Arch. Biol. Sci., Belgrade, 65 (3), 863-876, 2013 DOI:10.2298/ABS1303863H

EXPRESSION, DIVERGENCE AND EVOLUTION OF THE CALEOSIN GENE FAMILY IN *BRASSICA RAPA*

LIZONG HU, SHUFEN LI and WUJUN GAO

College of Life Sciences, Henan Normal University, Xinxiang, Henan, China

Abstract - Caleosins (CLO) are oil body-associated proteins encoded by a small gene family. To investigate the expression, functional diversity and evolutionary modes of *CLO* genes, we isolated and integrally analyzed *in silico* a total of 11 *CLO* genes from *Brassica rapa*. According to phylogeny and sequence analyses, 11 *BrCLO* genes were classified into 3 groups, and each group shared highly conserved sequence features. Syntenic analysis revealed that all members of the *BrCLO* gene family distributed on 7 chromosomes were expanded mainly by segmental duplications. Evolutionary analysis showed that CLO proteins were controlled by purifying selection in *B. rapa*. Interestingly, functional divergence studies indicated that site-specific relaxed functional constraints were present between the different clusters of caleosins. Expression pattern suggested that 6 *BrCLO* genes were potentially associated with oil body formation. Our findings provide valuable clues for an investigation of the evolutionary history and cellular functions of the *CLO* gene family in plants.

Key words: Brassica rapa, caleosin, functional divergence, evolution, expression

INTRODUCTION

Caleosin, which plays key roles in the oil-body formation, stability and integration, is generally localized on the surface of oil bodies (Murphy, 1993; Naested et al., 2000; Tzen, 2012). The typical structure features of the CLO proteins are the presence of the proline-knot motif and three conserved domains, which are the hydrophilic calcium-binding domain near the N-terminal end, the central hydrophobic oil body anchoring domain and the hydrophilic phosphorylation domain near the C-terminal end, respectively (Chen et al., 1999). In plants, *OsEFA27* was the first identified caleosin from *Oryza sativa* (Frandsen et al., 1996). Subsequently, many homologous caleosin isoforms were also identified from other plant species such as *Sesamum indicum* (Chen et al., 1999), *Brassica napus* (Hernandez-Pinzon et al., 2001), *Hordeum vulgare* (Liu et al., 2005), *Lilium longiflorum* Thunb. (Jiang et al., 2007), *Cycas revolute* (Jiang et al., 2009), *Olea europaea* (Zienkiewicz et al., 2010) and *Chlorella* sp. (Lin et al., 2012). No homologous *CLO* gene was observed in animals. Therefore, *CLO* genes could be widely distributed in true fungi, unicellular microalgae, and higher plants (Tzen et al., 1993; Naested et al., 2000; Purkrtova et al., 2007; Partridge and Murphy 2009; Lin et al., 2012; Tzen, 2012).

Recent studies revealed that caleosins were encoded by multiple genes which constitute a small gene family in plant species (Partridge and Murphy, 2009; Wei et al., 2011). In *A. thaliana*, microarray and EST data revealed that seven caleosin genes contained six active *CLO* genes and one caleosin-like gene pseudogene (Gierke et al., 2000). *AtCLO1* and *AtCLO2* were mainly expressed in developing seeds. *AtCLO3* (RD20) was responsive to a range of environmental stresses, especially in leaves and roots

(Aubert et al., 2010). The caleosin isoforms *AtCLO4* and *AtCLO5* displayed low levels of expression in non-stressed vegetative tissues (Gierke et al., 2000; Partridge and Murphy 2009). In *O. sativa*, the 6 identified *CLO* genes were named *OsCLO-1~6*, and they were divided into two groups that originated before the split of gymnosperms and angiosperms. Expression analysis revealed that rice *CLO* genes displayed distinct expression patterns in sampled tissues. *OsCLO-2*, *OsCLO-3* (*OsEFA27*) and *OsCLO-6* were drought-inducible genes, but *OsCLO-1*, *OsCLO-4* and *OsCLO-5* were not induced by drought stress (Wei et al., 2011). In addition, the localization analysis of CLO proteins demonstrated that caleosin isoforms map to at least two subcellular compartments (Naested et al., 2000; Liu et al., 2005; Purkrtova et al., 2007). Overall, the results above implied that caleosins could be directly or indirectly involved in a variety of biological processes, such as oil-body synthesis and stability, lipid trafficking, signal transduction, seed germination, plant-pathogen recognition, symptom development and abiotic stress responses (Naested et al., 2000; Poxleitner et al., 2006; Feng et al., 2011; Tzen, 2012).

Brassica rapa (2n=2x=20, AA genome) is a relatively simple diploid species from the Cruciferae family. It is a good genetic material that allows us to investigate duplicated gene fate, gene origin and expansion, gene dosage effects, and gene rearrangement after paleopolyploidizations (Mun et al., 2009; Cheng et al., 2011). Microarray-based transcriptome studies in *B. rapa* provided evidences for the expression profiling of some caleosin genes because their corresponding EST or cDNA could be identified from multiple tissues (Lee et al., 2008). However, the copy number, sequence features and evolutionary history of *B. rapa CLO* genes were largely unclear. In this study, a total of 11 *CLO* genes were identified from the entire *B. rapa* genome, and their sequence features were investigated in detail. Subsequently, we analyzed the phylogenetic relationship, functional divergence, and evolutionary dynamics of *CLO* genes in all sampled species. In addition, we examined the tissue expression patterns of 9 *CLO* genes in *B. rapa*. Our findings may contribute to better understanding of the functional diversity of this family of proteins and selective pressure upon *CLO* genes in *B. rapa* and other plants.

MATERIALS AND METHODS

Species samples and data retrieval

Although *B. rapa* was selected as the targeted species in this study, another five model species, including *O. sativa*, *A. thaliana*, *Physcomitrella patens*, *Selaginella moellendorffii* and *Chlamydomonas reinhardtii*, were also selected as sampled species. First, previously characterized *CLO* genes were collected from numerous species (Hernandez-Pinzon et al., 2001; Partridge and Murphy, 2009; Wei et al., 2011), and were used as query genes to search for all possible *B. rapa CLO* genes using BlastP (E < 0.1) from the BRAD database (*B. rapa* ssp. *pekinensis* cv. Chiifu genome V1.0, http://*Brassicadb.org*). In addition, BlastP (E < 0.1) were also applied to retrieve all potential CLO proteins of the 6 sampled species from the phytozome database (http://www.phytozome.net/). Subsequently, the overlapping CLO family members were manually removed. Finally, Pfam (http://pfam.sanger.ac.uk/search) was used to screen these CLO proteins for the caleosin domain (PF05042) to confirm the accuracy of *CLO* genes.

Sequence features and phylogenetic analysis

Initially, we analyzed the sequence features of *CLO* genes in *B. rapa*. The GSDS (Guo et al., 2007), MEME (Bailey et al., 2006) and Pfam (Punta et al., 2012) were used to illustrate the gene structures, conserved motif organizations and domain architectures of *B. rapa CLO* genes, respectively. Predicted protein properties were further analyzed using the Sequence Manipulation Suite (www.bioinformatics.org/sms2/). Full-length sequences of CLO proteins in the 6 sampled species were aligned using Clustal X with defaulted parameters (Thompson et al., 1997). A phylogenetic tree was constructed using the neighbor-joining method (Saitou et al., 1987) with 100 bootstrap trials, and was further viewed by MEGA (Tamura et al., 2007).

Chromosomal mapping and expansion pattern analysis

The BRAD databases were applied for a BLAST-based search of the entire *B. rapa* genomic sequence to retrieve the exact physical locations of all *CLO* genes. Each of these *CLO* genes was manually visualized on a/the? *B. rapa* chromosome. With respect to expansion patterns, we focused mainly on segmental and tandem duplication between *B. rapa CLO* genes because it was difficult to identify the transposition events. According to the procedures described by Maher et al. (2006), we further analyzed the syntenic relationships between different *CLO* genes at the terminal nodes of the phylogenetic tree.

Adaptive evolution, functional divergence and mapping critical sites

To examine selection pressures of lineage-specific duplicated genes, their Ka/Ks ratios were calculated using a sliding window of 100 bp and a moving step of 10 bp (Nei and Gojobori, 1986), and JCoDA (Steinway et al., 2010) was used to display the distributions of the Ka/Ks values. Subsequently, we clustered all CLO proteins using BLASTCLUST (http://toolkit.tuebingen.mpg.de/blastclust/) with the lowest thresholds of coverage (90%) and 30% identity. After removing highly divergent sequences, the codon alignment of the remaining 22 CLO proteins was generated via PAL2NAL, and gaps were removed from the alignment (Suyama et al., 2006). Site-specific, free-ratio, and branch-site models were used to detect selective pressures on these CLO proteins (Yang and Nielsen, 2002; Yang et al., 2005; Zhang et al., 2005; Yang, 2007).

To investigate the functional divergence between different clusters, the coefficient of type I functional divergences was calculated using DIVERGE 2.0 (Gu, 1999). Subsequently, site-specific posterior probability analysis was performed to identify amino acid sites that resulted in the functional divergence. Moreover, the representative CLO proteins were aligned using Clustal X with default parameters. All the critical amino acid sites, including positive selection sites and functional divergence sites, were illustrated on the representative alignments.

RT-PCR analysis of 9 CLO genes in B. rapa

Seeds of *B. rapa* (Jietou 2) were grown in a greenhouse. Root, stem and leaf were sampled at the five-leaf stage. Flower and seed (40 DAP, days after pollination) were also sampled. Total RNA of different samples was extracted using the EASYspin RNA extraction kit (Aidlab Biotech, Beijing, China). After DNase I treatment, the SuperScript II Reverse Transcriptase Kit (Invitrogen Life Technologies, Carlsbad, CA, USA) was used to reversely transcribe total RNA into cDNA. Gene-specific primers were designated for semi-quantitative RT-PCR analysis of *CLO* genes in *B. rapa*. The *Actin-7* gene (GenBank, *JN120480.1*) served as an internal control. The PCR reaction was denatured at 95°C for 5 min, followed by 28 or 30 cycles at 95°C for 30 s, 56°C for 30 s, 72°C for 1 min, with a final extension of 10 min at 72°C. The PCR products of each sample were analyzed on 1% agarose gels and validated by sequencing. The experiment was repeated at least three times.

RESULTS

Identification and sequence features of CLO genes in B. rapa

Highly conserved caleosin domains facilitate the identification of all members of the *CLO* gene family. According to the protocol described above, a total of 11 *CLO* genes were extracted from the fully sequenced *B. rapa* genome. They were further named as *BrCLO1~11* based on the chromosomal order. The details of all *BrCLO* genes, including gene names, locus identifier, genome position and amino acid properties, were listed (Table 1). Gene structure analyses showed that the coding sequences of all the *BrCLO* genes were disrupted by introns, and all of *BrCLO1~9* had five introns except for *BrCLO10* and *BrCLO11* (Fig. 1A). The lengths of all the introns for the *BrCLO* genes were highly variable. However, *BrCLO1~9* all had the same length of the third, fourth, fifth exons, suggesting that these exons were highly

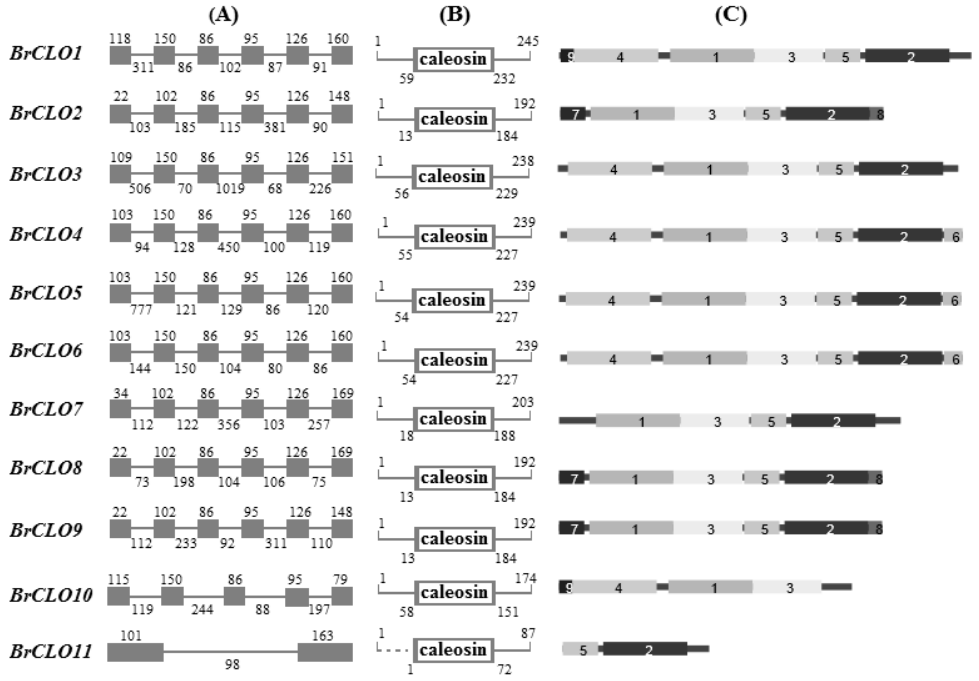

Fig. 1. The sequence features of the CLO gene family in *B. rapa*. In gene structures (A), intron, exon and sequence length are represented by lines, boxes and number, respectively. Three domains are highlighted by different boxes (B). Arabic numbers indicate different conserved motifs (C).

conserved. Interestingly, BrCLO10 and BrCLO11 were truncated and only contained part of the caleosin domain, but BrCLO1~9 all contained the entire caleosin domain (Fig. 1B). Moreover, BrCLO1~9 all shared conserved motif 1, motif 3, motif 5 and motif 2, and had a highly similar motif organization. Motif 4 was found in BrCLO1/3/4/5/6/10; motif 6 was located near the C-terminal end of BrCLO4/5/6; motif 7 and motif 8 co-occurred in BrCLO2/8/9; and motif 9 was observed in BrCLO1 and BrCLO10. The truncated BrCLO10 and BrClO11 had motifs 9/4/1/3 and motifs 5/2 respectively (Fig. 1C).

Phylogenetic relationships of CLO genes in plants

To shed light on the phylogenetic relationships of plant CLO family, we constructed a phylogenetic tree using the 39 CLO full-length amino acid sequences from yeast and six green plants (Fig. 2). Yeast CLO protein (UniProt, *J9NKK6*) was used as an outgroup, and all CLO proteins could be divided into three groups which were named as groups A, B, and C, respectively. In addition, three algal CLO proteins were not divided into any of the three groups. Group A only contained five CLO proteins from moss and lycophyte species. In contrast, Group B and Group C only contained CLO proteins from monocot and eudicot species. Eight sister pairs of paralogous *CLO* genes were found in 5 representative plants, implying that lineage-specific expansions occurred in this gene family.

Chromosomal localization and expansion patterns of CLO genes in B. rapa

The 11 *BrCLO* genes were localized on the 7 chromosomes of *B. rapa* by retrieving the physical positions of these genes from the BRAD database (Fig. 3). A maximum number of 3 genes including *BrCLO7/8/9* were present on chromosome A07; *BrCLO2/3* and *BrCLO10/11* were mapped on chromosome A02 and A10 respectively; and chromosome A01, A03, A04,

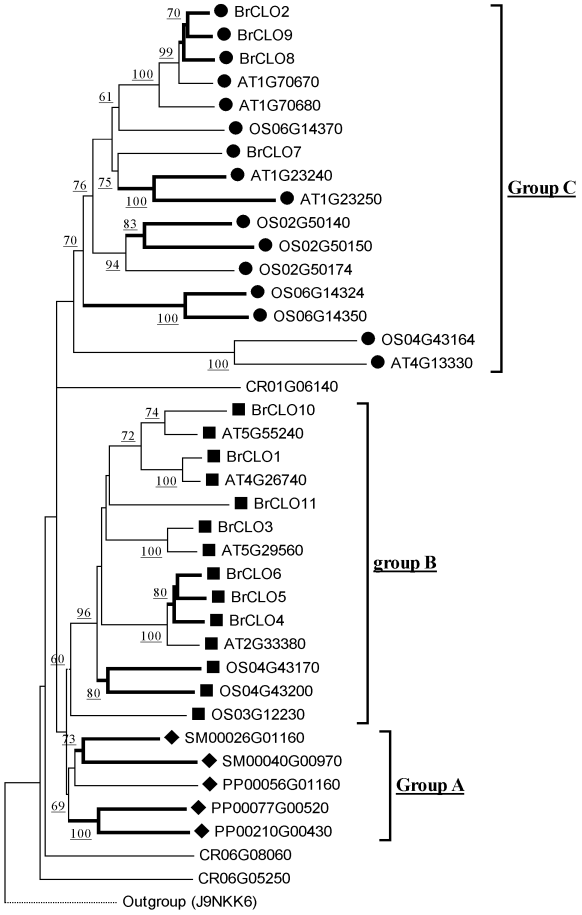

Fig. 2. The phylogenetic tree constructed using the CLO genes from plants. Genes in the same group were highlighted using identical symbols. Species-specific paralog pairs are highlighted by thick branches. The yeast CLO gene (UniProt, *J9NKK6*) was used as outgroup.

and A05 harbored *BrCLO1*, *BrCLO4*, *BrCLO5* and *BrCLO6*, respectively.

BrCLO10 and *BrCLO11* were tightly co-localized on the same regions, and they were separated by a 4.8 kb fragment. Therefore, they might be generated by the tandem duplication of the *BrCLO* gene. Following the phylogenetic relationships, we identified two paralog pairs in *B. rapa*, including *BrCLO2/8/9* and *BrCLO4/5/6*. Many conserved protein-coding genes flanking the paralogous *BrCLO* gene were observed, and this indicated that *BrCLO2/8/9* and *BrCLO4/5/6* had fine syntenic relationships. The result strongly

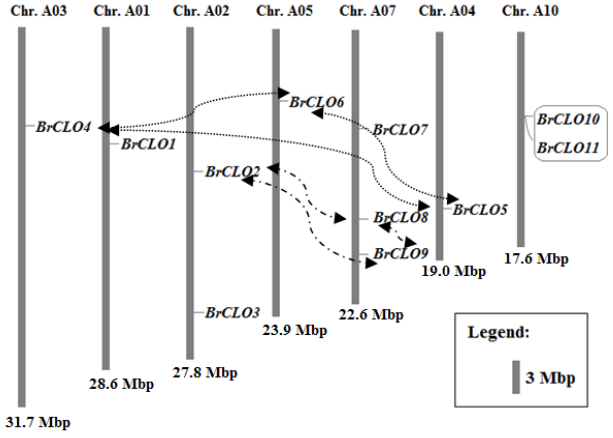

Fig. 3. Chromosomal mapping of the members of the CLO gene family in *B. rapa*. Paralog pairs are highlighted by the curved line with double arrow. Gene pairs of tandem duplication are included in square.

suggested that segmental duplication was responsible for the formation of *BrCLO2/8/9* and *BrCLO4/5/6* (Fig. 4).

Adaptive evolution of CLO genes in plants

We classified 8 caleosin paralogous gene pairs into four groups, representing four different types of plants. We further visualized the distributions of Ka/Ks values for each pair of the eight paralogous genes (Fig. 5). The results showed that all paralog pairs from lycophytes (Fig. 5A), mosses (Fig. 5B), eudicots (Fig. 5C) and monocots (Fig. 5D), respectively, had Ka/Ks<1, indicating that lineage-specific duplicated genes were mainly controlled by purifying selection. Subsequently, three pairs of nested site models, including M0 versus M3, M1a versus M2a, and M7 versus M8, were used for an investigation of selective pressures acting on CLO proteins. The results demonstrated that no positively selected sites were found using site models at the significant level ($P < 0.05$) (Table 2). Besides, the likelihood ratio test (LRT) of free-ratio and one-ratio models provided evidences for heterogeneous selection among different branches ($2\Delta l = 217.14$, $df = 40$, $P < 0.01$), and Ka, Ks and Ka/Ks values were visualized. Additionally, the free-ratio model identified 8 specific

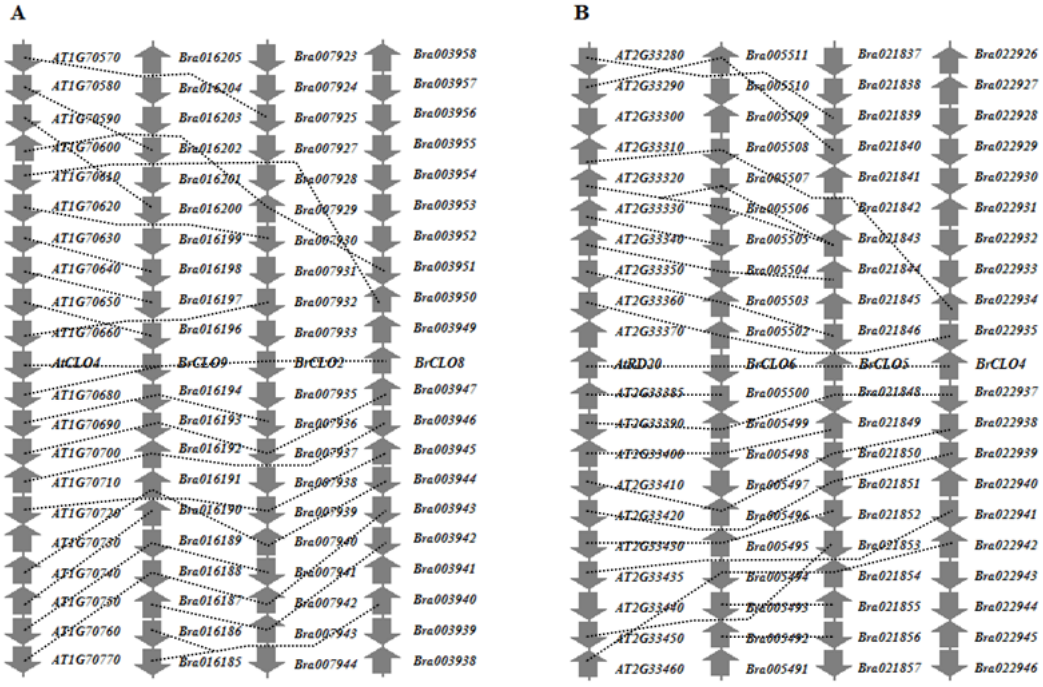

Fig. 4. Syntenic relationships of two pairs of BrCLO genes. Solid arrows represent gene loci, and the orientation of the arrows indicate the transcript direction. Syntenic genes are connected by dashed lines.

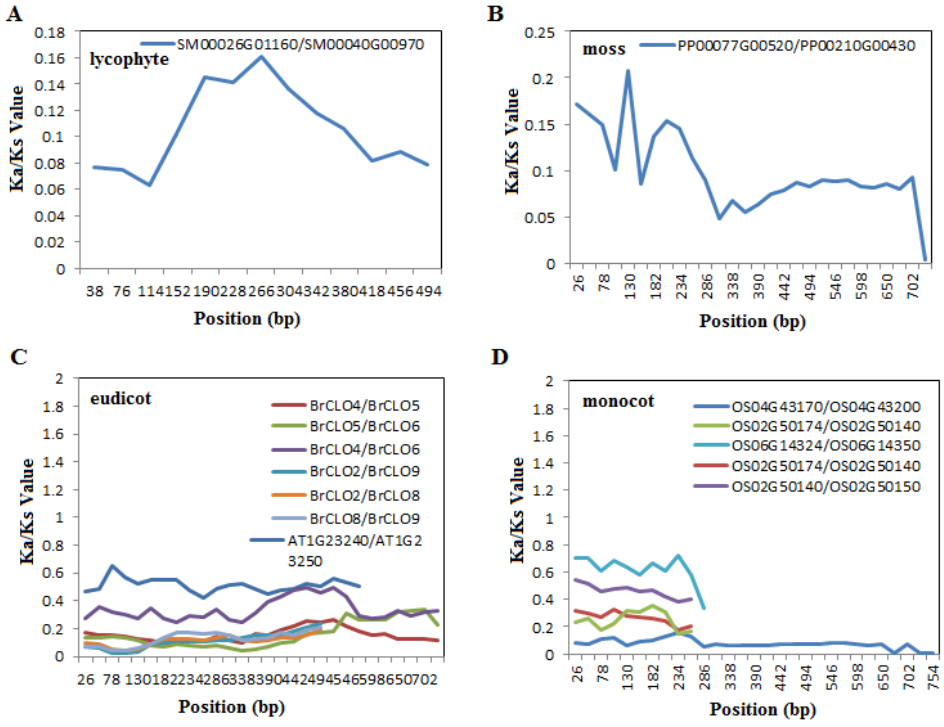

Fig. 5. The Ka/Ks distribution of *CLO* paralog pairs from lycophyte (A), moss (B), eudicot (C) and monocot (D) using a sliding window of 100 bp and a step size of 10 bp.

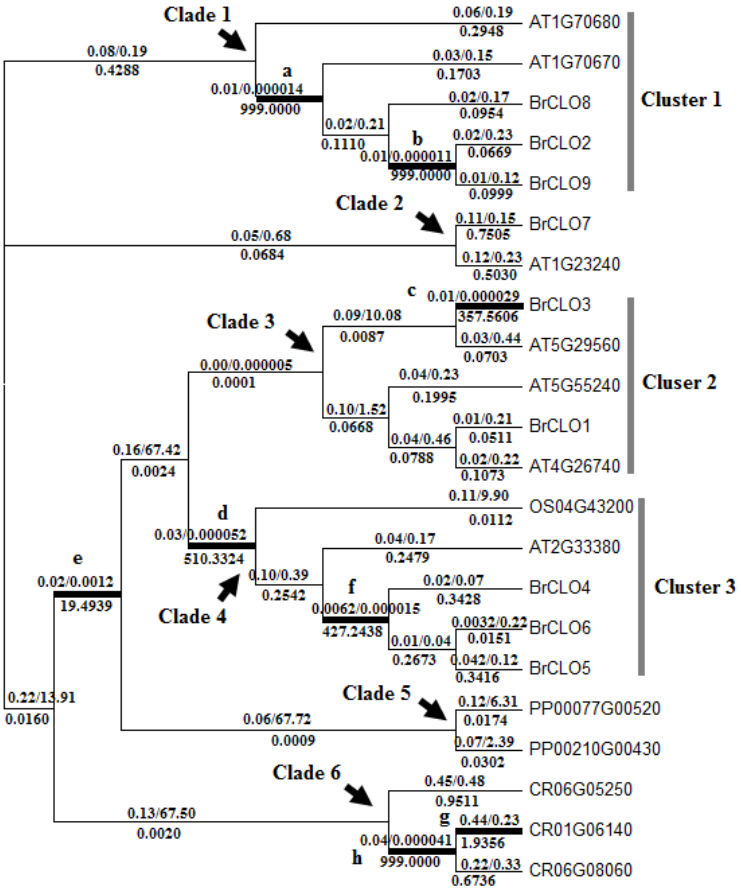

Fig. 6. Selective parameters estimated by the free-ratio model of the CLO gene family in plants. Branches with ω >1 are highlighted as thick lines (a~h). The Ka/Ks ratios are listed above the branches and the ω values are listed below the branches. Clade 1~6 are highlighted using solid arrows. Cluster 1-3 are labeled with thick lines.

branches with ω>1, and these branches, named as a~h, were labeled by thick lines (Fig. 6).

Positive selection might drive the evolution of some sites along a specific branch or clade of the phylogenetic tree. Therefore, we selected branch a~h and clade 1~6 as a foreground (Fig. 6), and the branch-site models were used to identify this type of sites. The LRT results suggested that positive selection contributed to the evolution of branches e, g, h and clade 6 (C6) (Table 2). However, in branches e and h, we found no positively selected sites at the significant level ($P < 0.05$). In contrast, 28 and 55 positively selected sites from branch g and C6, respectively, were under positive selection.

Functional divergence of CLO genes in plants

CLO proteins from *B. rapa* could be divided into 3 clusters (Fig. 6). DIVERGE2 was used to estimate the type I (θ_I) functional divergence between different CLO clusters, and some parameters and critical amino acid sites were listed (Table 3). The LRT results showed that the θ_I values of cluster 1/cluster 2 and cluster 1/cluster 3 comparisons were greater than zero at the significant level ($P < 0.01$). However, the θ_I value of the cluster 2/cluster 3 comparison was not greater than zero at the significant level ($P < 0.05$). To identify type I functional divergent sites, the distributions of site-specific posterior probabilities of three pairwise comparisons were visualized

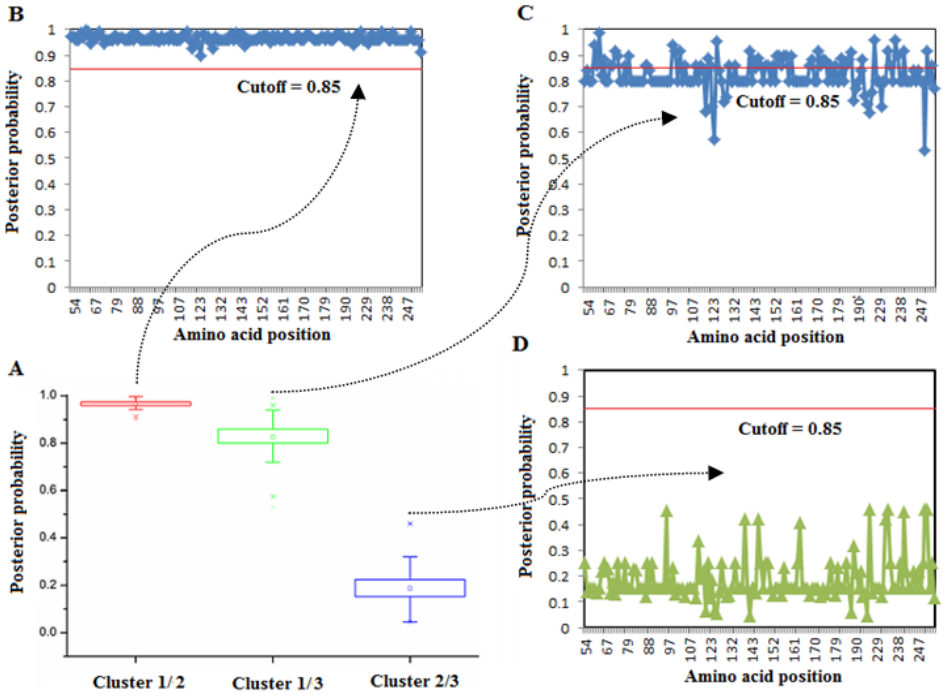

Fig. 7. Posterior probability profiles of the site-specific type I functional divergence between the different CLO clusters. The alignment positions with gaps were deleted, the red line indicates a cutoff = 0.85.

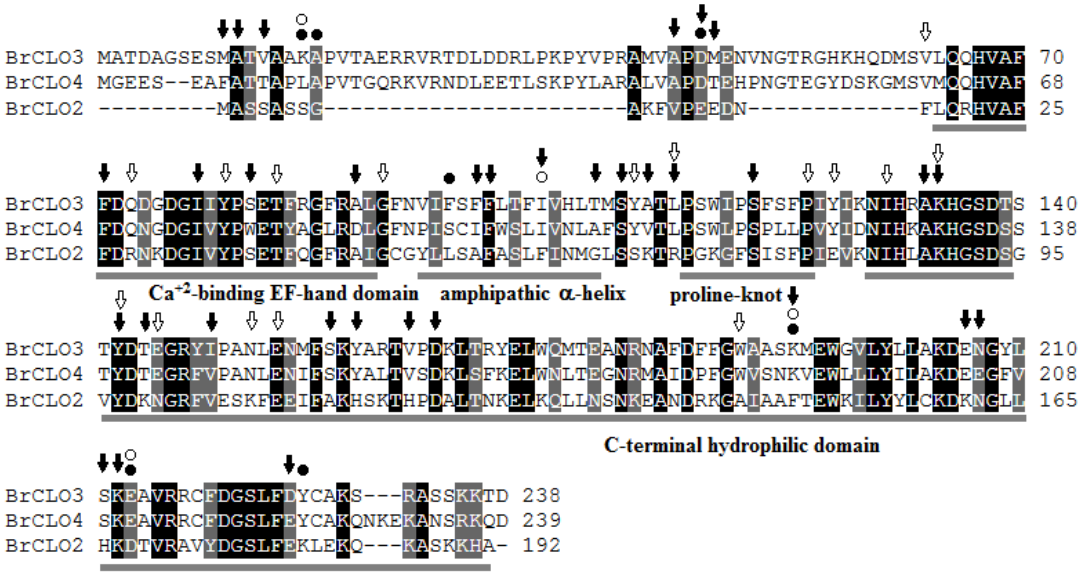

Fig. 8. Mapping of critical amino acids on the alignments of three representative CLO proteins in *B. rapa*. The amino acid sites below the solid and dashed circles are type I functional divergence sites in cluster 1/2 and cluster 1/3 comparisons, respectively. Positively selected sites on the CrCLO proteins from clade 6 and branch g are also highlighted using the solid and dashed circles.

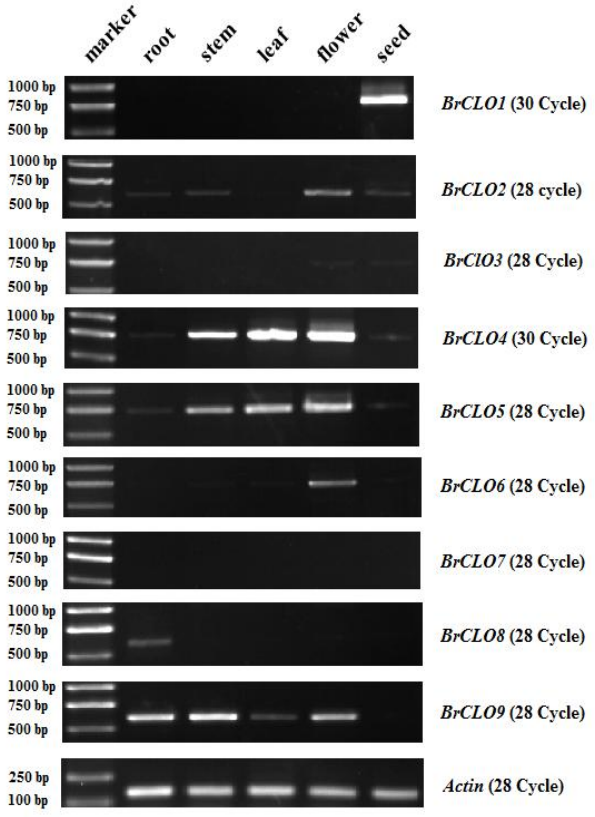

Fig. 9. Tissue-specific expression patterns of 9 CLO genes were analyzed using RT-PCR in *B. rapa*.

(Fig. 7A). If Q_k >0.85 was used as a cutoff, 152 functionally divergent sites were identified in the cluster 1/cluster 2 comparison (Fig. 7B); 58 functionally divergent sites were found in the cluster 1/cluster 3 comparison (Fig. 7C), and no functionally divergent sites were observed in the cluster 2/cluster 3 comparison (Fig. 7D).

Location of critical amino acids on the representative alignments

Initially, the representative alignment was generated by comparing three BrCLO proteins that contained BrCLO2, BrCLO3 and BrCLO4. The critical amino acid sites were 33 positively selected sites on three *C. reinhardtii* CLO proteins identified from C6 (P <0.01), 16 positively selected sites on one *C. reinhardtii* CLO protein identified from the g branch (P <0.01), 7 functionally divergent sites between cluster 1 and cluster 2 (P <0.01), and 4 functionally divergent sites between cluster 1 and cluster 3 (P <0.05). To shed light on the distribution of critical amino acids sites, we took the BrCLO2 protein as a reference and mapped these sites on this representative alignment (Fig. 8). The Ca^{+2}-binding EF-hand domain contained 7 positively selected sites. The amphipathic α-helix contained 2 functional divergence sites and 4 positively selected sites. The proline-knot contained 2 positively selected sites. Moreover, 3 functional divergence sites and 19 positively selected sites were co-located on the C-terminal hydrophilic domain. The remaining critical sites were mapped outside the functional domains.

Tissue expression patterns of 9 CLO genes in B. rapa

To investigate the relative expression levels of 9 *CLO* genes in various *B. rapa* tissues, semiquantitative RT-PCR was performed under normal growth conditions. Initially, we designed 9 pairs of gene-specific primers (Table 4). RT-PCR results showed that 9 *BrCLO* were constitutively or selectively expressed in the sampled tissues (Fig. 9). No transcription activity of *BrCLO7* was detected in any tissue. *BrCLO1* and *BrCLO8* were exclusively expressed in seed and root, respectively. The BrCLO2 gene was expressed in root, stem, flower and seed, and it was relatively higher in the flower. The BrCLO3 gene was expressed only in flower and seed. *BrCLO4*, *BrCLO5* and *BrCLO6* had similar expression patterns, and *BrCLO4* and *BrCLO5* were expressed in all sampled tissues. For five sampled tissues, *BrCLO6* was not expressed in root and seed, and *BrCLO9* was not expressed in seed.

DISCUSSION

To our knowledge, CLO genes encoding for caleosin anchored on the surface of the oil body are widely distributed in plants, fungi and algae (Chen et al., 1999; Naested et al., 2000; Liu et al., 2005; Purkrtova et al., 2007; Lin et al., 2012). The availability of fully sequenced genomes for model plants facilitates an understanding of detailed information about the CLO gene family at a genome-wide level. For

Table 1. The putative CLO gene family members in *B. rapa*

Name	[a]Gene model	[b]Chr: geome matched region	Number of Exons	Predicted protein properties		
				Length(aa) [e]	MW(kDa) [f]	pI[g]
BrCLO1	Bra026407	ChrA01: 9511963-9513377 (-1)[c]	6	245	28.13	6.16
BrCLO2	Bra007934	ChrA02: 11954251-11955703 (1)[d]	6	192	21.51	9.92
BrCLO3	Bra020623	ChrA02: 24142825-24145430 (-1)[c]	6	238	27.12	8.20
BrCLO4	Bra022936	ChrA03: 7853746-7855356 (-1)[c]	6	239	26.96	4.86
BrCLO5	Bra021847	ChrA04: 14855557-14857509 (-1)[c]	6	239	26.96	4.87
BrCLO6	Bra005501	ChrA05: 5831053-5832336 (1)[d]	6	239	26.94	5.37
BrCLO7	Bra012369	ChrA07: 8146540-8148101 (1)[d]	6	203	22.98	10.01
BrCLO8	Bra003948	ChrA07: 15922243-15923377 (-1)[c]	6	192	21.50	8.64
BrCLO9	Bra016195	ChrA07: 18935582-18937018 (1)[d]	6	192	21.64	8.24
BrCLO10	Bra002921	ChrA10: 6734335-6735507 (1)[d]	5	174	19.72	6.58
BrCLO11	Bra002920	ChrA10: 6740368-6740729 (1)[d]	2	87	10.37	4.77

Note: [a] From CoGe; [b] Chromosome; [c] Minus strand; [d] Plus strand; [e] Amino acid length; [f] Molecular weight; [g] Isoelectric point.

Table 2. Parameters estimation and likelihood ratio tests for site-specific and branch-site model.

Model	[a]np	Estimates of parameters	[b]lnL	[c]LRT pairs	[d]df	[e]2ΔlnL
		Site-specific models				
M0: one ratio	43	$\omega 0 = 0.12$	-6953.47			
M3: discrete	47	$P_0 = 0.10$, $p_1 = 0.48$, p2 = 0.42, $\omega_0 = 0.00029$, $\omega_1 = 0.06$, $\omega_2 = 0.26$	-6789.04	M0/M3	4	328.86**
M1a: neutral	44	$p_0 = 0.85$, $p_1 = 0.15$, $\omega_0 = 0.12$, $\omega_1 = 1.00$	-6920.77			
M2a: selection	46	$P_0 = 0.85$, $p_1 = 0.02$, p2 = 0.13, $\omega_0 = 0.12$, $\omega_1 = 1.00$, $\omega_2 = 1.00$	-6920.77	M1a/M2a	2	0
M7: beta	44	p = 0.90, q = 5.46	-6792.76			
M8: beta and ω	46	p = 0.90, q = 5.46, $p_0 = 0.99$, p_1= 0.00001, ω = 1.00	-6792.76	M7/M8	2	0
		Free ratio model				
Fr: free ratio	83	(see Fig. 6)	-6844.90	M0/Fr	40	217.14**
		Branch-site models				
model A (e) estimated ω2	46	$\omega_0 = 0.12$, $P_0 = 0.79$, $\omega_1 = 1.00$, $P_1 = 0.15$, $\omega_{2a\ fore} = 999.00$, $\omega_{2a\ back}$=0.12, P_{2a}=0.05, $\omega_{2b\ fore}$= 999.00, $\omega_{2b\ back} = 1.00$, P_{2b}=0.009	-6919.27			
model A (e) fixed ω2	45	$\omega_0 = 0.12$, $P_0 = 0.79$, $\omega_1 = 1.00$, $P_1 = 0.14$, $\omega_{2a\ fore} = 1.00$, $\omega_{2a\ back} = 0.12$, $P_{2a} = 0.06$, $\omega_{2b\ fore}$=1.00, $\omega_{2b\ back} = 1.00$, $P_{2b} = 0.01$	-6920.70	A (estimated) vs. A (fixed)	1	2.86*
model A (g) estimated ω2	46	$\omega_0 = 0.13$, $P_0 = 0.44$, $\omega_1 = 1.00$, $P_1 = 0.12$, $\omega_{2a\ fore} = 999.00$, $\omega_{2a\ back}$=0.13, P_{2a}=0.34, $\omega_{2b\ fore}$= 999.00, $\omega_{2b\ back} = 1.00$, P_{2b}=0.10	-6885.09			
model A (g) fixed ω2	45	$\omega_0 = 0.13$, $P_0 = 0.36$, $\omega_1 = 1.00$, $P_1 = 0.09$, $\omega_{2a\ fore} = 1.00$, $\omega_{2a\ back} = 0.13$, $P_{2a} = 0.43$, $\omega_{2b\ fore}$=1.00, $\omega_{2b\ back} = 1.00$, $P_{2b} = 0.11$	-6903.46	A (estimated) vs. A (fixed)	1	35.82**
model A (h) estimated ω2	46	$\omega_0 = 0.12$, $P_0 = 0.77$, $\omega_1 = 1.00$, $P_1 = 0.13$, $\omega_{2a\ fore} = 999.00$, $\omega_{2a\ back}$=0.12, P_{2a}=0.08, $\omega_{2b\ fore}$= 999.00, $\omega_{2b\ back} = 1.00$, P_{2b}=0.01	-6917.82			
model A (h) fixed ω2	45	$\omega_0 = 0.12$, $P_0 = 0.69$, $\omega_1 = 1.00$, $P_1 = 0.12$, $\omega_{2a\ fore} = 1.00$, $\omega_{2a\ back} = 0.12$, $P_{2a} = 0.16$, $\omega_{2b\ fore}$=1.00, $\omega_{2b\ back} = 1.00$, $P_{2b} = 0.03$	-6920.40	A (estimated) vs. A (fixed)	1	5.16*
model A (C6) estimated ω2	46	$\omega_0 = 0.11$, $P_0 = 0.27$, $\omega_1 = 1.00$, $P_1 = 0.04$, $\omega_{2a\ fore} = 1.80$, $\omega_{2a\ back}$=0.11, P_{2a}= 0.60, $\omega_{2b\ fore}$= 1.80, $\omega_{2b\ back} = 1.00$, P_{2b}= 0.09	-6876.88			
model A (C6) fixed ω2	45	$\omega_0 = 0.11$, $P_0 = 0.29$, $\omega_1 = 1.00$, $P_1 = 0.04$, $\omega_{2a\ fore} = 1.00$, $\omega_{2a\ back} = 0.11$, $P_{2a} = 0.58$, $\omega_{2b\ fore}$=1.00, $\omega_{2b\ back} = 1.00$, $P_{2b} = 0.09$	-6879.84	A (estimated) vs. A (fixed)	1	5.92**

Note: [a]np, number of free parameters; [b]lnL, log likelihood; [c]LRT, likelihood ratio test; [d]df, degrees of freedom; [e]2ΔlnL, twice the log-likelihood difference of two nested models compared; The significant tests at 5% cutoff are labeled with * and at 1% cutoff are labeled with **.

Table 3. The coefficient of type I functional divergence (θ_I) from pairwise comparisons between CLO groups

Pairwise comparisons	Type-I functional divergence			
	$\theta_I \pm$ S.E.[a]	LRT	Sig.[b]	Q_k threshold
Cluster1/2	0.96 ± 0.17	32.21**	P < 0.01	7 sites (Q_k > 0.99)
Cluster1/3	0.83 ± 0.17	24.73**	P < 0.01	4 sites (Q_k > 0.95)
Cluster2/3	0.19 ± 0.14	1.86	P > 0.05	0 site (Q_k > 0.85)

Note: [a] Standard error; [b] Significant level;

Table 4. Primers used for RT-PCR analysis of CLO genes in *B. rapa*

Gene	Forward Primer(from 5' to 3' end)	Reverse Primer(from 5' to 3' end)
BrCLO1	ATGAGTACGGCGACTGAG	TTAGTAGTAGGCTGTCTTG
BrCLO2	ATGGCTTCCTCTGCATCCTC	TTAAGCATGTTTCTTGGAAG
BrCLO3	ATGGCGACAGATGCAGGATC	TTAATCCGTCTTTTTAGAAG
BrCLO4	ATGGGAGAAGAGTCAGAGGC	TTAGTCTTGCTTGCGAGAATTG
BrCLO5	ATGGGAGAAGAGTCAGAGGC	TTAGTCTTGCTTGCGAGAATC
BrCLO6	ATGGGAGACGCGTCAGAAGC	TTAGTCTTGCTTGCGAGAATTAG
BrCLO7	ATGGCTCCTTCCGCTGCCTC	TTAACGTCTTCTTTTTTC
BrCLO8	ATGGCTTCCTCTGCATCACC	TTATGGATGCTCCTTAGAAG
BrCLO9	ATGGCTTCTTCTGAATCCAC	TCATGGATGTTTCTTAGAAG
Actin-7	TGTGACAATGGAACTGGAAT	GGACTGAGCTTCATCACCA

example, *Arabidopsis* and the rice CLO gene family have been well characterized in previous studies (Partridge and Murphy, 2009; Wei et al., 2011). Here, a total of 11 CLO genes were identified in the *B. rapa* genome, and their detailed information was listed (Table 1). Except for *BrCLO10* and *BrCLO11* genes, all BrCLO genes shared highly conserved exon-intron organizations and motif patterns (Fig. 1). According to sequence features, we predicted that *BrCLO10* and *BrCLO11* were generated by inserting a fragment into the intermediate region of the normal CLO gene, which could be paralogous *BrCLO1*. Based on phylogenetic relationships, the fact that *BrCLO10*, *BrCLO11* and *BrCLO1* were clustered together provided evidence for the this prediction (Fig. 2). Moreover, this result was also supported by the fact that *BrCLO10* and *BrCLO11* were localized on the same position in tandem (Fig. 3). Like a pseudogene in *A. thaliana* (Gierke et al., 2000), *BrCLO10* and *BrCLO11* could be pseudogenes that did not function properly. Phylogenetic analysis revealed that certain *Arabidopsis* CLO genes had no expected orthologous genes from *B. rapa*, which was inconsistent with the fact that a whole genome duplication event could be observed in *B. rapa* after its divergence from *Arabidopsis* lineage (Mun et al., 2009; Cheng et al., 2011). Therefore, we concluded that gene loss events had occurred during *BrCLO* evolution. To explore the complex expansive pattern of 11 BrCLO genes localized on 7 different chromosomes, syntenic analysis was performed. The results showed that *BrCLO2*/8/9 and *BrCLO4*/5/6 could be generated via segmental duplication (Fig. 4), whereas the fracture of one paralogous CLO gene of *BrCLO1* could be responsible for the formation of *BrCLO10* and *BrCLO11* as described above. The segmental duplication was consistent with an additional whole-genome triplication, named 4R, which occurred in *Brassica* diploid species (Lukens et al., 2003; Mun et al., 2009; Cheng et al., 2012).

To determine whether lineage-specific paralog pairs underwent adaptive evolution, the distributions of Ka/Ks values of 14 paralog pairs were visualized using JCoDA (Steinway et al., 2010) (Fig. 5). The results strongly suggested that these paralog

pairs were mainly subject to purifying selection after species-specific expansions. In addition, we applied site-specific, free-ratio, and branch-site models (Yang and Nielsen, 2002; Yang et al., 2005; Zhang et al., 2005; Yang, 2007) to detect selective pressures on the CLO proteins (Table 2). The site models found no amino acid sites under positive selection ($P < 0.05$). LRT analysis of the free-ratio model identified heterogeneous selective pressures among the branches, and revealed that the ω values of branches a~h were higher than 1 (Fig. 6). However, the branch-site model identified 28 and 55 positively selected sites from the branch g and C6, respectively, indicating that *C. reinhardtii* CLO proteins underwent rapid evolution. For *B. rapa* CLO proteins, they were mainly controlled by purifying or relaxed purifying selection. In addition, functional divergence analysis revealed that 152 and 58 site were responsible for type I functional divergence in cluster 1/cluster 2 (Fig. 7B) and cluster 1/cluster 3 (Fig. 7C) comparisons, respectively ($Q_k > 0.85$). This indicated that the *B. rapa* CLO proteins from different clusters underwent relaxed functional constraints. Interestingly, the representative amino acids, including positively selected sites and functionally divergent sites, were dispersed at poorly aligned regions and four potentially functional regions, including the Ca^{+2}-binding EF-hand, amphipathic α-helix, proline-knot, C-terminal hydrophilic domains (Fig. 8). Therefore, these critical sites could be responsible for the diverse roles of BrCLO genes in several biological processes, which is supported by some evidences in previous studies (Frandsen et al., 2001; Liu et al., 2005; Hanano et al., 2006; Aubert et al., 2010; Wei et al., 2011; Schoot et al., 2011; Lin et al., 2012; Tzen, 2012; Blee et al., 2012).

The expression profiles of 9 BrCLO genes were examined using RT-PCR in root, stem, leaf, flower and seed (Fig. 9). The *BrCLO1* gene was specifically and highly expressed in seed, implying that it could be tightly associated with oil bodies. Interestingly, *AtCLO1*, as an orthologous gene of *BrCLO1*, displayed similar seed-specific expression (Naested et al., 2000; Partridge and Murphy, 2009). *BrCLO4*, *BrCLO5* and *BrCLO6* were co-orthologous to *Arabidopsis RD20* that was induced by various abiotic stresses such as drought, salt stress, cold and wounding (Kant et al., 2008; Aubert et al., 2010). Interestingly, a similar expression pattern was detected only in the BrCLO6 gene, whereas *BrCLO4* and *BrCLO5* had a distinct expression. This demonstrated that changes in expression had occurred during the evolutionary process of the BrCLO gene family. Similarly, changes in expression were also found in *BrCLO2*, *BrCLO8* and *BrCLO9*. Moreover, EST and cDNA data supported the expression patterns of some BrCLO genes in this study (Lee et al., 2008). Overall, the transcript activities of *BrCLO1-6* were detected in seeds, which suggested that these CLO genes in *B. rapa* could be potentially associated with oil bodies.

In conclusion, *B. rapa* CLO proteins could be divided into three major groups, and their sequence features showed a high degree of similarity in the same group. Chromosomal mapping and syntenic analyses strongly suggested that segmental duplications contributed to CLO gene amplication. *B. rapa* CLO proteins evolved under purifying selection, and paralog pairs were subject to purifying selection after species-specific duplication. Interestingly, a total of 210 type I functional divergence sites were identified between cluster 1/2 and cluster 1/3 comparisons ($Q_k > 0.85$), implying that cluster 1 had split to a great extent apart from cluster 2/3. Expression analysis showed that 6 *BrCLO* genes could be potentially responsible for the formation of oil bodies.

Acknowledgments – The authors are grateful to the editor and anonymous reviewers for their useful suggestions. This work was supported by grants from the National Natural Science Foundation of China (30970211).

REFERENCES

Aubert, Y., Vile, D., Pervent, M., Aldon, D., Ranty, B., Simonneau, T., Vavasseur, A., and *J.P. Galaud* (2010). RD20, a stress-inducible caleosin, participates in stomatal control, transpiration and drought tolerance in *Arabidopsis thaliana*. *Plant Cell Physiol.* **51**, 1975-1987.

Bailey, T.L., Williams, N., Misleh, C., and *W.W. Li* (2006). MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res.* **34**, 369-373.

Blee, E., Flenet, M., Boachon, B., and *M.L. Fauconnier* (2012). A non-canonical caleosin from *Arabidopsis* efficiently epoxidizes physiological unsaturated fatty acids with complete stereoselectivity. *FEBS J.* **279**, 3981-3995.

Chen, J.C.F., Tsai C.C.Y., and *J.T.C. Tzen* (1999). Cloning and secondary structure analysis of caleosin, a unique calcium-binding protein in oil bodies of plant seeds, *Plant Cell Physiol.* **40**, 1079-1086.

Cheng, F., Liu, S., Wu, J., Fang, L., Sun, S., Liu, B., Li, P., Hua, W., and *X.W. Wang* (2011). BRAD, the genetics and genomics database for Brassica plants. *BMC Plant Biol.* **11**, 136.

Cheng, F., Wu, J., Fang, L., Sun, S., Liu, B., Lin, K., Bonnema, G., and *X.W. Wang* (2012). Biased gene fractionation and dominant gene expression among the subgenomes of *Brassica rapa*. *PLoS One* **7**, e36442.

Feng, H., Wang, X., Sun, Y., Wang, X., Chen, X., Guo, J., Duan, Y., Huang, L., and *Z. Kang* (2011). Cloning and characterization of a calcium binding EF-hand protein gene *TaCab1* from wheat and its expression in response to *Puccinia striiformis* f. sp. *tritici* and abiotic stresses. *Mol. Biol. Rep.* **38**, 3857-3866.

Frandsen, G.I., Mundy, J., and *T.C.Z. Jason* (2001). Oil bodies and their associated proteins, oleosin and caleosin. *Physiol. Plant* **112**, 301-307.

Frandsen, G., Müller-Uri, F., Nielsen, M., Mundy J., and *K. Skriver* (1996). Novel plant Ca^{2+}-binding protein expressed in response to abscisic acid and osmotic stress. *J. Biol. Chem.* **271**, 343-348.

Gierke, T., Todd, J., Ruuska, S., White, J., Benning, C., and *J. Ohlrogge* (2000). Microarray analysis of developing *Arabidopsis* seeds. *Plant Physiol.* **124**, 1570-1581.

Gu, X. (1999). Statistical methods for testing functional divergence after gene duplication. *Mol. Biol. Evol.* **16**, 1664-1674.

Guo, A.Y., Zhu, Q.H., Chen, X., and *J.C. Luo* (2007). GSDS: a gene structure display server. *Yi Chuan* **29**, 1023-1026.

Hanano, A., Burcklen, M., Flenet, M., Ivancich, A., Louwagie, M., Garin, J., and *E. Blée* (2006). Plant seed peroxygenase is an original heme-oxygenase with an EF-hand calcium binding motif. *J. Biol. Chem.* **281**, 33140-33151.

Hernandez-Pinzona, I., Patela K., and *D.J. Murphy* (2001). The *Brassica napus* calcium-binding protein, caleosin, has distinct endoplasmic reticulum- and lipid body-associated isoforms. *Plant Physiol. Biochem.* **39**, 615-622.

Jiang, P.L., Wang, C.S., Hsu, C.M., Jauh, G.Y., and *J.T. Tzen* (2007). Stable oil bodies sheltered by a unique oleosin in lily pollen. *Plant Cell Physiol.* **48**, 812-821.

Jiang, P.L., Chen, J.C., Chiu, S.T., and *J.T. Tzen* (2009). Stable oil bodies sheltered by a unique caleosin in cycad megagametophytes. *Plant Physiol. Biochem.* **47**, 1009-1016.

Kant, P., Gordon, M., Kant, S., Zolla, G., Daydov, O., Heimer, Y.M., Chalifa-caspi, V., Shaked, R., and S. Barak (2008). Functional-genomics-based identification of genes that regulate arabidopsis responses to multiple abiotic stresses. *Plant Cell Environ.* **31**, 697-714.

Lee, S.C., Lim, M.H., Kim, J.A., Lee, S.I., Kim, J.S., Jin, M., Kwon, S.J., Mun, J.H., Kim, Y.K., Kim, H.U., Hur, Y., and *B.S. Park* (2008). Transcriptome analysis in *B. rapa* under the abiotic stresses using *Brassica* 24K oligo microarray. *Mol. Cells* **26**, 595-605.

Lin, I.P., Jiang, P.L., Chen C.S., and *J.T.C. Tzen* (2012). A unique caleosin serving as the major integral protein in oil bodies isolated from *Chlorella sp.* cells cultured with limited nitrogen. *Plant Physiol. Biochem.* **61**, 80-87.

Liu, H., Hedley, P., Cardle, L., Wright, K.M., Hein, I., Marshall D., and *R. Waugh* (2005). Characterisation and functional analysis of two barley caleosins expressed during barley caryopsis development. *Planta* **221**, 513-522.

Lukens, L., Zou, F., Lydiate, D., Parkin, I., and *T. Osborn* (2003). Comparison of a *Brassica oleracea* genetic map with the genome of *Arabidopsis thaliana*. *Genetics* **164**, 359-372.

Maher, C., Stein, L., and *D. Ware* (2006). Evolution of *Arabidopsis* microRNA families through duplication events. *Genome Res.* **16**, 510-519.

Mun, J.H., Kwon, S.J., Yang, T.J., Seol, Y.J., Jin, M., Kim, J.A., Lim, M.H., Kim, J.S., Baek, S., Choi, B.S., Yu, H.J., Kim, D.S., Kim, N., Lim, K.B., Lee, S.I., Hahn, J.H., Lim, Y.P., Bancroft, I., and *B.S. Park* (2009). Genome-wide comparative analysis of the *Brassica rapa* gene space reveals genome shrinkage and differential loss of duplicated genes after whole genome triplication. *Genome Biol.* **10**, R111.

Murphy, D.J. (1993). Structure, function and biogenesis of storage lipid bodies and oleosins in plants. *Prog. Lipid Res.* **32**, 247-280.

Naested, H., Frandsen, G.I., Jauh, G.Y., Hernandez-Pinzon, I., Nielsen, H.B., Murphy, D.J., Rogers J.C., and *J. Mundy* (2000). Caleosins: Ca^{2+}-binding proteins associated with lipid bodies. *Plant Mol. Biol.* **44**, 463-476.

Nei, M., and *T. Gojobori* (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **3**, 418-426.

Partridge, M. and *D.J. Murphy* (2009). Role of membrane-bound caleosin and putative peroxygenase in biotic and abiotic stresses responses in *Arabidopsis*. *Plant Physiol. Biochem.* **47**, 796-806.

Poxleitner, M., Rogers, S.W., Samuels, A.L., Browse, J., and *J.C. Rogers* (2006) A role for caleosin in the degradation of oil-body storage lipid during seed germination. *Plant J.* **47**, 917-933.

Punta, M., Coggill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., Heger, A., Holm, L., Sonnhammer, E.L.L., Eddy, S.R., Bateman A., and *R.D. Finn* (2012). The Pfam protein families database. *Nucleic Acids Res*. **40**, 290-301.

Purkrtova, Z., d'Andrea, S., Jolivet, P., Lipovova, P., Kralova, B., Kodicek M., and *T. Chardot* (2007). Structural properties of caleosin: a MS and CD study. *Arch. Biochem. Biophys.* **464**, 335-343.

Saitou, N., and *M. Nei* (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol*. **4**, 406-425.

Schoot, C.V.D., Paul, L.K., Paul, S.B., and *P.L.H. Rinne* (2011). Plant lipid bodies and cell-cell signaling. *Plant Signaling and Behavior* **6**, 1732-1738.

Steinway, S.N., Dannenfelser, R., Laucius, C.D., Hayes, J.E., and *S. Nayak* (2010). JCoDA: a tool for detecting evolutionary selection. *BMC Bioinformatics* **11**, 284.

Suyama, M., Torrents, D., and *P. Bork* (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34**, 609-612.

Tamura, K., Dudley, J., Nei, M., and *S. Kumar* (2007). MEGA4: molecular evolutionary genetics analysis software version 4.0. *Mol. Biol. Evol*. **24**, 1596-1599.

Thompson, J.D., Gibson T.J., and *F. Plewniak* (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**, 4876-4882.

Tzen, J.T.C., Cao, Y.Z., Laurent, P., Ratnayake, C., and *A.H.C. Huang* (1993). Lipids, proteins, and structure of seed oil bodies from diverse species. *Plant Physiol.* **101**, 267-276.

Tzen, J.T.C. (2012). Integral proteins in plant oil bodies. *I.S.R.N. Bot.* doi:10.5402/2012/173954.

Wei, Z., Ma, H., and *X.C. Ge* (2011). Phylogenetic analysis and drought-responsive expression of the rice caleosin gene family. *Chinese Sci. Bull.* **56**, 1612-1621.

Yang, Z., and *R. Nielsen* (2002). Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.* **19**, 908-917.

Yang, Z., Wong W.S., and *R. Nielsen* (2005). Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* **22**, 1107-1118.

Yang, Z. (2007). PAML4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586-1591.

Zhang, J., Nielsen, R., and *Z. Yang* (2005). Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol. Biol. Evol.* **22**, 2472-2479.

Zienkiewicz, K., Castro A.J., Alche Jde D., Zienkiewicz A., Suarez, C. and *M.I. Rodriguez-Garcia* (2010). Identification and localization of a caleosin in olive (*Olea europaea* L.) pollen. *J. Exp. Bot.* **61**, 1537-1546.