GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Identification and expression analysis of *YABBY* family genes associated with fruit shape in tomato (*Solanum lycopersicum* L.)

H.Q. Han, Y. Liu, M.M. Jiang, H.Y. Ge and H.Y. Chen

Department of Plant Science, School of Agriculture and Biology, Shanghai Jiao Tong University, Shanghai, China

Corresponding author: H.Y. Chen
E-mail: chhy@sjtu.edu.cn

Genet. Mol. Res. 14 (2): 7079-7091 (2015)
Received June 6, 2014
Accepted January 12, 2015
Published June 29, 2015
DOI http://dx.doi.org/10.4238/2015.June.29.1

ABSTRACT. *YABBY* family genes play important roles in the development of leaf, flower, and fruit. The purpose of this research was to integrate all the *YABBY* genes and analyze the correlation between gene expression and fruit shape in tomato. Scanning of 24 genomes of sequenced species demonstrated that *YABBY* genes were very normal and stable in flowering plants except the seedless plants. Nine *YABBY* genes in tomato were computationally and experimentally characterized. The phylogeny was constructed based on whole proteins or the YABBY domain, and five distinct clades were observed as described in other angiosperms. A comparison of the expression patterns in tomatoes with large differences in fruit shape and/or size suggested that during the fruit development, *YABBY* genes had both negative and positive functions. The obtained information could provide a deeper understanding of the evolution of *YABBY* genes and can also be useful for tomato yield and shape breeding.

Key words: Tomato; *YABBY* gene; Fruit shape; Genomic scanning; Gene expression

INTRODUCTION

Tomato is a very important fruit worldwide and is used as a model plant for biology research. Fruit weight, size, and shape are significant economical characteristics for production and customer selection. Previous studies have demonstrated that *SUN* (Xiao et al., 2008), *OVATE* (Liu et al., 2002), locule number (Muños et al., 2011), and fasciated (*FAS*) (Cong et al., 2008) played a role in controlling tomato fruit shape or size (Rodríguez et al., 2011). Among them, the *FAS* gene is a seed plant-specific transcription factor that controls the carpel number. *FAS* gene belongs to the YABBY family, which is involved in the development of leaf, shoot, flower, and fruit (Siegfried et al., 1999; Bowman, 2000; Bartholmes et al., 2012).

The YABBY family includes 6 members in *Arabidopsis*, namely, FILAMENTOUS FLOWER (*FIL*), CRABS CLAW (*CRC*), INNER NO OUTER (*INO*), *YABBY2*, *YABBY3*, and *YABBY5* (Bowman and Smyth, 1999). Eight *YABBY* genes have been found to form a small family in rice (Toriba et al., 2007). Nine *YABBY* genes have been described in the tomato genome, and their distribution and expression patterns were reported (Huang et al., 2013). The *FIL* gene (or *YAB1*) is a transcription regulator that is required for flower formation and the establishment of floral meristem identity in *Arabidopsis* flower development (Chen et al., 1999; Sawa et al., 1999). *OsYABBY4* acts as a regulator during the development of vasculature and has been grouped into the *FIL* subfamily (Liu et al., 2007). Ectopic expression of the *OsYAB1* gene, the homolog of *YAB2*, results in extra stamens and carpels in rice (Jang et al., 2004). The *FAS* gene regulates the extreme fruit size during the process of tomato domestication. *CRC* regulates carpel and nectary development in *Arabidopsis* (Bowman and Smyth, 1999; Alvarez and Smyth, 2002) and is also involved in floral meristem termination, gynoecium differentiation, and ovule initiation in California poppy (Orashakova et al., 2009). Drooping leaf (*DL*) regulates carpel specification and midrib development in rice (Yamaguchi et al., 2004; Ohmori et al., 2008). As an ortholog of the *CRC*/*DL* genes, the lily *LiYAB1* gene can rescue the phenotype of the rice *dl* mutant (Wang et al., 2009). Similar to *CRC* in *Arabidopsis*, the altered expression of the *TmFIL* gene is related to the peltate leaf structure in *Tropaeolum majus* (Gleissberg et al., 2005). The *INO* gene plays a significant role in the formation and asymmetric growth of the ovule outer integument in *Arabidopsis* (Villanueva et al., 1999).

YABBY genes have been selected as the target genes for tomato fruit shape study not only because *FAS* causes the fruit size to change during tomato domestication but also because of paralogs such as *YAB1*, *CRC*, *DL*, and *INO* play critical roles in fruit development. It is interesting to research whether other YABBY members besides the *FAS* gene play a role in the development of fruit shape in tomato. Although *YABBY* genes have previously been reported in tomato, we identified some misannotations. Therefore, it was necessary to integrate all the information about YABBY members and analyze the gene distribution and evolution in different species. Moreover, detailed expression analysis would be helpful to correlate their role in tomato fruit size/shape.

MATERIAL AND METHODS

Plant material, phenotype, and fruit

Tomato plants were grown in the greenhouse of the Space Breeding Center in Pujiang Town, Shanghai, in 2012. Two types of tomato cultivars were used - one had fruit with fewer

than 3 carpels and was regarded as a small tomato, and the other one had fruit with more than 6 carpels and was regarded as a big tomato. Samples including leaves, ovaries 5 days before anthesis (5DBA), ovaries 0 days after anthesis (flower), ovaries 5 days after anthesis (5DAA), young fruit, and red fruit from big and small tomatoes were collected. A UMAX color scanner was used to take crosscut fruit digital images shown in **Figure S1**. Big and small tomato cultivars were chosen to compare gene sequences and analyze gene expression.

RNA extraction and cDNA synthesis

The total RNA of different samples was extracted with the UNIQ-10 kit (Sangon Biotech, China). A PrimeScript reverse transcription polymerase chain reaction (RT-PCR) kit (Takara, Japan) was used to synthesize the first-strand cDNA with the oligo dT primer. The cDNAs for quantitative real-time PCR were synthesized by PrimeScript RT reagent with gDNA Eraser based on the equal amount of total RNA.

Genome-wide scanning of YABBY family members

The annotated proteins of 24 genome-sequenced plants were downloaded from the Ensembl database (Kersey et al., 2013), and tomato protein sequences were obtained from the Sol Genomics Network (Bombarely et al., 2011). The HMM profile of the YABBY family was downloaded from the pfam database (Punta et al., 2012). The HMMER software (Finn et al., 2011) was used to search a customized protein sequence database from 25 plant genomes for matches to the YABBY HMM profile with the significance threshold (E-value) of 0.01 for sequence matches and 0.03 for hit matches. The proteins without a YABBY domain were discarded. Splice variants were checked in *Arabidopsis*, rice, and tomato.

Comparison of tomato YABBYs and gene isolation

The YABBY proteins obtained from genome-wide scanning of both ITAG2.3 and Ensembl annotations were analyzed with multiple alignments. Pairwise alignment was used to compare 2 sequences from the previous step.

The amplification of the coding sequence for YABBY family members in big and small tomato cultivars was performed with PrimeSTAR® Max DNA Polymerase (Takara). The primers were designed according to the coding DNA sequence (CDS) from both annotation databases. Gene-specific primers for isolating *YABBY* genes are listed in Table 1. PCR products were purified with PCR Product Purification kit (Sangon Biotech) and the A-tailing reaction with the DNA A-Tailing kit (Takara). A-tailed PCR products were cloned into the pMD19-T vector and transformed into JM109 *Escherichia coli* competent cells. Colony PCR was used to check the positive transformants. The sequencing service was supplied by Life Technology. The gene names were given according to the known information from GenBank.

Comparison of the YABBY domain, sequence, and structure

The coding sequences of each YABBY member were compared in big and small tomatoes. The coding sequences and the genome sequences from SL2.40 were analyzed using

FancyGene (Rambaldi and Ciccarelli, 2009). The YABBY domains of 9 tomato YABBY proteins were compared, and their identities were calculated.

Table 1. Primers used for gene isolation and quantitative real-time polymerase chain reaction (q-PCR).

Primer name	Primer sequence (5'-3') for cloning	Primer sequence (5'-3') for q-PCR
SlycFAS_Forward	ATGTCATTCGATATGACTTTTTCTTC	TGGTTCATCATCATCATCCTCTTCA
SlycFAS_Reverse	CTATTTGTTGCCCTCCAGCTTG	TGAATCGGTTGTACGCAGAAGG
SlycYab1_Forward	ATGATCAATCAATCAATCAAATCAAG	CAAAGCCACCAGTCGCAAACAGA
SlycYab1_Reverse	GTAAGGAGATACACCAATGTTTGCTG	CGTTGGATCTCGTCCTTGATGAATCG
SlycYab2_Forward	ATGTCACTTGACATGACATATTC	CACTTGACATGACATATTCCTCCTCTTC
SlycYab2_Reverse	TTAATAGAGACCAATTGTTTTCTGAG	ATGGAACGCTAACCGCAAGAATTG
SlycCRC_Forward	ATGGATTATGTTCAATCTTCTGAG	GCTTCCATCTGCCTACAATCG
SlycCRC_Reverse	AACATTGTTGGTATTTCCAGAAT	GCTTCTCTATGTGGTATCTCTGG
SlycDL_Forward	ATGTCTTCCTCATCTCCTAATTCCTCTTG	AGCGTATTAAAGCAGCACATCCACAGA
SlycDL_Reverse	CTTCTCAGCCAAGGTCCCATTTG	TCACTTCTCAGCCAAGGTCCCATTTG
SlycINO_Forward	ATGCAGCCACCAGAAGTTGACAAG	CCAACATGACTCACAAGCAAGC
SlycINO_Reverse	AGGAATCAAACCGTTGCTATCTCTTG	CTCCAAGACTACAGCCTCCTCTAT
SlycYab3_Forward	GGAAGTGTCAAAGACATCAATCAAATC	CACCTTCCTCCTCCACCTCCTT
SlycYab3_Reverse	TTGATTCTCATCCCATCATCAGTAAG	GTCCACATCGAACCGTCACAGT
SlycYab5a_Forward	ATGGCAAGTTGCATTGATGTTGCTTC	TATTGTTCTTGCGGTGAGTGTTCCAT
SlycYab5a_Reverse	CACACACAGCTTCATGAAATGAAGATC	GCCACGAAGCAGAAGAAGAGGAC
SlycYab5b_Forward	ATGTCATCAAGCTACATTGATTCTAC	TTGTGGTCCGTTAATATGGCTGCT
SlycYab5b_Reverse	GAAGGTGAAGGTCTTTATTTTTGG	GAGTAGAGTTGTTTGTGATGGATGATGAAG
SlActin_Forward		CATTGTGCTCAGTGGTGGTTCA
SlActin_Reverse		CATCTGCTGGAAGGTGCTAAGTG

Phylogeny tree

YABBY protein sequences from *Arabidopsis*, rice, and tomato were collected for the phylogeny tree construction. Multiple-sequence alignment was carried out with Clustal Omega (Sievers and Higgins, 2014). A maximum-likelihood tree was constructed with amino acids using MEGA 5.1 bita4 (Tamura et al., 2011) with 1000 bootstrap replications.

Quantitative real-time PCR (q-PCR)

Transcriptional levels of 9 tomato *YABBY* genes were quantified in leaf, flower, and different stages of fruit development from small and big tomatoes with a SYBR® Premix Ex Taq™ kit (Takara) by the FTC-2000 q-PCR System (Funglyn Biotech, Canada). The q-PCR primers were designed in the unconserved and unique region of each member as shown in Table 1. The tomato β-actin (SGN-U580609) gene was used as an internal control. Each set of experiments was repeated 3 times.

RESULTS

YABBY gene distribution in different plant species

An HMMER search of a customized database containing the annotated proteins of 24 genome-sequenced species including 10 eudicotyledon species, 10 Liliopsida species, and 4 seedless species was used to identify predicted YABBY proteins. Information about the *YABBY* genes is shown in Table 2. In eudicotyledons, the highest number of YABBY proteins

was 17 in *Glycine max*, and the lowest was 6 in *Arabidopsis lyrata*. In Liliopsida, the highest number of YABBY proteins was 25 in *Musa acuminata*, and the lowest was 5 in *Aegilops tauschii*. Interestingly, no *YABBY* genes were found in the seedless plants including *Selaginella moellendorffii*, *Physcomitrella patens*, *Chlamydomonas reinhardtii*, and *Cyanidioschyzon merolae* with a very high E-value. The number of detected YABBY proteins might include splice variants. Three and 7 splice variants were found in *Arabidopsis* and rice, respectively, while none were found in tomato.

Table 2. List of the YABBY family members detected in different species.

	Species	No. detected	Genome size (Mb)	No. of haploid chromosomes	Chromosome ploidy	Best 1 domain E-value
Eudicotyledons	*Arabidopsis lyrata*	6	206.67	8	Diploid	9.9e-40
	Arabidopsis thaliana	9 (3)	119.67	5	Diploid	1.3e-39
	Brassica rapa	11	283.98	10	Diploid	2.2e-40
	Glycine max	17	973.9	20	Tetraploid	1.1e-11
	Medicago truncatula	8	314.48	8	Diploid	2.6e-08
	Nicotiana benthamiana	15	2630.35	19	Paleopolyploids	4.1e-20
	Populus trichocarpa	13	485.67	19	Paleopolyploids	2.4e-47
	Solanum lycopersicum	9 (0)	781.51	12	Diploid	7.7e-23
	Solanum tuberosum	11	705.78	12	Doubled-monoploid	5.4e-23
	Vitis vinifera	7	486.26	19	Diploid	2.5e-42
Liliopsida	*Aegilops tauschii*	5	3313.65	7	Diploid	6.4e-12
	Brachypodium distachyon	13	272.06	5	Diploid	7.2e-26
	Hordeum vulgare	18	1868.64	7	Diploid	5.6e-11
	Musa acuminata	25	472.24	11	Doubled-haploid	1.1e-18
	Oryza sativa	15 (7)	426.34	12	Diploid	1.5e-34
	Sorghum bicolor	8	739.15	10	Diploid	1.1e-10
	Setaria italic	9	405.74	9	Diploid	3e-17
	Triticum aestivum	11	3800.33	7	Hexaploid	4.4e-34
	Triticum urartu	6	0.116	7	Diploid	6e-15
	Zea mays	10	2066.91	10	Diploid	2e-15
Lycopodiophyta	*Selaginella moellendorffii*	0	212.5	-	-	0.0023
Bryophyte	*Physcomitrella patens*	0	477.95	27	-	1.4
Chlorophyta	*Chlamydomonas reinhardtii*	0	120.41	17	-	5e-05
Rhodophyta	*Cyanidioschyzon merolae*	0	16.55	20	-	0.17

Dashes represent uncertain or not mentioned data. The proteins detected with default E-value but without the YABBY domain were discarded. The numbers in parentheses represent the numbers of splice variants.

Isolation of tomato *YABBY* genes

Nine YABBY family members were detected in the annotation database supported by Ensembl and the International Tomato Annotation Group release 2.3 (ITAG2.3) based on the SL2.40 genome. Multiple alignments were performed using all 18 sequences from both annotations. We found that 18 sequences could be divided into 9 pairs, with each pair containing sequences from separate annotations, but the sequences were not identical in each pair.

A comparison was carried out between the sequences in each pair. In the *FAS* pair, the 2 sequences were not completely identical at the 5'- and 3'-ends, and Solyc11g071810.1.1 was 86 and 52 bp longer at each end, respectively. Solyc11g071810.1.1 was identified to be the *FAS* gene. The second pair, named *YAB1*, contained Solyc01g091010.2.1 and FGENESHPRED00000062711, which was 72 bp shorter at the 5'-end. The *YAB2* pair consisted of Solyc06g073920.2.1 and FGENESHPRED00000004743; these sequences were not identical at the 5'-end and in the middle because of a 51-bp insertion in the second

sequence. The *CRC* pair was the only pair that included 2 completely identical sequences, Solyc05g012050.2.1 and FGENESHPRED00000035500. The fifth pair, like the rice *DL* gene, included Solyc01g010240.2.1 and FGENESHPRED00000038556; these sequences were totally different with a 2-bp nucleotide substitution and an 8-bp insertion leading to an open reading frame (ORF) change between the sequences. The *INO* pair did not have the same sequence because of a difference at the 5'-end. In the *YAB3* pair, a 21-bp deletion was present in Ensembl annotation. In the *YAB5a* pair, a difference in the 5'-end and a 5-bp insertion were found in the Ensembl annotation. For the last *YAB5b* pair, a 159-bp deletion was detected in the 5'-end of FGENESHPRED00000013783.

We designed gene-specific primers to isolate *YABBY* genes according to the annotated CDS. All 9 *YABBY* genes were cloned and identified. Seven genes showed the same sequence as the annotation from ITAG2.3, but neither sequence of the annotations was completely identical with that of the *DL* gene because of indels. A 5-bp deletion changed the ORF in the ITAG2.3 annotation sequence, leading to a frameshift mutation. At the same time, a 66-bp deletion was found in the Ensembl annotation sequence. For the *INO* gene, both sequences shared most of the *INO* CDS at the 3'-end, but the position of the initiation codon changed the ORF.

We used the Basic Local Alignment Search Tool of the National Center for Biotechnology Information with isolated sequences as the query and found the matched gene and protein sequences whose accession numbers are shown in Table 3. The comparison results were shown in **Supplementary material**, which included the sequences we isolated and the annotations in the ITAG2.3 release and Ensembl, respectively.

Table 3. Tomato *YABBY* gene and protein information from ITAG2.3, Ensembl annotation, and the National Center for Biotechnology Information (NCBI).

Gene name	ITAG2.3 release	Ensembl ID	Matching of 2 annotations	Length (bp)	NCBI gene accession	NCBI protein accession
FAS	Solyc11g071810.1.1	FGENESHPRED00000023487	N	534	NM_00124761	NP_001234390
YAB1	Solyc01g091010.2.1	FGENESHPRED00000062711	N	729	XM_004229745	XP_004229793
YAB2	Solyc06g073920.2.1	FGENESHPRED00000004743	N	579	XM_004241308	XP_004241356
CRC	Solyc05g012050.2.1	FGENESHPRED00000035500	Y	477	XM_004238984	XP_004239032
DL	Solyc01g010240.2.1	FGENESHPRED00000038556	N	522	XM_004228801	XP_004228849
INO	Solyc05g005240.1.1	FGENESHPRED00000005160	N	369	XM_004239291	XP_004239339
YAB3	Solyc08g079100.2.1	FGENESHPRED00000021258	N	651	XM_004245689	XP_004245737
YAB5a	Solyc07g008180.2.1	FGENESHPRED00000071753	N	543	XM_004242730	XP_004242778
YAB5b	Solyc12g009580.1.1	FGENESHPRED00000013783	N	576	XM_004251674	XP_004251722

YABBY domain and gene sequence and structure comparison

The coding sequences of big and small tomatoes were compared. There was no significant nucleotide substitution in the 9 *YABBY* genes. This result showed that *YABBY* genes were conserved at the sequence level.

Exon and intron organizations of tomato *YABBY* members are shown in Figure 1. *INO* was the only gene with 5 exons, and *FAS*, *YAB2*, *CRC*, *DL*, and *YAB5a* had 6 exons, while *YAB1*, *YAB3*, and *YAB5b* had 7 exons. Among tomato *YABBY* genes, the number of exons ranged from 5 to 7, and there were remarkable changes in the size of introns. The first intron of *FAS*, *YAB2*, and *YAB5b* was about 2000 bp or larger, but the smallest intron was only 200 bp. The exon sizes varied less than the intron sizes in different genes. *FAS* and *YAB2* had a very similar gene structure.

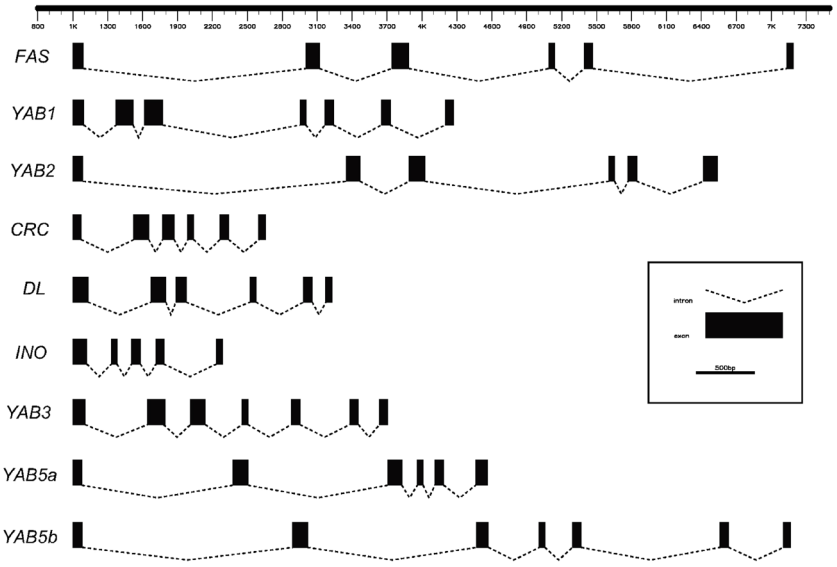

Figure 1. Schematic representation of tomato *YABBY* genes.

The YABBY domain as defined in pfam PF04690 consisted of 17 amino acids. Pairwise comparisons of whole proteins and the YABBY domains were carried out. The amino acid identities are shown in Table 4. The evolution of the whole protein and the YABBY domain was different, but the trend of identities between the protein and domain was very similar. The multiple alignment of the YABBY domain demonstrated high conservation in this family as shown in Figure 2.

Table 4. Amino acid identities among tomato YABBY proteins.

	FAS	YAB1	YAB2	CRC	DL	INO	YAB3	YAB5a	YAB5b
FAS		39.26	66.15	44.07	45.2	23.73	39.35	53.33	48.17
YAB1	91.07		40.91	30.17	33.06	17.77	59.09	40.91	38.43
YAB2	96.43	89.29		39.02	38.02	23.96	45.83	52.6	45.31
CRC	64.29	64.29	62.5		64.74	24.05	33.33	38.33	35.6
DL	64.29	67.87	62.5	85.71		21.39	35.65	38.89	38.74
INO	66.07	64.29	64.29	64.29	55.36		23.61	26.11	26.18
YAB3	91.07	94.64	89.29	67.86	67.86	66.07		42.13	40.74
YAB5a	92.86	91.07	91.07	64.29	64.29	69.65	92.86		66.49
YAB5b	87.5	85.71	85.71	64.29	64.29	69.65	91.07	94.64	

The upper numbers indicate identities (%) among whole proteins, and the lower numbers indicate those among YABBY domains.

Phylogeny analysis

To understand the phylogenetic relationships of YABBY family members in *Arabidopsis*, rice, and tomato, we constructed phylogenetic trees based on the whole protein sequence and the conserved YABBY domain as shown in Figure 3. In the tree based on the whole protein, *YAB2* and *YAB5* formed sister clades, and *INO* and *CRC* formed sister clades.

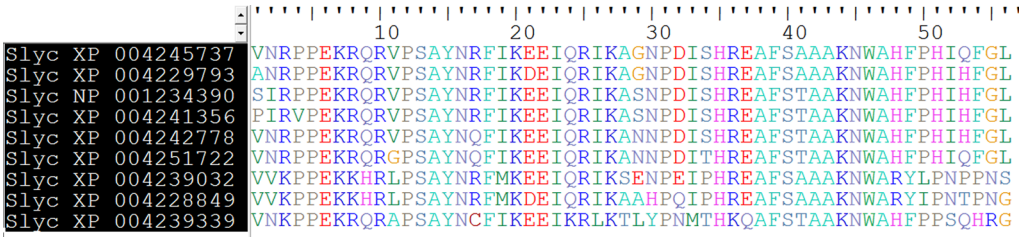

Figure 2. Multiple alignment of the YABBY domain in tomato.

The 2 big clades grouped together to form a sister clade with *YAB1/3*. However, in the tree based on the YABBY domain, the inner sister clades were *YAB1/3* and *YAB5*. The clade was successively followed by *YAB2*, *INO*, and *CRC*. Both the whole protein tree and the YABBY domain tree showed that tomato YABBY family members clustered with their respective orthologs and/or paralogs from *Arabidopsis* and rice.

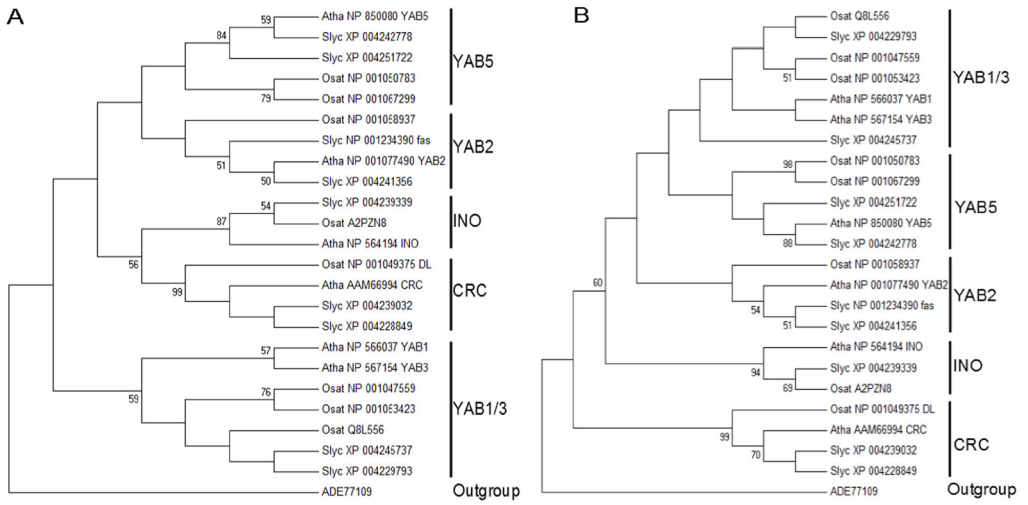

Figure 3. Phylogenetic trees of YABBY whole proteins (**A**) and YABBY domains (**B**). The phylogenetic trees were constructed using the whole proteins of *Arabidopsis* (Atha), rice (Osat), and tomato (Slyc) YABBYs by the neighbor-joining method. The ADE77109 protein of *Picea sitchensis* in the gymnosperms was chosen as a single outgroup. Low bootstrap support (<50%) is not shown.

To investigate whether orthologs/paralogs would show any correlation in different species, the number of genes in each clade was counted as shown in Figure 4. In the *YAB1/3* clade, tomato and *Arabidopsis* had 2 paralogs, but rice had 3 paralogs. In the *YAB2* clade, tomato had 2 paralogs, but *Arabidopsis* and rice had 1 ortholog. The *CRC* clade showed total identities to the *YAB2* clade. In the *INO* clade, each species had only 1 ortholog as in Papaveraceae (Bartholmes et al., 2012), and no gene duplication happened. In the *YAB5* clade, *Arabidopsis* had a single ortholog, but rice and tomato had 2 paralogs. These results revealed that gene duplication events happened in these 3 species at different times.

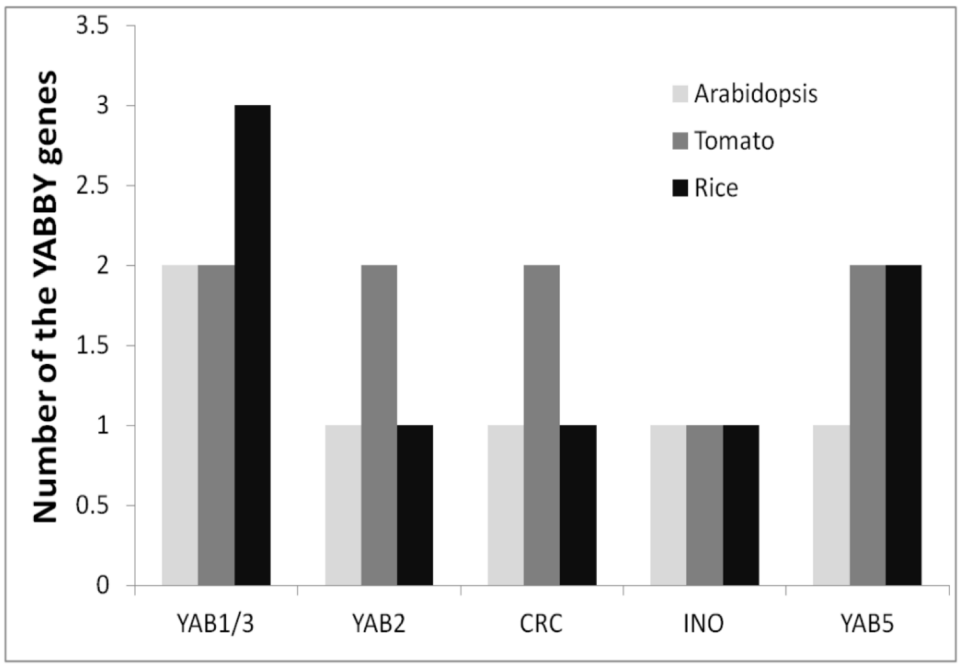

Figure 4. Number of orthologs and/or paralogs in the 5 *YABBY* gene clades.

Expression profiles

We analyzed the expression pattern of each *YABBY* gene in different tissues of 2 tomato cultivars with different fruit shapes and sizes as shown in Figure 5. The *FAS* gene expression patterns were very similar in big and small tomatoes, but the expression in 5DAA of small tomatoes was higher than twice the value in big tomatoes. The expression profiles of *YAB1* in big and small tomatoes were parallel except the high expression that was observed in the ovary of flower and 5DAA in small tomatoes. The expression of the *YAB2* gene was high in 5DBA and then decreased in small tomatoes, but in big tomatoes, expression in 5DAA was the highest and other stages were very similar. *CRC* was highly expressed in 5DBA and then decreased dramatically in small tomatoes, while expression increased in 5DBA and flower in big tomatoes and then decreased like in small tomatoes. For the *DL* gene, vegetative expression was not detected in either tomato. The expression peak in reproductive tissues occurred at the 5DBA, and, after that, the expression was very low or lost. *INO* gene expression was different in leaves of big and small tomatoes. The highest expression of *INO* was detected in 5DBA and then decreased until it disappeared in small tomatoes, but in big tomatoes, moderate expression was observed in the leaf, 5DBA, and flower. *YAB3* gene expression in big tomatoes was always lower than that in small tomatoes. The *YAB5a* gene was always steadily expressed in big tomatoes, but in small tomatoes, it went up in 5DBA, reached the highest expression in 5DAA, and then immediately dropped off. The *YAB5b* gene had high vegetative expression but very low or undetected productive expression.

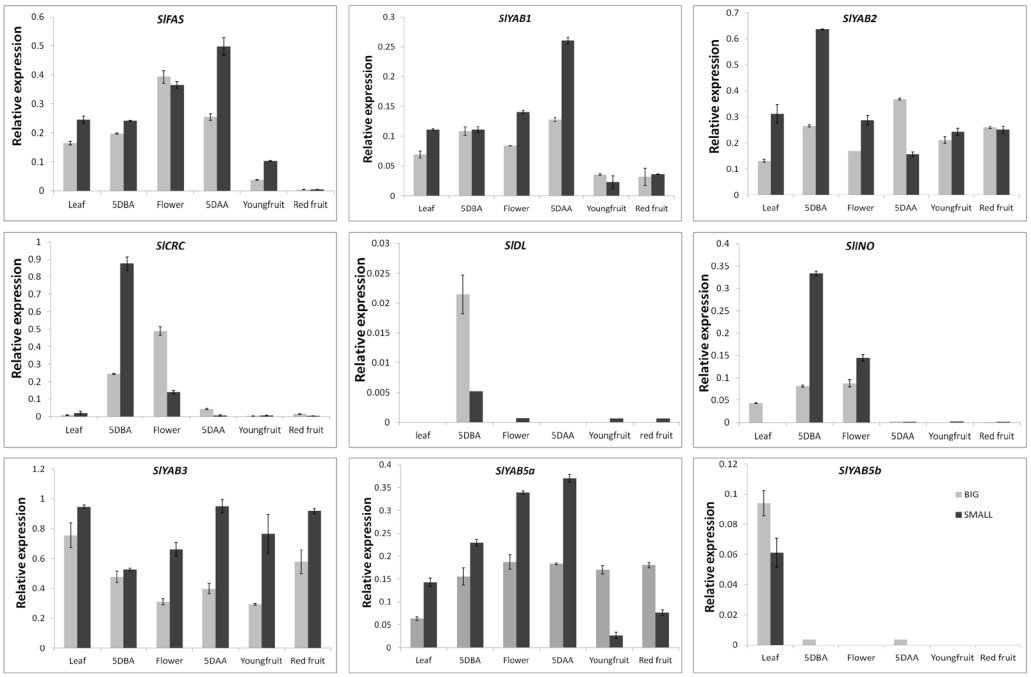

Figure 5. Tomato *YABBY* gene expression profiles. BIG: high-locule-number cultivar; SMALL: low-locule-number tomato cultivar; 5DBA: 5 days before anthesis; Flower: 0 days of anthesis; 5DAA: 5 days after anthesis; Youngfruit: 15 days after anthesis; Red fruit: fruit at breakers-turning stage.

DISCUSSION

The YABBY domain was named after the mutation phenotype of the *Arabidopsis CRC* gene (Bowman and Smyth, 1999). More functions of *YABBY* genes have been proven in different species (Sawa et al., 1999; Villanueva et al., 1999; Yamaguchi et al., 2004; Ohmori et al., 2008; Tanaka et al., 2012). In this study, information about the whole gene family might enhance our understanding of *YABBY* genes with respect to their distribution, evolution, expression pattern, and function. We found that seedless plants did not have proteins with the YABBY domain (Eckardt, 2010), but they had proteins with the high-mobility group box. The specific YABBY motifs might play a very important role in the process of evolution (Shamimuzzaman and Vodkin, 2013). The number of YABBY proteins in different species varied with genome size (*Arabidopsis* had 6), chromosome number (*Populus trichocarpa*), and chromosome ploidy (*Musa acuminata*). The number of genes increased with increasing genome size, and the complexity of the gene function increased with increased duplication during evolution (Yamaguchi et al., 2004). However, the *CRC/DL* gene had a new function in the development of the nectary in *Arabidopsis* (Lee et al., 2005) or midrib in rice (Yamaguchi et al., 2004) that was not obtained by gene duplication.

All 9 tomato genes with a conserved YABBY domain have been experimentally characterized. While a significant number of *YABBY* genes annotated by ITAG are correct and led to the successful amplification of 7 *YABBY* genes, a considerable number of genes annotated

by Ensembl are incorrect, resulting the failure to isolate them. Two genes, *DL* and *INO*, had different sequences in ITAG2.3 and Ensembl. Three of the 9 *YABBY* gene sequences were identical in the two annotations, but 6 were not. Misannotaion might be caused by using sequence similarity methods (Jones et al., 2007). This phenomenon showed that the annotation based on the computational method was reasonable but not always accurate.

Analysis of the tomato YABBY sequences showed high identities in the zinc-finger and YABBY domains as described in *Arabidopsis* (Bowman and Smyth, 1999). This indicated that the YABBY domain was extraordinarily conserved even in different subfamilies during the process of evolution. This might imply that the YABBY domain plays crucial roles in unique functions. The evolution of the gene size and structure, in part, might account for the diversity of gene functions, like that of the first *FAS* intron in tomato (Cong et al., 2008) or the fourth *DL* intron in rice (Ohmori et al., 2008).

Both the whole protein and YABBY domain trees yielded 5 clades as reported in other angiosperms (Bowman, 2000; Nakayama et al., 2010). The members of each clade were very stable in both phylogenic trees. We attributed the differences between the 2 trees to the different conservation of the genes. The phylogenic tree constructed from the YABBY domain would not represent the whole gene from a gene evolution perspective, but, as a part of the whole gene, they could present specific evolution clues.

YAB1, *YAB2*, *FAS*, *YAB3*, and *YAB5a* were expressed during leaf, flower, and fruit development as a function of abaxial cell fate (Siegfried et al., 1999). In contrast, *CRC*, *DL*, *INO*, and *YAB5b* were expressed specifically in reproductive organs (Villanueva et al., 1999; Yamada et al., 2003), and *YAB1*, *FAS*, and *YAB3* were expressed as reported in *Arabidopsis* (Siegfried et al., 1999). The most obvious characteristic was that expression in 5DAA was remarkably higher in small tomatoes than in big tomatoes, which was identical to results in previous reports (Cong et al., 2008; Bonaccorso et al., 2012), because a 6000- to 8000-bp insertion in the first intron changed the transcription in big tomatoes. The expression patterns of *YAB2* and *FAS* were not similar despite the 2 genes having high similarity at the sequence level. The expression patterns of *CRC*, *DL*, *INO*, and *YAB5b* were similar to results in a previous report (Huang et al., 2013). *FAS*, *YAB1*, *INO*, *YAB3*, and *YAB5* expression was negatively related to fruit shape or size, whereas *DL* expression was positively related. The expression of *YAB2* and *CRC* changed during the fruit expanding in both tomatoes. *YAB5b* gene expression was restricted in vegetative tissues. As conserved transcription factors, no change happened at the sequence level, but there were some changes at the expression level. These changes in expression levels might regulate fruit shape or size like the *FAS* gene.

Studies of the gene distribution, sequence changes, structure variation, and gene expression provide valuable insights into the role of *YABBY* genes in controlling fruit shape in tomato. The integration of *YABBY* genes from the tomato genome and the identification of predicted genes accelerate the utilization of huge amounts of data. Combining bioinformatic methods and experimental technique advances help to discern the underlying high-quality data from garbage.

ACKNOWLEDGMENTS

We thank Dr. Lida Zhang for technical assistance and Yue Wang, Ge Bai, and Kunpeng Jia for critical comments. Research supported by the National High Technology Research and Development Program of China ("863" Program; #2012AA100104), the China Spark Program (#2012GA680003), and the Grant of Shanghai Agricultural Committee (#5 2013).

Supplementary material

REFERENCES

Alvarez J and Smyth DR (2002). CRABS CLAW and SPATULA genes regulate growth and pattern formation during gynoecium development in *Arabidopsis thaliana*. *Int. J. Plant Sci.* 163: 17-41.

Bartholmes C, Hidalgo O and Gleissberg S (2012). Evolution of the YABBY gene family with emphasis on the basal eudicot *Eschscholzia californica* (Papaveraceae). *Plant Biol.* 14: 11-23.

Bombarely A, Menda N, Tecle IY, Buels RM, et al. (2011). The Sol Genomics Network (solgenomics.net): growing tomatoes using Perl. *Nucleic Acids Res.* 39: D1149-D1155.

Bonaccorso O, Lee JE, Puah L, Scutt CP, et al. (2012). FILAMENTOUS FLOWER controls lateral organ development by acting as both an activator and a repressor. *BMC Plant Biol.* 12: 176.

Bowman JL (2000). The YABBY gene family and abaxial cell fate. *Curr. Opin. Plant Biol.* 3: 17-22.

Bowman JL and Smyth DR (1999). CRABS CLAW, a gene that regulates carpel and nectary development in *Arabidopsis*, encodes a novel protein with zinc finger and helix-loop-helix domains. *Development* 126: 2387-2396.

Chen Q, Atkinson A, Otsuga D, Christensen T, et al. (1999). The *Arabidopsis* FILAMENTOUS FLOWER gene is required for flower formation. *Development* 126: 2715-2726.

Cong B, Barrero LS and Tanksley SD (2008). Regulatory change in YABBY-like transcription factor led to evolution of extreme fruit size during tomato domestication. *Nat. Genet.* 40: 800-804.

Eckardt NA (2010). *YABBY* genes and the development and origin of seed plant leaves. *Plant Cell* 22: 2103.

Finn RD, Clements J and Eddy SR (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 39: W29-W37.

Gleissberg S, Groot EP, Schmalz M, Eichert M, et al. (2005). Developmental events leading to peltate leaf structure in *Tropaeolum majus* (Tropaeolaceae) are associated with expression domain changes of a YABBY gene. *Dev. Genes Evol.* 215: 313-319.

Huang Z, Van Houten J, Gonzalez G, Xiao H, et al. (2013). Genome-wide identification, phylogeny and expression analysis of *SUN*, *OFP* and *YABBY* gene family in tomato. *Mol. Genet. Genomics* 288: 111-129.

Jang S, Hur J, Kim SJ, Han MJ, et al. (2004). Ectopic expression of *OsYAB1* causes extra stamens and carpels in rice. *Plant Mol. Biol.* 56: 133-143.

Jones CE, Brown AL and Baumann U (2007). Estimating the annotation error rate of curated GO database sequence annotations. *BMC Bioinformatics* 8: 170.

Kersey PJ, Allen JE, Christensen M, Davis P, et al. (2013). Ensembl Genomes 2013: scaling up access to genome-wide data. *Nucleic Acids Res.* 42: D546-D552.

Lee JY, Baum SF, Oh SH, Jiang CZ, et al. (2005). Recruitment of CRABS CLAW to promote nectary development within the eudicot clade. *Development* 132: 5021-5032.

Liu HI, Xu YY, Xu ZH and Chong K (2007). A rice *YABBY* gene, *OsYABBY4*, preferentially expresses in developing vascular tissue. *Dev. Genes Evol.* 217: 629-637.

Liu J, Van Eck J, Cong B and Tanksley SD (2002). A new class of regulatory genes underlying the cause of pear-shaped tomato fruit. *Proc. Natl. Acad. Sci. U. S. A.* 99: 13302-13306.

Muños S, Ranc N, Botton E, Bérard A, et al. (2011). Increase in tomato locule number is controlled by two single-nucleotide polymorphisms located near WUSCHEL. *Plant Physiol.* 156: 2244-2254.

Nakayama H, Yamaguchi T and Tsukaya H (2010). Expression patterns of *AaDL*, a CRABS CLAW ortholog in *Asparagus asparagoides* (Asparagaceae), demonstrate a stepwise evolution of *CRC*/*DL* subfamily of *YABBY* genes. *Am. J. Bot.* 97: 591-600.

Ohmori Y, Abiko M, Horibata A and Hirano HY (2008). A transposon, Ping, is integrated into intron 4 of the DROOPING LEAF gene of rice, weakly reducing its expression and causing a mild drooping leaf phenotype. *Plant Cell Physiol.* 49: 1176-1184.

Orashakova S, Lange M, Lange S, Wege S, et al. (2009). The CRABS CLAW ortholog from California poppy (*Eschscholzia californica*, Papaveraceae), *EcCRC*, is involved in floral meristem termination, gynoecium differentiation and ovule initiation. *Plant J.* 58: 682-693.

Punta M, Coggill PC, Eberhardt RY, Mistry J, et al. (2012). The Pfam protein families database. *Nucleic Acids Res.* 40: D290-D301.

Rambaldi D and Ciccarelli FD (2009). FancyGene: dynamic visualization of gene structures and protein domain architectures on genomic loci. *Bioinformatics* 25: 2281-2282.

Rodríguez GR, Muños S, Anderson C, Sim SC, et al. (2011). Distribution of *SUN*, *OVATE*, *LC*, and *FAS* in the tomato

germplasm and the relationship to fruit shape diversity. *Plant Physiol.* 156: 275-285.
Sawa S, Watanabe K, Goto K, Kanaya E, et al. (1999). FILAMENTOUS FLOWER, a meristem and organ identity gene of *Arabidopsis*, encodes a protein with a zinc finger and HMG-related domains. *Genes Dev.* 13: 1079-1088.
Shamimuzzaman M and Vodkin L (2013). Genome-wide identification of binding sites for NAC and YABBY transcription factors and co-regulated genes during soybean seedling development by ChIP-Seq and RNA-Seq. *BMC Genomics* 14: 477.
Siegfried KR, Eshed Y, Baum SF, Otsuga D, et al. (1999). Members of the *YABBY* gene family specify abaxial cell fate in *Arabidopsis*. *Development* 126: 4117-4128.
Sievers F and Higgins DG (2014). Clustal Omega, accurate alignment of very large numbers of sequences. *Methods Mol. Biol.* 1079: 105-116.
Tamura K, Peterson D, Peterson N, Stecher G, et al. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28: 2731-2739.
Tanaka W, Toriba T, Ohmori Y, Yoshida A, et al. (2012). The *YABBY* gene TONGARI-BOUSHI1 is involved in lateral organ development and maintenance of meristem organization in the rice spikelet. *Plant Cell* 24: 80-95.
Toriba T, Harada K, Takamura A, Nakamura H, et al. (2007). Molecular characterization the *YABBY* gene family in *Oryza sativa* and expression analysis of *OsYABBY1*. *Mol. Genet. Genomics* 277: 457-468.
Villanueva JM, Broadhvest J, Hauser BA, Meister RJ, et al. (1999). INNER NO OUTER regulates abaxial-adaxial patterning in *Arabidopsis* ovules. *Genes Dev.* 13: 3160-3169.
Wang A, Tang J, Li D, Chen C, et al. (2009). Isolation and functional analysis of *LiYAB1*, a *YABBY* family gene, from lily (*Lilium longiflorum*). *J. Plant Physiol.* 166: 988-995.
Xiao H, Jiang N, Schaffner E, Stockinger EJ, et al. (2008). A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. *Science* 319: 1527-1530.
Yamada T, Ito M and Kato M (2003). Expression pattern of INNER NO OUTER homologue in *Nymphaea* (water lily family, Nymphaeaceae). *Dev. Genes Evol.* 213: 510-513.
Yamaguchi T, Nagasawa N, Kawasaki S, Matsuoka M, et al. (2004). The YABBY gene DROOPING LEAF regulates carpel specification and midrib development in *Oryza sativa*. *Plant Cell* 16: 500-509.

