Journal of Applied Biology & Biotechnology Vol. 4 (04), pp. 011-030, July-August, 2016
Available online at http://www.jabonline.in
DOI: 10.7324/JABB.2016.40402

Genome-wide identification and expression analysis along the leaf developmental gradient of the sigma factor gene family in foxtail millet (*Setaria italica*)

Hongyun Liu[1], Jinjin Cheng[1], Siyuan Cheng[2], Hui Fan[1], Bo Wen[3,4], Zheng Liu[1,5]*

[1]College of Life Sciences, Hebei University, Baoding 071002, P.R., China. [2]College of Biology, Hunan University, Changsha 410082, P.R., China. [3]School of Horticulture, Anhui Agricultural University, Hefei 230036, P.R., China. [4]Key Laboratory of Genetic improvement and Ecophysiology of Horticultural Crop, AnHui Province (Institute of Horticulture, Anhui Academy of Agricultural Sciences), Hefei 230039, P.R., China. [5]Hebei Song-Ta-Po Agricultural Development Corporation Limited, Quyang, Baoding, 073100, P.R., China.

ARTICLE INFO

Article history:
Received on: 24/05/2016
Revised on: 10/06/2016
Accepted on: 01/07/2016
Available online: 26/08/2016

Key words:
Setaria italica, sigma factors, gene family, phylogenetic tree, protein domain, digital gene expression

ABSTRACT

Sigma factors are necessary for the initiation of transcription by RNA polymerase in bacteria and plastids of plants. In plants, a small family of nuclear genes is responsible for encoding the sigma factor proteins. In this study, a genome-wide identification and expression analysis of leaf gradient in millet (*Setaria italica*) were performed to characterize sigma factor genes and their proteins. By applying several bioinformatics tools, we identified chromosome locations of seven sigma factor genes in millet and their protein 3D structures. All these proteins contained three conserved domains of σ-70 family. These sigma factor genes have a closer phylogenetic relationship with their orthologs in maize than that in rice. The digital gene expression (DGE) analysis along the millet leaf developmental gradient indicated that *Sisig1*, *Sisig5*, *Sisig6* showed extremely high expression levels in leaf middle and tip regions. Combining the conservation analysis of residues of each sigma factor protein with the DGE profiles of these proteins, it reveals that *Sisig5* plays the housekeeping role compared with other Sisig proteins. Our study will facilitate the future research on crop evolution and the functional studies of sigma factor genes in millet.

1. INTRODUCTION

RNA polymerases play an important role in the first step of gene expression -transcription. As one of RNA polymerase subunits, sigma factors (Sigs) are necessary for RNA polymerase to recognize and bind the promoter and for controlling the rate of gene transcription [1, 2]. Sigs have been grouped into σ-70 families and σ-54 families [3]. So far, all known plant plastid Sigs belong to the σ-70 group [4]. Although plant Sigs are encoded by a small family of nuclear genes, they determine whether the plastid RNA polymerase can bind to a specific promoter and thus set the first checkpoint to control the plastome gene expression [5]. Multiple Sigs have been found in the plastids of higher plants [4]. In the model dicot plant *Arabidopsis thaliana*, six Sigs (SIG1-SIG6) have been identified and characterized [6]. In model monocot crop rice (*Oryza sativa*), six Sigs, OsSIG1 (Os-SigA), OsSIG2A, OsSIG2B, OsSIG3, OsSIG5 and OsSIG6, have also been found [7, 8, 9] or predicted from the rice genome and the full-length cDNA sequencing project [10]. However, only two *Sig* genes (*GRMZM2G143392* and *GRMZM5G830932*) have been identified in maize (*Zea mays*) [11, 12, 13]. In this study, six different gene loci representing seven distinct *Sig* genes of *Setaria italic* (*Sisigs*) were found through blast search of the genome of foxtail millet. Foxtail millet is self pollinating and its genome is relatively small, and therefore very suitable for whole genome research. In 2012, the complete genome sequencing of foxtail millet was successfully completed, making it as a model for the study of C_4 photosynthesis and facilitating the study of crop morphology, physiological and biochemical and comparative functional genomics.

* *Corresponding Author*
Email: liuzhengxp@yeah.net

Moreover, research has been done on millet by using the second generation of high-throughput sequencing technology [14]. The results provide genomic data for gene discovery and the genetic improvement of millet. The results also greatly enrich the genomic research containing comparative genetics and functional genomics [15]. All these studies provide reliable data resources for the study of *Sisig* genes. However, compared with the main crops of rice, wheat and maize, research on molecular genetics of millet started relatively late and most research findings are preliminary[16]. In order to obtain a better understanding of millet Sisigs that are critical for the expression of plastid genes, we performed a bioinformatic study and high-throughput digital gene expression (DGE) profiling on *Sisig* gene family in this important crop. The data from our studies will facilitate the future molecular and genetic studies in foxtail millet.

2. MATERIALS AND METHODS

2.1. Identification of the sigma factor genes in *Setaria italica*

The protein sequences of six *Arabidopsis* Sigs were downloaded from Phytozome (http://phytozome.jgi.doe.gov/pz/portal.html). The ID numbers of six rice *Sig* genes were obtained from oryzabase (http://www.shigen.nig.ac.jp/rice/oryzabase/) and their protein sequences were downloaded from NCBI database (http://www.ncbi.nlm.nih.gov/). To obtain all the *Sisig* genes, BLASTP searches were conducted in the Phytozome (http://www.phytozome.net/) and NCBI databases by using the rice and *Arabidopsis S*ig proteins as queries. Full-length genomic DNA, CDS (Coding DNA Sequence) and protein sequences of Sisigs of *Setaria italica* were downloaded from Phytozome. Besides, nine *sig* genes of maize were also obtained from Phytozome. According to the phylogenetic relationships between Sig protein sequences of *A. thaliana*, millet, rice as well as maize, the *Sig* genes of maize, rice and millet were renamed.

2.2. Computational and bioinformatic analysis of the sigma factor genes and proteins

The Size (aa), MW(Da) and PI of corresponding protein sequences were computed on the website of expasy (http://web.expasy.org/compute_pi/). The information of chromosomal location of *Sisig* genes were obtained from Phytozome and the chromosome location image of sigma factor genes was generated by MapInspect software [17]. To predict the exon-intron structure of the sigma factor genes, the genomic sequence of each gene was compared with its coding sequence (CDS) and was identified on GSDS (http://gsds.cbi.pku.edu.cn/) [18]. For a more intuitive understanding, these protein sequences of *Sisig* genes were submitted to the 3D LigandSite server (http://www.sbg.bio. ic.ac.uk/3dligandsite/) predict three-dimensional models [19]. The Structural evaluation and stereochemical analyses were assessed by using RAMPAGE Ramachandran plot analysis (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php) [20]. Furthermore, the TargetP server (http://www.cbs.dtu.dk/services/TargetP/) was used to predict the subcellular location of Sisig proteins [21]. The conserved motifs of Sisig protein sequences of foxtail millet were analyzed on MOTIF search server (http://www.genome.jp/tools/motif/) with the parameters: E-value of Pfam is 3e-10. Besides, the conserved sites of sigma factor protein sequences were analyzed on the Consurf server (http://consurf.tau.ac.il) with the default parameters. MEGA 4.0 was used to compute the evolutionary distances and construct the phylogenetic trees of Sigs in different plant species by using Clustal W for the alignment of amino acid sequences of Sigs. The neighbor-joining (NJ) method was applied to this analysis [22].

2.3. Expression analysis of millet sigma factor genes along leaf development gradient by Digital Gene Expression Profile (DGE)

Total RNA was extracted from leaf sheath (LS), leaf base area (LB), leaf middle area (LM) and leaf tip area (LT) by using TRIzol (Life Technologies, USA) in accordance with the manufacture's protocol. After RNase-free DNase treatment (Life Technologies, USA), the total RNA was checked for protein contamination and reagent contamination with a Nanodrop spectrophotometer and for RNA purity and degradation by agarose gel electrophoresis. mRNA was enriched by oligo(dT) magnetic beads (Theromo-fisher, USA). Three separate replicates for each leaf region were used.

Provided by a service from LC Sciences (Houston, USA), 12 DGE libraries in total were constructed in parallel using Illumina RNA ligation method (Illumina, San Diego, USA). A library with average length 350 bp, were fixed onto Illumina sequencing chip for cluster generation and performed deep-sequencing using Illumina Genome Analyzer.

The raw data containing adaptor sequences, tags with low quality sequences and unknown nucleotides N were filtered to obtain clean reads with 36 nt in length. Clean reads were then conducted for quality assessment, saturation evaluation and statistical analysis for experimental repeatability. These include classification of total and distinct reads, sufficiency analysis of transcripts coverage and correlation analysis of three parallel biological replicates. For annotation, all clean tags were mapped to the transcripts sequence of *S. italica* from the JGI Comparative Plant Genomics Portal (http://phytozome.jgi.doe.gov) by bowtie, only 1 bp mismatch is allowed. The number of perfect clean reads corresponding to each gene was calculated and normalized to the number of reads per kilobase of exon model per million mapped reads (RPKM). The RPKMs of the seven millet sigma factor genes were then picked out and compared statistically.

3. RESULTS AND DISCUSSION

3.1. Genome-wide characterization of the sigma factor genes in foxtail millet

After the foxtail millet genome database was carefully searched, six gene loci were defined to contain seven *Sisig* genes

(Table 1 and Fig. 1). It should be mentioned that there is one gene locus encoding two genes, i.e. locus *Si026193m.g* encodes the genes *Sisig2Aa* and *Sisig2Ab*. The amino acid sequence alignment by Clustal W revealed that the latter contains additional six successive amino acids (QLNLCF at position 440-445) compared with the former one. In general, the size of these seven proteins ranged from 484 to 566aa. When the structures of *Sisig* genes of foxtail millet were analyzed on the website of GSDS (Fig. 2), it was revealed that the number of exons ranged from six to nine: the *Sisig1* and *Sisig6* had 9 exons; the *Sisig3* contained 7 exons; the other three *Sisig* genes had six exons each. The *Sisig2Aa* and *Sisig2Ab* displayed the same gene structure. Based on gene duplication analysis, there was no segmental duplication event was identified. According to the subcellular location analyses of Sisig proteins, all the seven *Sisig* genes had chloroplast transit peptides (cTP) (Table 2). This indicated that all the products of these seven *Sisig* genes were targeted into chloroplast. This was consistent with the results by previous studies [4, 5]. Sisig2Aa and Sisig2Ab contained the same cTP sequence, consisting of 36 amino acids. Among seven Sisig proteins, Sisig3 had the longest cTP sequence of 69 amino acids. Sisig1, Sisig2B, Sisig5 and Sisig6 each contained 65, 39, 44, 46 amino acids, respectively.

Table 1: Sigma factor genes identified in *Setaria italica* and their protein properties.

Gene name	Gene loci	Transcripts	Protein size(aa)	pI	MW(Da)
Sisig1	Seita.6G049900	*Seita.6G049900.1*	499	9.31	55700.48
Sisig2Ab	Seita.8G112700	*Si026195m*	537	9.50	59572.87
Sisig2Aa	Seita.8G112700	*Si026193m*	543	9.47	60291.74
Sisig2B	Seita.9G461200	*Seita.9G461200.1*	532	9.35	59299.6
Sisig3	Seita.3G124300	*Seita.3G124300.1*	566	9.61	64546.09
Sisig5	Seita.3G126900		484	9.90	55190.18
Sisig6	Seita.9G002700	*Seita.9G002700.1*	554	8.95	63066.89

NOTE: Seita.8G112700 has two splice variants (*Si026195m* and *Si026193m*). aa: amino acid; pI:Isoelectric point ; MW: protein molecular weight.

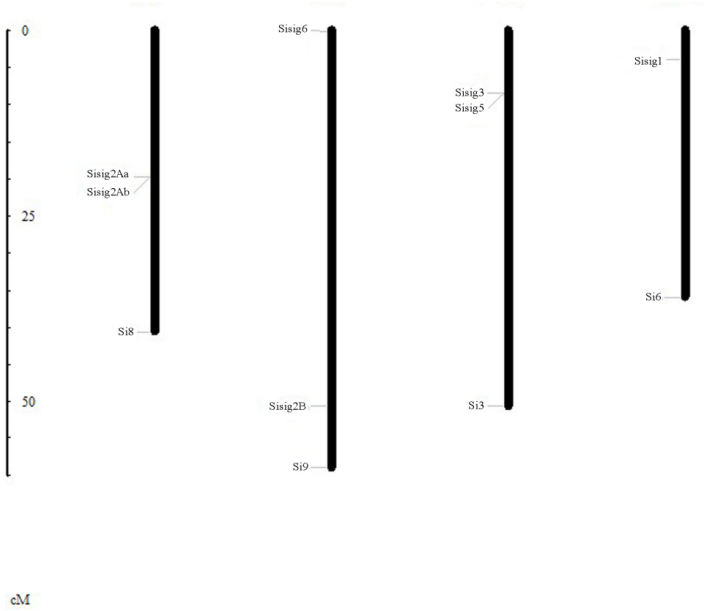


Fig. 1: Chromosome locations of the sigma factor genes in millet.

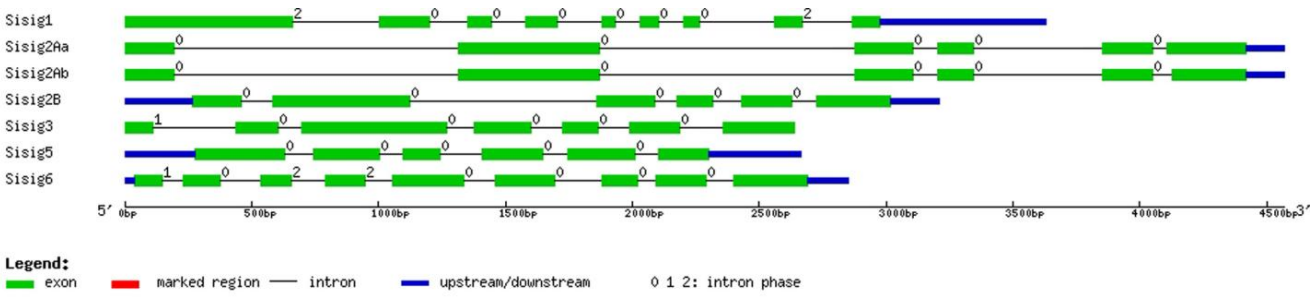


Fig. 2: Gene structures of the millet sigma factor gene family. Exons and introns are shown by filled boxes and single lines, respectively. Untranslated Regions (UTRs) are displayed by thick blue lines at both ends. Intron phases 0, 1, and 2 are indicated by numbers 0, 1, and 2.

3.2 Three-dimensional (3D) structures of millet sigma factor proteins

The three-dimensional models of the seven Sisig proteins were obtained from the 3D LigandSite server (Fig. 3). Through alignment with the protein sequences already existed in the database, the model images of these Sisig proteins were produced and coloured by rainbow from N to C terminus. These images provided a more intuitive understanding of the structure of Sisig proteins. In general, the 3D images of Sisig1, Sisig2Aa, Sisig2Ab, Sisig2B and Sisig5 showed the similar dimensional structures, looking like a "U" or "C". However, the 3D structures of Sisig3 and Sisig6 were irregular. In order to validate that the 3D models of Sisigs were displayed precisely, the ramachandran plot analyses using the RAMPAGE server were conducted.

The results showed that 87.9%, 87.6%, 89.8%, 89.4%, 89.0%, 90.8%, and 85.5% of amino acid residues were in the favored region; 7.8%, 8.8%, 7.3%, 7.6%, 9.4%, 5.6%, and 10.4% in the allowed region in Sisig1, Sisig2Aa, Sisig2Ab, Sisig2B, Sisig3, Sisig5, and Sisig6, respectively (Fig.4).

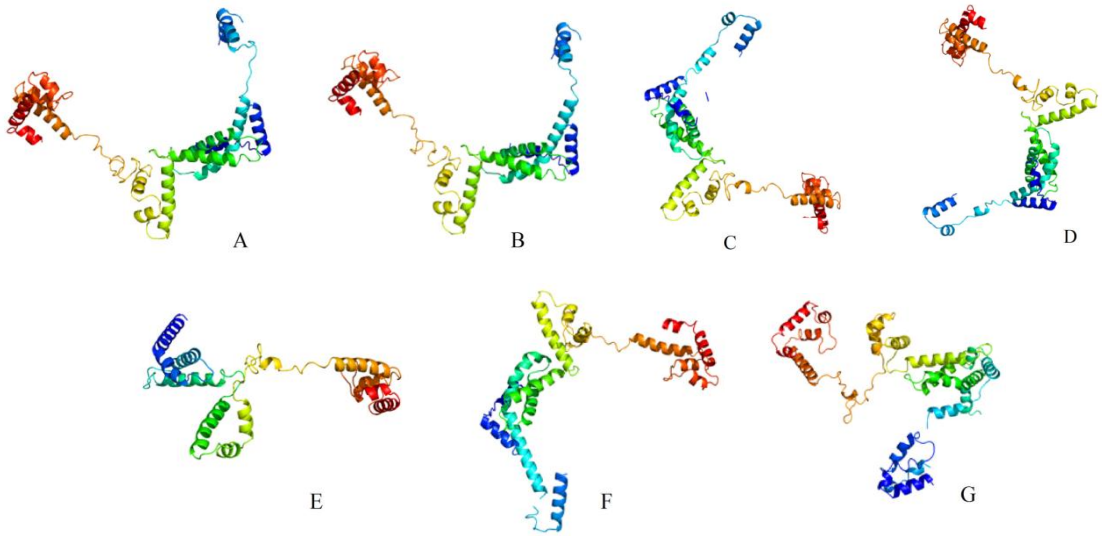


Fig. 3 The three-dimensional models of the millet sigma factor proteins
(A: Sisig1; B: Sisig2Aa; C: Sisig2Ab; D: Sisig2B; E: Sisig3; F: Sisig5; G: Sisig6)

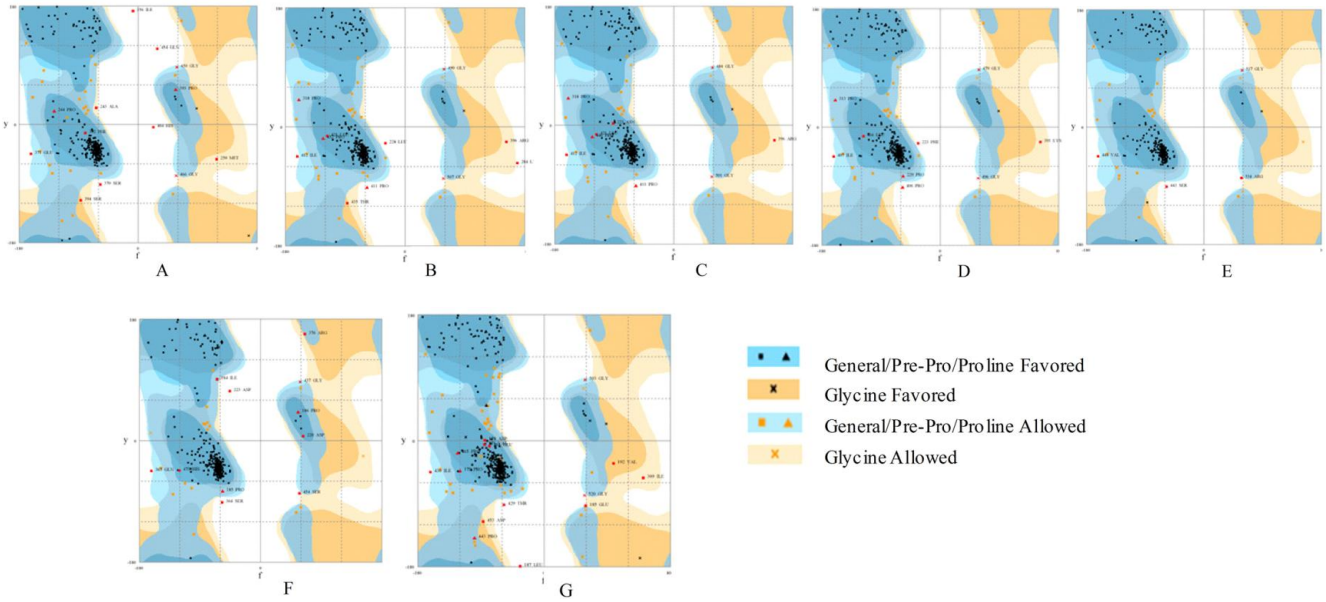


Fig. 4 The Ramachandran plot analyses on 3D models of Sisig proteins
(A: Sisig1; B: Sisig2Aa; C: Sisig2Ab; D: Sisig2B; E: Sisig3; F: Sisig5; G: Sisig6)

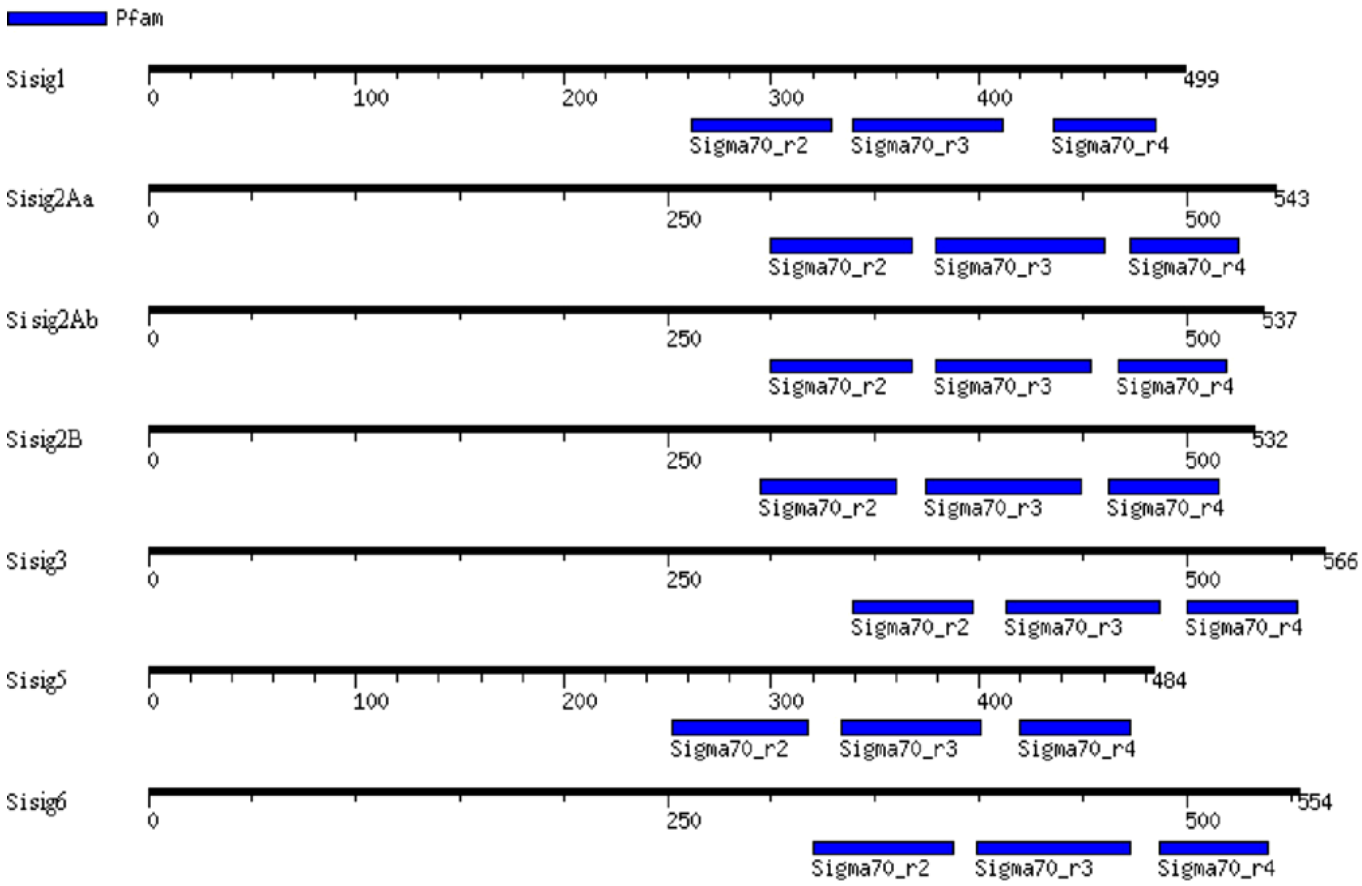


Fig. 5 The conserved motifs of sigma factor proteins in millet.
(Sigma70_r2, Sigma70_r3 and Sigma70_r4, respectively, represented in boxes with blue color).

3.3. Analysis on the conserved motifs of the sigma factor proteins in millet

The positions of the conserved motifs were generated by MOTIF (Fig. 5 and Suppl. 1). All the seven protein sequences contained three sigma-70 conserved domains, namely region 2, 3 and 4. Moreover, they were distributed on the C-terminal region of the proteins, all locating behind the amino acid site of 250. The conservation of these conserved regions correspond one-to-one with the same regions of bacterial sigma-70 factor and both species contain a large non-conservative N-terminal region [5], indicating similar evolutionary origins.

3.4. Evolutionary conservation of amino acid positions in each Sisig protein

The evolutionary conservation of amino acid positions in each Sisig protein which was based on the calculation of phylogenetic relations between homologous sequences was analyzed by the Consurf server (http://consurf.tau.ac.il/) with the default parameters (Table 3 and Suppl. 2-8). The result indicated that there existed differentially conserved degrees of residues in each protein. For instance, for the highest conserved scale of residues (100%), Sisig5 accounted for the largest proportion (36.98%) in the total residues of the protein,

Table 3: The evolutionary conservation of the residues in each sigma factor protein in millet.

Conservation scale	Percentage of residues in each protein (%)						
	Sisig1	**Sisig2Aa**	**Sisig2Ab**	**SiSisig2B**	**Sisig3**	**Sisig5**	**Sisig6**
100%	26.45	13.44	11.73	11.47	27.03	36.98	8.84
90% - 100%	15.43	10.50	9.87	11.47	6.54	0.00	13.72
80% - 90%	8.62	7.55	7.64	6.77	22.61	11.16	9.93
70% - 80%	11.82	6.81	6.89	8.27	12.90	7.64	10.65
60% - 70%	12.22	12.52	12.29	8.27	7.95	19.21	13.00
50% - 60%	9.42	14.92	14.53	16.54	10.60	10.95	14.44
40% - 50%	5.61	18.05	18.06	18.98	6.54	6.41	11.55
30% - 40%	7.02	9.95	11.55	11.65	3.18	5.58	9.57
20% - 30%	2.81	5.89	6.14	6.02	2.47	2.07	7.58
10% - 20%	0.60	0.37	1.30	0.56	0.18	0.00	0.72
Total	100	100	100	100	100	100	100

while Sisig6 accounted for the least proportion (8.84%) among *Sisig* gene family. Similarly, when calculating the most variable residues (conservation scale ≤30%) in each protein, it was also found that Sisig5 accounted for the least proportion (2.07%) in the total residues of the protein and Sisig6 accounted for the largest proportion (8.30%) among these proteins, indicating that the protein Sisig5 was evolutionary conserved compared with the protein Sisig6 when considering the conservation of the amino acid residues of the proteins. This also implied that the Sisig6 was a newly evolved protein. In all, analysis of the conserved sites of each protein would provide some information on the conservation degree of each protein and would further give us some hints on the evolutionary events of *Sisig* genes.

3.5. Phylogenetic tree construction

The *Sig* genes from millet, rice and maize were renamed according to their phylogenetic relationships with those of *Arabidopsis* (Suppl. 9). For example, the phylogenetically closest ortholog of *Arabidopsis Atsig1* in millet was renamed as *Sisig1*. It was found that there are 9 sigma factor genes in maize, which are 1.5 times in number than that in *Arabidopsis*, rice and millet. It is speculated that additional steps of gene duplication would occur in maize during evolution.

Phylogenetic tree construction of millet, rice, maize and *Arabidopsis* Sig proteins was generated by the software of Mega4.0 (Fig. 6). According to the phylogenetic relationship between *Arabidopsis* and millet, *Sisig* genes of millet were clarified into five types (Table 5): two *Sisig2* genes, and *Sisig1*, *Sisig3*, *Sisig5* and *Sisig6* each in one. Additionally, the orthologs of *Atsig4* were not found in rice, maize and millet, possibly *Atsig4* was a newly evolved gene after the monocot-dicot divergence. As a whole, millet *Sisig* genes had close phylogenetic relationships with rice and maize. Moreover, *Sisig* genes had closer relationships with their orthologs in maize. It is probably due to that fact that millet and maize belong to C_4 plants, while rice belongs to C_3 plant. Thus the phylogenetic relationship of *Sig* genes could, to some extent, reflect the evolutionary relationships of these plants. Rice is the typical model gramineous plant for comparative genomic studies because of its slow evolutionary rate, making it better to keep the traits of ancestral genome sequences [23]. Maize and millet both belong to C_4 monocotyledonous crops. So far, the whole genome sequencing works on rice, maize and millet have been completed [24, 25, 26], and some *Sig* genes from *Arabidopsis*, rice and maize were well characterized. Recent studies on *Arabidopsis* provided some evidences on specific roles of each *Atsig* gene [5, 6, 27, 28]. In rice and maize, some results on the studies of *Sig* genes were also obtained [7, 9, 10, 29, 30]. According to the phylogenetic relationship of Sig proteins between the millet and the other three plants, specific roles of *Sisig* genes could be inferred to some extent. Atsig1 is thought to be involved in light response as high-light stress changes the binding between Atsig1 protein and its target promoters [6]. It was found that accumulation of rice *Ossig1* transcripts is significantly higher in green shoots than in dark-grown shoots or in roots [7]. Further, Ossig1 regulates some gene expressions which are critical for the establishment of photosystem I in rice mature chloroplasts, thus participating in the maintenance of photosynthesis [10]. Similarly, the expression pattern of *Zmsig1* is aligned with the light-triggered plastid development processes [29, 30]. The close phylogenetic relationship between millet *Sisig1* and *sig1* genes in *Arabidopsis*, rice and maize suggests that Sisig1 would also be involved in light-regulated development and photosynthesis. Indeed, promoter analysis of *Sisig1* revealed that there exist a considerable number of *cis*-acting elements related to light regulation (Suppl. 8). Atsig5 was identified to bind some target promoters in response to blue light and various abiotic stresses (high light, high salinity and osmotic pressure as well as low temperature) [31, 32, 33]. For example, Atsig5 specifically binds the *psbD* blue-light responsive promoter in *Arabidopsis*. Similarly, *Sisig5* promoter contains elements related to stress signals, such as low temperature, salt and water stresses (Suppl. 8). In addition, Atsig4 plays a specific role in the transcription of *ndhF* [34], a plastid gene encoding a subunit of the plastid NDH complex. However, no *sig4* genes were found in rice, maize and millet. It would be interesting to reveal how *ndhF* genes are regulated in these important monocotyledonous crops. Besides, some specific Sigs such as SIG1, SIG2, and SIG6 are necessary for the expression of cytomin induced chloroplast genes in *Arabidopsis* [35, 36]. However, in millet, only Sisig3, which is phylogenetically distant to these *Arabidopsis* Sigs, was found to have promoter elements related to cytokinin. Thus, it is worthwhile to pay attention to the special function of Sisig3 on cytomin regulated expression of plastid genes.

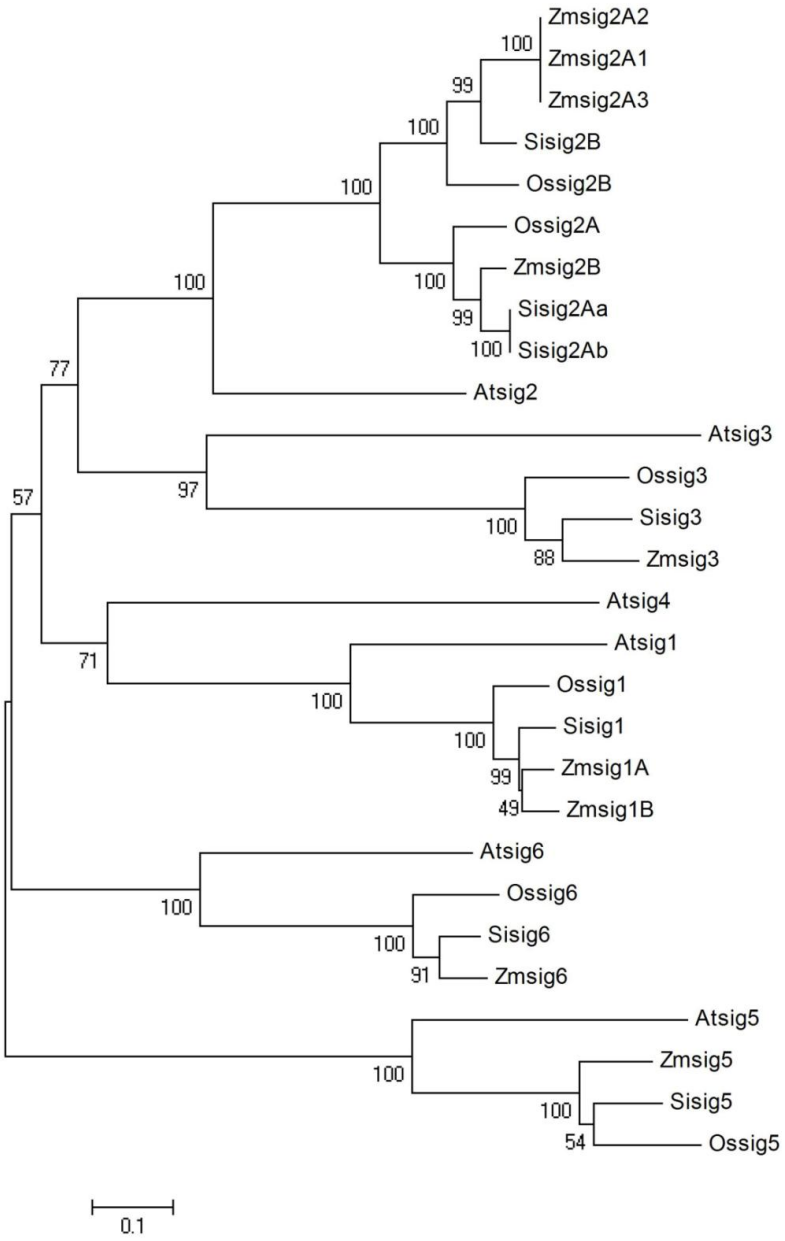


Fig. 6 Phylogenetic tree of foxtail millet, *Arabidopsis*, rice and maize Sig proteins. The joint unrooted tree was generated using MEGA4.0 by the neighbor joining method. Bootstrap values from 1000 replicates were indicated at each branch.Abbreviations: Os, *Oryza sativa*; At, *Arabidopsis thaliana*; Si, *Setaria italica* and Zm, *Zea mays.*

3.6. Expression profiling of millet sigma factor genes along leaf development gradient

The continuous developmental pattern from leaf base to tip has been well studied in developing grass leaves, such as maize and rice [37, 38]. Comparative gene expression analysis of C_4 and C_3 crops along leaf gradient will be helpful for addressing how C_4 traits are developed. Although sigma factors play essential roles in the initiation of expression of chloroplast genes and therefore in the efficiency of photosynthesis, all developmental gradient studies neglected the leaf sheath and none of these reports mentioned the behavior of the sigma factor gene family along such gradient except the one by Aubry et al [39]. In our study (Fig. 7), we found that six out of seven members of millet *Sisig* genes showed the lowest expression levels in leaf sheath area.

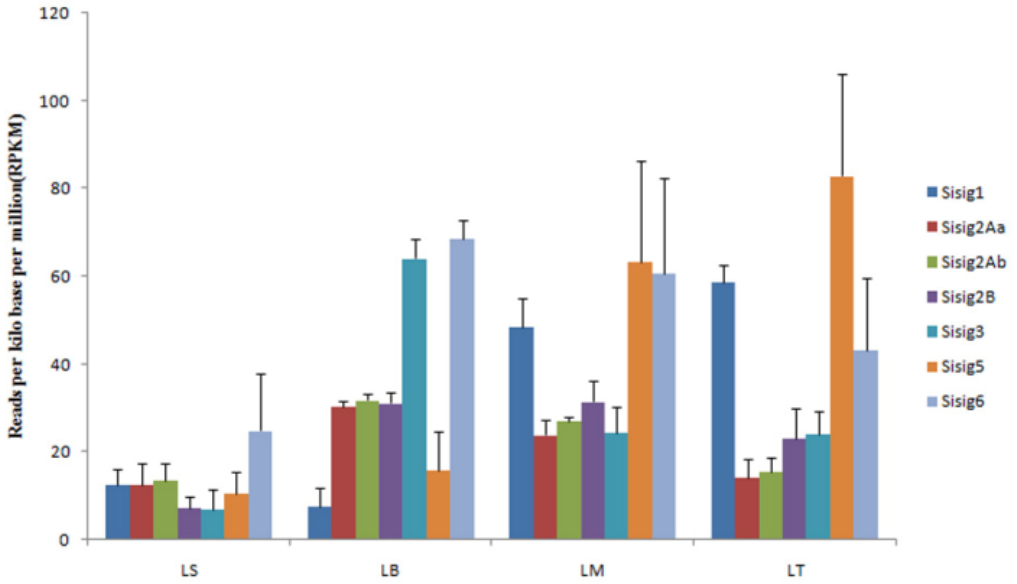


Fig. 7: Expression profiles of millet sigma factor genes along leaf developmental gradient (LS: leaf sheath; LB: leaf base; LM: leaf middle; LT: leaf tip)

Although it was found four sigma factor genes exhibit escalating pattern along leaf gradient in two independent C_4 lineages (*Cleome gynandra* and maize) [39], only one gene (i.e. *Sisig5*) fits this pattern in millet leaf. The gene *Sisig1* exhibited the similar expression pattern as that of *Sisig5* in general, although its expression level in leaf sheath area was 1.5 times as high as that in leaf base area. Five *Sisig* genes (i.e. *Sisig2Aa*, *Sisig2Ab*, *Sisig2B*, *Sisig3* and *Sisig6*) shared the same expression patterns along developmental gradient. They all showed significantly increased expression levels from leaf sheath area towards leaf basa area, i.e. the expression levels *Sisig2Aa*, *Sisig2Ab*, *Sisig2B*, *Sisig6* in leaf base area were 3 times as high as in leaf sheath area, even that of Sisig3 reached 9 fold change. Then the expression levels finally decrease gradually from leaf base area towards leaf tip area, suggesting that functional divergence of these genes compared with that of *Sisig1* and *Sisig5*. Additionally, Fig.7 also showed that expressions of *Sisig1*, *Sisig5* and *Sisig6* were significantly up-regulated in leaf middle and tip areas as their RPKM values reached over 50, suggesting that these genes are main genes involved in photosynthesis. The *Sisig5* was identified to show the highest expression levels (64 and 83 RPKM values, respectively) among the seven sigma factors in leaf middle and tip areas, revealing that it plays a critical role in regulating genes responsible for photosynthesis and leaf development. Combining the residue conservation analysis which identified that Sisig5 protein was the most conserved protein in amino acid sequence among Sisig protein family, it could be concluded that *Sisig5* plays the housekeeping function to maintain chloroplast development in millet. To our knowledge, this is the first study to identify the gene expression patterns of millet *Sisig* gene family along leaf development gradient. This work could provide a foundation for the functional identification of millet sigma factor gene family and for the study of the relationships among leaf morpho gradient, developmental gradient and gene expression patterns of this gene family in the future.

4. CONCLUSION

Although the functional domains of the seven Sisig proteins all contain conserved region 2, 3, 4, their 3D figures displayed the slightly different spatial structures among them which may result in their functional differences. The phylogenetic study of the sigma factor genes displayed us the phylogenetic distances of the *Sisig* genes of millet with their orthologues other plant species, and from this, specific functions of a *Sisig* gene could be inferred from the previous studies on the phylogenetically closed gene in other plant species.

Millet is a model plant for the research on C_4 photosynthesis and it maintains strong tolerance to drought and barren conditions. Through the study of expression profiling of millet sigma factor genes along leaf development gradient, it was found that the gene *Sisig1* exhibited the similar expression pattern as that of *Sisig5* in general. The highest expression levels of *Sisig5* in leaf middle and tip regions among all *Sisig* genes and the most conservation of amino acids of Sisig5 protein among these sigma factor proteins inferred that Sisig5 plays the housekeeping function to maintain chloroplast development in millet.

With the whole genome of millet been sequenced already, it facilitates us with in-depth study of millet *Sisig* gene family, which would be very helpful for our comprehensive understanding on the expressions and regulation of the chloroplast-related genes in this important crop. As the genomic analyses of *Sisig* gene family were fulfilled by this study, the next step should concentrate on characterization of these genes through molecular and biochemical assays.

5. ACKNOWLEDGEMENTS

We thank the Natural Science Foundation of Hebei Province (No. C2015201176), the Scientific Research Foundation for the Returned Overseas Chinese Scholars by State Education Ministry of China (No. 2011–1139), the Project for Improvement of Integrated Strength of Universities in Central and Western Areas of China (Plant Diversity and Resources Utilization), the Cutting-edge and Characteristic Disciplines of Biology (Botany), the Research Innovation Program for Postgraduates of Hebei Province (S2016022), the Opening Grant in Key Laboratory of Genetic improvement and Ecophysiology of Horticultural Crop, Anhui Province (Institute of Horticulture, Anhui Academy of Agricultural Sciences) (No.AHZDSYS201401)for funding.

6. REFERENCES

1. Borukhov S, Severinov K. Role of the RNA polymerase sigma subunit in transcription initiation. Res Microbiol. 2002; 153(9):557–562.
2. Lysenko EA. Analysis of the evolution of the family of the *Sig* genes encoding plant sigmafactors. Russ J Plant Physiol. 2006; 53(5):605-614.
3. Ishizaki Y, Tsunoyama Y, Hatano K, Ando K., Kato K, Shinmyo A, Kobori M, Takeba G, Nakahira Y, Shiina T. A nuclear-encoded sigma factor, Arabidopsis SIG6, recognizes sigma-70 type chloroplast promoters and regulates early chloroplast development in cotyledons. Plant J. 2005; 42(2):133–144.
4. Lerbs-Mache S. Function of plastid sigma factors in higher plants: regulation of gene expression or just preservation of constitutive transcription? Plant Mol Biol. 2011; 76(3-5):235–249.
5. Lysenko EA. Plant sigma factors and their role in plastid transcription. Plant Cell Rep. 2007; 26(26):845–859.
6. Hanaoka M, Kato M, Anma M, Tanaka K. SIG1, a sigma factor for the chloroplast RNA polymerase, differently associates with multiple DNA regions in the chloroplast chromosomes in *vivo*. Int J Mol Sci. 2012; 13(10):12182–12194.
7. Tozawa Y, Tanaka K, Takahashi H, Wakasa K. Nuclear encoding of a plastid σ factor in rice and its tissue- and light-dependent expression. Nucleic Acids Res. 1998; 26(2):415–419.
8. Kasai K, Kawagishi-Kobayashi M, Teraishi M, Ito Y, Ochi K, Wakasa K, Tozawa Y. Differential expression of three plastidial sigma factors, OsSIG1, OsSIG2A, and OsSIG2B, during leaf development in rice. Biosci Biotech Bioch. 2004; 68(4):973–977.
9. Kubota Y, Miyao A, Hirochika H, Tozawa Y, Yasuda H, Tsunoyama Y, Niwa Y, Imamura S, Shirai M, Asayama M. Two novel nuclear genes, OsSIG5 and OsSIG6, encoding potential plastid sigma factors of RNA polymerase in rice: tissue-specific and light-responsive gene expression. Plant Cell Physiol. 2007; 48(48):186–192.
10. Tozawa Y, Teraishi M, Sasaki T, Sonoike K, Nishiyama Y, Itaya M, Miyao A, Hirochika H. The plastid sigma factor SIG1 maintains photosystem I activity via regulated expression of the psaA operon in rice chloroplasts. Plant J. 2007; 52(1):124–132.
11. Tan S, Troxler RF. Characterization of two chloroplast RNA polymerase sigma factors from *Zea mays*: Photoregulation and differential expression. Proc Natl Acad Sci USA. 1999; 96(96):5316–5321.
12. Beardslee TA, Roy-Chowdhury S, Jaiswal P, Buhot L, Lerbs-Mache S, Stern DB, Allison LA. A nucleus-encoded maize protein with sigma factor activity accumulates in mitochondria and chloroplasts. Plant J. 2002; 31(2):199–209.
13. John CR, Smith-Unna RD, Woodfield H, Covshoff S, Hibberd JM. Evolutionary convergence of cell-specific gene expression in independent lineages of C_4 grasses. Plant Physiol. 2014; 165(1):62–75.
14. Jia GQ, Huang XH, Zhi H, Zhao Y, Zhao Q, Li WJ, Chai Y, Yang LF, Liu KY, Lu HYet al. A haplotype map of genomic variations and genome-wide association studies of agronomic traits in foxtail millet (*Setaria italica*). Nat Genet. 2013; 45(8):957–961.
15. Yuan M, Qu LJ, Wang XJ, Qian Q, Yang WC, Wang T, Kong HZ, Jiang GM, Chong K. Research advances on plant science in China in 2013. Chinese Bulletin of Botany. 2014; 49:347–406.
16. Feng XL, Zhao ZH, Wang XM, Qiu FC, Song GL, Wang DQ, Su X, Zhang XL, Wang F. Recent Research Progress in Foxtail Millet (*Setaria italica*). Agr Sci & Technol. 2014; 15:564–570, 575.
17. Lin YX, Jiang HY, Chu ZX, Tang XL, Zhu SW, Cheng BJ. Genome-wide identification, classification and analysis of heat shock transcription factor family in maize. BMC Genomics. 2011; 12(1):76.
18. Guo AY, Zhu QH, Chen X, Luo JC. GSDS: a gene structure display server. Hereditas (Beijing). 2007; 29(8):1023–1026.
19. Wass MN, Kelley LA, Sternberg MJ. 3DLigandSite: predicting ligand-binding sites using similar structures. Nucleic Acids Res. 2010; 38 (Suppl.):W469–473.
20. Lovell SC, Davis IW, Arendall WB, de Bakker PIW, Word JM, Prisant MG, Richardson JS, Richardson DC. Structure validation by Cα geometry: ϕ, ψ and Cβ deviation. Proteins. 2003; 50:437–450.
21. Emanuelsson O, Nielsen H, von Heijne G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol. 2000; 300(4):1005–1016.
22. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol. 2007; 24(8):1596–1599.
23. Zhang Q, Wang ZY, Ma XL, Nie LM, Wang HL, Wang JP. Comparative analysis of paleopolyploidy evolution in genomes of *Setaria italica* and *Zea mays*. J Henan Agric Sci. 2014; 43(6):10–15.
24. Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, et al. The B37 maize genome: Complexity, diversity and dynamics. Science. 2009; 326(5956):1112–1115.
25. Bennetzen JL, Schmutz J, Wang H, Percifield R, Hawkins J, Pontaroli AC, Estep M, Feng L, Vaughn JN, Grimwood J, et al. Reference genome sequence of the model plant *Setaria*. Nat Biotechnol. 2012; 30(6):555-561.
26. Zhang GY, Liu X, Wang J. Genome sequence of foxtail millet (*Setaria italica*) provides insights into grass evolution and biofuel potential. Nat Biotechnol. 2012; 30(6):549–556.
27. Isono K, Shimizu M, Yoshimoto K, Niwa Y, Satoh K, Yokota A, Kobayashi H. Leaf-specifically expressed genes for polypeptides destined for chloroplasts with domains of σ^{70} factors of bacterial RNA polymerases in *Arabidopsis thaliana*. Proc Natl Acad Sci USA. 1997; 94(26):14948–14953.
28. Schweer J, Türkeri H, Kolpack A, Link G. Role and regulation of plastid sigma factors and their functional interactors during chloroplast transcription–Recent lessons from *Arabidopsis thaliana*. Eur J Cell Biol. 2010; 89(12):940–946.
29. Lahiri SD, Yao J, McCumbers C, Allison LA. Tissue-specific and light-dependent expression within a family of nuclear-encoded σ-like factors from *Zea mays*. Mol Cell Biol Res Commun. 1999; 1(1):14–20.
30. Lahiri SD, Allison LA. Complementary expression of two plastid-localized sigma-like factors in maize. Plant Physiol. 2000; 123(3): 883–894.
31. Yao J, Roy-Chowdhury S, Allison LA. AtSig5 is an essential nucleus-encoded *Arabidopsis* σ-like factor. Plant Physiol. 2003; 132(2):739-47.
32. Tsunoyama Y, Ishizaki Y, Morikawa K, Kobori M, Nakahira Y, Takeba G, Toyoshima Y, Shiina T. Blue light-induced transcription of plastid-encoded *psbD* gene is mediated by a nuclear-encoded transcription initiation factor, AtSig5. Proc Natl Acad Sci USA. 2004; 101(9):3304–3309.
33. Nagashima A, Hanaoka M, Shikanai T, Fujiwara M, Kanamaru K, Takahashi H, Tanaka K. The multiple-stress responsive plastid sigma factor, SIG5, directs activation of the *psbD* blue light-responsive promoter (BLRP) in *Arabidopsis thaliana*. Plant Cell Physiol. 2004; 45(4):357–368.
34. Favory JJ, Kobayahi M, Tanaka K, Peltier G, Kreis M, Valay JG, Lerbs-Mache S. Specific function of a plastid sigma factor for ndhF gene transcription. Nucleic Acids Res. 2005; 33(18):5991–5999.

35. Borsellino L. Influence of light and cytokinin on organellar phage-type RNA polymerase transcript levels and transcription of organellar genes in *Arabidopsis thaliana*. PhD thesis, Humboldt Universität zu Berlin. 2011.
36. Cortleven A, Schmülling T. Regulation of chloroplast development and function by cytokinin. J Exp Bot. 2015; 66(16):4999–5013.
37. Pick TR, Bräutigam A, Schlüter U, Denton AK, Colmsee C, Scholz U, Fahnenstich H, Pieruschka R, Rascher U, Sonnewald U, Weber AP. Systems analysis of a maize leaf developmental gradient redefines the current C4 model and provides candidates for regulation. Plant Cell. 2011; 23(12):4208–4220.
38. Wang L, Czedik-Eysenberg A, Mertz RA, Si Y, Tohge T, Nunes-Nesi A, Arrivault S, Dedow LK, Bryant DW, Zhou W, et al. Comparative analyses of C4 and C3 photosynthesis in developing leaves of maize and rice. Nat Biotechnol. 2014; 32(11):1158–1165.
39. Aubry S, Kelly S, Kumpers BMC, Smith-Unna RD, Hibberd JM. Deep evolutionary comparison of gene expression identifies parallel recruitment of trans-factors in two independent origins of C_4 photosynthesis. PLOS Genet. 2014; 10(6):e1004365.

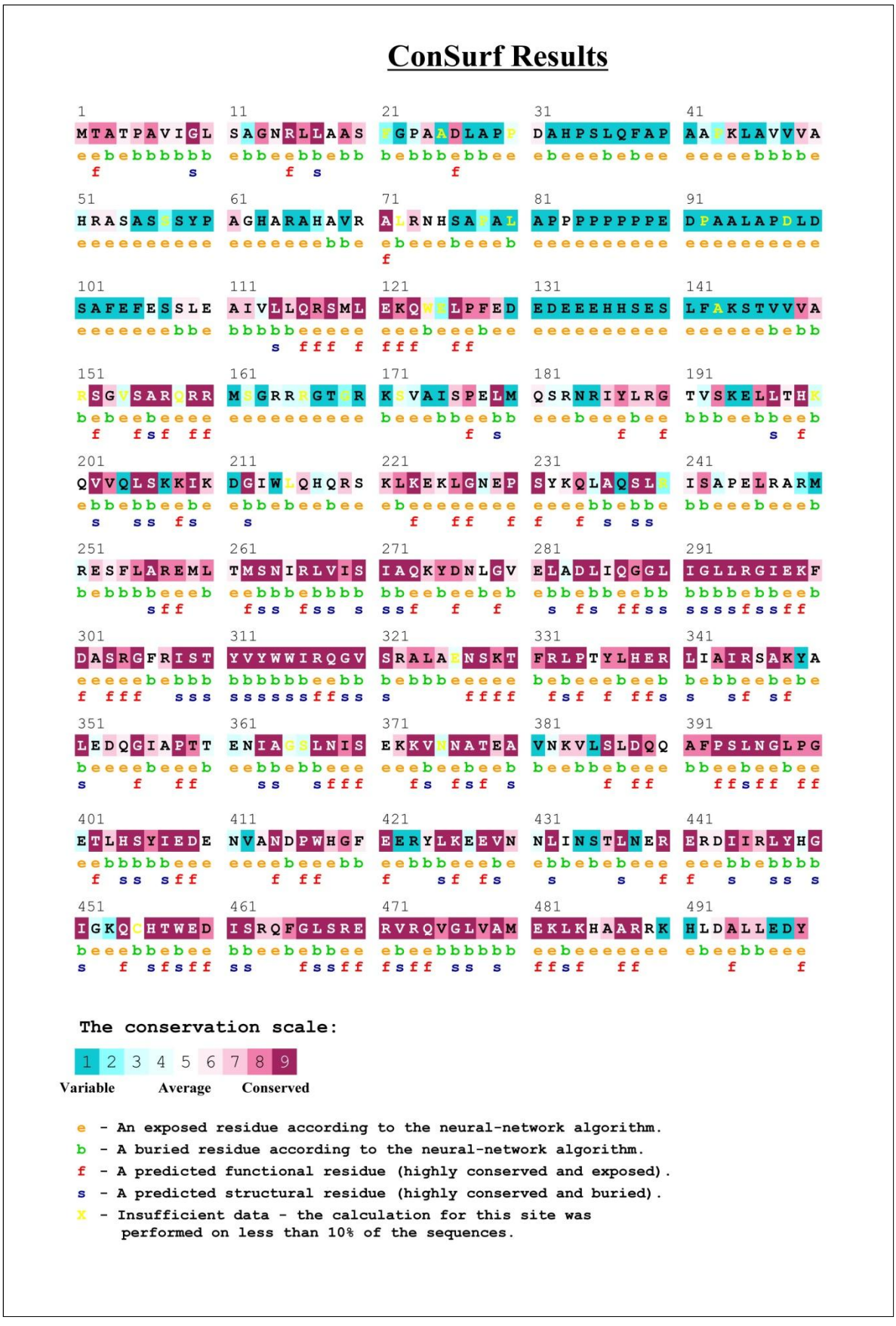
How to cite this article:

Hongyun Liu, Jinjin Cheng, Siyuan Cheng, Hui Fan, Bo Wen, Zheng Liu., Genome-wide identification and expression analysis along the leaf developmentalgradient of the sigma factor gene family in foxtail millet (*Setaria italica*). J App Biol Biotech. 2016; 4 (04): 011-030. DOI: 10.7324/JABB.2016.40402

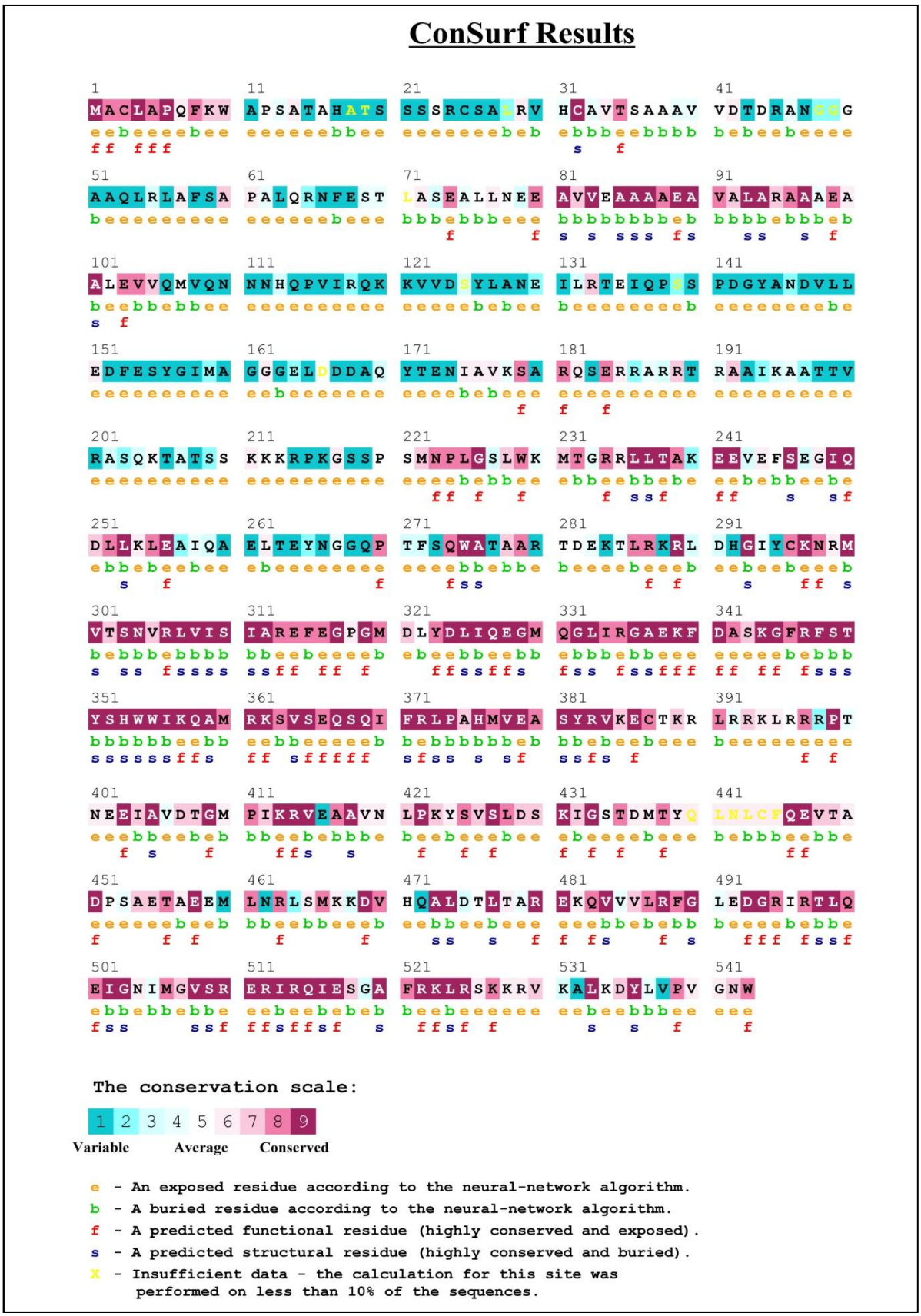
7. SUPPLEMENTARY DATA

Suppl. 1 The position information of the conserved motifs in sigma factor proteins of millet.

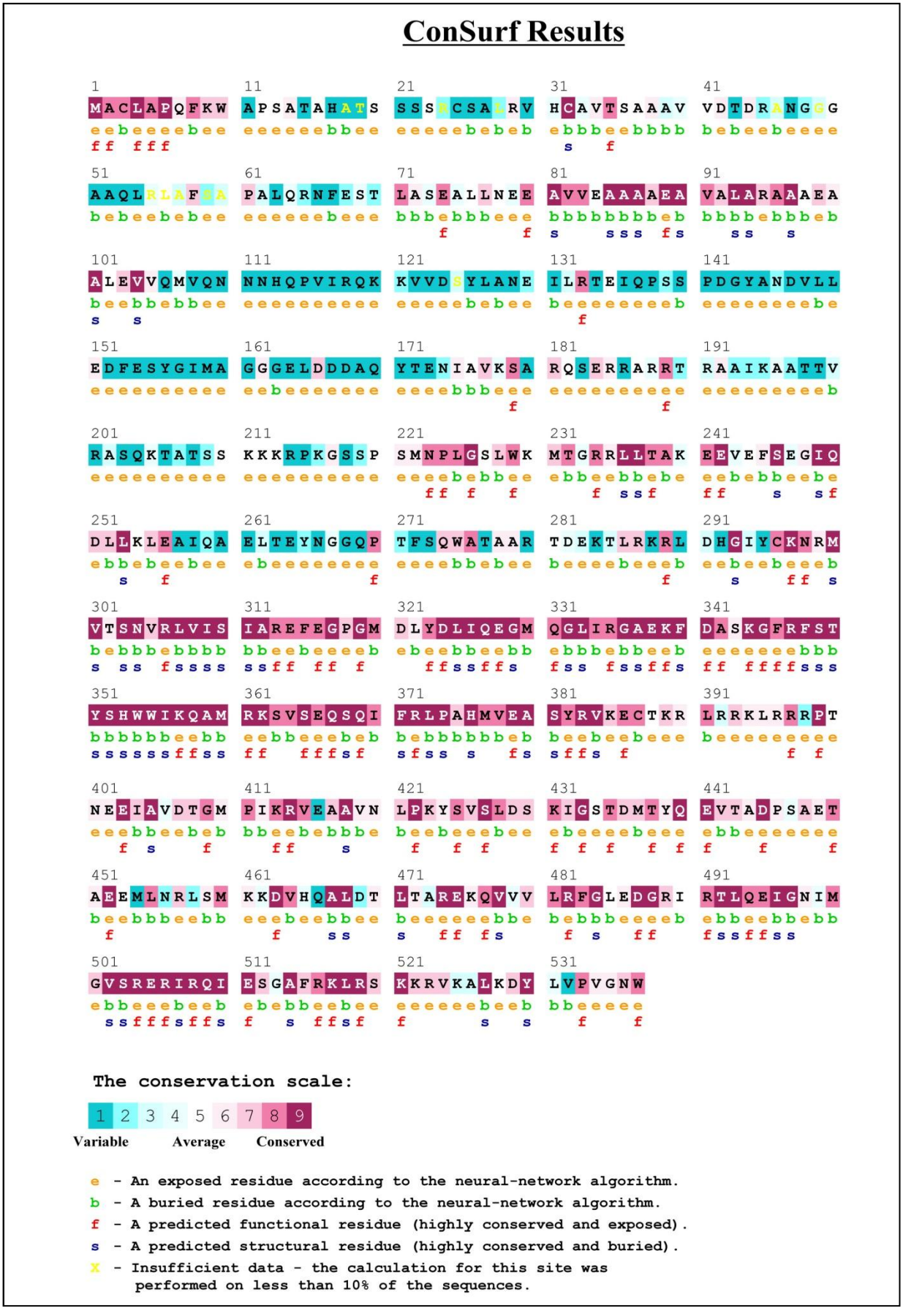
Query	Pfam	Position (independent E-value)	Description
Sisig 1	Sigma70_r3	339..411(1.2e-12)	PF04539, Sigma-70 region 3
	Sigma70_r2	262..329(2.9e-15)	PF04542, Sigma-70 region 2
	Sigma70_r4	436..485(2.1e-10)	PF04545, Sigma-70 region 4
Sisig 2Aa	Sigma70_r4	473..525(2.3e-20)	PF04545, Sigma-70 region 4
	Sigma70_r2	300..368(4.3e-16)	PF04542, Sigma-70 region 2
	Sigma70_r3	379..460(3.3e-14)	PF04539, Sigma-70 region 3
Sisig 2Ab	Sigma70_r4	467..519(2.2e-20)	PF04545, Sigma-70 region 4
	Sigma70_r3	379..454(4.6e-18)	PF04539, Sigma-70 region 3
	Sigma70_r2	300..368(4.3e-16)	PF04542, Sigma-70 region 2
Sisig 2B	Sigma70_r4	462..515(1.5e19)	PF04545, Sigma-70 region 4
	Sigma70_r3	374..449(2.6e-17)	PF04539, Sigma-70 region 3
	Sigma70_r2	295..360(2.2e-16)	PF04542, Sigma-70 region 2
Sisig 3	Sigma70_r4	500.553(3.3e-12)	PF04545, Sigma-70 region 4
	Sigma70_r3	413..487(1.8e-12)	PF04539, Sigma-70 region 3
	Sigma70_r2	339..397(6.3e-11)	PF04542, Sigma-70 region 2
Sisig 5	Sigma70_r3	334..401(6.2e-14)	PF04539, Sigma-70 region 3
	Sigma70_r4	240.473(5.9e-12)	PF04545, Sigma-70 region 4
	Sigma70_r2	252..318(8.9e⁻11)	PF04542, Sigma-70 region 2
Sisig 6	Sigma70_r3	399.473(3e-13)	PF04539, Sigma-70 region 3
	Sigma70_r2	320..388(2.6e-19)	PF04542, Sigma-70 region 2
	Sigma70_r4	487..539(3.3e-16)	PF04545, Sigma-70 region 4


ConSurf Results
The conservation scale:
1 2 3 4 5 6 7 8 9
Variable
Average
Conserved
e - An exposed residue according to the neural-network algorithm.
b - A buried residue according to the neural-network algorithm.
f - A predicted functional residue (highly conserved and exposed).
s - A predicted structural residue (highly conserved and buried).
X - Insufficient data - the calculation for this site was performed on less than 10% of the sequences.

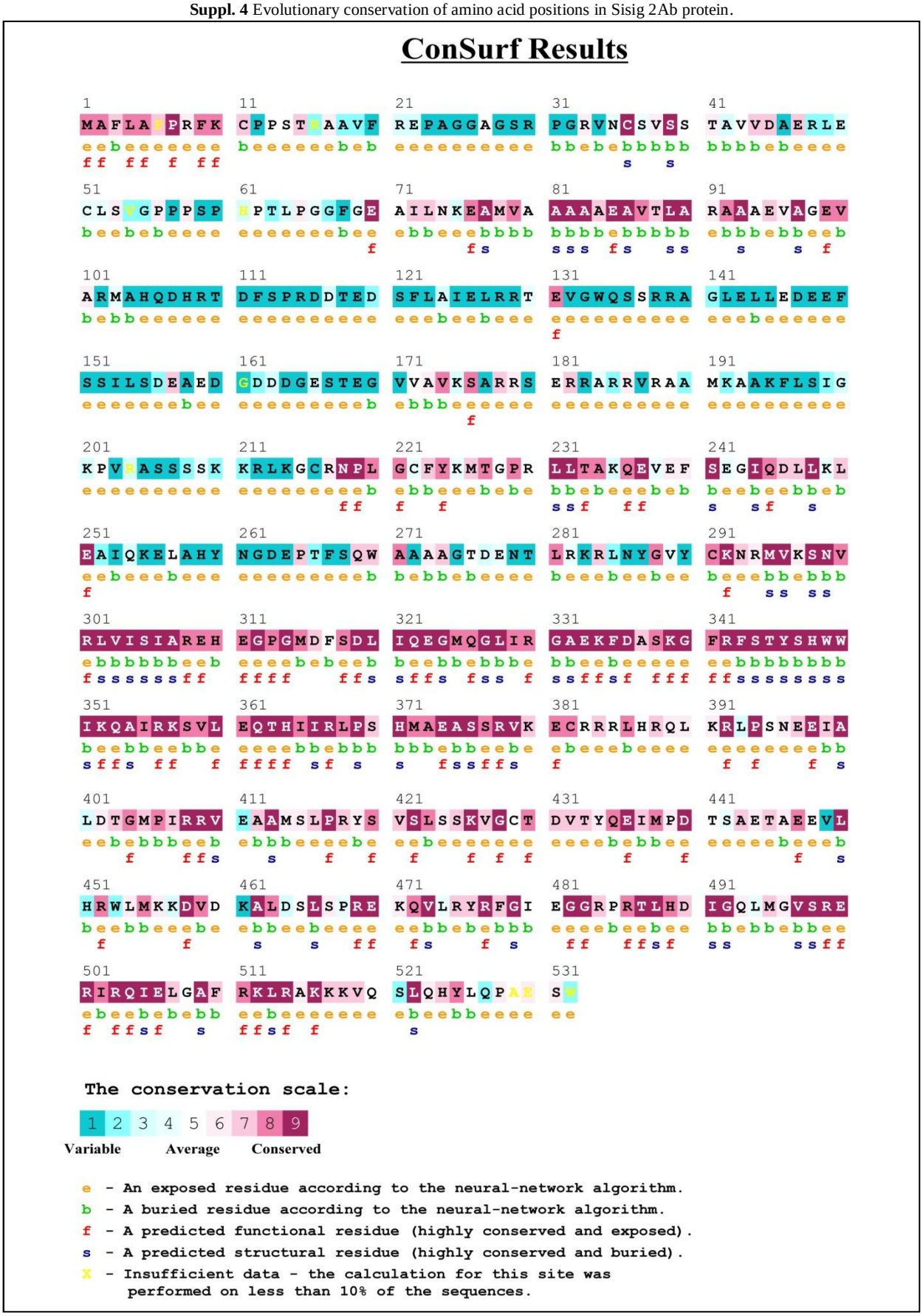
Suppl. 2 Evolutionary conservation of amino acid positions in Sisig 1 protein.



Suppl. 3 Evolutionary conservation of amino acid positions in Sisig 2Aa protein.

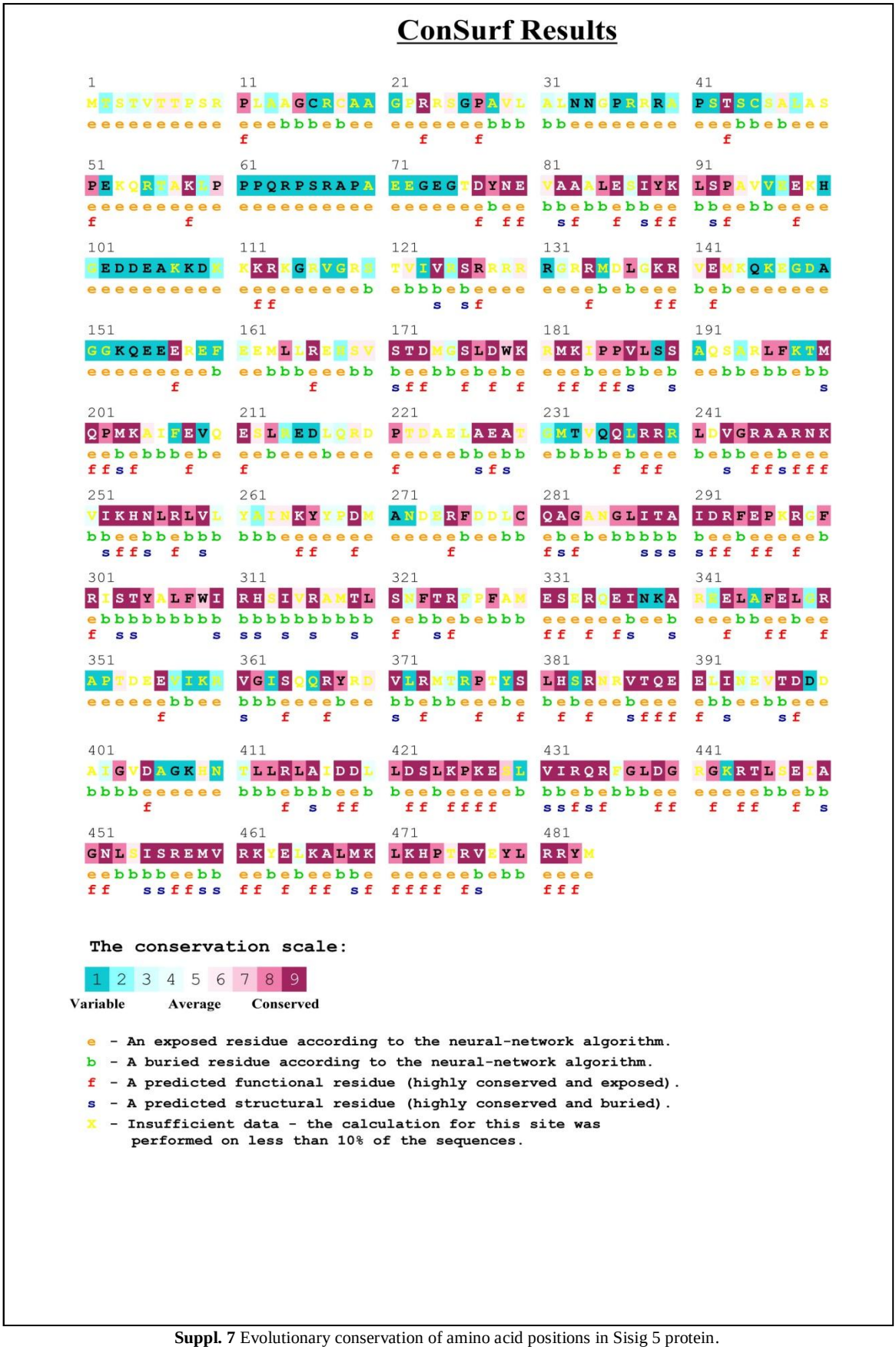

ConSurf Results
The conservation scale:
1 2 3 4 5 6 7 8 9
Variable
Average
Conserved
e - An exposed residue according to the neural-network algorithm.
b - A buried residue according to the neural-network algorithm.
f - A predicted functional residue (highly conserved and exposed).
s - A predicted structural residue (highly conserved and buried).
X - Insufficient data - the calculation for this site was performed on less than 10% of the sequences.

Suppl. 4 Evolutionary conservation of amino acid positions in Sisig 2Ab protein.

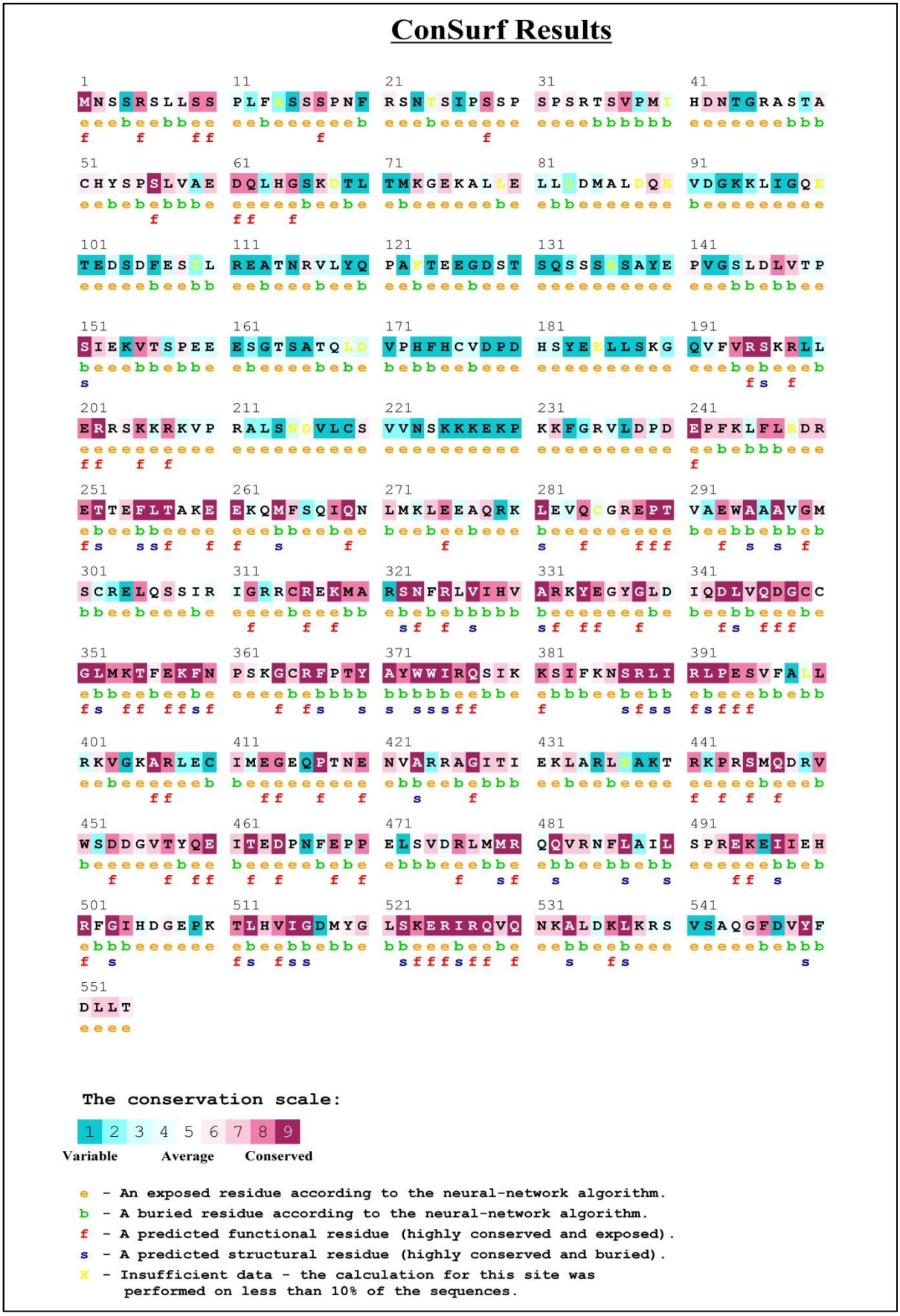


Suppl. 5 Evolutionary conservation of amino acid positions in Sisig 2B protein.

Suppl. 6 Evolutionary conservation of amino acid positions in Sisig 3 protein.



Suppl. 7 Evolutionary conservation of amino acid positions in Sisig 5 protein.



Suppl. 8 Evolutionary conservation of amino acid positions in Sisig 6 protein.

Suppl. 9 the sigma factor genes identified in Arabidopsis thaliana, rice, and maize.

SPECIES	ARABIDOPSIS THALIANA	ORYZA SATIVA	ZEA MAYS
SIG1	AT1G64860	OS08G0163400	GRMZM2G543629 GRMZM2G006736
SIG2	AT1G08540	OS11G0448400	GRMZM2G100086
			GRMZM2G003182 GRMZM2G143392
SIG3	AT3G53920	OS05G0589200	GRMZM5G830932
SIG4	AT5G13730		
SIG5	AT5G24120	OS05G0586600	GRMZM2G077436
SIG6	AT2G369900	OS08G0242800	GRMZM2G144196

Suppl. 10 The promoter *cis*-acting elements of the sigma factor genes in millet.

	Signals	si013608m Sisig1	Si026193m Si026195m Sisig2A	Si035112m Sisig2B	Si021619m Sisig3	Si021904m Sisig5	Si034994m Sisig6
Hormone	Auxin	TGTCTC YTGTCWC	TGACG GAGAC TGTCTC YTGTCWC ACTTTA	TGACG GAGAC ACTTTA	TGACG GAGAC YTGTCWC	TGACG GAGAC	TGACG GAGAC TGTCTC YTGTCWC
	ABA	ACACNNG CAAT	CACATG CAAACACC CAAT YACGTGGC ACGTGKC CACGTGGC	CACATG CAAT WAACCA	TGCCACCG G	ACACNNG CAAT ACGTSSSC	CAAT WAACCA ACGTSSSC ACGTGKC
	Cytokinin			TATTAG			
	GA	TGAC	TGAC	TGAC TAACAAR	TGAC TAACAAA TAACAAR	TGAC	TGAC
	SA	GRWAAW	TGACG GRWAAW	TGACG TTGAC	TGACG GRWAAW	TGACG GRWAAW GATAAG	TGACG GRWAAW
	JA	GCCGCC			GCCGCC		AACGTG
	Ethylene	GCCGCC			GCCGCC		
Environment	Water stress	ACGTG ACGT	CACATG CNGTTR ACGTG ACGT	WAACCA CACATG ACGTG ACGT	CNGTTR CATGTG	CNGTTR CATGTG ACGTG ACGT	WAACCA YAACKG ACGTG ACGT
	Light regulation	GCCAC GGGCC GRWAAW GATAA SCGAYNR NNNNNNN NNNNNNN NHD YTCANTY Y ATACGTG T YTYYMM CMAMCM MC GATA GATAA	CTCCTAATT TGACG GGGCC GRWAAW GATAA ACGTGGCA ACGTGGC CCGTCC GATA GATAA	TGACG ACTTTG GATAA SCGAYNRN NNNNNNN NNNNNNN HD YTCANTYY GATA GATAA	TGACG ACTTTG GCCAC GRWAAW GATAA CCGTCC GATA GATAA	TGACG GGGCC GRWAAW GATAA YTCANTYY GATAAG GATA GATAA	TGACG GGGCC GRWAAW GATAA SCGAYNRN NNNNNNN NNNNNNN HD TATTCT ACGTGGC GATA GATAA

	Copper response	GTAC		GTAC	GTAC	GTAC	GTAC
	Calcium response	VCGCGB MACGYGB	MACGYGB	MACGYGB	VCGCGB MACGYGB	VCGCGB MACGYGB	VCGCGB MACGYGB
	Iron deficiency		CACGTGG			CACGTGG	
	Sulfurresponse		GAGAC	GAGAC	GAGAC	GAGAC	GAGAC
	phosphorus starvation			GNATATNC			
	Salt stress	GAAAAA	GAAAAA			GAAAAA	GAAAAA
	Oxygen response	GTAC	GTAC	GTAC	GTAC	GTAC	GTAC
	CO_2 response			GANTTNC	GANTTNC	GANTTNC	
	Low temperature	CANNTG CANNTG	CANNTG CANNTG	CANNTG CANNTG	CANNTG CANNTG	CANNTG CANNTG	CANNTG CANNTG
Protein	heatshock protein	CCAAT	CCAAT	CCAAT		CCAAT	CCAAT
	Dof protein	AAAG	AAAG	AAAG	AAAG	AAAG	AAAG
	histone	CCGTCG				ACGTCA	ACGTCA
	storage protein		CAAACAC CNAACAC				
	cyclin				AACGG	AACGG	AACGG
	Chloroplast ribosomal protein			ATGGTA ATGGTATT			
Enzyme	α-amylase	CGACG	TGACGT AATTAAA TATCCA	TGACGT	TAACAAA TAACARA AATAAA	AATAAA	AATAAA TGACGT CGACG
	β-amylase			TACTATT			
	isoamylase	TGACT	TGACT	TGACT			
	glutamine synthetase	TTATTT	TTATTT				TTATTT
Pigment	phytochrome	CAAT	CAAT	CAAT AACCAA	AACCAA CGGATA	CAAT	CAAT AACCAA
Sp	tissue	CAAT CACGTG	CAAT CACGTG	CAAT CACGTG	GATA	CAAT CACGTG	CAAT GATA
ecificity		GATA	GATA	GATA		GATA	
	Mesophyll cell	YACT	YACT	YACT	YACT	YACT	YACT
	Fruit		TGTCACA				
	Fibrils	KCACGW	KCACGW	KCACGW		KCACGW	
	Stimulation induced		CTGAC	TTGACC	CTGAC		

Others	Pollen	GTGA	GTGA AGAAA	GTGA AGAAA	GTGA	GTGA AGAAA	GTGA AGAAA
	Root nodule	CTCTT	AAAGAT CTCTT	AAAGAT CTCTT	AAAGAT CTCTT	AAAGAT	CTCTT
	Polyadenylation			AATAAT			
	Oxidative phosphorylation	TGGGCY			TGGGCY	TGGGCY	TGGGCY
	napA		CAAACAC CNAACAC				
	Disease-resistant		TGTCA		TGTCA		
	Sugar repressive		TTATCC		TTATCC		
	MYB		GGATA	GGATA	MACCWAM C GGATA		
	MYC						CAACGTG
	E2F	TYTCCCG CC				TYTCCCGC C	
	RAV1		CAACA		CACCTG	CAACA	
	Flavonoid biosynthesis		CNGTTR		CNGTTR	CNGTTR	
	Splice junction						TGCAGG
	Injured		NGATT	NGATT	NGATT	NGATT	NGATT