYCCV/SYTG      CPCEPPCDD.DACCKIENDACVQEN      DYIINMMCNO;         LusPLA2II       IRTCKYCCV/GYTG      CPCEPPCDD.DACCKIENDACVQEN      DYIINMCNO;         LusPLA2II       IRTCKYCCV/GYGG      CPCEPPCDD.DACCKIENDACVGEN      DYINMCNO;         RATEP 48118       IRTCKYCCV/GYGG      CPCEPPCDD.DACCKIENDACVENEN      DYINMCNO;         TubDA009026       IRTCKYCCV/GYGG      CPCEPPCDD.DACCKIENDACVENEN      CILDVCGENEN         TubDA009026       IRTCKYCCV/GYGG      CPCEPPCDD.DACCKIENDACVENEN      CILDVCGENEN         TubDA00926       IRTCKYCCV/GYGG      CPCEPPCDD.DACCKIENDACVENEN      CLLDVCGENEN         MTAB277094      SSTCCTVTG      CPCEPPCDD.DACCRIENCUDENCKGEN      CHVCHK         MtAB277096       IRTCKTCCVCHYG      CPCEPPCDD.DACCRIENCUDENCKGEN      CHVCHK         MtAB277096       IRTCKTCCCVCHYG      CPCEPPCDD.DACCKIENDCVCKEN      CHVCHK         MtAB277096       IRTCKTCCCCHYGG      CPCEPPCDD.DACCKIENDCVCKEN      CHVNCHK         MtAB277096       IRTCKTCCCCHYGG      CPCEPPCDD.DACCKIENDCVCKEN      CMVNCHK         MtAB2877093       IRTCKTCCCCHYGG		ZmAFW64886	LMRYCKYCGVSYTC.	CPGEAPCDALDACOMLHDACV	QATDNDYLNMWCNOSIL
XIA       Ttacop5022       LMRYCKYCCVSYTG       CPCEMPCDALDACCK HEDCVG(M, C)       DIJINMMCNO, C)         RodEP4211       INYCKYCCVGNSG       CPCEMPCDDLDACCK HEDCVGNK, C)       DIJINMCNO, C)         RodEP4818       INYCKYCCVGNSG       CPCEMPCDDLDACCK HEDCVGNK, C)       DIJINMCNO, C)         RodEP4818       INYCKYCCVGNSG       CPCEMPCDDLDACCK HEDCVGNK, C)       DIJINMCNO, C)         NtBAD9026       INYCKYCCVGNSG       CPCEMPCDDLDACCK HEDCVGNK, C)       DINVCHE         TtacD95019       INYCKYCCVGNSG       CPCEMPCDDLDACCK HEDECVGNK, C)       DINVCHE         Samm63876       INYCKYCCVGNSG       CDCEMPCDDLDACCK HEDECVGNK, C)       DINVCHE         OSBAT0536       INYCKYCCVGNSG       CDCEMPCDDLDACCK HEDECVGNK, C)       DINVCHE         MtAES77094       SYCCTHYTG       CPCEMPCDDLDACCK HEDECVGNK, C)       DINVCHE         MtAES77093       INYCKYCCVGNSG       CPCEMPCDDLDACCK HEDECVGNK, C)       DINVCHE         <		TuEMS65617	LMRYGKYCGVSYTC.	CPGEPPCDALDACOMLHDACV	QATDDYLNMWCNQSLL
XIA       INTERCOLUMNO       INTERCOLUMNO       INTERCOLUMNO       INTERCOLUMNO       INTERCOLUMNO         XIA       INTERCOLUMNO       INTERCOLUMNO <t< th=""><th></th><th>TEACD95022</th><th>LMRYCKYCGVSYTC.</th><th>CPGEPPCDALDACOMLHDACV</th><th>QATDDYLNMWCNQSUL</th></t<>		TEACD95022	LMRYCKYCGVSYTC.	CPGEPPCDALDACOMLHDACV	QATDDYLNMWCNQSUL
XIA       REEF48118 TEBC30527 TEAC95019 INTCACCUCWSC       INTCRYCCYCWSC       CPCRPCDDLDACCKIEDXCWSK       CILINUXCHER TEAC95019 INTCRYCCYCWSC       CPCRPCDDLDACCKIEDXCWSK       CILINUXCHER SUBJECT         XIA       ALEF74917 ALEF77096 INTCRYCCYCWSC       CPCRPCDDLDACCRBDRCVCKKK       CILINUXCHER SUCKCYCWSC       CDCRPCDDLDACCRBDRCVCKKK       CILINUXCHER SUCKCYCWSC         XIA       ALEF76924 INTCRYCCYCWSC       CDCRPCDDLDACCRBDRCVCKKK       CILINUXCHER SUCKCYCWSC       CDCRPCDDLDACCRBDRCVCKKK       CILINUXCHER SUCKCYCWSC         XIA       ALEF7693 INTCRYCCYCWSC       CPCRPCDDLDACCRBDRCVCKKK       CILINUXCHER SUCKCYCWSC       CPCRPCDDLDACCRBDRCVCKKK       CILINUXCHER SUCKCYCWSC         XIA       ALEF7693 INTCRYCCYCWSC       CPCRPCDDLDACCRBDRCVCKKK       CMTHVXCHER SUCKCYCWSC       CPCRPCDDLDACCRBDRCVCKKK       CMTHVXCHER SUCKCYCWSC         XIA       ALEF76286 INTCRYCCYCWSC       CPCRPCDDLDACCMABDCVCKKK       CMTHVXCHER SUCKCYCWSC       CPCRPCDDLDACCMABDCVCKKK       CMTHVXCHER SUCKCYCWSC         XIA       ALEF6531 INTCRYCCYCWSC       CPCRPCDDLDACCMABDCVCKKK       CMTHVXCHER SUCKCYCWSC       CPCRPCDDLDACCMABDCVCKKK       CMTHVXCHER SUCKCYCWSC         ALEF6531 ALEF65330 INTCRYCCYCCYCCCCCCCCCCCC       CPCRPCDDLDACCMABDCVCKKK       CMTHVXCHER SUCKCYCWSC       CMTHVXCHER SUCKCYCKKKC       CMTHVXCHER SUCKCYCKKKC       CMTHVXCHER SUCKCYCKKKC       CMTHVXCHER SUCKCYCKKKK       CMTHVXCHER SUCKCYCKKKK       CMTHVXCHER SUCKC		Cappe55750	LEYCKYCCVCKSC	CPCERPCDDLDACOKTHDDCV	GENGUTNIKCHERPK
XIA       TGEOT30527 NEEAD9026 TAYGRYCGVCWSG CPCEXPCDLDACCKINDECVEXX NEAD9026 TAYGRYCGVCWSG CECEPCDLDACCREDBCVCXXX CHAC95019 ZAAT#6376 MAEE77094 SYCCIPCVCWSG CDCEXPCDLDACCREDBCVCXXX MAEE77094 SYCCIPCVCWSG CDCEXPCDLDACCREDBCVCXXX CMXSVXCHEI SYCCIPCVCWSG CDCEXPCDLDACCREDBCVCXXX CMXSVXCHEI SYCCIPCVCWSG CDCEXPCDLDACCREDBCVCXXX CMXSVXCHEI MAEE77094 SYCCIPCVCWSG CPCEPCDLDACCREDBCVCXXX CMXSVXCHEI SYCCIPCVCWSG CPCEPCDLDACCREDBCVCXXX CMXSVXCHEI MAEE77094 SYCCIPCVCWSG CPCEPCDLDACCMAEDCVCXXX CMXSVXCHEI MAEE77093 IXYGXYCGICHSG CPCEPCDLDACCMAEDCVCXXX ALEE06914 IXYGXYCGICHSG CPCEPCDLDACCMAEDCVCXXXX CMXSVXCHXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX		RCEEF48118	IRYCKYCCVCWTC.	CPGENPCDDLDACCKIHDDCV	EKK
XIA       NtBAD90926       IRYCRYCCVCWSG       CPCRKPCDDLDTCCKTHDDCVEKN       CMTYNYCHE         ZmAFW63876       IRYCRYCCVCWSG       CDCREPCDDLDACCRDHDRCVERN       CMMSVKCHE         OsBAF10536       IRYCRYCCVCWSG       CDCREPCDDLDACCRDHDRCVERN       CMMSVKCHE         OsBAF10536       IRYCRYCCVCWSG       CDCREPCDDLDACCRDHDRCVERN       CMMSVKCHE         MtAES77094       ISYCRYCCVCWSG       CDCREPCDDLDACCRDHDRCVERN       CMMSVKCHE         MtAES77095       IXYCRYCCVCWSG       CPCRKPCDDLDACCRDHDRCVERN       CMMSVKCHE         MtAES77096       IXYCRYCCVCWSG       CPCRKPCDDLDACCRCHDRCVERN       CMMSVKCHE         MtAES77095       IXYCRYCCVCWSG       CPCRKPCDDLDACCWERN       CMMSVKCHE         MtAES77096       IXYCRYCCVCWSG       CPCRKPCDDLDACCWERN       CMMSVKCHE         MtAES77095       IXYCRYCCVCWSG       CPCRKPCDDLDACCMENDCVERN       CMMTWKCHKI         MtAES77096       IXYCRYCCVCWSG       CPCREPCDDLDACCMENDCVERN       CMMTWKCHKI         MtAES77091       IXYCRYCCUCHSG       CPCREPCDDLDACCWENDCVERN       CMMTWKCHKI         ALBEC60513       IRYCRYCCUCHSG       CPCREPCDDLDACCWENDCVERN       CMMTWKCHKI         ALBEC6503       IRYCRYCCUCHSG       CPCREPCDDLDACCWENDCVERN       CMMTWKCHKI         ALAEE65634       IRYCRYCCUCHSG       CPCREPCD		TCE0Y30527	IRYCKYCCVCWSC.	CPGENPCDDLDACCKIHDKCV	EKK
XIA       TTAG095019       INTGEFCCVGWSG.       CEGERPCDDLDACGRNHDHCWGKK.       GIMSTECHN         XIA       STGG       INTGEFCCVGWSG.       CDGERPCDDLACGRNHDHCWGKK.       GIMSTECHN         XIA       STGGINTTC.       CPDERPCDLACGRNHDHCWGKK.       GIMSTECHN         XIA       ALEF62286       INTGEFCCVGWSG.       CPGERPCDDLACGRNHDHCWGKK.       GIMSTECHN         MALES77093       INTGEFCCVGWSG.       CPGERPCDDLACGRNHDHCWGKK.       GIMSTECHN         MALES77093       INTGEFCCVGWSG.       CPGERPCDDLACGRNHDHCWGKK.       GIMSTECHN         MALES77093       INTGEFCCVGYSG.       CPGERPCDDLACGRNHDHCWGKK.       GIMSTECHN         MALES77093       INTGEFCCVGYSG.       CPGERPCDDLACGRNHDHCWGKK.       GIMSTECHN         MALES77093       INTGEFCCVGYSG.       CPGERPCDDLACGRNHDNCYLIN.       GIMSTECHN         ALEFF62266       INTGEFCCIGHSG.       CPGERPCDDLACGRNHDNCYLIN.       GIMSTECHN         ALEFF62366       INTGEFCCIGHSG.       CPGERPCDDLACGRNHDNCYLIN.       GIMSTECHN         BacD00314       INTGEFCCIGHSG.       CPGERPCDDLACGRNHDNCYLIN.       GIMSTECHN         BacD014655       INTGEFCCIGHSG.       CPGERPCDDLACGRNHDNCYLIN.       GIMSTECHN         BacD014665       INTGEFCCIGHSG.       CPGERPCDDLACGRNHDNCYLIN.       GIMSTECHN         BacD14060		NLBAD90926	IRYGKYCGVCWSC.	CPGEKPCDDLDTCCKIHDECV	EKNGMTNVKCHEKFK
XIA       CDGRPCDD_LACGRNHDRCWENN       CLMNSVKCHEI         XIA       ALEFA77094       SSYCCI RYTG       CPDINPCDD_LACGRNHDRCWCKF       CHMSVKCHEI         MHAES77094       SSYCCI RYTG       CPDINPCDD_LACGRNHDRCWCKF       CMTNVKCHKI         MHAES77095       IXYGKYCGVCYKG       CPGINPCDD_LACGRNHDRCWCKF       CMTNVKCHKI         MHAES77093       IXYGKYCGVCYKG       CPGINPCDD_LACGRNHDRCWCKF       CMTNVKCHKI         MHAES77093       IXYGKYCGVCYKG       CPGINPCDD_LACGRNHDRCWCKF       CMTNVKCHKI         MHAES77093       IXYGKYCGVCYKG       CPGINPCDD_LACGRNHDRCWLN       CMTNVKCHKI         ALEFG62286       INYGKYCGICHSG       CPGINPCDD_LACGKINDHCYLN       CMTNVSCHKI         ALEFG62286       INYGKYCGICHSG       CPGINPCDD_LACGKINDHCYLN       CMTNVSCHKI         ALEFG42286       INYGKYCGICHSG       CPGINPCDD_LACGKINDHCYLN       CMTNVSCHKI         ALEFG42286       INYGKYCGICHSG       CPGINPCDD_LACGKINDHCYLN       CMTNVSCHKI         ALEFG42865       INYGKYCGICHSG       CPGINPCDD_LACGKINDHCYLN       CMTNVSCHKI         BnCDY0329       INYGKYCGICHSG       CPGINPCDD_LACGKINDNCVDLK       CMTNVSCHKI         ALEFF43665       INYGKYCGICYCG       CPGINPCDD_LACGMINDNCVDLK       CMTNVSCHKI         ALEFF43656       INYGKYCGICYCG       CPGINPCDD_LACGMINDNCVDLK <th></th> <th>TEACD95019</th> <th>IRYCKFCGVCWSC.</th> <th>CEGEEPCDDLDACORDHDHCV</th> <th>CKKCLMSIKCHEKFK</th>		TEACD95019	IRYCKFCGVCWSC.	CEGEEPCDDLDACORDHDHCV	CKKCLMSIKCHEKFK
XIA		ZnAFW63876	IRIGKICGVCWSC.	CDGEEPCDDLDACORDHDRCV	ERKGMMSVKCHERFK
XIA MLAES77096 IKTCRYCCGUTKG. CPGERPCDDIDACCMCHDECVDRF. CMTHVKCHKI MLAES77093 IKTCRYCCGUTKG. CPGERPCDDIDACCMCHDECVDRF. CMTHVKCHKI ALBER62866 IRYCKYCCGUSG. CPGERPCDDIDACCMCHDHCVEIN. CMTHISCHKI ALBER62861 IRYCKYCCGUSG. CPGERPCDDIDACCMCHDHCVEIN. CMTHISCHKI BnCDX82712 IRYCKYCCGUSG. CPGERPCDDIDACCMCHDHCVEIN. CMTHISCHKI BnCDX82712 IRYCKYCCGUSG. CPGERPCDDIDACCMCHDHCVEIN. CMTHISCHKI BnCDX03349 IRYCKYCCGUSG. CPGERPCDDIDACCMCHDHCVEIN. CMTHISCHKI ALBER653 IRYCKYCCGUSG. CPGERPCDDIDACCMCHDHCVEIN. CMTHISCHKI ALBER653 IRYCKYCCGUFG. CPGERPCDDIDACCMCHDHCVEIN. CMTHISCHKI ALBER653 IRYCKYCCGUFG. CPGERPCDDIDACCMCHDHCVEIN. CMTHISCHKI ALBER653 IRYCKYCCGUFG. CPGERPCDDIDACCMCHDHCVEIN. CMTYNCHKI ALBER653 IRYCKYCCGUFG. CPGERPCDDIDACCMCHDHCVEIN. CMTYNCHKI ALBER653 IRYCKYCCGUFG. CPGERPCDDIDACCMCHDNCVDIK. CMTYNCHKI ALAE8563 IRYCKYCCGUFG. CPGERPCDDIDACCMCHDNCVDIK. CMTYNCHKI BnCDY69584 IRYCKYCCGUFG. CPGERPCDDIDACCMTHDNCVDIK. CMTYNCHKI BnCDY4210 IRYCKYCCGUFG. CPGERPCDDIDACCMTHDNCVDIK. CMTYNCHKI BnCDY45584 IRYCKYCCGUFG. CPGERPCDDIDACCMTHDNCVDIK. CMTYNCHKI BnCDY45584 IRYCKYCCGUFG. CPGERPCDDIDACCMTHDNCVDIK. CMTYNCHKI BnCDX97477 AFLS0LFPRYCUTCCPNXSSCKDCGSIVWDORPTINLDHCCTCHDIGYDTHDO. AEXILKADU BnCDY4221 AFLS0LFPRYCUTCCPNXSSCKDCGSIVWDORPTINLDHCCTCHDIGYDTHDO. AEXILKADU BnCDY35249 AFLS0LFPRYCUTCCPNXSSCKDCGSIVWDORPTINLDHCCTCHDIGYDTHDO. AEXILKADU BnCDY32893 AFLS0LFPRYCUTCCPNXSSCKDCGSIVWDORPTINLDHCCTCHDIGYDTHDO. AEXILKADU TCEOY32890 AFLS0LFPRYCUTCCPNXSSCKDCGSIVWDORPTINLDHCCTCHDIGYDTHDO. AEXILKADU TCEOY3289 AFLS0LFPRYCUTCCPNXSSCKDCGSIVWDORPTINLDHCCTCHDIGYDTHDO. AEXILKADU TCEOY3289 AFLS0LFPRYCUTCCPNXSSCKDCGSIVWDRRPTINLDFCCTCHDIGYDTHDO. EXILKADU TCEOY3289 AFLS0LFPRYCUTCCPNXSSCKDCGSIVWDRRPTINLDFCCTCHDIGYDTHDO. EXILKADU MKKEH6448 AFLS0LFPRYCUTCCPNXSSCKDCGSIVWDRRPTINLDFCCTCHDIGYDTHDO. EXILKADU		0584F10536	SSYCCIPYTO	CROSSPCDDLDACORDHDHCV	CKR CMTNVKCHKKRK
XIA       NEAES77093       IKYGKYCGVGYSG       CPGEKPCDD IDACGMAHDDCVGKF       GMTHVKCHKI         ALEF66286       IRYGKYCGIGHSG       CPGEPCDD LDACGMAHDDCVGKF       GMTHVKCHKI         ALEC06913       IRYGKYCGIGHSG       CPGEPCDD LDACGKIHDMCVEIN       GMTHVKCHKI         BnCDX82712       IRYGKYCGIGHSG       CPGEPCDD LDACGKIHDMCVEIN       GMTNISCHKI         BnCDX8349       IRYGKYCGIGHSG       CPGEPCDD LDACGKIHDMCVEIN       GMTNISCHKI         ALEF86563       IHYGKYCGIGHSG       CPGEPCDD LDACGMIHDMCVEIN       GMTNISCHKI         ALEF86564       IRYGKYCGIGYFG       CPGEPCDD LDACGMIHDMCVEIN       GMTVNCHKI         ALCAR9704       IRYGKYCGIGYFG       CPGEPCDD LDACGMIHDMCVD LK       GMTVNCHKI         BnCDY06806       VRYCKYCGIGYFG       CPGEPCDD LDACGMIHDMCVD LK       GMTVNCHKI         BnCDY06954       IRYGKYCGIGYFG       CPGEPCDD LDACGMIHDMCVD LK       GMTYVNCHKI         ALEE85633       AFLSOLFPRYCHYCGPNKSSCKDGGSLVMDORP ID LDACGUTHENCVD LK       GMTYVNCHKI         BnCDY14210       IRYGKYCGIGYFG       CPGEPCDD LDACGMIHENCVD LK       MTYVNCHKI         BnCDY14210       IRYGKYCGIGYFG       CPGEPCD LDACGMIHENCVD LK       MTYVNCHKI         BnCDY14210       IRYGKYCGIGYFG       CPGEPCD LDACGMIHENCVD LK       MTYVNCHKI         BnCDY14210       IRYGKYCGIGYFG		MLAES77096	IKYCKYCCVCYWC.	CPGENPCDDIDACOMONDECV	DRF
XIA       ALEPB62286       IRYGKYCGIGHSG       CPGEPCDDLDACGWHDNCWELN       CMTNISCHKI         ALAEC06913       IRYGKYCGIGHSG       CPGEPCDDLDACGKIHDHCVELN       CMTNISCHKI         BnCDX82712       IRYGKYCGIGHSG       CPGEPCDDLDACGKIHDHCVELN       CMTNISCHKI         BnCDX03349       IRYGKYCGIGHSG       CPGEPCDDLDACGKIHDHCVELN       CMTNISCHKI         BnCDY0349       IRYGKYCGIGHSG       CPGEPCDDLDACGWHDSCVEAN       CMTNISCHKI         ALEP83634       IRYGKYCGIGHSG       CPGEPCDDLDACGWHDNCVEAN       CMTNISCHKI         ALCAB79704       IRYGKYCGIGHSG       CPGEPCDDLDACGWHDNCVEAN       CMTNISCHKI         BnCDY06906       VRYGKYCGIGHFG       CPGEPCDDLDACGWHDNCVEAN       CMTVNCHKI         BnCDY06906       VRYGKYCGIGHFG       CPGEPCDDLDACGWTHDNCVDLK       CMTYVNCHKI         BnCDY06906       VRYGKYCGIGHFG       CPGEPCDDLDACGWTHDNCVDLK       CMTYVNCHKI         BnCDY06906       VRYGKYCGIGHFG       CPGEPCDDLDACGWTHDNCVDLK       MTYVNCHKI         ALAEE85635       IRYGKYCGIGHFG       CPGEPCDDLDACGWTHDNCVDLK       MTYVNCHKI         ALAEE85635       IRYGKYCGIGHFG       CPGEPCDDLDACGWTHDNCVDLK       MTYVNCHKI         BnCDY05984       IRYGKYCGIGHFG       CPGEPCDDLDACGWTHDNCVDLK       MTYVNCHKI         ALAEE85635       AFLSQLFPRYGHYCGPNKSGKDGGSLWDORP IDWL		MtAES77093	IKYGKYCGVCYSG.	CPGENPCDDIDACCMAHDDCV	GKF GMTHVKCHKKFK
PLA2       IRYGKYCGICHSGCPGREPCDDLDACCKIRDBCVELNGMTNISCHKI         BnCDX82712       IRYGKYCGICHSGCPGREPCDDLDACCMIRDBCVEANGMTNISCHKI         BnCDX82712       IRYGKYCGICHSGCPGREPCDDLDACCMIRDBCVEANGMTNISCHKI         BnCDX0339       IRYGKYCGICHSGCPGREPCDDLDACCMIRDBCVEANGMTNISCHKI         AlEF43665       IHYGKYCGICYFGCPGREPCDDLDACCMIRDNCVEANGMTNISCHKI         ALEF43665       IRYGKYCGICYFGCPGREPCDDLDACCMIRDNCVEANGMTYVNCHKI         ALARE05806       VRYGKYCGICYFGCPGREPCDDLDACCMIRDNCVDLKGMTYVNCHKI         BnCDY06806       VRYGKYCGICYFGCPGRPCDGLDACCMIRDNCVDLKGMTYVNCHKI         BnCDY06806       VRYGKYCGICYFGCPGRPCDGLDACCMIRDNCVDLKGMTYVNCHKI         BnCDY06806       VRYGKYCGICYFGCPGRPCDGLDACCMIRDNCVDLKGMTYVNCHKI         BnCDY06806       VRYGKYCGICYFGCPGRPCDGLDACCMIRDNCVDLKGMTYVNCHKI         BnCDY06806       VRYGKYCGICYFGCPGRPCDGLDACCMIRDNCVDLKGMTYVNCHKI         BnCDY06806       VRYGKYCGICYFGCPGRPCDGLDACCCMIRDNCVDLKGMTYVNCHKI         BnCDY06807       IRYGKYCGICYFGCPGRPCDDLDACCMIRDNCVDLKGMTYVNCHKI         BnCDY06806       VRYGKYCGICYFGCPGRPCDDLDACCMIRDNCVDLKGMTYVNCHKI         BnCDY06807       IRYGKYCGICYFGCPGRPCDDLDACCMIRDNCVDLKGMTYVNCHKI         BnCDY069584	XIA	A1EFE62286	IRYGKYCGICHSC.	CPGEEPCDDLDACCMVHDNCV	ELNGMTNISCHKKFQ
PLA2_like PLA3655 PLA2006914 PLA200690 PLA2006900 PLA2006900 PLA2006900 PLA2006900 PLA2006900 PLA2006900 PLA2006900		AtAEC06913	IRYCKYCCICHSC.	CPGEEPCDDLDACCKIHDHCV	ELNCMTNISCHKKFQ
PLA2_like PLA32_like PLA32_like PLA32_like PLA33349 PLA32 PLA3349 PLA3452 PLA3452 PLA345 PLA3 PLA3 PLA3 PLA3 PLA3 PLA3 PLA		BoCDX82712	LEXCEXCCICHSC	CROENPEDDLDACOKIHDHCV	CMTNISCHKKNO
PLA2_like PLA3655 PL32Like PL32Lik		BnCDY03349	LBYCKYCCICHSC	CPCEEPCDDLDACCMVHDNCV	CMINI SCHKKEK
PLA2_like       AtAEE85634       IRYGKYCGICYFG       CPGEPPCDD LDACCMTHDMCVDLK       GMTYVNCHK         AtAEE85635       IRYGKYCGICYFG       CPGEPPCDD LDACCMTHDMCVDLK       GMTYVNCHK         BnCDY06806       VRYGKYCGICYFG       CPGEPPCDD LDACCMTHDMCVDLK       GMTYVNCHK         BnCDY06806       VRYGKYCGICYFG       CPGEPPCDD LDACCMTHDMCVDLK       GMTYVNCHK         BnCDY069584       IRYGKYCGICYFG       CPGEPPCDD LDACCMTHDMCVDLK       GMTYVNCHK         AtAEE85635       IRYGKYCGICYFG       CPGEPPCDD LDACCMTHDMCVDLK       GMTYVNCHK         VenomBA36       SWWD.FADYGCYCGRCGSG       TPVDD LDRCCVHDNCYDLK       GMTYVNCHK         BhCDY0637       ALLSOLFPRYCHYCGPNKSSGKDGCSNVWDQRPIDWLDRCCYCHDICYDTHDO       AEL LKADM         BhCDY14221       AFLSOLFPRYCHYCGPNKSGKDGCSIVWDQRPIDWLDRCCYCHDICYDTHDO       AEM LKADM         BhCDY14221       AFLSOLFPRYCHYCGPNKSGKDGCSIVWDQRPIDWLDRCYCHDICYDTHDO       AEM LKADM         BhCDY14221       AFLSOLFPRYCHYCGPNKSGKDGCSIVWDQRPIDWLDRCYCHDICYDTHDO       AEL LKADM         BhCDY1637       ALLSOLFPRYCHYC		Alefe43665	IHYGKYCGICYFC.	CRGEPPCDDLDACOMTHDNCV	DIK GNTYVNCHKOFK
PLA2_like       AtCAB79704       IRYGKYCGICYFG.       CPGEPPCDDLDACCMTHEDNCVDLK.       GMTYVNCHK         BnCDY6806       VRYGKYCGICYFG.       CPGEXPCDGLDACCMTHEDNCVDLK.       GMTYVNCHK         BnCDY69584       IRYGKYCGICYFG.       CPGEXPCDGLDACCMTHEDNCVDLK.       GMTYVNCHK         AtAEE85635       IRYGKYCGICYFG.       CPGEXPCDGLDACCMTHEDNCVDLK.       GMTYVNCHK         VenomBA36       SWMD.FADYGCYCGRGSG.       IPVDDLDBCCQVHDNCYNEAEKISGCWPYFKTYSPCSQC         VenomBA36       SWMD.FADYCCYCGRGSG.       IPVDDLDBCCQVHDNCYNEAEKISGCWPYFKTYSPCSQC         BnCDY635249       AFLSQLFPRYGHYCGPNWSSGKDGCSLVWDQRPIDWLDHCCYCHDIGYDTHDQ.       AEN.LKADM.         BnCDY6337       ALLSQLFPRYGHYCGPNWSSGKDGCSLVWDQRPIDWLDHCCYCHDIGYDTHDQ.       AEN.LKADM.         BnCDY6337       ALLSQLFPRYGHYCGPNWSSGKDGCSPIWDQRPIDWLDHCCYCHDIGYDTHDQ.       AEN.LKADM.         BnCDY6337       ALLSQLFPRYGHYCGPNWSSGKDGCSPIWDQRPIDWLDHCCYCHDIGYDTHDQ.       AEL.LKADM.         BnCDY6337       ALLSQLFPRYGHYCGPNWSSGKDGCSPIWDQRPIDWLDHCCYCHDIGYDTHDQ.       AEL.LKADM.         BnCDY63249       AFLSQLFPRYGHYCGPNWSSGKDGCSLWWDRPIDWLDHCCYCHDIGYDTHDQ.       AEL.LKADM.         TcG0732893       AFLSQLFPRYGHYCGPNWSSGKDGCSLWWDRPIDWLDHCCYCHDIGYDTHDQ.       AEL.LKADM.         TcG0732892       AFLSQLFPRYGHYCGPNWSSGKDGCSLWWDRRPIDWLDFCCYCHDIGYDTHDQ.       EKL.LKADL.         TcG0732893       AFLSQL		AtAEE85634	IRYGKYCGICYFC.	CPGEPPCDDLDACONTHDNCV	DLK GMTYVNCHKQFK
PLA2_like PLA36 PLA2 PLA36 PLA		AtCAB79704	IRYGKYCGICYFG.	CPGEPPCDDLDACONTHDNCV	DLKGMTYVNCHKQFK
PLA2_like PLA3 PLA3 PLA3 PLA3 PLA3 PLA3 PLA3 PLA3		BnCDY06806	VRYCKYCGICYFC.	CPGERPCDGLDACONTHDNCV	DIKCMTYVNCHKOFO
PLA2_like RCEF3635IRYCKYCGICYFCCPGEPPCDDLDDCCMTHDNCYDLKCMTYVDCHKU venomBA36 SWMDFADYCCYCGRCSGTPVDDLDRCCYVHDNCYNEAEKISGCWPYFKTYSYECSO AtAEE85583 AFLSOLFPRYCHYCGPNWSSCKDCGSNVWDORPIDWLDHCCYCHDICYDTHDOAEI.LKADM BnCDY14221 AFLSOLFPRYCHYCGPNWSSCKDGCSLVWDORPIDWLDHCCYCHDICYDTHDOAEI.LKADM BnCDY14221 AFLSOLFPRYCHYCGPNWSSCKDGCSLVWDORPIDWLDHCCYCHDICYDTHDOAEI.LKADM BnCDY14221 AFLSOLFPRYCHYCGPNWSSCKDGCSLVWDORPIDWLDHCCYCHDICYDTHDOAEI.LKADM BnCDY14221 AFLSOLFPRYCHYCGPNWSSCKDGCSLVWDORPIDWLDHCCYCHDICYDTHDOAEI.LKADM BnCDY1637 ALLSOLFPRYCHYCGPNWSSCKDGCSLWDORPIDWLDHCCYCHDICYDTHDOAEI.LKADM BnCDY163249 AFLSOLFPRYCHYCGPNWSSCKDGCSLWDORPIDWLDHCCYCHDICYDTHDOAEI.LKADM RCEEF36320 CFLSOLFPRYCHYCGPNWSSCKDGCSLWDRRPIDWLDFCCYCHDICYDTHDOAEI.LKADM TCGOY32893 AFLSOLFPRYCHYCGPNWSSCKDGCSLWWDRRPIDWLDFCCYCHDICYDTHDOEKI.LKADL TCGOY32890 AFLSOLFPRYCHYCGPNWSSCKDGCSLWWDRRPIDWLDFCCYCHDICYDTHDOEKI.LKADL TCGOY32890 AFLSOLFPRYCHYCGPNWSSCKDGCSLWWDRRPIDWLDFCCYCHDICYDTHDOEKI.LKADL TCGOY32899 AFLSOLFPRYCHYCGPNWSSCKDGCSLWWDRRPIDWLDFCCYCHDICYDTHDOEKI.LKADL TCGOY32890 AFLSOLFPRYCHYCGPNWSSCKDGCSLWWDRRPIDWLDFCCYCHDICYDTHDOEKI.LKADL TCGOY32890 AFLSOLFPRYCHYCGPNWSSCKDGCSLWWDRRPIDWLDFCCYCHDICYDTHDOEKI.LKADL TCGOY32890 AFLSOLFPRYCHYCGPNWSSCKDGCSLWWDRRPIDWLDFCCYCHDICYDTHDO		BnCDY69584		CRGEFFCDSDDACGDINDACY	MTYVNCHKOPK
PLA2_like RCEEF36320 GFLSOLFPRYCHYCCPNWSSCKDCCSLVWDCRPIDWLDFCCYCHDICYDTHDO		AtAEE85635	IRYGKYCGICYFC.	CPGEPPCDDLDDCCMTHDNCV	DIK GNTYVDCHKOFO
PLA2_like ALAEE85583 AFLSOLFPRYCHYCCPNWSSCKDCCSMVWDORPIDWLDHCCYCHDICYDTHDOAEL.LKADM. BnCDX97477 AFLSOLFPRYCHYCCPNWSSCKDCCSLVWDORPIDWLDHCCYCHDICYDTHDOAEM.LKADM. BnCDY06837 ALLSOLFPRYCHYCCPNWSSCKDCCSLVWDORPIDWLDHCCYCHDICYDTHDOAEM.LKADM. BnCDY06837 ALLSOLFPRYCHYCCPNWSSCKDCCSLVWDORPIDWLDHCCYCHDICYDTHDOAEM.LKADM. BnCDY35249 AFLSOLFPRYCHYCCPNWSSCKDCCSPIWDORPIDWLDHCCYCHDICYDTHDOAEL.LKADM. BnCDY35249 AFLSOLFPRYCHYCCPNWSSCKDCCSLWWDRPIDWLDHCCYCHDICYDTHDOAEL.LKADM. BnCDY35249 AFLSOLFPRYCHYCCPNWSSCKDCCSLWWDRRPIDWLDFCCYCHDICYDTHDOAEL.LKADM. TCEOY32892 AFLSOLFPRYCHYCCPNWSSCKDCCSLWWDRRPIDWLDFCCYCHDICYDTHDOEKL.LKADM. TCEOY32892 AFLSOLFPRYCHYCCPNWSSCKDCCSLWWDRRPIDWLDFCCYCHDICYDTHDOEKL.LKADM. TCEOY32890 AFLSOLFPRYCHYCCPNWSSCKDCCSLWWDRRPIDWLDFCCYCHDICYDTHDOEKL.LKADM. TCEOY32890 AFLSOLFPRYCHYCCPNWSSCKDCCSLWWDRRPIDWLDFCCYCHDICYDTHDOEKL.LKADM. TCEOY32890 AFLSOLFPRYCHYCCPNWSSCKDCCSLWWDRRPIDWLDFCCYCHDICYDTHDOEKL.LKADM. TCEOY32890 AFLSOLFPRYCHYCCPNWSSCKDCCSLWWDRRPIDWLDFCCYCHDICYDTHDOEKL.LKADM. TCEOY32890 AFLSOLFPRYCHYCCPNWSSCKDCCSLWWDRRPIDWLDFCCYCHDICYDTHDOEKL.LKADM. TCEOY32890 AFLSOLFPRYCHYCCPNWSSCKDCCSLWWDRRPIDWLDFCCYCHDICYDTHDOEKL.LKADM. TCEOY32890 AFLSOLFPRYCHYCCPNWSSCKDCCSLWWDRRPIDWLDFCCYCHDICYDTHDOEKL.LKADM. TCEOY32890 AFLSOLFPRYCHYCCPNWSSCKDCCSLWWDRRPIDWLDFCCYCHDICYDTHDOEKL.LKADM. TCEOY32890 AFLSOLFPRYCHYCCPNWSSCKDCCSLWWDRRPIDWLDFCCYCHDICYDTHDO		venomBAA36	SWWDFADYGCYCGRCGSC	TPVDDLDRCCQVHDNCY	NEAEKISGCWPYFKTYSYECSOCTL
PLA2_ like BacDy32497 AFLSQLFPRYCHYCCPNWSSCKDGCSLVWDQRPIDWLDHCCYCHDICYDTHDQAEM.LKADM. BacDy35249 AFLSQLFPRYCHYCCPNWSSCKDGCSLVWDQRPIDWLDHCCYCHDICYDTHDQAEM.LKADM. BacDy35249 AFLSQLFPRYCHYCCPNWSSCKDGCSPIWDQRPIDWLDHCCYCHDICYDTHDQAEL.LKADM. BacDy35249 AFLSQLFPRYCHYCCPNWSSCKDGCSPIWDQRPIDWLDHCCYCHDICYDTHDQAEL.LKADM. BacDy35292 AFLSQLFPRYCHYCCPNWSSCKDGCSLWWDRRPIDWLDYCCYCHDICYDTHDQAEL.LKADM. TCEOY32893 AFLSQLFPRYCHYCCPNWSSCKDGCSLWWDRRPIDWLDYCCYCHDICYDTHDQEKL.LKADM. TCEOY32890 AFLSQLFPRYCHYCCPNWSSCKDGCSLVWDRRPIDWLDFCCYCHDICYDTHDQEKL.LKADM. TCEOY32890 AFLSQLFPRYCHYCCPNWSSCKDGCSLVWDRRPIDWLDFCCYCHDICYDTHDQEKL.LKADM. TCEOY32890 AFLSQLFPRYCHYCCPNWSSCKDGCSLVWDRRPIDWLDFCCYCHDICYDTHDQEKL.LKADM. TCEOY32890 AFLSQLFPRYCHYCCPNWSSCKDGCSLVWDRRPIDWLDFCCYCHDICYDTHDQEKL.LKADM. TCEOY32890 AFLSQLFPRYCHYCCPNWSSCKDGCSLVWDRRPIDWLDFCCYCHDICYDTHDQEKL.LKADM. TCEOY32890 AFLSQLFPRYCHYCCPNWSSCKDGCSLVWDRRPIDWLDFCCYCHDICYDTHDQEKL.LKADM. TCEOY32890 AFLSQLFPRYCHYCCPNWSSCKDGCSLVWDRRPIDWLDFCCYCHDICYDTHDQEKL.LKADM. TCEOY32890 AFLSQLFPRYCHYCCPNWSSCKDGCSLVWDRRPIDWLDFCCYCHDICYDTHDQEKL.LKADM. TCEOY32890 AFLSQLFPRYCHYCCPNWSSCKDGCSLVWDRRPIDWLDFCCYCHDICYDTHDQEKL.LKADM. TCEOY32890 AFLSQLFPRYCHYCCPNWSSCKDGCSLVWDRRPIDWLDFCCYCHDICYDTHDQ	1	AtAEE85583	AFLSQLFPRYCHYCCPNWSSCH	CDGGSMVWDQRPIDWLDHCCYCHDICY	DTHDQAEL.LKADMAFL
PLA2_like PLA2_like		BnCDX97477	AFLSQLFPRYCHYCCPNWSSCH	COGS LVWDORP I DWLDHCOYCHDICY	DTHDOAEM.LKADMAFL
PLA2_like BBCD735249 AFLSQLFPRYCHYCCPNWSSCKDGCSLIWDCRPIDWLDFCCYCHDICYDTHDQAKI_LKADU RCEEF36320 CFLSQLFPRYCHYCCPNWSSCKDCCSLLWDKRPIDWLDFCCYCHDICYDTHDQAKI_LKADU TCD0732893 AFLSQLFPRYCHYCCPNWSSCKDCCSLWWDRRPIDWLDFCCYCHDICYDTHDQEKI_LKADU TCD0732890 AFLSQLFPRYCHYCCPNWSSCKDCCSLWWDRRPIDWLDFCCYCHDICYDTHDQEKI_LKADU TCD0732890 AFLSQLFPRYCHYCCPNWSSCKDCCSLWWDRRPIDWLDFCCYCHDICYDTHDQEKI_LKADU TCD0732890 AFLSQLFPRYCHYCCPNWSSCKDCCSLWWDRRPIDWLDFCCYCHDICYDTHDQEKI_LKADU TCD0732890 AFLSQLFPRYCHYCCPNWSSCKDCCSLWWDRRPIDWLDFCCYCHDICYDTHDQEKI_LKADU TCD0732890 AFLSQLFPRYCHYCCPNWSSCKDCCSLWWDRRPIDWLDFCCYCHDICYDTHDQEKI_LKADU TCD0732890 AFLSQLFPRYCHYCCPNWSSCKDCCSLWWDRRPIDWLDFCCYCHDICYDTHDQEKI_LKADU TCD0732890 AFLSQLFPRYCHYCCPNWSSCKDCCSLWWDRRPIDWLDFCCYCHDICYDTHDQEKI_LKADU TCD0732890 AFLSQLFPRYCHYCCPNWSSCKDCCSLWWDRRPIDWLDFCCYCHDICYDTHDQEKI_LKADU TCD0732890 AFLSQLFPRYCHYCCPNWSSCKDCCSLWWDRRPIDWLDFCCYCHDICYDTHDQ		BBCD114221	AFLSULFFRIGHICGFNWSSG	DCCSPTNDORPTDN:DRCCYCHDIGT	DTHDO ANT TRADAPL
PLA2_like REEEF36320 GFLS0LFPRYCHYCCPNWSSCKDRCSLLWDXRPIDWLDYCCYCHDICYDTHD0EKL.LKADLI TCB0732893 AFLS0LFPRYCHYCCPNWSSCKDCGSLVWDRRPIDWLDFCCYCHDICYDTHD0EKL.LKADLI TCB0732892 AFLS0LFPRYCHYCCPNWSSCKDCGSLVWDRRPIDWLDFCCYCHDICYDTHD0EKL.LKADLI TCB0732899 AFLS0LFPRYCHYCCPNWSSCKDCGSLVWDRRPIDWLDFCCYCHDICYDTHD0EKL.LKADLI TCB0732899 AFLS0LFPRYCHYCCPNWSSCKDCGSLVWDRRPIDWLDFCCYCHDICYDTHD0EKL.LKADLI TCB0732899 AFLS0LFPRYCHYCCPNWSSCKDCGSLVWDRRPIDWLDFCCYCHDICYDTHD0EKL.LKADLI TCB0732890 AFLS0LFPRYCHYCCPNWSSCKDCGSLVWDRRPIDWLDFCCYCHDICYDTHD0EKL.LKADLI TCB0732890 AFLS0LFPRYCHYCCPNWSSCKDCGSLVWDRRPIDWLDFCCYCHDICYDTHD0EKL.LKADLI TCB0732890 AFLS0LFPRYCHYCCPNWSSCKDCGSLVWDRRPIDWLDFCCYCHDICYDTHD0		BnCDY35249	AFLSOLEPRYCHYCCPNWSSC	UDGGSPINDORPIDNLDHCCYCHDICY	DTHDO AEL LKADVARL
TCEOY32893 AFLSOLFPRYCHYCCPNWSSCKDCGSLVWDRRPIDWLDFCCYCHDICYDTHDOEKL.LKADL TCEOY32892 AFLSOLFPRYCHYCCPNWSSCKDCGSLVWDRRPIDWLDFCCYCHDICYDTHDOEKL.LKADL TCEOY32890 AFLSOLFPRYCHYCCPNWSSCKDCGSLVWDRRPIDWLDFCCYCHDICYDTHDOEKL.LKADL TCEOY32889 AFLSOLFPRYCHYCCPNWSSCKDCGSLVWDRRPIDWLDFCCYCHDICYDTHDOEKL.LKADL MKEEH6448 AFLSOLFPRYCHYCCPNWSSCKDCGSLVWDRRPIDWLDFCCYCHDICYDTHDOAKL.LKADL	PLA <sub>2</sub> like	RCEEF36320	<b>GFLSQLFPRYGHYCGPNWSSCH</b>	DRGSLLWDKRPIDWLDYCCYCHDICY	DTHDQAKL.LKSDLEFL
TCEOY32892 AFLSOLFPRYCHYCCPNWSSCKDCCSLVWDRRPIDWLDFCCYCHDICYDTHDOEKLLKADL TCEOY32890 AFLSOLFPRYCHYCCPNWSSCKDCCSLVWDRRPIDWLDFCCYCHDICYDTHDOEKLLKADL TCEOY32889 AFLSOLFPRYCHYCCPNWSSCKDCCSLVWDRPIDWLDFCCYCHDICYDTHDOEKLLKADL MEKEH6448 AFLSOLFPRYCHYCCPNWSSCKDCCSLVWDRPIDWLDYCCYCHDICYDTHDOAKLLKADL	<b></b>	TCEOY32893	AFLSQLFPRYCHYCCPNWSSCH	CDGGSLVWDRRPIDWLDFCCYCHDIGY	DTHDOEKL.LKADLAFL
TCEOY32890 AFLSQLFPRYCHYCCPNWSSCKDGCSLVWDRRPIDWLDFCCYCHDICYDTHDQEKLLLKADL TCEOY32889 AFLSQLFPRYCHYCCPNWSSCKDGCSLVWDRPIDWLDFCCYCHDICYDTHDQEKLLLKADL MtKEH16448 AFLSQLF <u>PRYCHYCCPNWSS</u> CKDGCSL <u>VWDK</u> R <u>PIDWLD</u> YCCYCHDICYDTHDQAKLLLKADL		TCEOY32892	AFLSQLFPRYCHYCCPNWSSCH	CDGGSLVWDRRPIDWLDFCCYCHDIGY	DTHDQEKL.LKADLAFL
MEKEHI6448 AFLSOLF <u>PRYCHYCCPNWSS</u> CKDCCSL <u>VWDK</u> R <u>PIDWLD</u> YCCYCHLCBDHDOAKL.LKADL		TCEOY32890	AFLSQLFPRYCHYCCPNWSSCH	COGS LVWDRRP I DWLDFCCYCHDICY	DTHDOEKL.LKADLAFL
		Mt KER1 6449	AFLSOLEPRYCHYCCPNWSSC	COCCELENT AND A REAL PROPERTY OF THE REAL PROPERTY	DTHDO
				**	*

Fig. 1 Pairwise alignment of the amino acid sequences of LusPLA<sub>2</sub>I and LusPLA<sub>2</sub>II with multiple plant  $sPLA_2s$  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

were classified into three groups, viz., XI, PLA<sub>2</sub> like proteins, and venom sPLA<sub>2</sub> (Fig. 3). Among these three groups, snake venom sPLA<sub>2</sub> formed a separate group sharing common blood line with group XI sPLA<sub>2</sub>s. Earlier animal and plant sPLA<sub>2</sub>s have been reported to show similarities in the conserved domain region and evolution

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

(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





مدينة الملك عبدالعزيز KACST للعلوم والتفنية KACST ◄ Fig. 2 Visualization of similarity among flax and known plant sPLA<sub>2</sub>s. a % similarity among LusPLA<sub>2</sub>s (LusPLA<sub>2</sub>I and LusPLA<sub>2</sub>II) and known plant secretory PLA<sub>2</sub>s. b % similarity of LusPLA<sub>2</sub>s (LusPLA<sub>2</sub>I and LusPLA<sub>2</sub>II) to known sPLA<sub>2</sub>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<sub>2</sub>\_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<sub>2</sub>s.

Plant sPLA<sub>2</sub>s have been placed in group XI (Six and Dennis 2000) which is divided into sub-groups XIA and XIB as per PLA<sub>2</sub> superfamily numbering system (Burke and Dennis 2009). The sequence analysis (Fig. 1) revealed that LusPLA<sub>2</sub>s were placed in sub-group XIA and XIB. Alignment of LusPLA<sub>2</sub> protein sequences with known plant sPLA<sub>2</sub>s revealed that LusPLA<sub>2</sub>I is placed in group XIB along with isozymes of sPLA<sub>2</sub> from other analyzed plant species (Supplementary Table 1).

Isoenzymes of sPLA<sub>2</sub> from multiple plant species were sorted in group XIA along with LusPLA<sub>2</sub>II as per their similarity (Supplementary Table 2). Isoenzymes from *Arabidopsis thaliana*, *Brassica napus*, *Ricinus communis*,



**Fig. 3** Phylogenetic tree of sPLA<sub>2</sub> depicting evolutionary relationship of sPLA<sub>2</sub>s from flax with orthologous plant and snake venome sPLA<sub>2</sub>s. Un-rooted phylogenetic tree represents total 68 sPLA<sub>2</sub>s protein sequences belonging to flax, known plants, and snake venom. sPLA<sub>2</sub>s from multiple origin were divided into three groups such as XI, PLA<sub>2</sub> like proteins, and venom sPLA<sub>2</sub>. Flax sPLA<sub>2</sub> was grouped in cluster XI, LusPLA<sub>2</sub>I in XIB, and LusPLA<sub>2</sub>II in XIA. Venom







Fig. 4 Analysis of conserved domains and phylogeny of sPLA<sub>2</sub> belonging to cd04706 sub-family. **a** % similarity among the conserved domain region of flax and bacterial sPLA<sub>2</sub>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

*Theobroma cacao*, and *Medicago truncatula* formed a separate group of PLA<sub>2</sub>\_like proteins.

Plant sPLA<sub>2</sub>s display minor differences in their Ca<sup>2+</sup> binding domain and catalytic motif on the basis of which they are divided into sub-group XIA and XIB. sPLA<sub>2</sub>s of sub-group XIA contain Arg or Gly and group XIB contains Asn or Ser residues as third Ca<sup>2+</sup> coordinating amino acid (Fig. 1). In addition, plant sPLA<sub>2</sub>s grouped within the subgroup 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<sup>2+</sup> coordinating amino acid and His residue downstream to His/ASP dyad in LusPLA<sub>2</sub>II fits it into subgroup XIA. The presence of Leu residue as third Ca<sup>2+</sup> coordinating amino acid and Ser residue downstream to His/ASP dyad in LusPLA<sub>2</sub>I (Fig. 1) appropriately fits it into sub-group XIB. In addition to this, the Ala residue of

alignment of the conserved domain region of flax and bacterial sPLA<sub>2</sub>s. *Asterisks* denote the amino acid residues involved in Ca<sup>2+</sup> binding and *hashes* denote the amino acid involved in catalysis. **c** Phylogenetic tree of flax sPLA<sub>2</sub>s and bacterial sPLA<sub>2</sub>s of phylum Alphaproteobacteria and Firmicutes. Evolutionary analyses reveal that the members of cd04706 collection bear common ancestor

the conserved catalytic site motif ( $D\underline{A}CCxxHDxCV$ ) is replaced by Ser residue ( $D\underline{S}CCxxHDxCV$ ) in LusPLA<sub>2</sub>I. Such a replacement has also been observed in Arabidopsis (Lee et al. 2003).

Plant PLA<sub>2</sub>-like enzymes also contain the signature phospholipase  $A_2$  (PA2c) domain comprising the Ca<sup>2+</sup> binding loop and active site motif but show minor deviation from features of a standard sPLA<sub>2</sub>. They contain only six Cys residues instead of 12 Cys residues, a feature similar to animal sPLA<sub>2</sub> 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<sub>2</sub>\_like proteins with animal sPLA<sub>2</sub> sequence. This is in line with





**Fig. 5** Similarity and phylogenetic analysis of conserved domains of flax and representative sPLA<sub>2</sub>s of each sub-family of cd00618 collection. **a** % similarity among the conserved domain region of flax and representative sPLA<sub>2</sub>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<sub>2</sub> members of

the placement of  $PLA_2$ \_like proteins closer to snake venom in phylogenetic tree (Fig. 3).

Secretory phospholipase A<sub>2</sub> 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<sub>2</sub>s 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<sub>2</sub>s belonging to bacteria, fungi, plants, and animals. Plant sPLA<sub>2</sub>s are grouped in cd04706 collection along with sPLA<sub>2</sub>s of bacteria belonging to phylum Proteobacteria and Firmicutes.

Amino acid sequence comparison in the conserved domain region of flax and bacterial sPLA<sub>2</sub> revealed that despite of being assigned to same cd04706 sub-family, they



cd00125, cd04704, cd04705, cd04706, and cd04707 sub-families. The evolutionary analysis reveals that representative sPLA<sub>2</sub>s from different cd00618 sub-families evolved from a common ancestor. Plant and bacterial sPLA<sub>2</sub>s of cd04706 sub-family are grouped together. **c** Position of flax sPLA<sub>2</sub>s in the existing model of evolution of sPLA<sub>2</sub>s in the six kingdoms of life, viz., Animalia, Fungi, Plantae, Protista, Archeabacteria, and Eubacteria, belonging to different sub-families of cd00618 collection

showed minimal similarity among them (Fig. 4a). LusPLA<sub>2</sub>I showed high similarity (42%) to Clostridium botulinum of phylum firmicutes (Supplementary Table 3), whereas LusPLA<sub>2</sub>II showed similarity (42%) to Methylobacterium nodulans of phylum Alphaproteobacteria (Supplementary Table 4). While the sPLA<sub>2</sub> of flax contained 12 Cys residue that can form six intramolecular disulphide bonds, sPLA<sub>2</sub> 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<sub>2</sub>s of flax and bacteria. The calcium binding residues, Tyr, Gly, and Asp, are also conserved among the sPLA<sub>2</sub> of cd04706 sub-family (Fig. 4b). These features suggest that sPLA<sub>2</sub> of flax is structurally similar to sPLA<sub>2</sub> of bacteria. Phylogenetic analysis of sPLA<sub>2</sub> from plants and bacteria of the cd04706 collection suggests that eukaryote and prokaryote sequences share a common ancestor (Fig. 4c). Prokaryotic sPLA<sub>2</sub> and plant sPLA<sub>2</sub> form distinct groups of their own suggesting that sPLA<sub>2</sub> 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<sub>2</sub> of cd00125, cd04704, cd04705, cd04706, and cd04707 sub-families revealed that flax sPLA<sub>2</sub>s shows 70–50% similarity with known plant sPLA<sub>2</sub>, 45–35% similarity to bacterial sPLA<sub>2</sub> and 40–30% similarity to sPLA<sub>2</sub> from all other taxa (Fig. 5a; Supplementary Table 5, 6). The phylogenetic analysis of representative sPLA<sub>2</sub> of cd00618 collections belonging to prokaryotic taxa, eukaryotic taxa, and flax, revealed that eukaryotic sPLA<sub>2</sub> share a common origin with their homologs in bacteria (Fig. 5b). The conservation of amino acid residues involved in catalysis and Ca<sup>2+</sup> binding suggests that sPLA<sub>2</sub> of various cd00618 sub-families share a common ancestor and evolved rapidly by multiple point mutations and gene duplication.

LusPLA<sub>2</sub>s fits well in the model of evolution for sPLA<sub>2</sub> 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<sub>2</sub> and phylogenetic analysis of flax sPLA<sub>2</sub>s with representative of various cd00618 collections supports the placement of flax sPLA<sub>2</sub>s in the existing model of evolution.

Our investigation was focused on elucidation of the origin and evolution of flax sPLA<sub>2</sub>. Phylogenetic analysis revealed that flax sPLA<sub>2</sub>s share their evolutionary origin with bacteria of phylum Alphaproteobacteria and Firmicutes. Flax sPLA<sub>2</sub> fits well in the existing system of classification proposed for sPLA<sub>2</sub> belonging to different phyla. Similarity and positioning of flax sPLA<sub>2</sub> with sPLA<sub>2</sub> 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<sub>2</sub> 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<sub>2</sub>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<sub>2</sub> 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<sub>2</sub> 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<sub>2</sub> 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, McInerney 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<sub>2</sub> 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<sub>2</sub> 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<sub>2</sub> 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<sub>2</sub> 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<sub>2</sub> isoforms from soybean (*Glycine max*). Biochimie 94:2608–2619
- Nevalainen TJ, Cardoso JC, Riikonen PT (2012) Conserved domains and evolution of secreted phospholipases A<sub>2</sub>. FEBS J 279:636–649
- Pohnert G (2002) Phospholipase A<sub>2</sub> activity triggers the woundactivated 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 ELON-GATION1 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<sub>2</sub> superfamily and its group numbering system. Biochim Biophys Acta 1761:1246–1259
- Scherer GF, Ryu SB, Wang X, Matos AR, Heitz T (2010) Patatinrelated 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<sub>2</sub> 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<sub>2</sub> 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<sub>2</sub>S (PLA<sub>2</sub>s) are structurally related to the animal secretory PLA<sub>2</sub>s 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<sub>2</sub> 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

