HAYATI
Journal of Biosciences

Vol. 32 No. 3, May 2025 768-780
DOI:10.4308/hjb.32.3.768-780
ISSN: 1978-3019
EISSN: 2086-4094

Research Article

Comprehensive Characterization of Phospholipase C and D Families in Cocoa (*Theobroma cacao* L.): Identification, Phylogenetics, Gene Structure, and Transcriptomic Insights

Quy Phuong Nguyen[1], Huyen Thi Thanh Tran[2], Huong Thi Thanh Nguyen[1], Gioi Huy Dong[3], Quyen Thi Ha[4], Quyen Thi Xuan Vu[1], Chi Toan Le[5], Phi Bang Cao[1], Ha Duc Chu[4*]

[1]*Faculty of Natural Sciences, Hung Vuong University, Phu Tho Province 35000, Vietnam*
[2]*Faculty of Biology, Hanoi National University of Education, Hanoi City 122300, Vietnam*
[3]*Faculty of Biotechnology, Vietnam National University of Agriculture, Hanoi City 122300, Vietnam*
[4]*Faculty of Agricultural Technology, University of Engineering and Technology, Vietnam National University Hanoi, Hanoi City 122300, Vietnam*
[5]*Faculty of Biology, Hanoi Pedagogical University 2, Vinh Phuc Province 280000, Vietnam*

ARTICLE INFO

Article history:
Received December 18, 2024
Received in revised form January 29, 2025
Accepted January 31, 2025

KEYWORDS:
Cocoa,
phospholipase C,
phospholipase D,
gene family,
characterization,
expression profile

ABSTRACT

This study provides a detailed investigation of the phospholipase C (*PLC*) and phospholipase D (*PLD*) gene families in cocoa (*Theobroma cacao*), focusing on their identification, characterization, and expression patterns. A total of 10 *PLC* and 12 *PLD* genes was identified and systematically annotated based on their sequence homology, conserved domains, and functional classification, adhering to established nomenclature. Analysis of physicochemical properties revealed diversity in molecular weights, isoelectric points, and stability parameters, reflecting their structural and functional variability. Phylogenetic analysis classified the genes into distinct subfamilies and highlighted their evolutionary relationships with homologs in *Arabidopsis thaliana* and rice (*Oryza sativa*). Gene structure analysis demonstrated significant variation in exon-intron organization, indicating functional specialization and regulatory complexity within these gene families. Expression profiling during cocoa embryo development showed that certain genes, such as *TcNPC2, TcPI-PLC5*, and *TcPLDα1*, were highly expressed, while others exhibited stage-specific activity. In response to Phytophthora megakarya infection, several *PLC* and *PLD* genes displayed significant changes in expression across different time points and genotypes, including the upregulation of *TcPI-PLC2, TcPLDα5,* and *TcPLDζ2*, suggesting their roles in cocoa's stress responses and defense mechanisms. These findings offer new insights into the biological roles of *PLC* and *PLD* gene families in cocoa, particularly in growth, development, and stress adaptation, providing a solid foundation for further functional research and potential applications in cocoa improvement programs.

1. Introduction

Cocoa (*Theobroma cacao* L.), a perennial plant native to the tropical rainforests of the Amazon basin in South America (Lanaud *et al.* 2024), holds significant agricultural importance (Latif 2013; Wickramasuriya & Dunwell 2018). After its domestication, the cultivation of cocoa expanded globally to various tropical regions, including West Africa, Southeast Asia, and Central America. Serving as the fundamental ingredient in the chocolate industry, cocoa is economically vital, supporting the livelihoods of millions of smallholder farmers and contributing notably to the economies of producing nations (Latif 2013; Adeniyi 2019). The successful cultivation of cocoa hinges on specific environmental conditions characterized by warm temperatures, high humidity, and evenly distributed

* Corresponding Author
E-mail Address: cd.ha@vnu.edu.vn

rainfall throughout the year. However, cocoa plants are particularly sensitive to adverse environmental factors like drought, extreme temperatures, and pest and disease infestations (Schroth *et al.* 2016; Gateau-Rey *et al.* 2018). These stressors can significantly impair the plants' growth, development, and productivity, decreasing yields and diminishing bean quality. The challenges posed by climate change exacerbate these issues by increasing the frequency and intensity of environmental stresses, thereby threatening the sustainability of cocoa production systems (Gateau-Rey *et al.* 2018). It has been strongly believed that gaining a molecular-level understanding of how cocoa plants respond to environmental stresses is crucial for developing strategies to mitigate the negative impacts of these conditions. Investigating the molecular mechanisms underlying stress responses can identify key pathways and genes associated with stress tolerance (Romero Navarro *et al.* 2017; Pokou *et al.* 2019). This understanding is essential for breeding more resilient cocoa varieties and implementing effective management practices. Therefore, advancing research on the molecular responses of cocoa to stress conditions is imperative for enhancing the crop's adaptability and ensuring the long-term sustainability of cocoa production.

In plants, the phospholipase superfamily comprises enzymes that are essential for lipid metabolism and signal transduction (Ali *et al.* 2022). These enzymes hydrolyze phospholipids, leading to the generation of various bioactive molecules that influence cellular processes. Among them, phospholipase C (PLC) and D (PLD) are particularly significant due to their prominent roles in plant physiology, including growth, development, and stress responses (Apostolakos *et al.* 2008; Rupwate & Rajasekharan 2012; Yang *et al.* 2015; Ali *et al.* 2022; Fang *et al.* 2023). Particularly, PLC enzymes are characterized by conserved catalytic domains known as the X and Y domains, which are crucial for their enzymatic activity (Bill & Vines 2020). They also contain EF-hand motifs that enable calcium binding, indicating that intracellular calcium levels regulate their activity (Rupwate & Rajasekharan 2012), while PLD enzymes possess two conserved HKD motifs (named after the critical histidine, lysine, and aspartic acid residues) necessary for their catalytic function (Deepika & Singh 2022). The structural features of PLC and PLD allow them to interact with specific substrates and participate in intricate signaling networks within the plant cell (Bargmann & Munnik 2006; Rupwate & Rajasekharan 2012). For instance, these families have been comprehensively identified in *Arabidopsis thaliana* (Qin & Wang 2002), rice (*Oryza sativa*) (Singh *et al.* 2013), maize (*Zea mays*) (Chen *et al.* 2016; Zhu *et al.* 2020), two *Corchorus* species (Sadat *et al.* 2022), upland cotton (*Gossypium hirsutum*) (Song *et al.* 2017), sorghum (*Sorghum bicolor*) (Wei *et al.* 2022), tomato (*Solanum lycopersicum*) (Guo *et al.* 2024), peanut (*Arachis hypogaea*) (Zhang *et al.* 2023), three orchid species (Kanchan *et al.* 2021). However, little information on the PLC and PLD families in cocoa, even the genome and proteome databases of this important industrial crop species, have been published previously (Argout *et al.* 2011).

This study aims to comprehensively investigate the PLC and PLD enzyme families in cocoa, focusing on their identification, characterization, and functional insights. It elucidated the evolutionary relationships of these enzyme families through phylogenetic analysis and revealed structural variations in their exon-intron organization. we also investigated the expression patterns of *PLC* and *PLD* genes across various tissues and under biotic stress conditions, offering insights into their roles in stress responses and growth regulation. This research provides a comprehensive understanding of the *PLC* and *PLD* families in cocoa, offering a foundation for future functional studies and developing strategies to enhance stress resilience and productivity in this economically important crop.

2. Materials and Methods

2.1. Survey of PLCs and PLDs in Cocoa

To conduct a survey of the PLC and PLD families in the cocoa assembly, a systematic bioinformatics approach was undertaken, utilizing well-established PLCs and PLDs from *A. thaliana* (Qin & Wang 2002) as reference sequences. Subsequently, the retrieved *Arabidopsis* PLC and PLD sequences were employed as query sequences in homology searches against the cocoa genome and proteome databases (Argout *et al.* 2011). BLASTP and TBLASTN algorithms were used to identify potential homologs in cocoa, with criteria set for sequence similarity (E-value thresholds of 1e-5 or lower) to ensure the consistency of the matches. All candidate cocoa PLC and PLD sequences identified from the BLAST searches were then subjected to the Pfam server (Mistry *et al.* 2021) to confirm the presence of key functional domains characteristic of PLC and PLD enzymes, as previously provided (Rupwate & Rajasekharan 2012; Bill & Vines 2020; Deepika & Singh 2022).

2.2. Calculation of the Physicochemical Characteristics of PLCs and PLDs in Cocoa

To estimate the physicochemical characteristics of the PLC and PLD proteins in cocoa, a series of bioinformatics analyses were performed using computational tools designed to predict protein attributes as previously described (Chu *et al.* 2018, 2024a). Initially, amino acid sequences of the identified PLCs and PLDs were retrieved from the cocoa proteome database. These sequences were then analyzed using the ExPASy ProtParam tool (Gasteiger *et al.* 2003), which provides key physicochemical properties, including molecular mass (MM), theoretical isoelectric point (pI), aliphatic index (AI), and grand average of hydropathy (GRAVY).

2.3. Phylogenetic Tree Construction of PLCs and PLDs in Cocoa

To construct the phylogenetic tree of the PLCs and PLDs in cocoa, well-characterized PLC and PLD proteins from *A. thaliana* (At) (Qin & Wang 2002) and rice (Os) (Singh *et al.* 2013) were utilized as reference sequences. First, full-length amino acid sequences of cocoa PLC and PLD proteins were retrieved from genomic databases, while protein sequences of AtPLCs, AtPLDs, OsPLCs, and OsPLDs were obtained from recent studies (Singh *et al.* 2013; Qin & Wang 2002).

These protein sequences were aligned using Clustal Omega (Larkin *et al.* 2007; Thompson *et al.* 2002). A phylogenetic tree was then constructed using the Maximum Likelihood method implemented in MEGA software (Kumar *et al.* 2018), with a substitution model to optimize the fit of the evolutionary model to the data. Bootstrap analysis with 1,000 replicates was performed to evaluate the statistical robustness of the inferred tree topology.

2.4. Exon-intron Organization of *PLC*s and *PLD*s in Cocoa

To analyze the gene structure of *PLC* and *PLD* genes in cocoa, the genomic sequences and their corresponding coding sequences were obtained from publicly available cocoa genome databases (Argout *et al.* 2011). These sequences were used as input for the Gene Structure Display Server (Hu *et al.* 2015) to visualize the exon-intron organization of genes as previously described (La *et al.* 2022a; La *et al.* 2022b; Chu *et al.* 2024a). The genomic sequences were aligned with the coding sequences to identify and map the positions of exons, introns, and untranslated regions within each *PLC* or *PLD* gene. A graphical representation of the gene structure was then illustrated, displaying the number, length, and arrangement of exons and introns.

2.5. RNA-Seq Analysis of *PLC*s and *PLD*s in Response of Cocoa to *Phytophthora megakarya* Infection

We explored the recent RNA-Seq dataset related to biotic treatment as previously described (Chu *et al.* 2024b). The RNA-Seq dataset associated with GEO accession number GSE116041 was first accessed and downloaded from the NCBI GEO database (Barrett *et al.* 2013). This dataset described the transcriptomic response of cocoa to *Phytophthora megakarya* infection. Briefly, four-month-old grafted plants of two genotypes, SCA6 and NA32, were spray-inoculated with *P. megakarya* zoospores or distilled water (control). Leaf samples collected at 0, 6, 24, 48, and 72 hours after inoculation (hai) were utilized for RNA extraction. Libraries were sequenced on the Illumina HiSeq 2500 platform. Relative expression values of *PLC* and *PLD* genes were estimated using the *Actin 11* gene as previously described (Pinheiro *et al.* 2011). Genes were classified as upregulated or down-regulated based on a fold-change threshold of $|\text{fold-change}| \geq 1.5$ when comparing expression levels at 6, 24, and 72 hai to those at 0 hai (Chu *et al.* 2024b).

2.6. RNA-Seq Analysis of *PLC*s and *PLD*s During the Embryo Development

We explored the available RNA-Seq dataset during the somatic embryogenesis development as previously described (Chu *et al.* 2024b). The RNA-Seq dataset associated with GEO accession number GSE55476 (Maximova *et al.* 2014) was obtained from GEO NCBI (Barrett *et al.* 2013). Particularly, different types and stages of embryo development, including zygotic embryo tissues: torpedo (T-ZE), early-full (EF-ZE), late-full (LF-ZE), and mature (M-ZE) embryos; and somatic embryos: late torpedo (LT-SE) and mature (M-SE) stages, were collected for RNA-Seq. The relative expression levels of the *PLC* and *PLD* genes were normalized to the expression level of the *Actin 11* gene, as previously described (Pinheiro *et al.* 2011).

3. Results

3.1. Identification and Annotation of the PLCs and PLDs in Cocoa

To comprehensively survey PLCs and PLDs in the cocoa assembly, a homology-based method has been conducted by using PLCs and PLDs from *A. thaliana*. After carefully validating, the information on all putative PLCs and PLDs in cocoa was provided in Tables 1 and 2, respectively.

As expected, a total of 10 and 12 members of the PLCs and PLDs has been fully reported in the cocoa assembly. The nomenclature of PLC and PLD enzymes in cocoa is systematically designed to reflect their sequence homology, conserved domains, and functional characteristics, following conventions established in model plants like *A. thaliana*. The prefix "Tc" denotes the species abbreviation for *Theobroma cocoa*. For PLC enzymes, the naming is categorized into non-specific PLCs (NPCs) and phosphoinositide-specific PLCs (PI-PLCs). Based on the physical order in the cocoa genome, the NPCs are designated as TcNPC1 through TcNPC5 (e.g., TcNPC1, TcNPC2, TcNPC3, TcNPC4, TcNPC5), while the PI-PLCs are named TcPI-PLC1 through TcPI-PLC5 (e.g., TcPI-PLC1, TcPI-PLC2, TcPI-PLC3, TcPI-PLC4, TcPI-PLC5). For PLD enzymes, the general prefix "TcPLD" is followed by a lowercase letter corresponding to the PLD subclass transliterated from Greek letters to indicate the specific group, including α, β, γ, δ, and ζ. Numerical suffixes further distinguish individual genes within each subclass. For example, TcPLDα1 through TcPLDα5 (TcPLDα1, TcPLDα2, TcPLDα3, TcPLα4, TcPLDα5) represent five distinct genes within the PLDα subclass. Similarly, TcPLDδ1 through TcPLDδ3 (TcPLDδ1, TcPLDδ2, TcPLDδ3) denote genes within the PLDδ subclass, while TcPLDβ1, TcPLDγ1, and TcPLDζ1 through TcPLDζ2 (TcPLDζ1, TcPLDζ2) represent genes within the PLDβ, PLDγ, and PLDζ subclasses, respectively. This systematic nomenclature facilitates clear identification and classification of PLC and PLD family members in cocoa, enabling comparative analyses with other plant species.

Table 1. Annotation and characteristics of the PLCs in cocoa

Gene name	Clade	Locus name	Length (aa)	mW (kDa)	pI	GRAVY	AI
TcNPC1	NPC	Thecc.03G130600	521	57.80	6.96	-0.29	73.13
TcNPC2	NPC	Thecc.05G230600	513	57.60	5.09	-0.51	67.64
TcNPC3	NPC	Thecc.05G230700	516	58.56	5.61	-0.47	69.13
TcNPC4	NPC	Thecc.09G248300	536	60.23	6.23	-0.43	69.27
TcNPC5	NPC	Thecc.09G302700	534	59.73	8.95	-0.35	76.65
TcPI-PLC1	PI-PLC	Thecc.09G300800	586	67.33	7.11	-0.57	79.49
TcPI-PLC2	PI-PLC	Thecc.09G300900	596	68.12	7.57	-0.41	81.54
TcPI-PLC3	PI-PLC	Thecc.09G301100	591	67.11	6.30	-0.53	73.87
TcPI-PLC4	PI-PLC	Thecc.10G016100	586	66.65	5.93	-0.50	80.99
TcPI-PLC5	PI-PLC	Thecc.10G016200	597	67.62	6.18	-0.46	75.59

mW: Molecular weight, pI: Isoelectric point, GRAVY: Grand average of hydropathy, AI: Aliphatic index

Table 2. Annotation and characteristics of the PLDs in cocoa

Gene name	Clade	Locus name	Length (aa)	mW (kDa)	pI	GRAVY	AI
TcPLDα1	α	Thecc.06G021600	809	91.71	5.47	-0.39	82.42
TcPLDα2	α	Thecc.06G109800	765	87.32	7.37	-0.41	77.23
TcPLDα3	α	Thecc.09G025700	824	93.21	6.23	-0.41	80.15
TcPLDα4	α	Thecc.09G025800	829	93.89	6.58	-0.30	89.25
TcPLDα5	α	Thecc.09G127300	806	92.73	6.00	-0.42	83.52
TcPLDβ1	β	Thecc.01G370500	1118	124.96	6.41	-0.52	70.28
TcPLDγ1	γ	Thecc.03G268800	852	95.69	7.90	-0.37	81.98
TcPLDδ1	δ	Thecc.01G011000	853	96.91	6.60	-0.44	80.56
TcPLDδ2	δ	Thecc.02G090200	848	96.27	7.20	-0.37	81.72
TcPLDδ3	δ	Thecc.08G052300	845	95.82	7.38	-0.34	79.94
TcPLDζ1	ζ	Thecc.04G290500	1107	125.96	6.03	-0.39	82.46
TcPLDζ2	ζ	Thecc.05G317000	1109	126.17	5.68	-0.43	82.99

mW: Molecular weight, pI: Isoelectric point, GRAVY: Grand average of hydropathy, AI: Aliphatic index

3.2. Calculation of the Physicochemical Parameters of the PLCs and PLDs in Cocoa

The physicochemical parameters of the PLCs and PLDs in cocoa were systematically analyzed to provide insights into their molecular properties and potential functional roles. The MM scores of the identified PLC and PLD proteins ranged from 57.60 kDa to 68.12 kDa and 87.32 kDa to 126.17 kDa, respectively, indicating a variation in protein size across the PLC and PLD families. Similarly, the protein size of the PLCs and PLDs in cocoa ranged from 513 residues to 597 residues and 765 residues to 1118 residues, respectively. The theoretical pI values were calculated and found to range between 5.09 and 8.95 (in PLCs) and 5.47 and 7.90 (in PLDs), suggesting that the majority of these proteins are acidic, which may affect their solubility and stability under physiological conditions. Furthermore, the AI scores, which ranged from 67.64 to 81.54 (in PLCs) and 70.28 to 89.25 (in PLDs), indicate variability in thermostability, with higher values suggesting proteins are more stable in response to temperature fluctuations. The GRAVY scores of the PLC and PLD proteins varied from -0.29 to -0.57 and -0.30 to -0.52, respectively. All PLC and PLD proteins exhibited negative GRAVY values, indicating hydrophilicity. Together, these physicochemical and structural insights into cocoa PLC and PLD proteins enhance our understanding of their stability, solubility, and functional localization. They provide a foundation for further studies on their roles in lipid-mediated signaling and stress response mechanisms in cocoa.

3.3. Phylogenetic Insights Into the PLC and PLD Families in Cocoa

The phylogenetic analysis of PLC proteins in cocoa, based on well-characterized PLCs from *A. thaliana* (AtPLCs) and rice (OsPLCs), reveals a clear classification into distinct subfamilies corresponding to PI-PLCs and NPCs. Within the PI-PLC subfamily, cocoa PLCs cluster closely with homologs from *Arabidopsis* and rice, indicating conserved evolutionary relationships and functional similarities. For instance, TcPI-PLC2 and TcPI-PLC3 are closely related to AtPI-PLC4 and AtPI-PLC5. Similarly, TcPI-PLC5 aligns closely with OsPI-PLC3 (Figure 1). In the NPC subfamily, cocoa NPCs exhibit closer clustering with *Arabidopsis* NPCs, with TcNPC4 and TcNPC5 forming clades with AtNPC2 and AtNPC6, respectively. The close association of TcNPC1 and TcNPC3 with rice NPCs, such as OsNPC3 and OsNPC5, suggests shared evolutionary pressures and adaptations within monocots and dicots.

The classification of PLD proteins in cocoa, based on a phylogenetic comparison with PLDs from *A. thaliana* (AtPLDs) and rice (OsPLDs), reveals distinct clustering into the recognized PLD subfamilies: α, β, γ, δ, and ζ (Figure 2). Within the α subfamily, cocoa PLD proteins, such as TcPLDα1, TcPLDα2, TcPLDα3, TcPLDα4, and TcPLDα5 group closely with AtPLDα and OsPLDα proteins. The strong bootstrap values supporting these groupings highlight the reliability of the classification. In the β subfamily, TcPLDβ1 shows close relationships with AtPLDβ1, AtPLDβ2, and OsPLDβ1. The γ subfamily includes TcPLDγ1, which clusters alongside AtPLDγ1, AtPLDγ2, and AtPLDγ3. The δ subfamily includes three cocoa genes: TcPLDδ1, TcPLDδ2, and TcPLDδ3, which group with AtPLDδ and OsPLDδ. In the ζ subfamily, TcPLDζ1 and TcPLDζ2 align closely with AtPLDζ1, AtPLDζ2, and OsPLDζ1.

3.4. Gene Structure Analysis of the *PLC* and *PLD* Families in Cocoa

The gene structure analysis of *PLC* genes in cocoa reveals significant variation in exon-intron organization among different members of the PLC family. Genes encoding NPCs, such as *TcNPC1, TcNPC5, TcNPC4, TcNPC3,* and *TcNPC2*, exhibit relatively fewer exons compared to genes encoding PI-PLCs (Figure 3). For instance, *TcNPC1* and *TcNPC5* consist of four exons, indicating a simpler structure, while *TcNPC4* contains five exons, suggesting a slight increase in structural complexity. In contrast, the PI-PLC genes, including *TcPI-PLC1, TcPI-PLC2, TcPI-PLC3, TcPI-PLC4,* and *TcPI-PLC5*, display a more intricate exon-intron architecture. These genes typically have a higher number of exons, ranging from six to nine, which likely corresponds to their functional diversity and regulatory complexity. For example, *TcPI-PLC4* and *TcPI-PLC5* each have eight exons, while *TcPI-PLC2* and *TcPI-PLC3* contain six and seven exons, respectively.

Additionally, the intron lengths vary considerably among the *PLC* genes, with some, such as *TcNPC2*, showing relatively short introns. In contrast, others, like *TcPI-PLC1*, have longer introns, which may influence alternative splicing and gene expression regulation. The diversity in exon-intron organization reflects the evolutionary adaptation of *PLC* genes

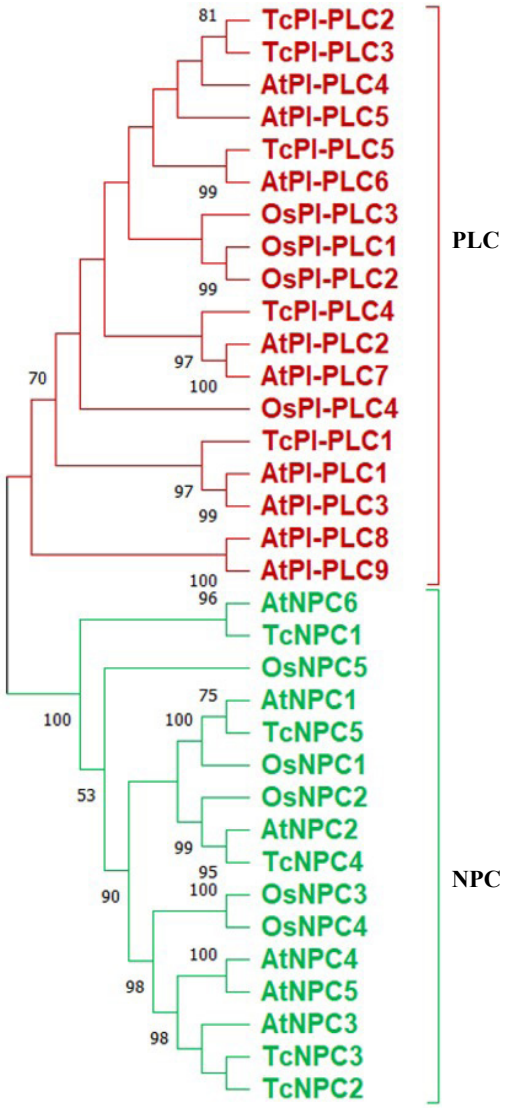

Figure 1. Phylogenetic relationship of the PLCs in cocoa, *Arabidopsis thaliana* and rice. Unrooted tree included non-specific phospholipase C (NPC) and phosphoinositide-specific phospholipase C (PI-PLC) subfamilies from cocoa (Tc), *A. thaliana* (At) and rice (Os). Bootstrap values are provided to indicate the reliability of major clades

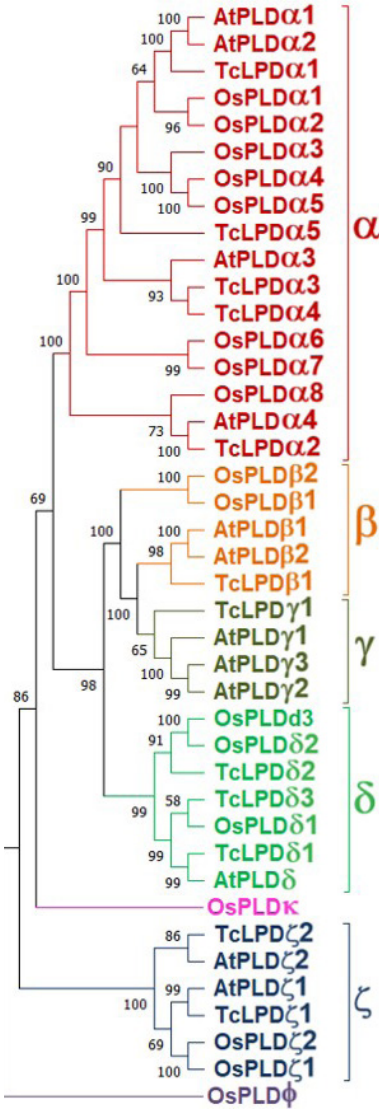

Figure 2. Phylogenetic relationship of the PLDs in cocoa, *Arabidopsis thaliana*, and rice. Phylogenetic tree showing the relationships of PLD enzymes from cocoa (Tc), *A. thaliana* (At), and rice (Os). Subfamilies (α, β, γ, δ, and ζ) are indicated on the branches. Bootstrap values are provided to demonstrate the reliability of the clustering

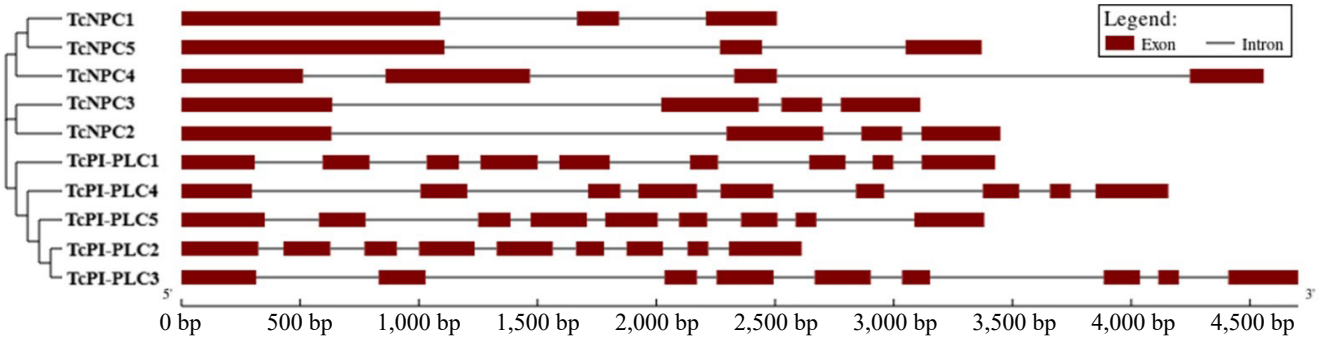

Figure 3. Gene structure of the *PLC*s in cocoa

in cocoa, likely driven by functional specialization to meet specific physiological and environmental demands. This structural variation provides a foundation for further studies on the regulatory mechanisms and functional roles of *PLC* genes in cocoa.

The gene structure analysis of *PLD* genes in cocoa reveals considerable variation in exon-intron organization among different members of the PLD family. As categorized in the phylogenetic tree (Figure 2), the *PLD* genes are categorized into subfamilies (α, β, γ, δ, and ζ), each displaying distinct structural patterns (Figure 4). Within the α subfamily, genes encoding PLD proteins in cocoa, such as *TcPLDα1, TcPLDα2, TcPLDα3, TcPLDα4,* and *TcPLDα5* exhibit diverse exon-intron configurations. Particularly, *TcPLDα1* has a relatively simple structure with fewer exons and shorter introns, whereas *TcPLDα4* and *TcPLDα5* show more complex arrangements, indicating functional diversification within this subfamily. The β subfamily, represented by *TcPLDβ1*, has a moderately complex structure characterized by multiple exons and longer introns, suggesting roles in more specialized lipid signaling pathways. The single member of the γ subfamily, *TcPLDγ1*, has a unique structural configuration, reflecting its functional divergence from other subfamilies. In the δ subfamily, genes such as *TcPLDδ1*, *TcPLDδ2*, and *TcPLDδ3* exhibit a conserved structural organization with relatively long introns that may be associated with alternative splicing and regulatory complexity. The ζ subfamily, including *TcPLDζ1* and *TcPLDζ2*, has the most complex gene structures, featuring a large number of exons and extensive introns.

3.5. Expression Dynamics of the *PLC* and *PLD* Families During Growth and Development in Cocoa

This study examined the expression patterns of the *PLC* and *PLD* genes across various stages of cocoa embryo development (Figure 5). Overall, the majority of the *PLC* and *PLD* genes were expressed at least one stage of embryo development, except for *TcPI-PLC1, TcPI-PLC2, TcPLDα2, TcPLDα3, TcPLDα4, TcPLDα5*. The expression levels of the identified *PLC* and *PLD* genes varied significantly across different developmental stages of the embryo. In the PLC gene family, three genes, including *TcNPC2, TcNPC3*, and *TcPI-PLC5*, were highly expressed in LT-SE and M-SE. Meanwhile, *TcNPC1, TcNPC4, TcNPC5, TcPI-PLC3,* and *TcPI-PLC4* were expressed in all stages of embryo development. In the PLD gene family, *TcPLDα1* was noted to be highly expressed in all tissues during the development of the embryo. In contrast, *TcPLDβ1, TcPLDγ1, TcPLDδ3, TcPLDζ1,* and *TcPLDζ2* were expressed in only one stage of embryo development.

Next, a considerable number of *PLC* genes exhibited differential expression in response to *P. megakarya* treatment across various time points and cocoa varieties (Figure 6). In the Nanay genotype, *TcNPC3* and *TcNPC2* were upregulated in treated leaf tissues at 6 and 72 hai by 2.01-fold and 2.13-fold, respectively, whereas *TcPI-PLC5* was down-regulated (-2.15-fold) in leaves at 24 hai. In the Scavina genotype, *TcPI-PLC2* showed a decrease in expression (-4.69-fold) at 24 hai but was significantly upregulated (9.44-fold) in leaf tissues at 72 hai. In the case of *PLD* genes, six genes, including *TcPLDα2,*

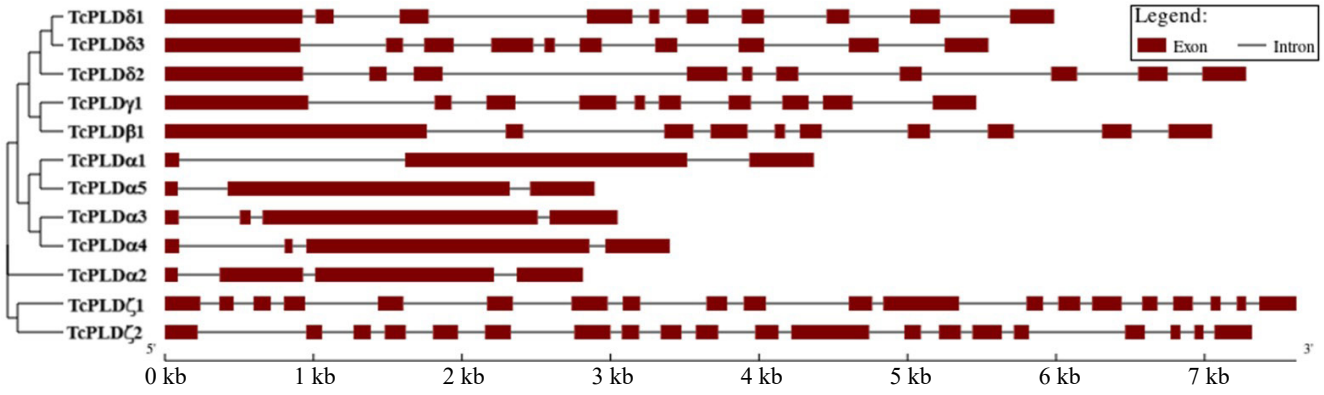

Figure 4. Gene structure of the *PLD*s in cocoa

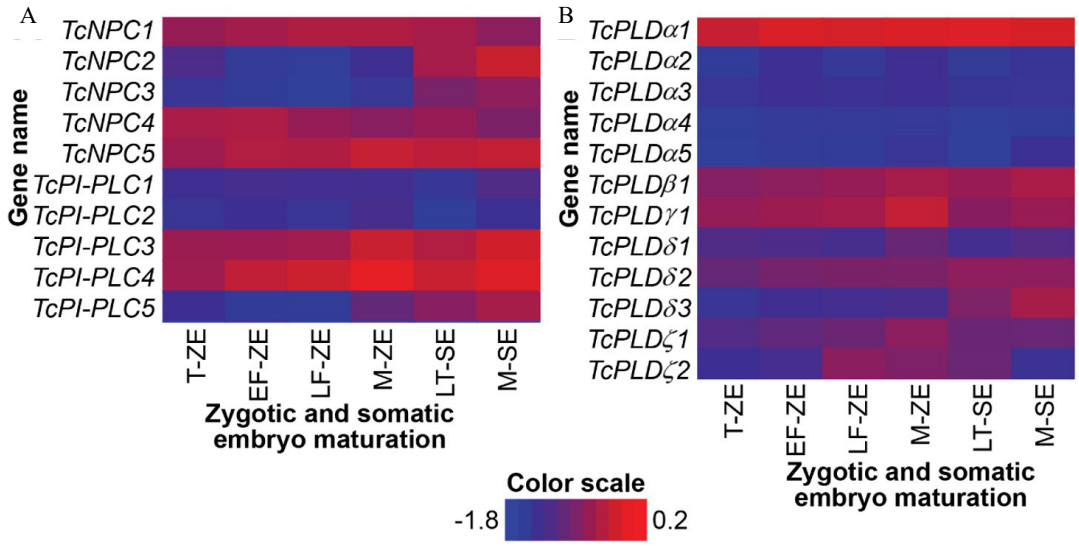

Figure 5. Tissue-specific expression of the *PLC*s and *PLD*s during zygotic and somatic embryo maturation. (A) Heatmap of *PLC* gene expression levels across different stages of embryo development, including torpedo (T-ZE), early-full (EF-ZE), late-full (LF-ZE), and mature zygotic embryos (M-ZE), as well as late torpedo (LT-SE) and mature somatic embryos (M-SE). (B) Heatmap of *PLD* gene expression levels during the same stages. Expression levels are normalized to the *Actin 11* gene and represented on a log2 scale. Colors indicate relative expression: red for high expression and blue for low expression

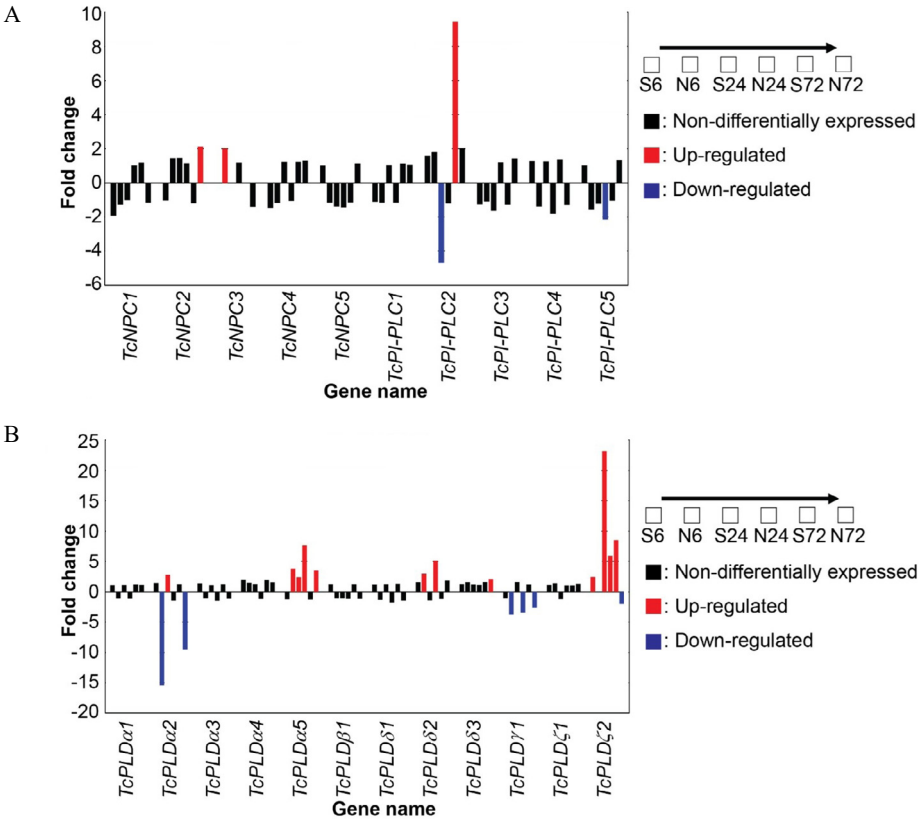

Figure 6. Expression profiles of the *PLC*s and *PLD*s in response to P. megakarya treatment. (A) Heatmap showing expression levels of *PLC* genes across two genotypes, Scavina (S) and Nanay (N), at 6 (S6/N6), 24 (S24/N24), and 72 (S72/N72) hours post-infection. (B) Heatmap showing expression levels of *PLD* genes under the same conditions. Colors indicated relative expression levels: red for upregulation and blue for downregulation

TcPLDα5, TcPLDδ2, TcPLDδ3, TcPLDγ1, and *TcPLDζ2*, were differentially expressed in response to treatments. Particularly, *TcPLDα2* and *TcPLDγ1* were reduced in leaves at least 2-time points of *P. megakarya* treatment. In contrast, *TcPLDα5* was induced in treated leaves at 6 (3.78-fold), 24 (7.63-fold), and 72 (3.51-fold) hai, while *TcPLDδ2* was induced in leaf tissues at 6 (3.01-fold) and 24 (5.11-fold) hai. *TcPLDδ3* was upregulated in leaves at 72 hai (2.07-fold), while *TcPLDζ2* was induced (5.92-fold) and reduced (-2.02-fold) in treated leaves at 24 and 72 hai, respectively. Additionally, *TcPLDα2* and *TcPLDα5* were upregulated in inoculated leaf tissues at 24 hai by 2.80-fold and 2.37-fold, respectively. Interestingly, *TcPLDζ2* was upregulated in leaf tissues at 6 (2.43-fold), 24 (23.16-fold), and 72 (8.51-fold) hai.

4. Discussion

4.1. Comparative Analysis of PLC and PLD Family Diversity Across Plant Species

The number of PLCs and PLDs varies significantly across plant species (Singh *et al.* 2013; Chen *et al.* 2016; Song *et al.* 2017; Zhu *et al.* 2020; Kanchan *et al.* 2021; Qin & Wang 2002; Sadat *et al.* 2022; Wei *et al.* 2022; Zhang *et al.* 2023; Guo *et al.* 2024), reflecting their evolutionary diversity and functional specialization (Ali *et al.* 2022; Fang *et al.* 2023). In *A. thaliana*, 9 PI-PLCs and 6 NPCs have been identified, along with 12 *PLD* genes classified into the α, β, γ, δ, and ζ subfamilies (Qin & Wang 2002). Similarly, in rice, the PLC family consists of 4 PI-PLCs and 5 NPCs, while the PLD family includes 17 members grouped into the α, β, δ, κ, ζ, and φ subfamilies (Singh *et al.* 2013). Maize has 5 PI-PLCs and 6 NPCs, alongside 13 *PLD* genes (Chen *et al.* 2016; Zhu *et al.* 2020). The tomato genome includes 14 *PLD* genes distributed across the α, β, δ, and ζ subtypes (Guo *et al.* 2024). In sorghum, 13 *PLD* genes are similarly categorized into various groups (Wei *et al.* 2022). The peanut genome exhibits extensive diversification, containing 46 *PLD* genes across subtypes such as α, β, γ, δ, ε, ζ, and η (Zhang *et al.* 2023). Furthermore, jute has 12 *PLD* genes in *C. capsularis* and 7 in *C. olitorius*, further emphasizing the variability in gene family expansion (Sadat *et al.* 2022). This diversity across species underscores the significance of the *PLC* and *PLD* gene families in lipid signaling and plant adaptation, likely contributing to physiological processes and abiotic stress resilience.

The PLC proteins in maize exhibited variations in length, Mw, and pI, which are critical for understanding their structure and potential functional roles (Zhu *et al.* 2020). The length of PLC proteins ranges between 259 and 606 residues, reflecting diversity among the members (Zhu *et al.* 2020). The Mw of these PLC proteins spans from 28.96 kDa to 68.23 kDa, indicating variability in protein size that might be associated with differences in enzymatic activity and substrate specificity (Zhu *et al.* 2020). The theoretical pI values of PLC proteins range from 5.74 to 9.09, suggesting that these proteins can exhibit varying charge properties under physiological conditions, which may influence their solubility, stability, and interactions with other biomolecules (Zhu *et al.* 2020). The protein characterization of PLC enzymes in orchids, including *Phalaenopsis equestris, Dendrobium catenatum*, and *Apostasia shenzhenica*, reveals conserved structural features across species (Kanchan *et al.* 2021). Particularly, the peptide lengths of PI-PLCs range from 588 to 604 residues, with an average Mw of approximately 67.8 to 68.1 kDa (Kanchan *et al.* 2021). These proteins are characterized by their pI, which ranges from 5.04 to 6.46, indicating a slightly acidic nature. The PI-PLC proteins also exhibit a high aliphatic index of around 77, suggesting structural stability and negative GRAVY values indicative of hydrophilicity (Kanchan *et al.* 2021). In contrast, NPC proteins in orchids have shorter peptide lengths, averaging between 481 and 531 residues, with Mw values of approximately 53.8 to 59.2 kDa (Kanchan *et al.* 2021). The pI values for PC-PLCs vary from 5.27 to 8.54, reflecting a broader range of charge properties (Kanchan *et al.* 2021). These proteins also show a hydrophilic nature, with negative GRAVY values and an aliphatic index averaging 74.18, which suggests moderate stability (Kanchan *et al.* 2021).

Previously, the phylogenetic relationships of the PLC family in maize were analyzed to understand their evolutionary divergence and classification within the PLC family (Zhu *et al.* 2020). A phylogenetic tree was constructed using full-length amino acid sequences of ZmPLCs and well-characterized homologs from other plant species, such as *A. thaliana* and rice, as reference sequences (Zhu *et al.* 2020). The results revealed that the ZmPLC family is divided into distinct subgroups corresponding to PI-PLCs and NPCs (Zhu *et al.* 2020). Next, the phylogenetic analysis of PLC enzymes in three species of orchids revealed that orchid PLCs are grouped into two main subfamilies, including PI-PLCs and NPCs (Kanchan *et al.* 2021). Within the PI-PLC subfamily,

orchid PLCs clustered closely with those of monocots like rice, while the NPC subfamily showed greater divergence, with orchid NPCs forming distinct clades that indicate potential species-specific adaptations (Kanchan *et al.* 2021). Taken together, the phylogenetic analysis of PLC proteins across various plant species reveals their evolutionary divergence into distinct subfamilies, including PI-PLCs and NPCs. These subfamilies are characterized by conserved catalytic domains, such as the X and Y domains in PI-PLCs, and exhibit both functional conservation and species-specific diversification. In model plants like *A. thaliana* and rice (*Oryza sativa*), PI-PLCs and NPCs are distributed across well-supported clades, reflecting their evolutionary relationships and lineage-specific adaptations. The clustering of orthologous PLCs among monocots and dicots indicates shared functional roles in fundamental cellular processes, such as lipid signaling and stress responses. At the same time, the expansion of certain subfamilies within specific species suggests adaptive evolution to unique environmental pressures. This phylogenetic framework highlights the conservation of PLC-mediated signaling pathways while providing insights into their functional specialization across plant species, paving the way for further comparative and functional studies.

Also, the gene structure of the *ZmPLC* family in maize was analyzed to elucidate their genomic organization and evolutionary characteristics (Zhu *et al.* 2020). The *ZmPLC* genes exhibit diverse structural features, including variations in the number and arrangement of exons and introns. The exon count among *ZmPLC* genes ranges from 6 to 12, reflecting significant differences in gene architecture within the family. Such variability in exon-intron organization is indicative of functional divergence and potential regulatory complexity among *ZmPLC* members (Zhu *et al.* 2020). The length of introns also varies considerably, suggesting differential splicing mechanisms that may generate multiple transcript variants, thereby increasing protein diversity. Conserved exon patterns were observed within subgroups of *ZmPLC*s, particularly among closely related genes, underscoring their shared evolutionary origins and functional similarities. This conserved exon-intron structure aligns with the maintenance of critical functional domains, such as the X and Y catalytic domains and EF-hand motifs, which are essential for enzymatic activity and calcium binding in phosphoinositide-specific PLCs (Zhu *et al.* 2020). Comparative analysis with *PLC* genes from other plant species reveals that the gene structures of *ZmPLC*s share similarities with those of monocots like rice (Singh *et al.* 2013). The gene structure analysis of *PLC* genes in orchids reveals significant variation in exon-intron organization, reflecting their evolutionary and functional diversity. The number of exons in orchid *PLC* genes ranges from 6 to 13, depending on the specific gene and subfamily. Among them, *PI-PLC*s generally exhibit more conserved exon-intron structures than NPCs, which show more significant variability (Zhu *et al.* 2020).

4.2. Tissue-specific and Stress-responsive Expression of *PLC* and *PLD* Genes Across Plant Species

Recently, the expression profiles of specific *ZmPLC* genes in maize highlight their diverse roles across tissues, developmental stages, and stress responses (Zhu *et al.* 2020). For instance, *ZmPI-PLC1* and *ZmPI-PLC2* are predominantly expressed in roots, suggesting their involvement in nutrient uptake and root growth processes (Zhu *et al.* 2020). Conversely, *ZmPI-PLC3* and *ZmNPC1* are highly expressed in leaves, indicating potential roles in photosynthetic regulation and the response to environmental conditions, such as light intensity and drought (Zhu *et al.* 2020). Under abiotic stress conditions, genes like *ZmPI-PLC4* and *ZmNPC3* are significantly upregulated in response to drought and salinity, emphasizing their roles in generating secondary messengers, such as inositol phosphates, to mediate stress signaling pathways (Zhu *et al.* 2020). Similarly, *ZmNPC2* shows enhanced expression during heat stress, suggesting its involvement in stabilizing membrane integrity and modulating cellular responses to temperature fluctuations (Zhu *et al.* 2020). In the context of biotic stress, genes such as *ZmPI-PLC5* and *ZmNPC5* are induced under pathogen attack, indicating their participation in plant defense mechanisms by triggering signal cascades that activate immune responses (Zhu *et al.* 2020). Developmentally, *ZmPI-PLC1* exhibits peak expression during seed germination, aligning with its potential role in mobilizing stored energy for early growth (Zhu *et al.* 2020).

In contrast, *ZmNPC4* shows elevated expression during grain filling, suggesting its involvement in lipid metabolism and storage processes critical for seed development (Zhu *et al.* 2020). The tissue-specific and stress-responsive expression patterns of these *ZmPLC* genes demonstrate their functional specialization, providing crucial insights into their roles in maize growth, development, and adaptation to environmental challenges (Zhu *et al.* 2020). Next, the spatio-temporal expression analysis of specific *PLC* genes in orchids demonstrated

distinct and dynamic expression patterns across tissues and developmental stages (Kanchan *et al.* 2021). For instance, *PePI-PLC1* and *DcPI-PLC1* exhibited high expression in floral tissues, suggesting their critical roles in flower development (Kanchan *et al.* 2021). Similarly, *PeNPC2* and *DcNPC3* showed predominant expression in roots, indicating their involvement in nutrient acquisition and root growth processes (Kanchan *et al.* 2021). Temporal expression analysis highlights that *PePI-PLC2* and *DcPI-PLC3* were strongly expressed during early seedling stages, consistent with their roles in cell signaling pathways that regulate early growth and cellular differentiation (Kanchan *et al.* 2021). Additionally, *PeNPC1* showed peak expression during flowering, indicating its likely involvement in reproductive processes. During fruit maturation, *DcNPC4* was upregulated, which may be related to lipid metabolism and signaling involved in seed and fruit development (Kanchan *et al.* 2021). Under environmental stress conditions, *PePI-PLC3* and *AsNPC2* are significantly induced in response to drought (Kanchan *et al.* 2021). Similarly, *PeNPC5* was upregulated under salinity stress (Kanchan *et al.* 2021).

In conclusion, this study, the comprehensive identification, characterization, and expression analysis of *PLC* and *PLD* gene families in coca provided critical insights into their evolutionary, structural, and functional roles. A total of 10 TcPLCs and 12 TcPLDs was identified and systematically annotated. The physicochemical properties of the identified proteins revealed variations in molecular weight, isoelectric points, and stability indices. Phylogenetic analysis classified the PLC and PLD enzymes into well-defined subfamilies, with close evolutionary relationships observed between cocoa and other plant species. The structural analysis further highlighted significant variation in exon-intron organization across *PLC* and *PLD* genes. The expression patterns of *PLC* and *PLD* genes during cocoa embryo development revealed their dynamic regulation across developmental stages. Several genes, such as *TcNPC2, TcNPC3*, and *TcPI-PLC5*, exhibited high expression in specific stages, while others, including *TcPLDα1*, were consistently expressed across all stages. In response to *P. megakarya* infection, a considerable number of *PLC* and *PLD* genes displayed differential expression across time points and cocoa varieties, with notable upregulation or downregulation of key genes, such as *TcPI-PLC2, TcPLDα5, TcPLDδ3*, and *TcPLDζ2*. Overall, these findings provide a comprehensive framework for understanding the roles of *PLC* and *PLD* genes in cocoa growth, development, and stress responses. The results lay the foundation for future functional studies to enhance cocoa resilience to biotic and abiotic stresses, contributing to the improvement of this economically important crop.

Acknowledgements

This work was funded by the fundamental research program of Hung Vuong University (Vietnam) under the project grant No. 06/2024/KHCN (HV06.2024).

References

Adeniyi, D., 2019. Diversity of cacao pathogens and impact on yield and global production, in: Aikpokpodion, P.O. (Eds.), *Theobroma cacao-Deploying Science for Sustainability of Global Cocoa Economy.* IntechOpen, Rijeka, pp. 1-20. DOI: 10.5772/intechopen.81993

Ali, U., Lu, S., Fadlalla, T., Iqbal, S., Yue, H., Yang, B., Hong, Y., Wang, X., Guo, L., 2022. The functions of phospholipases and their hydrolysis products in plant growth, development, and stress responses. *Prog Lipid Res*. 86, 101158. DOI:10.1016/j.plipres.2022.101158

Apostolakos, P., Panteris, E., Galatis, B., 2008. The involvement of phospholipases C and D in the asymmetric division of subsidiary cell mother cells of *Zea mays*. *Cell Motil Cytoskeleton*. 65, 863-875. DOI:10.1002/cm.20308

Argout, X., Salse, J., Aury, J.M., Guiltinan, M.J., Droc, G., Gouzy, J., Allegre, M., Chaparro, C., Legavre, T., Maximova, S.N., Abrouk, M., Murat, F., Fouet, O., Poulain, J., Ruiz, M., Roguet, Y., Rodier-Goud, M., Barbosa-Neto, J.F., Sabot, F., Kudrna, D., Ammiraju, J.S., Schuster, S.C., Carlson, J.E., Sallet, E., Schiex, T., Dievart, A., Kramer, M., Gelley, L., Shi, Z., Berard, A., Viot, C., Boccara, M., Risterucci, A.M., Guignon, V., Sabau, X., Axtell, M.J., Ma, Z., Zhang, Y., Brown, S., Bourge, M., Golser, W., Song, X., Clement, D., Rivallan, R., Tahi, M., Akaza, J.M., Pitollat, B., Gramacho, K., D'Hont, A., Brunel, D., Infante, D., Kebe, I., Costet, P., Wing, R., McCombie, W.R., Guiderdoni, E., Quetier, F., Panaud, O., Wincker, P., Bocs, S., Lanaud, C., 2011. The genome of *Theobroma cacao*. *Nat Genet*. 43, 101-108. DOI:10.1038/ng.736

Bargmann, B.O., Munnik, T., 2006. The role of phospholipase D in plant stress responses. *Curr Opin Plant Biol*. 9, 515-522. DOI:10.1016/j.pbi.2006.07.011

Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Holko, M., Yefanov, A., Lee, H., Zhang, N., Robertson, C.L., Serova, N., Davis, S., Soboleva, A., 2013. NCBI GEO: archive for functional genomics data sets - update. *Nucleic Acids Res*. 41, 991-995. DOI:10.1093/nar/gks1193

Bill, C.A., Vines, C.M., 2020. Phospholipase C. *Adv Exp Med Biol*. 1131, 215-242. DOI:10.1007/978-3-030-12457-1_9

Chen, L., Cao, B., Han, N., Tao, Y., Zhou, S.F., Li, W.C., Fu, F.L., 2016. Phospholipase D family and its expression in response to abiotic stress in maize. *Plant Growth Regulation*. 81, 197-207. DOI:10.1007/s10725-016-0197-4

Chu, H.D., Nguyen, K.H., Watanabe, Y., Le, D.T., Pham, T.L.T., Mochida, K., Tran, L.P., 2018. Identification, structural characterization and gene expression analysis of members of the Nuclear Factor-Y family in chickpea (*Cicer arietinum* L.) under dehydration and abscisic acid treatments. *Int J Mol Sci*. 19, 3290. DOI:10.3390/ijms19113290

Chu, H.D., Yen, T.T.H., Pham, C.T., Quynh, L.T.N., Huyen, T.T.T., Trung, N.Q., Gioi, D.H., Tien, T.V., Cao, P.B., 2024a. Identification of two enzymes for trehalose synthesis and their potential function in growth and development in peanut (*Arachis hypogaea*). *Journal of Tropical Life Science*. 14, 83-94. DOI:10.11594/jtls.14.01.10

Chu, N.T.B., Le, T.M., Chu, H.D., Tran, H.T.T., Tran, L.T.M., La, H.V., Vu, Q.T.X., Phung, H.H., Cao, P.B., 2024b. Mining GATA transcription factor encoding genes in the cocoa tree (*Theobroma cacao* L.) suggests their potential roles in embryo development and biotic stress response. *Journal of Tropical Biodiversity and Biotechnology*. 9, 88015. DOI:10.22146/jtbb.88015

Deepika, D., Singh, A., 2022. Plant phospholipase D: novel structure, regulatory mechanism, and multifaceted functions with biotechnological application. *Crit Rev Biotechnol*. 42, 106-124. DOI:10.1080/07388551.2021.1924113

Fang, Y., Jiang, J., Ding, H., Li, X., Xie, X., 2023. Phospholipase C: diverse functions in plant biotic stress resistance and fungal pathogenicity. *Mol Plant Pathol*. 24, 1192-1202. DOI:10.1111/mpp.13343

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D., Bairoch, A., 2003. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res*. 31, 3784-3788. DOI:10.1093/nar/gkg563

Gateau-Rey, L., Tanner, E.V.J., Rapidel, B., Marelli, J.P., Royaert, S., 2018. Climate change could threaten cocoa production: effects of 2015-16 El Nino-related drought on cocoa agroforests in Bahia, Brazil. *PLoS One.* 13, e0200454. DOI:10.1371/journal.pone.0200454

Guo, X., Zhu, W., Wang, F., Wang, H., 2024. Genome-wide investigation of the PLD gene family in tomato: identification, analysis, and expression. *Genes*. 15, 326. DOI:10.3390/genes15030326

Hu, B., Jin, J., Guo, A.Y., Zhang, H., Luo, J., Gao, G., 2015. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics*. 31, 1296-1297. DOI:10.1093/bioinformatics/btu817

Kanchan, M., Ramkumar, T.R., Himani, Sembi, J.K., 2021. Genome-wide characterization and expression profiling of the Phospholipase C (PLC) gene family in three orchids of economic importance. *J Genet Eng Biotechnol*. 19, 124. DOI:10.1186/s43141-021-00217-z

Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*. 35, 1547-1549. DOI:10.1093/molbev/msy096

La, H.V., Chu, H.D., Ha, Q.T., Tran, T.T.H., Van Tong, H., Van Tran, T., Le, Q.T.N., Bui, H.T.T., Cao, P.B., 2022a. SWEET gene family in sugar beet (*Beta vulgaris*): genome-wide survey, phylogeny and expression analysis. *PJBS.* 25, 387-395.

La, H.V., Chu, H.D., Tran, C.D., Nguyen, K.H., Le, Q.T.N., Hoang, C.M., Cao, B.P., Pham, A.T.C., Nguyen, B.D., Nguyen, T.Q., Van Nguyen, L., Ha, C.V., Le, H.T., Le, H.H., Le, T.D., Tran, L.P., 2022b. Insights into the gene and protein structures of the CaSWEET family members in chickpea (*Cicer arietinum*), and their gene expression patterns in different organs under various stress and abscisic acid treatments. *Gene*. 819, 146210. DOI:10.1016/j.gene.2022.146210

Lanaud, C., Vignes, H., Utge, J., Valette, G., Rhone, B., Garcia Caputi, M., Angarita Nieto, N.S., Fouet, O., Gaikwad, N., Zarrillo, S., Powis, T.G., Cyphers, A., Valdez, F., Olivera Nunez, S.Q., Speller, C., Blake, M., Valdez, F.J., Raymond, S., Rowe, S.M., Duke, G.S., Romano, F.E., Loor Solorzano, R.G., Argout, X., 2024. A revisited history of cacao domestication in pre-Columbian times revealed by archaeogenomic approaches. *Sci Rep*. 14, 2972. DOI:10.1038/s41598-024-53010-6

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and clustal X version 2.0. *Bioinformatics*. 23, 2947-2948. DOI:10.1093/bioinformatics/btm404

Latif, R., 2013. Health benefits of cocoa. *Curr Opin Clin Nutr Metab Care*. 16, 669-674. DOI:10.1097/MCO.0b013e328365a235

Maximova, S.N., Florez, S., Shen, X., Niemenak, N., Zhang, Y., Curtis, W., Guiltinan, M.J., 2014. Genome-wide analysis reveals divergent patterns of gene expression during zygotic and somatic embryo maturation of *Theobroma cacao* L., the chocolate tree. *BMC Plant Biol*. 14, 185. DOI:10.1186/1471-2229-14-185

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L.L., Tosatto, S.C.E., Paladin, L., Raj, S., Richardson, L.J., Finn, R.D., Bateman, A., 2021. Pfam: the protein families database in 2021. *Nucleic Acids Res*. 49, 412-419. DOI:10.1093/nar/gkaa913

Pinheiro, T.T., Litholdo, C.G., Jr., Sereno, M.L., Leal, G.A., Jr., Albuquerque, P.S., Figueira, A., 2011. Establishing references for gene expression analyses by RT-qPCR in *Theobroma cacao* tissues. *Genet Mol Res*. 10, 3291-3305. DOI:10.4238/2011.November.17.4

Pokou, D.N., Fister, A.S., Winters, N., Tahi, M., Klotioloma, C., Sebastian, A., Marden, J.H., Maximova, S.N., Guiltinan, M.J., 2019. Resistant and susceptible cacao genotypes exhibit defense gene polymorphism and unique early responses to *Phytophthora megakarya* inoculation. *Plant Mol Biol*. 99, 499-516. DOI:10.1007/s11103-019-00832-y

Qin, C., Wang, X., 2002. The Arabidopsis phospholipase D family. Characterization of a calcium-independent and phosphatidylcholine-selective PLD zeta 1 with distinct regulatory domains. *Plant Physiol*. 128, 1057-1068. DOI:10.1104/pp.010928

Romero Navarro, J.A., Phillips-Mora, W., Arciniegas-Leal, A., Mata-Quiros, A., Haiminen, N., Mustiga, G., Livingstone Iii, D., van Bakel, H., Kuhn, D.N., Parida, L., Kasarskis, A., Motamayor, J.C., 2017. Application of genome wide association and genomic prediction for improvement of cacao productivity and resistance to black and frosty pod diseases. *Front Plant Sci.* 8, 1905. DOI:10.3389/fpls.2017.01905

Rupwate, S.D., Rajasekharan, R., 2012. Plant phosphoinositide-specific phospholipase C: an insight. *Plant Signal Behav*. 7, 1281-1283. DOI:10.4161/psb.21436

Sadat, M.A., Ullah, M.W., Hossain, M.S., Ahmed, B., Bashar, K.K., 2022. Genome-wide in silico identification of phospholipase D (PLD) gene family from *Corchorus capsularis* and *Corchorus olitorius*: reveals their responses to plant stress. *J Genet Eng Biotechnol*. 20, 28. DOI:10.1186/s43141-022-00311-w

Schroth, G., Laderach, P., Martinez-Valle, A.I., Bunn, C., Jassogne, L., 2016. Vulnerability to climate change of cocoa in West Africa: patterns, opportunities and limits to adaptation. *Sci Total Environ*. 556, 231-241. DOI:10.1016/j.scitotenv.2016.03.024

Singh, A., Kanwar, P., Pandey, A., Tyagi, A.K., Sopory, S.K., Kapoor, S., Pandey, G.K., 2013. Comprehensive genomic analysis and expression profiling of phospholipase C gene family during abiotic stresses and development in rice. *PLoS One*. 8, e62494. DOI:10.1371/journal.pone.0062494

Song, J., Zhou, Y., Zhang, J., Zhang, K., 2017. Structural, expression and evolutionary analysis of the non-specific phospholipase C gene family in *Gossypium hirsutum. BMC Genomics*. 18, 979. DOI:10.1186/s12864-017-4370-6

Thompson, J.D., Gibson, T.J., Higgins, D.G., 2002. Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinformatics*. Chapter 2, Unit 2 3. DOI:10.1002/0471250953.bi0203s00

Wei, J., Shao, W., Liu, X., He, L., Zhao, C., Yu, G., Xu, J., 2022. Genome-wide identification and expression analysis of phospholipase D gene in leaves of sorghum in response to abiotic stresses. *Physiol Mol Biol Plants*. 28, 1261-1276. DOI:10.1007/s12298-022-01200-9

Wickramasuriya, A.M., Dunwell, J.M., 2018. Cacao biotechnology: current status and future prospects. *Plant Biotechnol J*. 16, 4-17. DOI:10.1111/pbi.12848

Yang, F., Abdelnabby, H., Xiao, Y., 2015. The role of a phospholipase (PLD) in virulence of Purpureocillium lilacinum (*Paecilomyces lilacinum*). *Microb Pathog*. 85, 11-20. DOI:10.1016/j.micpath.2015.05.008

Zhang, H., Yu, Y., Wang, S., Yang, J., Ai, X., Zhang, N., Zhao, X., Liu, X., Zhong, C., Yu, H., 2023. Genome-wide characterization of phospholipase D family genes in allotetraploid peanut and its diploid progenitors revealed their crucial roles in growth and abiotic stress responses. *Front Plant Sci*. 14, 1102200. DOI:10.3389/fpls.2023.1102200

Zhu, J., Zhou, Y., Li, J., Li, H., 2020. Genome-wide investigation of the phospholipase C gene family in *Zea mays. Front Genet*. 11, 611414. DOI:10.3389/fgene.2020.611414