Comparative transcriptome analysis between somatic embryos (SEs) and zygotic embryos in cotton: evidence for stress response functions in SE development

Fangyan Jin1,†, Lisong Hu1,†, Dajun Yuan1, Jiao Xu1, Wenhui Gao1, Liangrong He1,2, Xiyan Yang1,* and Xianlong Zhang1,*

1National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan, Hubei, China
2College of Plant Science, Tarim University, Alar, Xinjiang, China

Summary
As a product of asexual reproduction in plants, the somatic embryo (SE) differentiates into a new plantlet via a zygotic embryogenesis-like process. Here, we present the phenotypic and cellular differences between SEs and zygotic embryos (ZEs) revealed by histological section scanning using three parallel development stages of the two types of embryos of cotton (Gossypium hirsutum cv. YZ1), including globular, torpedo and cotyledonary-stages. To identify the molecular characteristics of SE development in cotton, the digital gene expression system was used to profile the genes active during SE and ZE development. A total of 4242 differentially expressed genes (DEGs) were identified in at least one developmental stage. Expression pattern and functional classification analysis based on these DEGs reveals that SE development exhibits a transcriptional activation of stress responses. RT-PCR analysis further confirmed enhanced expression levels of stress-related genes in SEs than in ZEs. Experimental stress treatment, induced by NaCl and ABA, accelerated SE development and increased the transcription of genes related to stress response, in parallel with decelerated proliferation of embryogenic calluses under stress treatment. Our data reveal that SE development involves the activation of stress responses, which we suggest may regulate the balance between cell proliferation and differentiation. These results provide new insight into the molecular mechanisms of SE development and suggest strategies that can be used for regulating the developmental processes of somatic embryogenesis.

Keywords: embryo development, Gossypium hirsutum, somatic embryo, zygotic embryo, transcriptional profiling, stress response gene.

Introduction
Embryogenesis is an important developmental process that represents the critical transition from vegetative to embryogenic growth during the life cycle of higher plants. This process, known as zygotic embryogenesis, starts with double fertilization to produce an endosperm and a zygote, which transits through globular-, torpedo- and cotyledonary-stages prior to germination. Single somatic cells, or groups of somatic cells, can be also be induced experimentally to form somatic embryos (SEs) that can develop into new plants. SEs resemble zygotic embryos (ZEcs) in many respects. First, both ZEs and SEs display cell polarity and asymmetric cell division during initiation (Dodeman et al., 1997). Second, both developmental processes undergo similar morphological changes, passing through globular- to cotyledonary-stages and finally developing into seedlings (Ikeda et al., 2006). The most significant similarity is that both processes show common expression of embryo-specific genes and signalling regulation. Somatic embryogenesis receptor-like kinase from Daucus carota (DcSERK) was the first identified marker gene shown to play a crucial role during somatic embryogenesis (Schmidt et al., 1997). The positive regulation of SERK in the induction of somatic embryogenesis has been widely demonstrated in monocots and dicots. Additionally, SERK genes are positively regulated during zygotic embryogenesis in various species (Hecht et al., 2001; Hu et al., 2005). Moreover, the identification of additional crucial genes in both SE and ZE regulation, such as ABSCISIC ACID INSENSITIVE3 (ABI3), FUSCA3 (FUS3) and LEAFY COTYLEDON1 (LEC1), has advanced the understanding of embryo development (Braybrook et al., 2006; To et al., 2006; Yang and Zhang, 2010). There are obvious differences between SEs and ZEs in the origin of embryogenesis. ZEs develop from fertilized eggs (Pasternak et al., 2002), but SEs originate from differentiated somatic cells that restructure to obtain embryogenic competence under suitable in vitro conditions (Dodeman et al., 1997). The induction of SE is a multifactorial event. Stress, as well as plant hormones, has been shown to play an important role during somatic embryogenesis (Karami and Saedt, 2010). It is known that excised explants used for SE induction, which are subjected to sterilization procedures, cut into pieces and cultured in an artificial nutritional environment, undergo considerable stress during this process. As the first report of SE formation in carrot cell suspension cultures (Reinert, 1959; Steward et al., 1958), the plant tissue culture system has been optimized continuously, which has enabled thousands of plant species to be regenerated. Increasing evidence has revealed that the stress response of cultured tissue plays an important role in SE induction (Pasternak et al., 2002; Thibaud-Nissen et al., 2003). The relevant stress factors include starvation, heat shock, osmotic stress resulting from high levels of salt, polyethylene glycol (PEG) or sucrose and abscisic acid (ABA) supply (Langhansova et al., 2004; Stasolla and Yeung, 2003; Touraev et al., 1996). There are also differences in physiological features, among which the most
important one is desiccation tolerance (Shiota et al., 1998). At the end of the developmental phase, the effects of ABA accumulation, growth inhibition and environmental stress cause ZEs to dehydrate; this effect results in a rapid decline in water content and increased storage protein synthesis, after which the embryo enters dormancy (Dodeman et al., 1997). Despite the synthesis of storage proteins during development, even at the early globular stage (Dodeman et al., 1997), SEs develop into seedlings, skipping over the process of dehydration and dormancy. Additionally, some genes have different expression change patterns in the two parallel embryonic pathways (Dong and Dunstan, 1999). Based on morphological resemblance, comparing the mechanistic differences between SE and ZE development might be a good way to understand the diversified process during somatic embryogenesis, especially the role of stress in SEs.

Molecular biological technology has made it possible to understand the molecular mechanisms regulating somatic embryogenesis. Microarray analysis has revealed expression of numerous genes associated with oxidative stress, response to stress and storage protein synthesis during soybean somatic embryogenesis (Thibaud-Nissen et al., 2003). Hundreds of differentially expressed genes (DEGs) were identified during cotton somatic embryogenesis (Zeng et al., 2006), suggesting candidates that might regulate the necessary balance between cell differentiation and cell proliferation (Hu et al., 2011; Zeng et al., 2007). More recently, a more complete picture of the cellular and molecular events during cotton somatic embryogenesis has been illustrated using an elite genotype *Gossypium hirsutum* cv. YZ1, using high-throughput RNA-Seq, sRNA-Seq and degradome sequencing technology (Yang et al., 2012, 2013). These studies describe the expression of a large number of genes during somatic embryogenesis and provided abundant resources for the elucidation of molecular mechanisms underlying somatic embryogenesis.

In this paper, we describe the use of next-generation sequencing (Illumina Digital Gene Expression II, DGE II) technology to compare gene expression changes during three parallel developmental stages of ZEs and SEs (globular-, torpedo- and cotyledon stage). We went on to test the hypothesis that experimental stress promotes SE development and prevents EC proliferation by regulating stress-related gene expression. This study brings new insights into the regulation of SE development.

**Results**

**Morphological and ultrastructural comparison between SEs and ZEs**

Somatic embryos and ZEs showed similar features in development, with both undergoing globular-, torpedo- and cotyledon stages (ZGE, zygotic globular-stage embryo; ZTE, zygotic torpedo-stage embryo; ZCE, zygotic cotyledon-stage embryo; SGE, somatic globular-stage embryo; STE, somatic torpedo-stage embryo; SCE, somatic cotyledon-stage embryo; Figure 1a–f). However, there were also differences in morphological characteristics. SEs were larger than ZEs in size at each stage, and the SE cotyledons were often irregular in shape and multiple, distinct from the two prominent cotyledons in cotton ZCEs.

To observe the subcellular structure of SEs and ZEs, the apical meristem tips were selected for transmission electron microscopy (TEM) analysis. The results revealed that the apical cells of SEs were vacuolated, with conspicuous mitochondria over all three stages, and numerous lipid droplets and starch grains at the torpedo- and cotyledon stages (Figure S1-a1–c2). In contrast, ZE cells had prominent nuclei, few small vacuoles and few subcellular organelles (Figure S1-d1–f2).

**Global analysis of DEGs**

Typical SEs (Figure 1a–c) and ZEs (Figure 1d–f) were selected for transcriptional analysis. We sampled three parallel developmental stages of SEs and ZEs (globular-, torpedo- and cotyledon stage) for sequencing, and each sample generated 0.55 GB of sequence data on average, including approximately 11.8 million of total reads of approximately 50 bp in length. After mapping these reads to the reference sequences, which were cotton unigenes collected from the National Centre for Biotechnology Information.

---

**Figure 1** Morphology of somatic embryos (SEs) and zygotic embryos (ZEs) at three developmental stages used for RNA-Seq analysis. (a–c) SE at the globular, torpedo-shaped and cotyledonal stages in order. (d–f) ZE at the globular, torpedo-shaped and cotyledonal stages in order. Bar = 0.2 mm in (a–c), 0.1 mm in (d–f).
After integrating the sequencing data of six samples, we detected a total of 20,220 genes. There were 9103 genes found to be highly expressed throughout embryo development, filtering with RPKM ≥ 25. These genes were then classified into two groups, (i) defined as DEGs (i.e. the genes with an absolute value log2 ratio ≥ 1.5849 which was equal to at least a 3.0-fold change) or (ii) defined as not DEGs (NDEGs). The NDEGs mainly included housekeeping genes, such as metabolism genes for maintaining cellular activity, but also embryogenesis-related genes, including BAM1 (BARELY ANY MERISTEM1), FUS3-complementing genes, AGPs (ARABINOGLUCAN GLYCOPROTEINS), SERKs (Table S1). These embryogenesis-related genes play roles in embryonic competence, cell division, polarity formation and normal embryo growth, which have been reported in both plant ZEs and SEs (Yang and Zhang, 2010). We focused our attention on the 4242 DEGs (Table S2). The number of DEGs at the three developmental stages is shown in Figure 2a. Compared with ZEs, the number of up-regulated genes was significantly higher than that of down-regulated genes at all SE stages. Moreover, the number of DEGs increased during the developmental process. Spatial analysis of the 4242 DEGs in the three developmental stages is summarized in Figure 2b. Of the DEGs, 1570 were expressed at all three developmental stages, 161 genes at the globular- and torpedo stages, 876 genes at the DEGs, 1570 were expressed at all three developmental stages. Type III genes, including 1565 DEGs, were down-regulated in SEs compared with ZEs. These genes were divided into three subclusters: Cluster 1 (573 genes) was almost equally expressed at the globular stage and down-regulated at the torpedo- and cotyledon stages; Cluster 2 (658 genes) was down-regulated at equal levels during all three stages; and Cluster 3 (334 genes) was down-regulated especially at the cotyledon stages. Type II genes, including 2368 DEGs, were up-regulated in SEs compared with ZEs. These genes were also divided into three subclusters: Cluster 4 (803 genes) was up-regulated at equal levels during all three stages; Cluster 5 (1064 genes), was up-regulated at the torpedo- and cotyledon stages and reached a maximal ratio at the cotyledon stage; and Cluster 6 (501 genes) showed an expression pattern similar to that of Cluster 5, but with a higher ratio than in Cluster 5. Type III (309 genes) DEGs were significantly up-regulated in SEs, with three clusters: Cluster 7 (94 genes) was highly up-regulated during all three stages with an extremely high ratio at the globular stage; Cluster 8 (118 genes) displayed an equal expression pattern during all three stages, but with a high log2 ratio (approximately 8) in SEs vs. ZEs; Cluster 9 (97 genes) was up-regulated to a large extent (with log2 ratios from approximately 7 to 16 in all three stages).

<table>
<thead>
<tr>
<th>No. of total reads</th>
<th>SGE</th>
<th>STE</th>
<th>SCE</th>
<th>ZGE</th>
<th>ZTE</th>
<th>ZCE</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. (%) of total mapped reads</td>
<td>11 484 979</td>
<td>11 647 501</td>
<td>12 214 281</td>
<td>11 850 158</td>
<td>11 488 161</td>
<td>11 144 188</td>
<td>11 799 878</td>
</tr>
<tr>
<td>No. (%) of unique match</td>
<td>5 221 570</td>
<td>5 317 179</td>
<td>5 472 276</td>
<td>5 892 850</td>
<td>5 675 152</td>
<td>6 429 916</td>
<td>5 688 157</td>
</tr>
<tr>
<td>No. (%) of multiposition match</td>
<td>5 892 850</td>
<td>5 317 179</td>
<td>5 472 276</td>
<td>5 892 850</td>
<td>5 675 152</td>
<td>6 429 916</td>
<td>5 688 157</td>
</tr>
<tr>
<td>No. (%) of total unmapped reads</td>
<td>4 877 960</td>
<td>4 970 786</td>
<td>5 126 186</td>
<td>5 358 593</td>
<td>5 209 080</td>
<td>5 924 477</td>
<td>5 244 513</td>
</tr>
<tr>
<td>No. (%) of detected unigenes</td>
<td>4 877 960</td>
<td>4 970 786</td>
<td>5 126 186</td>
<td>5 358 593</td>
<td>5 209 080</td>
<td>5 924 477</td>
<td>5 244 513</td>
</tr>
</tbody>
</table>

SCE, somatic cotyledon-stage embryo; SGE, somatic globular-stage embryo; STE, somatic torpedo-stage embryo; ZCE, zygotic cotyledon-stage embryo; ZGE, zygotic globular-stage embryo; ZTE, zygotic torpedo-stage embryo.

© 2013 Society for Experimental Biology, Association of Applied Biologists and John Wiley & Sons Ltd, Plant Biotechnology Journal, 12, 161-173
DEGs were assigned to the ontology of biological process. The GO functional classification analysis on each DEG type was performed to calculate the functional category distribution frequency based on biological process level 2 and level 3 (Figure 4). ‘Cellular process’ was the most highly represented active group in the biological process level 2 category, and ‘response to stimulus’ was the second, account for 20.1%. Other classifications found included ‘metabolic process’, ‘biological

Figure 2 Histogram and Venn diagram of differentially expressed genes (DEGs) between somatic and zygotic embryo (ZE) development partitioned into three groups: globular, torpedo-shaped and cotyledonal stages. (a) The number of DEGs up- or down-regulated at different stages, suggesting that there were more genes at higher expressional levels in somatic embryos (SEs) than ZEs. (b) Venn diagrams showing similarly or distinctly regulated genes over the three stages.

Figure 3 Cluster analysis of 4242 differentially expressed genes (DEGs) between somatic and zygotic embryos (ZEs) at the three developmental stages based on the K-means method. The DEGs were divided into three types, represented with three colours.
regulation’ and ‘developmental process’. Response to stimulus level 3 mainly comprised ‘response to chemical stimulus’, ‘response to stress’ and ‘response to abiotic stimulus’. Therefore, a prominent stress response was involved during the process of SE development, compared with ZE development.

Stress-related genes are active during cotton SE development

Combined with KEGG analysis and BLAST annotation, a large number of stress-induced genes were found in the DGE II profiling, and almost all of these genes showed higher expression levels in SEs. These genes mainly included hormone-related genes, kinase genes, genes encoding transcription factors and downstream stress responsive genes as shown in Table S3.

Hormone-related genes

A short-chain alcohol dehydrogenase, ABA2, involved in the synthesis enzyme of abscisic aldehyde, was highly expressed in SEs, belonging to Type III Cluster 7. Eight genes involved in ABA-mediated signalling had higher expression levels in SEs than in ZEs, including PYL, PP2CA, PP2C, SnRK2, ABF1, ABF2, ABR1 and ABR. Moreover, JAR1, JAZ1, JAZ2 and JAZ10, which respond to phytohormone jasmonic acid (JA), were also highly expressed in SEs.

Kinase genes

Two members of the CDPK family genes, seven CIPK family members, six MAP kinase family members and MAP3KA, totalling 16 kinases were all up-regulated in SEs compared with ZEs, belonging to Type II.

Genes encoding transcription factors

There were three NAC family members, including NAC002, NAC072 and NAC090, clustered to Type III and two other members belonging to Type II. There were 15 WRKY family members, six members belonging to Type II and others belonging to Type III. Four MYB family transcription factors emerged in Type II. Four genes encoding zinc finger family proteins were in Type II and Type III.

Downstream stress responsive genes

LEA1, LEA6, LEA14 and two LEA proteins were in Type III, and LEA5 belonged to Type II. Several genes, initially isolated from dehydrating control library and named as Early Responsive to Dehydration (ERD) and Responsive to Dehydration (RD), including ERD5, ERD7, ERD8, ERD15, RD2, RD20, RD22 and RD28 in the DEGs, were clustered to Type II. Among these, ERD5, ERD8, ERD15, RD22 and RD28 were identified as stress-responsive genes.

From all of the above stress-inducible genes in DEGs, 27 typical genes, including five genes from the ABA biosynthesis and signalling pathway, five LEAs, three RDs and 14 WRKY family members, were selected for validation by RT-PCR, and the results coincided with data from the transcription profile, as shown in Figures 5 and S3.

The stress response genes activated during SE development could finally be mapped to the convergent points

![Figure 4](image_url)
which are woven into an integrated stress-signalling network
drawn by Fujita et al. (2006), indicating that there are active
stress signalling pathways during SE development.

Stresses affect SE development and drive expression of
stress response genes

To verify the function of stress in the SE development, different
concentrations of NaCl (0–100 mM) and ABA (0–10 μM) were
added to MS medium. After 1 week of NaCl treatment, EC
growth was inhibited with the increased concentration of NaCl
on MS medium. Moreover, ECs died in the MS medium
with 100 mM NaCl. After 3 weeks, SEs could be observed
(Figure 6a–e). The fresh weight of ECs decreased as the concen-
trations of NaCl increased (Figure 6f), meaning that EC proliferation
was repressed under stress treatment. Conversely, the SE differen-
tiation was promoted and sustained when the concentration of
NaCl was increased in an appropriate range (0–75 mM). The
excessive concentration of NaCl (100 mM) interrupted EC prolifer-
ation and differentiation. The tendency towards proliferation and
differentiation of the ECs under NaCl treatment was summarized in
Figure 6f.

To further determine the expression of stress response genes
identified by the DGE II profiling during SE development under
stress treatment, we examined the expression of these genes by
qRT-PCR. The expression level of WRKY6 (zhu1_Ghi#S33830951), WRKY72 (zhu1_Ghi#S28672016), SAG20 (Senescence-
Associated Gene 20, zhu1_Ghi#S28687560) and two LEA proteins
(zhu1_Ghi#S28710690, zhu1_Ghi#S28660798) increased with
increasing concentrations of NaCl, which was significant at
75 mM NaCl (Figure 6g–k).

Similarly, EC proliferation decreased with increasing concen-
trations of ABA (Figure 7a–e). The number of SEs increased with
ABA concentration ranging from 0 to 5 μM, and the maximum
number was achieved at 5 μM of ABA. However, both proliferation
and differentiation were depressed at 10 μM (Figure 7f).

The stress response genes ABR (zhu1_Ghi#S33824883), RD22
(zhu1_Ghi#S28710342), LEA6 (zhu1_Ghi#S28710690) and other
two LEA proteins (zhu1_Ghi#S28710690, zhu1_Ghi#S28660798)

Figure 5 Detailed expression profiles of genes related to stress response. The relative expression level was obtained by RNA-seq after equation and logarithmic transformations of RPKM and by RT-PCR for data verification.
were induced by exogenous application of ABA. The highest levels were detected with treatments of high concentrations of ABA (5 and 10 μM) (Figure 7g–k). The results show that both ABA and NaCl treatments enhance the transcriptional activities of genes related to stress response and both (at high concentrations) repress EC proliferation and induce SE formation.

Discussion

Comparative analysis between SEs and ZEs at cellular and molecular levels

Somatic embryos undergo a similar developmental program to ZEs (Dodeman et al., 1997), and it can be hypothesized that the same set of genes might be operating in both cases to specify embryo development. Our analysis of zygotic and somatic embryogenesis in cotton shows that both progress passing through the globular-, torpedo- and cotyledon stages, with larger size in SEs than in ZEs (Figure 1), coinciding with the previous reports in dicots (Suhasini et al., 1997; Xu and Bewley, 1992). Despite the similarities described between the two types of embryogenesis, some key differences exist, namely the lack of surrounding by embryo sac and differentiation of endosperm in the case of SEs (Figure 1). These elements played a key role in bringing about the successful maturation of the ZEs. During tissue culture, morphogenesis of SEs occurred without the surrounding embryo sac and the simultaneous development of a normal endosperm tissue. These suggested that either those ZEs-related tissues did not play a crucial role in embryo development, or embryo culture condition interactions took over from those between the embryos and tissues.

At the subcellular level, lipid droplets, mitochondria and vacuoles largely increased during the development of ZEs in contrast to SEs (Figure 6). The developmental stage and environmental condition seem to influence these subcellular structures in both ZEs and SEs. However, the type and number of lipid droplets, mitochondria, and vacuoles do differ between SEs and ZEs (Figure 6a–e). The lipid droplets and mitochondria are crucial for cellular energy storage and metabolic processes, and their increased expression in ZEs suggests a higher metabolic demand compared to SEs. The vacuoles play a role in storing nutrients and maintaining cell turgor pressure; their increased expression in ZEs indicates a possible role in nutrient storage or cell expansion.

Figure 6  EC treated with NaCl. (a–e) EC treated with NaCl at 0, 25, 50, 75 and 100 mM, captured by stereo microscopy; white arrows show somatic embryos (SEs), and bars = 1 mm. (f) Number of SEs and fresh weight of calluses per plate; error bars represent the SD for three replicates. (g–k) Relative expression levels of stress-inducible genes under NaCl treatment, calculated by the equation ratio $= 2^{ΔCt}$ using GhUBQ7 as the control. Error bars represent the SD for three biological replicates and two technical replicates.
some plant species (Taylor and Vasil, 1996), and cotyledon cells of SEs contained numerous mitochondria and vacuoles in ginseng, which were the signs of active metabolic activity (Choi et al., 1999). In our study, both SEs and ZEs were characterized by increased number of lipid droplets, mitochondria and vacuoles during embryo development, suggesting that the metabolic activity in SE and ZE increased gradually (Figure S1). Moreover, we show in this paper that there were 4861 genes which were expressed highly but not differentially throughout SE and ZE development (Table S1). GO analysis showed that the NDEGs mainly participated in metabolic and cellular processes (Figure S4), involving numerous housekeeping genes. In addition, multiple embryogenic genes were categorized among the NDEGs. SERKS, which have been reported to fulfill specialized functions in embryogenesis, were expressed nondifferentially in SEs and ZEs at parallel stages (Hecht et al., 2001; Hu et al., 2005). BAM1, a receptor kinase-like protein gene, which is required for stem cell maintenance (DeYoung and Clark, 2008), was also expressed in both SEs and ZEs. These results suggested there were transcriptional, as well as morphological, similarities between cotton SE and ZE development.

However, it was concluded that SEs exhibited more metabolic activity than ZEs at parallel stages. Bigger vacuoles, more lipid droplets and mitochondria were present in the cells of SEs than in the cells of parallel development stages ZEs (Figure S1). Taylor and Vasil (1996) reported that SEs displayed cellular ultrastructure similar to later stages in ZEs. Meanwhile, there were larger numbers of genes differentially expressed in SEs than in ZEs at parallel stages.
parallel stages (Figure 2a). The results shown on the Venn diagram in Figure 2b indicate that there were 1570 DEGs common among the three stages. Moreover, the GO annotations of NDEGs and DEGs revealed that transcription activated by stress in SEs and ZEs during development was partially and integrally different (Figures 4 and S4). Six LEA genes (zhu1_Ghi#528710690, zhu1_Ghi#528660798, zhu1_Ghi#528710523, zhu1_Ghi#528710296, zhu1_Ghi#533810188, zhu1_Ghi#528710696) present in our DEGs were expressed throughout SE development (Table S2). Studies of storage protein synthesis in cotton SEs at the early globular stage confirm that protein synthesis and accumulation patterns mimic those reported for zygotic systems, but at much earlier stages and to a lesser degree (Shoemaker et al., 1987). We found that the gene encoding the ZS albumin protein (zhu1_Ghi#528710701) was expressed highly until cotyledon-stage in ZEs (RPKM, ZGE: 1.636984, ZTE: 99.69097, ZCE: 5471.502), whereas the transcript was detectable at each stage with considerable levels (SGE: 2319.779, STE: 1985.284, SCE: 2901.853) in SEs (Table S2). It was coinciding with the fact that the ZEs are surrounded by the embryo sac and fed by the endosperm in the ovule environment, while SEs exposed to in vitro culture system with significant stresses (Karami and Saidi, 2010). However, ZEs also undergo stresses following the severing of the vascular connection with the mother plants, and seed desiccation. At the end of ZE developmental stage, several stress response genes, such as HSPs and LEAs, are considered to participate in preparation for desiccation during maturation to ensure that ZEs survive dormancy (Dodeman et al., 1997; Huang et al., 2012).

**Stress signalling involved in SE development**

Various studies have been carried out to investigate the molecular mechanisms underlying SE development. Somatic embryogenesis in soybean and *Medicago truncatula* was reported to be accompanied with differential expression of stress-related genes (Mantiri et al., 2008; Thibaud-Nissen et al., 2003). Our RNA-Seq profiling, comparative analysis and GO annotations of DEGs also indicated that stress response-related transcription was notable during SE development (Figure 4). Additionally, several genes involved in SE development have been considered to be involved in stress signalling pathways, suggesting a relationship between SE development and stress signalling (Yang and Zhang, 2010). *SERK*, a key factor in regulating embryogenic competence, was demonstrated to regulate cell death and increase stress resistance in plants (He et al., 2007; Hu et al., 2005). Another essential gene for SE development, *MTSERF1*, an *ERF* family member, was induced by ethylene (Mantiri et al., 2008). Therefore, stress signalling has been occasionally identified in plants during somatic embryogenesis.

In the stress-signalling pathway, many genes may be expressed in embryos. ABA, which plays central role in stress signalling (Fujita et al., 2006; Shinozaki and Yamaguchi-Shinozaki, 2007), has been demonstrated to be important in development and maturation of ZEs (Chiwocha and von Aderkas, 2002) and SEs (Vahdati et al., 2008). Overexpression of *GhCIPK6* conferred salt and drought tolerance in transgenic *Arabidopsis* (He et al., 2013). It can also regulate somatic embryogenesis in cotton (X. Yang, J. Deng, L. He, X. Zhang, unpublished data). Various transcription factors, including members of the WRKY, MYB, NAC, *ERF*, zinc finger and basic-domain leucine zipper (bZIP) families, have been shown to regulate stress response and improve stress tolerance in plants by binding to cis-acting elements of downstream stress response genes, including ERDs, RDs, LEAs and osmotins (OSMs; Shinozaki and Yamaguchi-Shinozaki, 2007; Shinozaki et al., 2003). Data presented in the current work show that a large number of expressed stress response genes are involved in ABA biosynthesis and response (ABB, PYT, ABF1 and ABRI), JA response, some genes encoding kinases (CIPKs, CDPKs, MAPks), transcription factors (NACs, WRKYs, MYBs, ERFs, *Zinc Finger family proteins*) and downstream stress responsive genes (LEAs, ERDs, RDs, SAGs; Table S3).

Therefore, our data suggest a role for stress signalling during cotton SE development, similar to an integrated stress-signalling network proposed by Fujita et al. (2006). We can speculate that the in vitro tissue culture conditions may lead to the transcriptional activation of certain stress responsive signal transduction pathways, potentially involving phytohormone- and kinase-signalling pathway components, in turn activating stress-related transcription factors and downstream stress responsive genes.

**SE development is a process of adapting to in vitro stress**

Stress has been confirmed to play a key role in the induction of somatic embryogenesis in previous culture optimizing studies. For example, PEG and ABA treatment improve the efficiency of SE maturation in *Panax ginseng* (Langhansova et al., 2004). Rapid SE development has been achieved through metabolic stresses caused by reducing concentrations of nutrients and dehydration in cotton (Kumria et al., 2003). In our study, NaCl (0–75 mM) and ABA (0–5 μM) promoted SE development, coinciding with the previous studies, but higher concentrations of NaCl and ABA had adverse effects (Figures 6 and 7), indicating that severe but not fatal stress treatment could accelerate SE development. Besides SE, cultured immature ZEs, with already-existing embryogenically competent cells, could develop into seedlings in a simulated in vitro ovule environment, by skipping the later stages (like SEs) compared with in vivo (Fuller et al., 2011; Raghavan, 2003). In this way, stress responses observed during SE development might reflect the lack of the natural in vivo environment, suggesting the ability of stress responses to accelerate SE development. Therefore, the cells of cultured SEs and ZEs underwent stress stimulation by exogenous compounds: hormones, salt ions, glucose, vitamins and others in *in vitro*, and embryos and seedlings were interpreted as outcome of an *in vitro* adaption process to the culture environment.

However, the exact relationship between stress and SE development remains to be determined. In soybean, transcript profiles of the genetic response to oxidative stress and cell division suggest that SE development depends on a balance between cell proliferation and cell death (Thibaud-Nissen et al., 2003). Embryogenic competence in cotton was acquired from somatic cells through dedifferentiation and redifferentiation in tissue culture conditions, during which the cells were stimulated by exogenous compounds in the culture medium (Yang and Zhang, 2010). These culture conditions might cause stress response in cells, which presumably redirect intracellular signalling to modulate cellular patterning and differentiation (Zeng et al., 2007). In our study, results of NaCl and ABA treatment indicate that these treatments promote SE development (Figures 6 and 7). During the suitable stress treatment, the tendency of proliferation and differentiation showed that embryogenic cells have to confront a fate choice: continue proliferating or transform into embryos to survive. The stress factors acted as selectors to regulate the balance between proliferation and differentiation that determined SE development. In summary, we propose that...
SE development is a process whereby embryogenic cells respond to local environmental (e.g. ‘stress’) conditions resulting in changes in cell growth and differentiation, from uncontrolled proliferation in callus to embryogenesis.

**Stress response, possibly interacting with auxin signalling, results in cotton SE development**

In addition to the possibility that stress plays a key role in somatic embryogenesis, auxin has been considered to be critical during somatic embryogenesis based on the accumulated data (Fehér et al., 2003). First, application of exogenous auxin is vital for *in vitro* cultures. During this process, auxin might have several roles, acting as auxin directly or modifying intracellular indole acetic acid (IAA) metabolism and/or as a ‘stressor’ (Fehér et al., 2003; Pasternak et al., 2002). In our previous study, the endogenous auxin concentration reached a peak at the auxinogenic callus stage during somatic embryogenesis (Yang et al., 2012). In addition, the state of rapid cell proliferation and active aerobic metabolism occurring at the presence of exogenous auxin led to an oxidative burst in the tissue through generating reactive oxygen species (Pfeiffer and Höftberger, 2001).

Following the auxin surge, numerous auxin responsive genes were differentially expressed during SE development (Yang et al., 2012). Meanwhile, some auxin responsive genes, tryptophan synthase beta-subunit 2 (*TSB2*), IAA amido synthetase (*GH3*), Aux/IAAs and auxin response factors were accumulated in our profiling (Table S2). Other than that, many genes that were induced by auxin also found to be responsive to various abiotic stresses. For example, *Dchsp-1*, one of the *Hsps*, was defined as an auxin-induced gene that exhibited constant expression throughout carrot SE development (Kitamiya et al., 2000). In soybean and potato, SEs were induced in cotyledons by exogenous auxin and were associated with up-regulation of oxidative stress and defence genes (Sharma et al., 2008; Thibaud-Nissen et al., 2003). Thus, one hypothesis on the mechanism involved in SE development highlighted the interaction between auxin signalling and stress response (Fehér et al., 2003). Pasternak et al. (2005) reported that oxidative stress could enhance auxin responsiveness. Moreover, the MAP kinase pathway might link stress response to auxin signalling in plants (Hirt, 2000). However, further experiments were needed to evaluate the interaction between stress signalling and auxin signalling during SE development.

Embryogenesis is a complex process. Despite the progress achieved during the last few years in understanding the mechanisms involved in SE, there are still many aspects that are not fully understood and need to be studied in more detail. Clearly, molecular markers are required to follow specific events during SE development, with the ultimate goal of unraveling the regulatory networks that operate. The present comparative profiling provided important insights into the mechanisms underlying embryogenesis. The process of somatic embryogenesis was characterized by the induction of many stress-related genes, which leads to the hypothesis that SE is an extreme stress response of cultured plant cells. Based on the relationship between stress and SE development, SE might be an excellent model to study the mechanism of stress resistance in plants. Despite attempts to create suitable cultures for somatic embryogenesis, the artificial culture is an abnormal process for *in vitro* plant cell. Those artificial and extreme conditions might result in stress response in cells that would provide diverse resources for stress tolerance research.

**Experimental procedures**

**Plant material**

Zygotic embryos and SEs of *G. hirsutum* cv (YZ1) were collected at the globular-, torpedo- and cotyledonary stages, as shown in Figure 1. The cotton plants, used for ZE collection, were cultivated in the field under normal farming conditions during growing seasons from 2010 to 2012 in Wuhan, China. ZEs at the three development stages were excised carefully from developing bolls in August of 2010, August 2011 and August of 2012, as three replicate samples. The samples collected in August of 2010 were used for RNA-seq, and all the three replicate samples were used for RT-PCR validation. SEs were sampled from tissue culture on MSB medium using cotton hypocotyls as explants, as previously described (Yang et al., 2012), with at least three replicates. All samples were frozen immediately in liquid nitrogen and stored at –80 °C before use.

**Histocytological analysis**

We carried out a comparative structural investigation on the two types of embryos, and the tiny ZEs were visualized under a light microscope (HAL 100; Zeiss, Oberkochen, Germany) after paraffin sectioning, while SEs were observed under a stereomicroscope (MZ FLIII; Leica, Wetzlar, Germany). Embryo sacs pregnant with ZEs at different stages were fixed in FAA for at least 24 h. After dehydration in an alcohol gradient and infiltration with a chloroform gradient, the samples were embedded in paraffin and sliced into 8-μm-thick sections (Yang et al., 2010). Finally, the sections were observed by light microscopy.

Somatic embryos and ZEs were sampled and fixed with 2.5% glutaraldehyde for TEM analysis to investigate the difference in ultrastructures. The TEM method was described in detail (Hu et al., 2011). Several external layers of cells from the apex of embryos, for example, the cotyledon cells of cotyledon embryos, were cut into ultrathin sections with a ultra-microtome (Leica UC6; Leica) and observed under a light microscope (H-7650; Hitachi, Japan) at 80 KV, with a Gatan 832 CCD camera (Gatan 832, Pleasanton, CA) to capture images.

**RNA extraction, sequencing and raw data processing**

Total RNA of different developmental stages of ZEs (ZGE, ZTE and ZCE) and SEs (SGE, STE and SCE) were isolated using a modified guanidine thiocyanate method (Tu et al., 2007a).

Approximately, 20 μg of total RNA was sent to the Beijing Genomics Institute (Shenzhen, China), where the tag libraries were prepared and sequenced. The sequencing technology was DGE II. Before construction of the library, RNA quality and quantity were determined by an Agilent 2100 bioanalyzer with an rRNA ratio (28s/18s) > 1.5 and an RNA Integrity Number > 7. Then, mRNA was enriched using the oligo (dT) magnetic beads and interrupted by adding the fragmentation buffer to short segments (approximately 200 bp) that were used as templates to synthesize double-strand cDNA tags with random hexamer primer. After purification, washing for end repair and single-nucleotide A (adenine) addition, and attachment the sequencing adapters, the library fragments were enriched by PCR amplification. Millions of raw reads with a length of approximately 50 bp were generated via Illumina HiSeq™ 2000 (Illumina, San Diego, CA). Clean reads occupying >98% of raw reads were produced after filtering dirty raw reads. By using SOAPaligner/soap2 (Li et al., 2011), all clean reads were mapped
to the reference sequences, supporting up to two base mismatches in the alignment (Zenoni et al., 2010). For better expression analysis, transcription levels were calculated in RPKM, which was able to eliminate the disturbance of different gene lengths and sequencing discrepancies on the calculation of gene expression (Mortazavi et al., 2008). Sequence data from this study have been deposited in the National Centre for Biotechnology Information Gene Expression Omnibus under GEO Series accession number GSE45671.

Identification and annotation of DEGs

For differential expression analysis between SEs and ZEs, the fold changes between parallel development stages were assessed by taking the log2 ratio of RPKM. We estimated that the statistical analysis was dependable when the gene expression value RPKM ≥ 25 in at least two of six samples to enhance accuracy (Yang et al., 2012). P-value was calculated as described (Audic and Claverie, 1997), and the false discovery rate (FDR) was adjusted to determine the threshold of P-value in multiple tests. Based on these statistical analyses, the genes with FDR ≤ 0.001 and an absolute value log2 ratio ≥ 1.5849, which was equal to at least a 3.0-fold change in at least one of three log2 ratios, were considered differentially expressed.

A BLAST search was carried out against the TAIR9 protein database of Arabidopsis thaliana (http://www.arabidopsis.org/) and the GenBank nonredundant protein database using the BlastX program (http://blast.ncbi.nlm.nih.gov/Blast.cgi) (E-value ≥ 10^-5) to annotate DEGs. Then, all the DEGs were mapped to GO terms in the database (http://www.geneontology.org/) for GO functional classification analysis. Additionally, Cluster analysis of gene expression patterns was accomplished by Genesis (http://genome.tugraz.at/) based on the K-means method (Sturm et al., 2002).

ABA and NaCl stress treatments

To further investigate the effect of stress treatment on SE formation and development, the homogeneous ECs were cultured on MSB medium, supplemented with different concentrations of ABA and NaCl. ECs were homogenized using a suspension culture procedure modified from an earlier protocol (Cao et al., 2008). Briefly, ECs in suspension were filtered through double-layered stain less steel sieves (30 and 50 mesh) successively, and then, homogeneous calluses over the 50 mesh-size-sieve were inoculated into 90-mm-diameter Petri dishes with 30 mL of solid MSB medium. ABA was added to the autoclaved medium in the following concentrations: 0, 0.2, 1, 5, 10 μM (Garcia-Martin et al., 2005), and NaCl was added prior to autoclaving, as the following treatments: 0, 25, 50, 75, 100 mM (Al-Khayri, 2002). About 0.03 g of ECs was distributed into 32 points in every Petri dishes. After 2 weeks, we sampled the cultures for expression analysis. After 1 month, the number of SEs and the fresh weight of ECs were recorded. All experiments were conducted in triplicate.

RT-PCR and qRT-PCR analysis

Multiple DEGs identified from RNA-seq were selected for validation by reverse transcriptase-polymerase chain reaction (RT-PCR).

Gene-specific primers (Table S4) were designed using Primer Premier 5.0 with a primer length ≥ 25 bp and synthesized commercially (Genscript Bioscience). The cotton UBP7 (GenBank accession number: DQ116441) gene was used as an internal control (Tu et al., 2007b). The PCR products (10 μL) were separated on 1.4% agarose gel containing ethidium bromide. RT-PCR analysis was performed in three biological replicates.

EC treated by NaCl and ABA after 2 weeks were sampled for gene expression analysis by quantitative reverse transcriptase-PCR (qRT-PCR).

Funding from the National High-tech R and D Program of China (863 Program) (No. 2013AA102601-4) and the Fundamental Research Funds for the Central Universities (2010QC034) were appreciated.

Acknowledgements

The authors thank Prof. Keith Lindsey (The Integrative Cell Biology Laboratory, School of Biological and Biomedical Sciences, Durham University, Durham, UK) for critical reading of the manuscript and language corrections. We also thank the three anonymous reviewers for helpful comments that improved this paper. The authors declare that they have no competing financial interests.

References


**Supporting information**

Additional Supporting information may be found in the online version of this article.

**Figure S1** Ultrastructure of somatic and zygotic embryos at three developmental stages.

**Figure S2** Sequence saturation analysis in RNA-seq.

**Figure S3** RT-PCR validation of two biological replicates.

**Figure S4** GO functional classification of not differentially expressed genes on biological process at level 2.

**Table S1** RPKM of 4861 highly but not differentially expressed genes.

**Table S2** RPKM of 4242 differentially expressed genes.

**Table S3** Detailed list of stress response genes in the profiling.

**Table S4** Primers used for RT-PCR.

**Table S5** Primers used for qRT-PCR.