1 Isolation and characterization of “GmScream” promoters that regulate
2 highly expressing soybean (*Glycine max* Merr.) genes

3
4 Ning Zhang^a, Leah K. McHale^b, John J. Finer^{a*}

5
6 ^aDepartment of Horticulture and Crop Science, The Ohio State University, 1680 Madison Ave.,
7 Wooster, OH 44691, USA

8
9 ^bDepartment of Horticulture and Crop Science, The Ohio State University, 2021 Coffey Rd,
10 Columbus, OH 43210, USA

11
12 *Corresponding author. Tel.: +1 330 263 3880; fax: +1 330 263 3887.

13 *E-mail* address: finer.1@osu.edu (J.J. Finer).

Abstract

To increase our understanding of the regulatory components that control gene expression, it is important to identify, isolate and characterize new promoters. In this study, a group of highly expressed soybean (*Glycine max* Merr.) genes, which we have named “GmScream”, were first identified from RNA-seq data. The promoter regions were then identified, cloned and fused with the coding region of the *green fluorescent protein (gfp)* gene, for introduction and analysis in different tissues using three tools for validation. Approximately half of the GmScream promoters identified showed levels of GFP expression comparable to or higher than the Cauliflower Mosaic Virus 35S (35S) promoter. Using transient expression in lima bean cotyledonary tissues, the strongest GmScream promoters gave over 6-fold higher expression than the 35S promoter while several other GmScream promoters showed 2- to 3-fold higher expression. The two highest expressing promoters, GmScreamM4 and GmScreamM8, regulated two different elongation factor 1A genes in soybean. In stably transformed soybean tissues, GFP driven by the GmScreamM4 or GmScreamM8 promoter exhibited constitutive high expression in most tissues with preferentially higher expression in proliferative embryogenic tissues, procambium, vascular tissues, root tips and young embryos. Using deletion analysis of the promoter, two proximal regions of the GmScreamM8 promoter were identified as contributing significantly to high levels of gene expression.

Keywords: Soybean, Strong promoters; GmScreamM8; RNA-seq

1. Introduction

The promoter is the most important component regulating the specificity and intensity of expression of both native genes and transgenes in plants [1]. With the burgeoning releases of plant genomic sequences [2-6] and the wealth of bioinformatics tools [7-10], much information is now available for selection and identification of novel plant promoters and their regulatory components. For production of transgenics for basic research, the Cauliflower Mosaic Virus 35S (35S) promoter is still the most widely used promoter [11]. Although this promoter has been very useful to the plant biotechnology community and gives relatively high levels of constitutive expression in many plants, it originates from a plant virus and may not be as useful as native plant promoters for uncovering gene expression regulation mechanisms and regulating transgenes. Use of virus and non-native heterologous promoters can result in unexpected transgene expression in some plant tissues while they also may not accurately reflect general promoter specificity and intensity in all plant genomes [12]. In comparison, promoter isolation, characterization and use in their native genomes could facilitate the discovery of potential elements involved in gene expression regulation and increase our understanding of gene regulation in plants [13].

Promoters are generally categorized as constitutive, spatiotemporal, or inducible, according to the type of gene expression they regulate [1]. Promoters can also be classified as weak or strong based on gene expression intensity. Strong promoters, in most cases, contain multiple enhancers or regulatory elements that can activate gene expression by recruiting necessary transcription factors to facilitate the formation of transcription initiation complex or transcription transition from initiation to elongation [14]. Although only a few enhancers or regulatory elements have been isolated and characterized from strong promoters, the G-box

element (CACGTG) [15] has been verified in many studies as contributing to high levels of gene expression [16-19]. In addition, many strong promoters also contain an intron in the 5' untranslated region (5'UTR) of the native gene [20, 21]. The intron may contain enhancers or regulatory elements that function like enhancers in promoter regions [22-24]. Alternatively, introns can increase gene expression via intron-mediated enhancement (IME), facilitating enhanced gene expression by an interaction between gene transcription machinery and intron splicing during RNA processing [21, 25-26].

Strong constitutive promoters can direct gene expression in most tissues, where they function independently of environmental and developmental factors [1]. As previously stated, the most widely used strong constitutive promoter is the 35S promoter, which was isolated from a plant pathogenic virus [11]. Some additional constitutive promoters have been isolated and characterized from plant genomes. A maize ubiquitin (ubi) promoter [27] and a rice actin 1 (Act-1) promoter [28] have been well characterized and used in the production of transgenic plants. Perhaps the best-studied highly expressed promoters in plants regulate polyubiquitin genes [29-33]. In soybean, ten different *Glycine max* polyubiquitin (Gmubi) promoters were isolated and characterized [34, 35] and 7 of the 10 promoters were predicted to contain introns in the 5'UTR. The importance of the leader intron was subsequently demonstrated in the strongest of the analyzed Gmubi promoter [22]. Characterization of additional novel strong constitutive plant promoters would be useful as this could not only increase our knowledge of gene expression regulation but also lead to higher and more stable transgene expression in transgenic plants.

To date, only a few soybean promoters have been isolated and characterized. Several well-characterized tissue-specific and inducible soybean promoters have been reported, such as the *rbcS* promoter [36], the seed-specific 7αP promoter [37], the auxin-inducible GH3 [38] and

SAUR promoters [39], the jasmonate-inducible VspB promoter [40], the stress-inducible GmCaM-4 promoter [41], the heat-inducible Gmhsp 17.5-E promoter [42], ten GmERF (*G. max* ethylene response factor) promoters [34], and somatic embryo-inducible actin, heat shock protein and ribosomal protein promoters [43]. In terms of constitutive promoters, although ten soybean Gmubi promoters were isolated and characterized [34, 35], only one of these Gmubi promoters was extensively characterized [22]. Considering the economic importance of soybean in the world, it would be useful to obtain and evaluate additional strong, constitutive soybean promoters.

In this research, a set of strong constitutive promoters were isolated from highly expressed soybean genes, which have been termed “*GmScream*” (*G. max Scream*) genes, regulated by their corresponding GmScream promoters. Among the 20 GmScream promoters cloned and studied here, 11 promoters showed comparable or higher activity compared to the 35S promoter. The strongest promoter GmScreamM8, which is a native promoter of soybean *elongation factor 1A (eEF1A)* gene, gave rise to over 6-fold higher expression than the 35S promoter with transient expression analysis. In addition, we identified two critical proximal regions of the GmScreamM8 promoter, which might contain enhancers or regulatory elements contributing to high activity of the promoter.

2. Methods

2.1. Identification of highly expressed soybean genes using RNA-seq

Two sets of RNA-seq data [44, 45] were used to identify highly expressed soybean genes. Ten candidate genes (Group I; GmScream1-10) with high expression in at least one tissue were selected from RNA-Seq Atlas of soybean, based on the heatmap of the top 500 most highly

expressed genes from soybean [46]. The expression profile of each gene in 14 different soybean tissues was created using SigmaPlot ver. 8.0 (Systat Software, San Jose, CA). In Group II, another ten candidate genes (Group II; GmScreamM1-M10, “M” represents “Multiple tissues”) with strong constitutive expression were selected by analyzing RNA-Seq expression data for soybean cultivar Williams 82 available from NCBI Short Read Archive (SRA) (<http://www.ncbi.nlm.nih.gov/sra>) [45]. All available RNA-seq data (31 libraries as of February, 2012) were downloaded from the SRA. Libraries were subsequently categorized into 22 sets based on tissue types or treatments. Reads were aligned to the soybean reference genome (Glyma.Wm82.a1) [4] with TopHat [47]. The expression levels of annotated genes was estimated according to normalized Fragments Per Kilobase of transcript per Million mapped fragments (FPKM) to each gene, calculated with Cufflinks [48]. A list of the top 1% of expressed annotated genes were generated from each of the 22 groups. Genes that were highly expressed in at least 18 of the 22 total groups were selected as Group II candidate genes.

2.2. DNA constructs

Upstream sequences of the GmScream genes were obtained from Phytozome [49] and all 5' UTR regions, including introns in 5'UTRs, were included in the sequences that were cloned via PCR. About 1.2 to 1.5 kb DNA fragments upstream of the predicted translation start site were PCR-amplified from soybean (*G. max* cv. ‘Jack’) genomic DNA with specific primers (Supplemental Table 1) containing a restriction site (*HindIII/PstI/NcoI/XhoI*) at the 5' end of each primer preceded by four or six additional nucleotides. All PCR amplifications were conducted using a FailsafeTM PCR Kit (Epicenter Biotechnologies, Madison, WI, USA) according to the manufacturer’s instructions. For introduction of the promoters into the promoter

cloning and analysis vector, promoter fragments were digested with appropriate restriction enzymes and fused to the *green fluorescent protein (gfp)* gene in pFLEV [34]. For one of the strongest promoters (GmScreamM8), a series of 5' extension and deletion constructs, GmScreamM8L (L=large, 2.65 kb), GmScreamM8M (M=medium, 1.2 kb), GmScreamM8S (S=small, 1 kb), and GmScreamM8C (C=core, 0.9 kb), were generated using specific forward primers paired with the same reverse primer GmScreamM8_R (Supplemental Table 1). Digested PCR fragments were purified, ligated to the *gfp* gene in pFLEV, and transformed into *Escherichia coli* DH5α by heat shock for sequence analysis. The sequence of all cloned promoters were confirmed via sequencing.

For subcloning of the promoter expression cassette into an *Agrobacterium* binary vector for hairy root production, the whole expression cassette containing the promoter, GFP coding region, and the NOS terminator was excised from pFLEV using appropriate restriction enzymes and subcloned into the multiple cloning site (MCS) of the binary expression vector pCAMBIA1300 (CAMBIA, Canberra, Australia).

2.3. Gene expression quantification and data analysis

For transient expression in lima bean cotyledonary tissues, lima bean (*Phaseolus lunatus* cv “Henderson Bush”) seeds were sterilized and germinated in moistened paper towels. After 5 d, cotyledons from germinating seeds were excised for DNA introduction via particle bombardment as previously described [34, 50]. Briefly, plasmid DNAs were precipitated onto tungsten particles and introduced into cotyledonary tissues using a particle inflow gun [51]. A robotic tracking system was used to image gene expression every hour for 100 h [50, 52]. The 100 sequential images for each promoter construct were first aligned using Adobe ImageReady.

The aligned images were resized to 800 x 600 pixels and a 400 x 300 pixel area was selected from the larger area for GFP quantification. The level of GFP expression was calculated by multiplying the mean grayscale value per pixel and the total number of GFP-expressing pixels in both the green and red channel of each image [50]. For each promoter-containing construct, three cotyledons were separately bombarded per experiment, with at least two independent biological replications.

For production of soybean hairy roots, promoter-containing pCAMBIA1300 vectors were first introduced into *Agrobacterium rhizogenes* strain K599, which was then used to induce stably transformed soybean hairy roots as previously described [34]. Promoter intensity was analyzed by quantification of GFP expression levels in soybean hairy roots using ImageJ [34]. For each promoter construct, at least 20 independent transgenic events were generated and analyzed, with at least two independent biological replications. Statistical analysis was performed using SAS 9.2 (SAS Institute Inc., Cary, NC, USA). Mean difference ($p < 0.05$) was tested using Tukey's Studentized Range (HSD).

The two strongest promoters, GmScreamM4 and GmScreamM8, were introduced into soybean (cvs "Jack" and "JackX") proliferative embryogenic tissue via particle bombardment as described previously [53]. Based on the consistency and strength of expression, event 2 (E2) of GmScreamM4 and event 6 (E6) of GmScreamM8 were selected as representative expressors, for more detailed analysis of GFP expression analysis in various tissues. GFP expression was characterized in stably transformed embryogenic tissues as well as in different vegetative and reproductive tissues/organs in T₁ transgenic soybean plants. All images were collected using a MZFLIII stereomicroscope (Leica, Heerbrugg, Switzerland) equipped with a Nikon Coolpix 990 digital camera and a GFP2 filter set (Excitation 480 ± 40 nm; Emission 510 nm).

3. Results

3.1. Identifying highly expressed soybean genes using RNA-seq

Twenty highly expressed genes were identified from soybean as GmScream candidates (Table 1). Group I candidate genes (GmScream1-10) were highly expressed in one or more soybean tissues (Fig. 1), while Group II candidate genes (GmScreamM1-M10) were highly expressed in more than 18 tissues or treatments (Table 2). Highly expressed candidate genes in both groups were derived from various gene families (Table 1), however, GmScreamM1, GmScreamM4, and GmScreamM8 each encoded an elongation factor 1A (eEF1A) protein and the promoter from each of these *eEF1A* genes contained a large intron in the 5'UTR.

3.2. Transient expression of GmScream promoters in lima bean cotyledonary tissues

Out of the 20 promoters, which were cloned from highly expressed soybean genes, six (GmScream5, GmScream7, GmScreamM1, GmScreamM4, GmScreamM8, GmScreamM9) showed higher activity than the 35S promoter in transient expression assays in lima bean cotyledonary tissues, while another five promoters displayed an intensity of gene expression comparable to or weaker than the 35S promoter (Fig. 2). Among the remaining nine GmScream promoters, which were not shown in the transient expression profile (Fig. 2), seven had very low activity and two (GmScream4 and GmScream10) did not show any activity (Table 3). The two strongest GmScream promoters were the GmScreamM8 and GmScreamM4 promoters, which displayed over 6-fold higher expression and 3-fold higher expression than the 35S promoter respectively at the peak transient expression time point (Fig. 2).

The GmScream1, GmScream5, GmScreamM8 and GmScreamM9 promoters displayed peak expression at about 24 h after DNA introduction, which was around 5-6 h later than that of the 35S promoter, while the peak GFP expression driven by the GmScreamM4 promoter occurred 14 h later than the peak expression of the 35S promoter (Fig. 2B). In addition to a delay in peak transient expression, the GmScreamM4 promoter also showed an extended expression profile or a delay of GFP expression decline, compared to all of the other promoters (Fig. 2B).

3.3. Stable expression of GmScream promoters in soybean hairy roots

For the GmScream promoters, the four strongest promoters assayed in stably transformed soybean hairy roots were GmScream5, GmScreamM1, GmScreamM4, and GmScreamM8, which gave rise to 2- to 4-fold higher GFP expression compared to the 35S promoter (Fig. 3). The highest expressing promoter in soybean hairy roots was the GmScreamM8 promoter, which was higher, but not significantly different from the high expression observed with the GmScreamM4 promoter (Fig. 3). The seven GmScream promoters displayed levels of activity that were not significantly different from the 35S promoter (Fig. 3). In addition, among the remaining nine GmScream promoters, which were not shown in the stable expression profile (Fig. 3), two (GmScream6 and GmScream7) had lower activity but not significantly different from the 35S promoter, five showed significant lower activity, and another two (GmScream4 and GmScream10) did not show any functionality (Table 3).

3.4. Extension and deletion analysis of the GmScreamM8 promoter

Evaluation of a 5' extension construct, GmScreamM8L (Fig. 4A), showed that the intensity of GFP expression driven by the GmScreamM8L promoter was not significantly

different from the standard 1.5 kb GmScreamM8 promoter using both transient expression and stable expression in soybean hairy roots (Figs. 4B & 4C). However, the transient expression profile of GmScreamM8L was different than other promoter-containing constructs, displaying a slower decline of GFP expression after the peak expression time point (Fig. 4B).

Three additional 5' deletion constructs, GmScreamM8M, GmScreamM8S, and GmScreamM8C (Fig. 4A), gave lower activity than the GmScreamM8 promoter using our transient expression and stable expression assays (Figs. 4B & 4C). The GmScreamM8M construct, which had a 362 bp deletion of GmScreamM8 in its 5' region, showed a reduction in GFP expression of almost 50% compared to the GmScreamM8 promoter. GmScreamM8S, a further truncation of GmScreamM8M, yielded a slightly higher level of GFP expression than GmScreamM8M. The smallest deletion construct GmScreamM8C, consisting of a core proximal promoter sequence with only 55 bp upstream of the 5'UTR, showed very low activity (Fig. 4B). For stable expression in soybean hairy roots, the GmScreamM8M promoter showed a significant decrease in activity in comparison with the larger GmScreamM8 promoter while GFP expression driven by the GmScreamM8S promoter showed 7.6-fold higher expression than the GmScreamM8C promoter fragment (Fig. 4C).

3.5. GFP expression driven by the GmScreamM4 and GmScreamM8 promoters in transgenic soybeans

In stably transformed soybean embryogenic tissues, high GFP expression was detected in five of seven GmScreamM4 events and thirteen out of fifteen GmScreamM8 events. Transgenic events driven by either the GmScreamM4 or GmScreamM8 promoter followed a similar expression pattern in embryogenic tissues (Fig. 5). Higher GFP expression was observed in

actively proliferating embryogenic tissues and somatic embryos in early developmental stages (Fig. 5). GFP expression diminished slowly during embryo development, with much lower expression in mature somatic embryos (Fig. 5).

Further analysis of GFP expression in T₁ transgenic soybean plants revealed that both GmScreamM4 and GmScreamM8 promoters regulated gene expression in most soybean tissues, but at varying intensities (Fig. 6). Transgenic seedlings containing the GmScreamM4 promoter regulating the *gfp* gene, exhibited high expression in cotyledons, plumules, hypocotyls, and roots, albeit at a relatively lower expression level compared to those containing the GmScreamM8 promoter (Fig. 6A). Three-day-old progeny seedlings containing the *gfp* gene regulated by the GmScreamM8 promoter showed high expression in cotyledons and plumules (Fig. 6A). High GFP expression was also observed in the vascular tissues of hypocotyls and roots in young plants (Fig. 6A). Young leaves containing either the GmScreamM4 or GmScreamM8 promoter displayed moderate GFP expression in the median and lateral veins on both abaxial and adaxial sides of the leaf (Fig. 6A).

Flowers from transgenic plants containing the GmScreamM8 promoter exhibited higher GFP expression than those containing the GmScreamM4 promoter. High GFP expression was also observed in stamens and pistils of plants containing either the GmScreamM4 or GmScreamM8 promoter although anthers of the control plants also seemed to fluoresce. Immature seeds containing either the GmScreamM4 or GmScreamM8 promoter showed high expression levels while seeds with the GmScreamM4 promoter showed much higher expression (Fig. 6B).

4. Discussion

4.1. *GmScream* promoters

Of the twenty promoters that were cloned for this research, the 3 strongest promoters (GmScreamM1, GmScreamM4, and GmScreamM8) all regulated soybean *eEF1A* genes encoding translational elongation factor proteins. High levels of GFP expression with the three soybean eEF1A promoters in transiently-transformed lima bean cotyledons and stably transformed hairy root tissue (Figs. 2 & 3) were not surprising, considering the critical role of eEF1As in protein synthesis in plants [54, 55]. Since high *gfp* gene expression was observed in most soybean tissues containing the GmScreamM4 and GmScreamM8 promoters, both promoters were considered strong and constitutive, consistent with the transcriptome data from NCBI SRA from soybean (Tables 1 & 2). Eukaryotic EF1As are typically encoded by multigene families in *Arabidopsis* [56, 57], maize [58, 59], rice [60], cotton [61], cassava [62] and soybean [63]. In contrast to the consistent high levels of expression of the soybean *eEF1A* genes we selected in this study, some plant *eEF1A* genes are differentially expressed in different tissues/organs [57, 59, 61, 62, 64]. For example, only five out of the nine *eEF1A* genes in cotton were highly expressed in young fibers [61] and in cassava, only one of the three eEF1A promoters was constitutively expressed using transient expression in tobacco seedlings [62]. In this study, we did not isolate and characterize all the soybean eEF1A promoters as only three were selected for constitutive, high expression according to the criteria applied to the RNA-seq analysis. Although there are a number of studies on plant *eEF1A* gene expression analysis, isolation and characterization of plant eEF1A promoters has only been reported in *Arabidopsis* [65, 66] and cassava [62]. Parts of the *Arabidopsis thaliana* eEF1A promoter, along with its native 5'UTR intron [56] have been used in the production of commercial Roundup-resistant crops [Roundup Ready 2 Yield, 67] although the reason for its use was not specified.

1
2
3
4 289 Use of RNA-seq to identify highly expressed genes in soybean was expected to yield all
5
6
7 290 strong promoters, but almost half of the cloned promoters from those genes showed low or no
8
9 291 activity in both transient expression and stable expression analyses (Tables 1, 2 & 3). Due to the
10
11 292 methods used for promoter isolation and analysis, we may have unintentionally eliminated
12
13
14 293 critical promoter elements or enhancers [68], or lost the DNA structural components [69], which
15
16 294 may influence expression from beyond the promoter regions isolated in this study. Although
17
18
19 295 most regulatory elements seem to be located in proximal promoter regions [1, 30, 70], additional
20
21 296 elements [71, 72] or factors [68] that affect promoter function [14] can be located well upstream
22
23
24 297 of the proximal promoter region, or downstream of the translational start site including the
25
26 298 coding region and the terminator [68]. Cloning and evaluation of promoters using heterologous
27
28
29 299 coding regions and terminators resulted in the removal of the promoter from its native context,
30
31 300 which could lead to changes in specificity and intensity of some promoter elements [13].
32
33 301 However, removal of a promoter from its native context is not likely to result in a complete loss
34
35
36 302 of basic promoter functionality. In cases of strong promoters, where elements can be found
37
38 303 greater than 1.5 kb upstream of the transcriptional start site (TSS), smaller promoter fragments
39
40
41 304 located nearby the TSS should show easily measureable activity [68]. Therefore, it seems
42
43 305 unlikely that the critical promoter elements needed to achieve basic levels of gene expression
44
45 306 were missed, due to the size (1.2 to 1.6 kb) of the cloned promoter fragments that were generated
46
47
48 307 in this study. As most of the Group I promoters, identified through analysis of the RNA-seq
49
50
51 308 atlas, were highly expressed in certain tissues or under specific development stages (Fig. 1) and
52
53 309 Group II promoters, identified through analysis of NCBI SRA, were strongly constitutive in most
54
55 310 tissues or treatments (Table 2), all promoters, if the TSS was appropriately annotated, should
56
57
58 311 display some base levels of GFP expression using our rapid validation tools. However, as we
59
60
61
62
63
64
65

1
2
3
4 312 relied on publicly available RNA-seq data, and did not perform an independent transcriptome
5
6 313 analysis, gene features, such as TSSs and UTRs, may have been incorrectly annotated, especially
7
8
9 314 for genes containing introns [73]. A recent change in the annotation of the soybean genome
10
11 315 eliminated four of the twenty highly-expressing genes whose promoters were being tested (Table
12
13
14 316 1). Two of these promoters (GmScream4 and GmScream10) showed no activity in our assays
15
16 317 further confirming that some of the GmScream genes might not have been correctly annotated.
17
18
19 318 As sufficient numbers of highly expressing promoters were recovered, efforts were not
20
21 319 undertaken to further pursue these underperforming or non-functional promoters.
22
23

24 320 25 26 321 *4.2. Characterization of GmScream promoters* 27

28
29 322 In terms of the intensity of GmScream promoters, overall consistent results were obtained
30
31 323 between transient expression in lima bean cotyledons (Fig. 2) and stable expression in soybean
32
33 324 hairy roots (Fig. 3), as reported previously for evaluation of other soybean promoters and
34
35
36 325 promoter variants [1, 22, 34, 50]. For transient expression analysis, all GmScream promoters and
37
38 326 GmScreamM8 promoter variants displayed similar expression profiles, with the exception of
39
40
41 327 GmScreamM4 (Fig. 2B) and GmScreamM8L (Fig. 4B), which showed an extended expression
42
43 328 and a delay of GFP expression decline. The specific *gfp* gene that was used in this research is an
44
45
46 329 early GFP variant with good fluorescence but some instability (sGFPS65T) [34, 74], which
47
48 330 resulted in declines in GFP expression, once the promoter became less active. Therefore, the
49
50
51 331 intensity and localization of fluorescence was likely reflective of promoter activity. For the
52
53 332 extended expression of the GmScreamM4 promoter observed during transient expression (Fig.
54
55 333 2B), this might occur if the promoter showed wound inducibility or tissue specificity in
56
57
58 334 cotyledonary tissues. Wound inducibility of the GmScreamM4 promoter was not observed in
59
60
61
62
63
64
65

GmScreamM4-containing transgenic plants during excision and manipulation of the tissue for analysis (data not shown); however, some of the highest levels of expression were observed in the seed (Fig. 6B), suggesting that high expression in seed tissue could contribute to the extended expression in bombarded lima bean cotyledonary tissues. For the extended expression observed with transient expression of the GmScreamM8L promoter, this profile was not observed with the standard GmScreamM8 promoter (Fig. 2B), indicating that the additional 1 kb of the 5' promoter sequence may contain either seed-specific elements or wound-inducible elements conferring the extended expression profile in bombarded cotyledons.

For the four highest expressing promoters in transgenic soybean hairy roots (GmScream5, GmScreamM1, GmScreamM4, GmScreamM8) (Fig. 3), three of the four promoters regulated soybean *eEF1A* genes, while the GmScream5 gene has not yet been functionally annotated (Table 1), although weak similarity of the GmScream5 gene to a *Medicago* eEF1A protein was found using protein blast. Although the GmScream5 promoter expressed at high levels when evaluated using transient expression (Fig. 2), the other three eEF1A promoters showed much higher transient expression levels. In addition, all these four promoters possessed leader introns (Table 1) with high IMETER scores (data not shown; [75]). In comparison, for those less active promoters (GmScreamM2, GmScreamM3, GmScreamM6, GmScreamM9, and GmScreamM10), no stimulating introns were included in the GFP fusions, although the genes from which they were derived contain introns with high IMETER scores in the coding region. Intron sequences, especially the first intron in the 5'UTR or in the coding region, could significantly increase gene expression when used with native or heterologous promoter sequences [22, 23, 27, 28, 76-78], although the mechanism of intron enhancement is still not fully understood [79].

4.3. *GmScreamM8* promoter variants

GFP expression under the regulation of the large *GmScreamM8L* promoter was slightly lower but not significantly different from the original *GmScreamM8* promoter, indicating that no enhancer or regulatory element contributing to high expression existed in the extended 1 kb sequence. Although most of the regulatory elements in promoters are found in the proximal promoter region [1, 30, 70], certainly some elements may be located well upstream of the TSS [14, 71]. A large (1.5 kb) 5' extension of a 1 kb soybean polyubiquitin promoter (*GmubiXL*) gave a 2-fold increase in expression levels over the original promoter [22]; however, the specific enhancer in this region was not identified. Considering the high activity of the original 1.5 kb *GmScreamM8* promoter, which was comparable to the *GmubiXL* promoter [22] in both transient expression and stably transformed soybean hairy roots, additional efforts were not put into cloning and identifying even more distal enhancers that could extend beyond the cloned 2.5 kb promoter fragment.

Of the two short proximal regulatory regions (-728 to -367; 160 to -60; relative to the TSS) identified in the *GmScreamM8* promoter (Fig. 4), enhancers or regulatory elements contributing to high activity of the promoter could exist within the regions. Sequences resembling G-box-like elements (TACGTG, GACGTG) were found in the first and second truncation fragment respectively. The canonical G-box (CACGTG) regulatory element [15] has been well studied and demonstrated to confer high levels of gene expression [16-19], suggesting that G-box-like sequences identified in the *GmScreamM8* promoter could modulate the high level of observed promoter activity. However, other regulatory elements may also contribute to the high expression observed with the *GmScreamM8* promoter. In the *Arabidopsis eEF1A*

promoter [65, 66], a TEF 1 box (AGGGGCANNNNGTAANT), located about 100 bp upstream of the TSS, was required for expression in rapidly growing meristematic tissues. A similar sequence (AGGGTAAATTAGTAAGG) was also found approximately 80 bp upstream of the TSS in the GmScreamM8 promoter, which may also direct high activity of the promoter.

In this study, promoters from highly expressing soybean genes were cloned and partially characterized. We believe that this is the first report of cloning promoters, based on high expression levels. Surprisingly, the three highest expressing promoters regulated *eEF1A* genes and all contained leader introns in the 5'UTR. The relative contribution of the intron, intron sequences, and proximal promoter elements to overall promoter activity is currently not known. GmScreamM4 and GmScreamM8 promoters were roughly equivalent to or a little stronger in intensity and specificity to GmubiXL but much higher than the 35S [22, 35]. Use of highly expressed soybean promoters could lead to higher, more consistent and stable transgene expression in transgenic soybeans, which could be very useful for both commercial soybean transgenics as well as basic research of gene expression regulation.

Acknowledgements

Salaries and research support were provided by the United Soybean Board, Bayer CropScience, and by State and Federal funds appropriated to The Ohio State University/Ohio Agricultural Research and Development Center. This work was also partly supported by the OARDC Director's Associateship Program Award in 2011-2012. This research was also supported, in

part, through The Consortium for Plant Biotechnology Research, Inc. by DOE Prime Agreement No. DEFG36-02G012026. This support does not constitute an endorsement by DOE or by The Consortium for Plant Biotechnology Research, Inc. of the views expressed in this publication. Mention of trademark or proprietary products does not constitute a guarantee or warranty of the product by OSU/OARDC and also does not imply approval to the exclusion of other products that may also be suitable.

References

- [1] C.M. Hernandez-Garcia, J.J. Finer, Identification and validation of promoters and *cis*-acting regulatory elements, *Plant Sci.* 217 (2014) 109-119.
- [2] A.H. Paterson, J.E. Bowers, R. Bruggmann, I. Dubchak, J. Grimwood, H. Gundlach, G. Haberer, U. Hellsten, T. Mitros, A. Poliakov, The *Sorghum bicolor* genome and the diversification of grasses, *Nature.* 457 (2009) 551-556.
- [3] P.S. Schnable, D. Ware, R.S. Fulton, J.C. Stein, F. Wei, S. Pasternak, C. Liang, J. Zhang, L. Fulton, T.A. Graves, The B73 maize genome: complexity, diversity, and dynamics, *Science.* 326 (2009) 1112-1115.
- [4] J. Schmutz, S.B. Cannon, J. Schlueter, J. Ma, T. Mitros, W. Nelson, D.L. Hyten, Q. Song, J.J. Thelen, J. Cheng, Genome sequence of the palaeopolyploid soybean, *Nature.* 463 (2010) 178-183.
- [5] X. Xu, S. Pan, S. Cheng, B. Zhang, D. Mu, P. Ni, G. Zhang, S. Yang, R. Li, J. Wang, Genome sequence and analysis of the tuber crop potato, *Nature.* 475 (2011) 189-195.
- [6] R. Brenchley, M. Spannagl, M. Pfeifer, G.L. Barker, R. D'Amore, A.M. Allen, N. McKenzie, M. Kramer, A. Kerhornou, D. Bolser, Analysis of the bread wheat genome using whole-genome shotgun sequencing, *Nature.* 491 (2012) 705-710.
- [7] M.G. Grabherr, B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Full-length transcriptome assembly from RNA-Seq data without a reference genome, *Nat. Biotechnol.* 29 (2011) 644-652.

- [8] H.C. Lee, K. Lai, M.T. Lorenc, M. Imelfort, C. Duran, D. Edwards, Bioinformatics tools and databases for analysis of next-generation sequence data, *Brief Funct. Genomics*. 11 (2012) 12-24.
- [9] K. Mutz, A. Heilkenbrinker, M. Lönne, J. Walter, F. Stahl, Transcriptome analysis using next-generation sequencing, *Curr. Opin. Biotech.* 24 (2013) 22-30.
- [10] S. Pabinger, A. Dander, M. Fischer, R. Snajder, M. Sperk, M. Efremova, B. Krabichler, M.R. Speicher, J. Zschocke, Z. Trajanoski, A survey of tools for variant analysis of next-generation genome sequencing data, *Brief Bioinform.* 15 (2014) 256-278.
- [11] J.T. Odell, F. Nagy, N.H. Chua, Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter, *Nature*. 313 (1985) 810-812.
- [12] K.M. Butaye, B.P. Cammue, S.L. Delaure, M.F. De Bolle, Approaches to minimize variation of transgene expression in plants, *Mol. Breed.* 16 (2005) 79-91.
- [13] P.J. Rushton, A. Reinstädler, V. Lipka, B. Lippok, I.E. Somssich, Synthetic plant promoters containing defined regulatory elements provide novel insights into pathogen-and wound-induced signaling, *The Plant Cell*. 14 (2002) 749-762.
- [14] E. Calo, J. Wysocka, Modification of enhancer chromatin: what, how, and why?, *Mol. Cell*. 49 (2013) 825-837.
- [15] G. Giuliano, E. Pichersky, V.S. Malik, M.P. Timko, P.A. Scolnik, A.R. Cashmore, An evolutionarily conserved protein binding sequence upstream of a plant light-regulated gene, *Proc. Natl. Acad. Sci.* 85 (1988) 7089-7093.
- [16] R.G. Donald, A.R. Cashmore, Mutation of either G box or I box sequences profoundly affects expression from the *Arabidopsis* rbcS-1A promoter, *EMBO J.* 9 (1990) 1717-1726.
- [17] F. Ishige, M. Takaichi, R. Foster, N. Chua, K. Oeda, AG-box motif (GCCACGTGCC) tetramer confers high-level constitutive expression in dicot and monocot plants, *Plant J.* 18 (1999) 443-448.
- [18] W.L. McKendree Jr, R.J. Ferl, Functional elements of the *Arabidopsis* Adh promoter include the G-box, *Plant Mol. Biol.* 19 (1992) 859-862.
- [19] K. Stålberg, M. Ellerstöm, I. Ezcurra, S. Ablov, L. Rask, Disruption of an overlapping E-box/ABRE motif abolished high transcription of the napA storage-protein promoter in transgenic *Brassica napus* seeds, *Planta*. 199 (1996) 515-519.

- [20] R.M. Sinibaldi, I.J. Mettler, Intron splicing and intron-mediated enhanced expression in monocots, *Prog. Nucleic Acid Res. Mol. Biol.* 42 (1992) 229-257.
- [21] A. Rose, Intron-mediated regulation of gene expression, in: A.S. Reddy, M. Golovkin (Eds.), *Nuclear pre-mRNA processing in plants*, Springer, 2008, pp.277-290.
- [22] C.M. De La Torre, J.J. Finer, The intron and 5'distal region of the soybean Gmubi promoter contribute to very high levels of gene expression in transiently and stably transformed tissues, *Plant Cell Rep.* 34 (2015) 111-120.
- [23] M. Gidekel, B. Jimenez, L. Herrera-Estrella, The first intron of the *Arabidopsis thaliana* gene coding for elongation factor 1 β contains an enhancer-like element, *Gene*. 170 (1996) 201-206.
- [24] K. Stark, D.L. Kirk, R. Schmitt, Two enhancers and one silencer located in the introns of regA control somatic cell differentiation in *Volvox carteri*, *Genes Dev.* 15 (2001) 1449-1460.
- [25] A. Furger, J.M. O'Sullivan, A. Binnie, B.A. Lee, N.J. Proudfoot, Promoter proximal splice sites enhance transcription, *Genes Dev.* 16 (2002) 2792-2799.
- [26] A.B. Rose, The effect of intron location on intron-mediated enhancement of gene expression in *Arabidopsis*, *Plant J.* 40 (2004) 744-751.
- [27] A.H. Christensen, R.A. Sharrock, P.H. Quail, Maize polyubiquitin genes: structure, thermal perturbation of expression and transcript splicing, and promoter activity following transfer to protoplasts by electroporation, *Plant Mol. Biol.* 18 (1992) 675-689.
- [28] W. Zhang, D. McElroy, R. Wu, Analysis of rice Act1 5' region activity in transgenic rice plants, *The Plant Cell*. 3 (1991) 1155-1165.
- [29] J.E. Garbarino, T. Oosumi, W.R. Belknap, Isolation of a polyubiquitin promoter and its expression in transgenic potato plants, *Plant Physiol.* 109 (1995) 1371-1378.
- [30] J. Wang, J. Oard, Rice ubiquitin promoters: deletion analysis and potential usefulness in plant transformation systems, *Plant Cell Rep.* 22 (2003) 129-134.
- [31] H. Wei, M. Wang, P.H. Moore, H.H. Albert, Comparative expression analysis of two sugarcane polyubiquitin promoters and flanking sequences in transgenic plants, *J. Plant Physiol.* 160 (2003) 1241-1251.
- [32] E. Sivamani, R. Qu, Expression enhancement of a rice polyubiquitin gene promoter, *Plant Mol. Biol.* 60 (2006) 225-239.

- [33] D.G. Mann, Z.R. King, W. Liu, B.L. Joyce, R.J. Percifield, J.S. Hawkins, P.R. LaFayette, B.J. Artelt, J.N. Burris, M. Mazarei, J.L. Bennetzen, W.A. Parrott, C.N. Stewart Jr, Switchgrass (*Panicum virgatum* L.) polyubiquitin gene (PvUbi1 and PvUbi2) promoters for use in plant transformation, BMC Biotechnol. 11 (2011) 74.
- [34] C. Hernandez-Garcia, R. Bouchard, P. Rushton, M. Jones, X. Chen, M. Timko, J. Finer, High level transgenic expression of soybean (*Glycine max*) GmERF and Gmubi gene promoters isolated by a novel promoter analysis pipeline, BMC Plant Biol. 10 (2010) 237.
- [35] C.M. Hernandez-Garcia, A.P. Martinelli, R.A. Bouchard, J.J. Finer, A soybean (*Glycine max*) polyubiquitin promoter gives strong constitutive expression in transgenic soybean, Plant Cell Rep. 28 (2009) 837-849.
- [36] H. Quandt, I. Broer, A. Pühler, Tissue-specific activity and light-dependent regulation of a soybean rbcS promoter in transgenic tobacco plants monitored with the firefly luciferase gene, Plant Sci. 82 (1992) 59-70.
- [37] Y. Fu, H. Zhou, P. Wang, Cloning and identification of the seed specific promoter from soybean, J Northwest A & F University. 12 (2009) 105-111.
- [38] Z.B. Liu, T. Ulmasov, X. Shi, G. Hagen, T.J. Guilfoyle, Soybean GH3 promoter contains multiple auxin-inducible elements, The Plant Cell. 6 (1994) 645-657.
- [39] N. Xu, G. Hagen, T. Guilfoyle, Multiple auxin response modules in the soybean SAUR 15A promoter, Plant Sci. 126 (1997) 193-201.
- [40] H.S. Mason, D.B. DeWald, J.E. Mullet, Identification of a methyl jasmonate-responsive domain in the soybean vspB promoter, The Plant Cell. 5 (1993) 241-251.
- [41] H.C. Park, M.L. Kim, Y.H. Kang, J.C. Jeong, M.S. Cheong, W. Choi, S.Y. Lee, M.J. Cho, M.C. Kim, W.S. Chung, Functional analysis of the stress-inducible soybean calmodulin isoform-4 (GmCaM-4) promoter in transgenic tobacco plants, Mol. Cells. 27 (2009) 475-480.
- [42] E. Czarnecka, J. Key, W.B. Gurley, Regulatory domains of the Gmhsp17. 5-E heat shock promoter of soybean, Mol. Cell. Biol. 9 (1989) 3457-3463.
- [43] Z. Zhang, J.J. Finer, Soybean actin, heat shock protein, and ribosomal protein promoters direct tissue-specific transgene expression in transgenic soybean, In Vitro Cell. Dev. Biol. Plant. 51 (2015) 9-18.

- [44] D. Grant, R.T. Nelson, S.B. Cannon, R.C. Shoemaker, SoyBase, the USDA-ARS soybean genetics and genomics database, *Nucleic Acids Res.* 38 (2010) D843-6.
- [45] E.W. Sayers, et al., Database resources of the National Center for Biotechnology Information, *Nucleic Acids Res.* 39 (2011) D38-D51.
- [46] A.J. Severin, J.L. Woody, Y.T. Bolon, B. Joseph, B.W. Diers, A.D. Farmer, G.J. Muehlbauer, R.T. Nelson, D. Grant, J.E. Specht, M.A. Graham, S.B. Cannon, G.D. May, C.P. Vance, R.C. Shoemaker, RNA-Seq Atlas of *Glycine max*: a guide to the soybean transcriptome, *BMC Plant. Biol.* 10 (2010) 160.
- [47] C. Trapnell, L. Pachter, S.L. Salzberg, TopHat: discovering splice junctions with RNA-Seq, *Bioinformatics.* 25 (2009) 1105-1111.
- [48] C. Trapnell, B.A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M.J. Van Baren, S.L. Salzberg, B.J. Wold, L. Pachter, Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation, *Nat. Biotechnol.* 28 (2010) 511-515.
- [49] D.M. Goodstein, S. Shu, R. Howson, R. Neupane, R.D. Hayes, J. Fazo, T. Mitros, W. Dirks, U. Hellsten, N. Putnam, D.S. Rokhsar, Phytozome: a comparative platform for green plant genomics, *Nucleic Acids Res.* 40 (2012) D1178-D1186.
- [50] C.M. Hernandez-Garcia, J.M. Chiera, J.J. Finer, Robotics and dynamic image analysis for studies of gene expression in plant tissues, *J Vis Exp.* 39 (2010). pii: 1733. doi 10.3791/1733.
- [51] J.J. Finer, P. Vain, M.W. Jones, M.D. McMullen, Development of the particle inflow gun for DNA delivery to plant cells, *Plant Cell Rep.* 11 (1992) 323-328.
- [52] J.M. Chiera, R.A. Bouchard, S.L. Dorsey, E.H. Park, M.T. Buenrostro-Nava, P.P. Ling, J.J. Finer, Isolation of two highly active soybean (*Glycine max* (L.) Merr.) promoters and their characterization using a new automated image collection and analysis system, *Plant Cell Rep.* 26 (2007) 1501-1509.
- [53] E.R. Santarém, J.J. Finer, Transformation of soybean [*Glycine max* (L.) Merrill] using proliferative embryogenic tissue maintained on semi-solid medium, *In Vitro Cell. Dev. Biol. Plant.* 35 (1999) 451-455.
- [54] W.C. Merrick, Mechanism and regulation of eukaryotic protein synthesis, *Microbiol. Rev.* 56 (1992) 291-315.

- [55] G.R. Andersen, P. Nissen, J. Nyborg, Elongation factors in protein biosynthesis, Trends Biochem. Sci. 28 (2003) 434-441.
- [56] M. Axelos, C. Bardet, T. Liboz, A. Le Van Thai, C. Curie, B. Lescure, The gene family encoding the *Arabidopsis thaliana* translation elongation factor EF-1 α : molecular cloning, characterization and expression, MGG. 219 (1989) 106-112.
- [57] W.D. Ransom-Hodgkins, The application of expression analysis in elucidating the eukaryotic elongation factor one alpha gene family in *Arabidopsis thaliana*, MGG. 281 (2009) 391-405.
- [58] T. Berberich, K. Sugawara, M. Harada, T. Kusano, Molecular cloning, characterization and expression of an elongation factor 1 α gene in maize, Plant Mol. Biol. 29 (1995) 611-615.
- [59] N.P. Carneiro, P.A. Hughes, B.A. Larkins, The *eEF1A* gene family is differentially expressed in maize endosperm, Plant Mol. Biol. 41 (1999) 801-814.
- [60] S. Kidou, S. Ejiri, Isolation, characterization and mRNA expression of four cDNAs encoding translation elongation factor 1A from rice (*Oryza sativa* L.), Plant Mol. Biol. 36 (1998) 137-148.
- [61] W. Xu, X. Wang, H. Wang, X. Li, Molecular characterization and expression analysis of nine cotton *GhEF1A* genes encoding translation elongation factor 1A, Gene. 389 (2007) 27-35.
- [62] S. Suhandono, A. Apriyanto, N. Ihsani, Isolation and characterization of three cassava elongation factor 1 alpha (MeEF1A) promoters, PloS one. 9 (2014) e84692.
- [63] F. Aguilar, P. Montandon, E. Stutz, Two genes encoding the soybean translation elongation factor eEF-1 α are transcribed in seedling leaves, Plant Mol. Biol. 17 (1991) 351-360.
- [64] D. Vijaykumar, T. Ramachander, L. Mahishi, R. Kaul, P. Pyati, B. Paul, S. Rawal, Molecular cloning, characterization and tissue specific expression of an elongation factor 1A gene in *Saccharum officinarum* L., Plant Sci. 162 (2002) 315-321.
- [65] C. Curie, T. Liboz, C. Bardet, E. Gander, C. Medale, M. Axelos, B. Lescure, *Cis* and *trans*-acting elements involved in the activation of *Arabidopsis thaliana* A1 gene encoding the translation elongation factor EF-1 alpha, Nucleic Acids Res. 19 (1991) 1305-1310.
- [66] C. Curie, M. Axelos, C. Bardet, R. Atanassova, N. Chaubet, B. Lescure, Modular organization and developmental activity of an *Arabidopsis thaliana* EF-1 α gene promoter, MGG. 238 (1993) 428-436.

- [67] K.L. Fincher, S. Flasiński, and J.Q. Wilkinson. "Glyphosate resistant plants using hybrid promoter constructs." U.S. Patent No. 7,141,722. Nov 28. 2006. Washington, DC: U.S. Patent and Trademark Office.
- [68] G.A. Maston, S.K. Evans, M.R. Green, Transcriptional regulatory elements in the human genome, *Annu. Rev. Genomics Hum. Genet.* 7 (2006) 29-59.
- [69] F. Spitz, E.E. Furlong, Transcription factors: from enhancer binding to developmental control, *Nat. Rev. Genet.* 13 (2012) 613-626.
- [70] B. Plesse, A. Durr, J. Marbach, P. Genschik, J. Fleck, Identification of a new *cis*-regulatory element in a *Nicotiana tabacum* polyubiquitin gene promoter, *MGG.* 254 (1997) 258-266.
- [71] M. Bulger, M. Groudine, Functional and mechanistic diversity of distal transcription enhancers, *Cell.* 144 (2011) 327-339.
- [72] A. Visel, E.M. Rubin, L.A. Pennacchio, Genomic views of distant-acting enhancers, *Nature.* 461 (2009) 199-205.
- [73] N.H. Syed, M. Kalyna, Y. Marquez, A. Barta, J.W. Brown, Alternative splicing in plants-coming of age, *Trends Plant Sci.* 17 (2012) 616-623.
- [74] W.L. Chiu, Y. Niwa, W. Zeng, T. Hirano, H. Kobayashi, J. Sheen J, Engineered GFP as vital reporter in plants. *Curr. Biol.* 6 (1996) 325-330.
- [75] G. Parra, K. Bradnam, A.B. Rose, I. Korf, Comparative and functional analysis of intron-mediated enhancement signals reveals conserved features among plants, *Nucleic Acids Res.* 39 (2011) 5328-5337.
- [76] J. Callis, M. Fromm, V. Walbot, Introns increase gene expression in cultured maize cells, *Genes Dev.* 1 (1987) 1183-1200.
- [77] D. Mascarenhas, I.J. Mettler, D.A. Pierce, H.W. Lowe, Intron-mediated enhancement of heterologous gene expression in maize, *Plant Mol. Biol.* 15 (1990) 913-920.
- [78] P. Vain, K.R. Finer, D.E. Engler, R.C. Pratt, J.J. Finer, Intron-mediated enhancement of gene expression in maize (*Zea mays* L.) and bluegrass (*Poa pratensis* L.), *Plant Cell Rep.* 15 (1996) 489-494.
- [79] J.E. Gallegos, A.B. Rose, The enduring mystery of intron-mediated enhancement, *Plant Sci.* 237 (2015) 8-15.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Table 1. GmScream promoters, Gene IDs, and Gene annotations for their corresponding *GmScream* genes

	Promoter	Size (bp)	Gene ID		Gene functional annotation
			Wm82.a1.v1	Wm82.a2.v1	
Group I	GmScream1	1225	Glyma03g32020	no correspondence	Cupin
	GmScream2 ^b	1391	Glyma03g32030	Glyma.03g163500	Cupin
	GmScream3 ^b	1229	Glyma10g04280	Glyma.10g037100	Cupin
	GmScream4 ^c	1434	Glyma10g35080	no correspondence	Extensin
	GmScream5 ^a	1683	Glyma10g35870	Glyma.10g214500	Unknown
	GmScream6 ^b	1144	Glyma10g39150	Glyma.10g246300	Cupin
	GmScream7	1361	Glyma14g06680	Glyma.14g061500	Major intrinsic protein
	GmScream8	1519	Glyma17g01720	Glyma.17g013600	Peroxidase
	GmScream9	1233	Glyma19g13060	Glyma.19g063400	Oleosin
	GmScream10 ^c	1413	Glyma19g36440	no correspondence	Unknown
Group II	GmScreamM1 ^a	1263	Glyma05g11630	Glyma.05g089000	Elongation factor 1A
	GmScreamM2 ^{ab}	1527	Glyma18g44850	Glyma.18g216000	Ubiquitin-conjugating enzymes
	GmScreamM3	1574	Glyma09g04950	Glyma.09g044200	Translationally controlled tumor protein
	GmScreamM4 ^a	1420	Glyma05g24110	Glyma.05g114900	Elongation factor 1A
	GmScreamM5 ^b	1246	Glyma13g04580	no correspondence	Unknown
	GmScreamM6 ^{ab}	1331	Glyma09g40960	Glyma.09g273100	Ubiquitin-conjugating enzymes
	GmScreamM7 ^b	1410	Glyma19g39070	Glyma.19g203300	Ribosomal protein S28e
	GmScreamM8 ^a	1576	Glyma17g23900	Glyma.17g186600	Elongation factor 1A
	GmScreamM9	1488	Glyma08g43290	Glyma.08g320500	Ubiquitin family
	GmScreamM10	1481	Glyma15g15910	Glyma.15g148900	Translationally controlled tumor protein

^aThe promoters contain a leader intron in the 5' untranslated region (5'UTR).

^bThe promoters showed little activity.

^cThe promoters showed no activity.

Table 2. Expression analysis of GmScream candidate genes (Group II) identified from NCBI SRA for soybean cultivar ‘Williams 82’.

Promoters Groups of tissues/treatments	#GmScream									
	M 1	M 2	M 3	M 4	M 5	M 6	M 7	M 8	M 9	M 10
1. Dry Whole Seed	X	X	X	X	X	X	X	X	X	X
2. Whole seedling 6 days after imbibition	X	X	X	X	X	X	X	X	X	X
3. Globular stage whole seed		X								
4. Heart stage whole seed	X	X	X	X	X	X		X	X	X
5. Cotyledon stage whole seed	X	X	X	X	X	X	X	X	X	X
6. Cotyledons of seedlings, 6 days after imbibition		X	X	X	X	X		X	X	X
7. Embryonic cotyledons from late-maturation (AA1) stage embryos	X	X	X	X	X	X		X	X	X
8. Embryonic cotyledons of mid-maturation (B1) stage embryos	X	X	X	X	X	X	X	X	X	X
9. Apical meristem	X	X	X	X	X	X	X	X	X	X
10. Trifoliate leaves	X	X	X	X	X		X	X	X	X
11. Primary leaves, 3-week-old, 0 dai <i>Phakopsora pachyrhizi</i>			X		X		X	X	X	
12. Primary leaves, 3-week-old, 10 dai <i>Phakopsora pachyrhizi</i>	X	X	X		X		X	X	X	X
13. Stem	X	X	X	X	X	X	X	X	X	X
14. Root	X	X	X	X	X	X	X	X	X	X
15. Root tip	X		X	X	X	X	X	X	X	X
16. Stripped roots, mock inoculation	X	X	X	X	X	X	X	X	X	X
17. Root hairs, mock inoculation	X	X	X	X	X	X	X	X	X	
18. Root hairs, inoculation with <i>Bradyrhizobium japonicum</i>	X		X	X	X	X	X	X	X	X
19. Nodule	X	X	X		X	X	X	X	X	X
20. Floral bud	X	X	X	X	X	X	X	X	X	X
21. Flower (open)	X	X	X	X	X	X	X	X	X	X
22. Green pods	X	X	X	X	X	X	X	X	X	X
# of groups that the gene was categorized in the highest 1% of expressing genes	19	19	21	18	21	18	18	21	21	19

“X” indicates the gene was categorized in the highest 1% of expressing genes.

Table 3. Grouping of GmScream promoters compared to the 35S promoter based on transient GFP expression in lima bean cotyledons and stable expression in hairy roots

Relative Expression level	GmScream#			
	High	Medium	Low	No activity
Transient expression	5, 7, M1, M4, M8, M9	1, 8, 9, M3, M10	2, 3, 6, M2, M6, M5, M7	4, 10
Stable expression	5, M1, M4, M8	1, 6, 7, 8, 9, M3, M6, M9, M10	2, 3, M2, M5, M7	4, 10

Fig. 1. Expression profiles of GmScream candidate genes (Group I) identified through RNA-seq Atlas.

Fig. 2. Transient GFP expression driven by GmScream promoters in lima bean cotyledons. (A) Images from peak expression time. (B) Profiles of transient GFP expression driven by GmScream promoters over 100 h. GFP expression levels for each promoter were reported as the percentage of peak expression of the 35S promoter. The order of the promoters in the legend is reflected in the height of the curves. Bar = 1 mm.

Fig. 3. GFP expression in stably-transformed soybean hairy roots harboring different GmScream promoter constructs. (A) GFP expression in 5 independent roots obtained from each promoter. (B) Mean GFP expression intensity ($\pm$ SD) in soybean hairy roots transformed with GmScream constructs. Columns followed by the same letter are not significantly different at $p < 0.05$ using Tukey's Studentized Range (HSD) test. Bar = 1 mm.

Fig. 4. Analysis of GFP intensity transformed with GmScreamM8 extension and deletion constructs. (A) A schematic structure of all the extension and deletion constructs. (B) Profiles of transient GFP expression in lima bean cotyledon tissues. (C) Mean GFP expression intensity ($\pm$ SD) in soybean hairy roots transformed with different constructs. Columns followed by the same letter are not significantly different at $p < 0.05$ using Tukey's Studentized Range (HSD) test.

Fig. 5. GmScreamM4 or GmScreamM8-mediated GFP expression in different developmental stages of primary soybean embryogenic tissues. Bars = 1 mm.

Fig. 6. GFP expression in T₁ transgenic soybean plants containing the GmScreamM4 and GmScreamM8 promoter. (A) GFP expression in vegetative tissues of T₁ transgenic soybeans. (B) GFP expression in reproductive tissues of T₁ transgenic soybeans. Bars = 1 mm.

Figure 2

Figure 3

Figure 4

Figure 5

Figure 6