Identification and Bioinformatics Analysis on Phosphatidylcholine Diglyceride Choline Phosphotransferase Family Genes in Plants

Xiaoru Ran, Jun Hong, Fazhe Yan, Jianxin Shi ✉

Joint International Research Laboratory of Metabolic & Developmental Sciences, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, 200240, China

✉ Corresponding author email: jianxin.shi@sjtu.edu.cn

Abstract Phosphatidylcholine diglyceride choline phosphotransferase (PDCT) is involved in the synthesis of seed triglyceride (TAG) in plant, catalyzing the transfer between phosphatidylcholine (PC) and diacylglycerol (DAG), and the loss of the PDCT activity reduces significantly the accumulation of seed polyunsaturated fatty acids (PUFAs). In this study, taking advantage of public available genomes of six oil crops (peanut, rape, soybean, cotton, sesame and sunflower), two staple food crops (rice and maize) and one aquatic plant (Chinese lotus), we identified in total 20 *PDCT* orthologues using bioinformatics, and further investigated into physical and protein chemical property, chromosome distribution, phylogenetic evolution, gene structure and promoter sequence. Our results showed that the *PDCT* family members have conserved domains and gene structures, as well as stable protein properties and structures, implying certain evolutionary conservation among them. In addition, copy numbers of *PDCT* were significantly different in tested plant species, while crop plants and Arabidopsis had one single copy, oil crops almost have at least 2 copies. Furthermore, the existence of many transcription core promoter elements and photo-response cis-acting elements in the promoter regions of these identified *PDCT* genes indicated that the expression of *PDCT* genes is likely affected by multiple developmental and environmental stress signals. In sum, this study provided a bioinformatics reference for further study on the potential application of *PDCT* family genes in crop breeding.

Keywords PDCT; Protein structure; Evolutionary analysis; Promoter analysis

Plant seed oil plays an important role in human's daily life. It is not only an important source of dietary fat, but also an important source of biofuels and other industrial applications (Cahoon et al., 2007; Lu et al., 2011). The function, properties and suitability of different plant seed oils for specific applications mainly depend on the composition and content of their fatty acids (Wickramarathna et al., 2015). According to the number of double bonds, the unsaturated fatty acids in plant seed oil can be divided into monounsaturated oleic acid (18:1; OA) and polyunsaturated fatty acids (PUFA). Studies have shown that increasing the ratio of OA (18:1) in plant seed oil can improve oxidation stability and has significant health benefits for human body, which is also very important for the production of biodiesel and other renewable resources (Kinney, 1996). At the same time, high PUFA is a viable nutritional option in human diet plans (Hunter, 1990; Williams and Burdge, 2006). For this reason, as the main source of essential fatty acids in the diet, the synthesis of triacylglycerol in plants and the study of related enzymology have been the field of concern to scientists.

During the development of oil crop seeds, most of the monounsaturated oleic acid (18:1) formed in plasts enters phosphatidylcholine (PC) (Shanklin and Cahoon, 1998, Bates et al., 2007). Then, through the desaturation of fatty acid desaturase 2 (FAD2) and FAD3, it forms linoleic acid (18:2) and linolenic acid (18:3) (Browse et al., 1993, Wallis et al., 2002). Since then, PC has been incorporated into TAG, becoming a common form of fat storage in many plant seeds and a major source of essential fatty acids in human dietary nutrition choices (Lu et al., 2009).

Previous studies have identified a key enzyme in seed oil formation, phosphatidylcholine diester choline phosphotransferase (PDCT), which catalyzed the exchange of choline phosphate heads during seed TAG synthesis, leading to the interconversion of PC and DAG (Lu et al., 2009). As a gate enzyme, PDCT provides an important way for 18:1 entering PC desaturation to become 18:2 and 18:3, and for 18:2 and 18:3 desaturation products to return to DAG. In Ninanjie (*Arabidopsis thaliana*), defects in PDCT reduced the content of polyunsaturated fatty acids (PUFA) in seeds by 40% (Lu et al., 2009). Experiments in Youcai (*Brassica napus*) also demonstrated the effect of the loss of PDCT activity on the reduction of PUFA accumulation in rapeseed oil. In addition, allogeneously expressing the *PDCT* gene of Yama (*Linum usitatissimum* L.) in yeast and *Arabidopsis thaliana* increased the PUFA level of the mutant (Wickramarathna et al., 2015). These results indicate that PDCT plays an important role in PUFA accumulation in plant seeds. However, little research has been done on PDCTS in other species to see if they have any other function other than the aforementioned role in PUFA accumulation in seeds. In this study, by means of bioinformatics, using six published oil crops-- Huasheng (*Arachis hypogea*), Ganlanxing Youcai (*Brassica napus*), Dadou (*Glycine max*), Ludimian (*Gossypium hirsutum*) and Zhima (*Sesamum indicum*), Xiangrikui (*Helianthus annuus*), and two food crops, Shuidao (*Oryza sativa* Japonica Group), Yumi (*Zea mays*), and an aquatic plant, Zhongguolian (*Nelumbo nucifera*), twenty homologous genes encoding PDCT were identified. Their protein physicochemical properties, gene structure and evolutionary relationship, promoter elements and so on were analyzed. The results of this study provide bioinformatics reference for further exploration of breeding applications of this family gene.

1 Results and Analysis

1.1 Identification of the PDCT family

The protein sequence of PDCT (AtROD1) reported in *Arabidopsis thaliana* was used as query, and blast was used to search the protein data of *Glycine max*, *Brassica napus*, *Helianthus annuus*, *Gossypium hirsutum*, *Arachis hypogea*, *Sesamum indicum*, *Oryza sativa* Japonica Group, *Zea mays* and *Nelumbo nucifera*. A total of 31 initial candidate members were obtained, and then 20 candidate PDCT families were obtained by using conservative domain analysis to remove redundancy. The NCBI database revealed 20 genes belonging to the PAP2_Like Superfamily, which is consistent with the *PDCT* gene found in Bima (*Ricinus communis* L.). In addition, the four *PDCT* homologous genes identified in *Brassica napus* were consistent with the reported identification results, which further confirmed the reliability of our results. According to their positions on chromosomes, we named 20 *PDCT* homologous genes successively in the form of "species +PDCT+ number". The nucleotide sequence of *PDCT* gene family was significantly different. The shorter ones were *NnPDCT1* (1 630 bp) and *NnPDCT2* (1 391 bp) of *Nelumbo nucifera*, and the longest was *BnPDCT4* (8 875 bp) of *Brassica napus* (Table 1).

1.2 Analysis of basic physicochemical properties of PDCT family proteins

We then characterized the basic physicochemical properties of 20 PDCT family proteins (Table 2). Although the length of *PDCT* gene varied greatly among the nine tested species, the protein sequences encoded by *PDCT* gene ranged from 224 aa to 307 aa. The molecular weight of 20 PDCT protein family members of nine species was 25.13~33.20 kDa, and the amino acids with positive charge were more than those with negative charge. The theoretical isoelectric points are all greater than 7 and alkaline, ranging from 8.15 to 10.06, indicating that the PDCT family proteins are rich in basic amino acids. The results of ProtScale protein analysis showed that the PDCT protein lipid solubility index of nine species was more than 80, and the size was between 89.19 and 107.01, showing high lipid solubility. The hydrophobic index was greater than 0, indicating that PDCT family proteins were hydrophobic. The value of instability coefficient is bounded by 40. If the protein is less than 40, it is stable; if the protein is more than 40, it is unstable. From this, we can see that in the nine species, 8 PDCT family proteins showed instability, and the remaining 12 proteins showed relatively stable.

Table 1 Basic information of PDCT gene family members in tested 9 plant species

Species	Gene	Gene locus	Accession number	Chrome	Starting point	Ending point	Gene length (bp)
Arachis hypogaea	*AhPDCT1*	LOC112731416	XM_025780857	Chrom02	1033677	1035241	1565
Brassica napus	*AhPDCT2*	LOC112709786	XM_025761757	Chrom09	1213630	1215987	2358
	AhPDCT3	LOC112775887	XM_025819783	Chrom19	1553651	1555665	2015
	BnPDCT1	LOC106439422	XM_013880863	ChromA3	25986445	25989201	2757
Glycine max	*BnPDCT2*	LOC106404645	XM_013845347	ChromA5	26976798	26979150	2353
	BnPDCT3	LOC106399461	XM_013839946	ChromC5	36716346	36718659	2314
	BnPDCT4	LOC106406541	XM_013847231	ChromC5	17487755	17496629	8875
	GmPDCT1	LOC100775686	XM_003528267	Chrom07	2380709	2382657	1949
Gossypium	*GmPDCT2*	LOC100790665	XM_003531670	Chrom08	17187643	17190146	2504
hirsutum	*GhPDCT1*	LOC107906462	XM_016833457	Chrom19	61265562	61267597	2036
Helianthus annuus	*GhPDCT2*	LOC107906463	XM_016833458	Chrom19	61268836	61270393	1558
	GhPDCT3	LOC107937381	XM_016870252	Scaffold1852.1	85439	87900	2462
	GhPDCT4	LOC107937381	XM_016870253	Scaffold1852.1	85439	88477	3039
	HaPDCT1	LOC110927280	XM_022170941	Chrom02	176500432	176506642	6211
Sesamum indicum	*HaPDCT2*	LOC110915673	XM_022160404	Chrom16	10384900	10388902	4003
	SiPDCT	LOC105158376	XM_011075120	Chrom03	14323919	14326719	2801
Nelumbo nucifera	*NnPDCT1*	LOC104592473	XM_010251858	Scaffold00007	7976880	7978239	1360
Oryza sativa	*NnPDCT2*	LOC104586112	XM_010243243	Scaffold00153	151983	153373	1391
Japonica Group	*OsPDCT*	Os06g0607100	NM_001064564	Chrom06	25016353	25019393	3041
Zea mays	*ZmPDCT*	LOC100278435	NM_001366023	Chrom09	94503705	94505536	1832

Table 2 Basic physiological and chemical features of PDCT proteins identified

Protein name	Number of amino acids (aa)	Molecular weight (k Da)	pI	Number of negatively charged residues (Asp+Glu)	Number of positively charged residues (Arg+Lys)	Total number of atoms	Instability index	Aliphatic index	GRAVY
AhPDCT1	295	32.32	9.26	19	26	4550	45.55	90.98	0.209
AhPDCT2	306	33.20	9.78	17	27	4688	44.96	93.17	0.193
AhPDCT3	304	33.09	9.66	18	26	4666	45.93	92.8	0.199
BnPDCT1	282	31.32	9.23	19	25	4391	39.22	90.21	0.199
BnPDCT2	283	31.22	8.81	21	24	4382	42.26	91.94	0.204
BnPDCT3	283	31.13	9.05	20	24	4367	42.94	90.92	0.195
BnPDCT4	280	31.06	9.34	18	25	4362	38.09	90.86	0.227
GmPDCT1	279	30.51	9.14	17	22	4287	39.63	92.69	0.266
GmPDCT2	267	28.98	9.4	15	21	4081	38.57	92.81	0.28
GhPDCT1	291	32.33	10.06	14	30	4574	34.94	91.72	0.252
GhPDCT2	224	25.13	9.69	12	20	3579	38.42	107.01	0.536
GhPDCT3	275	30.21	9.77	17	28	4268	33.63	89.71	0.221
GhPDCT4	271	29.78	9.72	17	27	4206	34.61	90.66	0.238
HaPDCT1	274	30.14	9.08	15	19	4253	29.92	93.21	0.323
HaPDCT2	280	31.29	8.47	18	20	4413	43.67	96.04	0.305
SiPDCT	294	32.19	10.01	17	30	4558	35.98	90.99	0.102
NnPDCT1	283	31.60	9.96	19	31	4469	36.95	92.33	0.166
NnPDCT2	280	30.87	9.95	19	29	4373	37.07	96.11	0.28
OsPDCT	307	31.49	9.04	19	25	4417	46.76	89.19	0.37
ZmPDCT	282	29.98	8.15	21	23	4214	44.85	93.23	0.326

1.3 Structure and phosphorylation site analysis of PDCT family proteins

We also analyzed the structure and phosphorylation sites of 20 PDCT family proteins (Table 3). The results of PDCT protein secondary structure prediction showed that they had similar characteristics in composition: The α-spiral and random coil structures accounted for 32.86%~52.3% and 33.48%~48.41%, respectively. Transmembrane analysis showed that PDCT protein had four transmembrane helical structures in all dicotyledonous plants. In the monocotyledonous plants rice and maize, the transmembrane structure of PDCT proteins was shown to be 3. Phosphorylation site analysis showed that the phosphorylation sites of PDCT family proteins were relatively small, ranging from 13 to 31, and mainly concentrated in Serine (Ser) and Threonine (Thr), while Tyrosine (Tyr) phosphorylation sites were less (Table 3). In addition, no signal peptides were detected in 20 PDCT family proteins, indicating that PDCT proteins were probably non-secreted proteins.

1.4 Functional domain analysis of PDCT family protein sequences

We also predicted the functional domains of 20 PDCT family proteins, and found that there were 15 conserved moieties with high similarity in the PDCT protein sequences (Figure 1A). Among the 15 conserved motifs, motifs 1-6 showed high conserved distribution. In general, although the species differed greatly, the number of conserved motifs varied little, ranging from 7 to 10 (Figure 1B). SMART online detection of conservative motifs found that motif 1 was SOCS_box domain, where motif 2 was PAP2_3 Domain.

1.5 Chromosome localization of *PDCT* gene in nine species

According to the genome sequence information of PDCT in nine species, its distribution on chromosomes was also mapped (Figure 2). The number of *PDCT* genes identified was small, and the distribution of PDCT genes on the chromosomes of each species was relatively uniform. In peanut, *AhPDCT* gene was distributed on chromosome 2 (Chrome02), chromosome 9 (Chrome09) and chromosome 19 (Chrome19), respectively. In *Brassica napus*, *BnPDCT1* and *BnPDCT2* were located on genome A, *BnPDCT1* was located on chromosome 3 (ChromeA3), *BnPDCT2* was located on chromosome 5 (ChromeA5). Both *BnPDCT3* and *BnPDCT4* were located on chromosome 5 of the C genome (ChromeC5), and more distributions were shown in Figure 2.

Table 3 Phosphorylation sites and structures of PDCT proteins

| Protein name	Alpha-helix (%)	Extended strand (%)	Random coil (%)	Number of predicted TMHs	Threonine	Serine	Tyrosine	Total sites
AhPDCT1	49.49	14.24	36.27	4	11	11	0	22
AhPDCT2	50.98	7.52	41.5	4	12	11	1	24
AhPDCT3	40.79	14.8	44.41	4	12	13	0	25
BnPDCT1	39.36	16.67	43.97	4	9	15	5	29
BnPDCT2	36.4	17.31	46.29	4	11	14	3	28
BnPDCT3	32.86	18.73	48.41	4	10	16	3	29
BnPDCT4	42.14	12.86	45	4	9	16	5	30
GmPDCT1	45.88	12.9	41.22	4	9	12	1	22
GmPDCT2	46.82	14.23	38.95	4	8	12	1	21
GhPDCT1	34.71	19.93	45.36	4	13	17	1	31
GhPDCT2	45.09	21.43	33.48	4	10	13	2	25
GhPDCT3	41.09	15.27	43.64	4	13	15	1	29
GhPDCT4	42.07	13.65	44.28	4	12	14	1	27
HaPDCT1	38.32	13.87	47.81	4	13	13	2	28
HaPDCT2	36.79	22.86	40.36	4	10	11	1	22
SiPDCT	41.84	11.9	46.26	4	12	10	3	25
NnPDCT1	52.3	11.66	36.04	4	14	13	0	27
NnPDCT2	46.79	12.86	40.36	4	10	11	1	22
OsPDCT	45.93	9.77	44.3	3	4	6	0	10
ZmPDCT	48.23	7.09	44.68	3	8	5	0	13
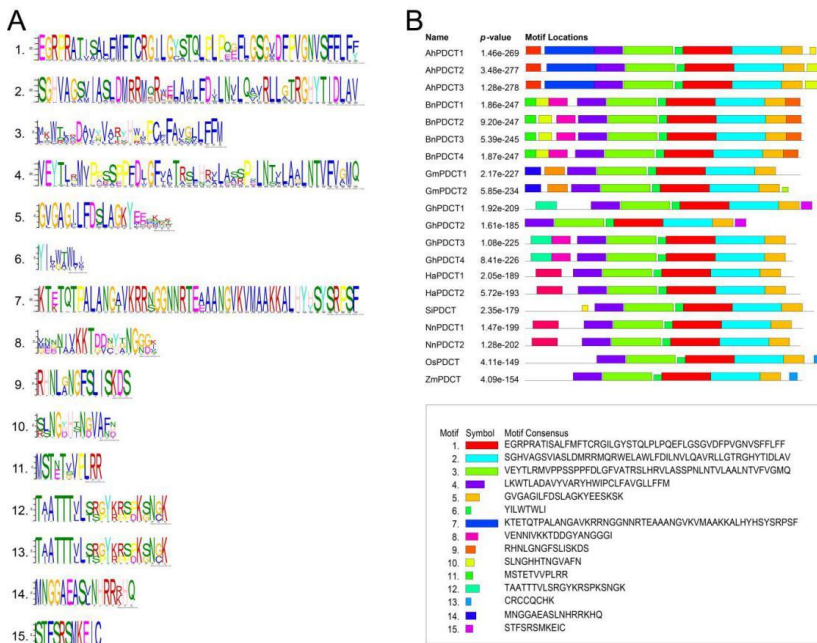
Figure 1 Prediction of conserved motifs in PDCT family proteins

Note: A: Highly conserved motif patterns of PDCT gene families; B: The distribution and sequence of motifs
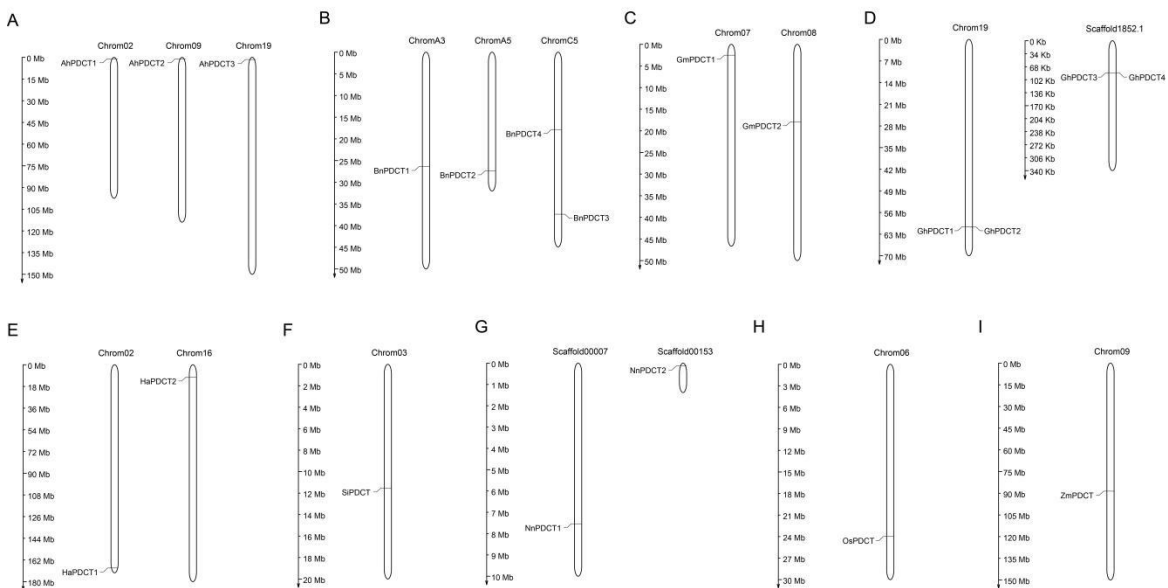

Figure 2 Chromosomal distribution of *PDCT* genes

Note: A: *Arachis hypogaea*; B: *Brassica napus;* C: *Glycine max*; D: *Gossypium hirsutum*; E: *Helianthus annuus;* F: *Sesamum indicum*; G: *Nelumbo nucifera*; H: *Oryza sativa Japonica Group*; I: *Zea mays*

1.6 Evolutionary relationship and gene structure analysis of *PDCT* gene family in nine species

In order to explore the evolution of the *PDCT* gene family, we constructed an evolutionary relationship tree based on the PDCT protein sequences in the nine species identified in this study and the previously reported PDCT protein sequences in *Arabidopsis thaliana*, *Linum usitatissimum* L. and *Ricinus communis* L.. The results showed that the *PDCT* genes in the same species were relatively close in evolutionary distance. Further analysis showed that the evolution of *AhPDCT2* and *AhPDCT3* subtypes in peanut was relatively close to that of *AtPDCT1*. In *Gossypium hirsutum* Linn., *GhPDCT1* and *GhPDCT2*, *GhPDCT3* and *GhPDCT4* were the closest evolutionary distances, respectively. In *Brassica napus*, *BnPDCT1* and *BnPDCT4*, *BnPDCT2* and *BnPDCT3* were the closest evolutionary distances, respectively. In addition, the evolutionary relationship also shows that compared with

other dicotyledons, sesame and sunflower are closer to monocotyledon rice and corn in evolutionary distance (Figure 3).

In order to further understand the evolutionary relationship of the *PDCT* gene family, we analyzed the structure of 20 *PDCT* genes identified in this study. The results showed that the genetic structure of the same plant had higher similarity in composition, which was consistent with the results of evolutionary analysis. The *PDCT* genes of dicotyledonous land plants, such as *Brassica napus* and *Arachis hypogaea*, mostly contain 3 exons and 2 introns, while monocotyledonous rice and maize and aquatic lotus contain 2 exons and 1 intron. The intron phase of *PDCT* genes in all nine species was 0, indicating a high degree of conservation (Figure 4).
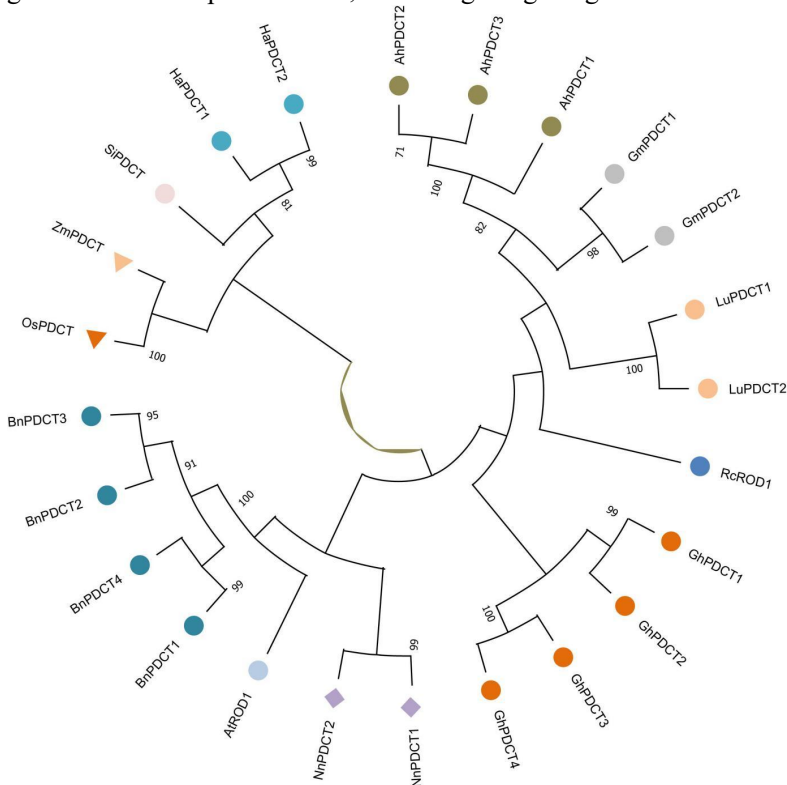

Figure 3 Phylogenetic analysis of *PDCT* genes

Note: *hypogaea*: Kelly dots; *Brassica napus:* Dark cyan dots; *Glycine max*: Grey dots; *Gossypium hirsutum*: Brown dots; *Helianthus annuus*: Light cyan dots; *Sesamum indicum*: Pink dots; *Ricinus communis*: Dark blue dots; *Arabidopsis thaliana*: Light blue dot; *Linum usitatissimum*: Orange dots; *Nelumbo nucifera*: Purple square; *Oryza sativa Japonica Group*: Brown triangle; *Zea mays*: Orange triangle
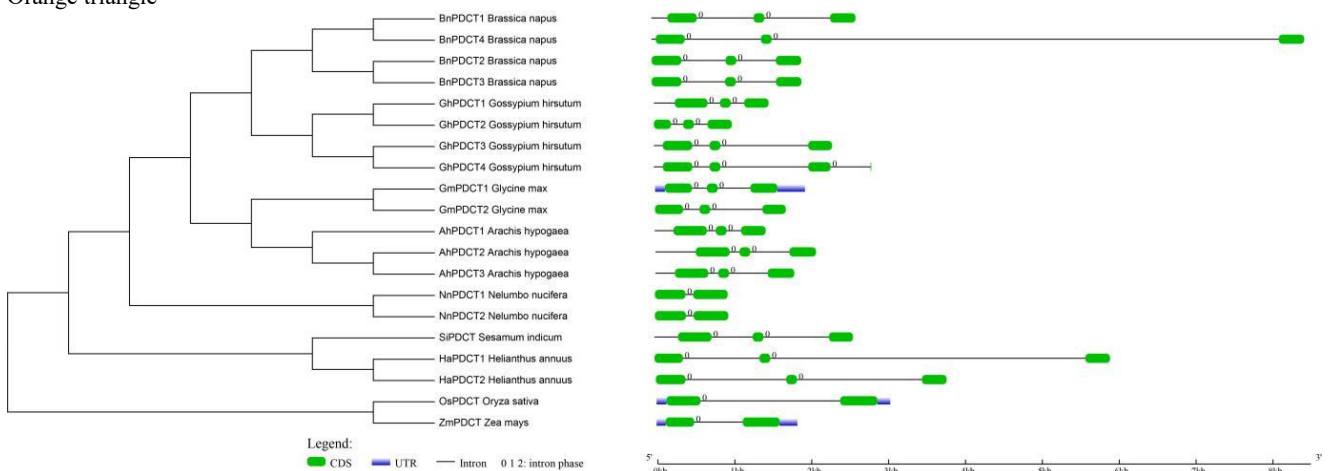

Figure 4 Gene structure analysis of *PDCT*

Note: The green and purple bar represent the CDS and the un-translated region (UTR), respectively, while the black line represents the intron; Numbers (0, 1, 2) stand for the intron phase

1.7 Analysis of cis-acting elements of *PDCT* gene promoter

Finally, we conducted online analysis of promoter sequences of 20 *PDCT* genes to search for possible cis-acting elements. The results showed that the promoters of 20 genes contained 2 836 cis-acting elements, excluding the elements of unknown function. Among them, TATA-box (974) and CAAT-box (720), which were the core promoter elements of transcription, accounted for the largest proportion, followed by stress response elements MYB (112) and MYC (73). In addition, we found 206 cis-acting elements associated with light response, including 14 species of G-box, Box 4, GT1-motif distributed in all identified *PDCT* genes. The number of elements related to light response of each gene ranged from 4 to 9, among which *AhPDCT1* had the largest number (Figure 5).
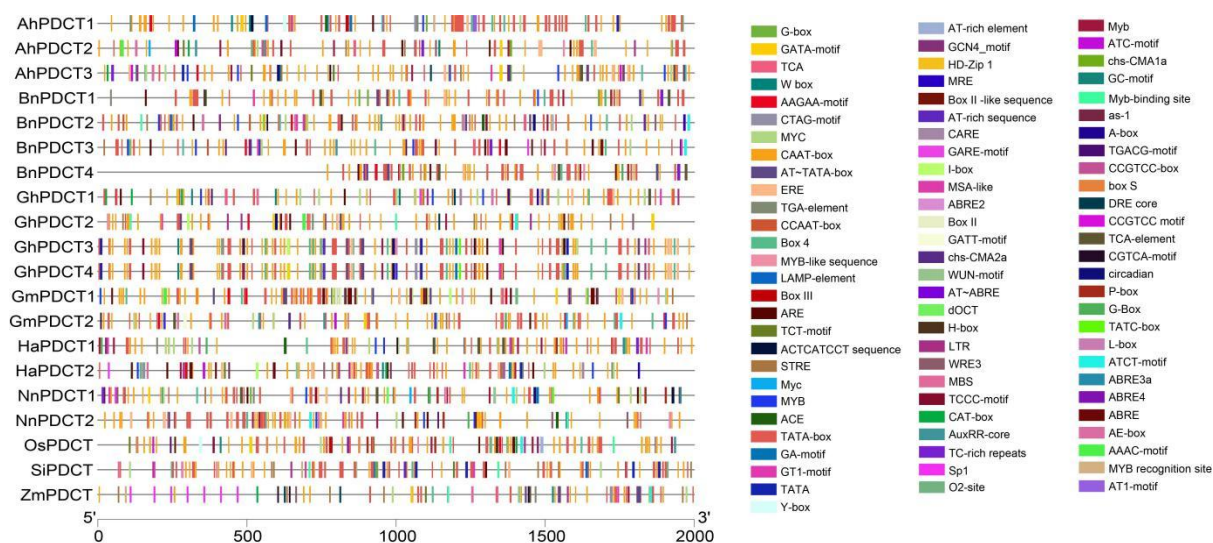

Figure 5 Cis-acting elements structures in the promoter region of PDCT family members

2 Discussion

Unsaturated fatty acids, especially PUFA, play an important role in human healthy diet. Therefore, controlling the synthesis of unsaturated fatty acids to improve the quality of edible oil in crop breeding has become a new direction for oil crop improvement in the future. *PDCT* has been shown to be a key enzyme for PUFA synthesis in plant seeds during TAG accumulation. For example, defects of *PDCT* in *Arabidopsis thaliana* seeds lead to a 40% reduction in PUFA content in mutant seeds (Lu et al., 2009). Studies on *Linum usitatissimum* and *Brassica napus* also confirmed the important role of *PDCT* in unsaturated fatty acid synthesis of oilseed plants (Wickramarathna et al., 2015). At present, *PDCT* gene identification and preliminary functional studies are limited to a few model plants, and the presence and function of *PDCT* in other plants have not been analyzed. Therefore, the identification and functional studies of *PDCT* in plants, especially in oil crops, need to be further strengthened. In recent years, the continuous improvement of species genome information has made it possible to identify and analyze various gene families including the PDCT family in the whole genome. In this study, a total of 20 *PDCT* genes were identified in 9 different plants including *Arachis hypogaea*, *Brassica napus*, *Glycine max*, *Gossypium*, *Sesamum indicum*, *Helianthus annuus*, *Nelumbo nucifera*, *Oryza sativa* and *Zea mays* by bioinformatics methods. The four *PDCT* genes identified in *Brassica napus* were consistent with the published results. The results of this study provide important biological information data for further study of gene function of *PDCT* family and possible breeding applications.

Our results show that members of the PDCT family have a certain degree of conservation in evolution, which is mainly manifested as: (1) The stability of the properties and structure of the PDCT proteins: the 20 identified proteins are not only similar in the number of amino acids, but also about 30 kDa in molecular weight, showing basic characteristics. The positive charge residue was more than the negative charge residue, and all showed hydrophobic proteins. In addition, PDCT family proteins were similar in the proportion of secondary structure, with α helices and random curls occupying the majority, and threonine and serine dominating the phosphorylation

site analysis. (2) Conserved gene structure: Despite the large differences in sequence length, PDCT family genes are relatively similar in structure. Dicotyledonous land plants all have 3 exons and 2 introns, while aquatic plants lotus and monocotyledonous plants rice and maize all have 2 exons and 1 and intron, and the intron phase is highly conserved with 0. (3) Conservation of evolutionary distance: The phylogenetic tree results were consistent with the intraspecific relationships of PDCT family proteins. (4) Conserved domain: PAP2_3 domain exists in both monodicotyledonous plants and terrestrial aquatic plants, and is located near the C-terminal of the protein.

Our results also showed that there were significant differences in the copy number of *PDCT* family genes in the nine plants tested, but all of them were within 4. The *PDCT* in food crops (*Oryza sativa* and *Zea mays*) and model plants (*Arabidopsis thaliana*) were single-copy, while the *PDCT* in oil crops (*Arachis hypogaea*, *Brassica napus*, *Glycine max*, *Gossypium*, and *Helianthus annuus*) were multi-copy (2 or more). Interestingly, a single copy of *PDCT* was found in sesame seeds, but this result deserves further clarification. The results of *PDCT* gene copy number found in this study were consistent with those previously reported in *Arabidopsis thaliana* (Lu et al., 2009), *Linum usitatissimum* (Wickramarathna et al., 2015) and *Brassica napus*.

Phylogenetic tree analysis showed that these single-copy species are relatively close in evolution, and the PDCT of *Sesamum indicum* is relatively distant from that of other typical oil crops, such as *Ricinus communis*, *Linum usitatissimum*, *Arachis hypogaea* and *Glycine max*. These results suggest that PDCT may play a role in the evolution or differentiation of oil plants and food crops, and on the other hand, PDCT in food crops may have different enzyme activities or functions than PDCT in oil crops. Interestingly, comparing the evolutionary distance between PDCT of the only aquatic crop, *Nelumbo nucifera*, and that of other terrestrial crops, PDCT may also be related to the differentiation of terrestrial and aquatic plants. Such information provides valuable reference for the further study of *PDCT* gene function and evolution.

Our results also showed that there are a large number of stress response elements and light response elements in the promoters of *PDCT* family genes, suggesting that the expression of PDCT may be influenced by many growth and environmental stress signals, especially light response signals. Therefore, it is reasonable to believe that *PDCT* may also play an important role in plant growth and environmental stress, especially in light response.

3 Materials and Methods

3.1 Data download and verification

In this study, proteome sequences, genome sequences and gene annotation files of *Arachis hypogea* (V3), *Brassica napus* (V2), *Glycine max* (V2.1), *Gossypium hirsutum* (V1), *Helianthus annuus* (V1), *Sesamum indicum* (V1), *Nelumbo nucifera* (V1.1), *Oryza sativa* Japonica Group (V4) and *Zea mays* (V4) were downloaded from NCBI (National Center of Biotechnology Information) database (https://ftp.ncbi.nlm.nih.gov/genomes/), respectively.

The *Arabidopsis thaliana* PDCT (AtROD1) protein sequence reported in the literature was used as query and Blast 2.5.0 software (Camacho et al., 2009) was used to search the PDCT gene homologous proteins of nine species (E-value=1e^{-10}). After manual de duplication of screening results, the initial protein of PDCT family was obtained, and then the initial protein sequence was uploaded to CDD online website (http://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi) and SMART website (http://smart.embl-heidelberg.de/) for preliminary functional domain detection and analysis. After manual de redundancy, candidate members of PDCT family were obtained for subsequent analysis.

3.2 Analysis of PDCT Family Proteins

The candidate proteins obtained by the above steps were analyzed, and the molecular weight and theoretical isoelectric point of the obtained 20 PDCT family proteins were predicted using Compute pI/MW (https://web.expasy.org/compute_pi/) from ExPASy. Online software ProtParam (https://web.expasy.org/protparam/) predicted first-order physicochemical properties such as charge residual base, total atomic number, instability coefficient, fat solubility index and hydrophobic index. PHD (https://npsaprabi.ibcp.fr/cgibin/npsa_automat.pl?page=/NPSA/npsa_phd.hml) method was used to predict the

secondary structure of proteins in the PDCT family. SignalP-5.0 (http://www.cbs.dtu.dk/services/SignalP/) realized the detection of protein signal peptide. TMHMM website (http://www.cbs.dtu.dk/services/) achieved the analysis of transmembrane structure. NetPhos 3.1 Server (www.cbs.dtu.dk/services/NetPhos/) provided phosphorylation site detection. Finally, MEME (http://meme-suite.org/) was used to predict and analyze the conservative motif of PDCT protein sequences of nine species.

3.3 Gene structure, chromosome localization, evolutionary relationship and promoter analysis of PDCT
The location information of PDCT genes in nine species was mapped on chromosomes by TBtools software. MEGA-X software (Kumar et al., 2018) realized the construction of phylogenetic tree: After multi-sequence alignment of PDCT protein sequences by ClustalW (Gap Opening Penalty=10, Gap Extension Penalty=0.2, Delay Divergent Cutoff=30%), the evolutionary tree was obtained by maximum likelihood method (Bootstrap=1000, Partial deletion).

The genetic structure of PDCT was mapped using GSDS (http://gsds.cbi.pku.edu.cn/) online software. PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) implemented the analysis of promoter cis-acting elements. And using TBtools software to draw the structure diagram.

Authors' contributions
RXR was the performer of the experimental research and the writer of the paper draft. HJ and YFZ participated in data collation, results discussion and paper draft writing. SJX was the project leader, experimental designer and modifier of the paper. All authors read and approved the final manuscript.

References
Bates P.D., Ohlrogge J.B., and Pollard M., 2007, Incorporation of newly synthesized fatty acids into cytosolic glycerolipids in pea leaves occurs via acyl editing, J. Biol. Chem., 282(43): 31206-31216
https://doi.org/10.1074/jbc.M705447200
PMid:17728247

Browse J., McConn M., James D., and Miquel M., 1993, Mutants of Arabidopsis deficient in the synthesis of alpha-linolenate. Biochemical and genetic characterization of the endoplasmic reticulum linoleoyl desaturase, J. Biol. Chem., 268(22): 16345-16351
https://doi.org/10.1016/S0021-9258(19)85427-3
PMid:8102138

Cahoon E.B., Shockey J.M., Dietrich C.R., Gidda S.K., Mullen R.T., and Dyer J.M., 2007, Engineering oilseeds for sustainable production of industrial and nutritional feedstocks: solving bottlenecks in fatty acid flux." Curr. Opin. Plant Biol., 10(3): 236-244
https://doi.org/10.1016/j.pbi.2007.04.005
PMid:17434788

Camacho C., Coulouris G., AvagyanV., Ma N., Papadopoulos J., Bealer K., and Madden T.L., 2009, BLAST+: architecture and applications, BMC Bioinformatics, 10: 421
https://doi.org/10.1186/1471-2105-10-421
PMid:20003500 PMCid:PMC2803857

Hunter J.E., 1990, n-3 fatty acids from vegetable oils, Am. J. Clin. Nutr., 51(5): 809-814.
https://doi.org/10.1093/ajcn/51.5.809
PMid:1970702

Kinney A.J., 1996, Designer oils for better nutrition, Nat. Biotechnol., 14(8): 946.
https://doi.org/10.1038/nbt0896-946
PMid:9631031

Kumar S., Stecher G., Li M., Knyaz C., and Tamura K., 2018, MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms, Mol. Biol. Evol., 35(6): 1547-1549
https://doi.org/10.1093/molbev/msy096
PMid:29722887 PMCid:PMC5967553

Lu C.F., Napier J.A., Clemente T.E., and Cahoon E.B., 2011, New frontiers in oilseed biotechnology: meeting the global demand for vegetable oils for food, feed, biofuel, and industrial applications, Curr. Opin. Biotechnol., 22(2): 252-259
https://doi.org/10.1016/j.copbio.2010.11.006
PMid:21144729

Lu C., Xin Z., Ren Z., Miquel M., and Browse J., 2009, An enzyme regulating triacylglycerol composition is encoded by the ROD1 gene of Arabidopsis, Proc Natl Acad Sci USA, 106(44): 18837-18842
https://doi.org/10.1073/pnas.0908848106
PMid:19833868 PMCid:PMC2774007

Shanklin J. and Cahoon E.B., 1998, Desaturation and Related Modifications of Fatty Acids1, Annu Rev. Plant Physiol. Plant Mol. Biol., 49: 611-641
https://doi.org/10.1146/annurev.arplant.49.1.611
PMid:15012248

Wallis J.G., Watts J.L., and Browse J., 2002, Polyunsaturated fatty acid synthesis: what will they think of next? Trends Biochem. Sci., 27(9): 467
https://doi.org/10.1016/S0968-0004(02)02168-0
PMid:12217522

Wickramarathna A.D., Siloto R.M., Mietkiewska E., Singer S.D., Pan X., and Weselake R.J., 2015, Heterologous expression of flax PHOSPHOLIPID:DIACYLGLYCEROL CHOLINEPHOSPHOTRANSFERASE (PDCT) increases polyunsaturated fatty acid content in yeast and Arabidopsis seeds, BMC Biotechnol., 15: 63
https://doi.org/10.1186/s12896-015-0156-6
PMid:26123542 PMCid:PMC4486708

Williams C.M. and Burdge G., 2006, Long-chain n-3 PUFA: plant v. marine sources, Proc. Nutr. Sci. 65(1): 42-50
https://doi.org/10.1079/PNS2005473
PMid:16441943