Vegetable Research

ARTICLE

Open Access

https://doi.org/10.48130/VR-2023-0005
Vegetable Research **2023**, 3:5

Comprehensive analysis of CYP78A family genes reveals the involvement of *CYP78A5* and *CYP78A10* in fruit development in eggplant

Mengya Zhou[#], Liying Zhang[#], Shuangxia Luo, Lijun Song, Shuxing Shen, Jianjun Zhao[*], Qiang Li[*] and Xueping Chen[*]

College of Horticulture, State Key Laboratory of North China Crop Improvement and Regulation, Key Laboratory of Vegetable Germplasm Innovation and Utilization of Hebei, Collaborative Innovation Center of Vegetable Industry in Hebei, Hebei Agricultural University, Baoding, Hebei 071000, China

[#] These authors contributed equally: Mengya Zhou, Liying Zhang

[*] Corresponding authors, E-mail: jjz1971@aliyun.com; yylq@hebau.edu.cn; chenxueping@hebau.edu.cn

Abstract

The CYP78A family is a plant-specific family, members of which have been considered as promising targets for yield improvement due to their important roles in regulating organ size. Eggplant is an important vegetable cultivated worldwide. However, little information about the eggplant CYP78As (SmCYP78As) limits the potential utilization of SmCYP78As for crop improvement. In this study, we identified six CYP78A genes in the eggplant genome named *SmCYP78A5* to *SmCYP78A10* according to the phylogenetic relationships to Arabidopsis CYP78As. The phylogenetic analysis of CYP78As from eggplant, Arabidopsis, rice and tomato classified the 27 CYP78As into five clades. SmCYP78As were found in three of the five clades. This classification is consistently supported by their gene structures, domains and conserved motifs. Segmental duplication events were found to contribute to the expansion of the SmCYP78A family. Comparative syntenic analysis provided further insight into the phylogenetic relationships of CYP78A genes from the four plants. qRT-PCR analysis revealed that the expression of the six SmCYP78As was detected in at least one of the eight tissues, showing a tissue-specific pattern. Notably, SmCYP78A5 and SmCYP78A10 were highly expressed in developing ovaries, indicating the involvement of fruit development in eggplant. Co-expression clustering and GO enrichment analysis suggested that *SmCYP78A5* and *SmCYP78A10* regulate fruit development likely through different pathways. In addition, six transcription factors were identified as promising candidates that may directly bind promoters of *SmCYP78A5* and *SmCYP78A10*. This study provides a comprehensive overview of the SmCYP78As family, which would lay a foundation for further understanding of evolution and function of the SmCYP78A family.

Citation: Zhou M, Zhang L, Luo S, Song L, Shen S, et al. 2023. Comprehensive analysis of CYP78A family genes reveals the involvement of *CYP78A5* and *CYP78A10* in fruit development in eggplant. *Vegetable Research* 3:5 https://doi.org/10.48130/VR-2023-0005

INTRODUCTION

Cytochrome P450 (P450) is an important superfamily in plants, the members of which play important roles in a wide range of biochemical pathways to produce phytohormones, including auxin, brassinosteroids (BRs) and gibberellins (GAs), as well as secondary metabolites, such as phenylpropanoids and fatty acids[1,2]. Plant P450s were grouped into 11 clans in two categories: multi-family clans (CYP71, CYP72, CYP85, CYP86) and single-family clans (CYP51, CYP74, CYP97, CYP710, CYP711, CYP727, CYP746)[2,3].

The CYP78A family is one of the families in the CYP71 clan and a plant-specific gene family[3]. The first CYP78A gene, CYP78A5/KLUH, was characterized in Arabidopsis, which is a positive regulator of organ size by promoting cell proliferation. While loss-of-function of Arabidopsis *CYP78A5/KLUH* results in smaller leaves and floral organs, over-expression of *CYP78A5/KLUH* produces larger organs, including leaves, seeds and petals[4,5]. Recently, AtKLUH was shown to play positive roles in regulating leaf longevity and drought tolerance by promoting cytokinin signaling and proline metabolism[6]. In addition, *cyp78a5* mutants also exhibit reduced plastochron length and early flowering[7,8], indicating the pleiotropic roles of AtKLUH in regulating Arabidopsis growth and development. There are five other Arabidopsis CYP78A members, CYP78A6, CYP78A7, CYP78A8, CYP78A9 and CYP78A10. CYP78A7 and CYP78A5 play redundant roles in regulating plastochron length, leaf size and apical dominance[7,9]. CYP78A6 (EOD3) and CYP78A9 were shown to redundantly regulate seed size and leaf senescence[9,10]. The functions of CYP78A10 are still unknown.

The orthologs of AtCYP78A5 have been shown to regulate seed or fruit size in other plants. For example, increased expression of tomato *KLUH* leads to larger fruits and seeds by stimulating cell division[11–14]. Over-expression of maize *PLASTOCHRON1* (*ZmPLA1*) stimulates leaf growth by extending cell division duration, leading to increased seed yield[15]. Recent studies in wheat showed that constitutive overexpression of *TaCYP78A5* enhances grain weight by accumulating auxin[16,17]. In addition, soybean CYP78A10 and CYP78A72, sweet cherry CYP78A9 and rice GIANT EMBRYO (GE; CYP78A13) have also been demonstrated as key regulators of organ size[18–24], showing great potential in yield improvement.

Eggplant is an important vegetable crop that is cultivated worldwide. Eggplant is also known as an important medical plant due to its high phenolic and alkaloid contents, including

chlorogenic acid and acetylcholine (ACh), which can be used to treat human diseases, such as diabetes and high blood pressure[25,26]. With increasing worldwide population, the yield of eggplants is needed to be increased to meet the demands. Considering the potential of CYP78A family members in yield improvement, it would be useful to identify and characterize the CYP78A family in the eggplant genome. However, none of the CYP78A genes have been cloned or characterized in eggplant to date.

In the present study, the CYP78A gene family was identified in the eggplant genome. Detailed information of the SmCYP78As, including chromosomal locations, phylogenetic relationships, gene structures, conserved motifs, synteny and candidate transcription factors which might directly bind the promoter of SmCYP78As, were investigated. qRT-PCR was performed to analyze the tissue-specific expression patterns of the *SmCYP78A*s. Expression levels of *SmCYP78A*s in young flower buds and developing ovaries were also analyzed using RNA-seq data. Co-expression clustering and GO enrichment analysis were conducted to gain further insights into the functions of *SmCYP78A*s. The results of the present study will be helpful for further functional study of the SmCYP78A genes.

MATERIALS AND METHODS

Identification of CYP78A genes in eggplant

Amino acid sequences of Arabidopsis CYP78As were downloaded from The Arabidopsis Information Resource (TAIR) database (www.arabidopsis.org). The six AtCYP78A proteins were used as query sequences and searched against the eggplant proteins in the Eggplant Genome Database (http://eggplant-hq.cn/Eggplant/home/index). Then, the Pfam (http://pfam.janelia.org/) and Simple Modular Architecture Research Tool (SMART) (http://smart.embl-heidelberg.de/) programs were used to confirm the existence of the P450 domain. The number of amino acids, isoelectric points (PIs) and molecular weights (MWs) of the 27 CYP78A proteins were determined using ExPASy (https://web.expasy.org/protparam/), and their subcellular localizations were predicted with Cell-PLoc 2.0 (www.csbio.sjtu.edu.cn/bioinf/Cell-PLoc-2/).

Phylogenetic analysis, gene structure and conserved motif recognition

Multiple sequence alignments were performed and the phylogenetic trees were constructed with the full protein sequences of the 27 CYP78As from eggplant, Arabidopsis, rice and tomato using MEGA 7.0. The neighbor-joining (NJ) method was used for the construction of the phylogenetic tree with the following parameters: Poisson correction, pairwise deletion and 1000 bootstrap replicates. Gene structures were analyzed using TBtools[27]. Conserved motifs were identified using MEME (Multiple Expectation Maximization for Motif Elicitation)[28].

Chromosomal distribution and synteny analysis

All *SmCYP78A* genes were mapped to chromosomes based on physical location information from the Eggplant Genome Database using Circos[29]. Multiple Collinearity Scan toolkit (MCScanX) is employed for scanning multiple genomes to align syntenic blocks[30].

Plant materials

An inbred line '14-345' with round and purple black fruits was used in this study for gene expression analysis and RNA-seq. The plants were grown in the greenhouse at Hebei Agricultural University in Baoding (38° N, 115° E), China.

Real-time quantitative RT-PCR (qRT-PCR)

Roots, stems, leaves, young flower buds, petals and sepals of 0 DPA (Days post anthesis) flowers, pericarp of 0 DPA ovary, fruit flesh of 10 DPA fruits were collected separately from at least 2–3 individual plants as one biological replicate. Three biological replicates for each tissue sample were carried out for qRT-PCR. Total RNAs were extracted using TRIzol reagent (Invitrogen, USA) and treated with DNase I (Fermentas, Canada) following the manufacturer's protocol. First strand cDNA was synthesized from 1 μg total RNA using PrimeScript 1st Strand cDNA Synthesis Kit (TaKaRa, Japan). qRT-PCR was performed on LightCycler 96 (Roche, Switzerland) using the ChamQ Universal SYBR qPCR Master Mix (Vazyme, China). *Clathrin adaptor complexes medium subunit* (*CAC*) gene (*Smechr0800014*) was used as the housekeeping gene to normalize the gene expression. Primers used for qRT-PCR are listed in Supplemental Table S1. The relative expression level was calculated using the $2^{-\Delta\Delta C_T}$ method.

Generation and analysis of RNA-seq data

Young flower buds and developing ovaries, including 10 DBA (Days before anthesis), 7 DBA and 0 DPA, were collected with four biological replicates for each tissue sample. The cDNA library preparation and sequencing were conducted by the Novogene Bioinformatics Technology Company (Beijing, China) on the Illumina HiSeq TM4000 platform (Illumina Inc., San Diego, CA, USA). The read mapping was performed using the latest version of the Tuxedo protocol with HISAT2 and StringTie[31]. The clean reads from each library were aligned to the eggplant genome of 'HQ-1315' using HISAT2. The mapping results were normalized *via* Stringtie to compute TPM (Transcripts Per Kilobase Million) values of genes.

Co-expression genes were clustered using fuzzy C means in the Mfuzz package[32] and gene clusters were visualized by plotting of the normalized expression profiles of each cluster using ggplot2 package in R[33]. For GO enrichment analysis, eggplant geneIDs were converted to Arabidopsis geneIDs using BLASTP and then GO enrichment analysis was performed using clusterProfiler with cnetplot function in R[34]. The significant GO terms were identified with the adjusted value smaller than 0.05.

RESULTS

Genome-wide identification and analysis of CYP78A genes in eggplant

The six AtCYP78A proteins were employed as a query to search the SmCYP78As in the eggplant genome via BlastP program, resulting in six putative SmCYP78A genes. The presence of the P450 domains in the six putative SmCYP78A proteins was confirmed via Pfam and SMART, indicating that the six proteins are members of the eggplant CYP78A family (Table 1).

The information of the six *SmCYP78A*s, including chromosomal locations, amino acids number (length), PIs, MWs and predicted subcellular localizations (PSL), was listed in Table 1. To gain further insights into the CYP78A family genes in plants, the information of *CYP78A*s from Arabidopsis, rice and tomato was also included in Table 1. The amino acids number of

Table 1. Summary information of CYP78A family genes in eggplant, Arabidopsis, rice and tomato.

Gene ID	Gene name	Location			Deduced polypeptide			PSL
		Chr	Start	End	Length (aa)	MW (KDa)	pI	
Smechr0101302	CYP78A6	Chr1	12640581	12643245	538	60.35	8.90	ER
Smechr0302971	CYP78A5	Chr3	89886800	89889184	516	58.19	6.79	ER
Smechr0400048	CYP78A7	Chr4	420566	422920	525	58.98	9.23	ER
Smechr0500049	CYP78A8	Chr5	783272	785346	528	59.76	9.30	ER
Smechr0502180	CYP78A9	Chr5	74969262	74971177	537	60.55	7.52	ER
Smechr1100733	CYP78A10	Chr11	10534819	10536967	551	61.70	8.23	ER
AT1G01190	CYP78A8	Chr1	83045	84946	541	60.91	8.22	ER
AT1G13710	CYP78A5	Chr1	4702657	4704694	518	57.64	8.57	ER
AT1G74110	CYP78A10	Chr1	27866667	27868368	538	60.18	7.84	ER
AT2G46660	CYP78A6	Chr2	19153328	19155579	531	59.57	8.29	ER
AT3G61880	CYP78A9	Chr3	22905868	22907958	556	62.62	9.04	ER
AT5G09970	CYP78A7	Chr5	3111945	3114239	537	59.49	6.69	ER
LOC_Os10g26340	CYP78A11/PLA1	Chr10	13658790	13660543	556	59.08	7.06	ER
LOC_Os11g29720	CYP78A5	Chr11	17234285	17238178	539	59.64	10.19	ER
LOC_Os03g04190	CYP78A9	Chr3	1920043	1921896	516	55.80	8.10	ER
LOC_Os03g30420	CYP78A6/GL3.2	Chr3	17340415	17342284	516	56.05	9.00	ER
LOC_Os03g40600	CYP78A7	Chr3	22567670	22568685	194	20.29	7.99	ER
LOC_Os03g40610	CYP78A8	Chr3	22572706	22574008	308	32.84	10.12	ER
LOC_Os07g41240	CYP78A13/GE	Chr7	24713778	24715813	526	55.89	8.68	ER
LOC_Os08g43390	CYP78A15/BSR2	Chr8	27420501	27422836	552	59.79	9.38	ER
LOC_Os09g35940	CYP78A10	Chr9	20691306	20693116	554	60.74	9.07	ER
Solyc01g096280	CYP78A6	Chr1	79622266	79624618	539	61.05	8.60	ER
Solyc03g114940	CYP78A5/KLUH	Chr3	59217389	59219730	517	58.38	6.21	ER
Solyc05g015350	CYP78A8	Chr5	10475028	10479267	332	37.85	6.08	ER
Solyc05g047680	CYP78A7	Chr5	58506390	58508292	532	60.49	9.20	ER
Solyc10g009310	CYP78A9	Chr10	3224421	3227123	526	61.64	9.05	ER
Solyc12g056810	CYP78A10	Chr12	62510941	62512676	537	60.71	7.50	ER

SmCYP78A proteins varies from 516 (*SmCYP78A5*) to 551 (*SmCYP78A10*), the PI ranges from 6.79 (*SmCYP78A5*) to 9.30 (*SmCYP78A8*) and the MW ranges from 58.19 (*SmCYP78A5*) to 61.70 KDa (*SmCYP78A10*). Interestingly, the amino acids lengths of rice CYP78A7 and CYP78A8 as well as tomato CYP78A8 (SlCYP78A8) were much smaller than other CYP78As from Arabidopsis, rice, tomato and eggplant. It would be interesting to know whether the three short CYP78A proteins show similar functions with other CYP78As. Notably, the subcellular localizations of all CYP78As were predicted to be localized in endoplasmic reticulum (ER), which is in agreement with the subcellular localizations of AtCYP78A5 and TaCYP78A5 *in vivo*[4,16].

To elucidate the evolutionary relationships of SmCYP78As, an unrooted neighbor-joining (NJ) phylogenetic tree was constructed using the full protein sequences from the six AtCYP78As, nine OsCYP78As, six SlCYP78As and six SmCYP78As. The resulting tree contained five distinct clades (C1-C5) (Fig. 1a). Phylogenetic analysis revealed that C1 and C4 were shared in all the four species. There was equal number of the CYP78A proteins from eggplant, Arabidopsis, tomato and rice in C1 (Fig. 1a). C4 contained two SmCYP78As and one CYP78A from each of tomato, Arabidopsis and rice, indicating the expansion of SmCYP78As in this clade compared with the other three species. C2 and C3 didn't contain SmCYP78As. C2 was shared by rice and tomato, but not in Arabidopsis and eggplant, suggesting the unique roles of the CYP78As in C2 that were likely acquired or expanded in tomato and rice after divergence from the last common ancestor with eggplant and Arabidopsis. It is worth noting that C3 only contains rice CYP78As and OsCYP78A8 didn't fit into any clades, which may have evolved following divergence and have special roles in rice. Moreover, C5 didn't include any rice CYP78As but only members from eggplant, tomato and Arabidopsis, suggesting that the CYP78As in C5 may have been lost in rice during evolution.

Gene structure analysis showed that the number of exons in the 27 CYP78A genes was conserved and most of them contain two exons except *OsCYP78A7* and *AtCYP78A9*, which contains only one and three exons, respectively (Fig. 1b). In addition, the introns of the 27 CYP78As are a phase 0 intron (Fig. 1b), further suggesting the highly conservation of CYP78A genes during the evolution of the four plants.

Ten conserved motifs that are shared among the 27 CYP78A proteins were identified using the MEME (Fig. 1c; Supplemental Fig. S1). Twenty-four CYP78A proteins contain all 10 motifs with motif 6, 2, 9, 7 and 10 at N terminal and motif 1, 5, 3, 4 and 8 at C terminal (Fig. 1c), suggesting the similar function of these CYP78As. The other three CYP78A proteins with shortest amino acid length did not include some motifs (Fig. 1c). For example, SlCYP78A8 does not have motif 2, 6 and 10. While OsCYP78A7 does not include motif 1, 6, 2, 7, 9 and 10, OsCYP78A8 does not contain motif 1, 5, 3, 4 and 8. Further studies are required to investigate the roles of these motifs regarding the functions of CYP78As.

Synteny analysis of CYP78A genes from eggplant, Arabidopsis, tomato and rice

The six *SmCYP78A* genes were mapped on five chromosomes, i.e. E01, E03, E04, E05 and E11 (Fig. 2a). Interestingly, all the six *SmCYP78A* genes were located at the end of the five chromosomes. Similar locations of CYP78As were also found in Arabidopsis, tomato and rice genomes (Fig. 2). Syntenic

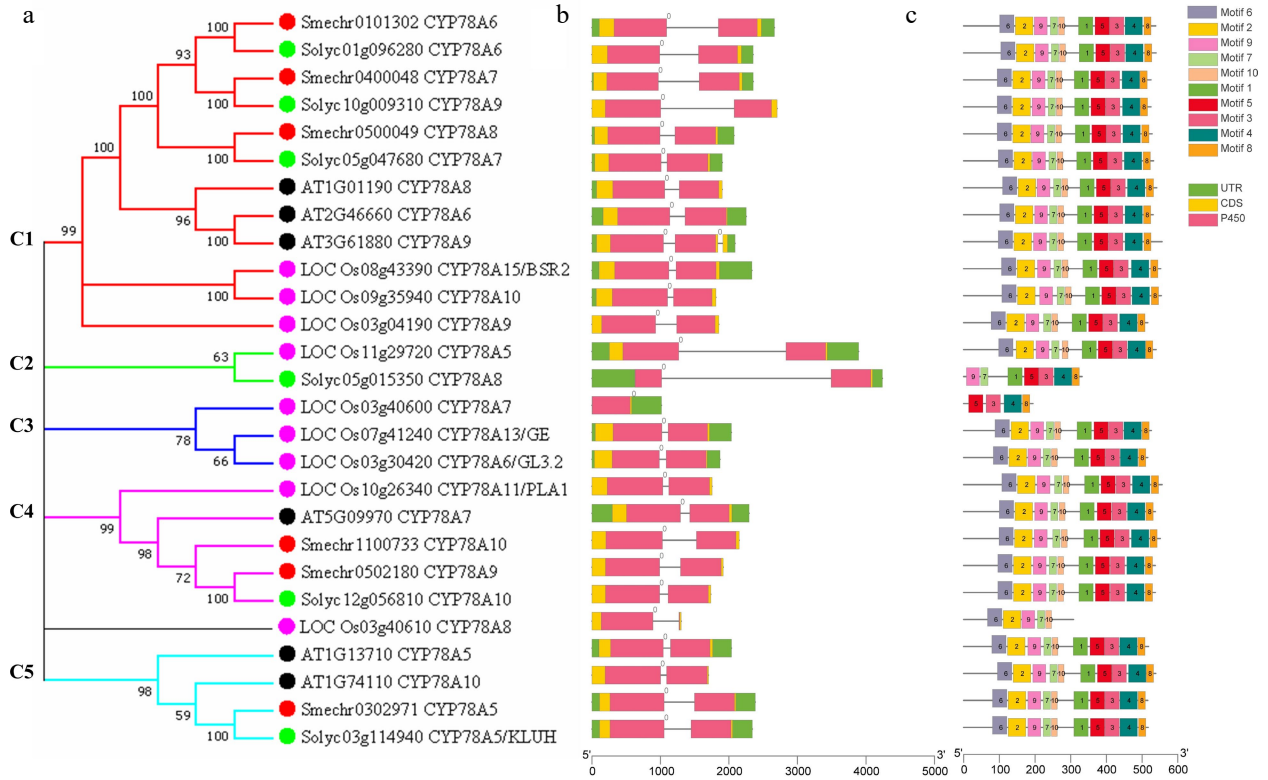

Fig. 1 Phylogenetic relationships, gene structure and conserved protein motifs of CYP78A genes from eggplant, Arabidopsis, rice and tomato. (a) The phylogenetic tree was constructed based on the full-length protein sequences of six AtCYP78As, nine OsCYP78As, six SlCYP78As and six SmCYP78As proteins using MEGA 7.0 software. Eggplant, Arabidopsis, rice and tomato CYP78As were labeled by red, black, pink and green dots. (b) Exon-intron structure of CYP78A genes. Black lines indicate introns. The number indicates the phases of corresponding introns. (c) The motif composition of CYP78A proteins. The motifs, numbers 1–10, are displayed in different colored boxes. The sequence logos and E values for each motif are given in Supplemental Fig. S1.

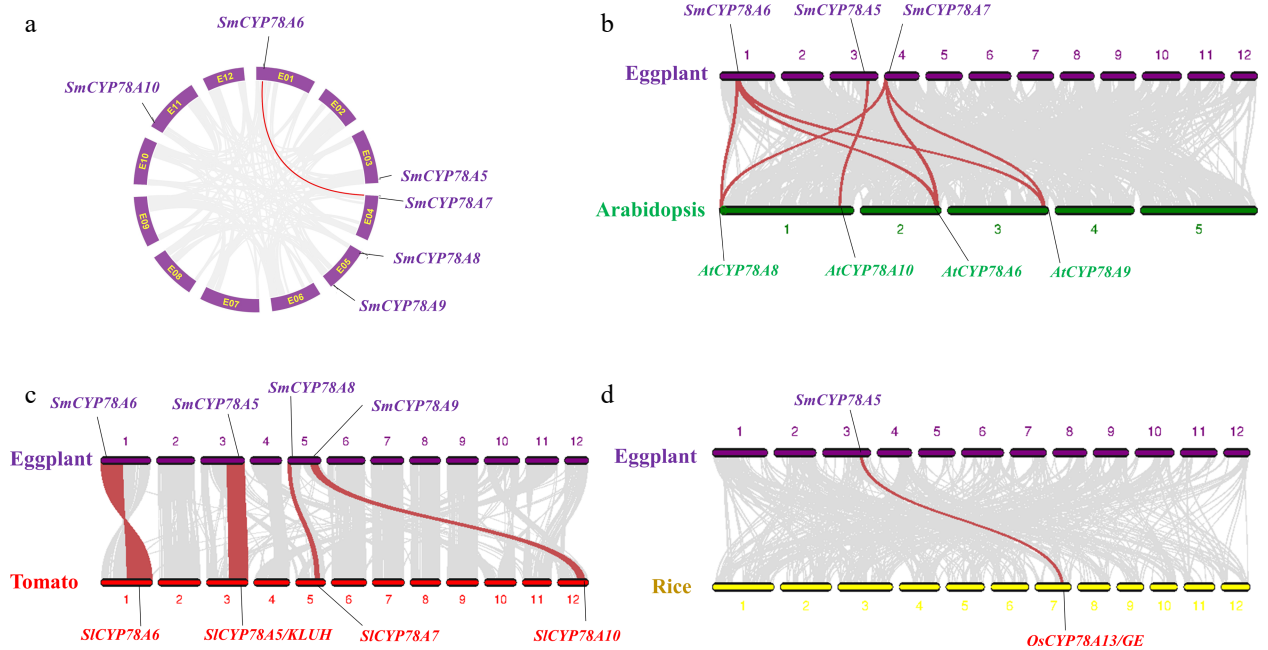

Fig. 2 Gene duplication and synteny analysis of *SmCYP78A* genes. (a) Schematic representations for the chromosomal distribution and interchromosomal relationships of *SmCYP78A* genes. Gray lines indicate all synteny blocks in the eggplant genome, and the red lines indicate segmental duplicated *SmCYP78A* gene pairs. (b) Synteny analysis of *CYP78A* genes between eggplant and Arabidopsis. (c) Synteny analysis of *CYP78A* genes between eggplant and tomato. (d) Synteny analysis of *CYP78A* genes between eggplant and rice. Gray lines in the background indicate the collinear blocks between genomes, while the red lines highlight the syntenic blocks harboring *CYP78A* gene pairs.

analysis of the eggplant genome were performed using MCscanX to identify duplication events among SmCYP78As. Only one gene pair, *SmCYP78A6* and *SmCYP78A7*, were identified in the eggplant genome, indicating that segmental duplication contributes to the expansion of the CYP78A family in eggplant.

Comparative syntenic analyses of eggplant genome were performed with genomes of Arabidopsis, tomato and rice. Three (*SmCYP78A5*, *SmCYP78A6* and *SmCYP78A7*), four (*SmCYP78A5*, *SmCYP78A6*, *SmCYP78A5*, *SmCYP78A8* and *SmCYP78A9*) and one (*SmCYP78A5*) SmCYP78A gene show syntenic relationships with those in Arabidopsis, tomato and rice, respectively (Fig. 2). Interestingly, *SmCYP78A6* and *SmCYP78A7* were syntenic with three Arabidopsis CYP78A genes (*AtCYP78A6*, *AtCYP78A8* and *AtCYP78A9*), respectively. Notably, *SmCYP78A5* showed a syntenic relationship with *AtCYP78A10*, *SlCYP78A5/KLUH* and *OsCYP78A13/GE*, indicating that these orthologous pairs likely have existed before the ancestral divergence with conserved functions.

Expression profiles for *SmCYP78A* genes in different tissues

Real-time quantitative RT-PCR were used to detect the expression patterns for the six *SmCYP78A* genes in the roots, stems, leaves, young flower buds, petals, sepals, pericarp and fruit flesh. The six *SmCYP78A* genes showed different patterns of tissue-specific expression and exhibited relatively low expression levels in most tissues (Fig. 3a). *SmCYP78A5*, *SmCYP78A7*, *SmCYP78A8*, *SmCYP78A9* and *SmCYP78A10* was specifically expressed in young flower buds, roots, petals, roots and stems, respectively (Fig. 3a). *SmCYP78A6* showed high levels of transcript abundance in roots and pericarp (Fig. 3a). The different expression patterns of the *SmCYP78A*s indicate their distinct roles in various aspects of physiological and developmental processes.

Considering the important roles of *CYP78A* genes in fruit development and the fact that fruit size was largely determined at the early developmental stages, we analyzed the RNA-seq data of young flower buds and developing ovaries in eggplant. The six *SmCYP78A* genes showed different expression patterns in developing ovaries (Fig. 3b). While *SmCYP78A5* and *SmCYP78A10* showed high expression, the other four *SmCYP78A* genes were barely expressed in young flower buds and developing ovaries (Fig. 3b). Moreover, *SmCYP78A5* showed highest expression in young flower buds and gradually decreased with the development of eggplant ovary and showed no expression at 0 DPA (Fig. 3b). Interestingly, *SmCYP78A5* showed similar expression patterns with tomato *KLUH*, the closest ortholog of *SmCYP78A5*, in developing ovaries[11], indicating their conserved roles in regulating fruit size. *SmCYP78A10* abundantly expressed in developing ovaries with the expression peak at 10 DBA and very low expression in 0 DPA (Fig. 3b).

Analysis of co-expression genes of *SmCYP78A5* and *SmCYP78A10*

The high expression of *SmCYP78A5* and *SmCYP78A10* in developing ovaries (Fig. 3b) indicated their important roles in regulating fruit development in eggplant. To gain further insight into the functions of *SmCYP78A5* and *SmCYP78A10*, co-expression analysis was performed using fuzzy C-means clustering. Twelve co-expressed clusters were identified with Cluster 6 and 11 representing *SmCYP78A10* and *SmCYP78A5*, respectively (Fig. 4; Supplemental Table S2).

Cluster 6 represented genes that expressed at higher levels in ovaries at 10 DBA and 7 DBA than young flower buds and ovaries at 0 DPA (Fig. 4). Cluster 6 was significantly enriched with genes involved in cellular processes, such as 'Cell cycle process', 'Organelle fission' and 'Microtubule-based process' (Fig. 5). Genes involved in these processes included *SmCYP78A10* and putative orthologs of Arabidopsis *SUN1*, *TON1*, *TUA6* and *NEK1* (Fig. 5). Genes in Cluster 11 showed highest expression in young flower buds and low expression in developing ovaries (Fig. 4). Cluster 11 was enriched with genes involved in photosynthesis related processes, including 'photosynthesis', 'carbon fixation' and 'response to high light intensity'. Genes involved in these processes included *NDF*s, *PRK* and *PPH1* (Fig. 5). The GO enrichment analysis indicated

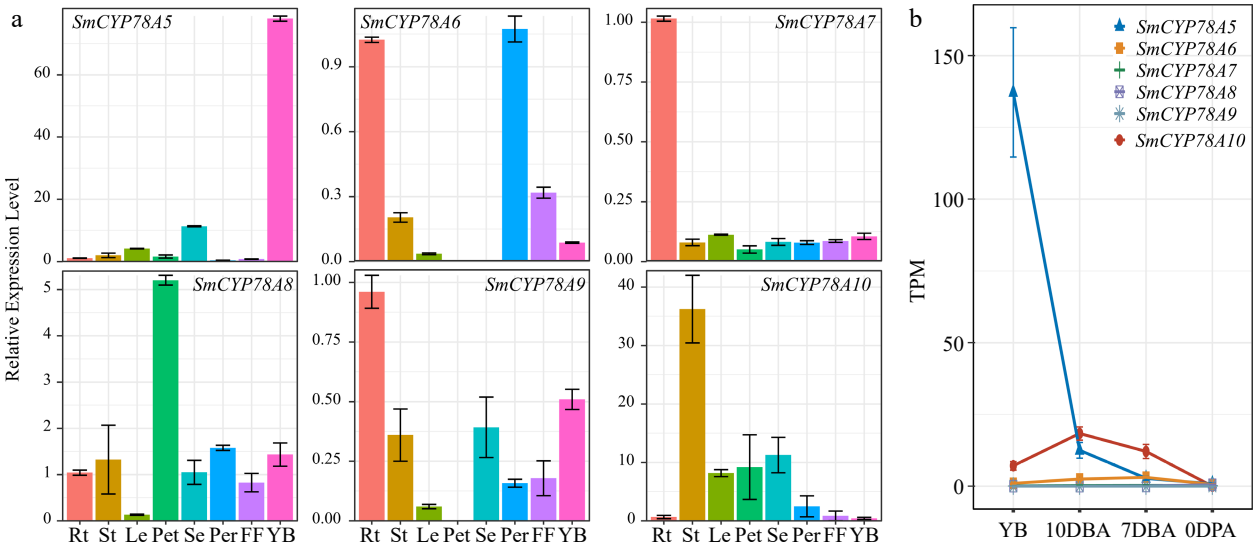

Fig. 3 Expression profiles of the six *SmCYP78A* genes in different tissues. (a) Relative transcript abundances of the *SmCYP78A* genes examined by qRT-PCR. (b) Expression of the *SmCYP78A* genes in young flower buds and developing ovaries detected by RNA-seq. Rt, Root; St, Stems; Le, Leaf; Pet, Petal; Se, Sepal; Per, Pericarp; FF, Fruit flesh; YB, Young flower buds; DBA, Days before anthesis; DPA, Days post anthesis.

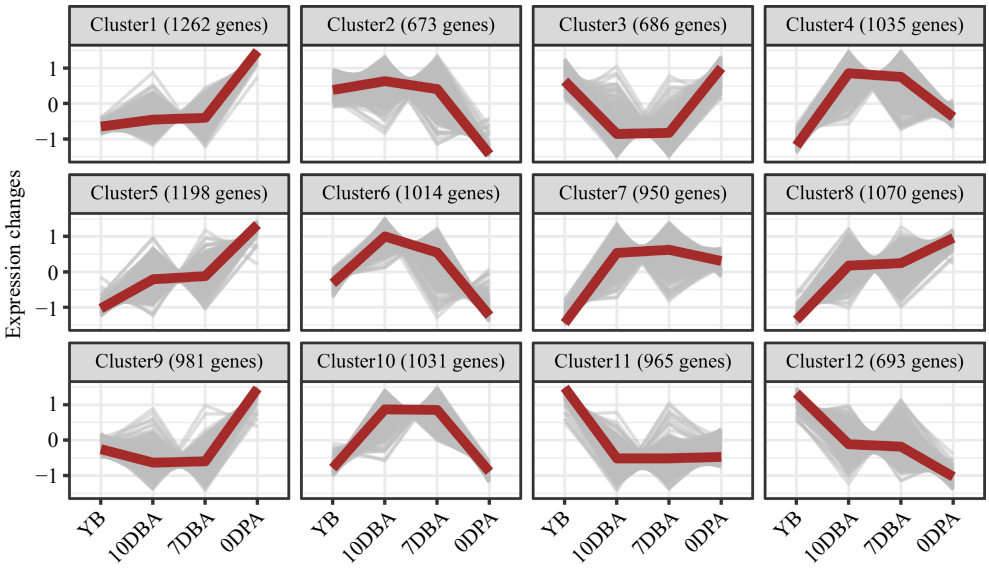

Fig. 4 Twelve co-expressed clusters are clustered using fuzzy C-means clustering in Mfuzz with normalized expression values (z-scores). The red lines represent the average of expression values, whereas the gray lines represent the expression values of the co-expressed genes. YB, Young flower buds; DBA, Days before anthesis; DPA, Days post anthesis.

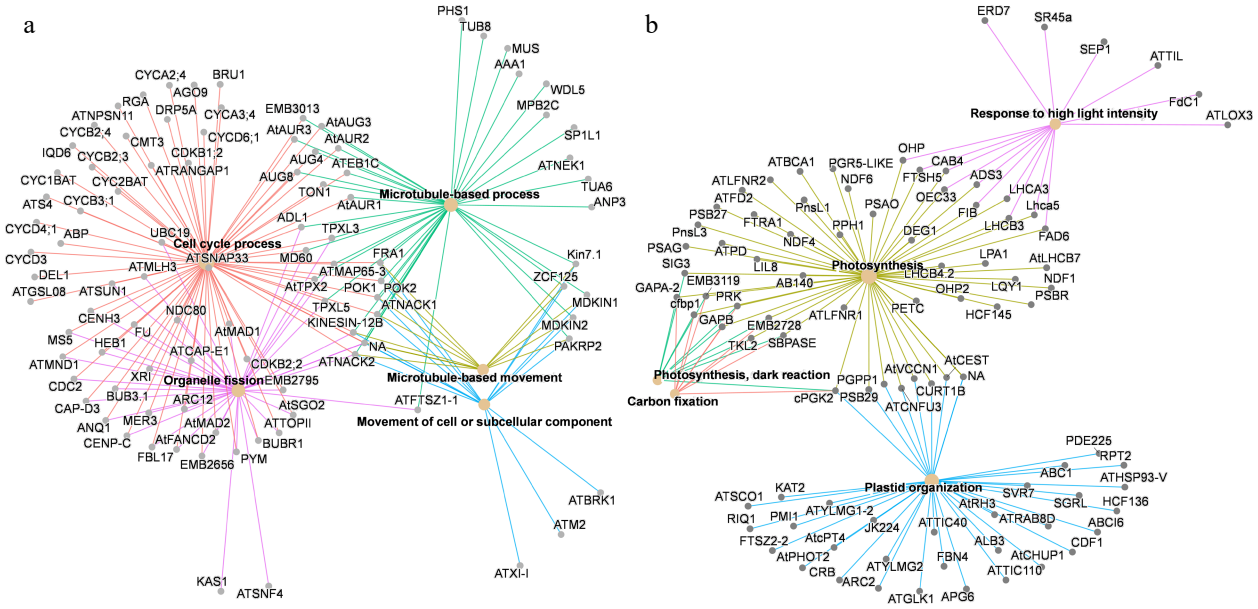

Fig. 5 Significantly enriched GO terms (biological process) of co-expression genes in (a) Cluster 6 and (b) Cluster 11. Only the top five enriched GO terms are shown. The color of lines represents different GO terms.

that *SmCYP78A10* and *SmCYP78A5* regulate fruit development likely through different mechanisms.

Since transcription factors (TFs) are the main regulators of gene expression, we sought out the TFs in the two clusters. Cluster 6 harbored 77 TFs (7.60%) which were classified into 29 families (Fig. 6a; Supplemental Table S3). The 10 most abundant TF families in cluster 6 were HB (8), GRAS (6), MYB (5), bHLH (5), B3 (5), ERF (4), zf-HD (3), NAC (3), MYB-related (3) and GRF (3) (Fig. 6a; Supplemental Table S3). The Cluster 11 contained 63 TFs (6.52%) mainly from families classified as HB (8), bHLH (5), MIKC (4), MYB (4), NF-YA (4), bZIP (3), C2C2-CO-like (3), C2C2-YABBY (3), C3H (3) and HSF (3) (Fig. 6b; Supplemental Table S4). Interestingly, HB, MYB and bHLH TFs were found in both Cluster 6 and 11, suggesting that HB, MYB and bHLH TFs might play important roles in regulating the expression of *CYP78A*s in eggplant.

Transcription factor binding site analysis in the promoters of *SmCYP78A*s

To gain further insight into the transcriptional regulation of the *SmCYP78A5* and *SmCYP78A10*, we selected a 1.5 kb regulatory region upstream of the ATG of *SmCYP78A5* and *SmCYP78A10* (Supplemental Table S5) to scan transcription factor binding sites (TFBSs) using PlantRegMap. Interestingly, two HB TFs, *Smechr0402062* and *Smechr0101299*, that are co-expressed with *SmCYP78A5* in Cluster 11 were predicted to directly target *SmCYP78A5* (Table 2). *SmCYP78A10* was identified as candidate target of *Smechr0402092* (AP2), *Smechr0902218* (Cysteine-rich polycomb-like protein, CPP),

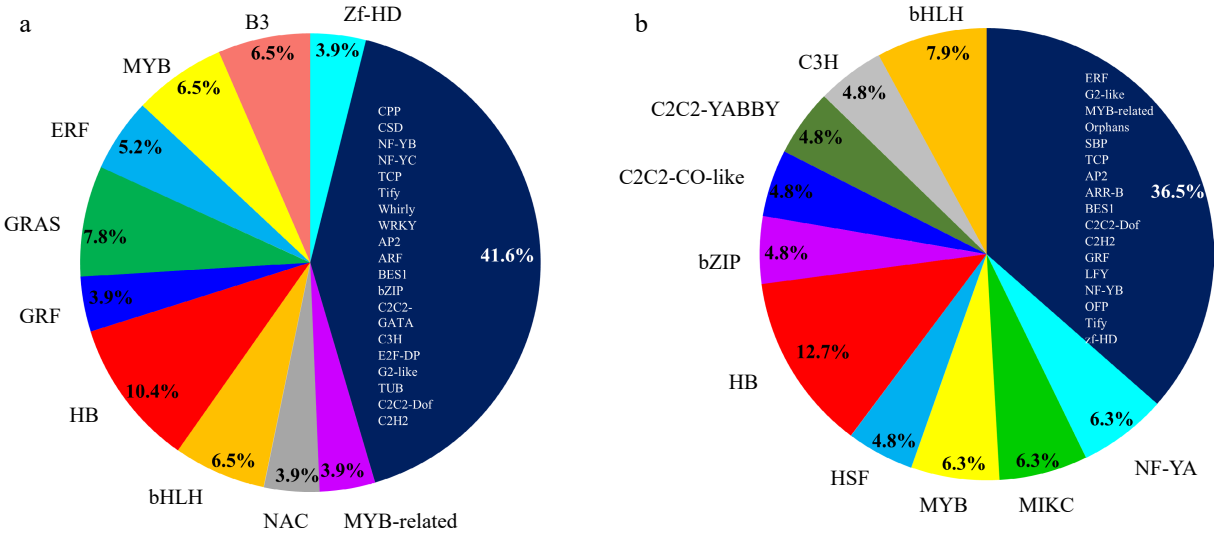

Fig. 6 Overview of distribution of TF families that were co-expressed with (a) *SmCYP78A10* in Cluster 6 and (b) *SmCYP78A5* in Cluster 11. The Plant Transcription Factor Database v5.0 (http://planttfdb.gao-lab.org) was used to identify TFs in the eggplant genome.

Table 2. Candidate transcription factors binding promoters of SmCYP78As identified by PlantRegMap.

Gene ID	TF family	Arabidopsis ortholog	Binding sequence	Strand	P value
Smechr0402062	HB	AT4G08150	CACTTCCCTTCTCTCTCTCT	+	1.71E-05
Smechr0101299	HB	AT2G46680	TCATTTATTGAAC	–	9.07E-05
			GGAATGATTGTAA	–	9.88E-05
Smechr0402092	AP2	AT4G37750	CATCACAAATTCCAAAATCCC	+	2.73E-05
			AAACACTCTCCCCCACGTATA	–	7.73E-05
Smechr0902218	CPP	AT4G14770	TAAAATTTTAAAA	–	7.34E-05
			TGAAATTTAAAAA	–	8.37E-05
			TCAAATTTAAAAA	+	8.47E-05
Smechr0801604	MYB		CTTGAAGACCGTTGA	+	9.42E-05
Smechr0201168	TCP	AT3G27010	TTGCCCCAC	+	5.27E-05

Smechr0801604 (MYB) and *Smechr0201168* (TCP) that are co-expressed genes of *SmCYP78A10* in Cluster 6. Some orthologs of the TFs were known from other studies to be involved in organ size regulation in plants. For example, AINTEGUMENTA (ANT) is an ortholog of Smechr0402092 in Arabidopsis and has been demonstrated as a positive organ size regulator by stimulating cell proliferation and modulating auxin biosynthesis[35,36]. *Smechr0902218* encodes a CPP TF and is closely related to Arabidopsis *TCX2/SOL2* that has been reported to regulate both cell fate and cell division[37,38]. Smechr0201168 is a putative ortholog of Arabidopsis TCP20 which has been proposed to control cell division and growth by directly binding to the GCCCR element in the promoters of cyclin CYCB1;1[39]. In addition, studies from Arabidopsis have shown the important roles of HB TFs in regulating organ size[40,41]. Therefore, the TFs may function as regulators of eggplant fruit development by directly binding the promoters of *CYP78As*.

DISCUSSION

The members of P450 superfamily encodes enzymes presenting in the kingdoms of life with functional diversity[2]. CYP78A is a plant-specific P450 family and has been well studied in Arabidopsis, in which CYP78A genes play important roles in growth and development, including plastochron and organ size[9]. However, no related information has been reported in eggplant. In the present study, we identified six CYP78A members in eggplant genome (Table 1), which is same to the number of CYP78A genes in Arabidopsis and tomato. However, compared to rice, the size of CYP78A families was small in Arabidopsis, tomato and eggplant. Considering the number of CYP78A family members was not correlated with genome size, suggesting differential expansion events occurred during the evolution of the CYP78A family between rice and Arabidopsis, tomato and eggplant. Gene duplication events contribute to the expansion and evolution of gene families[42] and one segmental duplication event (*SmCYP78A6* and *SmCYP78A7*) was identified in eggplant genome (Fig. 2a), which may result from the ancient whole genome duplication (WGD) in eggplant before the divergence of asterids and rosids[43].

Although CYP78A genes have been reported in many plant species, genome-wide identification of CYP78A family were only performed in Arabidopsis[9]. To better understand the phylogenetic relationships of CYP78As in eggplant with those of model plant species, such as Arabidopsis, tomato and rice, we also identified CYP78A genes in the genomes of tomato and rice (Table 1). Phylogenetic trees combining eggplant, Arabidopsis, tomato and rice CYP78As were constructed, which divided the 27 CYP78As into five clades and the six SmCYP78A members into three clades (Fig. 1a). Interestingly, lineage-specific gene loss and expansions were observed in some clades, indicating the divergence of the clades during the evolution of the species. For example, C2 and C3 did not

include any eggplant CYP78As, suggesting that the two clades were either lost in eggplant or were acquired in rice and tomato after divergence from the last common ancestor. Similar reasons could explain why none of the rice CYP78As were found in C5 (Fig. 1a).

To further obtain insight of the evolutionary relationship and diversity/conservativeness of CYP78A genes in eggplant, Arabidopsis, tomato and rice, the gene structure, P450 domain and motif analyses were also conducted (Fig. 1b & c). The results showed that CYP78A members from the four species showed similar gene structure, P450 domain and motif compositions. In addition, all the CYP78A proteins were proposed to be ER-localized (Table 1). These results indicated that the CYP78A family was highly conserved in plants. More importantly, many studies from other plants, such as rice, tomato and maize, also suggested the highly conserved roles of CYP78As in plants[11,15,44].

Notably, phylogenetic analysis showed that SmCYP78A5 was grouped into the same clade with AtCYP78A5, AtCYP78A10 and SlCYP78A5/KLUH (Fig. 1a). Moreover, syntenic analysis indicated that SmCYP78A5 was an ortholog of AtCYP78A10, SlCYP78A5/KLUH and OsCYP78A13/GE (Fig. 2), which have been identified as key positive organ size regulators in Arabidopsis, tomato and rice, respectively. Therefore, it is reasonable to hypothesize that SmCYP78A5 may positively control organ size in eggplant. SmCYP78A6, SmCYP78A7 and SmCYP78A8 were classified into the same clade with AtCYP78A6, AtCYP78A8 and AtCYP78A9 (Fig. 1a). Furthermore, SmCYP78A6 and SmCYP78A7 were syntenic orthologs of AtCYP78A6, AtCYP78A8 and AtCYP78A9 (Fig. 2b). AtCYP78A6 and AtCYP78A9 have been found to redundantly regulate seed size and leaf senescence, whereas AtCYP78A8 was shown to be involved in seed color regulation but not leaf senescence[9,10,45]. It would be interesting to know whether SmCYP78A6, SmCYP78A7 and SmCYP78A8 play roles in the regulation of organ size, seed color and leaf senescence. Phylogenetic analysis indicated the closest relationships between SmCYP78A9 and SmCYP78A10 with AtCYP78A7 and OsPLA1 (Fig. 1a). Loss of function of *AtCYP78A7* and *OsPLA1* led to short plastochron in Arabidopsis and rice, respectively[9,46], indicating the important roles of SmCYP78A9 and SmCYP78A10 in plastochron regulation in eggplant, which needs to be confirmed in future studies. In addition, we found that *SmCYP78A6*, *7*, *8* and *9* were barely expressed in developing ovaries but highly expressed in other tissues, such as roots and petals, indicating that they may play important roles in these tissues.

Tomato *KLUH* underlies fruit weight QTL *fw3.2* and is highly expressed in vegetative meristems and young flower buds[11]. Our previous study indicated that tomato KLUH positively regulates fruit size by promoting cell proliferation in the pericarp at the early stages of fruit development, which has been proposed to be associated with lipid metabolism and photosynthesis[12]. In sweet cherry, *PaCYP78A9* showed highest expression in floral organs and was shown to regulate fruit size by affecting cell proliferation and expansion in pericarp. Interestingly, *SmCYP78A5* showed similar expression pattern with *SlCYP78A5*/*KLUH* and *PaCYP78A9* (Fig. 3b)[11,12]. Moreover, gene co-expression analysis in the present study suggested the tight links between SmCYP78A5 and photosynthesis, further supporting the conserved roles of SmCYP78A5 and SlKLUH in regulating fruit development. In the present study, although *SmCYP78A5* and *SmCYP78A10* had high expression in developing ovaries, they showed different expression patterns. Moreover, GO enrichment of the co-expression genes of *SmCYP78A5* and *SmCYP78A10* suggested that the two genes regulate fruit development through different pathways. These results indicated the functional divergence of eggplant CYP78A family members. Over-expression and knock-out of *SmCYP78A5* and *10* were required to confirm their roles in regulating organ size, especially in fruit size or length in eggplant.

It has been shown that AtKLUH regulate organ size in non-cell autonomous manner by producing a mobile molecule[4,5,9]. Several lines of evidence suggest that the mobile signal might be fatty acid-derived molecules[5,12,47,48]. A recent study in Arabidopsis showed that AtKLUH promotes organ growth as well as drought tolerance mainly by affecting cytokinin signaling and proline metabolism[6]. However, the studies from maize, wheat and rapeseed indicated the involvement of CYP78As in auxin metabolism[49]. Therefore, more direct evidence is needed to determine the mobile signal generated by CYP78As.

Identification of TFs that directly target *CYP78A*s would be helpful to gain further insight into the transcriptional regulation of the pathway mediated by CYP78As in eggplant. In Arabidopsis, *suppressor of da1-1* (*SOD7*) encodes B3 domain transcription factor NGATHA-like protein 2 (NGAL2). SOD7/NGAL2 directly binds the promoter of *KLUH* to repress its expression and thereby negatively regulates seed size[50]. In sweet cherry, AGAMOUS-LIKE transcription factor PavAGL15 regulates fruit size by directly repressing the expression of *PavCYP78A9*[22]. In this study, several putative TFs binding to the promoters of *SmCYP78A*s were identified through bioinformatic prediction and co-expression analysis (Fig. 4; Fig. 6; Table 2). More interestingly, the orthologs of these TFs in Arabidopsis, including ANT, TCX2/SOL2 and TCP20, have been shown to be associated with organ size regulation[35–41]. However, the involvement of these TFs in fruit development needs to be confirmed in eggplant and further studies, such as Yeast One Hybrid and Electrophoretic Mobility Shift Assay (EMSA), are required to dissect the relationships between the TFs and *SmCYP78A*s.

CONCLUSIONS

In this work, we identified six CYP78A family genes in the eggplant genome and provided comprehensive analysis of CYP78A genes from eggplant, Arabidopsis, rice and tomato. The results indicated the close evolutionary relationship and functional conservation of CYP78A genes in plants. The high expression of *SmCYP78A5* and *SmCYP78A10* in young flower buds and developing ovaries suggested their important roles in controlling fruit development. Co-expression clustering, GO enrichment analysis and TF binding site analysis indicated the different mechanisms underlying fruit development regulation between *SmCYP78A5* and *SmCYP78A10* and identified six potential upstream TFs that directly bind to the promoters of *SmCYP78A5* and *SmCYP78A10*.

ACKNOWLEDGMENTS

This work was supported by Natural Science Foundation of Hebei Province (C2021204015), 2021 Project for the Introduc-

tion of Oversea Students in Hebei (C20210510), Science and technology research projects of colleges and universities in Hebei Province (ZD2022111), the Introduction of Talents Start-up fund of State Key Laboratory of North China Crop Improvement and Regulation (NCCIR2020RC-13), the Introduction of Talents Start-up fund of Hebei Agricultural University (YJ2020067), Hebei Fruit Vegetables Seed Industry Science and Technology Innovation Team Project (21326309D) and Vegetable Innovation Team Project of Hebei Modern Agricultural Industrial Technology System (HBCT2018030203).

Conflict of interest

Jianjun Zhao is an Editorial Board member of the journal Vegetable Research. He was blinded from reviewing or making decisions on the manuscript. The article was subject to the journal's standard procedures, with peer-review handled independently of the Editorial Board member and his research group.

Supplementary Information accompanies this paper at (https://www.maxapress.com/article/doi/10.48130/VR-2023-0005)

Dates

Received 19 November 2022; Accepted 27 December 2022; Published online 14 February 2023

REFERENCES

1. Mizutani M, Ohta D. 2010. Diversification of P450 genes during land plant evolution. *Annual Review of Plant Biology* 61:291–315
2. Hansen CC, Nelson DR, Møller BL, Werck-Reichhart D. 2021. Plant cytochrome P450 plasticity and evolution. *Molecular Plant* 14:1244–65
3. Nelson D, Werck-Reichhart D. 2011. A P450-centric view of plant evolution. *The Plant Journal* 66:194–211
4. Adamski NM, Anastasiou E, Eriksson S, O'Neill CM, Lenhard M. 2009. Local maternal control of seed size by *KLUH*/*CYP78A5*-dependent growth signaling. *PNAS* 106:20115–20
5. Anastasiou E, Kenz S, Gerstung M, MacLean D, Timmer J, et al. 2007. Control of plant organ size by *KLUH*/*CYP78A5*-dependent intercellular signaling. *Developmental Cell* 13:843–56
6. Jiang L, Yoshida T, Stiegert S, Jing Y, Alseekh S, et al. 2021. Multi-omics approach reveals the contribution of KLU to leaf longevity and drought tolerance. *Plant Physiology* 185:352–68
7. Wang J, Schwab R, Czech B, Mica E, Weigel D. 2008. Dual effects of miR156-targeted SPL genes and *CYP78A5*/*KLUH* on plastochron length and organ size in Arabidopsis thaliana. *The Plant Cell* 20:1231–43
8. Poretska O, Yang S, Pitorre D, Poppenberger B, Sieberer T. 2020. AMP1 and CYP78A5/7 act through a common pathway to govern cell fate maintenance in *Arabidopsis thaliana*. *PLoS Genetics* 16:e1009043
9. Nobusawa T, Kamei M, Ueda H, Matsushima N, Yamatani H, et al. 2021. Highly pleiotropic functions of CYP78As and AMP1 are regulated in non-cell-autonomous/organ-specific manners. *Plant Physiology* 186:767–81
10. Fang W, Wang Z, Cui R, Li J, Li Y. 2012. Maternal control of seed size by *EOD3*/*CYP78A6* in *Arabidopsis thaliana*. *The Plant Journal* 70:929–39
11. Chakrabarti M, Zhang N, Sauvage C, Muños S, Blanca J, et al. 2013. A cytochrome P450 regulates a domestication trait in cultivated tomato. *PNAS* 110:17125–30
12. Li Q, Chakrabarti M, Taitano NK, Okazaki Y, Saito K, et al. 2021. Differential expression of *SlKLUH* controlling fruit and seed weight is associated with changes in lipid metabolism and photosynthesis-related genes. *Journal of Experimental Botany* 72:1225–44
13. Li Q, Feng Q, Snouffer A, Zhang B, Rodríguez GR, et al. 2022. Increasing fruit weight by editing a *cis*-regulatory element in tomato *KLUH* promoter using CRISPR/Cas9. *Frontiers in Plant Science* 13:879642
14. Alonge M, Wang X, Benoit M, Soyk S, Pereira L, et al. 2020. Major impacts of widespread structural variation on gene expression and crop improvement in tomato. *Cell* 182:145–161.e23
15. Sun X, Cahill J, Van Hautegem T, Feys K, Whipple C, et al. 2017. Altered expression of maize *PLASTOCHRON1* enhances biomass and seed yield by extending cell division duration. *Nature Communications* 8:14752
16. Guo L, Ma M, Wu L, Zhou M, Li M, et al. 2022. Modified expression of *TaCYP78A5* enhances grain weight with yield potential by accumulating auxin in wheat (*Triticum aestivum* L.). *Plant Biotechnology Journal* 20:168–82
17. Zhou M, Peng H, Wu L, Li M, Guo L, et al. 2022. *TaKLU* plays as a time regulator of leaf growth via auxin signaling. *International Journal of Molecular Sciences* 23:4219
18. Wang X, Li Y, Zhang H, Sun G, Zhang W, et al. 2015. Evolution and association analysis of *GmCYP78A10* gene with seed size/weight and pod number in soybean. *Molecular Biology Reports* 42:489–96
19. Dai AH, Yang SX, Zhou HK, Tang KQ, Li G, et al. 2018. Evolution and expression divergence of the *CYP78A* subfamily genes in soybean. *Genes* 9:611
20. Zhao B, Dai A, Wei H, Yang S, Wang B, et al. 2016. *Arabidopsis KLU* homologue *GmCYP78A72* regulates seed size in soybean. *Plant Molecular Biology* 90:33–47
21. Qi X, Liu C, Song L, Li Y, Li M. 2017. *PaCYP78A9*, a cytochrome P450, regulates fruit size in sweet cherry (Prunus avium L.). *Frontiers in Plant Science* 8:2076
22. Dong Y, Qi X, Liu C, Song L, Ming L. 2022. A sweet cherry AGAMOUS-LIKE transcription factor *PavAGL15* affects fruit size by directly repressing the *PavCYP78A9* expression. *Scientia Horticulturae* 297:110947
23. Nagasawa N, Hibara KI, Heppard EP, Vander Velden KA, Luck S, et al. 2013. *GIANT EMBRYO* encodes CYP78A13, required for proper size balance between embryo and endosperm in rice. *The Plant Journal* 75:592–605
24. Yang W, Gao M, Yin X, Liu J, Xu Y, et al. 2013. Control of rice embryo development, shoot apical meristem maintenance, and grain yield by a novel cytochrome p450. *Molecular Plant* 6:1945–60
25. Yarmohammadi F, Ghasemzadeh Rahbardar M, Hosseinzadeh H. 2021. Effect of eggplant (*Solanum melongena*) on the metabolic syndrome: A review. *Iranian Journal of Basic Medical Sciences* 24:420–27
26. Ogunsuyi OB, Olagoke OC, Afolabi BA, Loreto JS, Ademiluyi AO, et al. 2022. Effect of *Solanum* vegetables on memory index, redox status, and expressions of critical neural genes in *Drosophila melanogaster* model of memory impairment. *Metabolic Brain Disease* 37:729–41
27. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, et al. 2020. TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Molecular Plant* 13:1194–202
28. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, et al. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research* 37:W202–W208
29. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, et al. 2009. Circos: an information aesthetic for comparative genomics. *Genome Research* 19:1639–45
30. Wang Y, Tang H, DeBarry JD, Tan X, Li J, et al. 2012. *MCScanX*: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research* 40:e49

31. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. 2016. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Protocols* 11:1650
32. Futschik ME, Carlisle B. 2005. Noise-robust soft clustering of gene expression time-course data. *Journal of Bioinformatics and Computational Biology* 3:965–88
33. Wickham H. 2016. ggplot2: elegant graphics for data analysis. New York: Springer. https://doi.org/10.1007/978-0-387-98141-3
34. Yu G, Wang LG, Han Y, He QY. 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics:a journal of integrative biology* 16:284–7
35. Randall RS, Sornay E, Dewitte W, Murray JAH. 2015. *AINTEGUMENTA* and the D-type cyclin CYCD3;1 independently contribute to petal size control in *Arabidopsis*: evidence for organ size compensation being an emergent rather than a determined property. *Journal of Experimental Botany* 66:3991–4000
36. Li YJ, Yu Y, Liu X, Zhang XS, Su YH. 2021. The Arabidopsis MATERNAL EFFECT EMBRYO ARREST45 protein modulates maternal auxin biosynthesis and controls seed size by inducing *AINTEGUMENTA*. *The Plant Cell* 33:1907–26
37. Simmons AR, Davies KA, Wang W, Liu Z, Bergmann DC. 2019. SOL1 and SOL2 regulate fate transition and cell divisions in the *Arabidopsis* stomatal lineage. *Development* 146:dev171066
38. Noh M, Shin JS, Hong JC, Kim SY, Shin JS. 2021. Arabidopsis TCX8 functions as a senescence modulator by regulating *LOX2* expression. *Plant Cell Reports* 40:677–89
39. Li C, Potuschak T, Colón-Carmona A, Gutiérrez RA, Doerner P. 2005. Arabidopsis TCP20 links regulation of growth and cell division control pathways. *PNAS* 102:12978–83
40. Bueso E, Muñoz-Bertomeu J, Campos F, Brunaud V, Martínez L, et al. 2014. *ARABIDOPSIS THALIANA HOMEOBOX25* uncovers a role for Gibberellins in seed longevity. *Plant physiology* 164:999–1010
41. Vigeolas H, Hühn D, Geigenberger P. 2011. Nonsymbiotic hemoglobin-2 leads to an elevated energy state and to a combined increase in polyunsaturated fatty acids and total oil content when overexpressed in developing seeds of transgenic Arabidopsis plants. *Plant physiology* 155:1435–44
42. Leister D. 2004. Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance genes. *Trends in genetics* 20:116–22
43. Wei Q, Wang J, Wang W, Hu T, Hu H, et al. 2020. A high-quality chromosome-level genome assembly reveals genetics for important traits in eggplant. *Horticulture Research* 7:153
44. Mimura M, Itoh JI. 2014. Genetic interaction between rice *PLASTOCHRON* genes and the gibberellin pathway in leaf development. *Rice* 7:25
45. Sotelo-Silveira M, Cucinotta M, Chauvin AL, Chávez Montes RA, Colombo L, et al. 2013. Cytochrome P450 *CYP78A9* is involved in Arabidopsis reproductive development. *Plant Physiology* 162:779–99
46. Miyoshi K, Ahn BO, Kawakatsu T, Ito Y, Itoh JI, et al. 2004. *PLASTOCHRON1*, a timekeeper of leaf initiation in rice, encodes cytochrome P450. *PNAS* 101:875–80
47. Imaishi H, Matsuo S, Swai E, Ohkawa H. 2000. CYP78A1 preferentially expressed in developing inflorescences of *Zea mays* encoded a cytochrome P450-dependent lauric acid 12-monooxygenase. *Bioscience, Biotechnology, and Biochemistry* 64:1696–701
48. Kai K, Hashidzume H, Yoshimura K, Suzuki H, Sakurai N, et al. 2009. Metabolomics for the characterization of cytochromes P450-dependent fatty acid hydroxylation reactions in *Arabidopsis*. *Plant Biotechnology* 26:175–82
49. Shi L, Song J, Guo C, Wang B, Guan Z, et al. 2019. A CACTA-like transposable element in the upstream region of *BnaA9.CYP78A9* acts as an enhancer to increase silique length and seed weight in rapeseed. *The Plant Journal* 98:524–39
50. Zhang Y, Du L, Xu R, Cui R, Hao J, et al. 2015. Transcription factors SOD7/NGAL2 and DPA4/NGAL3 act redundantly to regulate seed size by directly repressing *KLU* expression in *Arabidopsis thaliana*. *The Plant Cell* 27:620–32