

# Conserved domains and evolution of secreted phospholipases A<sub>2</sub>

Timo J. Nevalainen<sup>1</sup>, João C. R. Cardoso<sup>2</sup> and Pentti T. Riikonen<sup>3</sup>

<sup>1</sup> Department of Pathology, University of Turku and Turku University Hospital, Finland

<sup>2</sup> Comparative Molecular Endocrinology, Centre of Marine Sciences, Universidade do Algarve, Campus de Gambelas, Faro, Portugal

<sup>3</sup> Department of Information Technology, University of Turku, Finland

## Keywords

conserved domains; eukaryotes; evolution; prokaryotes; secreted PLA<sub>2</sub>s

## Correspondence

Timo J. Nevalainen, Department of Pathology, University of Turku, Kiinamyllynkatu 10, FIN-20520 Turku, Finland  
Fax: +358 2 3337456  
Tel: +358 2 3337500  
E-mail: timneva@utu.fi

(Received 1 September 2011, revised 8 December 2011, accepted 13 December 2011)

doi:10.1111/j.1742-4658.2011.08453.x

Secreted phospholipases A<sub>2</sub> (sPLA<sub>2</sub>s) are lipolytic enzymes present in organisms ranging from prokaryotes to eukaryotes but their origin and emergence are poorly understood. We identified and compared the conserved domains of 333 sPLA<sub>2</sub>s and proposed a model for their evolution. The conserved domains were grouped into seven categories according to the *in silico* annotated conserved domain collections of 'cd00618: PLA<sub>2</sub>-like' and 'pfam00068: Phospholip\_A2\_1'. PLA<sub>2</sub>s containing the conserved domain cd04706 (plant-specific PLA<sub>2</sub>) are present in bacteria and plants. Metazoan PLA<sub>2</sub>s of the group (G) I/II/V/X PLA<sub>2</sub> collection exclusively contain the conserved domain cd00125. GIII PLA<sub>2</sub>s of both vertebrates and invertebrates contain the conserved domain cd04704 (bee venom-like PLA<sub>2</sub>), and mammalian GIII PLA<sub>2</sub>s also contain the conserved domain cd04705 (similar to human GIII PLA<sub>2</sub>). The sPLA<sub>2</sub>s of bacteria, fungi and marine invertebrates contain the conserved domain pfam09056 (prokaryotic PLA<sub>2</sub>) that is the only conserved domain identified in fungal sPLA<sub>2</sub>s. Pfam06951 (GXII PLA<sub>2</sub>) is present in bacteria and is widely distributed in eukaryotes. All conserved domains were present across mammalian sPLA<sub>2</sub>s, with the exception of cd04706 and pfam09056. Notably, no sPLA<sub>2</sub>s were found in Archaea. Phylogenetic analysis of sPLA<sub>2</sub> conserved domains reveals that two main clades, the cd- and the pfam-collection, exist, and that they have evolved via gene-duplication and gene-deletion events. These observations are consistent with the hypothesis that sPLA<sub>2</sub>s in eukaryotes shared common origins with two types of bacterial sPLA<sub>2</sub>s, and their persistence during evolution may be related to their role in phospholipid metabolism, which is fundamental for survival.

## Introduction

### Phospholipases A<sub>2</sub>

Phospholipases A<sub>2</sub> (PLA<sub>2</sub>s; [EC 3.1.1.4](#)) are a group of lipolytic enzymes that hydrolyze the *sn*-2 bond of phospholipids, such as phosphatidylcholine and phosphatidylethanolamine, resulting in the release of fatty acid and lysophospholipid. They have been isolated

from organisms ranging from bacteria to mammals and are suggested to have emerged early in evolution. In general, PLA<sub>2</sub>s are classified into three broad categories: Ca<sup>2+</sup>-dependent secreted PLA<sub>2</sub>s (sPLA<sub>2</sub>s); Ca<sup>2+</sup>-dependent cytosolic PLA<sub>2</sub>s; and Ca<sup>2+</sup>-independent cytosolic PLA<sub>2</sub>s. Cytosolic PLA<sub>2</sub>s are known to play an important role in cellular signalling and

### Abbreviations

EST, expressed sequence tag; G, group; ML, maximum likelihood; NJ, neighbour-joining; OtoC, N-terminal region of otoconin; OtoN, N-terminal region of otoconin; PLA<sub>2</sub>, phospholipase A<sub>2</sub>; sPLA<sub>2</sub>, secreted phospholipase A<sub>2</sub>.

prostanoid metabolism [1,2]. Secreted PLA<sub>2</sub>s are components of various body fluids, including blood plasma, pancreatic juice, tears, seminal fluid, and snake and other venoms, and they participate in diverse physiological and pathological functions, such as in the digestion of dietary phospholipids, in inflammatory reactions and in the defence against bacteria and other pathogens [1–7].

### Secreted PLA<sub>2</sub>s

Secreted PLA<sub>2</sub>s are the product of distinct genes and they have been classified, according to their molecular structure, into the following groups: IA, IB, IIA, IIB, IIC, IID, IIE, IIF, III, V, IX, X, XIA, XIB, XIIA, XIIB, XIII and XIV [1]. Evidence of structural similarity among group (G)I, GII, GV and GX members was taken to suggest that they may form a distinct GI/II/V/X PLA<sub>2</sub> collection [8]. The GXIV PLA<sub>2</sub>s of bacteria and fungi differ in their primary structure and folding from the other sPLA<sub>2</sub>s [1].

Secreted PLA<sub>2</sub>s are small-molecular-mass proteins (14–18 kDa, 120–135 amino acid residues) that require the presence of Ca<sup>2+</sup> at millimolar concentrations for their catalytic activity. They share a conserved 3D structure that is stabilized by five to eight disulfide bonds. The structure of the GI, GII and GX PLA<sub>2</sub>s consists of three  $\alpha$ -helices, a two-stranded  $\beta$ -sheet (the  $\beta$ -wing) and a conserved Ca<sup>2+</sup>-binding loop [9–13]. The conserved catalytic network of GIB and GIIA PLA<sub>2</sub>s consists of hydrogen-bonded side-chains formed by histidine, which is localized in the long  $\alpha$ -helix1, tyrosine and aspartic acid residues, as well as the hydrophobic wall that shields it [14]. The sPLA<sub>2</sub>s of plants have a different 3D structure. In the GXIB PLA<sub>2</sub> of rice (*Oryza sativa*) the  $\beta$ -wing is absent and the C-terminal  $\alpha$ -helix3 has a different orientation [15]. The sPLA<sub>2</sub>s of prokaryotes and fungi are characterized by a dominant  $\alpha$ -helical fold [16]. The 3D structure of GIII PLA<sub>2</sub> differs from that of GI/II/V/X PLA<sub>2</sub>s but they share identical motifs [17].

### Conserved domains of sPLA<sub>2</sub>s and their evolution

Functional motifs of sPLA<sub>2</sub>s that are well conserved include the Ca<sup>2+</sup>-binding and catalytic sites, and the conserved cysteine residues and disulfide bond pattern [1,2]. *In silico* database annotations identified two conserved signature patterns: the 'PA2\_HIS (PS0018, Phospholipase A<sub>2</sub> histidine active site C-C-{P}-x-H-{LGY}-x-C', where x represents a nonconserved amino acid, and amino acids within brackets are not

allowed) that contains the histidine residue of the sPLA<sub>2</sub> active site, and the 'PA2\_ASP (PS00119, Phospholipase A<sub>2</sub> aspartic acid active site [LIVMA]-C-{LIVMFYWPCST}-C-D-{GS}-{G}-{N}-x-{QS}-C', where x represents a nonconserved amino acid, amino acids within curly brackets are not allowed and amino acids within square brackets are allowed) centred on the active site aspartic acid residue and localized towards the C-terminal portion of the molecule [18]. The Ca<sup>2+</sup>-binding motif, which, for example in mammalian GIB sPLA<sub>2</sub>s, is Y-x-G-x-G (where x represents a nonconserved amino acid) is localized before the histidine catalytic site towards the N-terminus.

Studies of sPLA<sub>2</sub>s have focussed mainly on eukaryotes [1–4]. However, the availability of molecular data from phylogenetically distant organisms, and the identification of new sPLA<sub>2</sub>s, challenges the current classification system, especially when applied to invertebrate and prokaryote sPLA<sub>2</sub>s [1,19–21]. Despite their potential evolutionary relationships [22,23], the origin and emergence of sPLA<sub>2</sub>s is still intriguing. Although all sPLA<sub>2</sub>s share the same catalytic mechanism involving the canonical histidine residue, there is considerable variation in their sequence identity. Based on their identity, members of the GI/II/V/X PLA<sub>2</sub> collection have been designated 'conventional sPLA<sub>2</sub>s' and are believed to be evolutionally close, whereas GIII and GXII PLA<sub>2</sub>s are classified as 'atypical sPLA<sub>2</sub>s' and are evolutionally more distant [4,8]. In bacteria and fungi there are sPLA<sub>2</sub>s that show only limited sequence identity with other sPLA<sub>2</sub>s [1,16], and to date the evolutionary connection between these structurally distant proteins is lacking.

The present study aimed to contribute to the understanding of the origin and evolution of sPLA<sub>2</sub>s based upon the identification of their conserved domains in a representative collection of prokaryotes and eukaryotes. Recently, based on published data and the collection of well-annotated multiple sequence alignment models of their conserved domains, the family hierarchy of sPLA<sub>2</sub>s was comprehensively classified in two general collections: 'cd00618 PLA<sub>2</sub>\_like: Phospholipase A<sub>2</sub>, a superfamily of secretory and cytosolic phospholipases A<sub>2</sub>' and 'pfam00068: Phospholip\_A2\_1, Phospholipase A<sub>2</sub>' [24,25]. In the present study, secreted PLA<sub>2</sub> protein sequences retrieved from public databases were classified according to their conserved domain structures. Their sequences were compared and specific motifs within each subfamily group were identified, and, based upon their sequence and structure similarity, a model for their origin and evolution was proposed.

## Results

### Repertoire of prokaryotic and eukaryotic sPLA<sub>2</sub>s

A total of 333 sPLA<sub>2</sub> sequences from a wide range of taxa were retrieved from publicly available databases, and 358 conserved domains within the cd00618 (PLA<sub>2</sub>-like) and pfam00068 (Phospholip\_A2\_1) collections were identified (Tables 1 and S1). Members of the cd-collection identified are grouped into five distinct sub-families: the 'cd04706 PLA<sub>2</sub>\_plant: Plant-specific sub-family of Phospholipase A<sub>2</sub>'; 'cd00125 PLA<sub>2</sub>c: Phospholipase A<sub>2</sub>, a family of secretory and cytosolic enzymes'; 'cd04707 otoconin\_90: Phospholipase A<sub>2</sub>-like domains present in otoconin-90 and otoconin-95'; 'cd04704 PLA<sub>2</sub>\_bee\_venom\_like: A subfamily of Phospholipase A<sub>2</sub>, similar to bee venom PLA<sub>2</sub>' and 'cd04705 PLA<sub>2</sub>\_group\_III\_like: A subfamily of Phospholipase A<sub>2</sub>, similar to human group III PLA<sub>2</sub>'. The pfam-collection members are subdivided into 'pfam06951: PLA<sub>2</sub>G12, group XII secretory phospholipase A<sub>2</sub> precursor' and 'pfam09056 Phospholip\_A2\_3: Prokaryotic phospholipase A<sub>2</sub> domain found in PLA<sub>2</sub>s of bacteria and fungi' [24,25, <http://www.ncbi.nlm.nih.gov/cdd>].

All conserved domains of sPLA<sub>2</sub>s identified are present in representatives of the Kingdom Animalia with the exception of cd04706 of the sPLA<sub>2</sub>s of bacteria and plants (Table 1). Within this Kingdom, sPLA<sub>2</sub>s containing the conserved domain cd00125 are widespread in organisms ranging from basal metazoa (Porifera, Placozoa, Cnidaria and Rotifera), protostomes (Insects, Nematodes, Molluscs and Arthropods) and early deuterostomes (Echinodermata, Cephalocordata and Tunicata) to teleosts and tetrapods (Amphibia, Aves and Mammalia). The conserved domain cd00125 is typical of GIA, IB, IIA, IIB, IIC, IID, IIE, IIF, V and X sPLA<sub>2</sub>s. The conserved domain cd04707 was found in the N- and C-terminal regions of vertebrate otconins. The conserved domains

cd04704 and cd04705 are present in metazoan GIII PLA<sub>2</sub>s, the latter domain exclusively in mammals. The sPLA<sub>2</sub>s of bacteria contain either cd04706 or pfam09056 (present in XIV PLA<sub>2</sub>s) and exceptionally pfam06951. In contrast, no sPLA<sub>2</sub>s were found in archaea. Pfam09056 is present in the sPLA<sub>2</sub>s of fungi. The majority of plant sPLA<sub>2</sub>s contained the conserved domain cd04706 (identified in XIA and XIB PLA<sub>2</sub>s). The conserved domain pfam06951 is present in the GXIIA PLA<sub>2</sub>s of unicellular and multicellular organisms, including marine algae and bacteria (Tables 1 and S1).

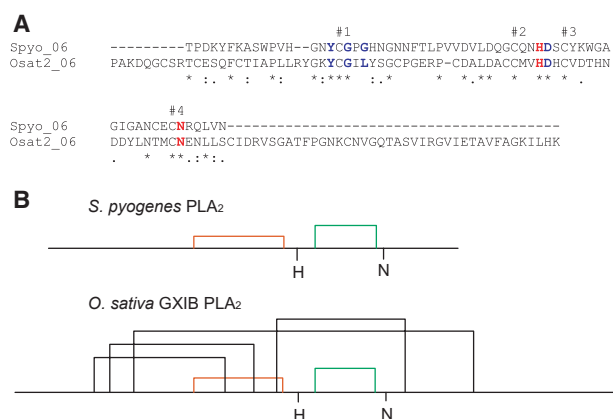
### Conserved domain cd04706 of PLA<sub>2</sub>s of bacteria and plants

The sPLA<sub>2</sub>s of bacteria and plants share a number of conserved structural features for the histidine and aspartic acid catalytic motifs and Ca<sup>2+</sup>-binding domains (Fig. 1). The conserved domain cd04706 was identified in the sPLA<sub>2</sub>s of bacteria of the class Alphaproteobacteria of the phylum Proteobacteria and also in those of the phylum Firmicutes, which includes both Gram-negative and Gram-positive bacteria, such as the human pathogens *Streptococcus pyogenes*, *Clostridium perfringens*, *Clostridium botulinum* and *Bacillus cereus* [26] (Table S1). Secreted PLA<sub>2</sub>s containing cd04706 were also identified in numerous plants. The sPLA<sub>2</sub>s of *O. sativa* are well characterized and classified as GXI PLA<sub>2</sub>s [15,27]. In general, several GXI PLA<sub>2</sub> isoforms were identified in plants, which may have resulted from gene-duplication events, for example, four GXI PLA<sub>2</sub>s were identified in *O. sativa* (Table S1). The GXIB PLA<sub>2</sub> of *O. sativa* is a 16.6 kDa protein in which the Ca<sup>2+</sup>-binding site contains tyrosine, glycine and aspartic acid residues, and the histidine residue of the active site is centred in the catalytic site motif (Fig. 1A).

**Table 1.** Distribution of the conserved domains of sPLA<sub>2</sub> in the Kingdoms of Life. +, presence; –, absence.

| PLA <sub>2</sub> conserved domain | Group number <sup>a</sup>                  | Archaea | Bacteria | Viridiplantae | Fungi | Animalia |
|-----------------------------------|--|---------|----------|---------------|-------|----------|
| cd00125                           | IA, IB, IIA, IIB, IIC, IID, IIE, IIF, V, X | –       | –        | –             | –     | +        |
| cd04704                           | III  | –       | –        | –             | –     | +        |
| cd04705 <sup>b</sup>              | III  | –       | –        | –             | –     | +        |
| cd04706                           | XIA, XIB                                   | –       | +        | +             | –     | –        |
| cd04707                           | Otoconin                                   | –       | –        | –             | –     | +        |
| pfam06951                         | XIIA, XIIB                                 | –       | +        | +             | +     | +        |
| pfam09056                         | XIV  | –       | +        | –             | +     | +        |

<sup>a</sup> Group number according to the classification in Schaloske and Dennis [1]. <sup>b</sup> cd04705 is found only in mammalian GIII PLA<sub>2</sub>s.



**Fig. 1.** Alignment of the conserved domain cd04706 sequences of *Streptococcus pyogenes* sPLA<sub>2</sub> (Spyo\_06) and *Oryza sativa* GXIB PLA<sub>2</sub> (Osat2\_06) (A) and the positions of the disulfide bonds (B). In panel A the four conserved cysteine (C) residues are indicated with '#' and the conserved putative catalytic histidine (H) and asparagine (N) residues are highlighted in red and the Ca<sup>2+</sup>-binding residues tyrosine (Y), glycine (G) and aspartic acid (D) are highlighted in blue. The consensus symbols denote the degree of conservation: '\*', identical residues; ':', conserved substitutions; and '.', semiconservative substitutions. In (B) the conserved disulfide bonds C<sub>1</sub>-C<sub>2</sub> and C<sub>3</sub>-C<sub>4</sub> are red and green, respectively. H and N mark the histidine and asparagine residues of the catalytic sites, respectively. The solid horizontal line represents the mature peptide region.

Sequence comparisons revealed that the plant (*O. sativa*) and bacterial (*S. pyogenes*) sPLA<sub>2</sub>s share 17% amino acid sequence identity. In both proteins, there is an asparagine residue in the C-terminal portion of the molecule, which probably contributes to the catalytic function [15], instead of the more commonly occurring aspartic acid residue. In the GXIB PLA<sub>2</sub> of *O. sativa* there are 12 cysteine residues that form six disulfide bonds [15], while the conserved domain of *S. pyogenes* sPLA<sub>2</sub> contains five cysteine residues, four of which (C<sub>1</sub>-C<sub>2</sub> and C<sub>3</sub>-C<sub>5</sub>), align with the plant GXIB PLA<sub>2</sub> cysteines (C<sub>4</sub>-C<sub>8</sub> and C<sub>9</sub>-C<sub>10</sub>) and form conserved disulfide bonds, and are suggestive of structural similarity (Fig. 1B).

### Conserved domain cd00125 of PLA<sub>2</sub> of animals

The conserved domain cd00125 was identified in the basal invertebrates Placozoa up to mammals and was prevalent in the vertebrate GI/II/V/X PLA<sub>2</sub> collection where the functional sites and cysteine residues are highly conserved (Fig. 2). Members of this collection share 26–50% amino acid sequence identity within the conserved domain region. Among the members of the GI/II/V/X PLA<sub>2</sub> collection, the catalytic histidine is centred in the active site motif and

the aspartic acid residue of the active site is present in the C-terminal portion of the molecule. In the Ca<sup>2+</sup>-binding site, Ca<sup>2+</sup> is bound by tyrosine and two glycines and an aspartic acid adjacent to the catalytic histidine (Fig. 2).

Homologues of GI/II/V/X PLA<sub>2</sub> collection members are present also in invertebrate genomes, such as the sea anemone *Nematostella vectensis*, where six genes were identified (Table S1) [28]. The sPLA<sub>2</sub> of the sea anemone *Adamsia carciniopados* (also called *Adamsia palliata*) contains the pancreatic loop characteristic of vertebrate GI PLA<sub>2</sub>s and lacks the C-terminal extension found in GII PLA<sub>2</sub>s but shares homology for both GI and GII PLA<sub>2</sub>s [19]. Comparison of the disulfide bond positions of the *A. carciniopados* sPLA<sub>2</sub> with those of the human GI/II/V/X PLA<sub>2</sub> collection indicates that the sea anemone sPLA<sub>2</sub> shares five conserved disulfide bonds with the vertebrate homologues, suggesting structural, and possibly functional, conservation across the phylogenetically distant organisms of Cnidaria and Mammalia (Fig. 3).

### Conserved domain cd04707 of otoconins

Otoconins containing the conserved domain cd04707 were identified in a number of vertebrates, from fish to mammals, and two conserved domains within the N-terminal (OtoN) and C-terminal (OtoC) regions of the same collection were identified within the mature protein sequence (Tables 1 and S1). The human otoconin-90 is a 53 kDa protein, and the OtoN and OtoC domains are 37% identical and structurally closely related to the members of the GI/II/V/X PLA<sub>2</sub> collection, suggesting common ancestry (Fig. 2). The human OtoN and OtoC domains are 36% and 34% identical, respectively, when compared with GIB PLA<sub>2</sub>. Moreover, the disulfide bond patterns of human otoconin and GIB PLA<sub>2</sub> are identical (Fig. 3). Both OtoN and OtoC domains contain a histidine residue within the conserved catalytic site. However, they are believed to be catalytically inactive as a result of mutations in the Ca<sup>2+</sup>-binding sites (Fig. 2), leading to loss of the usual Ca<sup>2+</sup>-binding residues in the OtoN domain, and in the OtoC domain the conserved second glycine of the mammalian GIB sPLA<sub>2</sub>s (Y-x-G-x-G) is replaced with a glutamic acid, although PLA<sub>2</sub> activity remains to be investigated [29].

### Conserved domain cd04704 of group III PLA<sub>2</sub>s

Secreted PLA<sub>2</sub>s, containing the conserved domain cd04704 (bee venom-like PLA<sub>2</sub>), were identified in arthropods, reptiles and vertebrates, including humans.



**Fig. 2.** Multiple sequence alignment of the conserved domains of human sPLA<sub>2</sub>s. HsapIB\_25 (GIB PLA<sub>2</sub>), HsapIIA\_25 (GIIA), HsapIIC\_25 (GIIC), HsapIID\_25 (GIID), HsapIIE\_25 (GIII), HsapIIF\_25 (GIIF), HsapV\_25 (GV) and HsapX\_25 (GX) PLA<sub>2</sub>s contain the conserved domain cd00125 and belong to the structurally related GI/II/V/X PLA<sub>2</sub> collection. HsapIII\_04 and HsapIII\_05 (GIII PLA<sub>2</sub>) represent cd04704 and cd04705, respectively. HsapXIIA\_51 and HsapXIIB\_51 (GXIIA and GXIIB PLA<sub>2</sub>s) contain pfam06951. HsapOtoN\_07 and HsapOtoC\_07 contain the N- and C-terminal sections of otoconin-90, respectively. The catalytic histidine (H), aspartic acid (D) and asparagine (N) residues are highlighted in red, and the Ca<sup>2+</sup>-binding amino-acid residues tyrosine (Y), glycine (G), histidine (H), phenylalanine (F), tryptophan (W), proline (P), leucine (L) and aspartic acid (D) are highlighted in blue.

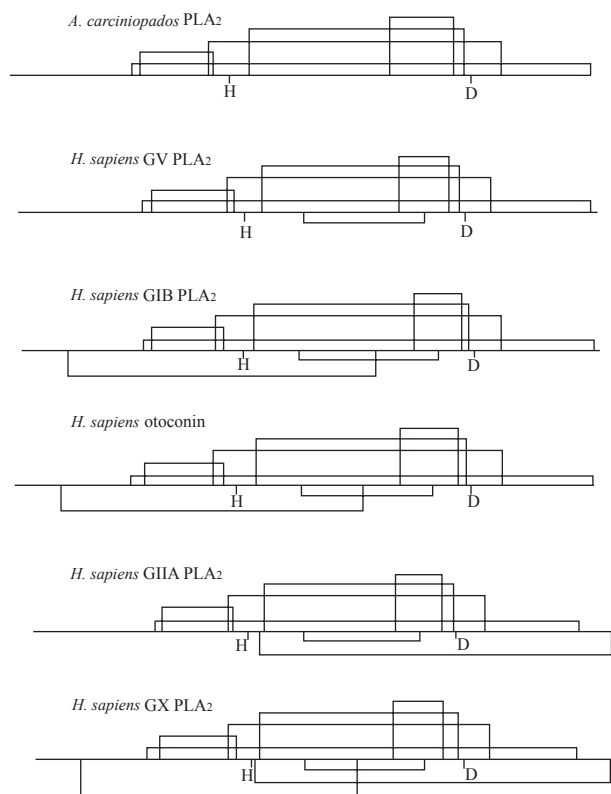
The honeybee *Apis mellifera* venom GIII PLA<sub>2</sub> shares considerable sequence identity with the members of the GI/II/V/X PLA<sub>2</sub> collection (e.g. 44% identity with the human GIB PLA<sub>2</sub>) and is related in its 3D structure and catalytic mechanism to GI and GII PLA<sub>2</sub>s [11,17]. The backbone of the GIII PLA<sub>2</sub> molecule contains the conserved Ca<sup>2+</sup>-binding loop with tryptophan, glycine and aspartic acid residues and the conserved catalytic histidine and aspartic acid residues (Fig. 4). Human GIII PLA<sub>2</sub> is a 57 kDa protein in which cd04704 is localized in the middle part of the molecule and flanked by N- and C-terminal extensions [30]. The cd04704 of human GIII PLA<sub>2</sub> displays features similar to the arthropod GIII PLA<sub>2</sub>s, including the Ca<sup>2+</sup>-binding and catalytic site residues and also the 10 conserved cysteines that form five disulfide bonds at similar locations, and shares 44% sequence identity with honeybee venom GIII PLA<sub>2</sub>.

### Conserved domain cd04705 of group III PLA<sub>2</sub>

Proteins containing the conserved domain cd04705 (similar to human group III PLA<sub>2</sub>) were found only in mammals (Table 1). The conserved domain cd04705 is localized in the C-terminal part of the GIII PLA<sub>2</sub> molecule and is structurally unrelated to cd04704 (Fig. 2). In contrast to cd04704, cd04705 is considered to be catalytically inactive and its function is unknown [30].

### Conserved domain pfam09056 of PLA<sub>2</sub>s of bacteria, fungi and animals

The conserved domain pfam09056 (prokaryotic PLA<sub>2</sub>) was identified in the sPLA<sub>2</sub>s of Gram-positive and Gram-negative bacteria, protists, fungi and marine invertebrates of the phyla Cnidaria and Mollusca. Bacterial sPLA<sub>2</sub>s containing pfam09056 are markedly



**Fig. 3.** Comparison of the disulfide bond positions in the cnidarian *Adamsia cariniopados* and human GV, GIB, GIIA and GX PLA<sub>2</sub>s and otoconin-90. Five disulfide bonds are conserved from the cnidarian to the human. Histidine (H) and aspartic acid (D) are the active-site amino acid residues of the PLA<sub>2</sub>s, and the corresponding residues are present in otoconin. The basal line represents the mature peptide region.

```

# # #
AmelIII_04 I I Y P G T L W C G H G N K S S C P N E L G R F K H T D A C C R T H D M C P D V M S A G E S K H G L T N T A S H T R L S
HsapIII_04 - T M P G T L W C G V G D S A G N S S E L G V F G P D L C C R E H D R C P Q N I S P L Q Y N Y G I R N Y R F H T I S H
***** * : : : : : * * : * * * * * : : : : : * * *
# # #
AmelIII_04 C D C D D K F Y D C L K N S A D T I S S Y F V G K M Y F N L I D T K C Y K L
HsapIII_04 C D C D T R F Q C L Q N Q H D I S - D I V G V A F F N W L I P C F V L
***** * : : : : : * * * * * : * * : : : * * :

```

**Fig. 4.** Alignment of the conserved domains cd04704 of the *Apis mellifera* GIII PLA<sub>2</sub> (AmelIII\_04) and human GIII PLA<sub>2</sub> (HsapIII\_04). The Ca<sup>2+</sup>-binding residues tryptophan (W), glycine (G) and aspartic acid (D) are highlighted in blue, the catalytic histidine (H) and aspartic acid (D) are highlighted in red, and the conserved cysteine (C) residues are indicated with '#'. The asterisks indicate conserved residues.

different in their molecular structure from the bacterial sPLA<sub>2</sub>s containing cd04706. Comparison of the conserved domains of *Streptomyces violaceoruber* (pfam09056) and *S. pyogenes* (cd04706) sPLA<sub>2</sub>s indicates that they share only 8% sequence identity. The catalytic motif of the sPLA<sub>2</sub> of the Gram-positive bacterium *S. violaceoruber* [16,31] contains the conserved histidine and aspartic acid residues (Figs 2 and 5). The conserved

```

#1 #2
Svio_56 ----SASSQNAWLAANRNQSAWAAAYEFDWSTDLCSQAPDNPFPGFP--FNTACARHDFGY
Tbor_56 TDRLLYSTAMP AFLTAKR NK---NPNLDWSDDGCSKSPDRPAGFN--FLDSCRHDFGY
Nvec11_56 -----KYINGCSVPGDLFPFYKKTFTFACRHRDVCY
Myes_56 -----DGCTTFPNI D I PFRNSLTFACRHRDICY
C. magus -----QXPSTAELCKINSNACSVFPSX-IPCQKXFLAACDRHDTCY
          : * : . . * * * *
          #3 #4
Svio_56 RNYKAAGSF-DANKSRIDSAFYEDMKRVCTGYTG--E--KNTACNSTAWTYQAVK---
Tbor_56 RNYKKQHRFTANRKRIDNFKKDLNECAKYSGL-ESWKGVACRKLANTYDAVTFPG
Nvec11_56 RCFN--VYDWKGD--SCDAVFKANMYSLCKEYGPSSVFRVRI CRRAADGYHLAVAKLG
Myes_56 KCGNNAMYAVLHGRYRCDLFYRNMKAI CKRR---SLFRRTCMVNARLYTTSVRTL G
C. magus HCGK--HFGFKQD--DCDAFFRDMTALCAHGTD----DEGKCPX-----
          : : . * * : : *

```

**Fig. 5.** Comparison of the conserved domains pfam09056 of *Streptomyces violaceoruber* (Svio\_56), *Tuber borchii* (Tbor\_56), *Nematostella vectensis* (Nvec11\_56) and *Mizuhopecten yessoensis* (Myes\_56) sPLA<sub>2</sub>s with the GIX PLA<sub>2</sub> of *Conus magus*. Four cysteine residues (C, marked with '#') and the histidine (H) and aspartic acid (D) residues of the active site (highlighted in red) are conserved. The putative Ca<sup>2+</sup>-binding residues aspartic acid (D), glycine (G) and leucine (L) are shown in blue. There are two putative disulfide bonds between the conserved cysteines C<sub>1</sub>-C<sub>2</sub> and C<sub>3</sub>-C<sub>4</sub>.

Ca<sup>2+</sup>-binding domain is absent and the binding of Ca<sup>2+</sup> is mediated via the aspartic acid and leucine residues upstream of the catalytic site and by the aspartic acid residue adjacent to the catalytic histidine residue (Figs 2 and 5). There are four cysteines that form two putative disulfide bonds, including a bond that connects the catalytic and the Ca<sup>2+</sup>-binding sites (Fig. 5).

In addition to bacteria, sPLA<sub>2</sub>s containing pfam09056 were also identified in both unicellular and multicellular fungi. The sPLA<sub>2</sub> of the fungus *Tuber borchii* [32] is a 23 kDa protein that contains two conserved disulfide bonds analogous to those of the sPLA<sub>2</sub> of *S. violaceoruber* (Fig. 5), and the two conserved domains are 42% identical.

There are pfam09056-containing sPLA<sub>2</sub>s in aquatic invertebrates of the phylum Cnidaria, including the sea anemone *N. vectensis* and the hydrozoan *Hydra magnipapillata*. Cnidarian sPLA<sub>2</sub>s have the conserved catalytic histidine and aspartic acid residues but contain more numerous cysteine residues and putative disulfide bonds than the corresponding bacterial and fungal sPLA<sub>2</sub>s (Fig. 5).

Recently, expressed sequence tags (ESTs), coding for potential sPLA<sub>2</sub>s containing the conserved domain pfam09056, were identified in the protist *Astrammia rara* (phylum Foraminifera) and the scallop *Mizuhopecten yessoensis* (phylum Mollusca) [33]. Previously, a related sPLA<sub>2</sub> was isolated from the venom of another mollusc – the marine snail *Conus magus* – and classified as GIX PLA<sub>2</sub> [34] but no conserved domain was identified for this protein in the present study, probably because of the incompleteness of its sequence. However, conserved histidine and aspartic acid active-site motifs are present in the pfam09056-containing sPLA<sub>2</sub>s of the scallop *M. yessoensis* and the sea anemone *N. vectensis* and in the GIX PLA<sub>2</sub> of the

**Fig. 6.** Consensus bootstrap phylogenetic tree of the sPLA<sub>2</sub> conserved domains. The tree was constructed with 53 taxa (see Table S1 for details) and using the NJ method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches, and bootstrap branches lower than 50 were collapsed. Acap\_56, *Ajellomyces capsulata* (Fungi); AmellIII\_04, *Apis mellifera* (Insecta); Aory2\_56, *Aspergillus oryzae* (Fungi); Bbac\_56, *Bdellovibrio bacteriovorus* (Deltaproteobacteria); BtaullIII\_05, *Bos taurus* (Mammalia); CfamIII\_05, *Canis familiaris* (Mammalia); Cglo\_56, *Chaetomium globosum* (Fungi); Cimm\_56, *Coccidioides immitis* (Fungi); Crei\_06, *Chlamydomonas reinhardtii* (Viridiplantae); Dcar\_06, *Dianthus caryophyllus* (Viridiplantae); Dgeo\_56, *Deinococcus geothermalis* (Deinococci); Dmel\_25, *Drosophila melanogaster* (Insecta); DmelXIIA\_51, *Drosophila melanogaster* (Insecta); Fsp\_56, *Frankia sp.* (Actinobacteria); GgalOtoN\_07, *Gallus gallus* (Aves); Hmag\_56, *Hydra magnipapillata* (Cnidaria); HsapIII\_05, *Homo sapiens* (Mammalia); HsapOtoN\_07, *Homo sapiens* (Mammalia); HsapXIIA\_51, *Homo sapiens* (Mammalia); Hsp\_56, *Helicosporium sp.* (Fungi); HsusIII2\_04, *Heloderma suspectum* (Anguimorpha); IpacXIIA\_51, *Ixodes pacificus* (Arachnida); Iscall\_04, *Ixodes scapularis* (Arachnida); MdomXIIA\_51, *Monodelphis domestica* (Mammalia); Medu\_25, *Mytilus edulis* (Mollusca); Mext2\_06, *Methylobacterium extorquens* (Alphaproteobacteria); Mgr1\_56, *Magnaporthe grisea* (Fungi); MmusIIA\_25, *Mus musculus* (Mammalia); MmusIII\_05, *Mus musculus* (Mammalia); MmusOtoC\_07, *Mus musculus* (Mammalia); MmusOtoN\_07, *Mus musculus* (Mammalia); MmusXIIA\_51, *Mus musculus* (Mammalia); Mnod2\_06, *Methylobacterium nodulans* (Alphaproteobacteria); Mrad\_06, *Methylobacterium radiotolerans* (Alphaproteobacteria); Mtru\_06, *Medicago truncatula* (Viridiplantae); Ntab\_2\_06, *Nicotiana tabacum* (Viridiplantae); Nvec10\_56, *Nematostella vectensis* (Cnidaria); Nvec11\_56, *Nematostella vectensis* (Cnidaria); Nvec12\_65, *Nematostella vectensis* (Cnidaria); Nvec13\_56, *Nematostella vectensis* (Cnidaria); OanalB\_25, *Ornithorhynchus anatinus* (Monotremata); OhanIA\_25, *Ophiophagus Hannah* (Reptilia); Osat2\_06, *Oryza sativa* (Viridiplantae); PperIII\_04, *Phlebotomus perniciosus* (Insecta); Scoe1\_56, *Streptomyces coelicolor* (Actinobacteria); Sequ\_06, *Streptococcus equi* (Firmicutes); Sery\_56, *Saccharopolyspora erythraea* (Actinobacteria); Spyo\_06, *Streptococcus pyogenes* (Firmicutes); Svio\_56, *Streptomyces violaceoruber* (Actinobacteria); TgutXIIA\_51, *Taeniopygia guttata* (Aves); TnigOtoC\_07, *Tetraodon nigroviridis* (Teleostei); XlaeXIIA\_51, *Xenopus laevis* (Amphibia); XtrollIII\_04, *Xenopus tropicalis* (Amphibia).

snail *C. magus* (Fig. 5). Furthermore, the GIX PLA<sub>2</sub> of *C. magus* shares 13–33% sequence identity with the bacterial, fungal and cnidarian pfam09056 PLA<sub>2</sub>s, but lower identity of 9–13% with the cd04706, cd04704 and cd00125 PLA<sub>2</sub>s.

### Conserved domain pfam06951 of GXII PLA<sub>2</sub>s

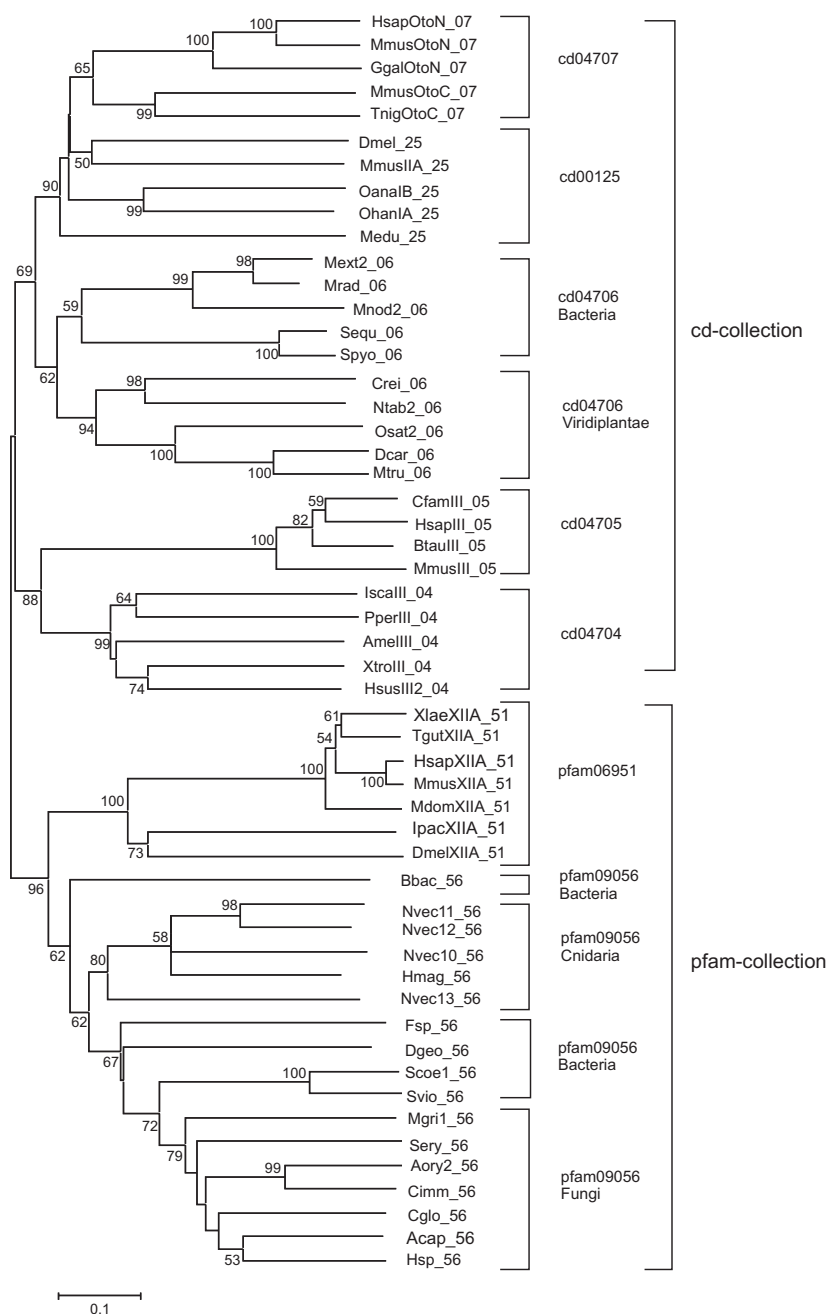
GXIIA PLA<sub>2</sub> was first cloned from human [35]. In the present study, the conserved domain pfam06951 was identified in the GXII PLA<sub>2</sub>s of a large number of vertebrate and invertebrate species, including the basal metazoans *N. vectensis* (Cnidaria), *Trichoplax adhaerens* (Placozoa) and *Brachionus plicatus* (Rotifera), as well as in the nonmetazoan eukaryotes (protists), such as the amoeba *Naegleria gruberi* (Heterolobosea), *Monosiga ovata* (Choanoflagellida), *Euglena gracilis* (Euglenozoa), *Phytophthora infestans* (Oomycetes), *Thalassiosira pseudonana* (Bacillariophyta), *Capsaspora owczarzewski* (Ichthyosporidia), *Chromera velia* (Alveolata) and *Thecamonas trahens* (Apusozoa). Furthermore, pfam06951 was also identified in the sPLA<sub>2</sub>s of the marine algae *Micromonas* (Viridiplantae) and the prokaryote *Planctomyces maris* (Bacteria) (Table S1). These observations demonstrate that the conserved domain pfam06951 is widely distributed not only in the sPLA<sub>2</sub>s of higher animals but also in the sPLA<sub>2</sub>s of simple eukaryotes and prokaryotes. Sequence alignment of the GXIIA PLA<sub>2</sub> conserved domains of organisms ranging from protists to mammals reveals high conservation of the Ca<sup>2+</sup>-binding and catalytic sites. The canonical his-

tidine catalytic site C-C-x-x-H-x-x-C motif is highly conserved. In the aspartic acid catalytic motif, the cysteine and aspartic acid residues are also conserved, with the exception of *E. gracilis* (Euglenozoa), *Micromonas pusilla* (Viridiplantae) and *P. infestans* (Oomycetes), in which aspartic acid is replaced by glutamic acid and the first cysteine is absent in all the sequences (Fig. S1). Comparison of the human GXIIA PLA<sub>2</sub> (pfam06951) with GIB PLA<sub>2</sub> (cd00125) indicated 37% amino acid sequence identity.

Another gene product closely related to GXIIA PLA<sub>2</sub> is GXIIB PLA<sub>2</sub> [36]. Human GXIIB PLA<sub>2</sub> shares 46% sequence identity with GXIIA PLA<sub>2</sub>. GXIIB PLA<sub>2</sub> is a catalytically inactive protein as a result of the substitution of histidine with leucine in the catalytic site (Fig. 2).

### Phylogenetic analysis of sPLA<sub>2</sub>s

Phylogenetic analysis of the conserved domains of sPLA<sub>2</sub>s, carried out with the maximum likelihood (ML) and neighbour-joining (NJ) methods, produced similar tree topologies, suggesting that the members of this protein family have a common and ancient evolutionary origin (Fig. 6). Two major sPLA<sub>2</sub> groups – the cd-collection (which includes the cd00125, cd04704, cd04705, cd04706 and cd04707 of unicellular and multicellular organisms) and the pfam-collection (which contains the sPLA<sub>2</sub>s with the annotated pfam06951 and pfam09056 domains) – exist and underwent distinct trajectories during evolution.



The sPLA<sub>2</sub> forms of bacteria cluster within each PLA<sub>2</sub>s subgroup, suggesting that the eukaryote and prokaryote sPLA<sub>2</sub> repertoire share a PLA<sub>2</sub>-like common ancestor molecule, which remains to be identified. The cd-collection contains the majority of the sPLA<sub>2</sub> sequences identified, and the clustering observed indicates that discrete relationships between the sPLA<sub>2</sub> protein groups exist (Fig. 6). Members of the cd04704 and cd04705 groups seem to have evolved separately from the other cd-collection members, such as cd00125 and cd04707, and have only been identified in animals.

## Discussion

PLA<sub>2</sub> activity was first reported in canine pancreatic juice [37] and sPLA<sub>2</sub>s have now been isolated from a large number of snake and other venoms, and also from cells, tissues and body fluids of various unicellular and multicellular organisms [38–40]. Initially, sPLA<sub>2</sub>s were classified into two major groups of GI and GII PLA<sub>2</sub>s [41], but subsequent sequence homology and conserved disulfide bonds were observed within the other members of the GI/II/V/X PLA<sub>2</sub> col-



lection and the related protein otoconin [42]. This classification has been expanded and refined over the past two decades as information from phylogenetically distinct organisms has accumulated at an accelerating rate [1,2,8,23]. The recent systematic identification of the conserved domains of protein molecules [24,25] allows a novel classification of the sPLA<sub>2</sub>s ('cd/pfam classification') based on the identification of the conserved domains, and a meaningful phylogenetic analysis of the whole range of sPLA<sub>2</sub>s.

The aim of the present study was to elucidate the origin and evolution of the sPLA<sub>2</sub>s. The enormous evolutionary span between the organisms expressing sPLA<sub>2</sub>s (ranging from prokaryotes to metazoan eukaryotes) by necessity introduces a degree of uncertainty in the comparison of protein sequences between such distant phyla. When PLA<sub>2</sub> sequences containing conserved domains of different groups are compared, the molecular structure of a particular PLA<sub>2</sub> reflects the divergent or convergent evolution of the conserved domain structure and the evolutionary history of the organism in question. In the present study, we investigated the structural variability of the conserved domains of sPLA<sub>2</sub>s in a wide range of organisms, from bacteria to mammals. A total of 358 conserved domains were identified among 333 sPLA<sub>2</sub> sequences. In viruses, sPLA<sub>2</sub>-like proteins (GXIII PLA<sub>2</sub>) have also been reported but their molecular structure, including the conserved domain and the enzymatic activity, differ from those of sPLA<sub>2</sub>s [43] and thus were not included in the current analysis.

Two main distinct forms of sPLA<sub>2</sub>s were identified in bacteria: those containing cd04706 (bacteria/plant-specific subfamily of PLA<sub>2</sub>) and those containing pfam09056 (prokaryotic/fungal PLA<sub>2</sub>). Notably, no sPLA<sub>2</sub>s were identified in the Archaea. A marked difference between Bacteria and Archaea is that the predominant lipid constituents of the archaean membranes are prenyl ether lipids, whereas the bacterial membranes contain acyl ester lipids [44–46]. The phospholipid metabolism of bacteria is driven by phospholipases that specifically hydrolyze the acyl ester bonds but are incapable of hydrolyzing the prenyl ether lipids. In light of this observation, it is hypothesized that the sPLA<sub>2</sub>s of the higher organisms may have their evolutionary origin in Bacteria rather than in Archaea.

Members of the sPLA<sub>2</sub>s are proposed to share a common ancestry and to have emerged early in evolution, and several theories based upon their sequence similarities and predicted molecular structure have suggested that they evolved rapidly. The GI and GII sPLA<sub>2</sub>s of Elapidae and Crotalidae snake venoms,

respectively, have distinct molecular structures and it has been proposed that they share a common ancestor [22]. Snake venom sPLA<sub>2</sub>s are suggested to have evolved at an accelerated rate that has resulted in the presence of many variant PLA<sub>2</sub> molecules produced by the venom glands [40,47–49].

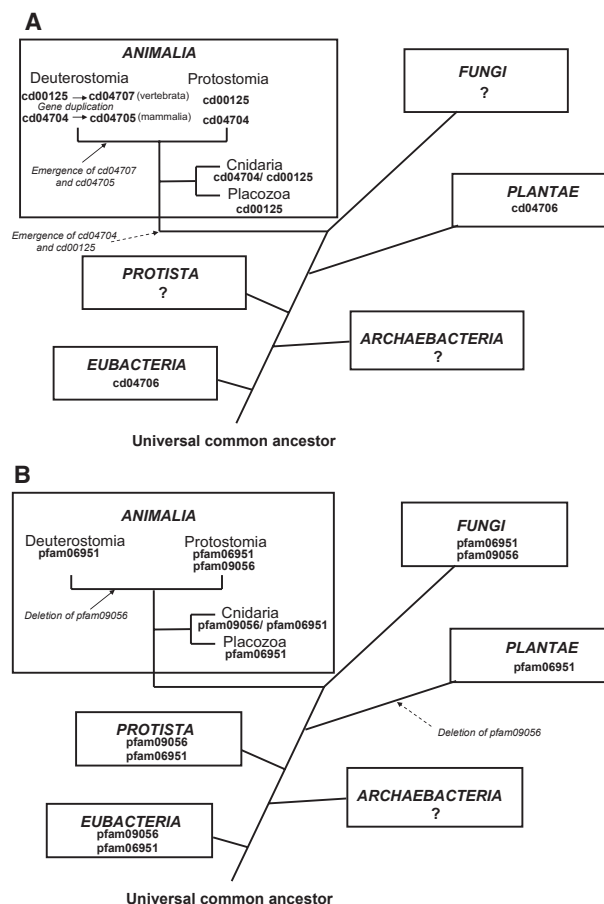
A recent example of the rapid evolution of sPLA<sub>2</sub> genes is the bovine GIID PLA<sub>2</sub>. In cattle, five duplications have been identified, while a single gene copy is present in the human and rodent genomes. The bovine GIID PLA<sub>2</sub>s are expressed in the mammary gland and up-regulated during the lactating period, and are suggested to participate in the innate immune response [50]. In human and mouse, *GIIA*, *GIIC*, *GIID*, *GIIIE*, *GIIIF* and *GV PLA<sub>2</sub>* genes are also the result of gene-duplication events within the same chromosome. In human they are localized in chromosome 1, whereas *GIB* and *GIX* genes map to chromosomes 12 and 16, respectively. However, comparisons between the gene homologues in human and mouse reveal that species-specific events may occur; for example, the human *GIIC PLA<sub>2</sub>* is a pseudogene, whereas the mouse homologue encodes an enzymatically active protein [51]. Another example of the functional diversity within the members of the GI/II/V/X PLA<sub>2</sub> collection are the human and mouse *GIIA PLA<sub>2</sub>*s, which are efficient bactericidal enzymes involved in the innate immune response, whereas the closely related digestive enzyme *GIB PLA<sub>2</sub>* is only marginally bactericidal [6,52]. For example, *in vitro* assays demonstrated that Gram-positive bacteria are killed by sPLA<sub>2</sub>s and that human *GIIA PLA<sub>2</sub>* is highly potent in this respect, whereas *GIB PLA<sub>2</sub>* is of low efficacy [52].

In Cnidaria, a sister group of the vertebrate clade that diverged more than 500 million years ago [53,54], sPLA<sub>2</sub>s have been reported that structurally resemble the vertebrate GI and GII PLA<sub>2</sub>s [19–21,28]. Our current observations indicate that a cnidarian sPLA<sub>2</sub> [19] containing cd00125 has disulfide bonds at locations identical to those of the human *GV PLA<sub>2</sub>* and conserved in other members of the GI/II/V/X PLA<sub>2</sub> collection. Group I PLA<sub>2</sub>s of elapid snake venoms have lost the ancestral pancreatic loop present in the cnidarian sPLA<sub>2</sub> and also in the mammalian *GIB PLA<sub>2</sub>*. Such structural changes have resulted in the appearance of novel functions, including toxicity of the venom sPLA<sub>2</sub>s [55]. However, the number of cysteine residues and disulfide bonds of sPLA<sub>2</sub>s vary among closely related animals such as the sea anemones *Adamsia carcinipados*, *Urticina crassicornis*, *Condylactis gigantea* and *N. vectensis* [19–21,28,56]. The variation in the disulfide patterns and other structural features seems to preclude the

exact placement of these cnidarian sPLA<sub>2s</sub> in the currently recognized groups of the GI/II/V/X PLA<sub>2</sub> collection [19–21].

In the current study, homologues of the vertebrate GXII members with highly conserved domain regions were identified for the first time in unicellular organisms, and their function in such organisms remains to be established. In human, GXIIA PLA<sub>2</sub> is expressed in T lymphocytes and seems to be involved in the regulation of the immune response [57]. Human GXIIB PLA<sub>2</sub> was recently shown to be involved in the triglyceride metabolism in the liver [58] and is proposed to activate specific receptors that remain to be identified [36].

Life as we know it can be divided into the prokaryotic (cellular organisms that lack a nucleus) Domains of Bacteria (Eubacteria) and Archaea (Archaeobacteria), and the Domain of Eukaryota (organisms consisting of nucleated cells, such as animals, plants and fungi) [44]. Prokaryotes are the oldest cellular life forms on Earth, dating back 3.5–4 billion years and predating eukaryotes by 1 billion years. The current phylogenetic analysis of the conserved domains of sPLA<sub>2s</sub> of the representatives of the major prokaryote and eukaryote taxa supports the hypothesis that the sPLA<sub>2s</sub> of the eukaryotes, including the Metazoa, Viridiplantae and Fungi, may have shared a common origin with their homologues in bacteria. Two sPLA<sub>2</sub> groups (the cd-collection and the pfam-collection) emerged early in evolution and underwent distinct evolutionary trajectories. Based upon the retrieved data and phylogenetic relatedness, a model for the evolution of the two sPLA<sub>2s</sub> group members is proposed centred around gene-duplication and gene-deletion events (Fig. 7A,B). While the two members of the pfam-collection are present in representatives of the Eubacteria and Animalia kingdoms and maintained throughout evolution (Fig. 7B), the cd-collection members seem to have mainly emerged in the Animalia kingdom (Fig. 7A) and their expansion may be associated with the gene duplications that are proposed to have contributed to the increase in organismal complexity during eukaryote evolution [59]. The cd04706 protein members are exclusively found in Eubacteria and Plantae kingdoms and were lost from other life forms. Despite the lack of data from representatives of all clades and the failure to identify homologues it can be hypothesized that in the kingdom Animalia, sPLA<sub>2s</sub> with the conserved domains cd00125 and cd04704 were the first members to emerge. Subsequently, several independent gene or genome-duplication events occurred and resulted in the emergence of two novel family members of vertebrate cd04704 and mammalian cd04705. The sequence



**Fig. 7.** Proposed evolution of sPLA<sub>2s</sub> in the six Kingdoms of life (Animalia, Fungi, Plantae, Protista, Archaeobacteria and Eubacteria). Models were constructed using the retrieved data obtained and phylogenetic relationships. (A) Hypothetical evolution of the sPLA<sub>2</sub> cd-collection members. (B) Hypothetical evolution of the pfam-collection members. Dashed arrows indicate unclear evolutionary trajectories and '?' not yet identified.

similarity observed between the vertebrate sPLA<sub>2s</sub> containing conserved domains cd04707 and cd00125 and also the mammalian cd04705 with cd04704 suggests that they have a shared common origin and that cd04707 emerged from a gene-duplication event of the cd00125-like ancestral gene at the time of vertebrate emergence and that cd04705 resulted from a later gene-duplication event of the cd04704-like ancestral gene precursor within the mammalian lineage (Fig. 7A). In contrast, only fungi, protists and aquatic invertebrates, such as those of the phyla Cnidaria and Mollusca, have acquired pfam09056-containing sPLA<sub>2s</sub> and the homologue gene seems to have been deleted from Protostomes and Deuterostome genomes (Fig. 7B).

The present analysis is based on the highly conserved peptide motifs directly involved in the catalytic function of sPLA<sub>2</sub>s, including the Ca<sup>2+</sup>-binding site and the catalytic centre. However, the functional roles of the surrounding domains are at present less well understood (e.g. the functions of the domains flanking the central catalytically active cd04704 of mammalian GIII PLA<sub>2</sub>s are unknown) [3] and their study may provide novel insight into the catalytic and noncatalytic functions of sPLA<sub>2</sub>s. For instance, binding of sPLA<sub>2</sub>s to specific cellular receptors is independent of their catalytic activity, and the protein domains involved in the receptor activation are not yet fully resolved [4]. Other examples are the toxicity of some snake venom PLA<sub>2</sub>s, which does not correlate with their catalytic activity [8], and the bactericidal effects of sPLA<sub>2</sub>s lacking catalytic activity [60].

It is concluded that the sPLA<sub>2</sub>s of eukaryotes share their evolutionary origin with two distinct types of bacterial sPLA<sub>2</sub>s. Their evolution and prevalence in genomes seems to be related to the functional constraints of phospholipid metabolism which is a fundamental and conserved process in organisms. Although relatively little is known about the function of prokaryotic sPLA<sub>2</sub>s, the large number of distinct sPLA<sub>2</sub> isoforms in metazoans reflects the wide variation of substrate types encountered in the extracellular environment where the enzyme plays many important roles in nutrition, reproduction and immunity.

## Material and methods

### Database mining and data collection

Secreted PLA<sub>2</sub> sequences were retrieved from the publicly available protein databases of NCBI (<http://www.ncbi.nlm.nih.gov>) and Swiss-Prot (<http://www.uniprot.org>) using the Basic Local Alignment Search Tool (BLASTp) algorithm [61] and default settings. Database searches were performed using the peptide sequences of the human GIB PLA<sub>2</sub> ([P04054](#)), honeybee venom GIII PLA<sub>2</sub> ([P00630](#)), human GXIIA PLA<sub>2</sub> ([Q9BZM1](#)), *O. sativa* GXIB PLA<sub>2</sub> ([Q9XG81](#)) and *S. violaceoruber* PLA<sub>2</sub> ([Q6UV28](#)). In addition, PLA<sub>2</sub> and PLA<sub>2</sub>-like protein sequences were identified in the Microbial and Eukaryotic Genome database (<http://www.ncbi.nlm.nih.gov/genome>) following a similar strategy. Searches were also performed on nucleotide database data and covered all completed genomes in the NCBI genome database (1359 bacterial, 79 archaeal and 231 eukaryotic genomes; October 2010 release), and also available EST data using the tBLASTn and sequence matches with an e-value of < 10 were retrieved and their sequence analysed. The deduced protein sequences were obtained using the BCM Search Launcher (<http://searchlauncher.bcm.tmc.edu/>

[seq-util/Options/sixframe.html](#)) and compared with available homologue data.

### *In silico* sequence annotations

The conserved domains of sPLA<sub>2</sub> were identified using the NCBI Conserved Domains Database CDD-27036 PSSMs (<http://www.ncbi.nlm.nih.gov/cdd>) annotation by using the sequences identified in this study as queries. The result includes an alignment between the query and the search model consensus sequence, the expected-value for the alignment, the identity (name) of the conserved domain and the location of the conserved domain in the query sequence [25]. Histidine active-site and aspartic acid active-site protein motifs were identified based on the PROSITE database (<http://au.expasy.org/prosite>) pattern annotation. The localization of disulfide bonds was retrieved from the Swiss-Prot database and from published data.

### Sequence comparisons and phylogenetic analysis

Pairwise and multiple protein sequence alignments were carried out using the ClustalW2 program [62] available from EBI (<http://www.ebi.ac.uk/Tools/msa/clustalw2>) and the default parameters. The percentage of sequence similarities (based on the observed substitutions of one amino acid for another in homologous proteins) and identities between the sPLA<sub>2</sub>s were calculated based upon protein alignments using the GeneDoc interface (<http://www.psc.edu/biomed/genedoc>). Phylogenetic analysis of the conserved domain sequences of bacterial, fungal and metazoan PLA<sub>2</sub>s was performed using 53 taxa representatives. The protein alignment produced was submitted to PROTTEST analysis (<http://darwin.uvigo.es/software/prottest.html>) to select the best model of protein evolution that fits the data set [63]. Phylogenetic analyses were conducted using the NJ [64] and ML methods, and reliability for internal branching was assessed using the bootstrap method [65]. NJ analysis was performed using MEGA4 programme [66] and the p-distance amino acid model with 1000 bootstrap replicates. All positions containing alignment gaps and missing data were eliminated (pairwise deletion option) and a total of 215 positions were analyzed in the final data set. The ML tree (PHYML, v3.0 aLRT) [67] was constructed in Phylogeny.fr web interface (<http://phylogeny.lirmm.fr/phylo.cgi/index.cgi>). The WAG substitution model was selected assuming an estimated proportion of invariant sites (of 0.167) and four gamma-distributed rate categories to account for rate heterogeneity across sites. The gamma shape parameter was estimated directly from the data ( $\gamma = 1.804$ ) and analysis was performed using 100 bootstrap replicates. Graphical representation and edition of the phylogenetic tree were performed with TREEDYN (v198.3, <http://www.treedyn.org>).

Both methods produced similar tree topologies and the NJ bootstrap consensus tree was selected and taxa clades with support values < 50% collapsed.

## Acknowledgements

The authors thank Deborah Power for critical reading of the manuscript and anonymous referees for valuable suggestions. Supported by the Research Fund of Turku University Hospital and the Portuguese National Science Foundation (FCT)/CCMAR pluriannual grant.

## References

- Schaloske RH & Dennis EA (2006) The phospholipase A<sub>2</sub> superfamily and its numbering system. *Biochim Biophys Acta* **1761**, 1246–1259.
- Six DA & Dennis EA (2000) The expanding superfamily of phospholipase A<sub>2</sub> enzymes: classification and characterization. *Biochim Biophys Acta* **1488**, 1–19.
- Murakami M, Taketomi Y, Girard C, Yamamoto K & Lambeau G (2010) Review. Emerging roles of secreted phospholipase A<sub>2</sub> enzymes: lessons from transgenic and knockout mice. *Biochimie* **92**, 561–582.
- Murakami M, Taketomi Y, Miki Y, Sato H, Hirabayashi T & Yamamoto K (2011) Recent progress in phospholipase A<sub>2</sub> research: from cells to animals to humans. *Prog Lipid Res* **50**, 152–192.
- Nevalainen TJ, Haapamäki MM & Grönroos JM (2000) Roles of secretory phospholipases A<sub>2</sub> in inflammatory diseases and trauma. *Biochim Biophys Acta* **1488**, 83–90.
- Nevalainen TJ, Graham GG & Scott KF (2008) Antibacterial actions of secreted phospholipases A<sub>2</sub>: review. *Biochim Biophys Acta* **1781**, 1–9.
- Birts CN, Barton CH & Wilton DC (2010) Catalytic and non-catalytic functions of human IIA phospholipase A<sub>2</sub>. *Trends Biochem Sci* **35**, 28–35.
- Valentin E & Lambeau G (2000) Increasing diversity of secreted phospholipases A<sub>2</sub> and their receptors and binding proteins. *Biochim Biophys Acta* **1488**, 59–70.
- Dijkstra BW, Kalk KH, Hol WG & Drenth J (1981) Structure of bovine pancreatic phospholipase A<sub>2</sub> at 1.7 Å resolution. *J Mol Biol* **147**, 97–123.
- Brunie S, Bolin J, Gewirth D & Sigler PB (1985) The refined crystal structure of dimeric phospholipase A<sub>2</sub> at 2.5 Å. Access to a shielded catalytic center. *J Biol Chem* **260**, 9742–9749.
- White SP, Scott DL, Otwinowski Z, Gelb MH & Sigler PB (1990) Crystal structure of cobra-venom phospholipase A<sub>2</sub> in a complex with a transition-state analogue. *Science* **250**, 1560–1563.
- Wery JP, Schewitz RW, Clawson DK, Bobbitt JL, Dow ER, Gamboa G, Goodson T Jr, Hermann RB, Kramer RM, McClure DB *et al.* (1991) Structure of recombinant human rheumatoid arthritic synovial fluid phospholipase A<sub>2</sub> at 2.2 Å resolution. *Nature* **352**, 79–82.
- Steiner RA, Rozeboom HJ, de Vries A, Kalk KH, Murshudov GN, Wilson KS & Dijkstra BW (2001) X-ray structure of bovine pancreatic phospholipase A<sub>2</sub> at atomic resolution. *Acta Crystallogr D Biol Crystallogr* **57**, 516–526.
- Renetseder R, Brunie S, Dijkstra BW, Drenth J & Sigler PB (1985) A comparison of the crystal structures of phospholipases A<sub>2</sub> from bovine pancreas and *Crotalus atrox* venom. *J Biol Chem* **260**, 11627–11634.
- Guy J, Ståhl U & Lindqvist Y (2009) Crystal structure of a class XIB phospholipase A<sub>2</sub>: rice (*Oryza sativa*) isoform-2 PLA<sub>2</sub> and an octanoate complex. *J Biol Chem* **284**, 19371–19379.
- Matoba Y, Katsube Y & Sugiyama M (2002) The crystal structure of prokaryotic phospholipase A<sub>2</sub>. *J Biol Chem* **277**, 20059–20069.
- Scott DL, Otwinowski Z, Gelb MH & Sigler PB (1990) Crystal structure of bee venom phospholipase A<sub>2</sub> in a complex with a transition-state analogue. *Science* **250**, 1563–1566.
- Hulo N, Bairoch A, Bulliard V, Cerutti L, Cuche B, De Castro E, Lachaize C, Langendijk-Genevaux PS & Sigrist CJA (2008) The 20 years of PROSITE. *Nucleic Acid Res* **36**, D245–D249.
- Talvinen KA & Nevalainen TJ (2002) Cloning of a novel phospholipase A<sub>2</sub> from the cnidarian *Adamsia cariniopados*. *Comp Biochem Physiol B Biochem Mol Biol* **132**, 571–578.
- Razpotnik A, Križaj I, Šričar J, Kordiš D, Maček P, Frangež R, Kem WR & Turk T (2010) A new phospholipase A<sub>2</sub> isolated from the sea anemone *Urticina crassicornis* – its primary structure and phylogenetic classification. *FEBS J* **277**, 2641–2653.
- Romero L, Marcussi S, Marchi-Salvador DP, Silva FP Jr, Fuly AL, Stábeli RG, da Silva SL, Gonzáles J, del Monte A & Soares AM (2010) Enzymatic and structural characterization of a basic phospholipase A<sub>2</sub> from the sea anemone *Condylactis gigantea*. *Biochimie* **92**, 1063–1071.
- Davidson FF & Dennis EA (1990) Evolutionary relationships and implications for the regulation for phospholipase A<sub>2</sub> from snake venom to human secreted forms. *J Mol Evol* **31**, 228–238.
- Dennis EA (1994) Diversity of group types, regulation, and function of phospholipase A<sub>2</sub>. *J Biol Chem* **269**, 13057–13060.
- Finn RD, Tate J, Mistry J, Coghill PC, Sammut SJ, Hotz HR, Ceric G, Forslund K, Eddy SR, Sonnhammer ELL *et al.* (2008) The pfam protein families database. *Nucleic Acid Res* **36**, D281–D288.

- 25 Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR *et al.* (2010) CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acid Res* **39**, D225–D229.
- 26 Beres SB, Sylva GL, Barbian KD, Lei B, Hoff JS, Mammarella ND, Liu MY, Smooth JC, Porcella SF, Parkins LD *et al.* (2002) Genome sequence of a serotype M3 strain of group A *Streptococcus*: phage-encoded toxins, the high-virulence phenotype, and clone emergence. *PNAS* **99**, 10078–10083.
- 27 Ståhl U, Lee M, Sjö Dahl S, Archer D, Cellini F, Ek B, Iannaccone R, MacKenzie D, Semeraro L, Tramontano E *et al.* (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.
- 28 Nevalainen TJ (2008) Phospholipase A<sub>2</sub> in the genome of *Nematostella vectensis*. *Comp Biochem Physiol D Genom Proteom* **3**, 226–233.
- 29 Pote KG, Hauer CR III, Michel H, Shabanovitz J, Hunt DF & Kretsinger RH (1993) Otoconin-22, the major protein aragonitic frog otoconia, is homolog of phospholipase A<sub>2</sub>. *Biochemistry* **32**, 5017–5024.
- 30 Valentin E, Ghomashchi F, Gelb MH, Lazdunski M & Lambeau G (2000) Novel human secreted phospholipase A<sub>2</sub> with homology to the group III bee venom enzyme. *J Biol Chem* **275**, 7492–7496.
- 31 Sugiyama M, Ohtani K, Izuhara M, Koike T, Suzuki K, Imamura S & Misaki H (2002) A novel prokaryotic phospholipase A<sub>2</sub>. Characterization, gene cloning, and solution structure. *J Biol Chem* **277**, 20051–20058.
- 32 Soragni E, Bolchi A, Balestrini R, Gambaretto C, Percudani R, Bonfante P & Ottonello S (2001) A nutrient-regulated, dual localization phospholipase A<sub>2</sub> in the symbiotic fungus *Tuber borchii*. *EMBO J* **20**, 5079–5090.
- 33 Meng X, Chang Y, Qui X & Wang X (2010) Generation and analysis of expressed sequence tags from adductor muscle of Japanese scallop *Mizuhopecten yessoensis*. *Comp Biochem Physiol Part D Genomics Proteomics* **5**, 288–294.
- 34 McIntosh JM, Ghomashchi F, Gelb MH, Dooley DJ, Stoehr SJ, Giordani AB, Naisbitt SR & Olivera BM (1995) Conodipine-M, a novel phospholipase A<sub>2</sub> isolated from the venom of the marine snail *Conus magus*. *J Biol Chem* **270**, 3518–3526.
- 35 Gelb MH, Valentin E, Ghomashchi F, Lazdunski M & Lambeau G (2000) Cloning and recombinant expression of a structurally novel human secreted phospholipase A<sub>2</sub>. *J Biol Chem* **275**, 39823–39826.
- 36 Rouault M, Bollinger JG, Lazdunski M, Gelb MH & Lambeau G (2003) Novel mammalian group XII secreted phospholipase A<sub>2</sub> lacking enzymatic activity. *Biochemistry* **42**, 11494–11503.
- 37 Bókay A (1877) Ueber die Verdaulichkeit des Nucleins und Lecithins. *Ztschr Physiol Chem* **1**, 157–164.
- 38 Burke JE & Dennis EA (2009) Phospholipase A<sub>2</sub> biochemistry. *Cardiovasc Drug Ther* **23**, 49.
- 39 Tan PTJ, Khan MA & Brusica V (2003) Bioinformatics for venom and toxin sciences. *Brief Bioinform* **4**, 53–62.
- 40 Lynch VJ (2007) Inventing an arsenal: adaptive evolution and neofunctionalization of snake venom phospholipase A<sub>2</sub> genes. *BMC Evol Biol* **7**, 2.
- 41 Heinrikson RL, Krueger ET & Keim PS (1977) Amino acid sequence of phospholipase A<sub>2</sub>-alpha from the venom of *Crotalus adamanteus*. A new classification of phospholipases A<sub>2</sub> based on structural determinants. *J Biol Chem* **252**, 4913–4921.
- 42 Wang Y, Kowalski PE, Thalman I, Ornitz DM, Mager DL & Thalmann L (1998) Otoconin-90, the mammalian otoconial matrix protein, contains two domains of homology to secretory phospholipase A<sub>2</sub>. *Proc Natl Acad Sci USA* **95**, 15345–15350.
- 43 Zádori Z, Szelei J, Lacoste MC, Li Y, Gariépy S, Raymond P, Allaire M, Nabi IR & Tijssen P (2001) A viral phospholipase A<sub>2</sub> is required for parvovirus infectivity. *Dev Cell* **1**, 291–302.
- 44 Woese CR, Kandler O & Wheelis ML (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eukarya. *PNAS* **87**, 4576–4579.
- 45 Hemmi H, Shibuya K, Takahashi Y, Nakayama T & Nishino T (2004) (S)-2,3-Di-O-geranylgeranylglycerol phosphate synthase from the thermoacidophilic archeon *Sulfolobus solfataricus*. Molecular cloning and characterization of a membrane-intrinsic prenyltransferase involved in the biosynthesis of archaeal ether-linked membrane lipids. *J Biol Chem* **279**, 50197–50202.
- 46 Cavalier-Smith T (2009) Review. Predation and eukaryote cell origins: a coevolutionary perspective. *Int J Biochem Cell Biol* **41**, 307–322.
- 47 Nakashima K, Ogawa T, Oda N, Shimohigashi Y, Hattori M, Sakaki Y, Kihara H & Ohno M (1994) Darwinian evolution of *Trimeresurus flavoviridis* venom gland phospholipase A<sub>2</sub> isozymes. *Pure Appl Chem* **66**, 715–720.
- 48 Kini RM & Chan YM (1999) Accelerated evolution and molecular surface of venom phospholipase A<sub>2</sub> enzymes. *J Mol Evol* **48**, 125–132.
- 49 Fry BG (2005) From genome to “venome”: molecular origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences and related body proteins. *Genome Res* **15**, 403–420.
- 50 Golik M, Cohen-Zinder M, Loor JJ, Drackley JK, Band MR, Lewin HA, Weller JI, Ron M & Seroussi E (2006) Accelerated expansion of group IID-like

- phospholipase A<sub>2</sub> genes in *Bos taurus*. *Genomics* **87**, 527–533.
- 51 Tischfield JA (1997) A reassessment of the low molecular weight phospholipase A<sub>2</sub> gene family in mammals. *J Biol Chem* **272**, 17247–17250.
- 52 Koduri RS, Grönroos JO, Laine VJO, Le Calvez C, Lambeau G, Nevalainen TJ & Gelb MH (2002) Bactericidal properties of human and murine groups I, II, V, X, and XII secreted phospholipases A<sub>2</sub>. *J Biol Chem* **277**, 5849–5857.
- 53 Chen JY, Oliveri P, Gao F, Dornbos SQ, Li CW, Bottjer DJ & Davidson EH (2002) Precambrian animal life: probable developmental and adult cnidarian forms from Southwest China. *Dev Biol* **248**, 182–196.
- 54 Wood RA, Grotzinger JP & Dickson JA (2002) Proterozoic modular biomineralized metazoan from the Nama group, Namibia. *Science* **296**, 2383–3286.
- 55 Armugam A, Gong NL, Li XJ, Siew PY, Chai SC, Nair R & Jeyaseelan K (2004) Group IB phospholipase A<sub>2</sub> from *Pseudonaja textilis*. *Arch Biochem Biophys* **421**, 10–20.
- 56 Putnam NH, Srivastava M, Hellsten U, Dirks B, Chapman J, Salamov A, Terry A, Shapiro H, Lindquist E, Kapitonov VV *et al.* (2007) Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* **317**, 86–94.
- 57 Ho I, Arm JP, Bingham CO, Choi A, Austen KF & Glimcher LH (2001) A novel group of phospholipase A<sub>2</sub>s preferentially expressed in type 2 helper T cells. *J Biol Chem* **276**, 18321–18326.
- 58 Guan M, Qu L, Tan W, Chen L & Wong C (2011) Hepatocyte nuclear factor-4 alpha regulates liver triglyceride metabolism in part through secreted phospholipase A<sub>2</sub> GXIIB. *Hepatology* **53**, 458–466.
- 59 Crow KC & Wagner GP (2006) What is the role of genome duplication in the evolution of complexity and diversity? *Mol Biol Evol* **23**, 887–892.
- 60 Páramo L, Lomonte B, Pizarro-Cerdá J, Bengoechea JA, Gorvel JP & Moreno P (1998) Bactericidal activity of Lys-49 and Asp-49 myotoxic phospholipase A<sub>2</sub> from *Bothrops asper* snake venom: synthetic Lys49 myotoxin II-(115-129)-peptide identifies its bactericidal region. *Eur J Biochem* **253**, 452–461.
- 61 Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W & Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402.
- 62 Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R *et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948.
- 63 Abascal F, Zardoya R & Posada D (2005) ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* **21**, 2104–2105.
- 64 Saitou N & Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406–425.
- 65 Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791.
- 66 Tamura K, Dudley J, Nei M & Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**, 1596–1599.
- 67 Guindon S & Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**, 696–704.

## Supporting information

The following supplementary material is available:

**Fig. S1.** Alignment of the conserved domain pfam06951 sequences of 20 GXIIA PLA<sub>2</sub>s.

**Table S1.** Conserved domains of 333 prokaryotic and eukaryotic secreted phospholipases A<sub>2</sub>.

This supplementary material can be found in the online version of this article.

Please note: As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.