

New insights on the function of plant acyl carrier proteins from comparative and evolutionary analysis

Xuezhen Yang

Yangtze University

Xiaoxue Liu

Yangtze University

Yanchen Zhou

Yangtze University

Fan Zhang

Yangtze University

Lan Huang

Yangtze University

Jun Wang (✉ wangjagri@yangtzeu.edu.cn)

Yangtze University <https://orcid.org/0000-0003-2756-1476>

Jian Song (✉ songjian@yangtzeu.edu.cn)

Yangtze University

Lijuan Qiu

Chinese Academy of Agricultural Sciences Institute of Bast Fiber Crops

Research article

Keywords: angiosperm, acyl carrier protein, evolution, differentiation, expression

Posted Date: December 16th, 2019

DOI: <https://doi.org/10.21203/rs.2.18739/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Version of Record: A version of this preprint was published on January 1st, 2021. See the published version at <https://doi.org/10.1016/j.ygeno.2020.11.015>.

Abstract

Background

Acyl carrier proteins (ACPs) play a central role in both plastidial and mitochondrial Type II fatty acid synthesis in plant cells. However, a large proportion of plant ACPs remain functionally uncharacterized, and their evolutionary history remains elusive.

Results

In present study, 97 putative ACPs were identified from ten angiosperm species examined. Based on phylogenetic analysis, ACP genes were grouped into plastidial (cpACP: ACP1/2/3/4/5) and mitochondrial (mtACP: mtACP1/mtACP2/mtACP3) ACPs. Protein sequence (motifs and length), tertiary structure, and gene structure (exon number, average intron length, and intron phase) were highly conserved in different ACP subclades. The differentiation of ACPs into distinct types occurred 85–98 and 45–57 million years ago. A limited proportion of ACP genes experience tandem or segmental duplication, corresponding to two rounds of whole genome duplication. Ka/Ks ratios revealed that duplicated ACP genes underwent a purifying selection. Regarding expression patterns, most ACPs were expressed constitutively and tissue-specifically. Notably, the average expression levels of ACP1, mtACP3, and mtACP1 were positively correlated with those of ACP3, ACP4, and mtACP2, respectively. Analysis of cis-elements showed that seven motifs (CACTFTPPCA1, DOFCOREZM, GT1CONSENSUS, CAATBOX1, ARR1AT, POLLEN1LELAT52, and GATABOX) related to tissue-specific, ABA, and light-mediated gene regulation were ubiquitous in all ACPs investigated, which shed new light on the regulation patterns of these central enzymatic partners of the FAS system.

Conclusions

This study presents a thorough overview of angiosperm ACP gene families and provides informative clues for the functional characterization of plant ACPs in the future.

Background

Fatty acids (FAs) are essential organic components acting as precursor molecules for lipids and as crucial substrates for oxidation and cellular energy production in all known organisms [1]. Functionally, in addition to their involvement in different types of lipid metabolism, FAs are involved in the response to abiotic and biotic stress conditions [2, 3]. *De novo* FA synthesis is accomplished by either the multifunctional FA synthase (FAS) polyprotein complex or by coordinated discrete monofunctional FA synthase enzymes termed Type I FAS and Type II FAS, respectively [4]. There are two distinct Type II FAS in plant cells, one located in the plastids (cpFAS), whose enzymes are responsible for most of the FA production, and the other in the mitochondria (mtFAS), whose enzymes are believed to manufacture the FA precursor necessary for lipoic acid synthesis [5, 6]. Acyl carrier proteins (ACPs) found within the basic FA synthesis apparatus function by shuttling acyl intermediates between the active sites of each enzyme component [7, 8]. Each ACP is originally expressed as an inactive apo-ACP subsequently activated to a holo-ACP by transfer of a phosphopantetheine group from CoA to a specific serine residue on the ACP backbone via a thioester bond [8-10].

The first ACP was isolated from *Escherichia coli* and characterized as a small, soluble acidic protein [11], subsequently identified in almost all living organisms [12, 13]. Within the plant kingdom, there are usually multiple ACP isoforms, e.g., in soybean (*Glycine max*) [14], spinach (*Spinacea oleracea*) [13, 15, 16], rapeseed (*Brassica napus*) [17, 18], rapeseed (*Brassica campestris*) [19], barley (*Hordeum vulgare*) [20], cigar flower (*Cuphea lanceolata*) [21, 22], and peanut (*Arachis hypogaea* L.) [23]. To date, eight ACPs have been identified in *Arabidopsis thaliana*, five plastidial and three mitochondrial ACPs [24].

Plant ACPs often show both constitutive and tissue-specific expression patterns [3]. *ACP1* and *ACP2* from *Arabidopsis* were preferentially expressed in seeds and roots, respectively [3]. In spinach, *ACP-I* was detected in the leaves, and *ACP-II* was constitutively expressed in the roots, leaves, and seeds [15, 16]. In peanut, *mtACP* (*AhACP2/3*) was highly expressed in the flower tissues, which was distinct from *cpACP* (*AhACP1/4/5*), where *AhACP1* was predominantly expressed in the seeds and *AhACP4* and *AhACP5* were abundant in the roots [23]. Changes in the expression levels of ACPs can result in altered FA composition in different tissues. Overexpression of *ACP1* in *Arabidopsis* resulted in elevated expression in the leaves rather than in the seeds and altered the FA composition in the leaves by reducing hexadecatrienoic acid (16:3) and increasing linolenic acid (18:3) [25]. Overexpression of *AtACP5* modulates FA composition and enhances salt stress tolerance in *Arabidopsis* and resulted in the reduction of oleic acid (C18:1) in the seedlings [26]. Reduction of an *ACP* isoform from soybean (*Glyma18g47950*) decreased both palmitic (16:0) and stearic acid (18:0) in root and resulted in the reduction of nodule number in symbiosis [14]. Ectopic expression of an *ACP* from olive (*OeACP*) in tobacco leaves significantly increased the content of oleic acid (18:1) and linolenic acid (18:3) but reduced hexadecadienoic acid (16:2) and hexadecatrienoic acid (16:3), implying that *OeACP* might be specifically involved in the regulation of FA chain length from C16 to C18 and desaturation from 18:0 to 18:1 and 18:3 as well [27]. Plant ACPs reportedly respond to environmental stimuli (e.g., light) and are involved in nutrient (e.g., phosphorous) uptake efficiency [3, 28].

Despite intensive study on ACPs in different plant species, the evolutionary history and expression patterns of plant ACPs have not been characterized. The availability of multiple genome sequencing allows the retrieval of putative *ACP* ortholog and paralog genes, and RNA-seq data allows us to investigate the comparative relevance of *ACP* genes from a functional perspective. Thus, we are now able to comprehensively understand the evolution of gene function. Here we collected *ACP* gene and protein sequences from ten plant species, including dicots and monocots, based on genome annotation information. First, we inferred the phylogenetic relationships and then used these to further investigate the conservation of primary amino acid sequence and tertiary structure, family expansion patterns, temporal and spatial expression profiles, and putative transcriptional *cis*-elements for each *ACP*. This study provides detailed information on *ACP* genes in plant lineages from an evolutionary perspective, which would aid the further functional study of *ACPs* in plants.

Results

Identification of acyl carrier protein (ACP) in different plant species

Although high functional significance has been attached to ACP during FA synthesis, evolutionary history of this gene family has not been characterized yet. Previously most ACPs of *A. thaliana* have been cloned and functionally characterized [3, 24-26, 29, 30], eight ACPs from this model plant lineage were used as query sequences to perform a BLASTP search against protein databases of nine plant species covering six families

(i.e. *Cruciferae*, *Malvaceae*, *Fabaceae*, *Euphorbiaceae*, *Vitaceae*, *Linaceae*, and *Gramineae*), including both monocot and eudicot. All ACP sequences were then subjected to functional annotation by InterProScan. A total of 97 non-redundant ACPs were identified with 5-17 ACP members in each plant species (Fig. 1, Additional file 1: Table S1). The Identified ACPs spanned from 78 to 254 amino acids in length with 134.8 amino acids on average, and 95.9% were less than 170 amino acids. Furthermore, most of these ACPs (88.7%) were acidic with a pI (isoelectronic point) of less than 7.0 (Additional file 1: Table S1), which was assumed to be vital for higher ACP structures [31]. We classified these ACPs into eight putative groups, namely ACP1-ACP5 (plastidial ACP, also termed cpACP) and mtACP1-mtACP3 (mitochondrial ACP), based on the sequence similarity with these eight ACPs in *A. thaliana* (Fig. 1; Additional file 1: Table S1). Notably, ACP4, mtACP1, and mtACP2 were present in all 10 plant species examined, with ACP4 being the most abundant, accounting for more than 1/3 of all ACPs detected (Fig. 1). This classification contradicts the *in silico* prediction of subcellular localization by WoLF pSORT [32], according to which ~96.4% of ACP1-5 is preferentially localized in chloroplast, whereas only 59.5% of putative mtACPs should be localized in mitochondria (Additional file 1: Table S1). Nonetheless, the classification system was still adopted in this study for convenience.

Phylogenetic relationships and gene structural analysis among plant ACPs

To further investigate the evolutionary relationship among ACPs from ten different plant species, phylogenetic trees were generated by calculating Maximum Likelihood and Monte Carlo Markov Chain (MCMC) topology based on the alignments of 97 full-length ACP sequences, rooted by ACP from *Chlamydomonas reinhardtii* (*CrmtACP2/CrACP4*). The results showed that ACPs were grouped into two distinct clades (I and II), which were further partitioned into two and three subclades, respectively (Fig. 2). Most branch topology (91/97) generated by two independent methods matched to each other perfectly, with few disorders (6/97) within the same branch (Fig. 2), which clearly indicated the robustness of the two tree topologies. Results showed that subclade Ia included ACP1-5, while subclades Ib, IIa, IIb, and IIc exclusively included ACP4, mtACP1, mtACP2, and mtACP3, respectively (Fig. 2). Both clades contained ACPs from all ten plant species examined, with at least eight species in each subclade. Specifically, eight ACPs from *Arabidopsis* (AtACPs, AtmtACPs) were included in all subclades. AtACP1, 2, 3 and 5 were grouped together within subclade Ia, while AtACP4 and AtmtACPs were distributed separately in different subclades (Fig. 2, indicated in red).

With regard to protein sequence and gene structure, the exon number of most ACPs within clade I was four, whereas that of clade II was two (Additional file 2: Figures S1A, Student's *t* test $P < 0.001$). As for protein length, mtACP exhibited an average length of 126.79 amino acids, which was much shorter than that of ACPs from clade I (142.73 amino acids) (Additional file 2: Figures S1B, Student's *t* test $P < 0.001$). Significant differences were observed in terms of molecular weight (MW), average intron length, and pI (isoelectronic point) as well (Student's *t* test $P < 0.05$) (Additional file 2: Figures S1C, S1D, S1E). With regard to the intron phase, nearly 95% of exons are disrupted by intron between codons (0), only 4.1% and 2.7% of exons are spaced between first (1) and second (2) nucleotide of code respectively (Additional file 1: Table S1), and no bias was observed between clades.

Structural conservation among different subclades

Protein function largely depends on primary structure, i.e., amino acid sequence; therefore, functionally important sites (often related to substrate binding and reaction) tended to be retained during evolution to

maintain a certain physiological role. To gain new insights into the extent of ACP sequence conservation, multiple alignment was performed on these sequences using MUSCLE, and displayed using Weblogo3. Generally, the N-terminal is less conserved (lower percentage of amino acids with overall height of bits score > 2.0) than that of the C-terminal in different subclades (Fig. 3a), except for subclade IIa. In a previous study, the Asp-Ser-Leu (DSL) motif was assigned the essential roles of activating ACPs before accepting an acyl group and functioning as a recognition site for phosphopantetheinyl transferase (PPTase) [10]. In our study, the DSL motif was conserved in 93.9% of all identified ACPs. However, this extent of conservation was much higher than the average conservation levels in subclades Ia, Ib, IIb, and IIc, while only 78.6% of ACPs retained the DSL motif in subclade IIa (Fig. 3a, indicated by red dashed rectangle), indicating that ACPs in subclade IIa (mtACP1) were less conserved.

ACPs are structurally flexible, both in the secondary structure and spatial conformation, which provides the protein with the capacity to alternatively sequester and release the acyl moiety to the active sites of partner enzymes [12]. We hypothesized that the consensus sequence of proteins from the same subclade might better reflect their function than that by a consensus of all ACP sequences from a certain species. Here, we first aligned ACP sequences within the same subclades consisting of ACPs from different plant species, after which we constructed a majority consensus sequence to model the tertiary structure. The resulting structures indicated that all ACPs contained 4 α helices in each subclade, including three longer helices and a shorter helix (Fig. 3b, c, d, e, f). In clade I (cpACPs), the shorter helix was located between helix I and II, whereas in clade II (mtACPs), the shorter helix was found between helix II and III (Fig. 3 b, c, d, e, f), suggesting that cpACPs and mtACPs might have diverged in tertiary structure, but the helix location was conserved to some extent.

ACP family expansion within species

Most flowering plants are estimated to experience one or more rounds of whole genome duplication during their evolutionary history [33-35]. Besides this, tandem duplication, segmental duplication, chromosomal rearrangement, and transposition contribute to genome permutation as well, thus, resulting in gene gain or loss, and consequently, leading to the expansion or retention of gene family members [36]. To better understand ACP gene expansion history, we investigated potential tandem and segmental duplication events within different plant genomes. Based on the genomic location of ACPs we identified five gene pairs neighboring each other separated by only ten or fewer extra genes from four different species, namely, Arabidopsis (one), cotton (one), *Linum* (two), and rice (one), indicating that these ACPs possibly experienced tandem duplication (Table 1). Thereafter, the sequence identities of proteins flanking the ACPs were calculated (thirty proteins located within, up-, and downstream for each. See details in the Materials and Methods). Eleven ACP gene pairs were identified as potentially derived from segmental duplication (Table 1). Among these ACPs originated by segmental duplication, soybean and *Linum* ACPs accounted for 54.5% and 36.3%, respectively. Notably, eight out of twelve ACPs from *Linum* underwent tandem or segmental duplication (Table 1). We also noticed that half (50%) of the tandem- or segmental duplication-derived ACPs were related to ACP4, mtACP1, and mtACP2. Interestingly, four ACPs (*LuACP4c*, *LuACP4d*, *LuACP4e*, and *LuACP4f*) on scaffold37 and scaffold475 of the *Linum* genome were estimated to have experienced both tandem and segmental duplication. We reasoned that likely, tandem duplication occurred prior to segmental duplication. However, ACPs of only six plant species examined, namely, *A. thaliana*, *G. max*, *G. raimondii*, *L. usitatissimum*, *P. vulgaris*, and *O. sativa*, probably experienced tandem- and segmental-duplication. Further, duplicated blocks were also observed among different legume species. A set of genes containing mtACP1 and mtACP2 was collinearly located on chromosomes spanning a physical distance

of 0.43 to 0.56Mb in soybean, medicago, and common bean (Additional file 3: Figures S2) genomes, suggesting that these ACPs diverged prior to the divergence of the legume species.

To map the divergence of these *ACP* genes, we calculated the K_s (synonymous substitution rate) of closely neighboring *ACP* gene pairs on the phylogenetic tree and estimated the time of divergence (Fig. 2a). We found that 45.8% of neighboring *ACP* gene pairs experienced tandem or segmental duplication, and the time of divergence of these genes is more recent than that of *ACP* genes without traces of duplications. As expected, *ACP* gene pairs that diverged within the last 10 million years were derived from the same subclades, whereas gene pairs that diverged more anciently were derived from different subclades. With regard to plant species, 52.9% of ACPs from soybean were involved in segmental duplication rather than tandem duplication and diverged at approximately 12.9 million years ago (Mya) on average, whereas 83.3% of ACPs from *Linum* underwent segmental and tandem duplication at approximately 5.1 Mya on average.

To elucidate the time between divergence events among clades, pairwise K_s between subclades were calculated. We found these K_s centered on at least eight major peaks (Additional file 4: Table S2) corresponding to the gene divergence event windows from ~237.7 Mya to ~16.4 Mya. All K_s among subclades share two peaks of 0.63 and 1.11, corresponding to ~51.6 Mya and ~91 Mya, respectively (Additional file 5: Figures S3). Similarly, we noticed that most *ACP* gene pairs (74.2%) exhibited $K_a/K_s < 1$, suggesting a purifying selection; conversely, only 12.9% experienced positive and neutral selections (Additional file 4: Table S2)

ACP gene expression profile

As remarkable differences were observed among ACPs in terms of clades, gene structure, and expansion patterns, we questioned whether these differences might also be reflected at the expression level. Rapidly progressing high-throughput sequencing technology, such as RNA-seq, provided a substantial amount of bulk data that was readily available for gene expression-profile investigation. We collected RNA-seq data of different tissues from four plant species, namely, *L. usitatissimum*, *G. max*, *A. thaliana*, and *M. truncatula*, to study the *ACP* gene expression patterns. Data from different resources were first normalized via the Z-score method, and then combined together for further analysis. Approximately 90.5% of these *ACP*s were detectable in seeds/embryos, leaves, stems, and roots of different species (Fig. 4a), whereas ACP2 was only detectable in *Arabidopsis*, indicating that most *ACP*s were constitutively expressed and actively functional. Hierarchical clustering analysis indicated that *ACP* of the same species tended to group together (Fig. 4a). We found three gene pairs derived from segmental- (*GmACP4a/GmACP4c*; *LuACP4a/LuACP4b*; *LumtACP1a/LumtACP1b*) and tandem-duplication (*AtACP2/AtACP3*; *LuACPe/LuACPf*; *LuACPc/LuACPd*), which were adjacently grouped and showed similar expression patterns (Table 1; Fig. 4a). Interestingly, seven among nine of these tandem- or segmental-duplication derived *ACP* gene pairs were of the same *ACP* type (ACP4, mtACP1, and mtACP3). Relevant network analysis showed that, regarding the expression pattern, 14 and 11 *ACP* gene pairs were positively and negatively correlated, respectively (Fig. 4b). ACPs from the same species tended to be positively correlated, and most ACPs common to both soybean and *Linum* were negatively correlated (Fig. 4b).

Regarding the average expression level of each *ACP* type from different species, positive correlations were observed among ACP1, ACP2, and ACP3, between mtACP1 and mtACP2, and between ACP4 and mtACP3 (Additional file 6: Figure S4A). In particular, the Pearson correlation coefficient (r) for the correlation between

ACP1 and ACP3 was 0.94, which was much higher than those for mtACP1 and mtACP2, or for ACP4 and mtACP3 (0.71 and 0.62, respectively). Both ACP1 and ACP3 exhibited relatively higher expression levels in the roots and seeds/embryos than in the stems and leaves (Additional file 6: Figure S4B). From the violin plot, we also noticed relatively higher expression levels of mtACP1 and mtACP2 than of mtACP3 (Student's *t* test $P < 0.001$). The widest distribution range was observed for ACP4 (Additional file 6: Figure S4B), probably owing to the fact that the ACP4 family is the largest among ACP families.

***Cis*-regulatory element analysis within the promoter region of ACP**

Cis-elements from gene promoter sequences are deemed specific binding sites by *trans*-acting transcription factors (TFs) and other additional co-factors to initiate gene transcription [38]. The type and number of *cis*-elements in the promoter region are assumed to be one of the key factors in coordinating the spatial and temporal gene expression pattern during plant growth and development and during responses to environmental stress conditions [37]. To better understand the *cis*-elements of these *ACP* genes, 1000bp upstream of the transcription start site (TSS) of the ACP coding region were analyzed *in silico* using the Plant *Cis*-acting Regulatory DNA Elements (PLACE) database [38]. In total, 249 putative *cis*-elements were identified, and each *ACP* gene contained 31-90 of these elements, among which only seven (2.8%)—CACTFTPPCA1, DOFCOREZM, GT1CONSENSUS, CAATBOX1, ARR1AT, POLLEN1LELAT52, and GATABOX—were ubiquitous in the promoter of all ACPs examined. Further, these seven *cis*-elements reached 10.3 times per gene on average, with variation among different ACPs. In terms of function, these seven ubiquitous *cis*-elements were related to organ-specific gene expression (i.e., the leaf, seed, and pollen), ABA, and light-mediated gene regulation (Fig. 5a; Additional file 7: Table S3). Furthermore, subclade-specific *cis*-elements were investigated as well. A total of 19, 7, 4, 9, 6 *cis*-elements belonging to subclade Ia, Ib, IIa, IIb, and IIc, respectively, were identified and found to be involved in tissue specific gene expression, phytohormone regulation, stress response, and light-mediated gene regulation (Fig. 5a). Regarding different species, there are 87 common *cis*-elements presenting in all species examined with function related to tissue-specific gene expression, phytohormone, and stress response etc. (Fig. 5b). Each species contained 1–19 species-specific *cis*-elements in one or more *ACP* gene promoter regions involved in the various functions mentioned above, but these differed among species (Fig. 5b). Statistically, all ACP (100%) promoters contained *cis*-elements related to the seeds, leaves, roots, shoots, ABA, auxin, biotic and abiotic stress, and light responses (Fig 5c). Only 23.47% of ACPs contained drought related *cis*-elements with 1.91 times/gene on average, suggesting some of these ACPs might be involved in response to drought stress. Interestingly, *cis*-elements were found to be slightly more abundant in clade II (1.16 elements per gene) than in clade I (1.09 elements per gene), but the difference was not significant (Student's *t*-test, $P > 0.05$), suggesting that both cpACP and mtACP might play equally important roles in plants.

Discussion

ACP is well conserved in plant lineages

ACP is not only present in *E. coli* as a small soluble acidic protein, but also identified to be ubiquitously present in almost all living organisms [11-13]. In this study we identified a total of 97 ACPs from ten plant lineages, with multiple members (5-17) in each plant lineage (Fig. 1, Table S1). These ACPs were found to possess protein sequences that were longer and with a wider range (from 78 to 254 amino acids with ~134.8 amino acids on

average) than microbial ACPs (77-100 amino acids residues) [12]. This is possibly due to the presence of leading sequence in the N terminal determining the ACP subcellular compartment in plants [30, 39].

Based on protein sequence homology, the 97 plant ACPs identified here were grouped into two distinct clades, cpACPs and mtACPs (Fig. 2). Each subclade, Ib, IIa, IIb, IIc exclusively contained ACP4, mtACP1, mtACP2, and mtACP3, respectively, except for a mixture of ACP1-5 in subclade Ia (Fig. 2), indicating a significant difference between cpACPs and mtACPs in terms of protein sequence. This was further evidenced by the observation that cpACPs exhibited increased protein length, greater exon numbers, and greater average intron length than those by mtACPs. Intron phase differed between cpACPs and mtACPs as well (Fig. 2; Additional file 1: Table S1). We also found that all ten-plant species contained ACP4, mtACP1, and mtACP2, of which ACP4 was the most abundant and accounted for more than 1/3 of total ACPs (Fig. 1), implying that these ACPs, especially ACP4, were well conserved in different plant lineages.

Protein function is directly correlated to its sequence, and the serine residual within the DSL motif is crucial for the attachment of phosphopantetheine to an apo-ACP for activation [10]. On average, 93.9% of ACPs contained the DSL motif, however only 78.6% of mtACP1 retained this motif (Fig. 3a), implying that the DSL motif was comprehensively conserved in flowering plants, but certain ACP types (mtACP1) might evolve toward diversification. In this study, the modeling of deduced consensus ACP sequences from each subclade revealed that ACPs (clade I) and mtACPs (clade II) differed in the extent of conservation of the shorter α helix location (Fig. 3b, c, d, e, f), suggesting that the shorter α -helix is probably important for either maintaining the basic active enzyme structure or directly fulfilling an ACP function [40-42].

However, in the mtACP clade, only ~38.6% showed consistency with *in silico* subcellular localization prediction models. This result was consistent with that of a previous study in which *GmmtACP2c* (*Glyma.18G244300*, previously termed *Glyma18g47950*) showed sequence similarity with *AtmtACP2* but was experimentally demonstrated to be localized in the plastids and to function in symbiosis [14]. Approximately 96.4% of ACPs from clade I were consistent with an *in silico* prediction model, probably because plastidial ACP is responsible for most FA synthesis [5, 6]. When considering the diversification in ACP N-terminal sequence among different subclades, these findings might also imply that ACP conservation potentially resides in signal peptide information as well.

ACP is a multiple gene family in many species, including plant lineages [43], indicating that an intensive expansion of *ACP* has occurred independently in different plant species. Our results showed that five *ACP* pairs were possibly derived from tandem duplications and eleven *ACP* pairs might have been derived from segmental duplications (Table 1). These *ACP* pairs were present in six plant species, the most abundantly in *Linum*, followed by soybean, *Arabidopsis*, cotton, common bean, and rice. This is likely owing to the high (35–50%) seed oil content found in *Linum* [44], which requires more *ACP* members for oil synthesis. However, the *ACP* derived from tandem and segmental duplication only accounted for a small portion (30/97) of all *ACPs*, the rest of which (68/97) were more likely a consequence of other means (e.g., transposition or polyploidy etc.), and thus, warrant further investigation. Among tandem- or segmental-derived *ACP* gene pairs, several showed similar expression patterns in the leaves, seeds, stems, and roots (Fig. 4a, b). This is consistent with the results of previous studies in which whole-genome duplication events generated gene pairs often showing similar expression patterns [34, 45, 46].

Historical expansion and differentiation of plant ACPs

Pairwise K_s value estimation (Table 1) indicated that ACPs derived from tandem and segmental duplication tended to diverge at a time scale in agreement with the recent whole genome duplication (WGD) events or later, i.e., soybean (~13 Mya), *Linum* (~5-9 Mya), common bean (~56.5 Mya), cotton (~16.6 Mya), Arabidopsis (~20-40 Mya), and rice (~40-50 Mya) [28, 47-51], suggesting that local genomic variations (tandem or segmental duplication) occurred simultaneously or slightly after WGD.

Pairwise K_s between subclades formed eight peaks (Additional file 4: Table S2), representing a time scale from ~16.4 to ~237.7 Mya, which is only slightly earlier than the time of convergence of seed plants at 310 Mya [52], implying an ancient history of ACP differentiation in plants. Further, we found that two of eight peaks (~0.63 and ~1.11) were present in all subclade-subclade comparisons (Additional file 4: Table S2), which represents two distinct opportunities at 85–98 Mya and 45-57 Mya, possibly signifying two putative, centralized ACP differentiation events. As is well known, core eudicots split from angiosperm at ~133 Mya [53]. Almost at the same time scale, a hexaploidy event (~125Mya), also termed gamma triplication, underlies the evolutionary past of most flowering plants [54, 55]. The first putative, centralized ACP differentiation event (85–98 Mya) is more recent than the hexaploidization event. This is probably because ACPs underwent subfunctionalization or neofunctionalization following the hexaploidization event. The time scale of the second putative, centralized differentiation event (45–57 Mya) neighbors the date of WGD in legumes, *G. max*, *P. vulgaris*, and *M. truncatula* around ~56–58 Mya [50, 56-58]. Further, the “split date” of maize from its progenitor (rice) that occurred around ~50 Mya [59, 60], might imply that ACP family differentiation occurred shortly after the large-scale genome wide duplication events ~50-60 Mya. This time scale is within proximity to the Cretaceous-Tertiary mass extinction event of ~65 Mya [61]. Potentially, the explosive expansion and differentiation of ACPs at this time was one of the means utilized for adaptation to severe environmental change.

ACP gene expression profiles

Expression data from four different plant species showed that most ACPs were constitutively expressed and abundantly presented in one or more plant organs (Fig. 4a; Additional file 1: Table S1). The tissue specific expression profile of ACPs agreed with the experimental findings in many plant species. ACP1 and ACP2 from Arabidopsis were preferentially expressed in seed and root, respectively [3]. For ACPs responsive to light, the only example was found in Arabidopsis leaves, where ACP4, rather than ACP1 and ACP2, was induced by light. Transcriptional initiation of certain genes is coordinated in various levels, including trans-acting factors, cis-elements, and other co-factors, among which cis-elements were centralized in the promoter region, where they play a vital role in mediating the binding of transcription factors [62]. In our study, among a total of 249 cis-elements identified in 97 ACP gene promoter regions (1000 bp up to the TSS), seven cis-elements were ubiquitous and highly conserved in all plant species investigated, with an abundance as high as ~10.3 repeats per gene. Further, these cis-elements were correlated to tissue-specific gene expression (i.e., the leaves, seeds, stems, or roots), ABA, and light-mediated gene regulation (Fig. 5a; Additional file 7: Table S3). However, no direct correlation between phytohormones and ACP expression has been established yet owing to the lack of documented experimental evidence, but the presence of ABA-related cis-elements in the ACP promoter region sheds new light on the regulation patterns of these central enzymatic partners of the FAS system.

In addition to the type and number of identified *cis*-elements (tissue specific, light induced, and ABA-related), our results also suggested a correlation with stress response (e.g. either biotic or abiotic), despite subclade-specific (ACP type) or species-specific expression pattern (Fig. 5a, b, c). However, the role of ACPs in stress tolerance or resistance has not been reported. As a product of FAS, FA reportedly shows extensive involvement in plant responses to stress [63, 64]. In this sense, ACPs probably act indirectly against adverse environments. However, upon deeper inspection, each ACP differentiated and exhibited a unique *cis*-element type and abundance, a finding that clearly supports the results reported herein and those previously reported indicating that ACP members are expressed differently in a tissue-specific manner [3, 15, 16, 23, 43]. We identified 87 commonly presented and 1–19 species-specific *cis*-elements in different species (Fig. 5b), a finding that confirms the high degree of conservation of the promoter region of ACPs, indicating the diversification of ACPs at the transcriptional level via *cis*-element evolution.

Conclusion

In this study we investigated the basic features of ACPs from ten plant species, including both dicot and monocot, and performed a comparative and evolutionary genomic study of ACP genes. We identified 97 ACPs from different plant species genome wide that displayed comprehensive conservation at different levels, e.g., isoelectronic point, protein length, and DSL motif. Phylogenetic topology revealed that putative cpACP (ACP1-5) and mtACP (mtACP1-3) belonged to different clades with distinct protein length, exon number, intron phase, average intron length, and tertiary structure, suggesting that ACPs in different plant lineages evolved independently, while remaining highly conserved. Evolutionary analysis revealed that a small proportion of ACPs diverged by tandem and segmental duplication, while most probably diverged by other means. The differentiation of ACPs into distinct types occurred approximately 85–98 and 45–57 Mya. Ka/Ks ratios revealed that duplicated *ACP* genes underwent purifying selections. Bulk analysis of RNA-seq data from four different plant species revealed that most *ACP* genes were constitutively expressed in different organs, and ACPs from *Linum* and soybean tended to be negatively correlated, and overtly ACP1-3, ACP4 and mtACP3, and mtACP1 and mtACP2 were positively correlated. *Cis*-element prediction of promoter sequences of different ACPs showed that seven *cis*-elements corresponding to organ specific gene expression, ABA, and light mediated gene regulation were constitutively presented in all ACPs with an average of ~10.3 repeats. In addition, species-specific and subclade-specific *cis*-elements were identified, which were probably related to the specific role of each ACP. This study provided detailed information on ACP genes in plant lineages in the view of evolution and expression.

Methods

Database search and sequence mining

The protein databases of ten plant species examined in this study were all downloaded from Phytozome database (www.phytozome.net). ACP of each plant species were selected by functional classification of both PF00550 (Pfam database) and PTHR20863 (PATHER database) [65, 66]. Then web based InterProScan program (<http://www.ebi.ac.uk/interpro/search/sequence-search>) was utilized for confirmation [67]. ACPs from different plant species were renamed based on the sequence similarity to eight ACPs from *Arabidopsis thaliana* (Additional file 8: ACP_raw_protein_sequence.docx). Gene structure and intron phase were calculated by GSDS2.0 [68]. Exon number, protein length, average intron length was obtained from www.phytozome.net.

Molecular weight and isoelectric point were predicted using programs from Sequence Manipulation Suite (<http://www.bioinformatics.org/sms2/>). Subcellular localization prediction of each ACP was performed using online server WoLF pSORT [32].

Multiple alignment and phylogenetic analysis

Deduced protein sequences of ACPs were aligned with MUSCLE3.8 standalone [69]. Aligned sequences were manually edited by Jalview [70], and the resulting alignments were shown in Additional file 9: Sequence. Aligned_sequence_for_tree.docx, The Bayesian phylogenetic tree was constructed using MrBayes v 3.2.1 and based on Bayesian information criterion (BIC), the VT+G4 model selected by ProtTest3.2 and ModelFinder [71-73]. Random trees were generated via four independent runs with four Markov chains by 5,000,000 generations and Bayesian trees were sampled every 100th generations with default settings. TRACER 1.7 [74] was used to assess the convergence of MCMC (Monte Carlo Markov Chain) tree using a 0.25% burn in value and FigTree v1.4.2 produced by BEAST [75], was used to display the phylogenetic tree. Posterior probabilities (pp) and the consensus trees were computed in MrBayes and labeled on the tree. A maximum likelihood tree was constructed by IQ-tree [76] with the VT+G4 model, clade support for the maximum likelihood tree was determined by bootstrap based on 1000 pseudoreplicates. Phylogenetic trees were modified by Adobe Illustrator CS3. Both MCMC and maximum likelihood trees were rooted by ACP from *Chlamydomonas reinhardtii* (XP_001693782.1 and XP_001699275.1)

Identification of conserved amino acids and tertiary structure modeling

Deduced amino acid sequences were first subjected to multiple alignments with the remove of gaps only presented in less than two ACP sequences using Jalview [70]. Then the conservation was calculated and displayed by WebLogo [77]. Conserved amino acids were defined by two criteria: height of the bit score is larger than 2.0 and percentage of certain amino acids to total ACP protein sequence aligned is greater than 90%. For tertiary structure modeling of each subclade, consensus protein sequence of each subclade was deduced by amino acids with most abundance at each position. Consensus protein sequence of each clade was then submitted for SWISS-MODEL (<http://www.swissmodel.expasy.org/>) for model-based tertiary structure prediction. Cartoon view of protein structure was viewed by PyMOL (v1.5.0.3).

Tandem and segmental duplication events identification

To determine the gene pairs whether they derived from tandem and segmental duplication events, we first collected protein sequence of 30 genes up and down stream of ACP gene on each chromosome to form an individual *.fasta file, then these protein sequence were aligned by Clustal W embedded in BioEdit (v7.0.9.0), after that pairwise identity were calculated using MatGat2.01 [78]. The protein pairs with identity score larger than 60% were deemed as putative paralogs or homologues. ACP genes with up and down stream protein pairs of putative paralogs or homologues larger than 10 were finally defined as segmental duplication events, and tandem duplication events was defined as ACP genes separated by less than 10 genes and with physical distance less than 200 kb distance [46]. Collinear distribution of ACP gene blocks among different species was identified according to the method described above and sketched in PowerPoint manually.

Ks calculation and divergent time estimation

Protein sequence of each ACP gene was first aligned by Muscle3.8.31 (Edgar, 2004), then CDS sequence of each ACP gene was aligned using RevTrans 2.0 [79] guided by protein sequence alignment. Ks was calculated by Yn00 program of PAML [80, 81]. Clade-clade pairwise Ks were plotted by Minitab 16.0. Divergent time (T) was estimated by $T = Ks / 2\lambda$, assuming a clock-like rate (λ) of 6.1×10^{-9} synonymous substitutions per site each year [82].

Transcriptional analysis ACP genes

Genome-wide transcriptional data of ACP genes in different plant species were downloaded from GEO datasets of NCBI (<http://www.ncbi.nlm.nih.gov/gds/>), detailed source could be found in Additional file 10: Table S4. For gene expression analysis, RNA-seq data (FPKM) of ACP genes from root, stem, leaf, and seed/embryo, in 4 different species were extracted from these transcriptional datasets, and first normalized by standard score (also termed as Z-score) method described by the formula of , in which z , x , μ , and σ represent normalized value, value before normalization, average value of each species, and standard deviation respectively. Normalized data were then transformed by $\log(2)$ and combined together for further viewing by MeV v4.8 [83].

Cis-regulatory element prediction and characterization

For *cis*-element prediction, promoter sequence (1000bp upstream sequence of the putative phytoACP coding genes) was first withdrawn from the genome database downloaded from www.phytozome.net. Then these sequence were submitted to search against PLACE database (<https://sogo.dna.affrc.go.jp/cgi-bin/sogo.cgi?lang=en&pj=640&action=page&page=newplace>) [38] to identify plant *cis*-acting regulatory DNA elements.

Abbreviations

ACP: Acyl carrier protein; FA: fatty acid; cp: chloroplast; mt: mitochondria.

Declarations

Acknowledgement

We would like to thank Editage [www.editage.cn] for English language editing.

Authors' contributions

JS, LQ, and JW, conceived and designed of this study. FZ, XL and XY collected the data. JW, JS and XL performed the phylogenetic analysis and the tertiary structure modeling, LH, XY and YZ conducted the *cis*-elements related analysis, XY, and XL, YZ worked with the gene expression profiling data. XY and JS drafted the manuscript. All authors read and approved the final manuscript.

Funding

Research was funded by grants of National Key R&D Program of China (Grant No.: 2016YFD0100304-6; 2016YFD0100201-14), the Open Fund of Engineering Research Center of Ecology and Agricultural Use of Wetland, Ministry of Education (Grant KF201808; KF201910).

Availability of data and material

All data and material were available in the additional files.

Ethics approval and consent to participate

Here is to clarify that the source of the plant materials used in our study was under any permissions necessary to collect such samples. Experimental research on plants (either cultivated or wild), including collection of plant material, complied with institutional, national, or international guidelines. Field studies were conducted in accordance with local legislation.

Consent for Publication

Not applicable

Competing interests

All authors have neither financial nor non-financial competing interests.

Declaration of interest statement

For the authors contributed for the manuscript, all authors have no conflict of interest. The contribution of authors is properly reflected by the order of author list and the "Author contribution statement". And the funds to support this work are also properly declared in the acknowledgement.

Besides, we declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the manuscript entitled "New insights on the function of plant acyl carrier proteins from comparative and evolutionary analysis"

References

1. Papackova Z, Cahova M. Fatty acid signaling: the new function of intracellular lipases. *Int J Mol Sci.* 2015;16(2):3831-3855.
2. Weber H. Fatty acid-derived signals in plants. *Trends Plant Sci.* 2002;7(5):217-224.
3. Bonaventure G, Ohlrogge JB. Differential regulation of mRNA levels of acyl carrier protein isoforms in *Arabidopsis*. *Plant physiol.* 2002;128(1):223-235.
4. Smith S, Witkowski A, Joshi AK. Structural and functional organization of the animal fatty acid synthase. *Prog Lipid Res.* 2003;42(4):289-317.
5. Wada H, Shintani D, Ohlrogge J. Why do mitochondria synthesize fatty acids? Evidence for involvement in lipoic acid production. *Proc Natl Acad Sci U S A.* 1997;94(4):1591-1596.
6. Yasuno R, von Wettstein-Knowles P, Wada H. Identification and molecular characterization of the beta-ketoacyl-[acyl carrier protein] synthase component of the *Arabidopsis* mitochondrial fatty acid synthase. *J Biol Chem.* 2004;279(9):8242-8251.
7. Kresge N, Simoni RD, Hill RL. The role of the acyl carrier protein in fatty acid synthesis: the work of P. Roy Vagelos. *J Biol Chem.* 2005;280(35):e32.

8. Hiltunen JK, Chen Z, Haapalainen AM, Wierenga RK, Kastaniotis AJ. Mitochondrial fatty acid synthesis-an adopted set of enzymes making a pathway of major importance for the cellular metabolism. *Prog Lipid Res.* 2010;49(1):27-45.
9. Lambalot RH, Walsh CT. Cloning, overproduction, and characterization of the *Escherichia coli* holo-acyl carrier protein synthase. *J Biol Chem.* 1995;270(42):24658-24661.
10. Mofid MR, Finking R, Marahiel MA. Recognition of hybrid peptidyl carrier proteins/acyl carrier proteins in nonribosomal peptide synthetase modules by the 4'-phosphopantetheinyl transferases AcpS and Sfp. *J Biol Chem.* 2002;277(19):17023-17031.
11. Majerus PW, Alberts AW, Vagelos PR. The acyl carrier protein of fatty acid synthesis: purification, physical properties, and substrate binding site. *Proc Natl Acad Sci U S A.* 1964;51:1231-1238.
12. Byers DM, Gong H. Acyl carrier protein: structure-function relationships in a conserved multifunctional protein family. *Biochem Cell Biol.* 2007;85(6):649-662.
13. Ohlrogge JB, Kuo TM. Plants have isoforms for acyl carrier protein that are expressed differently in different tissues. *J Biol Chem.* 1985;260(13):8032-8037.
14. Wang J, Toth K, Tanaka K, Nguyen CT, Yan Z, Brechenmacher L, et al. A soybean acyl carrier protein, GmACP, is important for root nodule symbiosis. *Mol Plant Microbe In.* 2014;27(5):415-423.
15. Scherer DE, Knauf VC. Isolation of a cDNA clone for the acyl carrier protein-I of spinach. *Plant Mol Biol.* 1987;9(2):127-134.
16. Schmid KM, Ohlrogge JB. A root acyl carrier protein-II from spinach is also expressed in leaves and seeds. *Plant Mol Biol.* 1990;15(5):765-778.
17. de Silva J, Loader NM, Jarman C, Windust JH, Hughes SG, Safford R. The isolation and sequence analysis of two seed-expressed acyl carrier protein genes from *Brassica napus*. *Plant Mol Biol.* 1990;14(4):537-548.
18. Safford R, Windust JH, Lucas C, De Silva J, James CM, Hellyer A, et al. Plastid-localised seed acyl-carrier protein of *Brassica napus* is encoded by a distinct, nuclear multigene family. *Eur J Biochem.* 1988;174(2):287-295.
19. Rose RE, DeJesus CE, Moylan SL, Ridge NP, Scherer DE, Knauf VC. The nucleotide sequence of a cDNA clone encoding acyl carrier protein (ACP) from *Brassica campestris* seeds. *Nucleic Acids Res.* 1987;15(17):7197.
20. Hansen L, von Wettstein-Knowles P. The barley genes *Acl1* and *Acl3* encoding acyl carrier proteins I and III are located on different chromosomes. *Mol Gen Genet.* 1991;229(3):467-478.
21. Schutt BS, Brummel M, Schuch R, Spener F. The role of acyl carrier protein isoforms from *Cuphea lanceolata* seeds in the de-novo biosynthesis of medium-chain fatty acids. *Planta.* 1998;205(2):263-268.
22. Voetz M, Klein B, Schell J, Topfer R. Three different cDNAs encoding acyl carrier proteins from *Cuphea lanceolata*. *Plant physiol.* 1994;106(2):785-786.
23. Li M-J, Wang X-J, Su L, Bi Y-P, Wan S-B. Characterization of five putative acyl carrier protein (ACP) isoforms from developing seeds of *Arachis hypogaea* L. *Plant Mol Biol Rep.* 2010;28(3):365-372.
24. Meyer EH, Heazlewood JL, Millar AH. Mitochondrial acyl carrier proteins in *Arabidopsis thaliana* are predominantly soluble matrix proteins and none can be confirmed as subunits of respiratory Complex I. *Plant Mol Biol.* 2007;64(3):319-327.

25. Branen JK, Chiou TJ, Engeseth NJ. Overexpression of acyl carrier protein-1 alters fatty acid composition of leaf tissue in *Arabidopsis*. *Plant physiol.* 2001;127(1):222-229.
26. Huang J, Xue C, Wang H, Wang L, Schmidt W, Shen R, et al. Genes of ACYL CARRIER PROTEIN family show different expression profiles and overexpression of ACYL CARRIER PROTEIN 5 modulates fatty acid composition and enhances salt stress tolerance in *Arabidopsis*. *Front Plant Sci.* 2017;8(987).
27. De Marchis F, Valeri MC, Pompa A, Bouveret E, Alagna F, Grisan S, et al. Overexpression of the olive acyl carrier protein gene (*OeACP1*) produces alterations in fatty acid composition of tobacco leaves. *Transgenic Res.* 2016;25(1):45-61.
28. Zhang D, Song H, Cheng H, Hao D, Wang H, Kan G, et al. The acid phosphatase-encoding Gene *GmACP1* contributes to soybean tolerance to low-phosphorus stress. *PLoS Genet.* 2014;10(1):e1004061.
29. Hlousek-Radojic A, Post-Beittenmiller D, Ohlrogge JB. Expression of constitutive and tissue-specific acyl carrier protein isoforms in *Arabidopsis*. *Plant physiol.* 1992;98(1):206-214.
30. Shintani DK, Ohlrogge JB. The characterization of a mitochondrial acyl carrier protein isoform isolated from *Arabidopsis thaliana*. *Plant physiol.* 1994;104(4):1221-1229.
31. Schulz H. On the structure-function relationship of acyl carrier protein of *Escherichia coli*. *J Biol Chem.* 1975;250(6):2299-2304.
32. Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, et al. WoLF PSORT: protein localization predictor. *Nucleic Acids Res.* 2007;35(Web Server issue):W585-587.
33. Bowers JE, Chapman BA, Rong J, Paterson AH. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature.* 2003;422(6930):433-438.
34. Blanc G, Wolfe KH. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell.* 2004;16(7):1667-1678.
35. Pfeil BE, Schlueter JA, Shoemaker RC, Doyle JJ. Placing paleopolyploidy in relation to taxon divergence: A phylogenetic analysis in legumes using 39 gene families. *Syst Biol.* 2005;54(3):441-454.
36. Freeling M. Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. *Annu Review Plant Biol.* 2009;60:433-453.
37. Wu C, Washida H, Onodera Y, Harada K, Takaiwa F. Quantitative nature of the Prolamin-box, ACGT and AACAA motifs in a rice glutelin gene promoter: minimal cis-element requirements for endosperm-specific gene expression. *The Plant J : for cell and molecular biology.* 2000;23(3):415-421.
38. Higo K, Ugawa Y, Iwamoto M, Korenaga T. Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Res.* 1999;27(1):297-300.
39. Takahashi T, Muneoka Y, Lohmann J, Lopez de Haro MS, Solleder G, Bosch TC, et al. Systematic isolation of peptide signal molecules regulating development in hydra: LWamide and PW families. *Proc Natl Acad Sci U S A.* 1997;94(4):1241-1246.
40. Kim Y, Prestegard JH. Refinement of the NMR structures for acyl carrier protein with scalar coupling data. *Proteins.* 1990;8(4):377-385.
41. Zornetzer GA, Fox BG, Markley JL. Solution structures of spinach acyl carrier protein with decanoate and stearate. *Biochemistry.* 2006;45(16):5217-5227.
42. Johnson MA, Peti W, Herrmann T, Wilson IA, Wuthrich K. Solution structure of *Asl1650*, an acyl carrier protein from *Anabaena* sp. PCC 7120 with a variant phosphopantetheinylation-site sequence. *Protein*

- science : a publication of the Protein Society. 2006;15(5):1030-1041.
43. Battey JF, Ohlrogge JB. Evolutionary and tissue-specific control of expression of multiple acyl-carrier protein isoforms in plants and bacteria. *Planta*. 1990;180(3):352-360.
 44. Soto-Cerda BJ, Diederichsen A, Ragupathy R, Cloutier S. Genetic characterization of a core collection of flax (*Linum usitatissimum* L.) suitable for association mapping studies and evidence of divergent selection between fiber and linseed types. *BMC Plant Biol*. 2013;13:78.
 45. Yim WC, Lee BM, Jang CS. Expression diversity and evolutionary dynamics of rice duplicate genes. *Mol Genet Genomics*. 2009, 281(5):483-493.
 46. Zan YJ, Ji Y, Zhang Y, Yang SH, Song YJ, Wang JH. Genome-wide identification, characterization and expression analysis of populus leucine-rich repeat receptor-like protein kinase genes. *BMC genomics*. 2013;14.
 47. Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. japonica). *Science*. 2002;296(5565):92-100.
 48. Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, et al. Genome sequence of the palaeopolyploid soybean. *Nature*. 2010;463(7278):178-183.
 49. Wang Z, Hobson N, Galindo L, Zhu S, Shi D, McDill J, et al. The genome of flax (*Linum usitatissimum*) assembled de novo from short shotgun sequence reads. *The Plant J : for cell and molecular biology*. 2012;72(3):461-473.
 50. Schmutz J, McClean PE, Mamidi S, Wu GA, Cannon SB, Grimwood J, et al. A reference genome for common bean and genome-wide analysis of dual domestications. *Nat Genet*. 2014;46(7):707-713.
 51. Li F, Fan G, Lu C, Xiao G, Zou C, Kohel RJ, et al. Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. *Nat Biotechnol*. 2015;33(5):524-530.
 52. Soltis DE, Bell CD, Kim S, Soltis PS. Origin and early evolution of angiosperms. *Ann NY Acad Sci*. 2008;1133:3-25.
 53. Jiao Y, Leebens-Mack J, Ayyampalayam S, Bowers JE, McKain MR, McNeal J, et al. A genome triplication associated with early diversification of the core eudicots. *Genome Biol*. 2012;13(1):R3.
 54. Vekemans D, Proost S, Vanneste K, Coenen H, Viaene T, Ruelens P, et al. Gamma paleohexaploidy in the stem lineage of core eudicots: significance for MADS-Box gene and species diversification. *Mol Biol Evol*. 2012;29(12):3793-3806.
 55. Zheng C, Chen E, Albert VA, Lyons E, Sankoff D. Ancient eudicot hexaploidy meets ancestral eurosid gene order. *BMC genomics*. 2013;14 Suppl 7:S3.
 56. Mudge J, Cannon SB, Kalo P, Oldroyd GE, Roe BA, Town CD, et al. Highly syntenic regions in the genomes of soybean, *Medicago truncatula*, and *Arabidopsis thaliana*. *BMC Plant Biol*. 2005;5:15.
 57. Lavin M, Herendeen PS, Wojciechowski MF. Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the tertiary. *Syst Biol*. 2005;54(4):575-594.
 58. Young ND, Debellé F, Oldroyd GED, Geurts R, Cannon SB, Udvardi MK, et al. The medicago genome provides insight into the evolution of rhizobial symbioses. *Nature*. 2011;480(7378):520-524.
 59. Kellogg EA. Evolutionary history of the grasses. *Plant physiol*. 2001;125(3):1198-1205.
 60. Xu JH, Messing J. Organization of the prolamin gene family provides insight into the evolution of the maize genome and gene duplications in grass species. *Proc Natl Acad Sci U S A*. 2008;105(38):14330-14335.

61. Fawcett JA, Maere S, Van de Peer Y. Plants with double genomes might have had a better chance to survive the Cretaceous–Tertiary extinction event. *Proc Natl Acad Sci U S A*. 2009;106(14):5737-5742.
62. Arnone MI, Davidson EH. The hardwiring of development: organization and function of genomic regulatory systems. *Development (Cambridge, England)* 1997;124(10):1851-1864.
63. Upchurch RG. Fatty acid unsaturation, mobilization, and regulation in the response of plants to stress. *Biotechnol Lett*. 2008;30(6):967-977.
64. Chaffai R, Seybou TN, Marzouk B, El Ferjani E. A comparative analysis of fatty acid composition of root and shoot lipids in *Zea mays* under copper and cadmium stress. *Acta Biol Hung*. 2009, 60(1):109-125.
65. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res*. 2016;44(D1):D279-D285.
66. Mi H, Muruganujan A, Casagrande JT, Thomas PD. Large-scale gene function analysis with the PANTHER classification system. *Nat Protoc*. 2013;8:1551.
67. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics*. 2014;30(9):1236-1240.
68. Hu B, Jin J, Guo A-Y, Zhang H, Luo J, Gao G. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics*. 2015;31(8):1296-1297.
69. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004;32(5):1792-1797.
70. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*. 2009;25(9):1189-1191.
71. Darriba D, Taboada GL, Doallo R, Posada D: ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics*. 2011;27(8):1164-1165.
72. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 2017;14(6):587-589.
73. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*. 2012;61(3):539-542.
74. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst Biol*. 2018;67(5):901-904.
75. Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, et al. BEAST 2: A software platform for Bayesian evolutionary analysis. *PLOS Comput Biol*. 2014;10(4):e1003537.
76. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 2015;32(1):268-274.
77. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. *Genome Res*. 2004;14(6):1188-1190.
78. Campanella JJ, Bitincka L, Smalley J. MatGAT: an application that generates similarity/identity matrices using protein or DNA sequences. *BMC Bioinformatics*. 2003;4:29.
79. Wernersson R, Pedersen AG. RevTrans: Multiple alignment of coding DNA from aligned amino acid sequences. *Nucleic Acids Res*. 2003;31(13):3537-3539.

80. Yang Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci.* 1997;13(5):555-556.
81. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 2007;24(8):1586-1591.
82. Lynch M, Conery JS. The evolutionary fate and consequences of duplicate genes. *Science.* 2000;290(5494):1151-1155.
83. Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, et al. TM4: a free, open-source system for microarray data management and analysis. *Biotechniques.* 2003;34(2):374-378.

Table

Due to technical limitations, Table 1 available in the Supplementary Files section below.

Figures

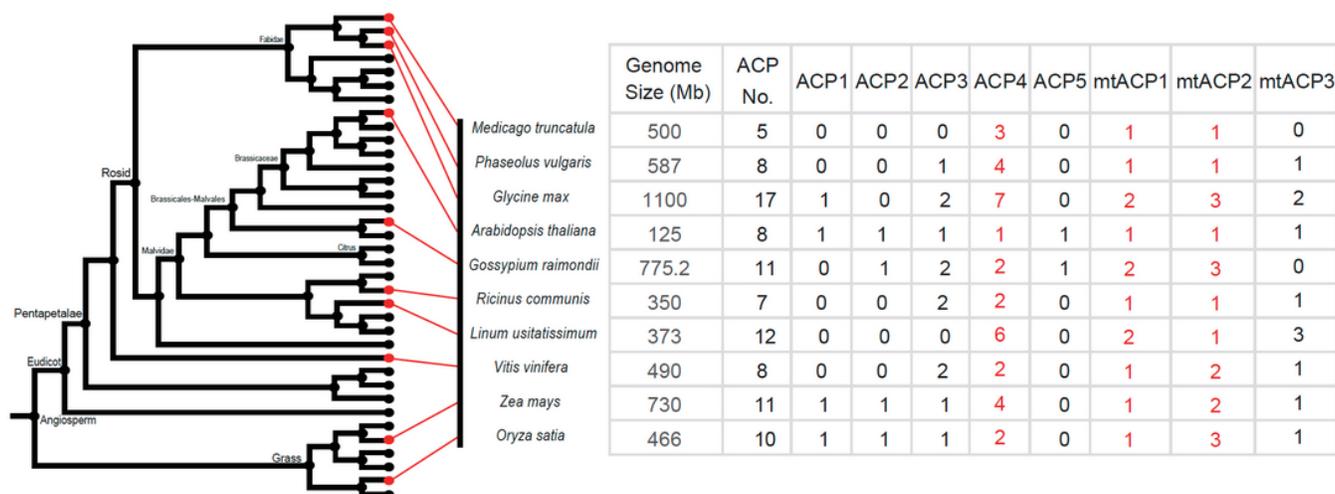


Figure 1

Different plant species sampled in this study are indicated by red space-filled circle. Numbers within brackets representing ACP proteins identified from each species. Phylogenetic tree was redraw based on that from www.phytozome.net. ACP classification was performed based on sequence similarity to ACPs from *Arabidopsis thaliana*.

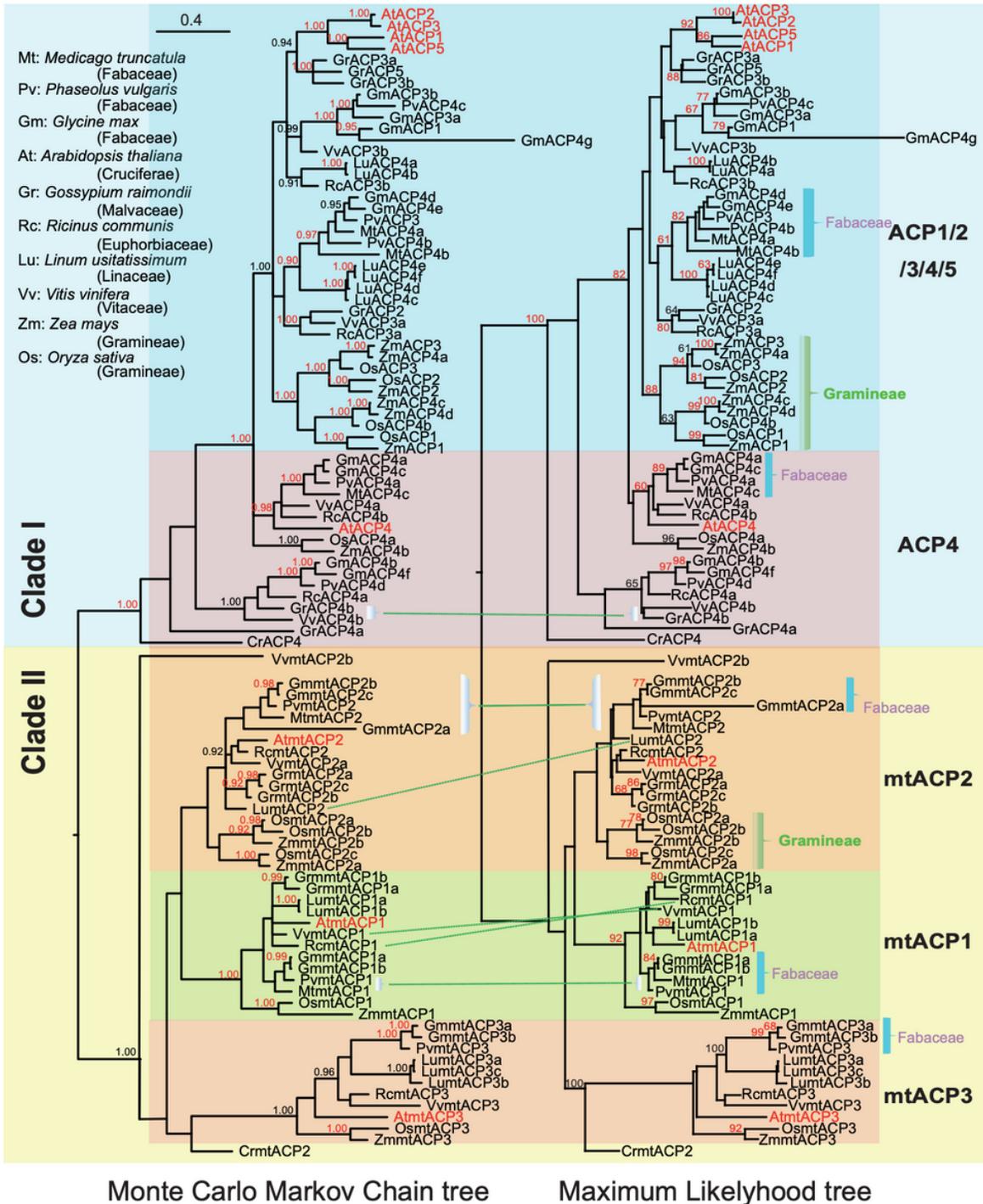


Figure 2

(A) MCMC (Monte Carlo Markov Chain) tree was generated MrBayes v 3.2.1. posterior probabilities value larger than 0.90 were placed on the tree. (B) Maximum likelihood tree was constructed by IQ-tree with 1000 pseudoreplicates, Bootstrap value larger than 60% were placed on the tree. Both MCMC and maximum likelihood tree were rooted by CrmtACP2 and CrACP4 from *Chlamydomonas reinhardtii*

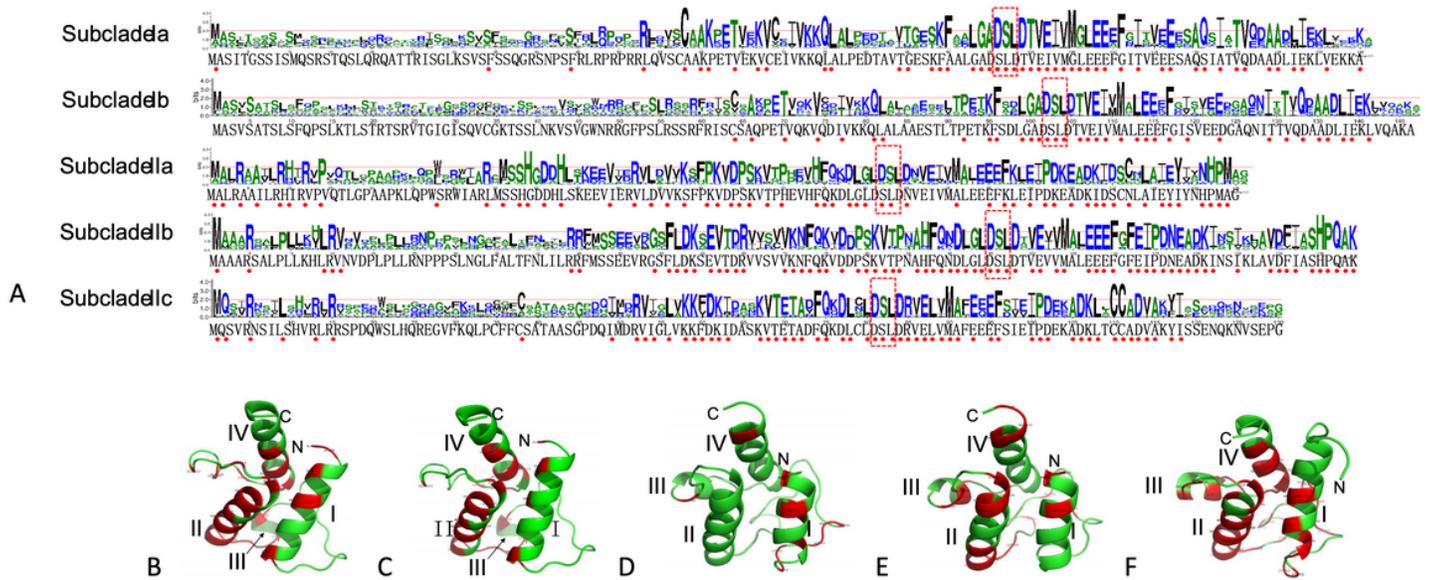


Figure 3

(A) Conserved motifs and amino acids identification within subclades. Sequence logo was generated by Weblogo 3.0 (<http://weblogo.threeplusone.com/>) based on alignment of ACPs from each subclade. Sequence conservation at each position (measured in bits) are indicated by the overall height of each stack, and the height of symbols within the stack represents the relative frequency of the corresponding amino acid at certain position. Consensus protein sequence was displayed under x-axis, DSL motif is indicated by red dashed rectangle. Conserved amino acids with over height of bits score > 2 and relative frequency > 50% is indicated by red space-filled circle. (B) - (F) Ribbon views of deduced consensus ACPs from subclade Ia, Ib, IIa, IIb, and IIc respectively. Tertiary structure of ACP was modeled by SWISS-MODEL and viewed by PyMOL (v1.5.0.3). Templates for modeling for subclades Ia~IIc are 2xz1.1.C (65.85% in similarity), 2ava.1.A (62.20%), 2m5r.1.A (62.82%), 2m5r.1.A (61.54%), 2dnw.1.A (42.86%) respectively. N, C terminal and helix I-IV were also indicated.

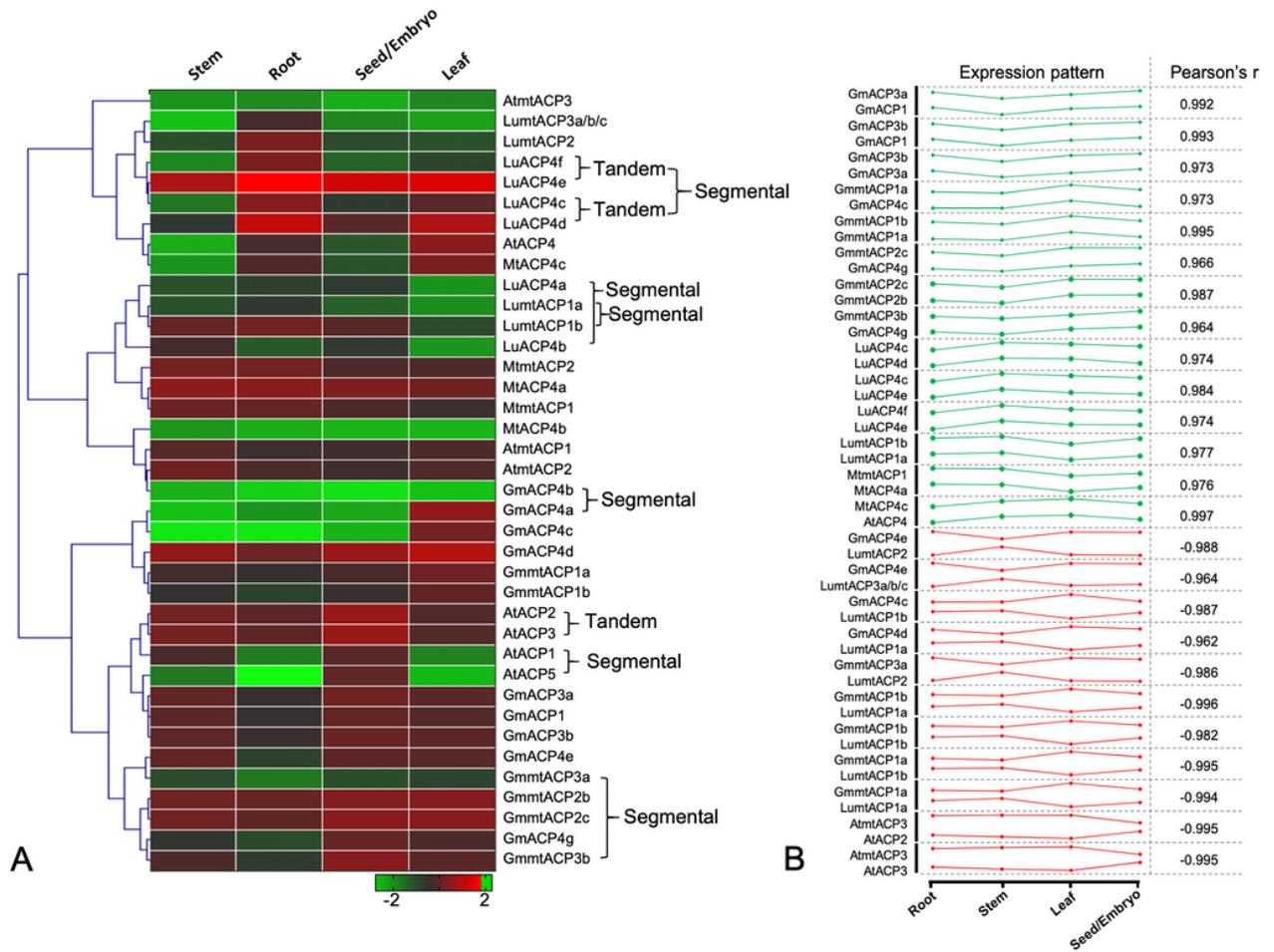


Figure 4

(A) Relative expression level (normalized by Z-score) of ACPs in stem, root, seed/embryo, and leaf from four different plant species (Arabidopsis, soybean, linum and medicago), ACPs were clustered by hierarchy clustering analysis. Tandem and segmental derived ACPs were indicated based on Table 1. (B) ACPs of similar expression pattern were identified by Pearson correlation analysis with Pearson's coefficient >0.96.

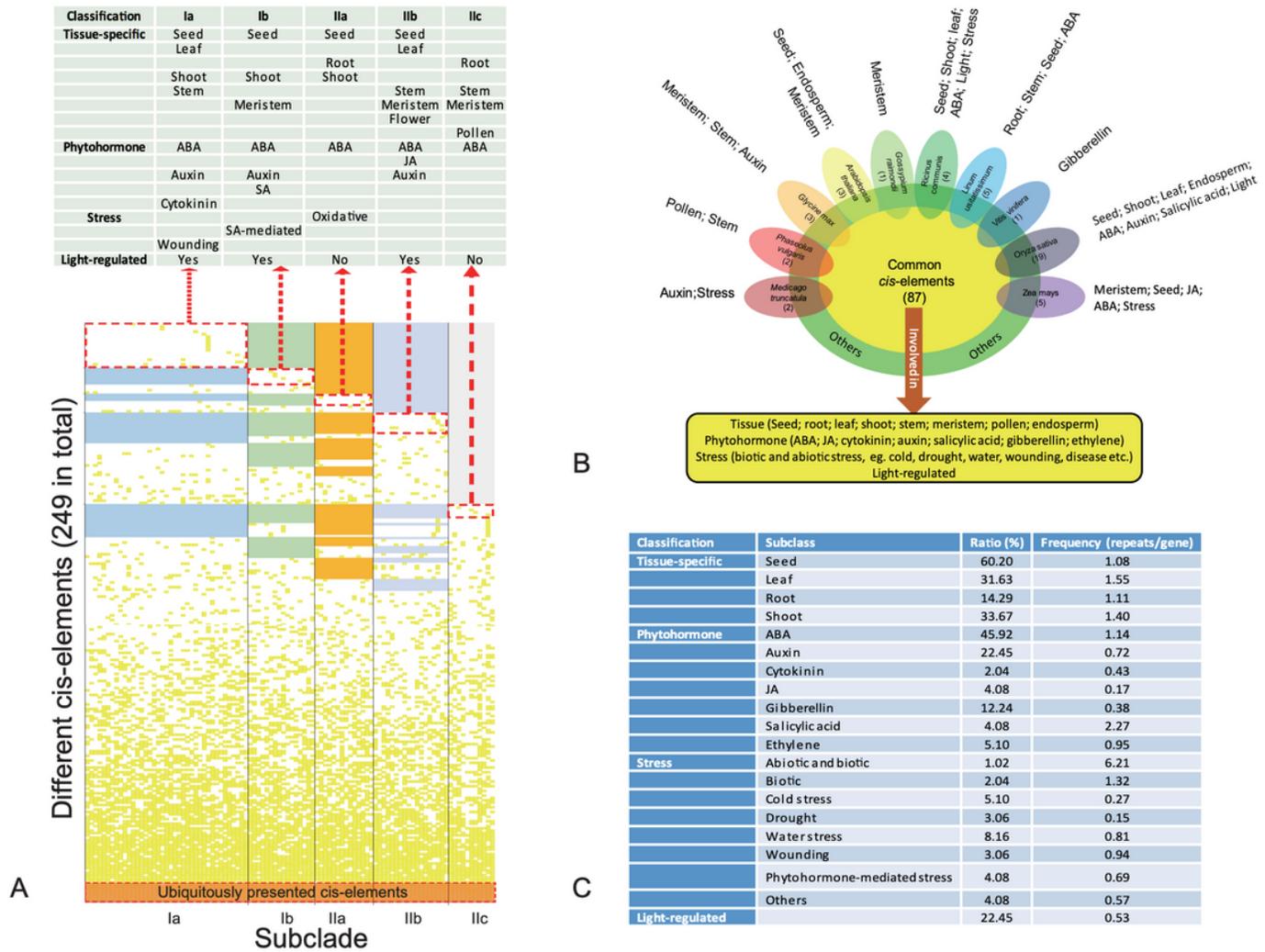


Figure 5

(A) Cis-elements of each ACP gene are displayed according to phylogenetic order, ubiquitously presented elements, clade specific elements were labeled in dashed rectangle. (B) Common cis-elements and species-specific cis-elements. Element numbers were showed in bracket. (C) Statistics of cis-element related to specific function.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFiles.zip](#)
- [Table1.xlsx](#)