Genome-wide identification, phylogeny and expression analyses of *SCARECROW-LIKE*(*SCL*) genes in millet (*Setaria italica*)

Hongyun Liu¹ · Jiajia Qin² · Hui Fan¹ · Jinjin Cheng¹ · Lin Li³ · Zheng Liu¹

Received: 30 November 2016/Revised: 13 May 2017/Accepted: 29 May 2017/Published online: 14 June 2017
© Prof. H.S. Srivastava Foundation for Science and Society 2017

Abstract As a member of the GRAS gene family, *SCARECROW-LIKE* (*SCL*) genes encode transcriptional regulators that are involved in plant information transmission and signal transduction. In this study, 44 *SCL* genes including two *SCARECROW* genes in millet were identified to be distributed on eight chromosomes, except chromosome 6. All the millet genes contain motifs 6–8, indicating that these motifs are conserved during the evolution. *SCL* genes of millet were divided into eight groups based on the phylogenetic relationship and classification of *Arabidopsis* *SCL* genes. Several putative millet orthologous genes in *Arabidopsis*, maize and rice were identified. High throughput RNA sequencing revealed that the expressions of millet *SCL* genes in root, stem, leaf, spica, and along leaf gradient varied greatly. Analyses combining the gene expression patterns, gene structures, motif compositions, promoter *cis*-elements identification, alternative splicing of transcripts and phylogenetic relationship of *SCL* genes indicate that these genes may play diverse functions. Functionally characterized *SCL* genes in maize, rice and *Arabidopsis* would provide us some clues for

future characterization of their homologues in millet. To the best of our knowledge, this is the first study of millet *SCL* genes at the genome wide level. Our work provides a useful platform for functional analysis of *SCL* genes in millet, a model crop for C₄ photosynthesis and bioenergy studies.

Keywords *Setaria italica* · GRAS family · *SCARECROW-LIKE* (*SCL*) genes · Bioinformatics · Gene expression patterns

Introduction

Transcription factors (TFs) are regulatory proteins that control the transcriptional level of a set of given target genes (Lee et al. 2008; Wang et al. 2016). GRAS is an important plant-specific gene family of transcription factors, named after the three known members (*GAI*, *RGA* and *SCR*) (Lee et al. 2008; Tian et al. 2004). Although only a few GRAS genes were well characterized from different plant species, they are functionally versatile in regulating plant growth and development, and are at least involved in C₄ Kranz initiation, gibberellins and phytochrome A signaling pathways, axillary meristem formation, root radial patterning, and the occurrence of male gametophyte (Tian et al. 2004).

As a member of GRAS protein family, *SCARECROW* (*SCR*) genes play crucial roles in the development of both roots and shoots (Gao et al. 2004, 2015; Heidstra et al. 2004; Kamiya et al. 2003; Lim et al. 2005; Sbabou et al. 2010; Solé et al. 2008; Wysocka-Diller et al. 2000). *SCR*s are essential for root cell radial patterning, the quiescent centre (QC) maintenance and the endodermis differentiation in *Arabidopsis*, maize (*Zea mays* L.) and rice (*Oryza*

Electronic supplementary material The online version of this article (doi:10.1007/s12298-017-0455-6) contains supplementary material, which is available to authorized users.

✉ Zheng Liu
liuzhengxp@yeah.net

- ¹ College of Life Sciences, Hebei University, Baoding 071002, People's Republic of China
- ² School of Physical Sciences, University of the Chinese Academy of Sciences, Beijing 100049, People's Republic of China
- ³ College of Biology, Hunan University, Changsha 410082, People's Republic of China

sativa L.) (Lim et al. 2000; Sbabou et al. 2010). SCRs are also necessary for the formation of root endodermis in *Arabidopsis* and rice. The interactions between SCR and SHORTROOT (SHR) proteins, lead SHR being combined to the endodermal cells nucleus, thus restrain their further migration to other root cell layers and that is why all plants have a single endodermis layer (Sbabou et al. 2010). Besides, SCR and SHR play crucial roles in the regeneration of the root tip both in *Arabidopsis* and maize after laser ablation or excision (Solé et al. 2008). Researches on *Arabidopsis* showed that the expression of *SCR* genes was closely related to chromatin assembly factor-1 (CAF-1). Mutation of FAS1 and FAS2, two *Arabidopsis* CAF-1 subunits, resulted in the disruption of *SCR* gene expression and led to the organizational derangement of both shoot and root apical meristem (SAM and RAM) (Gao et al. 2004). SCR protein was reported to be involved in the formation of accumulating starch sheath layer of the hypocotyls and formation of radial patterning in the hypocotyls and the inflorescence (Gao et al. 2004). The *SCR* genes also play a role in the establishment of Kranz structure in C₄ plants (e.g. maize). SCR mutation causes vascular bundle sheath cell proliferation, abnormal vascular bundle sheath chloroplast differentiation, vein disorientation, loss of small veins and decrease in vein density (Slewinski et al. 2012).

Some gene products are highly similar to the C terminus of the *SCR* gene products, and these genes are called *SCARECROW-LIKE* (*SCL*) (Pysh et al. 1999). They are also related to plant information transmission and signal transduction, for example, *SCL13* and *SCL21* play roles in signal transduction of phytochrome A (Torres-Galea et al. 2013). In addition, studies have shown that some *SCL* genes are involved in the regulation of plant abiotic stresses, such as drought or salt stress (Zhou et al. 2013). Over-expression of *Populus euphratica SCL7* in *Arabidopsis thaliana* significantly improves the tolerance to salt and drought (Ma et al. 2010; Zhou et al. 2013). *SCL* genes are well-known to be under transcriptional regulation. Interestingly, several *SCL* genes are also regulated by micro-RNAs (miRNAs) (Llave and Carrington 2002).

With the complete genome sequencing of foxtail millet accomplished, more and more researches had been done on millet by using the second generation high-throughput sequencing technology (Jia et al. 2013). All these provide reliable data resources for the study of millet *SCL* genes which are not investigated in this important crop to our knowledge. Compared with rice, *Arabidopsis* and maize, researches on molecular genetics of millet started relatively late and most researches are preliminary (Feng et al. 2014; Liu et al. 2016). Thus, we did bioinformatics research and expression analyses on *SCL* gene family in millet in order for a better understanding of these important genes.

Materials and methods

Identification of the *SCL* genes in *Setaria italica*

Arabidopsis SCL genes were obtained from the PlnTFDB database (<http://plntfdb.bio.uni-potsdam.de/v3.0/>) (Perezrodriguez et al. 2010). Then blastP of these genes was conducted on the website NCBI (<http://www.ncbi.nlm.nih.gov/>), Phytozome (<http://www.phytozome.net/>) and SIFGD (<http://structuralbiology.cau.edu.cn/SIFGD/>) (Qi et al. 2015) to find candidate *SCL* genes of *Setaria italica*. The candidate *SCL* genes were analyzed on SMART database (<http://smart.embl-heidelberg.de/>) (Letunic 2015) and CD Search (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi?>) with default parameters to inspect whether they have the conserved domain: GRAS (Pfam03514), and further the best alignments against the signature motifs (VHIID, PFYRE, and SAW) were performed to identify *SCL* genes as described by Pysh et al. (1999) and Lim et al. (2005). With the same methods, *SCL* genes of maize and rice were identified from NCBI and Phytozome.

Locations and structures of *SCL* genes in millet

The information of chromosomal locations of *SCL* genes were obtained from Phytozome and SIFGD, and the chromosome location image of *SCL* genes was generated by MapInspect (Liu et al. 2011). To predict the exon–intron structures of the *SCL* genes, GSDS (<http://gsds.cbi.pku.edu.cn/>) was used by comparing the genomic sequence of each gene with its coding sequence (CDS).

Motif display, promoter and phylogenetic analysis of the *SCL* genes in *Setaria italica*

The conserved motifs of *SCL* protein sequences of foxtail millet were analyzed on MEME server (<http://meme-suite.org/tools/meme>). Sequences of 1500 bp upstream of the transcription start sites of millet *SCL* genes were analyzed on PlantCARE (Lescot et al. 2002) (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>). MEGA 7.0 was used to compute the evolutionary distances and construct the phylogenetic trees of *SCL* genes in different plant species by using Clustal W for the alignment of amino acid sequences of *SCL* genes. The neighbor-joining (NJ) method was then applied to this analysis with 1000 replications of bootstrap.

Gene expression analyses of *SCL* genes in millet

Total RNAs were extracted from four gradient regions of millet young leaves, i.e. leaf sheath area (LS), leaf base area (LB), leaf middle area (LM) and leaf tip area (LT) by

the TRIzol reagent (Life Technologies, USA) in accordance with the manufacture's instruction manual. Then the total RNAs were treated with RNase-free DNase (Life Technologies, USA) to eliminate residual DNA. Nanodrop spectrophotometer (Thermo Scientific) was used to check protein and reagent contamination, and agarose gel electrophoresis to check RNA purity and degradation. mRNAs were enriched by oligo (dT) magnetic beads (Thermo-fisher, USA). Three separate replicates for each leaf gradient region were used for digital gene expression profiling (DGE) through Illumina deep-sequencing by LC Sciences (Hangzhou, China) to investigate the expression patterns of these *SCL* genes. The RPKM (reads per kilobase of exon model per million mapped reads) values of the millet *SCL* genes were then picked out. Heat map of the expression patterns of these *SCL* genes along leaf gradient were generated by using the HemI software (Deng et al. 2014). Moreover, The FPKM (Fragments Per Kilobase of exon model per million mapped reads) values of high-throughput RNA-seq on root, stem, leaf and spica of foxtail millet were obtained from SIFGD. Heat maps of the expression profiling of *SCL* genes in these four tissues were also generated by HemI.

Results and discussion

Transcription factors are a class of proteins which bind to a specific nucleotide sequence upstream of a gene to regulate the gene transcription. They are involved in the development of plant growth and morphogenesis (Balazadeh et al. 2008; Broun 2004) as well as are necessary for the plant defense responses (Chen and Zhu 2004; Chen et al. 2002; Singh et al. 2002). GRAS is a family of plant specific transcription regulation factors having been reported to regulate plant growth and development (Tian et al. 2004). *SCL* genes and *SCR* genes, a subfamily of *GRAS*, are highly similar on the C terminus of their amino acid sequences. Although *SCL* genes attracted great attentions on plant researchers and remarkable progresses have been made in some species of plants such as *Arabidopsis* (Fode et al. 2008; Ma et al. 2010; Zhang et al. 2011), cucumber (Liu et al. 2011), white lupin (Sbabou et al. 2010), pine (Laajanen et al. 2007), lily (Morohashi et al. 2003), wheat (Chen et al. 2015) and so on, *SCL* genes in millet, to our knowledge, have not been studied.

Locations and structures of the *SCL* genes in *Setaria italica*

In the study, there are 44 *SCL* genes including two *SCR* genes of millet (*Seita.7G305500* and *Seita.8G008700*) were identified in total based on the presence of the

signature motifs (VHIID, PFYRE, and SAW) (Suppl. 1, 2 and 3). These genes are distributed on eight chromosomes, and no *SCL* gene was found on the chromosome 6 (Fig. 1). Moreover, there are more than eight *SCL* genes clustering on chromosome 3, 7 and 9, respectively. In comparison, all chromosomes of *Arabidopsis* and maize contain *SCL* genes (Suppl. 4 and 5). For example, 43 *SCL* genes distributed on 10 chromosomes in maize and chromosome 4 contains the largest number of *SCL* genes, i.e. 8 *SCL* genes (Suppl. 5). To gain further insights into the structures of millet *SCL* genes, CDS sequences and their corresponding genomic sequences were used to generate exon–intron structures on the website of GSDS. The two millet *SCR* genes are highly similar in structure with two exons. While *Si025289m.g* contains the largest number of exons (i.e. five exons) among *SCL* genes in millet, up to 31 *SCL* genes (73.8%) have only one exon. Only two *SCL* genes (5.0%) have three exons and the other eight *SCL* genes (19.0%) have two exons (Fig. 2). Similarly, most *SCL* genes in *Arabidopsis* and maize have only one exon with only one and five genes containing more than one exon in *Arabidopsis* and maize, respectively (Suppl. 6, 7). Surprisingly, although *GRMZM2G089782* contains the largest number of introns among *SCL* genes in maize, its homologue in millet *Si008220m.g* has no intron (Fig. 2 and Suppl. 7). Alternative splicing analyses (Suppl. 8) revealed that *Si034417m.g* and *Si035052m.g* have more than one transcripts. *Si034417m* and *Si034418m* are two transcripts of *Si034417m.g*, and *Si034417m* contains longer 5' UTR than that of *Si034418m*. *Si035052m.g* has four transcripts, i.e. *Si035052m*, *Si035053m*, *Si035054m* and *Si035469m*. Among which, the longest 5' UTR can be found in *Si035469m*. Moreover, *Si035053m* and *Si035054m* have the shortest 3' UTR.

Motif display and promoter analyses on the *SCL* genes in *Setaria italica*

The 20 conserved motifs of millet *SCL* genes were identified by the MEME program (Fig. 3) and it was found that the motif 1–12 belong to GRAS super family according to CD search (Table 1). All *SCL* genes of millet contain the motif 6, 7 and 8, indicating these motifs are conserved during evolution. It was reported that in *Arabidopsis* (Pysh et al. 1999) and maize (Lim et al. 2005), *SCL* genes contain three conserved signature motifs of VHIID, PFYRE and SAW. In millet, VHIID spans the motifs 7, 2 and 11; PFTRE spans motifs 10, 8, 3 and 12; and SAW spans the motifs 1 and 4. By comparison with *Arabidopsis* *SCL* genes, the signature amino acids of VHIID in millet were found to contain P and N in sites 2 and 12 of motif 7, respectively; H, D, Q and L in sites 2, 5, 13 and 17 of motif 2, respectively; and V, R, I, T, G in sites from 9 to 13 of

Fig. 1 Locations of *SCL* genes on chromosomes of millet

motif 11 (Table 1). However, it was found that some signature amino acids in VHIID, for example, sites 2 and 12 of motif 7 and sites 2, 5, 13 and 17 of motif 2, are not 100% conserved in millet (Suppl. 1), although they are absolutely conserved in *Arabidopsis* (Pysh et al. 1999). This suggests that *SCL* genes experienced more times of mutation events in millet compared with their homologues in *Arabidopsis*. Similarly, the signature amino acids of PFTRE in millet are L, I, V, N from site 5 to 8 of motif 10; P in site 18 of motif 8; F and Y in site 5 and 16 of motif 3, respectively; K and E in site 3 and 9 of motif 12, respectively (Table 1). In this signature motif, only E in site 9 of motif 12 is observed to be absolutely conserved in millet *SCL* genes (Suppl. 2). Interestingly, three new 100% conserved sites were found in PFTRE of millet. These are sites 18 (D) and 19 (E) of motif 10 and site 5 (R) of motif 12, indicating that these three amino acid residues are specific signatures of millet *SCL* genes. Also, the signature amino acids of SAW in millet are designated as R, E, W and G in sites 11, 16, 21 and 29 of motif 1, respectively; and W in sites 10 and 21 of motif 4, respectively (Table 1). The absolutely conserved amino acids were identified as W and G only in motif 1 (Suppl. 3), although the others are highly conserved.

It was found that *Si024639m.g*, *Si012046m.g* and *Si012328m.g* contain the greatest number of motifs (i.e. 20 motifs) with the E-value calculation as zero. Interestingly, the motifs of these three genes exhibit similarity on sizes and arrangements along the linear sequence within these genes. Additionally, the number of a few of motifs is doubled in some *SCL* genes. For example, *Si034627m.g* contains two motif 2 and *Si039098m.g* contains two motif 6.

Furthermore, the analyses of *cis*-acting elements in the promoter regions of millet *SCL* genes revealed that they contained several categories of elements, e.g. light

responsive element, hormone responsiveness, cell cycle, elicitation, and so on (Suppl. 9). Existence of specific *cis*-acting elements upstream a *SCL* gene may suggest its specific functions under specific a condition. For example, *Si021635m.g* contains *cis*-acting element HSE which is involved in heat stress responsiveness, implying that this gene could respond to high temperature.

Phylogenetic analysis of the *SCL* genes

To evaluate the evolutionary and phylogenetic relationship among *SCL* genes in *Arabidopsis*, millet, rice and maize, multiple sequence alignment was performed by mega 7.0 by using full length peptide sequences of 44 millet, 24 *Arabidopsis*, 34 rice and 43 maize *SCL* genes. Then the alignment data were used to construct phylogenetic tree (Fig. 4). Based on the phylogenetic relationship and AtSCLs classification (Pysh et al. 1999; Lim et al. 2005), the tree was divided into eight groups, from group A to H, as monophyletic clades. The group A contains not only *SCR* genes from *Arabidopsis* (*AT3G54220* and *AT4G08250*), millet (*Seita.7G305500* and *Seita.8G008700*), rice (*Os12g022000* and *Os11g0124300*) and maize (*GRMZM2G015080* and *GRMZM2G131516*), but also contains five millet *SCL* genes which show high similarity on the motif distribution. The group D and F comprise the largest clades, both containing 24 members. As the group D contains eight millet genes, nine maize genes and six rice genes, but only one *Arabidopsis* gene, gene expansion in monocots is evident after monocot–dicot divergence. The group E contains the least members of *SCL* genes, only 11 genes including one millet gene. This group can be further divided into four subgroups, each containing no more than three members. The members that are closely related in phylogenetic tree share common

Fig. 2 Gene structures of the millet *SCL* genes. 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 (color figure online)

motif compositions as expected. For example, the group B contains four millet *SCL* genes which have the same motif patterns and share the motifs 1–12 and 14. The group C and D, however, contain 6 and 7 millet *SCL* genes, respectively, with similar motif patterns in each group. As a whole, millet *SCL* genes with similar motif patterns tend to cluster into the same group, indicating that protein sequence motifs can be signatures to show evolutionary distances of gene family members.

Expression analyses of millet *SCL* genes

The FPKM values of millet *SCL* genes in root, stem, leaf and spica through RNA-seq were obtained from SIFGD and used to generate heat map by the HemI (Fig. 5). Nearly 55% *SCL* genes in millet do not transcribe in these four tissues, indicating that these genes either work on millet specific developmental stages beyond the four detected organs, or are functionally invalid pseudogenes.

Fig. 3 Conserved motifs in millet SCL proteins characterized using MEME software. Different colored boxes represent different motifs, as indicated at the bottom of the figure (color figure online)

Phylogenetic tree exhibited that every group, from A to H, contains SCL genes which do not transcribe in the above four organs. Group D contains six genes which rarely show any transcript accumulation in the four organs, whereas group B, E and F contain only one gene respectively, which does not transcribe in these organs. For instance, maize

GRMZM2G106356 and *GRMZM2G023872* were specifically up-regulated in embryo and endosperm, respectively (Lu et al. 2013), however, their orthologue *Si024639m* in group D does not express in root, stem and leaf. As *Si024639m* did not also express in spica and it has the element of Skn-1_motif which is a *cis*-acting regulatory

Table 1 Motif sequences of SCL proteins identified by MEME tools

Motif	Width	Best possible match
1	41	LNVIACEGTD RMERHET YK QWQ ARMQ MAG FRQVPLNYYIVK
2	17	V HIID YGIMYGF QWPWL
3	29	NTPFFWTRFREALFYYSALFDMLDATIPR
4	21	EDNGWLLLG WKGRILY TVSTW
5	41	EQIETGYRLSKCAEMFGVPFEFNAIACKWETVDIEDLNID
6	21	GDPTQRLAHYFAEGLEARLVG
7	21	CPFFKFAYFFANQTILEAMKG
8	29	DITNPRDMVLNNIKKM NP KVFIHVEQNCN
9	29	LIHCAQAVAANNHTKANEILKQIKQHASP
10	21	PDEV LIV NCMYNFKHLMDES
11	15	KRPGGPPH VRIT GID
12	15	DSKERHLVERDCFGR
13	23	DMMLPYISRMLMEEDIDDKFKYH
14	21	TGSQIYKKLMCKRTSATDMLK
15	29	SCMDYVTMAFLKGME EANK FLPINNNLFR
16	21	PALLQAQKPFYQILSSPYPS
17	11	DKVKKCYHKDF
18	21	KETCAHEMFDKMMLCGYETCI
19	29	MATTPDELVEGLTEKSPQSPSVFLDVFS
20	41	RDELVNQMVSESSNHSGAKKRYARDDHIEEGEARRTSKSLM

Bold characters indicate the signature amino acid residues of *SCL* proteins

element required for endosperm expression, it suggests that *Si024639m* might be functional mainly during seed germination rather than seed maturation (Fig. 5). It was also found that some millet *SCL* genes only specifically express in one tissue. For example, *Si024734m* only specifically expresses in stem and *Si040237m* only in root. Even in the same tissue, *SCL* genes also showed differential transcription levels. *Si035054m* and *Si021635m* exhibited the highest transcription levels in leaf and in root respectively compared with other *SCL* genes. By checking the expression patterns of *GRMZM2G153333* in maize, the orthologue of *Si021635m*, it was found that it not only expresses on high level in primary root, but also in seed and anther (Stelpflug et al. 2015). Considering that *Si021635m.g* contains *cis*-acting elements of ABRE (abscisic acid responsiveness), JERE (jasmonate ethylene responsive) and Skn-1_motif (endosperm expression), *Si021635m.g* may also play versatile functions beyond root development during plant growth.

The expression levels of millet *SCL* genes along leaf gradient were shown in Fig. 6. As a whole, only 33.3% *SCL* genes can be detected to transcribe in either four parts (i.e. LS, LB, LM, or LT) of millet leaves. Five *SCL* genes, i.e. *Si000234m.g*, *Si016508m.g*, *Si024520m.g*, *Si021635m.g*, *Si035089m.g*, showed detectable transcription along leaf gradient, however, the transcripts of *Si000234m.g* in LS and of *Si016508m.g* and *Si024520m.g*

in LM and LT are low accumulated. *Si035089m.g* and *Si021635m.g* showed similar expression patterns along the leaf gradient. Both displayed high expressions in LS and LB with a sudden decline in LM and then a slight increase in leaf tip. Transcriptome assays in maize found that *GRMZM5G885274* and *GRMZM2G153333*, the orthologues of *Si035089m.g* and *Si021635m.g* in maize, respectively, do not show significant differential expressions along leaf developmental gradient (Wang et al. 2014), implying that their millet orthologues have been endowed new functions in leaf development during evolution. Phylogenetic analyses further revealed that although *Si035089m.g* and *Si021635m.g* are classified into the same group B, the evolutionary distance between them is not very close. Motif analyses indicated that *Si035089m.g* lacks the motif 9 located most closely to N-terminal region compared with the *Si021635m.g* although both genes contain the similar motif patterns and the similar motif arrangement. Further, Promoter analyses exhibited that the promoter of *Si021635m.g* contains more *cis*-acting elements compared with that of *Si035089m.g*. Thus, it is not surprising that these two genes accumulated their most abundant transcripts in different tissues, i.e. *Si035089m.g* in stem and *Si021635m.g* in root due to difference of promoters and protein motifs of these two genes (Fig. 5), and perhaps the additional motif 9 in *Si021635m* results in its evolutionary and functional divergence from

Fig. 4 Phylogenetic tree of SCL gene family of millet, rice, maize and *Arabidopsis*

Si035089m.g. Interestingly, *Si035089m.g* accumulated the highest transcript in stem compared with that in spica, leaf and root, suggesting that *Si035089m.g* may be involved in the formation of vascular tissue or vascular tissue related transportation rather than functioning on photosynthesis related events. Similarly, the orthologues of this gene in maize, *GRMZM5G885274* and in rice, *LOC_Os03g09280*, showed increased expression levels from LT towards LB (Wang et al. 2014), indicating that the similar function of

these orthologues may remain during evolution of monocots.

Untranslated regions (UTRs) always regulate the translation of mRNA in development and differentiation (de Moor et al. 2005; Zou et al. 2003). Furthermore, it has been demonstrated that UTRs could play an important role in directing cell specific expression and the UTRs of genes of C_3 plants can also direct bundle sheath specific expression in C_4 plants (Brown et al. 2011). In our study, it was found

Fig. 5 Expression patterns of millet *SCL* genes in root, stem, leaf and spica (*Si04417m* and *Si034418m* are the two alternative splicing transcripts of *Si034417m.g*; *Si035052m*, *Si035053m*, *Si035054m* and *Si035469m* are four transcripts of *Si035052m.g*)

that in four alternative splicing transcripts of *Si035052m.g*, *Si035054m* have the highest expression level in the four organs of spica, leaf, stem and root compared with the other three transcripts, and that non expression was detected in the four organs for *Si035052m*. Also, although two transcripts of *Si034417m.g* share the same expression pattern along leaf developmental gradient, the expression levels of *Si034417m* are higher than that of *Si034418m* in the four organs. As the main difference among the alternative splicing transcripts of these genes are UTR lengths, it suggests that the UTRs may play crucial roles in the strength of gene expression levels. Interestingly, *Si035052m.g* contains *cis*-acting elements of AC-I and AC-II, which are involved in negative regulation on phloem expression and responsible for restricting the vascular expression to the xylem, implying that this gene may be involved in differentiation of conductive tissues or related to the transport of materials.

Si035089m.g has close evolutionary relationships with the three genes (*AT1G21450*, *AT4G17230* and

AT1G50600) in *Arabidopsis* through phylogenetic analyses. Experiments showed that these three genes are induced by osmotic or drought stress in stem, leaf and root, especially for *AT4G17230* whose expression was significantly up-regulated during osmotic stress, and was obviously induced during the early drought treatment (Guo et al. 2009). Furthermore, *AT4G17230* has been found to be involved in red light signal transduction pathway dependent on the phytochrome A (Torres-Galea et al. 2006). Similarly, after checking the *cis*-elements of *Si035089m.g* promoter region according to PlantCARE database, it showed that this millet gene contains ABRE, G-Box, MBS motifs and motif IIb which are responsive to ABA or drought, and also contains Box4, LAMP-element and L-box which are light responsive elements. Therefore, detailed studies on the orthologous genes in *Arabidopsis* would provide some clues for the functional prediction of *Si035089m.g* in millet.

In *Arabidopsis*, *SCRs* are not only critical for the establishment of radial pattern in both root and shoot (Bolle

Fig. 6 Expression patterns of millet *SCL* genes along leaf gradient (LS leaf sheath, LB leaf base, LM leaf middle, LT leaf tip, *Si04417m* and *Si034418m* are the transcripts of *Si034417m.g*; *Si035052m*, *Si035053m*, *Si035054m* and *Si035469m* are transcripts of *Si035052m.g*)

2004), but also are important for functional differentiation of bundle sheath (BS) cells (Gao et al. 2014). For example, one *SCR* gene *AT3G54220* expresses mainly in the phloem-related BS cells to participate in sugar transport, while the other *SCR* gene *AT5G41920* is confined to accumulate its transcripts in the xylem-related BS cells and is necessary for mineral transport (Gao et al. 2014). Interestingly, their homologue in maize, *ZmSCR* (*GRMZM2G131516*), is also essential for Kranz formation (Slewiniski et al. 2012). Loss of *ZmSCR* function results in abnormal development of BS cells and of their chloroplasts in the cells (Slewiniski et al. 2012). Further, this gene can complement the *Arabidopsis* *SCR* mutant phenotype, indicating that *SCR* function remains conserved in dicots and monocots (Lim et al. 2005). Phylogenetic analyses indicate that two millet *SiSCR* genes (*Seita.7G305500* and *Seita.8G008700*) are closely related to their homologues in *Arabidopsis* (*AT3G54220*) and in maize (*GRMZM2G131516*). GO analysis of *SiSCRs* indicate that they both are involved in several biological processes, such as bundle sheath cell fate

specification, leaf development, radial pattern formation, gravitropism, asymmetric cell division, and so on (<http://plantfdb.cbi.pku.edu.cn/>). Thus, the genetic information on *Arabidopsis* *SCRs* and maize *ZmSCR* would provide us some clues for the characterization of the millet *SiSCRs*. Additionally, by comparing the CDS sequences and peptide sequences of the two *SiSCRs*, it was found that they both showed 98% identities in nucleotide coding sequences and in amino acid sequences (Suppl. 9 and 10). However, such 2% variation in sequences result in the differences on the number of motifs between these two genes as *Seita.8G008700* contains one additional motif 9 located most closely towards N-terminal region (Fig. 3). Therefore, it would be intriguing to investigate the functional differences between *Seita.7G305500* and *Seita.8G008700* in future study.

Another finding is that even though some millet *SCL* genes are closely related during evolution, they exhibit significantly different motif patterns and differential expression patterns in four millet organs and along leaf

gradient. For example, although *Si024520m.g* and *Si026116m.g* are most closely related *SCL* genes through phylogenetic analyses, both belonging to the same subgroup of group D (Fig. 4), *Si024520m.g* has moderate expression levels in spica, stem and root, whereas *Si026116m.g* only exhibits weak expression in root and nearly non expression in spica and stem (Fig. 5). This could be the reason that *Si024520m.g* contains the *cis*-acting regulatory element involved in seed-specific regulation, RY-element and gibberellin-responsive element, P-box. Furthermore, *Si024520m.g* has moderate expression levels in LS and LB along leaf gradient, however, hardly any transcript of *Si026116m.g* can be detected in the four leaf regions (Fig. 6). Not surprisingly, *Si026116m.g* contains five additional motifs compared with *Si024520m.g* (Fig. 3). The phenomenon that closely related genes in a gene family show differential expression patterns is not an isolated incident happening only in *SCL* genes as some paralogous genes of *Dof* family members with highly identical amino acid sequences in bamboo were also observed to exhibit totally opposite expression patterns (Wang et al. 2016). The results therefore revealed the functional diversity among *SCL* genes in millet.

Overall, expressions of large parts of millet *SCL* genes were not detected in the four tissues and along leaf gradient, indicating that these genes may be functional in specific organs during specific developmental stages rather than play house keeping functions during the millet whole life cycle. Other millet *SCL* genes showed different expression patterns from each other, implying that these *SCL* genes may play different roles during millet development and growth.

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 Postgraduate Research Innovation Fund Project of Hebei Province (S201622) for funding.

Author contributions HL performed the experiments, analyzed the data and wrote the manuscript; JQ, HF, JC, LL participated in analyzing data; ZL conceived the study, analyzed the data and revised the manuscript.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Balazadeh S, Riano-Pachon DM, Mueller-Roeber B (2008) Transcription factors regulating leaf senescence in *Arabidopsis thaliana*. Plant Biol (Stuttg) 10(Suppl 1):63–75

- Bolle C (2004) The role of GRAS proteins in plant signal transduction and development. Planta 18:683–692
- Broun P (2004) Transcription factors as tools for metabolic engineering in plants. Curr Opin Plant Biol 7:202–209
- Brown NJ, Newell CA, Stanley S, Chen JE, Perrin AJ, Kajala K, Hibberd JM (2011) Independent and parallel recruitment of preexisting mechanisms underlying C₄ photosynthesis. Science 331:1436–1439
- Chen WJ, Zhu T (2004) Networks of transcription factors with roles in environmental stress response. Trends Plant Sci 9:591–596
- Chen W, Provart NJ, Glazebrook J, Katagiri F, Chang HS, Eulgem T, Mauch F, Luan S, Zou G, Whitham SA, Budworth PR, Tao Y, Xie Z, Chen X, Lam S, Kreps JA, Harper JF, Si-Ammour A, Mauch-Mani B, Heinlein M, Kobayashi K, Hohn T, Dangl JL, Wang X, Zhu T (2002) Expression profile matrix of Arabidopsis transcription factor genes suggests their putative functions in response to environmental stresses. Plant Cell 14:559–574
- Chen K, Li H, Chen Y, Zheng Q, Li B, Li Z (2015) TaSCL14, a novel wheat (*Triticum aestivum* L.) gras gene, regulates plant growth, photosynthesis, tolerance to photo oxidative stress, and senescence. J Genet Genom 2:21–32
- de Moor CH, Meijer H, Lissenden S (2005) Mechanisms of translational control by the 3' UTR in development and differentiation. Semin Cell Dev Biol 16:49–58
- Deng W, Wang Y, Liu Z, Cheng H, Xue Y (2014) HemI: a toolkit for illustrating Heatmaps. PLoS ONE 9:e111988
- Feng X, Zhao Z, Wang X, Qiu F, Song G, Wang D, Su X, Zhang X, Wang F (2014) Recent research progress in Foxtail Millet (*Setaria italica*). Agric Sci Technol 15(564–570):575
- Fode B, Siemsen T, Thurow C, Weigel R, Gatz C (2008) The Arabidopsis GRAS protein SCL14 interacts with class II TGA transcription factors and is essential for the activation of stress-inducible promoters. Plant Cell 20:3122–3135
- Gao MJ, Parkin I, Lydiate D, Hannoufa A (2004) An auxin-responsive SCARECROW-like transcriptional activator interacts with histone deacetylase. Plant Mol Biol 55:417–431
- Gao X, Wang C, Cui H (2014) Identification of bundle sheath cell fate factors provides new tools for C₃-to-C₄ engineering. Plant Signal Behav. doi:10.4161/psb.29162
- Gao MJ, Li X, Huang J, Gropp GM, Gjetvaj B, Lindsay DL, Wei S, Coutu C, Chen ZX, Wan XC, Hannoufa A, Lydiate DJ, Gruber MY, Chen ZJ, Hegedus DD (2015) Scarecrow-like15 interacts with histone deacetylase19 and is essential for repressing the seed maturation programme. Nat Commun 6:7243
- Guo HJ, Jiao YN, Di C, Yao DX, Zhang GH, Zheng X, Liu L, Zhang QL, Guo AG, Su Z (2009) Discovery of Arabidopsis GRAS family genes in response to osmotic and drought stresses. Chin Bull Bot 44:290–299
- Heidstra R, Welch D, Scheres B (2004) Mosaic analysis using marked activation and deletion clones dissect Arabidopsis SCARECROW action in asymmetric cell division. Gene Dev 18:1964–1969
- Jia G, Huang X, Zhi H, Zhao Y, Zhao Q, Li W, Yang Chai Y, Yang L, Liu K, Lu H et al (2013) A haplotype map of genomic variations and genome-wide association studies of agronomic traits in foxtail millet (*Setaria italica*). Nat Genet 45:957–961
- Kamiya N, Itoh J, Morikami A, Nagato Y, Matsuoaka M (2003) The SCARECROW gene's role in asymmetric cell divisions in rice plants. Plant J 36:45–54
- Laajanen K, Vuorine I, Salo V, Juuti J, Raudaskoski M (2007) Cloning of *Pinus sylvestris* SCARECROW gene and its expression pattern in the pine root system, mycorrhiza and NPA-treated short roots. New Phytol 175:230–243
- Lee MH, Kim B, Song SK, Heo JO, Yu NI, Lee SA, Kim M, Kim DG, Sohn SO, Lim CE, Chang KS, Lee MM, Lim J (2008) Large-

- scale analysis of the GRAS gene family in *Arabidopsis thaliana*. *Plant Mol Biol* 67:659–690
- Lescot M, Déhais P, Moreau Y, De Moor B, Rouzé P, Rombauts S (2002) PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res* 30:325–327
- Letunic I (2015) Smart: recent updates, new developments and status in 2015. *Nucleic Acids Res* 43(Database issue):257–260
- Lim J, Helariutta Y, Specht CD, Jung J, Sims L, Bruce WB, Diehn S, Benfey PN (2000) Molecular analysis of the *SCARECROW* gene in maize reveals a common basis for radial patterning in diverse meristems. *Plant Cell* 12:1307–1318
- Lim J, Jung JW, Lim CE, Lee MH, Kim BJ, Kim M, Bruce WB, Benfey PN (2005) Conservation and diversification of *SCARECROW* in maize. *Plant Mol Biol* 59:619–630
- Liu S, Liu XH, Jiang L (2011) Genome-wide identification, phylogeny and expression analysis of the lipoxygenase gene family in cucumber. *Genet Mol Res* 10:2613–2636
- Liu HY, Cheng JJ, Cheng SY, Fan H, Wen B, Liu Z (2016) Genome wide identification and expression analysis along the leaf developmental gradient of sigma factor gene family in foxtail millet (*Setaria italica*). *J Appl Biol Biotechnol* 4:011–030
- Llave C, Carrington JC (2002) Cleavage of Scarecrow-like mRNA targets directed by a class of *Arabidopsis* miRNA. *Science* 297:2053–2056
- Lu X, Chen D, Shu D, Zhang Z, Wang W, Klukas C, Chen LL, Fan Y, Chen M, Zhang C (2013) The differential transcription network between embryo and endosperm in the early developing maize seed. *Plant Physiol* 162:440–455
- Ma HS, Liang D, Shuai P, Xia XL, Yin WL (2010) The salt- and drought-inducible poplar GRAS protein SCL7 confers salt and drought tolerance in *Arabidopsis thaliana*. *J Exp Bot* 61:4011–4019
- Morohashi K, Minami M, Takase H, Hotta Y, Hiratsuka K (2003) Isolation and characterization of a novel GRAS gene that regulates meiosis-associated gene expression. *J Biol Chem* 278:20865–20873
- Perezrodriguez P, Rianopachon DM, Correa LG, Rensing SA, Kersten B, Muellerroeber B (2010) PlnTFDB: updated content and new features of the plant transcription factor database. *Nucleic Acids Res*. doi:10.1093/nar/gkp805
- Pysh LD, Wysocka-Diller JW, Camilleri C, Bouchez D, Benfey PN (1999) The GRAS gene family in *Arabidopsis*: sequence characterization and basic expression analysis of the *SCARECROW-LIKE* genes. *Plant J* 18:111–119
- Qi Y, Zhang L, Xin Y, Zhang ZH, Xu WY, Zhen S (2015) SIFGD: *Setaria italica* functional genomics database. *Mol Plant*. doi:10.1016/j.molp.2015.02.001
- Sbabou L, Bucciarelli B, Miller S, Liu J, Berhada F, Filali-Maltouf A, Allan D, Vance C (2010) Molecular analysis of scarecrow genes expressed in white lupin cluster roots. *J Exp Bot* 61:1351–1363
- Singh K, Foley RC, Onate-Sanchez L (2002) Transcription factors in plant defense and stress responses. *Curr Opin Plant Biol* 5:430–436
- Slewinski TL, Anderson AA, Zhang C, Turgeon R (2012) Scarecrow plays a role in establishing Kranz anatomy in maize leaves. *Plant Cell Physiol* 53:2030–2037
- Solé A, Sánchez C, Vielba JM, Valladares S, Abarca D, Díaz-Sala C (2008) Characterization and expression of a *Pinus radiata* putative ortholog to the *Arabidopsis* SHORT-ROOT gene. *Tree Physiol* 28:1629–1639
- Stelpflug SC, Sekhon RS, Vaillancourt B, Hirsch CN, Buell CR, de Leon N, Kaeppler SM (2015) An expanded maize gene expression atlas based on RNA-sequencing and its use to explore root development. *Plant Genome*. doi:10.3835/plantgenome2015.04.0025
- Tian C, Ping W, Sun S, Li J, Chen M (2004) Genome-wide analysis of the GRAS gene family in rice and *Arabidopsis*. *PLoS ONE* 54:519–532
- Torres-Galea P, Huang LF, Chua NH, Bolle C (2006) The GRAS protein SCL13 is a positive regulator of phytochrome-dependent red light signaling, but can also modulate phytochrome A responses. *Mol Genet Genomics* 276:13–30
- Torres-Galea P, Hirtreiter B, Bolle C (2013) Two GRAS proteins, *SCARECROW-LIKE21* and *PHYTOCHROME A SIGNAL TRANSDUCTION 1*, function cooperatively in phytochrome A signal transduction. *Plant Physiol* 161:291–304
- Wang L, Czedit-Eysenberg A, Mertz RA, Si Y, Tohge T, Nunes-Nesi A, Arrivault S, Dedow LK, Bryant DW, Zhou W, Xu J, Weissmann S, Studer A, Li P, Zhang C, LaRue T, Shao Y, Ding Z, Sun Q, Patel RV, Turgeon R, Zhu X, Provart NJ, Mockler TC, Fernie AR, Stitt M, Liu P, Brutnell TP (2014) Comparative analyses of C4 and C3 photosynthesis in developing leaves of maize and rice. *Nat Biotechnol* 32:1158–1165
- Wang T, Yue JJ, Wang XJ, Xu L, Li LB, Gu XP (2016) Genome-wide identification and characterization of the dof, gene family in moso bamboo (*Phyllostachys heterocycla*, var. pubescens). *Genes Genom* 38:733–745
- Wysocka-Diller JW, Helariutta Y, Fukaki H, Malamy JE, Benfey PN (2000) Molecular analysis of *SCARECROW* functions reveals a radial patterning mechanism common to root and shoot. *Development* 127:595–603
- Zhang Z, Ogawa M, Fleet CM, Zentella R, Hu J, Heo JO, Lim J, Kamiya Y, Yamaguchi S, Sun TP (2011) *SCARECROW-LIKE 3* promotes gibberellins signaling by antagonizing master growth repressor *DELLA* in *Arabidopsis*. *Proc Natl Acad Sci USA* 108:2160–2165
- Zhou LJ, Zhang FC, Wang Y (2013) Research progress on the functional mechanism of GRAS family genes in plant growth, metabolism and stress. *Plant Physiol J* 49:855–860
- Zou Z, Eibl C, Koop HU (2003) The stem-loop region of the tobacco psbA 5'UTR is an important determinant of mRNA stability and translation efficiency. *Mol Genet Genomics* 269:340–349