



Papain-Like Cysteine Protease Gene Family in Fig (*Ficus carica* L.): Genome-Wide Analysis and Expression Patterns

Yanlei Zhai¹, Yuanyuan Cui¹, Miaoyu Song¹, Alexander Vainstein², Shangwu Chen³ and Huiqin Ma^{1*}

¹ College of Horticulture, China Agricultural University, Beijing, China, ² The Robert H. Smith Faculty of Agriculture, Food and Environment, The Hebrew University of Jerusalem, Rehovot, Israel, ³ College of Food Science and Nutritional Engineering, China Agricultural University, Beijing, China

OPEN ACCESS

Edited by:

Peter Ulvskov,
University of Copenhagen, Denmark

Reviewed by:

Birger Lindberg Møller,
University of Copenhagen, Denmark

Manuel Martinez,
Polytechnic University of Madrid,
Spain

Marko Novinec,
University of Ljubljana, Slovenia

*Correspondence:

Huiqin Ma
hqma@cau.edu.cn

Specialty section:

This article was submitted to
Plant Physiology,
a section of the journal
Frontiers in Plant Science

Received: 17 March 2021

Accepted: 03 May 2021

Published: 28 May 2021

Citation:

Zhai Y, Cui Y, Song M, Vainstein A,
Chen S and Ma H (2021) Papain-Like
Cysteine Protease Gene Family in Fig
(*Ficus carica* L.): Genome-Wide
Analysis and Expression Patterns.
Front. Plant Sci. 12:681801.
doi: 10.3389/fpls.2021.681801

The papain-like cysteine proteases (PLCPs) are the most abundant family of cysteine proteases in plants, with essential roles in biotic/abiotic stress responses, growth and senescence. Papain, bromelain and ficin are widely used in food, medicine and other industries. In this study, 31 PLCP genes (*FcPCLPs*) were identified in the fig (*Ficus carica* L.) genome by HMM search and manual screening, and assigned to one of nine subfamilies based on gene structure and conserved motifs. SAG12 and RD21 were the largest subfamilies with 10 and 7 members, respectively. The *FcPCLPs* ranged from 1,128 to 5,075 bp in length, containing 1–10 introns, and the coding sequence ranged from 624 to 1,518 bp, encoding 207–505 amino acids. Subcellular localization analysis indicated that 24, 2, and 5 PLCP proteins were targeted to the lysosome/vacuole, cytoplasm and extracellular matrix, respectively. Promoter (2,000 bp upstream) analysis of *FcPCLPs* revealed a high number of plant hormone and low temperature response elements. RNA-seq revealed differential expression of 17 *FcPCLPs* in the inflorescence and receptacle, and RD21 subfamily members were the major *PLCPs* expressed in the fruit; 16 and 5 *FcPCLPs* responded significantly to ethylene and light, respectively. Proteome analyses revealed 18 and 5 *PLCPs* in the fruit cell soluble proteome and fruit latex, respectively. Ficins were the major *PLCP* in fig fruit, with decreased abundance in inflorescences, but increased abundance in receptacles of commercial-ripe fruit. *FcRD21B/C* and *FcALP1* were aligned as the genes encoding the main ficin isoforms. Our study provides valuable multi-omics information on the *FcPCLP* family and lays the foundation for further functional studies.

Keywords: papain-like cysteine protease, gene architecture, transcriptome, proteome, *Ficus carica* L.

INTRODUCTION

Cysteine proteases, which contain a cysteine residue at their active catalytic site, catalyze the hydrolysis of peptides and proteins. Cysteine proteases can be divided into 11 clans with different evolutionary routes; different families in the same clan are diverse in sequence and structure (Bah et al., 2006; Rawlings et al., 2016). C1A is the largest family of the CA clan of papain-like

cysteine proteases (PLCPs) (Rawlings et al., 2010). Members of the PLCP family in plants, such as papain, chymopapain, caricain (*Carica papaya*), bromelain (*Ananas comosus*) and ficin (*Ficus carica*), have broad substrate specificity and strong proteolytic activity, and the enzymes are of high commercial value in cheese making (Mazri et al., 2018), meat tenderization (Sullivan and Calkins, 2010; Bekhit et al., 2014), beer stabilization, biscuit baking and leather softening, as versatile biocatalysts (Huang et al., 2008) and in making digestive drugs (Zhalehjo and Mostafaie, 2012).

In plants, PLCPs act as an immunity hub, playing critical roles in plant–pathogen/pest interactions and abiotic stress responses (Shindo and Van Der Hoorn, 2008; Misas-Villamil et al., 2016). PLCPs are first synthesized as inactive precursors with a signal peptide, an N-terminal self-inhibiting predomain, and a mature catalytic domain (Richau et al., 2012; Wang et al., 2014). The mature PLCPs are monomer proteases consisting of an α -helix and β -sheet domain of similar size, forming an active cleft that specifically binds with the substrate. The active catalytic site—the highly conserved catalytic triad Cys–His–Asn—is located at the cleft, which is the conserved characteristic of PLCPs (Polaina and MacCabe, 2007). Most PLCPs have small molecular masses, ranging from 20 to 35 kDa, and a few are 50–75 kDa. The optimum pH for catalytic activity is 5.0–8.0 (Dubey et al., 2007). In plants, PLCPs are divided into nine subfamilies according to the propeptide domain and characteristic motifs: RD21A-like, CEP1-like, XCP2-like, XBCP3-like, THI1-like, SAG12-like, RD19A-like, ALP-like, and CTB-like (Richau et al., 2012).

Fig (*Ficus carica* L.) is a latex-producing fruit crop; the milk-like latex flows out when fig tissue is wounded (Raskovic et al., 2016). The latex participates in defense against fungi and insects (Mnif et al., 2015) and has historically been used to treat skin diseases. On the other hand, the proteases in latex damage the skin of fig pickers and workers in the orchard, and in commercial-ripe fig fruit, the latex needs to be drained after harvesting (Flaishman et al., 2008).

PLCPs—dominated by ficin isoforms—are the major protein component of fig latex (Zare et al., 2013; Raskovic et al., 2014). Ficin (EC 3.4.22.3), also known as ficain, is widely found in *Ficus* species. Our previous study revealed that multiple cysteine proteases, i.e., ficins A, B, C, D, and cysteine protease RD21A, make up the large proportion of the fig fruit's soluble proteome in both commercial-ripe and tree-ripe fruit, but transcripts of ficin isoforms were not identified by RNA-seq (Cui et al., 2019).

Publication of the fig (*F. carica*) genome (Usai et al., 2020) has provided the necessary information for bioinformatics analyses of the FcPLCP family. In the present study, gene structures, conserved motifs, phylogenetic relationships, and promoter *cis*-elements of FcPLCPs were analyzed, and gene-transcription patterns and protein abundance in fig fruit were revealed. This combined genomic, transcriptomic and proteomic study provides a matrix of information on the PLCP family in fig, laying a sound foundation for the identification of important PCLPs for further studies of biological function.

MATERIALS AND METHODS

Identification of PLCP Genes From the Genome of *Ficus carica*

Genomic data of *F. carica* and *Morus notabilis* were downloaded from NCBI¹, genomic data of *Arabidopsis thaliana* were downloaded from the TAIR database², data of *Ficus hispida* and *Ficus microcarpa* were downloaded from the Genome Sequence Archive (GSA) and Genome Warehouse (GWH) database³. The gene and coding sequences (CDSs) were extracted from contig level sequences (BioProject: PRJNA565858⁴) using TBtools (Chen et al., 2020), according to gene location information in Usai et al. (2020). The protein sequences were translated based on the CDSs.

The Hidden Markov Model (HMM) file corresponding to the peptidase C1 domain (PF00112) was downloaded from the Pfam database (El-Gebali et al., 2019). HMMER was used to search for the PLCPs from *F. carica*, *A. thaliana*, *M. notabilis*, *F. hispida*, and *F. microcarpa* genome databases with default parameters. All candidate proteins were confirmed by Pfam and the conserved domain database (Marchler-Bauer and Bryant, 2004; NCBI CDD⁵). Redundant sequences, sequences with E-values less than 1E-20, and sequences with no peptidase C1 domain were eliminated.

Multiple Sequence Alignment, and Phylogenetic and Sequence Feature Analyses

The PLCP sequences of *F. carica*, *A. thaliana*, *M. notabilis*, *F. hispida*, and *F. microcarpa* were subjected to multiple sequence alignment using ClustalW with default parameters. An unrooted phylogenetic tree based on this alignment was constructed using the neighbor-joining method by MEGA X, with the following parameters: Poisson model, pairwise deletion, 1,000 bootstrap replications. PLCPs were named according to their homology with *Arabidopsis*.

Sequence length, molecular weight, isoelectric point (pI), and the grand average of hydropathicity (GRAVY) index of PLCP proteins were computed by ProtParam⁶ (Gasteiger et al., 2005). The signal peptide and subcellular location were predicted by SignalP⁷ (Armenteros et al., 2019) and DeepLoc⁸ (Armenteros et al., 2017).

Exon–intron positions were obtained by genome annotation. The conserved motifs of PLCPs of *F. carica* were computed by the MEME program (Bailey et al., 2009), with the following parameters: classic mode; site distribution, zero or one occurrence per sequence; number of motifs, 20; width of motifs,

¹<https://www.ncbi.nlm.nih.gov/>

²<https://www.arabidopsis.org/>

³<https://bigd.big.ac.cn/gsa/>

⁴<https://www.ncbi.nlm.nih.gov/Traces/wgs/VYVB01?display=history%EF%BC%89>

⁵<https://www.ncbi.nlm.nih.gov/cdd/>

⁶<http://web.expasy.org/protparam/>

⁷<https://services.healthtech.dtu.dk/service.php?SignalP-5.0>

⁸<https://services.healthtech.dtu.dk/service.php?DeepLoc-1.0>

between 6 and 50 residues. PROSITE (Sigrist et al., 2012) and NCBI CDD were used for motif analysis. The diagrams were drawn with TBtools (Chen et al., 2020).

Chromosomal Location, Gene Duplication, and Promoter Analysis

Papain-like cysteine proteases genes were mapped to *F. carica* chromosomes using TBtools (Chen et al., 2020). The Multiple Collinearity Scan toolkit (MCScanX) was used to conduct the syntenic analysis among *F. carica*, *F. hispida*, and *F. microcarpa* (Wang et al., 2012).

An all-against-all BLASTP alignment was run to reveal potential gene duplication. The criteria for duplicated pairs were: (a) length of the aligned sequence covers >75% of the longer gene, and (b) >75% similarity of aligned regions (Vatansever et al., 2016). Corresponding coding regions were aligned using ClustalW. The number of non-synonymous substitutions per non-synonymous site (K_a) and the number of synonymous substitutions per synonymous site (K_s) were calculated by KaKs_Calculator 2.0 (Wang et al., 2010). The gene duplications were dated using the formula $T = K_s/2r$; r , which is the rate of divergence for nuclear genes, was taken to be 7×10^{-9} synonymous substitutions per site per year according to a previous report in *A. thaliana* (Ossowski et al., 2010).

Promoter analysis was conducted using Plant CARE (Lescot et al., 2002), based on the 2000-bp sequence upstream of the gene.

Plant Materials

The common fig cultivar Zibao, planted at the Shangzhuang Experimental Station of China Agricultural University (Peking, China), served as the plant material. Based on the three fig development phases (Flaishman et al., 2008), we subdivided the fig development process into six stages: stages 1 and 2 belonging to phase I, a rapid growth stage; stages 3 and 4 belonging to phase II, during which fruit size and hardness remain almost unchanged; stages 5 and 6 belonging to phase III, the mature stage, where stage 5 corresponds to commercial ripeness. Inflorescences and receptacles at each stage were separated, marked as F1–F6 and R1–R6, respectively, and stored at -80°C for further RNA-seq and proteomic analyses.

Fig fruit latex was collected by cutting the stage 5 fruit peel with a scalpel; the latex that flowed out was collected into centrifuge tubes and frozen in liquid nitrogen, then stored at -80°C for protein identification.

RNA-Seq and Quantitative Real-Time PCR (qRT-PCR) Validation

RNA was isolated from samples using the modified CTAB method (Chai et al., 2014). Library construction and RNA-seq methods were as described previously (Wang et al., 2017). The gene-expression profiles during fig development were analyzed based on Illumina RNA sequencing data and annotated against the published fig genome sequence. The transcriptome data of fig fruit at the six development stages were deposited in NCBI (accession number: PRJNA723733). Transcript abundance of fig *PLCPs* was calculated as fragments per kilobase of exon model

per million mapped reads (FPKM) and is displayed in heat maps. Significant gene expression was defined by $p\text{-adjust} < 0.05$ and $|\log_2(\text{fold change})| \geq 1$.

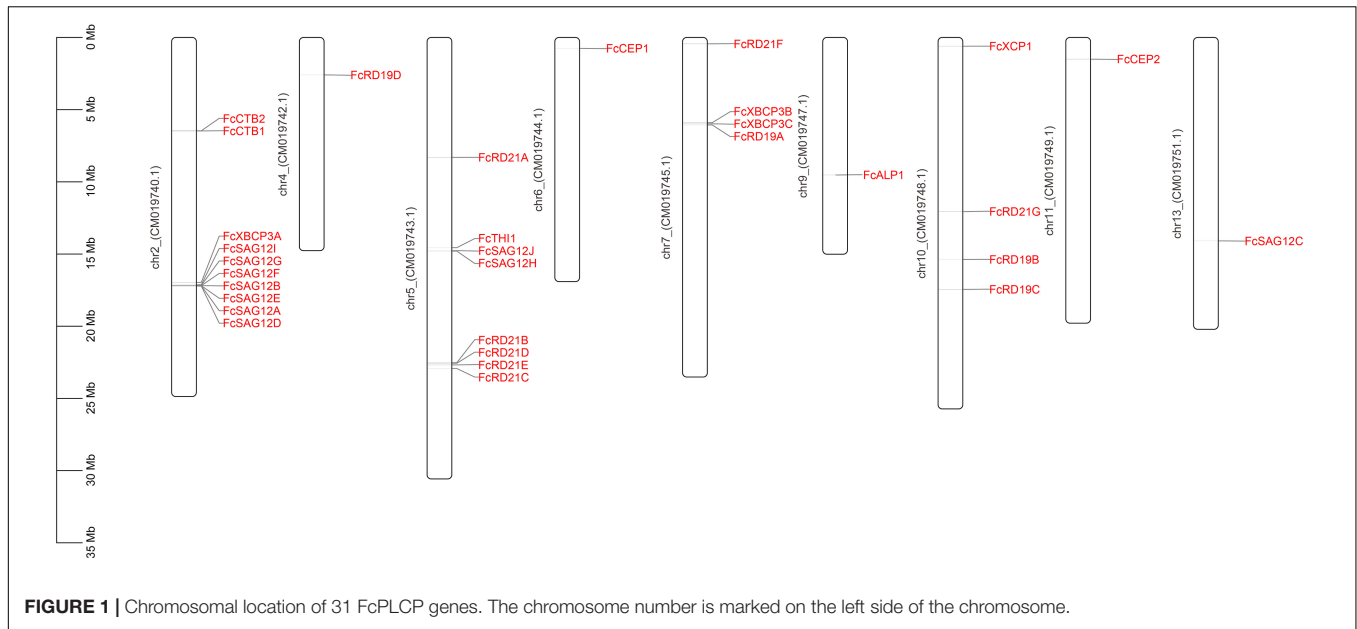
Transcriptome data of fig fruit after ethylene application (Cui et al., 2020) and under light deprivation (Wang et al., 2019) were re-mined to explore the changes in *PLCP* gene expression under these treatments. Briefly, fig fruit in phase II were injected with 1 mL of 250 mg/L ethephon solution through the ostiole, RNA-seq was carried out on the inflorescence and receptacle at 2, 4, and 6 days after treatment, and the transcriptome data were stored at NCBI (SRA accession: PRJNA606407) (Cui et al., 2020). For the light-deprivation treatment, fig fruit in stage 2 were covered with opaque paper bags and the transcriptomes of the inflorescence and receptacle of light-deprived and control fruit were determined in commercial-ripe fruit. The complete dataset can be found in the NCBI SRA database (accession number PRJNA494945) (Wang et al., 2019).

The expression of eight *PLCP* genes was validated by qRT-PCR. PrimeScriptTM RT reagent kit (RR037Q, Takara, Dalian, China) was used to reverse transcribe the total RNA. The 18S gene was used for normalization. Primers were designed by Beacon Designer 7 software (Supplementary Table 1). The qPCR was performed with ChamQ Universal SYBR qPCR Master Mix (Q711-02, Vazyme, Nanjing, China). A 15- μL reaction mixture was added to each well. The PCR program was as follows: 95°C for 30 s, and 40 cycles of: 95°C for 10 s, 60°C for 30 s. The $2^{-\Delta\Delta\text{Ct}}$ method (Livak and Schmittgen, 2001) was used for relative quantification analysis with three replicates for each sample.

Quantitative Proteomic Analysis

Stage 1, 3, and 5 inflorescences and receptacles were used for proteomic analysis. Protein extraction and quantitative analysis were as described in Cui et al. (2019). Three biological replicates were performed for each sample. The digested peptides were labeled with TMT10plexTM Isobaric Label Reagent Set (Thermo Scientific) according to the manufacturer's instructions. A Q-Exactive mass spectrometer (Thermo Fisher Scientific, Waltham, MA, United States) was used to detect peptide signals. The MS scans were run as described in our previous publication (Cui et al., 2019). The MS results were input into PD software (Proteome Discoverer 1.4, Thermo) to screen the spectra. The selected peptides were identified using Mascot (version 2.3.01) and annotated according to the UniProt database. Then the peptides were quantified by PD software based on their annotation and spectrum. ANOVA was performed to evaluate the significance of the differences. Proteins with a p -value less than 0.05, and fold change ≥ 1.2 or ≤ 0.83 were considered differentially abundant proteins (DAPs). The mass spectrometry proteomics data of inflorescences and receptacles were deposited in the ProteomeXchange Consortium via the PRIDE (Perez-Riverol et al., 2019) partner repository with the dataset identifier PXD025170.

The tris-phenol method (Xie et al., 2009) was used to extract latex protein. A 2-D Quant-kit was used to quantify the protein, and 20 μg protein was applied to SDS-PAGE. The latex protein was excised from the gel and cut into 10 pieces, digested with trypsin enzyme diluted with NH_4HCO_3



solution, and the peptides obtained from the digestion were separated by multidimensional liquid chromatography (Dionex Ultimate 3000 nano-LC system), then detected and analyzed by tandem MS (Thermo Fisher Q-Exactive). Mascot (version 2.3.01), MaxQuant (version 1.5.2.8), Thermo Scientific Proteome Discoverer (version 1.3/1.4), and UniProt database were used for protein identification and quantification. The mass spectrometry proteomics data of latex were deposited in the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD025485.

RESULTS

PLCP Gene Identification and Phylogenetic Analysis

Thirty-one PLCP genes were identified in the fig genome by HMM search and manual screening, and mapped to nine chromosomes. Chromosomes 2, 5, 7, and 10 had 10, 8, 4, and 4 PLCPs, respectively, and chromosomes 4, 6, 9, 11, and 13 each had 1 PLCP (Figure 1).

To investigate the evolutionary relationship among PLCPs, we constructed a phylogenetic tree using the fig PLCPs together with 32 PLCPs from *A. thaliana*, 24 PLCPs from *M. notabilis*, 35 PLCPs from *F. hispida* and 34 PLCPs from *F. microcarpa* (Supplementary Table 2). The PLCPs were clustered into nine subfamilies (Figure 2). Based on the homology to *A. thaliana*, we named the nine subfamilies as in Richau et al. (2012), i.e., RD21 (responsive to desiccation 21), CEP (cysteine endopeptidase), XCP (xylem cysteine peptidase), XBCP3 (xylem bark cysteine peptidase 3), THI1 (Th1 immune response-associated cysteine protease), SAG12 (senescence-associated gene 12), RD19 (responsive to desiccation 19), ALP (aleurain-like protease), and CTB (cathepsin B-like). These subfamilies can

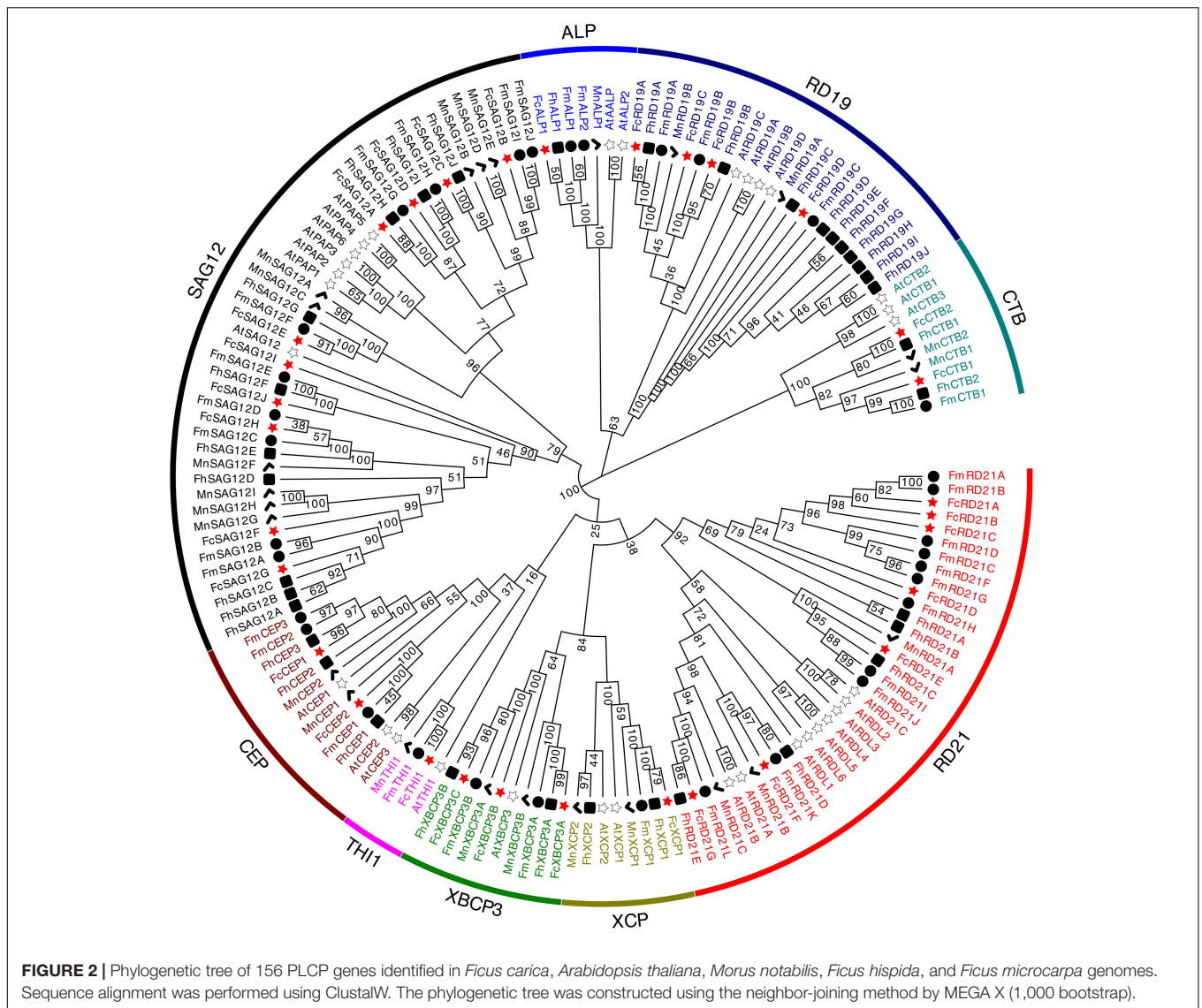
also be classified according to their closet animal counterparts: cathepsin B (CTB subfamily), cathepsin F (RD19 subfamily), cathepsin H (ALP subfamily), and cathepsin L (SAG12, THI1, CEP, XBCP3, XCP, and RD21 subfamilies) (Martinez and Diaz, 2008; Richau et al., 2012). SAG12 was the largest PLCP subfamily of fig and mulberry, with 10 and 9 members, respectively. RD21, with 7 members, was the second largest subfamily in fig.

K_a/K_s Analysis

Genome replication, segmental duplication, and tandem duplication are considered the main evolutionary forces. We identified six duplicated gene pairs in the fig PLCP family (Table 1). The K_a/K_s ratios of *FcRD21A/FcRD21B*, *FcRD21A/FcRD21C*, *FcRD21B/FcRD21C*, and *FcSAG12F/FcSAG12G* were <1 , suggesting negative selection on these four pairs. The K_a/K_s ratios of *FcRD19A/FcRD19B* and *FcRD19B/FcRD19C* were >1 , suggesting positive selection for these two pairs. In general, negative selection ($K_a/K_s < 1$) eliminates deleterious mutations, retaining the original function of the protein. Positive selection ($K_a/K_s > 1$) changes the protein, usually related to coevolution of immune system genes and parasites (Hurst, 2002). Duplication of these six PLCP gene pairs were calculated to have occurred between 1.76 and 59.98 million years ago.

Gene Structure and Sequence Features

Gene structure analysis showed that the PLCP genes of *F. carica*, *A. thaliana*, and *M. notabilis* differ in intron numbers, whereas the intron numbers within a subfamily were conserved (Supplementary Figure 1). Subfamilies RD21 and RD19 featured 2–5 introns and 2–4 introns, respectively. The other subfamilies had 1–10 introns. The full lengths of the *FcPLCPs* were from 1,128 to 5,075 bp with CDSs ranging from 624 to 1,518 bp.



Multiple sequence alignment revealed conserved motifs in the *FcPLCPs* (Table 2 and Supplementary Figure 2); most *FcPLCPs* in the same subfamily had similar motifs, indicating

TABLE 1 | K_a/K_s calculation and divergence time of the duplicated *FcPLCP* gene pairs.

Duplicated pair	K_a	K_s	K_a/K_s	Divergence time (MYA)
FcRD19A/FcRD19B	0.4155	0.1305	3.1840	9.32
FcRD19B/FcRD19C	1.0683	0.8398	1.2721	59.98
FcRD21A/FcRD21B	0.1051	0.1133	0.9277	8.09
FcRD21A/FcRD21C	0.1087	0.1908	0.5697	13.63
FcRD21B/FcRD21C	0.0840	0.1127	0.7459	8.05
FcSAG12F/FcSAG12G	0.0072	0.0246	0.2933	1.76

$K_a/K_s < 1$, negative selection; $K_a/K_s = 1$, neutral selection; $K_a/K_s > 1$, positive selection; MYA, million years ago.

that the protein structures were conserved (Figure 3). Motifs 5, 7, and 12 were identified as the inhibitor I29 domain (pfam08246), acting as propeptides that can inhibit protease activity. Most of the *FcPLCPs* had the inhibitor I29 domain, which contains the ERFNIN motif (Richau et al., 2012)—an interspersed amino acid motif in the N-terminal propeptide region with a highly conserved EX₃RX₂(V/I)FX₂NX₂IX₃N sequence. RD19 subfamily *FcPLCPs* carried a conserved ERFNAQ motif instead of ERFNIN. In the CTB subfamily, motif 17, characterized as a propeptide, replaced the inhibitor I29 domain at the N terminus.

Motifs 1, 2, 3, 4, 6, 8, 9, 10, 11, 13, 14, and 19 were identified as belonging to the peptidase C1 domain (PF00112). Motifs 2, 10, and 8 had Asn, Cys, and His catalytic sites, respectively. Motif 10 of some SAG12 subfamily members had a double Cys in the catalytic site. The distribution of the specific conserved motifs varied among subfamilies. In RD19, ALP, and CTB subfamilies, motif 16 replaced motifs 11 and 6. In the RD21 and XBCP3

TABLE 2 | Putative motifs of fig PLCPs identified by MEME program.

Number	Motif consensus	Description
Motif 1	AAVEGINKIVTGKLVSLSEQZLVDCD	
Motif 2	YWJVKNWGTSWGEGGYIRIQRNV	PS00640 THIOL_PROTEASE_ASIN
Motif 3	LPESVDWRKKGAVTPVKBQGS	
Motif 4	GCONGLMDYAFEYIKNNGGJT	
Motif 5	EEKEKRFEIFKDNLRFIEEHN	Inhibitor_I29
Motif 6	LLKAVAHQPVSVDAIDA	
Motif 7	KKNRTYKGLNRFADLTNEEF	Inhibitor_I29
Motif 8	LBHGVLAVGYG	PS00639 THIOL_PROTEASE_HIS
Motif 9	ZGLCGIAMEASYPTK	
Motif 10	CGSCWAFS	PS00139 THIOL_PROTEASE_CYS
Motif 11	ANSHAVTIDGYEDVPAN DEKA	
Motif 12	EASVEELFEQWMARHG KVYNS	Inhibitor_I29
Motif 13	FQFYSSGV	
Motif 14	BYPYTGKDGTC	
Motif 15	CFSWGCPCLESATCCDDNYSCCP HDYPVCDVDAGTCLKS	pfam00396 GRAN: Granulin superfamily
Motif 16	NFSVISLDEDQIAANLVKNGPLE VAINAV	
Motif 17	KLTSPILQDEIVKEVNNPEAGWKAA MNPFRFSNATVAEFKRLLVGKPTPK	
Motif 18	PPPPSPPTVCDKYYSCEGET CCCJYE	
Motif 19	FTGSCGTE	
Motif 20	CDPYFDNTGCSHPGC	

The conserved motifs exported by MEME program using the PLCP sequences of *F. carica*, *A. thaliana*, and *M. notabilis*. Letters in red are conserved residues.

subfamilies, motifs 18 and 15 were usually located at the C terminus (Figure 3).

Motif 15 appeared in two members of the RD21 subfamily and two members of the XBCP3 subfamily (Figure 3). It was identified as a granulin domain (PF00396). In plants, the granulin motif has been found at the C terminus of some cysteine proteases whose expression is upregulated under environmental stress (Bateman and Bennett, 2009). Not every XBCP3 and RD21 subfamily member carried the granulin motif, suggesting that the granulin polymorphism evolved by domain loss (Richau et al., 2012).

All CEP subfamily *FcPLCPs* had the KDEL sequence at the C terminus for retention in the endoplasmic reticulum. Than et al. (2004) found that removing the N-terminal propeptide and C-terminal KDEL sequence under acidic conditions results in ricinosome maturation. We also found that the N terminus of *FcALP1* carries an NPIR signal known as a vacuolar-targeting sequence.

Fig PLCPs are 207 to 505 amino acids in length, and have a molecular mass of 23.09–56.19 kDa and pI of 3.88 to 9.14 (Supplementary Table 2). The GRAVY index of *FcPLCPs* was from –0.59 to –0.14, indicating they are hydrophilic; 19 *FcPLCPs*, 31 *AtPLCPs*, and 20 *MnPLCPs* were predicted to carry N-terminal signal peptides (Figure 3). In several PLCP

sequences annotated in the present study, neither signal peptides nor propeptides were identified. This could be due to the assembly quality of the used fig genome (Usai et al., 2020, 74-fold coverage of the cv. Dottato haploid genome) and the possible existence of pseudogenes. Most *FcPLCPs* were predicted to be soluble proteins. Subcellular localization analysis showed that 24, 5, and 2 *FcPLCPs* were located in the lysosome/vacuole, extracellular matrix and cytoplasm, respectively. All 32 *AtPLCPs* were predicted to be localized to the lysosome/vacuole. *MnPLCPs* were located in the lysosome/vacuole (21 *PLCPs*), extracellular matrix (1 *PLCP*), plastid (1 *PLCP*), and cytoplasm (1 *PLCP*) (Supplementary Table 3).

Promoter Analysis

All *FcPLCP* promoters contained at least one putative biotic/abiotic stress response element; 25 *PLCPs* contained an abscisic acid response element, 22 contained a methyl jasmonate response element, 19 contained a salicylic acid response element, 17 contained a low-temperature response element, 16 contained an auxin response element, 14 had a gibberellin response element, 16 had a defense and stress response element and 14 had a drought response element (Figure 4), suggesting that *FcPLCPs* are involved in stress responses. Moreover, all of the *FcPLCP* promoters contained light response elements, and 29 of them contained anaerobic induction elements, indicating that their expression is regulated by light and oxygen (Supplementary Table 4).

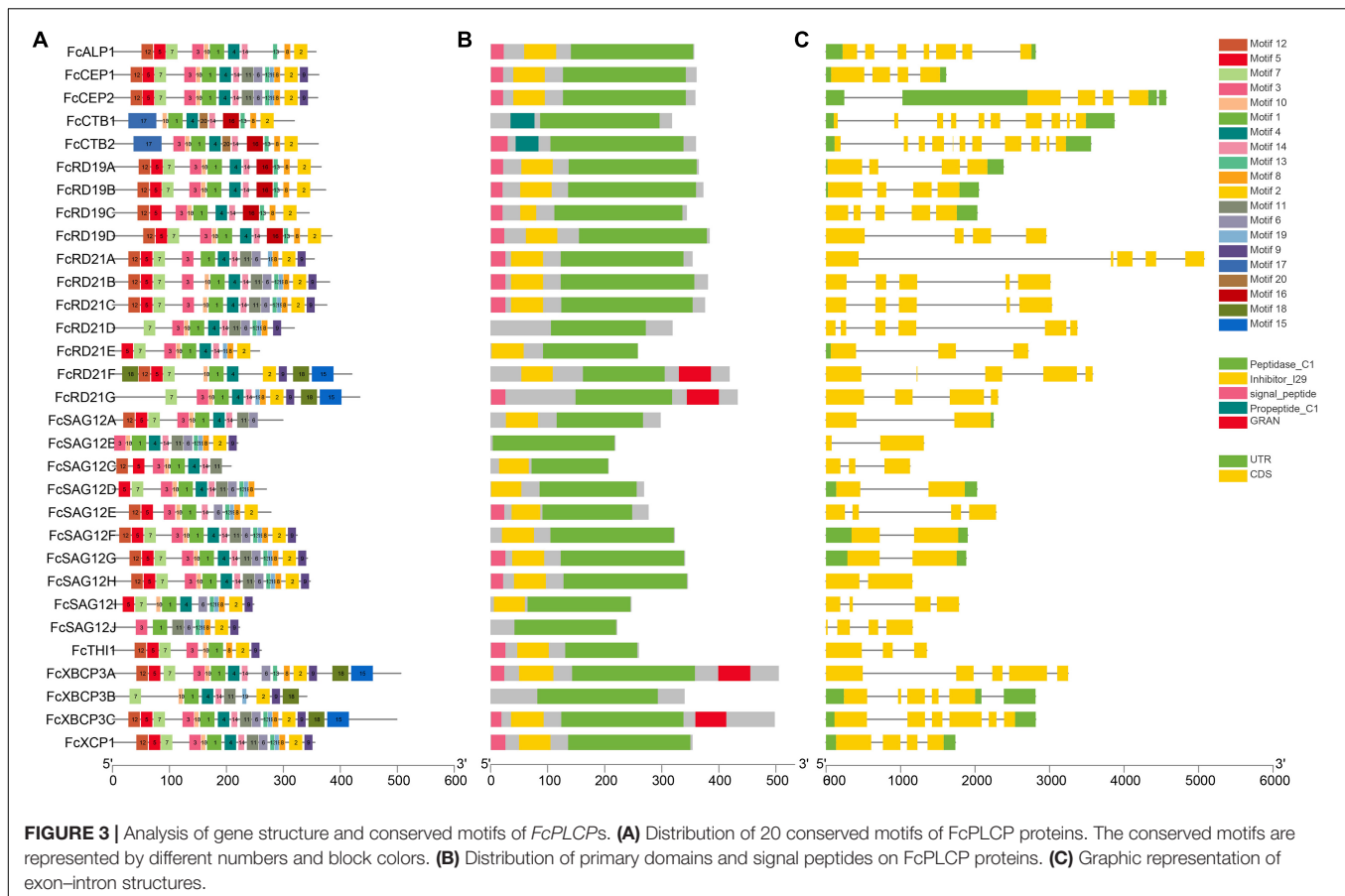
Expression Pattern of PLCPs in Fig Fruit Development

Seventeen *PLCPs* were identified in the inflorescence and receptacle transcriptomes of the six fig fruit developmental stages; no tissue-specific *PLCP* was found in the inflorescence or receptacle. The RD21 subfamily made up the largest proportion of detected expressed *PLCPs* (Figure 5).

Along inflorescence development, the expression of *c41414_g2* (*FcCEP1*) and *c47131_g1* (*FcRD21A*, *FcRD21B*, *FcRD21C*) decreased. In contrast, the expression of *c42827_g1* (*FcRD21F*), *c13313_g2* (*FcRD19C*, *FcRD19B*), *c13373_g1* (*FcXBCP3A*), *c43917_g2* (*FcRD21G*), and *c45335_g2* (*FcTH11*) increased. The expression of *c42134_g1* (*FcRD19A*) and *c78176_g1* (*FcALP1*) first increased and then decreased; *c41414_g2* (*FcCEP1*) expression was abundant at the early stage.

PLCP genes *c42827_g1* (*FcRD21F*), *c42134_g1* (*FcRD19A*), *c78176_g1* (*FcALP1*), *c13313_g2* (*FcRD19C*, *FcRD19B*), and *c13373_g1* (*FcXBCP3A*) were upregulated at the late stage of receptacle development, whereas *c47131_g1* (*FcRD21A*, *FcRD21B*, *FcRD21C*), annotated as ficin, was highly expressed from R1 to R4, and then dropped in R5 and R6. The expression of *c78176_g1* (*FcALP1*), which was also annotated as ficin, increased during receptacle development.

The changes in transcription abundance of eight *PLCPs* during fig fruit development were verified by qRT-PCR. The expression trends were consistent with the RNA-seq results, indicating their reliability (Supplementary Figure 3). The quality of the RNA-seq of fig fruit after ethylene treatment and under



light deprivation were validated by qRT-PCR when the data were analyzed for the original purposes of the studies (Wang et al., 2019; Cui et al., 2020).

Change in Abundance of PLCP Proteins During Fig Fruit Development

Protein identification and quantification were performed at three developmental stages (stages 1, 3, and 5) for inflorescences and receptacles. Eighteen PLCPs were annotated, and 13 of them were identified as DAPs. Their fold changes are shown in **Table 3** and **Supplementary Table 5**.

Through protein sequence alignment, ficin 2C was identified as FcALP1, ficin 1A, ficin 1B, and ficin 1C were FcRD21C, and ficin 3 was identified as FcRD21B. Moreover, ficin D, ficin 4, and ficin 5 exhibited 72.15, 74.54, and 78.13% sequence identity with FcRD21B, respectively, and ficin 6A had 78.33% sequence identity with FcRD21C (**Supplementary Figure 4**). Along fruit development, the abundance of ficins decreased in the inflorescence, e.g., ficin 2C was a downregulated DAP in F5 compared to F3 (F5/F3). In the receptacle, the abundance of these ficins decreased in stage 3 but increased in stage 5. Aside from ficin 2C (FcALP1), most ficin proteins were of higher abundance in the receptacle than in the inflorescence.

In addition to ficins, some other PLCPs were identified as DAPs, including FcXCP1 (increased in R5/F5), FcCTB2

(increased in R5/F5 and R5/R3), FcXBCP3C (increased in R5/R3), and FcCEP1 (decreased in R3/F3 and R5/F5).

PLCP Proteins in Fig Fruit Latex

Seventy-four proteins were identified in the latex, with molecular masses ranging from 8.9 to 206.1 kDa and pI values of 4.84 to 10.78. Latex contained many stress-response proteins, including trypsin-like protease inhibitor, chitinase, endochitinase, and pathogenesis-related (PR) protein isoforms, indicating that it plays an essential role in resistance to insects and microbes. PLCPs were the most abundant protein component in fig latex at the commercial-ripe stage, accounting for 38.93% of the total protein content. The identified PLCPs included ficin 4 (14.93%), ficin 1B (8.63%), ficin 1A (8.04%), W9RY43 (5.82%, FcRD21G), and ficin D (1.50%) (**Table 4** and **Supplementary Table 6**).

Expression Pattern of PLCPs in Fig Fruit Treated With Ethephon and Light Deprivation

Phase II fig fruit were treated with 1 mL of 250 mg/L ethephon (Cui et al., 2020). The RNA-seq data for the inflorescence and receptacle at 0, 2, 4, and 6 days after treatment were analyzed. Sixteen PLCPs showed differential expression (**Figure 6**). Eleven PLCPs were upregulated in the inflorescence or receptacle; among them, *c39535_g2* (*FcRD19D*), *c43917_g2*

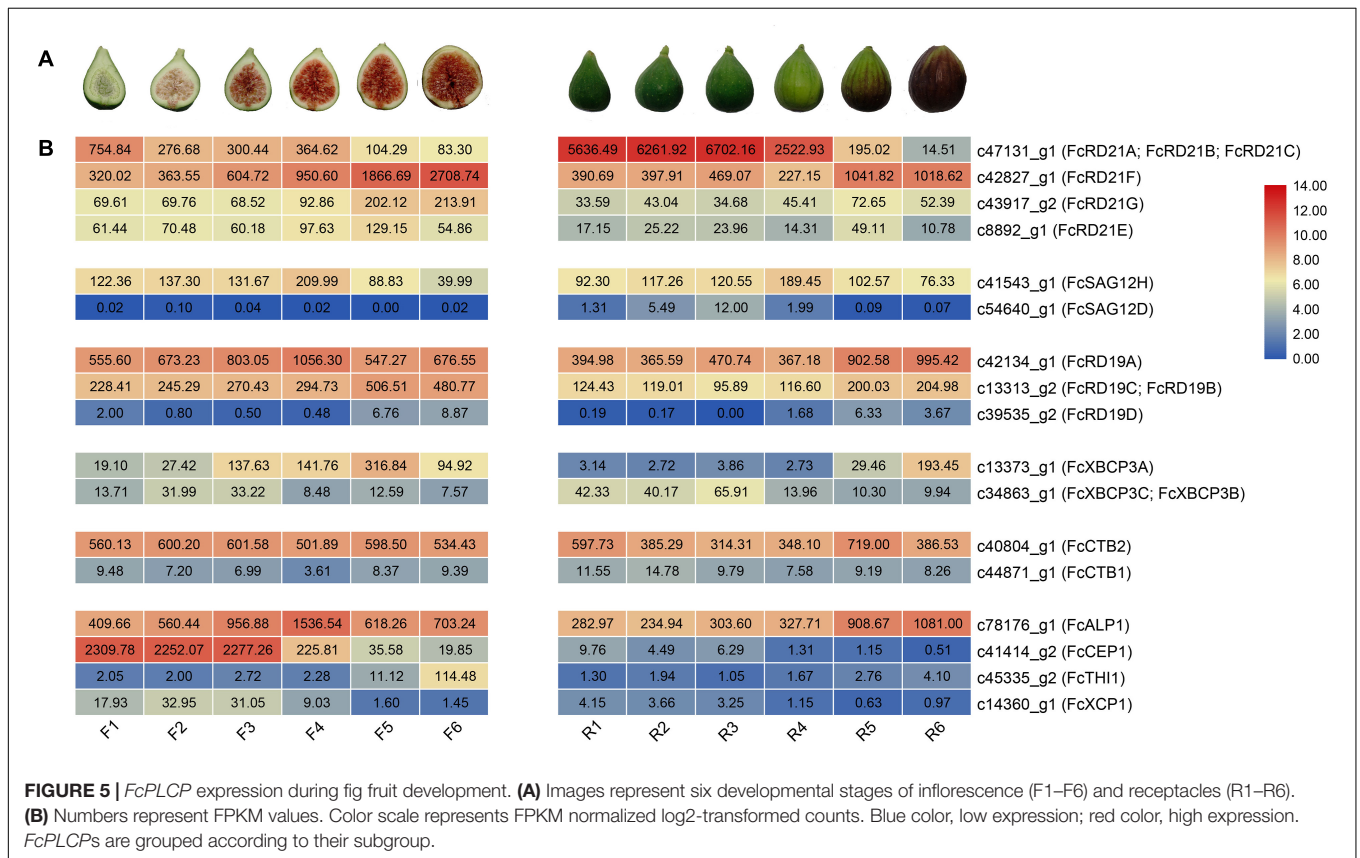
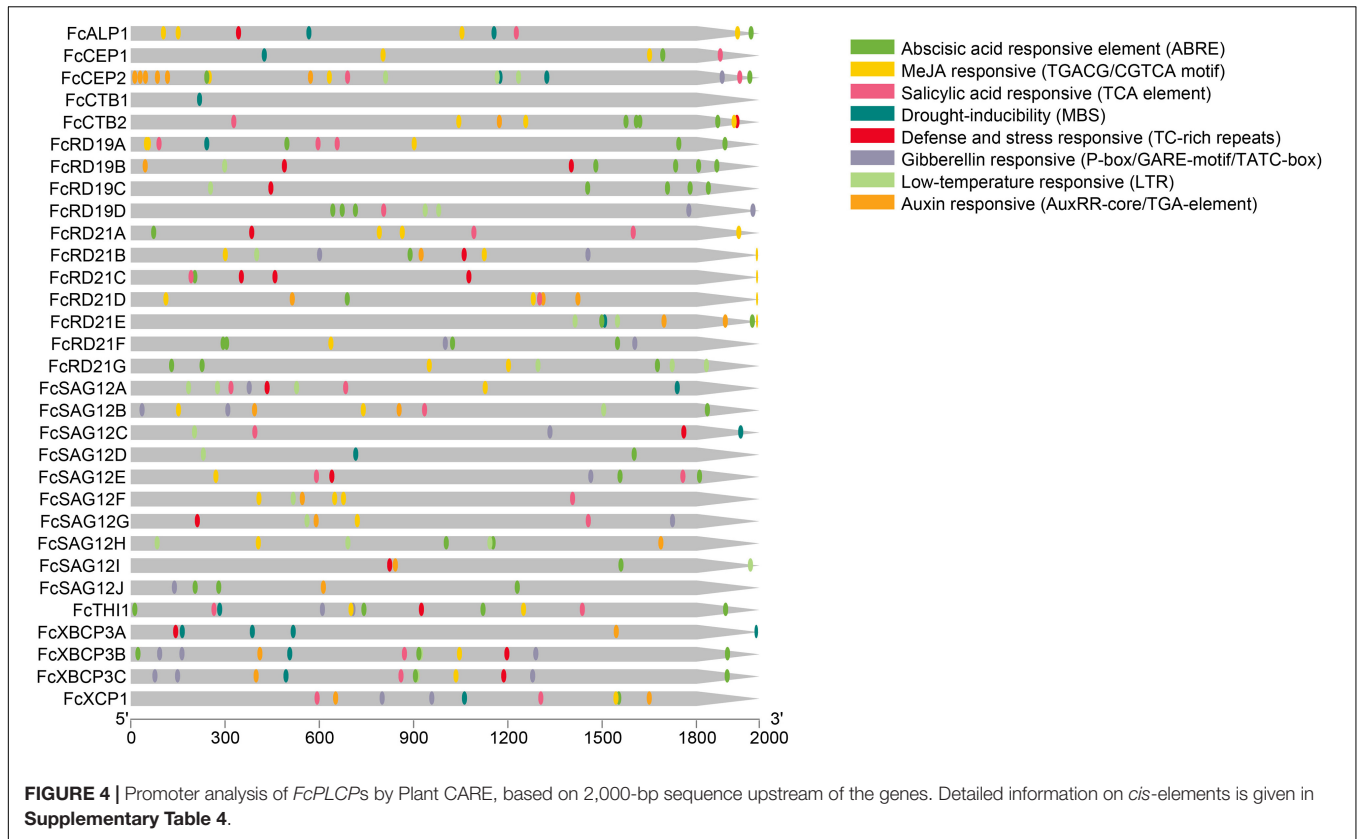


TABLE 3 | Papain-like cysteine proteases (PLCPs) identified in fig fruit soluble proteome and comparison of their abundance in the inflorescence and receptacle during fruit development.

Accession	Description	Gene ID	F3/F1	F5/F3	R3/R1	R5/R3	R1/F1	R3/F3	R5/F5
A0A2Z6DRM5	Ficin 2c	FcALP1	ns	0.451	ns	ns	0.779	ns	ns
A0A2Z6DRL4	Ficin 1a	FcRD21C	ns	ns	0.493	5.076	3.011	1.738	18.993
A0A2Z6DRT1	Ficin 1b	FcRD21C	ns	ns	0.505	ns	2.972	ns	8.625
A0A2Z6DRL5	Ficin 1c	FcRD21C	ns	0.533	ns	1.883	3.107	2.654	ns
A0A2Z6DRW8	Ficin 6a (Fragment)	FcRD21C	0.809	ns	ns	2.051	2.183	4.475	ns
A0A2Z6DRN1	Ficin 3	FcRD21B	ns	ns	0.767	ns	ns	ns	ns
A0A2Z6DRL6	Ficin 4	FcRD21B	0.657	ns	0.627	1.206	2.915	ns	ns
A0A2Z6DRM9	Ficin 5	FcRD21B	0.574	ns	ns	ns	2.105	ns	ns
A0A182DW11	Ficin isoform D	FcRD21B	ns	0.388	ns	2.604	ns	ns	31.287
W9QLF0	Xylem cysteine proteinase 1	FcXCP1	ns	ns	ns	ns	ns	ns	2.140
W9RJW0	Uncharacterized protein	FcCEP1	ns	ns	ns	ns	ns	0.273	0.439
W9RKF4	Cathepsin B	FcCTB2	ns	ns	ns	1.672	ns	ns	4.718
W9SBZ5	Oryzain alpha chain	FcXBCP3C	ns	ns	ns	1.621	ns	1.453	ns

Proteins with $p < 0.05$ and fold change ≥ 1.2 or ≤ 0.83 were identified as differentially abundant. ns, no significant difference in abundance; F, inflorescence; R, receptacle; 1, 3, and 5 represent fruit development stages corresponding to phase I (early stage of fruit development), late phase II (before fruit begins to ripen) and middle of phase III (commercial-ripe), respectively. Blue values represent downregulated; red values represent upregulated.

TABLE 4 | Papain-like cysteine proteases identified in the commercial-ripe fig fruit receptacle latex proteome.

Accession	Description	Score	Coverage (%)	PSMs	Intensity	Relative amount (%)
A0A2Z6DRL6	Ficin 4	1949.16	99.54	2680	145000000	14.93
A0A2Z6DRT1	Ficin 1b	1477.69	83.33	2389	83800000	8.63
A0A2Z6DRL4	Ficin 1a	1606.25	90.50	2169	78100000	8.04
A0A182DW11	Ficin isoform D	230.57	16.36	287	14600000	1.50
W9RY43	Cysteine proteinase RD21a	186.08	8.19	72	56500000	5.82

Detailed information can be found in **Supplementary Table 6**. PSMs, the peptide-spectrum matches.

(*FcRD21G*), *c42827_g1* (*FcRD21F*), *c45335_g2* (*FcTHI1*), *c13313_g2* (*FcRD19C*, *FcRD19B*) were significantly upregulated in both the inflorescence and receptacle. Seven PLCPs were downregulated in the inflorescence or receptacle after ethephon treatment. Among them, *c47131_g1* (*FcRD21A*, *FcRD21B*, *FcRD21C*) was the most downregulated transcript.

Fig fruit were deprived of light by bagging at stage 2 of development to study the biology of fig color development (Wang et al., 2019). The RNA-seq data for the inflorescences and receptacles of ripe fruit revealed 3 and 4 differentially expressed PLCPs, respectively: all of them were downregulated under light deprivation (Table 5). Genes *c47131_g1* (*FcRD21A*, *FcRD21B*, and *FcRD21C*) and *c8892_g1* (*FcRD21E*) were repressed in both the inflorescence and receptacle, whereas *c56852_g1* (*FcSAG12F*, *FcSAG12G*) was only downregulated in the inflorescence, and *c39535_g2* (*FcRD19D*) and *c13373_g1* (*FcXBCP3A*) were only downregulated in the receptacle.

DISCUSSION

PLCPs in *Ficus* Species

Fig belongs to the genus *Ficus* in the family Moraceae. The genus has about 700 species, most of them evergreens, including trees, shrubs and climbers growing under different climatic conditions

(Flaishman et al., 2008; Zhang et al., 2020). In three sequenced *Ficus* species genomes, 31, 35 and 34 PLCP genes were identified (*F. carica*, *F. hispida* and *F. microcarpa*, respectively). The number of PLCPs varied among species and subfamilies, possibly due to whole-genome duplication, tandem duplication, and large-scale segmental duplication (Liu et al., 2018). In eukaryotic species, gene duplications are estimated to occur at an average rate of 0.01 per gene per million years (Lynch and Conery, 2000).

PLCPs of the three *Ficus* species differed in the structure of their subfamilies. In *F. carica*, SAG12 (10 members) and RD21 (7) were the two largest subfamilies, whereas in *F. hispida*, SAG12 (10) and RD19 (10) were the two largest; the RD21 subfamily had 5 PLCPs, and no THI subfamily members were identified, possibly being lost during evolution: the THI subfamily has been reported as lost in poplar (Zou et al., 2017). In *F. microcarpa*, RD21 (11) and SAG12 (10) were the two largest subfamilies. The three *Ficus* species all had 10 SAG12 subfamily members, whereas in mulberry and *Arabidopsis*, SAG12 had 9 and 6 members, respectively. It is speculated that SAG12 PLCPs may have formed before the formation of the individual species. SAG12 PLCPs are senescence-associated genes and their expression increases in senescing leaves (James et al., 2018).

The RD19 subfamily showed significant expansion in *F. hispida*, and the RD21 subfamily showed significant expansion in *F. microcarpa*, which could be due to local

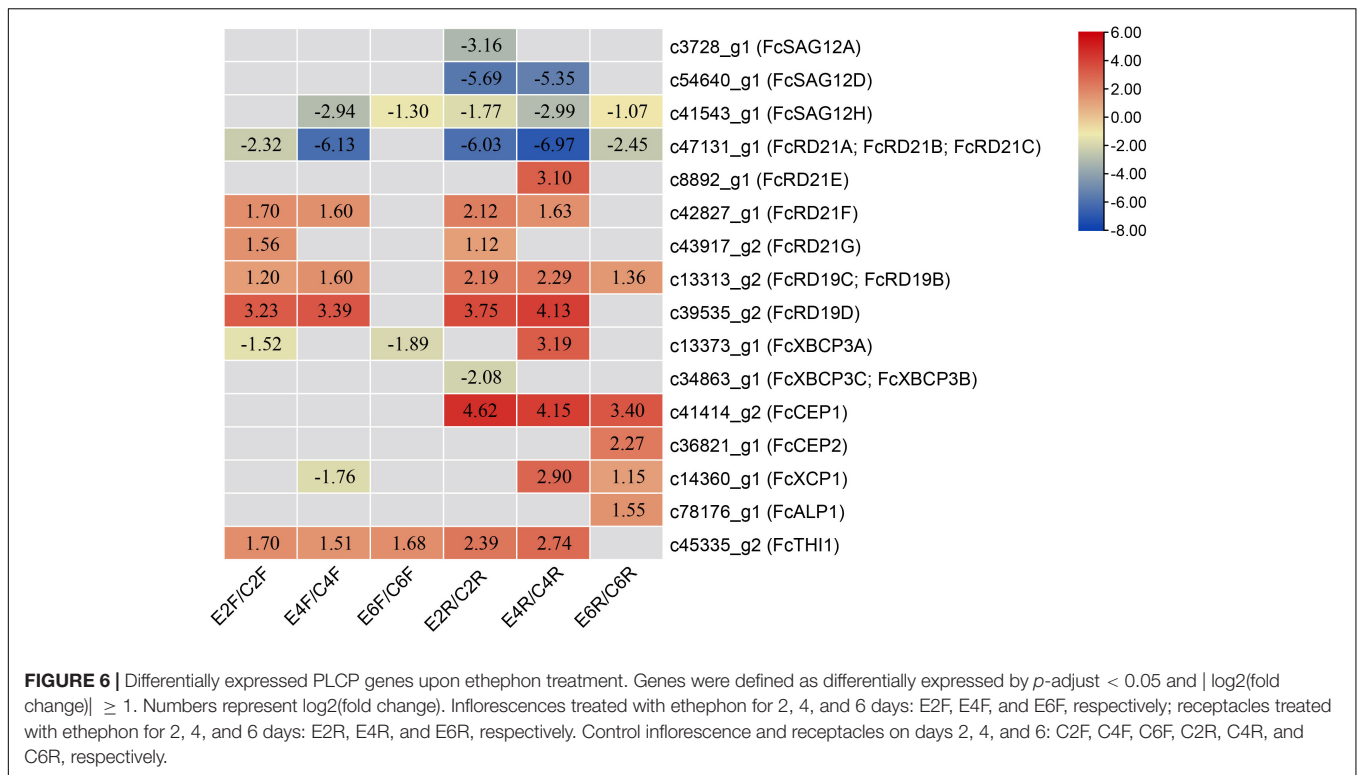


TABLE 5 | Differentially expressed PLCP genes in fig fruit subjected to light deprivation.

Gene ID	Inflorescence control (FC) (FPKM)	Inflorescence light deprivation (FL) (FPKM)	$\log_2\text{FC}(\text{FL}/\text{FC})$	Receptacle control (RC) (FPKM)	Receptacle light deprivation (RL) (FPKM)	$\log_2\text{FC}(\text{RL}/\text{RC})$
c56852_g1 (FcSAG12F; FcSAG12G)	4.98	0.41	-3.62	ns	ns	ns
c47131_g1 (FcRD21A; FcRD21B; FcRD21C)	174.26	31.18	-2.48	66.76	15.60	-2.09
c8892_g1 (FcRD21E)	18.58	5.67	-1.71	17.19	4.38	-1.97
c39535_g2 (FcRD19D)	ns	ns	ns	10.08	2.27	-2.15
c13373_g1 (FcXBCP3A)	ns	ns	ns	55.83	5.93	-3.23

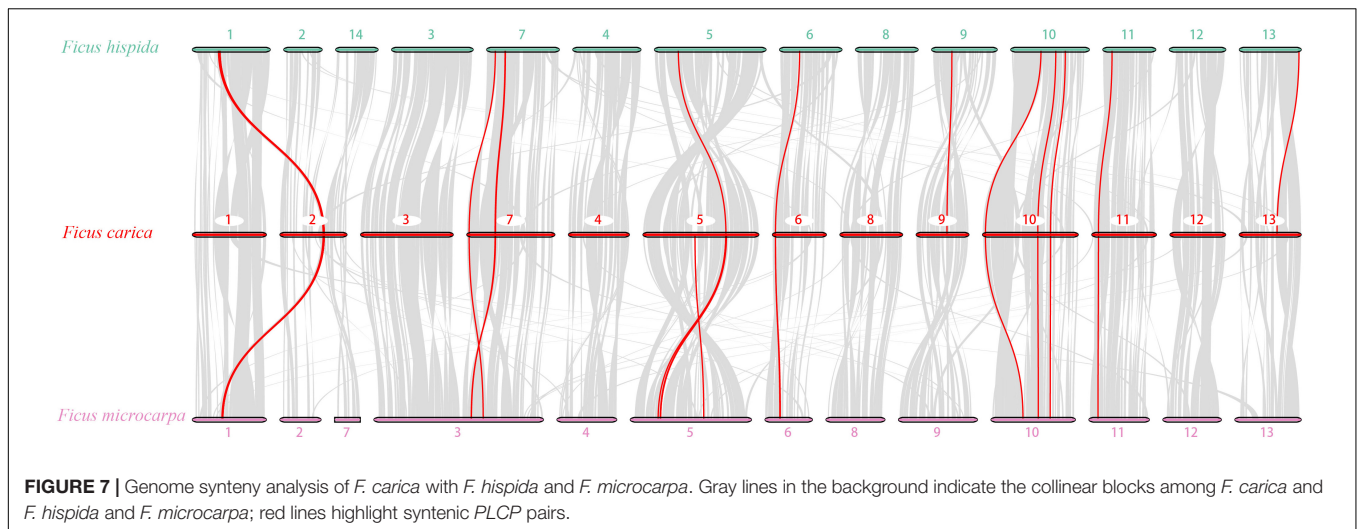
ns, no significant difference; Log_2FC (fold change).

gene duplication. Both RD19 and RD21 subfamilies are involved in disease resistance. *Arabidopsis rd19*-null mutants have been found to be more susceptible than the wild type to the bacterial pathogen *Ralstonia solanacearum*, whereas *rd21*-null mutants show impaired resistance to *Botrytis cinerea* (Misas-Villamil et al., 2016).

In *F. hispida*, 35 *PLCPs* were distributed on 10 of the 14 chromosomes (**Supplementary Figure 5**); 7 SAG12 subfamily members and 8 RD19 members formed gene clusters on chr1 (GWHALOG000000001) and chr4 (GWHALOG000000004), respectively. The formation of gene clusters may be related to the genes' tandem duplication. In *F. microcarpa*, 34 *PLCPs* were distributed on 9 of 13 chromosomes (**Supplementary Figure 6**); 7 SAG12 subfamily members and 9 RD21 members formed gene clusters on chr1 (GWHABKV000000001.1) and chr5 (GWHABKV000000005.1), respectively. In *F. carica*, 31

PLCPs were distributed on 9 of 13 chromosomes. Seven SAG12 members and 4 RD21 members formed gene clusters on chr2 (CM019740.1) and chr5 (CM019743.1), respectively. The genome distribution and similar clustering of the *PLCP* genes from the three *Ficus* species indicated mutual whole-genome duplications and species-specific tandem duplications of specific subfamilies, as reported in castor bean, physic nut (Zou et al., 2018) and cotton (Zhang et al., 2019).

To further analyze the *PLCP* family's phylogenetic mechanism, we constructed comparative syntenic maps of *F. carica* associated with *F. hispida* and *F. microcarpa*. The three *Ficus* species' genomes showed high homology (**Figure 7**). A total of 16 and 20 *FcPLCPs* showed syntenic relationships with those in *F. hispida* and *F. microcarpa*, respectively (**Supplementary Table 7**), indicating that *F. carica* may have a closer evolutionary relationship with *F. microcarpa* than with *F. hispida*. Some



syntenic gene pairs were conserved in the three *Ficus* species, including some SAG12 and RD21 subfamily members, indicating that their expansion occurred before the three species' divergence. A chromosome fusion or fission event occurred between *F. microcarpa* chr3 and its homologs, chr3 and chr7 of *F. carica* and *F. hispida*, respectively (Zhang et al., 2020). *FcRD19A*, *FcXBCP3C*, and *FcRD21F* on chr7 were identified as syntenic genes. *F. carica* chr2, *F. microcarpa* chr2 and 7, and *F. hispida* chr2 and 14 may also have undergone fusion or fission events. Several inversions occurred in the three genomes' chromosomal fragments, calling for further study.

PLCP Expression Patterns

Ficus is characterized by unique aggregate fruit that develop from an enclosed urn-shaped inflorescence; the receptacle serves as a physical barrier, protecting the enclosed inflorescence and small drupelets from disease and insects. A few recent studies have revealed that the inflorescence and surrounding receptacle of figs differ with respect to ripening (Freiman et al., 2015), and their response to gibberellin (Chai et al., 2018, 2019) and ethephon treatment (Cui et al., 2020).

In the present study, among the 31 sequences recruited as PLCP-encoding genes from the published fig genome, 17 PLCP transcripts and 18 PLCP proteins were identified, and their spatiotemporal expression/abundance pattern in the inflorescence and receptacle was revealed by transcriptome and proteome analysis, respectively. Limited by the fact that only fig fruit material was used in our study, we are not in a position to suggest which remaining sequences are transcribed/translated in other tissues or, in other words, are pseudogenes. The present results emphasize the divergent roles of the inflorescence and receptacle in fig fruit reproductive biology. The spatial expression pattern of PLCPs has been reported in *Arabidopsis* (Richau et al., 2012), rubber tree (Zou et al., 2017), cotton (Zhang et al., 2019), castor bean and physic nut (Zou et al., 2018).

Pollination of the *Ficus* inflorescence relies on species-specific wasps (*Blastophaga psenes* L.) (Flaishman et al., 2008;

Zhang et al., 2020). The wasps enter the enclosed fig fruit through the scale-covered ostiole, and pollination occurs inside the urn-shaped syconia at the end of phase I of fig fruit development. Bacterial and fungal pathogens and the diseases that they transmit have been reported in fig production: *Alternaria* and *Fusarium* were the two major pathogens, introduced by fig wasps and producing internal fig rot (Stover et al., 2007). Transcriptome analyses of the inflorescence demonstrated high *CEP1* levels in the young fruit. AtCEP1 has been reported to be essential in tapetum programmed cell death and pollen development (Liu et al., 2018). The relatively high expression of members of the *FcRD21* and *FcRD19* subfamilies, both involved in resistance to biotic stress, suggests that tissue-specific PLCP expression could play a role in supporting fig inflorescences' control of the biological risk associated with the wasp pollinator entering the syconium at the early stage of fruit development.

Among of the 18 PLCPs identified in the fig's soluble proteome, 6 were identified as DAPs in the inflorescence, all of them ficins. Significantly decreased abundance ($p < 0.05$, fold change ≤ 0.83) was found for ficin 1C, ficin 2C, ficin 4, ficin 5, ficin 6A and ficin isoform D during inflorescence development. When the fig fruit ripens, the decreased abundance of ficin and other PLCPs in the inflorescence—the major edible part of fig fruit—may indicate a decreased requirement for antibacterial agents, thereby facilitating the feeding of dispersing organisms.

A large number of PLCPs, especially ficins, are toxic to herbivorous insects (Konno et al., 2004). The biotic risk faced by the fig receptacle is different from that faced by the inflorescence, the latter being well protected from most insects by the receptacle and scales covering the ostiole. PLCPs of subfamilies RD21, RD19, ALP1, and CTB2 showed high transcription levels in the receptacle. In papaya fruit, subfamily III PLCP genes, including the papain gene *CpXCP5*, are expressed at high levels in stage I fruit; the papain and other major PLCPs in papaya latex provide defense against herbivorous insects as the papaya fruit develops (Liu et al., 2018). In our study, only one putative *FcXCP1* transcript was identified in fig fruit and it showed

basal expression in the receptacle, whereas one RD21 member (*c47131_g1*), previously shown to have roles in plant immunity and resistance to necrotrophic fungal pathogens and arthropod crustaceans (Rustgi et al., 2018), exhibited extremely high expression in the receptacle before fig fruit ripening.

Our fig fruit receptacle proteomic data also supported a significant increase in abundance of ficin 4, ficin 1A, ficin 1C, ficin 6A, ficin isoform D, CTB2 and XBCP3C from mid-stage fig development to near commercial ripeness. Latex collected from the receptacle of commercial-ripe figs was rich in ficin 4, ficin 1A, ficin 1B and ficin isoform D. The major PLCP components in the receptacle latex were similar to those found in the receptacle. The presence of ficins in the latex confirms their role in plant resistance to microbes and herbivores (Kitajima et al., 2018). High transcription and protein abundance of the major PLCPs in the fig receptacle in the commercial-ripe fruit suggest strong and persistent PLCP protection of the receptacle against biotic stresses.

Moreover, *PLCP* expression has been reported to be modulated by plant hormones and environmental stimuli. Ethephon is regularly applied to rubber trees to increase the yield of rubber latex (Zhu and Zhang, 2009). In our study, most of the PLCPs were upregulated following ethephon application. Commonly found light-responsive elements in *PLCP* promoters and a comparison of light-deprived and natural grown fig fruit transcriptomes strongly suggest that light is a positive signal in the expression of most *PLCPs*. In support of this, a study with smyrna-type fig cultivars found the highest protease activity in the late afternoon after long light exposure (Lazreg-Aref et al., 2018). Recently, differently changing patterns of protease activity in different fig types and cultivars have been reported. In the common-type cultivar Kahlia and the San Pedro-type cultivar Bither Abiadh, protease activity decreases with maturity, whereas in the smyrna-type cultivars Njali and Temri, and the caprifig cv. Besbessi, protease activity increases with maturity (Lazreg-Aref et al., 2018). This shows that in addition to the stage of development, cultivar may also be an important factor affecting *PLCP* expression in fig fruit, warranting further study.

CONCLUSION

In this study, the PLCP family was analyzed in fruit of the edible fig *F. carica* at the level of gene structure, sequence characteristics, promoter *cis*-elements, expression patterns and proteomics. Species-specific PLCP subfamily duplication was revealed, which could be relevant to the uniqueness of edible fig, being the only deciduous species of *Ficus*, which has been under long selection pressure by humans as one of the earliest domesticated fruit trees. High expression of disease- and herbivore-resistance/repelling PLCP genes and a high abundance of ficins in the inflorescence, receptacle and fruit latex provide valuable information on fig fruit developmental biology. Comparisons with other *Ficus* species and a comparison between cultivated figs (common fig) and wild-type figs (caprifig) are needed to better understand *Ficus* PLCPs and their role in plant development, and especially in biotic stress responses.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: NCBI, accession numbers: PRJNA606407, PRJNA494945, PRJNA723733, and ProteomeXchange consortium, via PRIDE partner repository: PXD025170 and PXD025485.

AUTHOR CONTRIBUTIONS

YZ, YC, and MS conducted the experiments and data analyses. YZ, AV, SC, and HM prepared the manuscript. All authors have read and approved the manuscript for publication.

FUNDING

This work was supported by National Natural Science Foundation of China project NSFC (31372007), and 111 Project (B17043).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2021.681801/full#supplementary-material>

Supplementary Figure 1 | Analysis of gene structure and conserved motifs of *Ficus carica*, *Arabidopsis thaliana*, and *Morus notabilis*.

Supplementary Figure 2 | Conserved motifs of fig PLCPs identified by MEME program. Letters represent amino acids, their size represents the conservation of residues at individual positions.

Supplementary Figure 3 | Transcriptomic changes during fig fruit development validated by qRT-PCR using 8 PLCP genes.

Supplementary Figure 4 | Sequence alignment of FcPLCPs and previously identified ficins.

Supplementary Figure 5 | Chromosomal location of *Ficus hispida* PLCP genes.

Supplementary Figure 6 | Chromosomal location of *Ficus microcarpa* PLCP genes.

Supplementary Table 1 | Primer sequences of 8 FcPLCP genes used for quantitative real-time PCR.

Supplementary Table 2 | Papain-like cysteine proteases (PLCP) gene and predicted protein features of *Arabidopsis thaliana*, *Morus notabilis*, *Ficus hispida* and *Ficus microcarpa*.

Supplementary Table 3 | Prediction of FcPLCPs' subcellular location.

Supplementary Table 4 | *Cis*-element analysis of FcPLCP promoters (2000 bp).

Supplementary Table 5 | Identification and relative quantification of PLCPs in fig soluble proteome.

Supplementary Table 6 | Proteins identified from commercial-ripe fig fruit receptacle latex proteome.

Supplementary Table 7 | Syntenic gene pairs within *F. carica*, *F. hispida*, and *F. microcarpa*.

REFERENCES

- Armenteros, J. J. A., Sønderby, C. K., Sønderby, S. K., Nielsen, H., and Winther, O. (2017). DeepLoc: prediction of protein subcellular localization using deep learning. *Bioinformatics* 33, 3387–3395. doi: 10.1093/bioinformatics/btx431
- Armenteros, J. J. A., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., et al. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* 37, 420–423. doi: 10.1038/s41587-019-0036-z
- Bah, S., Paulsen, B. S., Diallo, D., and Johansen, H. T. (2006). Characterization of cysteine proteases in Malian medicinal plants. *J. Ethnopharmacol.* 107, 189–198. doi: 10.1016/j.jep.2006.03.008
- Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202–W208.
- Bateman, A., and Bennett, H. P. (2009). The granulin gene family: from cancer to dementia. *Bioessays* 31, 1245–1254. doi: 10.1002/bies.200900086
- Bekhit, A. A., Hopkins, D. L., Geesink, G., Bekhit, A. A., and Franks, P. (2014). Exogenous proteases for meat tenderization. *Crit. Rev. Food Sci. Nutr.* 54, 1012–1031. doi: 10.1080/10408398.2011.623247
- Chai, L., Chai, P., Chen, S., Flaishman, M. A., and Ma, H. (2018). Transcriptome analysis unravels spatiotemporal modulation of phytohormone-pathway expression underlying gibberellin-induced parthenocarpic fruit set in San Pedro-type fig (*Ficus carica* L.). *BMC Plant Biol.* 18:100. doi: 10.1186/s12870-018-1318-1
- Chai, L., Li, Y., Chen, S., Perl, A., Zhao, F., and Ma, H. (2014). RNA sequencing reveals high resolution expression change of major plant hormone pathway genes after young seedless grape berries treated with gibberellin. *Plant Sci.* 229, 215–224. doi: 10.1016/j.plantsci.2014.09.010
- Chai, P., Dong, S., Chai, L., Chen, S., Flaishman, M., and Ma, H. (2019). Cytokinin-induced parthenocarp of San Pedro type fig (*Ficus carica* L.) main crop: explained by phytohormone assay and transcriptomic network comparison. *Plant Mol. Biol.* 99, 329–346. doi: 10.1007/s11103-019-00820-2
- Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., et al. (2020). TBtools – an integrative toolkit developed for interactive analyses of big biological data. *bioRxiv* 13, 1194–1202. doi: 10.1016/j.molp.2020.06.009
- Cui, Y., Wang, Z., Chen, S., Vainstein, A., and Ma, H. (2019). Proteome and transcriptome analyses reveal key molecular differences between quality parameters of commercial-ripe and tree-ripe fig (*Ficus carica* L.). *BMC Plant Biol.* 19:146. doi: 10.1186/s12870-019-1742-x
- Cui, Y., Zhai, Y., Flaishman, M., Li, J., Chen, S., Zheng, C., et al. (2020). Ethephon induces coordinated ripening acceleration and divergent coloration responses in fig (*Ficus carica* L.) flowers and receptacles. *Plant Mol. Biol.* 105, 347–364. doi: 10.1007/s11103-020-01092-x
- Dubey, V. K., Pande, M., Singh, B. K., and Jagannadham, M. V. (2007). Papain-like proteases: applications of their inhibitors. *Afr. J. Biotechnol.* 6, 1077–1086.
- El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., et al. (2019). The Pfam protein families database in 2019. *Nucleic Acids Res.* 47, D427–D432. doi: 10.1093/nar/gky995
- Flaishman, M. A., Rodov, V., and Stover, E. (2008). The fig: botany, horticulture, and breeding. *Hort. Rev.* 34, 113–196. doi: 10.1002/9780470380147.ch2
- Freiman, Z. E., Rosianskey, Y., Dasmohapatra, R., Kamara, I., and Flaishman, M. A. (2015). The ambiguous ripening nature of the fig (*Ficus carica* L.) fruit: a gene-expression study of potential ripening regulators and ethylene-related genes. *J. Exp. Bot.* 66, 3309–3324. doi: 10.1093/jxb/erv140
- Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D., and Bairoch, A. (2005). "Protein identification and analysis tools on the ExPASy server," in *The Proteomics Protocols Handbook*, ed. J. M. Walker (Totowa, NJ: Humana Press) 531, 571–607. doi: 10.1385/1-59259-58
- Huang, L., Qu, H. Z., Zhang, L., Du, S. S., Yang, S., Hao, D. Y., et al. (2008). Purification and characterization of a proteolytic enzyme from fig latex. *Chem. Res. Chin. Univ.* 24, 348–352. doi: 10.1016/s1005-9040(08)60073-2
- Hurst, L. D. (2002). The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends Genet.* 18:486. doi: 10.1016/s0168-9525(02)02722-1
- James, M., Poret, M., Masclaux-Daubresse, C., Marmagne, A., Coquet, L., Jouenne, T., et al. (2018). SAG12, a major cysteine protease involved in nitrogen allocation during senescence for seed production in *Arabidopsis thaliana*. *Plant Cell Physiol.* 59, 2052–2063. doi: 10.1093/pcp/pcy125
- Kitajima, S., Aoki, W., Shibata, D., Nakajima, D., Sakurai, N., Yazaki, K., et al. (2018). Comparative multi-omics analysis reveals diverse latex-based defense strategies against pests among latex-producing organs of the fig tree (*Ficus carica*). *Planta* 247, 1423–1438. doi: 10.1007/s00425-018-2880-3
- Konno, K., Hirayama, C., Nakamura, M., Tateishi, K., Tamura, Y., Hattori, M., et al. (2004). Papain protects papaya trees from herbivorous insects: role of cysteine proteases in latex. *Plant J.* 37, 370–378. doi: 10.1046/j.1365-313x.2003.01968.x
- Lazreg-Aref, H., Gaaliche, B., Ladhari, A., Hammami, M., and Hammami, S. O. (2018). Co-evolution of enzyme activities and latex in fig (*Ficus carica* L.) during fruit maturity process. *South Afr. J. Bot.* 115, 143–152. doi: 10.1016/j.sajb.2018.01.022
- Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., et al. (2002). PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* 30, 325–327. doi: 10.1093/nar/30.1.325
- Liu, J., Sharma, A., Niewiara, M. J., Singh, R., Ming, R., and Yu, Q. (2018). Papain-like cysteine proteases in *Carica papaya*: lineage-specific gene duplication and expansion. *BMC Genomics* 19:26. doi: 10.1186/s12864-017-4394-y
- Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2⁻ΔΔCT method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262
- Lynch, M., and Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes. *Science* 290, 1151–1155. doi: 10.1126/science.290.5494.1151
- Marchler-Bauer, A., and Bryant, S. H. (2004). CD-Search: protein domain annotations on the fly. *Nucleic Acids Res.* 32, W327–W331. doi: 10.1093/nar/gkh454
- Martinez, M., and Diaz, I. (2008). The origin and evolution of plant cystatins and their target cysteine proteinases indicate a complex functional relationship. *BMC Evol. Biol.* 8:198. doi: 10.1186/1471-2148-8-198
- Mazri, C., el Ain Soumia, Ha, and Siar, H. (2018). Characterization and application of phytochemicals substances of the fig tree: biological and sensory characterization of ficin and cheeses "Fresh and Soft". *Asian J. Appl. Sci. Eng.* 7, 71–78. doi: 10.18034/ajase.v7i2.226
- Misas-Villamil, J. C., van der Hoorn, R. A., and Doehlemann, G. (2016). Papain-like cysteine proteases as hubs in plant immunity. *N. Phytol.* 212, 902–907. doi: 10.1111/nph.14117
- Mnif, I. H., Siala, R., Nasri, R., Mhamdi, S., Nasri, M., and Kamoun, A. S. (2015). A cysteine protease isolated from the latex of *Ficus microcarpa*: purification and biochemical characterization. *Appl. Biochem. Biotechnol.* 175, 1732–1744. doi: 10.1007/s12010-014-1376-2
- Ossowski, S., Schneeberger, K., Lucas-Lledó, J. I., Warthmann, N., Clark, R. M., Shaw, R. G., et al. (2010). The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science* 327, 92–94. doi: 10.1126/science.1180677
- Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D. J., et al. (2019). The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* 47, D442–D450.
- Polaina, J., and MacCabe, A. P. (2007). *Industrial Enzymes – Structure, Function and Applications*. Germany: Springer.
- Raskovic, B., Bozovic, O., Prodanovic, R., Niketic, V., and Polovic, N. (2014). Identification, purification and characterization of a novel collagenolytic serine protease from fig (*Ficus carica* var. Brown Turkey) latex. *J. Biosci. Bioeng.* 118, 622–627. doi: 10.1016/j.jbiosc.2014.05.020
- Raskovic, B., Lazic, J., and Polovic, N. (2016). Characterisation of general proteolytic, milk clotting and antifungal activity of *Ficus carica* latex during fruit ripening. *J. Sci. Food Agric.* 96, 576–582. doi: 10.1002/jsfa.7126
- Rawlings, N. D., Barrett, A. J., and Bateman, A. (2010). MEROPS: the peptidase database. *Nucleic Acids Res.* 38, D227–D233. doi: 10.1093/nar/gkp971
- Rawlings, N. D., Barrett, A. J., and Finn, R. (2016). Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* 44, D343–D350. doi: 10.1093/nar/gkv1118
- Richau, K. H., Kaschani, F., Verdoes, M., Pansuriya, T. C., Niessen, S., Stüber, K., et al. (2012). Subclassification and biochemical analysis of plant papain-like cysteine proteases displays subfamily-specific characteristics. *Plant Physiol.* 158, 1583–1599. doi: 10.1104/pp.112.194001

- Rustgi, S., Boex-Fontvieille, E., Reinbothe, C., von Wettstein, D., and Reinbothe, S. (2018). The complex world of plant protease inhibitors: insights into a Kunitz-type cysteine protease inhibitor of *Arabidopsis thaliana*. *Commun. Integr. Biol.* 11:e1368599. doi: 10.1080/19420889.2017.1368599
- Shindo, T., and Van Der Hoorn, R. A. (2008). Papain-like cysteine proteases: key players at molecular battlefields employed by both plants and their invaders. *Mol. Plant Pathol.* 9, 119–125. doi: 10.1111/j.1364-3703.2007.00439.x
- Sigrist, C. J., De Castro, E., Cerutti, L., Cuče, B. A., Hulo, N., Bridge, A., et al. (2012). New and continuing developments at PROSITE. *Nucleic Acids Res.* 41, D344–D347. doi: 10.1093/nar/gks1067
- Stover, E., Aradhya, M., Ferguson, L., and Crisosto, C. H. (2007). The fig: overview of an ancient fruit. *HortScience* 42, 1083–1087. doi: 10.21273/hortsci.42.5.1083
- Sullivan, G. A., and Calkins, C. R. (2010). Application of exogenous enzymes to beef muscle of high and low-connective tissue. *Meat Sci.* 85, 730–734. doi: 10.1016/j.meatsci.2010.03.033
- Than, M. E., Helm, M., Simpson, D. J., Lottspeich, F., Huber, R., and Gietl, C. (2004). The 2.0 Å crystal structure and substrate specificity of the KDE1-tailed cysteine endopeptidase functioning in programmed cell death of *Ricinus communis* endosperm. *J. Mol. Biol.* 336, 1103–1116. doi: 10.1016/j.jmb.2003.12.075
- Usai, G., Mascagni, F., Giordani, T., Vangelisti, A., Bosi, E., Zuccolo, A., et al. (2020). Epigenetic patterns within the haplotype phased fig (*Ficus carica* L.) genome. *Plant J.* 102, 600–614. doi: 10.1111/tpj.14635
- Vatansever, R., Koc, I., Ozyigit, I. I., Sen, U., Uras, M. E., Anjum, N. A., et al. (2016). Genome-wide identification and expression analysis of sulfate transporter (SULTR) genes in potato (*Solanum tuberosum* L.). *Planta* 244, 1167–1183. doi: 10.1007/s00425-016-2575-6
- Wang, D., Zhang, Y., Zhang, Z., Zhu, J., and Yu, J. (2010). KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genom. Proteom. Bioinform.* 8, 77–80. doi: 10.1016/S1672-0229(10)60008-3
- Wang, W., Zhang, L., Guo, N., Zhang, X., Zhang, C., Sun, G., et al. (2014). Functional properties of a cysteine proteinase from pineapple fruit with improved resistance to fungal pathogens in *Arabidopsis thaliana*. *Molecules* 19, 2374–2389. doi: 10.3390/molecules19022374
- Wang, Y., Tang, H., DeBarry, J. D., Tan, X., Li, J., Wang, X., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40:e49. doi: 10.1093/nar/gkr1293
- Wang, Z., Cui, Y., Vainstein, A., Chen, S., and Ma, H. (2017). Regulation of fig (*Ficus carica* L.) fruit color: metabolomic and transcriptomic analyses of the flavonoid biosynthetic pathway. *Front. Plant Sci.* 8:1990. doi: 10.3389/fpls.2017.01990
- Wang, Z., Song, M., Li, Y., Chen, S., and Ma, H. (2019). Differential color development and response to light deprivation of fig (*Ficus carica* L.) syconia peel and female flower tissues: transcriptome elucidation. *BMC Plant Biol.* 19:217. doi: 10.1186/s12870-019-1816-9
- Xie, C., Wang, D., and Yang, X. (2009). Protein extraction methods compatible with proteomic analysis for the cotton seedling. *Crop Sci.* 49, 395–402. doi: 10.2135/cropsci2008.06.0367
- Zare, H., Moosavi-Movahedi, A. A., Salami, M., Mirzaei, M., Saboury, A. A., and Sheibani, N. (2013). Purification and autolysis of the ficin isoforms from fig (*Ficus carica* cv. Sabz) latex. *Phytochemistry* 87, 16–22. doi: 10.1016/j.phytochem.2012.12.006
- Zhalehjo, M., and Mostafaie, A. (2012). The application of fig proteases (ficin) as a digestive drug. *Res. Pharmaceut. Sci.* 7:610.
- Zhang, S., Xu, Z., Sun, H., Sun, L., Shaban, M., Yang, X., et al. (2019). Genome-wide identification of papain-like cysteine proteases in *Gossypium hirsutum* and functional characterization in response to *Verticillium dahliae*. *Front. Plant Sci.* 10:134. doi: 10.3389/fpls.2019.00134
- Zhang, X., Wang, G., Zhang, S., Chen, S., Wang, Y., Wen, P., et al. (2020). Genomes of the banyan tree and pollinator wasp provide insights into fig-wasp coevolution. *Cell* 183, 875–889. doi: 10.1016/j.cell.2020.09.043
- Zhu, J., and Zhang, Z. (2009). Ethylene stimulation of latex production in *Hevea brasiliensis*. *Plant Signal. Behav.* 4, 1072–1074. doi: 10.4161/psb.4.11.9738
- Zou, Z., Huang, Q., Xie, G., and Yang, L. (2018). Genome-wide comparative analysis of papain-like cysteine protease family genes in castor bean and physic nut. *Sci. Rep.* 8:331. doi: 10.1038/s41598-017-18760-6
- Zou, Z., Xie, G., and Yang, L. (2017). Papain-like cysteine protease encoding genes in rubber (*Hevea brasiliensis*): comparative genomics, phylogenetic, and transcriptional profiling analysis. *Planta* 246, 999–1018. doi: 10.1007/s00425-017-2739-z

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Zhai, Cui, Song, Vainstein, Chen and Ma. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.