INTERNATIONAL JOURNAL OF AGRICULTURE & BIOLOGY
ISSN Print: 1560–8530; ISSN Online: 1814–9596
17–0836/2018/20–6–1274–1282
DOI: 10.17957/IJAB/15.0626
http://www.fspublishers.org

Full Length Article

A Genome-Wide Analysis of GATA Transcription Factor Family in Tomato and Analysis of Expression Patterns

Qi Yuan, Chunli Zhang, Tingting Zhao, Meini Yao and Xiangyang Xu*
College of Horticulture & Landscape Architecture, Northeast Agricultural University, Harbin, P.R. China
*For correspondence: xxy709@126.com

Abstract

GATA proteins are a class of transcriptional regulators that can identify and bind to GATA motifs, which generally have a zinc finger structure. In this study, 30 GATA transcription factors in tomato were obtained, analyzed, and divided into 4 subgroups. For the tomato GATA zinc finger domain, most of the amino acid sites were highly conserved. The results of chromosomal localization showed that the GATA family members were distributed on 12 chromosomes of tomato. The tissue-specific expression patterns of 14 GATA genes were analyzed, and the expression of these genes in roots, stems, leaves, flowers and fruits was very different. The expression patterns under abiotic stress indicated that members of subfamily I were responsive to cold stress, drought stress and salt stress. However, the expression of members of subfamily I under salt treatment was not obvious. These results can provide some help and practical guidance for follow-up studies of the tomato GATA transcription factor family. © 2018 Friends Science Publishers

Keywords: Bioinformatics analysis; GATA; Gene expression; Tomato

Introduction

Tomatoes are one of the most important Solanaceous fruit crops in Solanaceae family, and it is an important model plant for plant genetics and breeding and in the developmental biology (fruit development and maturation), pathological molecular biology and other fields. On May 31, 2012, the research results of tomato genome sequencing were published in Nature (Sato *et al*., 2012). Its rich data content provided a good data platform for bioinformatics analysis of a specific gene family system in the tomato genome (Zhu *et al*., 2014).

A transcription factor is a protein that can be combined with a specific nucleotide sequence upstream of a gene; it is also known as a trans-acting factor. By combining and interacting with cis-acting elements in the promoter region of the related gene, a transcription factor can also interact with other proteins and regulate the start and transcription efficiency of gene transcription (Singh *et al*., 2002). In different stages of plant life, the regulation of gene expression is essential. Transcriptional regulation is indispensable for regulating gene expression. The nuclear localization signal in transcription factors can direct transcription factor synthesis in endochylema and then enter the nucleus through the nuclear membrane. Transcription factors are numerous and have many functions that can be combined with target gene promoters, and they can both positively and negatively regulate transcriptional activity (Jin *et al*., 2015). It is well known that there are many transcription factor families in plants, including bZIP, bHLH, FAR1, CAMTA, GRAS, NAC, and GATA. In 1988, Evans (Evans *et al*., 1988) reported the (T/A) GATA (A/G) sequence for the first time in a chicken globin gene promoter, Then, the transcription factor GATA-1 (Hannon *et al*., 1991) was found in 1991, and GATA-2, GATA-3, GATA-4, GATA-5, GATA-6 and many other GATA transcription factors were found. Together, they form a family of GATA transcription factors. Multiple group consensus sequences (T/A) GATA (A/G) were found in the regulatory and coding sequences of globin genes and other erythroid-specific genes, and the protein factor binding to this motif specificity was identified and named the GATA transcription factor. The DNA binding domain of the GATA transcription factor is composed of a class IV zinc finger structure ($C\text{-}X_2\text{-}C\text{-}X_{17\text{-}20}\text{-}C\text{-}X_2\text{-}C$) and the subsequent basic region (Reyes *et al*., 2004). Most of the GATA transcription factors in plants contain ($C\text{-}X_2\text{-}C\text{-}X_{18}\text{-}C\text{-}X_2\text{-}C$) and ($C\text{-}X_2\text{-}C\text{-}X_{20}\text{-}C\text{-}X_2\text{-}C$) zinc finger structures (Lowry and Atchley, 2000). The GATA family is a class of transcriptional regulators that can identify and bind to GATA motifs, which generally have a zinc finger structure. The common feature of the GATA family is their high affinity for a consistent sequence (T/A)GATA(A/G). Zinc finger proteins are among the most common transcription factors and exist widely in eukaryotes. A zinc finger is a universal protein structure element that is formed by a pair of cysteine and a pair of histidine complexes with Zn^{2+} that folds in on itself to form a relatively independent "finger"

To cite this paper: Yuan, Q., C. Zhang, T. Zhao, M. Yao and X. Xu, 2018. A genome-wide analysis of GATA transcription factor family in tomato and analysis of expression patterns. *Int. J. Agric. Biol*., 20: 1274−1282

tetrahedron structure. The GATA transcription factor is a DNA-binding domain containing 1 or 2 Cys2/Cys2 type zinc finger structures that recognizes the common DNA sequence 5′-(T/A)HGATA(A/G)-3′, which is present in the promoter region of targeted genes (Marzluf, 1997). A GATA factor is a kind of transcription factor that exists widely in eukaryotes and plays an important role in plant responses to light, chlorophyll synthesis, cytokinin responses, and carbon and nitrogen metabolism, which are biological processes that are closely related to crop production. Research on the GATA family can provide new theoretical bases for increasing crop yields (Ao *et al.,* 2015).

The widespread presence of family members has been found in animals, fungi, plants and other organisms. GATA transcription factors have been reported in many plants, such as *Arabidopsis* (Reyes *et al*., 2004), *Ricinus communis* (Ao *et al.,* 2015), *Ammopiptanthus mongolicus* (Shi *et al*., 2011) and rice (Davierwala *et al*., 2001), but there is very little analysis of GATA transcription factors in tomato. Therefore, this study focuses on the systematic bioinformatics analysis and expression pattern analysis of GATA transcriptional factors in tomato.

Materials and Methods

Identification of GATA Transcription Factor Family Members in Tomato

Amino acid sequences of the GATA family gene coding region were downloaded from the Plant Transcription Factor Database (http://planttfdb.cbi.pku.edu.cn) and SGN (https://solgenomics.net). Software HMMER (http://hmmer.org/) (Lizong *et al*., 2014) and SMART (http://smart. Embl -heidelberg.de) were used to identify all the members of the GATA transcription factor family in tomato, and 30 genes containing the GATA domain were obtained.

Evolutionary Analysis of the GATA Gene in Tomatoes

The GATA protein sequences of tomato and *Arabidopsis thaliana* were aligned using the Clustal X software. Phylogenetic tree analysis was carried out according to the GATA sequence of tomato using the phylogenetic tree analysis software MEGA and using neighbor-joining (N J). The parameters were Poisson correction, pairwise deletion, bootstrap and 1000 repeats (Zhang *et al*., 2014). The GATA gene family of *Arabidopsis thaliana* was introduced to classify the GATA gene family of tomato, and a phylogenetic tree was constructed.

Analysis of Conserved Domains of the GATA Protein in Tomato

DNAMAN software was used to analysis the conserved domain sequences of tomato GATA proteins (Wang *et al*., 2014).

Conserved Sequence Analysis and Gene Structure Analysis of the GATA Protein Family in Tomatoes

The online software MEME (http://meme-suite.org) was used to predict conserved motifs (Bailey and Elkan, 1994). The exon-intron structure analysis of tomato GATA genes were displayed by inputting sequences into GSDS (http://gsds.cbi.pku.edu.cn) (Hu *et al*., 2015). Exon-Intron information was downloaded from SGN.

Chromosomal Localization of the GATA Gene Family in Tomato

According to the genomic information of the retrieved GATA transcription factors, we used the software MapInspect (http://www.plant-breeding.wur.nl/UK) to make chromosome localization mapping by using the data downloaded from the Sol Genomics Network (SGN) and NCBI (https://www.ncbi.nlm.nih.gov/genome).

Tissue Specific Expression of Plant Materials

The tomato cultivar Moneymaker was grown in a greenhouse at the horticultural experimental station of Northeast Agricultural University. A total of approximately 100 seeds were sown in October 2016. The temperature of the greenhouse was controlled at 20−25°C, and the relative humidity was 50%, with an 11/13 h light/dark period. In this experiment, the young roots, stems, leaves, flowers and fruits were taken. Then, the plant materials were stored in a -80°C ultra-low temperature freezer until they were used for extracting RNA and real-time quantitative PCR (Yu *et al*., 2015).

Abiotic Stress Treatments of Plant Materials

When the Moneymaker tomato seedlings reached the four-leaf stage, tomato seedlings with the same growth conditions were treated with cold, drought and salt stress treatments (Zhang *et al*., 2014). For the cold stress treatment, the tomato seedlings were grown in an artificial climate box at 5°C. Then, young leaves were collected at 0, 2, 4, 8, 12 and 24 h after the cold stress treatment and frozen in liquid nitrogen. For the drought treatment, the tomato seedlings were treated with 10% of PEG at 0, 4, 6, 8 and 12 h. For the salt stress treatment, the tomato seedlings were treated with 100 mM NaCl at 0, 2, 4, 8, 12 and 24 h. Three biological replicates of each treatment were performed. Each repetition included sowing, treatment and sampling. Three repeated samples were used for RNA extraction and qRT-PCR.

qRT-PCR Analysis

In this study, we used the tomato Actin-7 gene as an internal reference (Niu *et al*., 2016). The special real-time

quantitative PCR primers of each SlGATA genes were designed using the Primer 5 software and the NCBI primer designing tool (https://www.ncbi.nlm.nih.gov/tools/primer-blast) (Malik *et al*., 2015). The specific primer sequences that were used are shown in Table 1. The qRT-PCR reaction was performed using an iQ5 system. The real-time PCR reaction mixture contained 10 μL of of SYBR® Green Master Mix, 1 μL of each primer, 1 μL of cDNA template, and sterile distilled water up to a total volume of 20 μL. The thermal conditions were as follows: 95°C for 10 min and 40 cycles of 95°C for 5 s, 56°C for 15 s, and 72°C for 30 s (Zhao *et al*., 2015). The data were analyzed using the 2-ΔΔCT method (Livak and Schmittgen, 2001).

Results

Basic Information Analysis of GATA Transcription Factor Family Members in Tomato

In this study, 30 candidate GATA genes were obtained. The basic properties of the GATA gene family members were analyzed using bioinformatics methods, including protein length, domain location, molecular weight and theoretical isoelectric point. We named the family members based on their arrangement order on chromosomes 1 to 12 in tomato. These details are shown in Table 2.

In the GATA gene family of tomato, the length of 30 GATA protein sequences varied greatly. The protein sequence of SlGATA14 was the longest, which was 632 aa. The protein sequence of SlGATA24 was the shortest, which was 89 aa. The average length was 312 aa. The theoretical isoelectric point ranged from 4.8449 (SlGATA5) to 10.6253 (SlGATA3). GATA family members were distributed on 12 tomato chromosomes. Chromosome 1 had the largest distribution, with 6 transcription factors. Chromosomes 7 and 11 had the fewest transcription factors, with 1 transcription factor each. The GATA domain of the tomato GATA transcription factor family member is located between 4 and 564. SlGATA14 has two domain locations, and the other members have a domain location. For SlGATA4, SlGATA14, and SlGATA27, the length of the GATA domain is 37, that of SlGATA5 is 36, and that in other transcription factors is 35. The most important is the combination of the phylogenetic tree of the family: in subfamily I, the GATA domain position is mostly in 167-287; in subfamily II, it is mostly in 4-181; in subfamily III, it is mostly in 199-564, and in subfamily IV, it is mostly in 7-42. From these data, we determined the domain locations in the same subfamily had similarity, and the accuracy of phylogenetic tree was confirmed. Specific information on tomato GATA transcription factors is listed in Table 2.

Phylogenetic Analysis of the GATA Gene Family in Tomato and *Arabidopsis thaliana*

In the phylogenetic tree of the GATA gene family, 60 GATA genes were divided into 4 subgroups, which were named for the corresponding subgroups. The specific distribution of GATA transcription factor family members in tomato in each branch is as follows: I(14) SlGATA2, SlGATA6, SlGATA7, SlGATA9, SlGATA11, SlGATA12, SlGATA13, SlGATA15, SlGATA17, SlGATA19, SlGATA22, SlGATA25, SlGATA26, SlGATA28; II (9) SlGATA1, SlGATA3, SlGATA8, SlGATA10, SlGATA16, SlGATA20, SlGATA24, SlGATA29, SlGATA3; III (4) SlGATA4, SlGATA5, SlGATA14, SlGATA27; and IV (3) SlGATA18, SlGATA21, SlGATA23. The subfamilies are shown in Fig. 1.

Identification and Analysis of Conserved Domains of GATA Genes in Tomato

DNAMAN software was used to identify and analyze the domain protein sequences of tomato GATA genes (Fig. 2). According to the previous standards, the secondary structure of the zinc finger domain was identified and included four β folds and one α helix. The comparison results showed that the zinc finger structure of the other three subfamilies was C-X_2-C-X_{18}-C-X_2-C, except the zinc finger structure of subfamily III was C-X_2-C-X_{20}-C-X_2-C, consistent with the structure reported in *Arabidopsis.* Moreover, as shown in Fig. 2, the tomato GATA zinc finger domains for most amino acid sites are highly conserved, such as Cys-7, Cys-10, Thr-17, Pro-18, Arg-21, Gly-23, Pro-24, and flanking sequences of the second cysteine pairs (LCNACG). According to phylogenetic tree classification, most of the amino acid sites in subfamily I are highly conserved, such as His-9, Lys-15, Trp-20, Gly-26, Lys-28, Thr-29, Val-36, and Arg-37. Partial amino acid sites in subfamily II are relatively conserved, such as Leu-19, Trp-20, Gly-26, Pro-27, Lys-28, Ser-29, and Ile-36. There are also some amino acid sites in subfamily III that are conserved, such as His-9, Met-20, Arg-22, Gly-26, Pro-27, Arg-28, and Leu-36. Some amino acid sites of subfamily IV are conserved, such as Gly-5, Pro-6, Tyr-8, His-9, Gly-11, Val-12, Thr-13, Ser-14, Leu-19, Trp-20, Asu-22, and Pro-25. Compared to the first three subfamilies, the amino acids of the IV subfamily are the most conserved. To sum up, the majority of amino acid sites in each subfamily are highly conserved. For example, His-9 exists in subfamily I, subfamily III and subfamily IV; Trp-20 is found in subfamily I, subfamily II and subfamily IV; and Gly-26 is found in subfamily I, subfamily II and subfamily III.

Conservative Sequence Analysis and Gene Structure Analysis of the Tomato GATA Protein Family

The conserved motif analysis of the tomato GATA gene sequence was performed using the MEME online tool. Through the MEME online software prediction, the results are shown in Fig. 3. It can be seen from the analysis results that the GATA protein domain of the 30 tomato GATA transcription factors is highly conserved and exists in each member. There are 10 conservative motifs.

Table 1: Primer sequences

Gene number	Upstream primer	Downstream primer
SlGATA2	CTAATTCTTCCGACGATTTCAC	CCATTCCAACTCCGCTACA
SlGATA26	GGATGTCTACGGACGGTTAA	CGCAGAGTTTGTCGGTGAA
SlGATA12	TCTAAACAATGGAAGCGTCGGA	GGAAGGGAGCATCGGTGAA
SlGATA22	TCCGAACACTACAAACCCG	CCTTCACTATTGCTACCCGTAT
SlGATA6	CTGTGAAACCGAGAAGCAAACG	GAACCTGACAATGAGTACACCGAC
SlGATA7	TGGGGTGAGTTATCGGTAGA	CCGGTGACTGTTCGAGTGT
SlGATA9	CGTTGACGACTTGCTTGACT	AATGGTGACTTTGCCCTTGT
SlGATA11	TTCACATTCTGGGTATTCGTTG	GCTGTCGTTGTTATAGTGGAGG
SlGATA15	CTATTTTCCTCAAGGCTTAGACTGC	AACGGGGTGGTTCATGGGT
SlGATA25	AGCCTTCTTCCCCACCCTGTA	CCGTTTCAAAACCACCCACC
SlGATA13	TCTACCCTGAGTACCGTCCTGC	GGTTTCTGCTCTGTCTGCTCCTAT
SlGATA17	ACGCCTGTGGTGTTCGTT	AGTTGGGTGGCTATTGCTG
SlGATA19	TGAAGATTGGGATGCGACGG	CGCAATATCCAAAGGAGGCT
SlGATA28	TCGTAGGCAGTATGCTTGGT	AATATGGGAGTGGACAGGTCA
Actin-7	ATTGGTGCTGAGAGGTTCCG	CGGGAAACAGACAGGACACT

Table 2: Information for GATA transcription factor family members in tomato

Original number in the SGN database	Accession	Protein length	Chr	PI	Mw (Da)	Position of GATA domain
Solyc01g060490.2.1	SlGATA1	171aa	1	8.5827	18569.7	68-103
Solyc01g090760.2.1	SlGATA2	260aa	1	6.8357	29246	167-202
Solyc01g100220.2.1	SlGATA3	149aa	1	10.6253	16112.9	21-56
Solyc01g106030.2.1	SlGATA4	326aa	1	6.3741	35018.7	244-281
Solyc01g106040.2.1	SlGATA5	379aa	1	4.8449	41204.8	218-254
Solyc01g110310.2.1	SlGATA6	325aa	1	6.5106	35912.3	243-278
Solyc02g062380.1.1	SlGATA7	289aa	2	7.0524	32109.2	213-248
Solyc02g062760.2.1	SlGATA8	254aa	2	7.7366	28161.4	146-181
Solyc02g084590.2.1	SlGATA9	323aa	2	6.5131	34786.7	244-279
Solyc02g085190.1.1	SlGATA10	252aa	2	8.438	27896.7	131-166
Solyc03g033660.2.1	SlGATA11	295aa	3	9.4449	33283.4	214-249
Solyc03g120890.2.1	SlGATA12	350aa	3	6.6926	37951.8	239-274
Solyc04g015360.2.1	SlGATA13	337aa	4	6.6144	37004.5	236-271
Solyc04g076530.2.1	SlGATA14	632aa	4	6.9525	70270.5	199-236/527-564
Solyc05g053500.2.1	SlGATA15	245aa	5	7.7809	27875.1	179-214
Solyc05g054400.2.1	SlGATA16	197aa	5	10.4196	21531.6	32-67
Solyc05g056120.2.1	SlGATA17	327aa	5	5.9327	35850.2	229-264
Solyc06g060940.1.1	SlGATA18	538aa	6	8.4118	59785.2	7-42
Solyc06g075140.2.1	SlGATA19	258aa	6	8.1167	28673.5	169-204
Solyc07g038160.2.1	SlGATA20	266aa	7	10.0042	29715.7	137-172
Solyc08g007190.2.1	SlGATA21	542aa	8	7.4817	60329.6	7-42
Solyc08g066510.2.1	SlGATA22	359aa	8	6.7917	40265	252-287
Solyc08g077960.2.1	SlGATA23	538aa	8	7.314	60252.4	7-42
Solyc09g075610.2.1	SlGATA24	89aa	9	10.4415	9817.73	4-39
Solyc09g091250.2.1	SlGATA25	304aa	9	8.0228	33761.4	195-230
Solyc10g018560.1.1	SlGATA26	256aa	10	7.5935	29195.1	172-207
Solyc10g047640.1.1	SlGATA27	311aa	10	5.5851	32827.1	225-262
Solyc11g069510.1.1	SlGATA28	326aa	11	9.161	35678.8	210-245
Solyc12g008830.1.1	SlGATA29	293aa	12	10.1661	32442.3	119-154
Solyc12g099370.1.1	SlGATA30	168aa	12	10.2961	18486.8	33-68

Motif5 exists in all members. Motif6 is present in all members of subfamily I, subfamily III and subfamily II, except for SlGATA8 and SlGATA10. Most members of subfamily I have Motif1, Motif3, Motif4, and almost all members of subfamily II have Motif3. Most members of subfamily III have Motif3 and Motif7. All members of subfamily IV have Motif1 to Motif10. Each subfamily of tomato GATA genes is more likely to have the same protein motif, and it also demonstrates the reliability of the tomato GATA phylogenetic tree.

The tomato GATA I subfamily intron number is 1 (SlGATA2, SlGATA6, SlGATA9, SlGATA12, SlGATA13, SlGATA15, SlGATA17, SlGATA19, SlGATA22, SlGATA26, SlGATA28) to 3 (SlGATA7), and most have 1 intron. The tomato GATA subfamily II intron number is 1 (SlGATA8, SlGATA10, SlGATA24) and 2 (SlGATA1, SlGATA3, SlGATA16, SlGATA20, SlGATA29, SlGATA30). The tomato GATA subfamily III intron number is 6 (SlGATA4, SlGATA27) and 14 (SlGATA14). The tomato GATA subfamily IV intron number is 7 (SlGATA18, SlGATA21, SlGATA23). From these subfamilies, the number of introns in tomato GATA is very different. The gene structure of the GATA transcription factor family members in the same subfamily is highly consistent (Fig. 4). The accuracy of the phylogenetic relationship and phylogenetic tree of the GATA gene in tomato has been further verified.

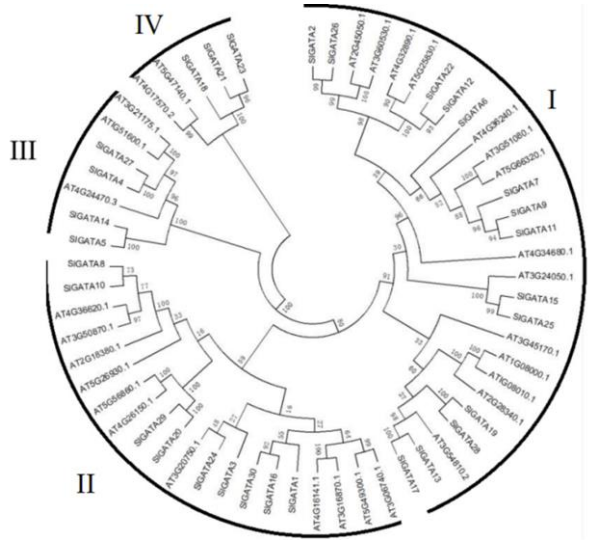

Fig. 1: Phylogenetic Tree of GATA Protein Family in Tomato and *Arabidopsis thaliana*

Phylogenetic analysis of the GATA protein family in tomato and *Arabidopsis thaliana*. The phylogenetic tree was constructed from a complete alignment of 60 GATA proteins using the NJ method with bootstrapping analysis (1000 replicates). Sixty GATA proteins were divided into 4 subgroups, designated subfamilies I, II, III, and IV

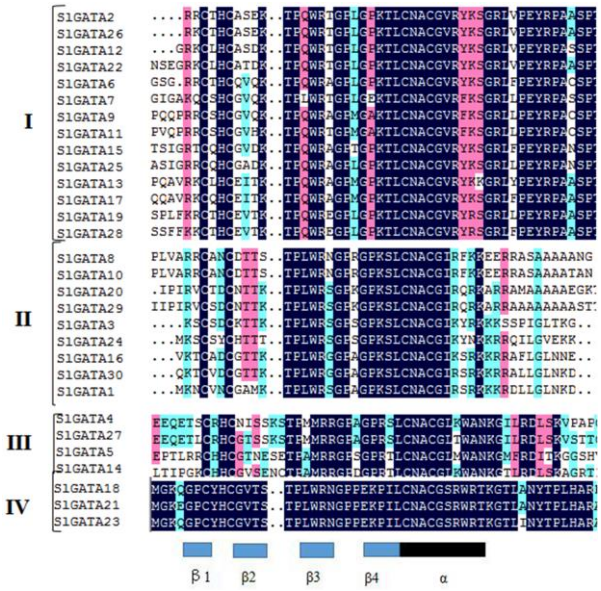

Fig. 2: Comparison of Conserved Domain Sequences of GATA Protein Family in Tomato

Analysis of the conserved domain sequences of tomato GATA protein was conducted with DNAMAN software. Thirty GATA proteins were divided into 4 subgroups, designated subfamilies I, II, III, and IV. The bottom frame indicates the secondary structure of the zinc finger domain, including the four β folds and one α helix

Chromosomal Localization of the GATA Gene Family in Tomato

The results of chromosomal localization showed that GATA family members were distributed on 12 tomato chromosomes. Chromosome 1 had the largest distribution with 6 gene sequences. Chromosomes 3, 4, 6, 9, 10, and 12 had two gene sequences. Four genes were distributed on chromosome 2. There were 3 gene sequences in chromosomes 5 and 8. Chromosomes 7 and 11 had the fewest genes, with 1 each. The longest chromosome was 1, which was 99 cM. The shortest chromosome was chromosome 6, which was only 50 cM. The distance between adjacent genes on the same chromosome was no more than 100 kb (Velasco *et al*., 2010; Hu and Liu, 2012). It can be concluded from Fig. 5 that there is no tandem duplication.

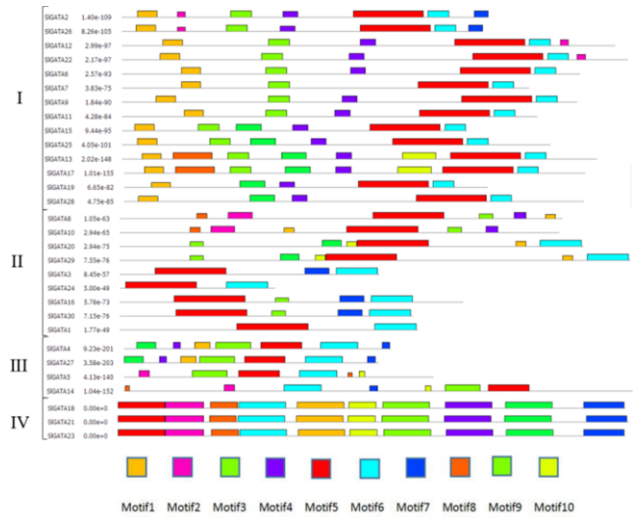

Fig. 3: Analysis of Conserved Motifs of GATA Gene Family in Tomato

Analysis of distribution of conserved motifs in tomato using the MEME software. Ten conserved motifs are displayed in different colors. Thirty GATA proteins were divided into 4 subgroups, designated subfamilies I, II, III and IV

Tissue Specific Expression Analysis of GATA Genes

In this study, all members of subfamily I were selected, including 14 transcription factors and real-time quantitative PCR was used to analyze their expression patterns. As shown in Fig. 6, the results showed that not all GATA genes were expressed in different tissues of tomato (roots, stems, leaves, flowers, fruits), for example, in the root, SlGATA11 and SlGATA28 were hardly expressed; The expression of SlGATA13 was very low in flowers. Moreover, the relative expression of the same gene in different tissues was very different, for example, the relative expression of SlGATA7 in flowers was 3 times as high as in the other four tissues; The relative expression of SlGATA17 in stems was 6 times that of other tissues. There are differences in the expression of different genes in the same tissue, for example, in the fruit the relative expression of SlGATA12 is 5 times that of SlGATA2 and 10 times of SlGATA22.

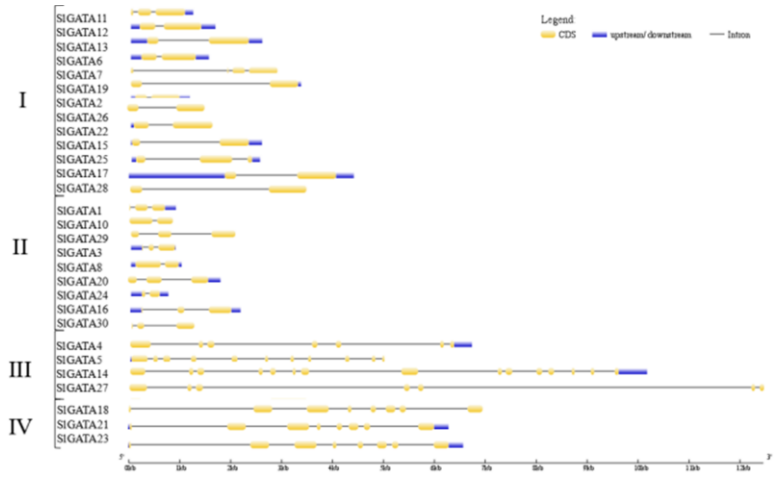

Fig. 4: Genetic Structure Analysis of GATA Gene Family in Tomato

Analysis of distribution of tomato introns and exons with GSDS. Thirty GATA proteins were divided into 4 subgroups, designated subfamilies I, II, III and IV. The gene structure of the GATA transcription factor family members in the same subfamily is highly consistent

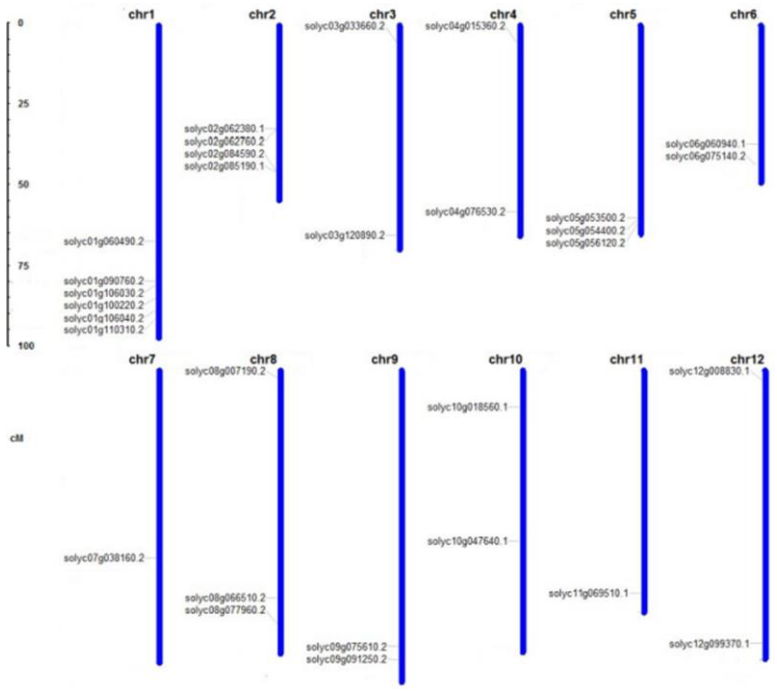

Fig. 5: Chromosomal Location of Tomato GATA Genes on all 12 Chromosomes

Thirty GATA family members were distributed on 12 tomato chromosomes. The chromosome number is indicated at the top of each chromosome. Chromosomal positions of the tomato GATA genes are indicated by gene name

The expression of 14 genes in different tissues of tomato is different, it may be implied that these genes play a different role in the growth and development of tomato.

Analysis of Expression Pattern of GATA Transcription Factors in Tomato under Abiotic Stress

As shown in Fig. 7, we can clearly see the expression of 14 GATA transcription factors under the three different conditions, including salt, cold and drought treatments. As shown in Fig. 7a, under salt treatment, most of the genes expressed in the 14 genes were comparable to the 0 h control, and the change was not very obvious at less than 2 times the expression, which included SlGATA6, SlGATA9, SlGATA25, SlGATA13, SlGATA28, SlGATA15, SlGATA11, and SlGATA19. However, there were some changes in gene expression with greater quantities, more than five times the difference, as observed for SlGATA2 and SlGATA7.

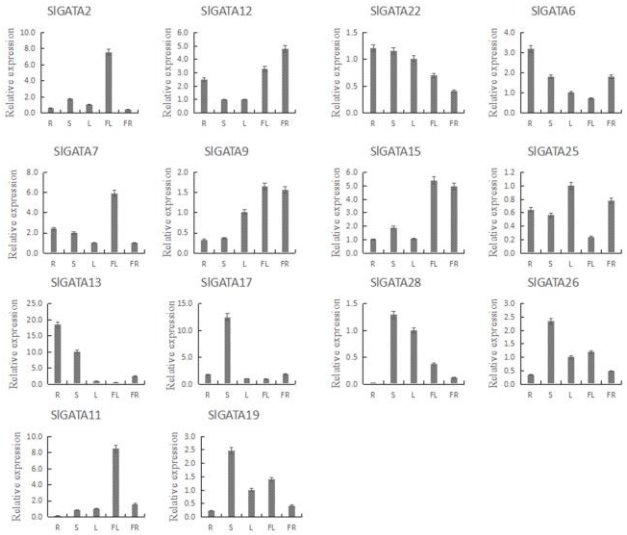

Fig. 6: Expression analysis of the SlGATA genes in different tissues of tomato

Samples were collected from roots (R), stems (S), leaves (L), flowers (FL), fruits (FR). The analysis was carried out by qRT-PCR

Among the 14 SlGATA genes, the expression of 9 genes (SlGATA7, SlGATA25, SlGATA2, SlGATA13, SlGATA28, SlGATA17, SlGATA22, SlGATA11, and SlGATA9) was up-regulated under salt stress. The expression of SlGATA2, SlGATA17, SlGATA22 and SlGATA11 first increased and then decreased. Additionally, the expression of SlGATA17 and SlGATA22 increased at 4 h and then decreased, whereas that of SlGATA2 and SlGATA11 increased at 8 h and 2 h and then decreased, respectively.

As shown in Fig. 7b, under cold conditions, most of the 14 transcription factors compared to the 0 h control, and the difference in the amount of expression was more than 10-fold, such as the expression of SlGATA7, SlGATA15, SlGATA26, SlGATA11, SlGATA19 and SlGATA11 at 8 h compared with that 0 h, which exhibited a more than 40-fold difference. However, in the 14 transcription factors, the expression of some genes changed very slightly, such as SlGATA12 and SlGATA6. Among the 14 SlGATA genes, the expression of 4 genes (SlGATA25, SlGATA13, SlGATA17, SlGATA15) were up-regulated under cold stress. The expression of SlGATA13, SlGATA17 and SlGATA25 first increased and then decreased. For example, the expression of SlGATA17 and SlGATA13 increased at 4 h and then decreased, whereas that of SlGATA25 increased at 2 h and then decreased.

As shown in Fig. 7c, under drought conditions, the expression of some of the 14 genes is less than 2 times compared to the 0 h expression, including SlGATA15, SlGATA25, SlGATA11, SlGATA6, and SlGATA28. However, in these 14 transcription factors, some members, such as SlGATA26, have a more than 20-fold difference in expression.

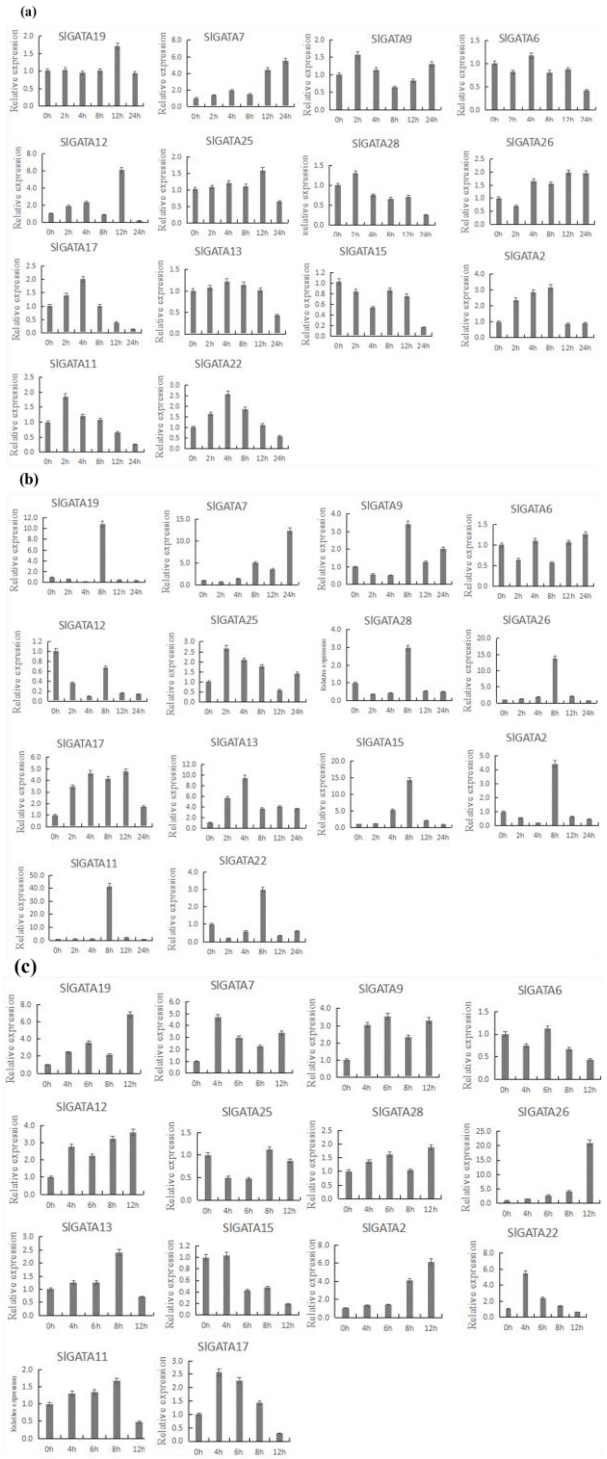

Fig. 7: Analysis of Expression of Tomato GATA Family Members under the Three Treatments

Relative expression of 14 transcription factors was analyzed with qRT-PCR under three different conditions, including salt, cold and drought conditions. (a) The Y-axis is the scale of the relative expression level. The X-axis is the time course of salt stress treatment. (b) The Y-axis is the scale of the relative expression level. The X-axis is the time course of cold stress treatment. (c) The Y-axis is the scale of the relative expression level. The X-axis is the time course of drought stress treatment

Among the 14 SlGATA genes, the expression of 6 genes (SlGATA7, SlGATA2, SlGATA13, SlGATA17, SlGATA22 and SlGATA11) was up-regulated under drought stress. The expression of SlGATA7, SlGATA17, SlGATA22, SlGATA13 and SlGATA11 first increased and then decreased. For example, the expression of SlGATA7, SlGATA17 and SlGATA22 increased at 4 h and then decreased, whereas that of SlGATA13 and SlGATA11 increased at 8 h and then decreased.

To sum up, in the three abiotic stress conditions, the expression under salt treatment showed the least expression difference. It can be inferred that compared with the other two treatments, the expression of GATA transcription factors is not related to salt treatment. As seen in Fig. 7, under the three different treatments, the expression patterns of all members of the GATA transcription factor subfamily I of tomato generally showed an increasing trend first and then decreased over time.

Discussion

In this paper, tomato GATA genes were divided into four subfamilies (I–IV) in a phylogenetic tree (Fig. 1). Previous studies showed that GATA genes in *Arabidopsis thaliana* (Reyes *et al*., 2004) and *R. communis* (Ao *et al.*, 2015) were also divided into four subfamilies. From the occurrence of four subfamilies in different species, we can speculate that GATA gene evolution is relatively conservative. Meanwhile，the results of this study showed that the GATA genes and protein structures of tomato and *Arabidopsis thaliana* were very similar and that their homology was high. For example, the zinc finger structure of tomato subfamily III is C-X_2-C-X_{20}-C-X_2-C, and the zinc finger structure of the other three subfamilies is C-X_2-C-X_{18}-C-X_2-C. This result is consistent with reports in *Arabidopsis* (Reyes *et al*., 2004). The similarity of the protein structures and the homology of the sequences are often consistent with the similarity of their functions. This indicates that tomato and *Arabidopsis* GATA proteins may have similar functions.

However, as already reported, there are great differences in the gene and protein structures of members of the GATA family in the monocotyledon rice and the dicotyledons *Arabidopsis* and tomato (Reyes *et al*., 2004). For example, the GATA protein of rice was divided into six subfamilies (subfamilies I, II, III, V, VI, and VII) without a subfamily IV. However, the GATA genes of tomato and *Arabidopsis thaliana* were divided into four subfamilies. In rice, Subfamily V contains a ZnF PMZ domain and FAR1 domain, and subfamily VI contains two GATA domains. This is also different from tomato and *Arabidopsis thaliana*. In addition, in rice, subfamilies V, VI and VII have different genetic structures than do *Arabidopsis* and tomato, and subfamily VII contains only one exon. These differences indicate that the GATA family may have structural and functional differentiation between monocotyledons and dicotyledons, but further studies are needed.

It has been reported that GATA families are regulated by many kinds of abiotic stress (Bi *et al*., 2005; Richter *et al*., 2013; Peng *et al*., 2015) in different plants. However, there have been few investigations of the GATA gene in tomato plants under abiotic stress. Therefore, we studied the expression patterns of GATA gene families in tomato under abiotic stress, including drought stress, cold stress and salt stress. Under drought stress, compared to the control at 0 h, the relative expression of a tomato GATA gene family (SlGATA2, SlGATA22, SlGATA17, SlGATA13, SlGATA7 and SlGATA11) increased over time. Another study indicated that, under drought stress, compared with the control, the expression of *A. mongolicus* AmZFPG increased with the extension of treatment time (Shi *et al*., 2011). This result is consistent with our result. It can be inferred that the expression of these genes is positively correlated with drought stress.

Similarly, under cold stress, compared with 0 h, the expression of SlGATA17 increased after treatment for 2 h. After 24 h of treatment, the expression decreased, but it was still higher than at 0 h. In *A. mongolicus*, after 2 h of treatment at 4°C, the expression of AmZFPG increased compared to the control. After 4 h of treatment, the expression level of AmZFPG decreased compared to 2 h of treatment, and the expression level was still higher than that of the control group. Thus, the results of this tomato GATA transcription factor study were the same as those in *A. mongolicus*. The results from cold stress indicate that the expression of these genes is associated with low temperature.

Finally, under salt stress, different genes showed increasing expression patterns at different response stages. For example, the relative expression level of SlGATA11 increased after treatment for 2 h, and we can speculate that the gene may be involved in the initial response. However, the relative expression level of SlGATA19 increased after treatment for 12 h, and we can speculate that the gene may be involved in the relative downstream response. This indicated that GATA genes may be involved in the whole process of cold response. However, the specific circumstances need to be investigated further to verify these predictions. Our results suggested that GATA genes played an important role in defense against and tolerance to three types of stress. This research would provide the basic information needed to further investigate tomatoes' response to different stress conditions.

Conclusion

In this study, 30 GATA transcription factors in tomato were obtained and analyzed using bioinformatics methods. In the phylogenetic tree of the GATA gene family, 30 GATA genes were divided into 4 subgroups. In the tomato GATA zinc finger domain, most amino acid sites are highly

conserved. Compared to the first three subfamilies, the amino acids of subfamily IV are the most conserved. The results of chromosomal localization showed that GATA family members were distributed on 12 tomato chromosomes. Chromosome 1 had the largest distribution. Chromosomes 7 and 11 had the fewest genes. In this study, all members of subfamily I were selected, including 14 transcription factors and real-time quantitative PCR was used to analyze their expression patterns. The tissue-specific expression patterns of 14 GATA genes were analyzed, and the expression of these genes in roots, stems, leaves, flowers and fruits was very different, and it showed obvious tissue specificity. The expression of 14 GATA transcription factors is related to the three different conditions, including the salt, cold and drought conditions. Compared with salt treatment, the expression of GATA transcription factors is more related to the other two treatments, including the cold and drought treatments. These results provide the basic information needed for further investigations of tomato GATA transcription factors.

Acknowledgments

This work was supported by the National Key R&D Program of China (2017YFD0101900); the China Agriculture Research System (CARS-23 -A-16); Natural Science Foundation of Heilongjiang Province (C2017024), and Youth Talent support program of Northeast Agricultural University (17QC07).

References

Ao, T., X.J. Liao, W. Xu and A.Z. Liu, 2015. Identification and characterization of GATA gene family in Castor Bean (*Ricinus communis*). *Plant Divers. Resour.*, 37: 453−462

Bailey, T.L. and C. Elkan, 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, 2: 28−36

Bi, Y.M., Y. Zhang, T. Signorelli, R. Zhao, T. Zhu and S. Rothstein, 2005. Genetic analysis of *Arabidopsis* GATA transcription factor gene family reveals a nitrate-inducible member important for chlorophyll synthesis and glucose sensitivity. *Plant. J.*, 44: 680−692

Davierwala, A.P., W. Ramakrishna and V. Chowdari, 2001. Potential of (GATA)n microsatellites from rice for inter- and intra-specific variability studies. *BMC. Evol. Biol.*, 1: 7

Evans, T., M. Reitman and G. Felsenfeld, 1988. An erythrocyte-specific DNA-binding factor recognizes a regulatory sequence common to all chicken globin genes. *Proc. Nat. Acad. Sci. USA*, 85: 5976−5980

Hannon, R., T. Evans, G. Felsenfeld and H. Gould, 1991. Structure and promoter activity of the gene for the erythroid transcription factor GATA-1. *Proc. Nat. Acad. Sci. USA*, 88: 3004−3008

Hu, B., J. Jin, A.Y. Guo, H. Zhang, J. Luo and G. Gao, 2015. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics*, 31: 1296−1297

Hu, L.F. and S.Q. Liu, 2012. Genome-wide analysis of the MADS-box gene family in cucumber. *Genome*, 55: 245−256

Jin, J.P., A.Y. Guo, K. He, H. Zhang, Q.H. Zhu, X. Chen, G. Gao and J.C. Luo, 2015. Classification, prediction and database construction of Plant transcription factors. *Biotechnol. Bull.*, 31: 68−77

Livak, K.J. and T.D. Schmittgen, 2001. Analysis of relative gene expression data using realtime quantitative PCR and the 2− ΔΔCT method. *Methods*, 25: 402−408

Lizong, H., G. Jinjie, Z. Yongfeng, Z. Liying, H. Yaqun and C. Jingtang, 2014. Molecular characterization and evolutionary analysis of the putative caleosin gene family in maize (*Zea mays*). *Int. J. Agric. Biol*., 16: 261−268

Lowry, J.A. and W.R. Atchley, 2000. Molecular evolution of the GATA family of transcription factors: conservation within the DNA-binding Domain. *J. Mol. Evol.*, 50: 103−115

Malik, W., A.A. Khan, H.M.N. Cheema, U. Aslam, M.Z. Iqbal, A. Qayyum, A. Yasmeen and N. Bibi, 2015. Transcriptome analysis of pigment related genes in colored cotton. *Int. J. Agric. Biol*., 17: 205−210

Marzluf, G.A., 1997. Genetic regulation of nitrogen metabolism in the fungi. *Microbiol. Mol. Biol. Rev.*, 61: 17−32

Niu, Y.L., X.M. Jiang and X.Y. Xu, 2016. Bioinformatic analysis of GRAS transcription factors in tomato. *Chin. Agric. Sci. Bull.*, 32: 106−116

Peng, X., Q. Wu, L. Teng, F. Tang, Z. Pi and S. Shen, 2015. Transcriptional regulation of the paper mulberry under cold stress as revealed by a comprehensive analysis of transcription factors. *BMC. Plant Biol.*, 15: 108

Reyes, J.C., M.I. Muro-Pastor and F.J. Florencio, 2004. The GATA family of transcription factors in *Arabidopsis* and rice. *Plant Physiol.*, 134: 1718−1732

Richter, R., E. Bastakis and C. Schwechheimer, 2013. Cross-repressive interactions between SOC1 and the GATAs GNC and GNL/CGA1 in the control of greening, cold tolerance, and flowering time in *Arabidopsis*. *Plant Physiol*., 162: 1992−2004

Sato, S., S. Tabata and H. Hirakawa, 2012. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature*, 485: 635−641

Shi, J.N., M.Q. Liu, J. Shi, Y.P. Wang, G.H. Zhi, Y.Z. Chen and C.F. Lu, 2011. Sequence analysis and expression pattern of Am ZFPG encoding a GATA type zinc finger protein in *Ammopiptanthus mongolicus*. Beijing Linye Daxue Xuebao. *J. Beijing For. Univ.*, 33: 21−25

Singh, K., R.C. Foley and L. Oñate-Sánchez, 2002. Transcription factors in plant defense and stress responses. *Curr. Opin. Plant Biol*., 5: 430−436

Velasco, R., A. Zharkikh and J. Affourtit, 2010. The genome of the domesticated apple (*Malus* × *domestica* Borkh.). *Nat. Genet.*, 42: 833−839

Wang, F., N. Qiu, Q. Ding, J. Li, Y. Zhang, H. Li and J. Gao, 2014. Genome-wide identification and analysis of the growth-regulating factor family in Chinese cabbage (*Brassica rapa* L. ssp. pekinensis). *BMC. Genom.*, 15: 807

Yu, C., X. Cai, Z. Ye and H. Li, 2015. Genome-wide identification and expression profiling analysis of trihelix gene family in tomato. *Biochem. Biophys. Res. Commun.*, 468: 653−659

Zhang, Z.Z., X.L. Chen, X. Guan, Y. Liu, H.Y. Chen, T.T. Wang, L.D. Mouekouba, J.F. Li and A.X. Wang, 2014. A genome-wide survey of homeodomain-leucine zipper genes and analysis of cold-responsive HD-Zip I members' expression in tomato. *Bioscience Biotechnol. Biochem.*, 78: 1337−1349

Zhao, T.T., G. Liu, S. Li, J.F. Li, J.B. Jiang, H. Zhang, L.G. Kang, X.L. Chen and X.Y. Xu, 2015. Differentially expressed gene transcripts related to the Cf-19-mediated resistance response to *Cladosporium fulvum* infection in tomato. *Physiol. Mol. Plant Pathol.*, 89: 8−15

Zhu, Y.Y., B. Xue, A.Q. Wang, W.J. Wang, A. Zhou, S.X. Huang and Y.S. Liu, 2014. Comprehensive bioinformatic analysis of bZIP transcription factors in *Solanum lycopersicum*. *Chin. J. Appl. Environ. Biol*., 20: 767−774

(Received 29 July 2017; Accepted 06 January 2018)