Heliyon 5 (2019) e01868

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.heliyon.com

In silico genome-wide identification and comprehensive characterization of the *BES1* gene family in soybean

Qing Li [a,b], Luqin Guo [c], Hong Wang [a], Yu Zhang [a,c], Chengming Fan [d,**], Yanting Shen [d,*]

[a] College of Life Sciences and Oceanography, Shenzhen University, Shenzhen 518060, China
[b] Key Laboratory of Optoelectronic Devices and Systems of Ministry of Education and Guangdong Province, College of Optoelectronic Engineering, Shenzhen University, Shenzhen 518060, China
[c] College of Horticulture, Henan Agricultural University, Zhengzhou 450002, China
[d] Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China

ARTICLE INFO

Keywords:
Bioinformatics
Plant biology
Molecular biology

ABSTRACT

The *BES1* transcription factor family play a central role in brassinosteroid signaling pathway that regulates a wide range of plant growth and developmental processes, as well as resistances to various stresses. However, no comprehensive study of the *BES1* gene family in soybean has been reported. In this work, 16 *GmBES1*-like genes were identified in soybean, which could be divided into two clades based on their phylogenetic relationships, gene structures and motif compositions. We then examined their duplication status and evolutionary models. The result showed that most of the *GmBES1*-like genes have duplicated counterparts generated from the recent Glycine WGD event, and these genes are originated from 6 distinct ancestors before the Gamma WGT event. We further studied the expression profiles of *GmBES1*-like genes, and found their spatio-temporal and stressed expression patterns varied tremendously. For example, *GmBES1-5* and *GmBES1-6* were highly expressed in almost every sample, whereas *GmBES1-7* and *GmBES1-8* were not expressed. Additionally, interaction network analysis revealed the presence of 3 clusters between *GmBES1*-like genes and other associated genes, implying that they have both the conserved and divergent functions. Lastly, we analyzed the genetic diversity of *GmBES1*-like genes in 302 resequenced wild, landrace and improved soybean accessions. It showed that most of these genes are well conserved, and they are not changed during domestication and improvement. These results provide insights into the characterization of *GmBES1* family and lay the foundation for further functional study of such genes.

1. Introduction

Brassinosteroids (BRs) are a group of plant-specific steroid hormones that play fundamental roles in a broad spectrum of plant growth and developmental processes, including cell elongation and division, photomorphogenesis, stomata formation, vascular differentiation, plant architecture, flowering, male fertility, seed germination, senescence and resistances to various biotic and abiotic stresses [1, 2, 3, 4]. To date, the core BR signaling pathway has been extensively defined, starting with the membrane-localized receptor BRASSINOSTEROID INSENSITIVE 1 (BRI1) and co-receptor BRI1-ASSOCIATED RECEPTOR KINASE 1 (BAK1), and ultimately culminating in the activation of BRI1-EMSSUPPRESSOR (BES1) and BRASSINAZOLE-RESISTANT 1 (BZR1) transcription factors, which direct a transcriptional network controlling the expression of thousands of genes enabling BRs to influence growth and stress programs [5]. BES1 and its homolog BZR1 are two key effectors that positively mediate BR responses. In the presence of BR signal, BES1 and BZR1 are rapidly dephosphorylated and activated by PROTEIN PHOSPHATASE 2A (PP2A) [6]. The activated BES1 and BZR1 translocate into the nucleus and bind to either E-box or BRRE motif of their target genes to directly regulate their expression, and then initiating a series of BR-responsive events [7, 8]. In the absence of BRs, BES1 and BZR1 are phosphorylated and inactivated by the kinase BRASSINOSTEROID-INSENSITIVE 2 (BIN2), which is a pivotal negative regulator of BR signaling [9, 10].

BES1 and BZR1 belong to the *BES1* gene family, which is a new class of plant-specific transcription factors with 6 more members in Arabidopsis according to PlantTFDB 4.0 (http://planttfdb.cbi.pku.edu.cn/)

* Corresponding author.
** Corresponding author.
E-mail addresses: cmfan@genetics.ac.cn (C. Fan), ytshen89@gmail.com (Y. Shen).

https://doi.org/10.1016/j.heliyon.2019.e01868
Received 7 April 2019; Received in revised form 23 May 2019; Accepted 29 May 2019

[11]. Among them, BES1 and BZR1 are more closely related to each other, with 88% sequence identity at the protein level [8]. *BES1* was identified via a gain-of-function mutant *bes1-D* that fully suppresses *bri1* dwarf phenotype and shows constitutive BR responses, including excessive stem elongation, resistance to a BR-biosynthesis inhibitor brassinazole (BRZ), as well as up-regulation of BR-induced gene expression [12]. *BZR1* was identified by a dominant mutant *bzr1-D*, which has a similar phenotype to *bes1-D* in the dark but a semi-dwarf phenotype in the light [13]. As the major transcription factors in BR signaling, BES1 and BZR1 have a huge number of direct target genes. Functional classification of these target genes further revealed that BES1 and BZR1 also regulate numerous molecular links between BR and other pathways, such as the crosstalk between BRs and other hormones, light-signaling, stress responses, as well as some cellular and developmental processes [14, 15]. *BES1*, *BZR1* and four other homologs *BEH1–4* have partially redundant functions in BR signaling [8]. The remaining two homologs BETA-AMYLASE 2 (BMY2) and BETA-AMYLASE 4 (BMY4) function as transcription factors controlling shoot growth and development through the crosstalk with BR signaling, and may compete with BES1/BZR1 for common targets in this process [16].

Presently, most of our understanding about *BES1* gene family are acquired from the model plant Arabidopsis. It is essential to extend *BES1* family-related study to crops, such as soybean which is an important crop for seed protein and oil content. In this study, a comprehensive genome-wide analysis was performed to characterize the *BES1* gene family in soybean. In total, 16 *GmBES1*-like genes were identified and classified according to their features including phylogenetic relationships, gene structures, as well as the conserved protein motifs. We further analyzed their duplication status, evolutionary models, spatio-temporal and stressed expression patterns, co-functional network and genetic variations. The obtained results provide valuable resources for the future research on the functions of *GmBES1*-like genes, as well as the BR signaling pathway in soybean. Moreover, this study may also contribute to our knowledge of the polyploid effect during the evolution of the *BES1* gene family in soybean.

2. Materials and methods

2.1. Identification of GmBES1-like genes

The Arabidopsis AtBES1/AtBZR1 protein sequence was used as query to search against the soybean genome in Phytozome (https://phytozome.jgi.doe.gov/pz/portal.html). The Pfam tool (http://pfam.xfam.org/) was further used to identify the retrieved GmBES1-like candidates with the conserved BES1_N domain [17]. The molecular weights and isoelectric points of BES1-like proteins were analyzed using Protparam (https://web.expasy.org/protparam/). The subcellular locations were predicted by Plant-mPLoc (http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/) [18].

2.2. Phylogenetic analysis of BES1-like members

A maximum likelihood (ML) phylogenetic tree was constructed using MEGA 6 with 200 bootstrap replications and an optimal JTT + G + I model [19]. The gene structures of the *BES1*-likes were acquired from their genomic DNA information. MEME (http://meme-suite.org/) [20] was used to identify the 8 conserved motifs of BES1-like proteins.

2.3. Collinearity analysis of the GmBES1-like genes

MCScanX toolkit was used to identify collinear regions containing *GmBES1*-like genes in soybean [21]. The collinearity relationships of *GmBES1*-like genes were drawn using Circos [22]. The divergence time of the homologous blocks were evaluated to investigate the evolution of *GmBES1*-like genes in soybean according to the synonymous substitution rate (Ks). If the Ks $>$ 1.5, the divergence time is after the Gamma whole-genome triplication (WGT) event; If the Ks $<$ 0.3, the divergence time is after the Glycine whole-genome duplication (WGD) event; and when the Ks is between 0.3 and 1.5, the divergence time is after the Legume WGD event but before the Glycine WGD event [23].

2.4. RNA sequence and data analysis

Our previously reported Illumina RNA-seq data for 28 samples were used for *GmBES1*-like genes spatio-temporal expression analysis [24]. The reads were mapped to the soybean reference genome (Wm82. a2.v1) using HISAT [25]. Transcripts assembly and their expression calculations were achieved utilizing StringTie [26]. The gene expression values were counted by fragments per kilobase of exon per million fragments mapped (FPKM). To study the expression patterns of *GmBES1*-like genes against abiotic stresses, the previously released Illumina RNA-seq data for dehydration and salt stresses were explored [27, 28]. The gene expression values were also counted by FPKM. Differential expression analyses were performed by comparing the expression of a gene at each time point to 0 hour. The heat maps with or without hierarchical clustering were visualized using Heml [29].

2.5. Establishment of gene co-functional network

Genes co-functioned with each *GmBES1*-like gene were accessed by SoyNet (www.inetbio.org/soynet) with a "find new members of a pathway" based method [30]. The generated file in sif format (.sif) containing closely connected genes to each *GmBES1*-like gene was downloaded. To construct the co-functional network, the sif files for all *GmBES1*-like genes were merged together and visualized with the Cytoscape tool [31].

2.6. SNP genotyping

The *GmBES1*-like genes in 302 accessions were genotyped using our released whole-genome resequencing data [32]. Read mapping and SNP calling were performed as a previously described method [33]. The genomic region was consisted of 5′UTR, exon, intron and 3′UTR according to the genome annotation. The SNPs were divided into non-synonymous SNPs (cause amino acid changes), synonymous SNPs (do not cause amino acid changes), alternative spliced SNPs (within 2 bp of a splicing junction) and premature SNPs (create a stop codon).

3. Results and discussion

3.1. Genome-wide identification of GmBES1-like genes in soybean

To identify the entire *GmBES1*-like genes, we used the Arabidopsis AtBES1/AtBZR1 protein sequence to perform a genome-wide search in soybean. Subsequently, the retrieved sequences were subjected to Pfam analyses (Fig. S1). By removing the alternative splice forms of the same gene, a total of 16 GmBES1-like members with the typical BES1_N domain were identified in soybean, being just twice the number of *BES1* genes in diploid Arabidopsis. This is not an accident considering that soybean is a palaeo-tetraploid plant, in which most genes are present in multiple copies due to the two WGD events [34]. To simplify, they were named as *GmBES1-1* to *GmBES1-16* in order (Table 1). It is worthy to note that the Glyco_hydro_14 domain was only found in GmBES1-13 to GmBES1-16 as well as AtBMY2 and AtBMY4 (Fig. S1). The 16 *GmBES1*-like genes were mapped onto the 11 chromosomes of soybean with a non-random distribution (Table 1). Their protein sizes ranged from 178 to 705 amino acids, the molecular weights extended from 19784.81 to 78842.49 Da, and the predicted isoelectric points changed from 5.16 to 9.84 (Table 1). These proteins varied considerably from 14.9% to 97.5% in sequence identity, although they shared the conserved N-terminal domain (Figs. S2, S3). All GmBES1-like proteins were predicted to be located in the nucleus in accordance with their function as transcription factors (Table 1).

Table 1
The basic information of *GmBES1*-like genes.

			Protein			
Gene name	Gene ID	Gene location	Length/aa	MW/Da	pI	Localization
GmBES1-1	Glyma.17G248900.1	Chr17:40417445–40419968 +	311	33986.19	9.22	nucleus
GmBES1-2	Glyma.06G034000.1	Chr06:2613465–2615567 -	320	34892.97	9.36	nucleus
GmBES1-3	Glyma.14G076900.1	Chr14:6448729–6452256 -	311	33929.09	8.95	nucleus
GmBES1-4	Glyma.04G033800.1	Chr04:2687701–2689839 -	320	34911.97	9.34	nucleus
GmBES1-5	Glyma.01G178000.1	Chr01:51435454–51437016 -	308	33284.31	9.37	nucleus
GmBES1-6	Glyma.11G064300.1	Chr11:4849799–4851673 +	310	33564.63	9.37	nucleus
GmBES1-7	Glyma.07G099100.1	Chr07:9347781–9349110 +	178	19784.81	9.84	nucleus
GmBES1-8	Glyma.14G127400.1	Chr14:20399619–20401627 -	299	33535.12	9.84	nucleus
GmBES1-9	Glyma.13G266600.1	Chr13:36972604–36976752 -	334	35522.69	8.99	nucleus
GmBES1-10	Glyma.12G231400.1	Chr12:39138374–39142508 +	334	35637.81	9.14	nucleus
GmBES1-11	Glyma.13G266500.1	Chr13:36960188–36965286 -	325	35132.12	8.49	nucleus
GmBES1-12	Glyma.12G231500.1	Chr12:39151416–39156034 +	322	34796.89	8.83	nucleus
GmBES1-13	Glyma.09G217000.1	Chr09:44013762–44019811 +	654	74123.42	5.87	nucleus
GmBES1-14	Glyma.01G151800.1	Chr01:48859327–48866275 +	656	74399.77	5.76	nucleus
GmBES1-15	Glyma.09G260100.1	Chr09:47841348–47848090 +	705	78842.49	5.16	nucleus
GmBES1-16	Glyma.18G232400.1	Chr18:52065790–52072872 -	704	78838.52	5.31	nucleus

3.2. Phylogenetic and classification analysis of GmBES1-like genes

To conduct the classification of *GmBES1*-like genes, we constructed a phylogenetic tree using the soybean and Arabidopsis BES1 family protein sequences (Fig. 1A). We found that the 16 *GmBES1*-like genes were clearly grouped into 2 clades (designated A and B) according to the bootstrap values and phylogenetic topology. Clade A included 12 members (*GmBES1-1* to *GmBES1-12*), and it was further divided into 2 distinct subgroups (designated A1 and A2). Subgroup A2 contained *GmBES1-7* and *GmBES1-8*, Subgroup A1 included others of clade A. Clade B had 4 members (*GmBES1-13* to *GmBES1-16*).

Previous research has shown that the exon/intron diversification among gene family members plays an important role in the evolution of multiple gene families mainly through three mechanisms: exon/intron gain/loss, exonisation/pseudoexonisation, and insertion/deletion [35]. Thus, the number and positions of exons and introns in *BES1*-like genes were determined by comparing full-length cDNA sequences and the corresponding genomic DNA sequences of each *BES1*-like gene (Fig. 1B). We found that all *BES1*-like genes contained multiple exons. Moreover, most members shared almost similar exon/intron structures and intron phases in the same clade or subgroup. In detail, the *BES1*-like genes in subgroup A1 and subgroup A2 had 2-3 and 5 exons respectively, whereas 10 exons were included in the *BES1*-like genes of clade B. There are much more exons in the *BES1*-like genes of clade B than those in clade A, implying that there may be a functional divergence among the *BES1*-like genes.

Proteins, which come from the same subfamily and share similar motif compositions, are likely to share similar functions [20]. Therefore, the 8 conserved motifs were predicted by the MEME program among the soybean and Arabidopsis BES1-like proteins (Fig. 1C, Fig. S4). As expected, a common motif composition of the closely related members in the phylogenetic tree was revealed, suggesting functional similarities among the BES1-like proteins within the same clade or subgroup. The result showed that the motifs 1, 3, 4 were shared by all BES1-like members. Moreover, each class of BES1-like proteins had their own motifs, such as motif 2 and 5, 6, 7 only existed in subgroup A1 and clade B respectively.

Comparative analysis of homologs between Arabidopsis and soybean may be helpful for the understanding of homologous gene functions in soybean. Our results showed that GmBES1-1 to 4, AtBES1 and AtBZR1 are clustered into a branch in the phylogenetic tree (Fig. 1), implicating that these proteins might be function as AtBES1/AtBZR1. A recent study

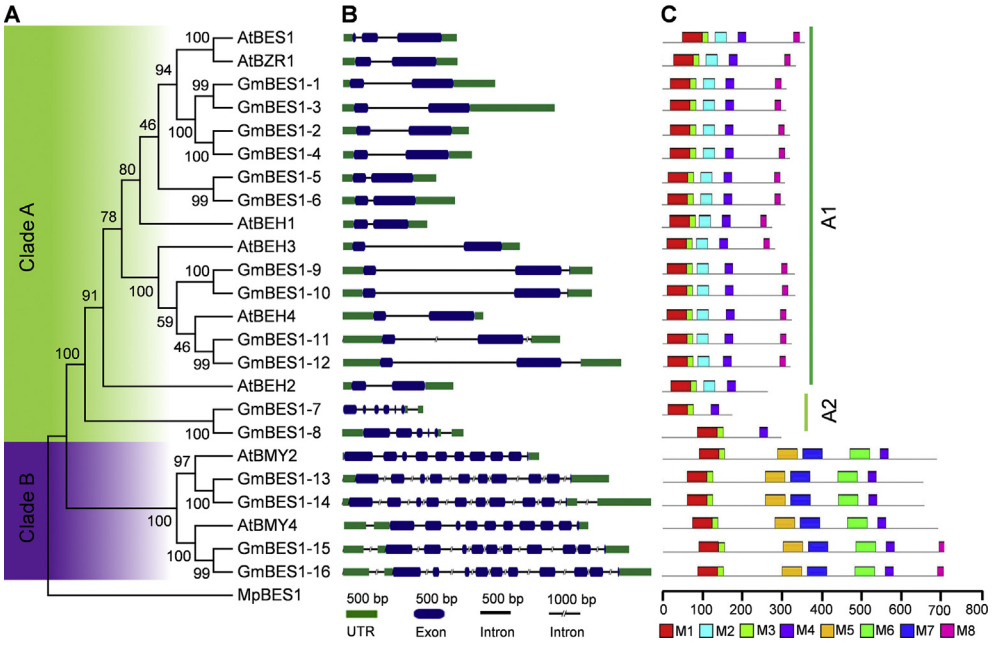

Fig. 1. Phylogenetic relationships, gene structures and conserved motifs of *BES1*-like genes in Arabidopsis and soybean. (A) The maximum likelihood tree of BES1-like proteins. The *Marchantia polymorpha* BES1-like (Mapoly0013s0054.1) protein was used as an outgroup. Clades A and B were shown with pale-green and purple, respectively. Subgroups A1 and A2 were indicated on the left side of the figure. (B) The structures of *BES1*-like genes. (C) The distribution of conserved motifs of BES1-like proteins. Different motifs were indicated by different colored boxes numbered M1–M8.

supports this speculation. The functional characterization of a soybean *BZR1*-like gene *GmBZL2* (here renamed as *GmBES1-3*) revealed that GmBZL2 are highly conserved with Arabidopsis BZR1 in the BR signaling pathway [36]. GmBES1-5 and GmBES1-6 also get together with AtBES1/AtBZR1 but form a larger group (Fig. 1), implying that they likely have other roles or partially redundant functions with AtBES1/AtBZR1. Consistent with this, GmBEHL1 (here termed as GmBES1-5), an ortholog of Arabidopsis BEH1, were demonstrated as a co-repressor to negatively regulate soybean nodulation, which is the first study to establish a direct link between the nodulation pathway and BR signaling [37]. Similarly, the remaining GmBES1-like homologs in clade A and GmBES1-13 to 16 in clade B of the phylogenetic tree might be have the functions as AtBEHs and AtBMY2/AtBMY4, respectively (Fig. 1).

3.3. Duplication status and evolutionary models of the GmBES1-like genes

Soybean is a paleopolyploid plant that has experienced the Gamma WGT (130–240 million years ago), Legume WGD (59 million years ago) and Glycine WGD (13 million years ago) events [23], resulting in a highly duplicated soybean genome with about 75% of the genes present in multiple copies [34]. The presence of duplicated gene pairs could give more opportunities for gene evolution [38, 39]. Thus, it would be valuable to determine the duplication status of *GmBES1*-like genes. Based on our collinearity analysis, we found that all the *GmBES1*-like genes, except for *GmBES1-7*, *GmBES1-8*, *GmBES1-9* and *GmBES1-12*, have duplicated counterparts generated from the Glycine WGD event (Fig. 2A). Moreover, although the duplicate copy of *GmBES1-7* was lost in the syntenic block of chromosome 9, the block containing *GmBES1-7* experienced Glycine WGD event (Fig. 2A). And the similar situation was happened to *GmBES1-8* (Fig. 2A). Besides the WGD duplication, *GmBES1-9* and *GmBES1-12* may experience tandem duplication event with *GmBES1-11* and *GmBES1-10* respectively, since they are adjacent in the same chromosome (Fig. 2A).

WGD, or polyploidy, is a common phenomenon in nature, particularly in plants [40, 41]. The duplicated genes generated from WGD provide important raw genetic material for adaptive evolution [42], including subfunctionalization via purifying selections (Ka/Ks $<$ 1) [43], neofunctionalization through positive selections (Ka/Ks $>$ 1) [44], and nonfunctionalization [38, 45]. The Ka/Ks values for 6 duplicated *GmBES1*-like gene pairs were less than 0.3 (Table S1), indicating that they are highly purifying selected and subfunctionalized in the soybean evolution process. Our results support the conclusion that most of duplicated genes in soybean are subfunctionalized, and only a small proportion of the duplicated genes have been neofunctionalized or nonfunctionalized [46].

Based on the average Ks values of the collinear blocks, the evolution history of *GmBES1*-like genes was predicted. These *GmBES1*-like genes were classified into 6 subgroups, *GmBES1-A* to *F* (Fig. 2B). For example, *GmBES1-A* firstly generated 3 copies after the Gamma WGT event, followed one copy was lost, the retained two copies were further doubled after the Legume WGD event, and then one of the 4 copies was lost too, the remaining 3 copies were duplicated after the Glycine WGD event at last, resulting in a production of 6 genes (*GmBES1-1* to *GmBES1-6*). In contrast, *GmBES1-B*, *GmBES1-C* and *GmBES1-D* generated two *GmBES1*-like genes in the same way after the 3 rounds of the evolution processes. Somewhat similar to *GmBES1-B* to *D*, *GmBES1-E* and *GmBES1-F* both produced two collinear blocks, although in which only one *GmBES1*-like gene was reserved, and the other *GmBES1*-like gene was lost. These results indicated that the *GmBES1* gene family originate from 6 distinct ancestors, at least prior to the Gamma WGT event.

3.4. The spatio-temporal and stressed expression profiles of GmBES1-like genes

As subfunctionalization through differential gene expression is a relatively rapid process compared to the classical model of subfunctionalization via mutations [47], a large proportion of duplicated genes in soybean exhibit subfunctionalization at the expression level [46]. To confirm this conclusion, we explored the spatio-temporal expression profiles of *GmBES1*-like genes in 28 soybean samples using our previously reported Illumina RNA-seq data [24]. As a result, these *GmBES1* duplicated genes, except *GmBES1-5* and *GmBES1-6*, showed differential expression in 28 samples, implying that they underwent expression subfunctionalization (Fig. 3A). On the whole, the expression levels and patterns of these genes varied considerably (Fig. 3A). For example, *GmBES1-5*, *GmBES1-6* and *GmBES1-9* exhibited constantly high expression in almost every sample. In contrast, *GmBES1-3*, *GmBES1-13* and *GmBES1-15* showed significantly lower expression in most tissues. Moreover, *GmBES1-7*, *GmBES1-8* and *GmBES1-14* were not detected in all samples due to their extremely low values, implying that they are the potential pseudogenes. The differential expression patterns of these *GmBES1*-like genes could provide important clues for their function, implicating that they possibly play diverse roles in regulation of soybean growth and development.

BRs have long been recognized as an important phytohormone involved in a wide range of stress responses. The *BES1* gene family are emphasized for their key roles in mediating plants with tolerances to drought and high-salt stresses [4, 48]. To study the potential roles of *GmBES1*-like genes under these stresses, we investigated their expression in response to treatment with dehydration or salt (NaCl) based on the

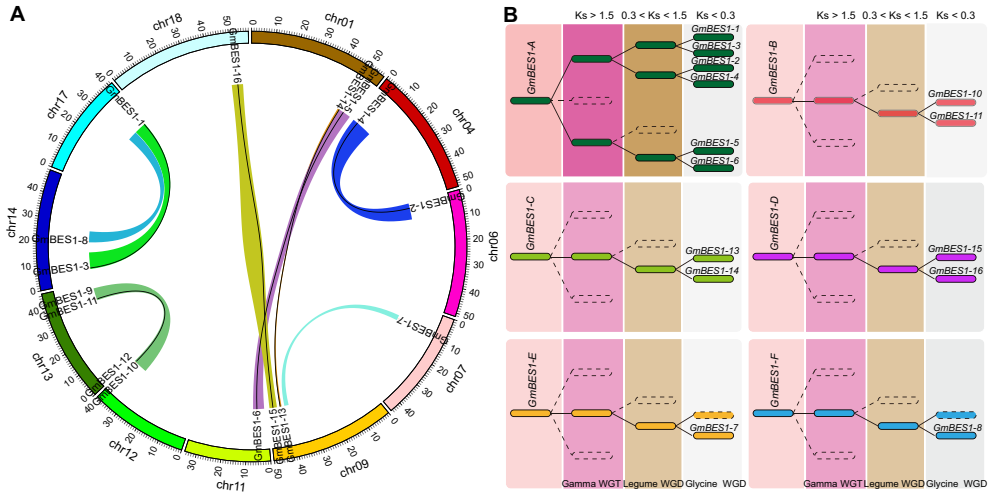

Fig. 2. The duplication status and evolutionary models of the *GmBES1*-like genes. (A) The collinear blocks containing *GmBES1*-like genes after the Glycine WGD event. The colored rainbows showed the collinear blocks, and the lines within these blocks displayed the location of *GmBES1*-like genes. The positions of *GmBES1-9* and *GmBES1-11* were hard to tell apart because they are adjacent in the same chromosome. Similar status was observed for *GmBES1-10* and *GmBES1-12*. (B) The deduced evolutionary models for *GmBES1*-like genes in the evolution process of soybean genome. The reserved and lost blocks in the corresponding evolution process were displayed by the solid and hollow blocks, respectively. The block lost *GmBES1*-like gene but with retained other genes was showed by the solid block with dotted outline.

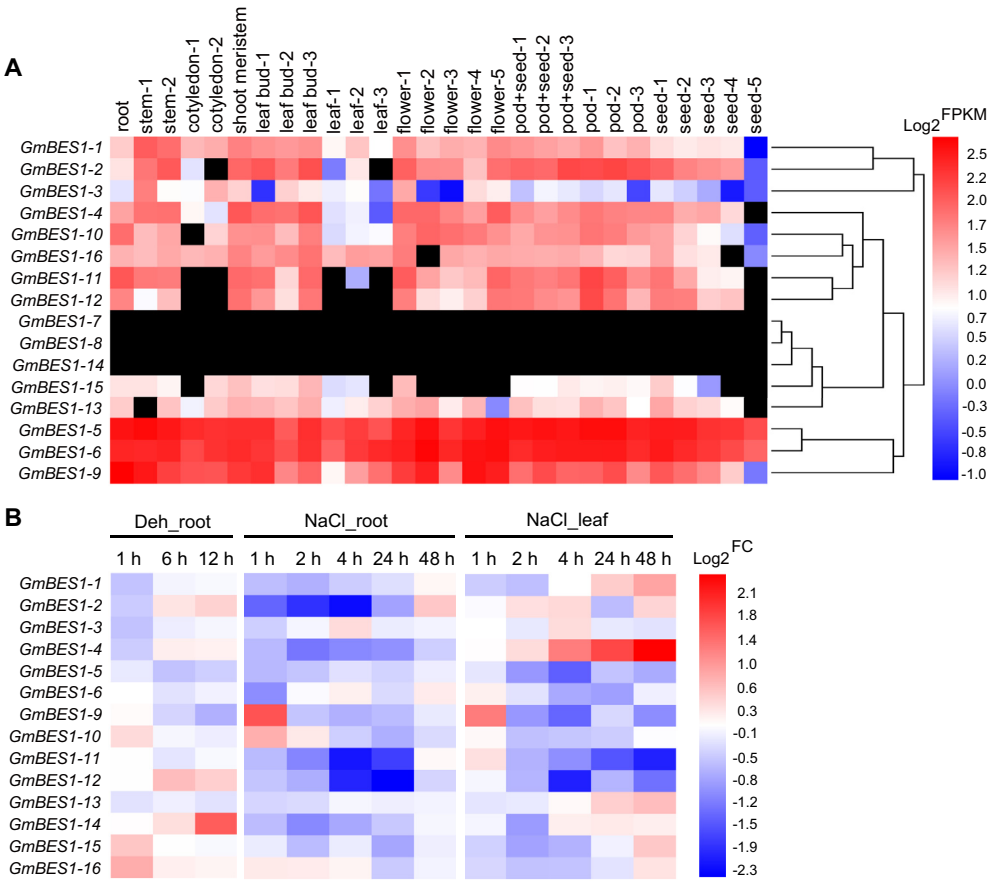

Fig. 3. The spatio-temporal (A) and stressed (B) expression patterns of *GmBES1*-like genes in soybean. The size of the number represented earlier to later developmental stages within the same tissue/organ. Gradient colors indicated log2 transformed FPKM values in different samples or fold-change (FC) values in gene expression at different treatment time points compared with control (0 h). The black color represented that the FPKM value is 0.

previously released Illumina RNA-seq data [27, 28]. The result provided a preliminary impression of the expression variations of these genes except *GmBES1-7* and *GmBES1-8*, which were not detected in almost every sample (Fig. 3B). Under dehydration stress, the expression of *GmBES1-12* was slightly up-regulated by 1.63-fold at 6 h and 1.46-fold at 12 h compared with control (0 h) in root. And the transcripts of *GmBES1-15* and *GmBES1-16* were also moderately increased but earlier by 1.54-fold and 1.82-fold at 1 h, respectively. Moreover, *GmBES1-14* was significantly induced more than 3-fold in response to 12 h dehydration stress. However, *GmBES1-9* was slightly down-regulated by 1.56-fold at 12 h. Under salt stress, the expression patterns of *GmBES1*-like genes were much more complicated, and most of these genes were up or down-regulated at least one time point compared with control in root and leaf. For example, *GmBES1-4, GmBES1-5, GmBES1-11*, *GmBES1-12* and *GmBES1-14* were decreased at almost every time point of NaCl treatment in root. And similar expression patterns for these genes were observed in leaf except for *GmBES1-4* and *GmBES1-14*. In contrast, *GmBES1-4* was gradually induced with the extension of salt stress time in leaf. Meanwhile, *GmBES1-14* was only slightly down-regulated at 2 h of salt stress in leaf. Interestingly, *GmBES1-9* was obviously increased at 1 h, but subsequently declined after prolonged time of salt stress both in root and leaf.

The transcription rate of a gene is determined by trans-acting transcription factors that bind to cis-acting regulatory elements in promoters, additional co-factors, and chromatin accessibility [49]. The analysis of the promoter regions of *BES1*-like genes from Arabidopsis and soybean revealed the presence of a variety of cis-acting elements, which were classified into four major classes: development-, hormone-, stress-, and light responsiveness-related cis-acting elements (Fig. S5). It is not surprising since *BES1* family act as an integration hub in multi-signal regulated growth and development, as well as various environment responses [4]. A comparison of the cis-acting elements among these *BES1*-like genes displayed both a similar and specific distribution of these elements, which may affect the responsiveness of *BES1*-like genes to the environment and development (Fig. S5). Considering all these factors together with the *BES1*-like gene expression patterns in the model plant Arabidopsis could provide more valuable clues for the functional study of these transcription factors in soybean. Alternatively, these *BES1*-like genes may require a post-transcriptional regulation mechanism, since *AtBES1* and *AtBZR1* were not transcriptionally regulated in response to BRs [9, 12]. Furthermore, the protein levels of AtBES1 and AtBZR1 are accumulated by BRs treatment through protein dephosphorylation [4].

3.5. Co-functional network between GmBES1-like genes and other genes

Co-functional networks are useful for identifying genes that are involved in a particular pathway or phenotype [30, 50]. To understand the interactions of *GmBES1*-like genes with their neighbors and to get more insights into their functions, a gene correlation network was constructed using SoyNet, which is a functional gene network and co-expression network database for soybean [30]. There were 1680 interacting gene pairs between *GmBES1*-like genes and other genes in soybean (Fig. 4), suggesting that GmBES1-like proteins are widely involved in the soybean metabolic network and regulated diverse biological processes and pathways. Interacting genes were classified into 3 major gene ontology classes using GO annotations: biological process, cellular component and molecular function (Table S2). Notably, the co-functional network was divided into 3 clusters (Fig. 4). In the largest cluster I, a total of 1581 of interacting gene pairs were found among *GmBES1*-like genes from clade A1 (not include *GmBES1-5* and

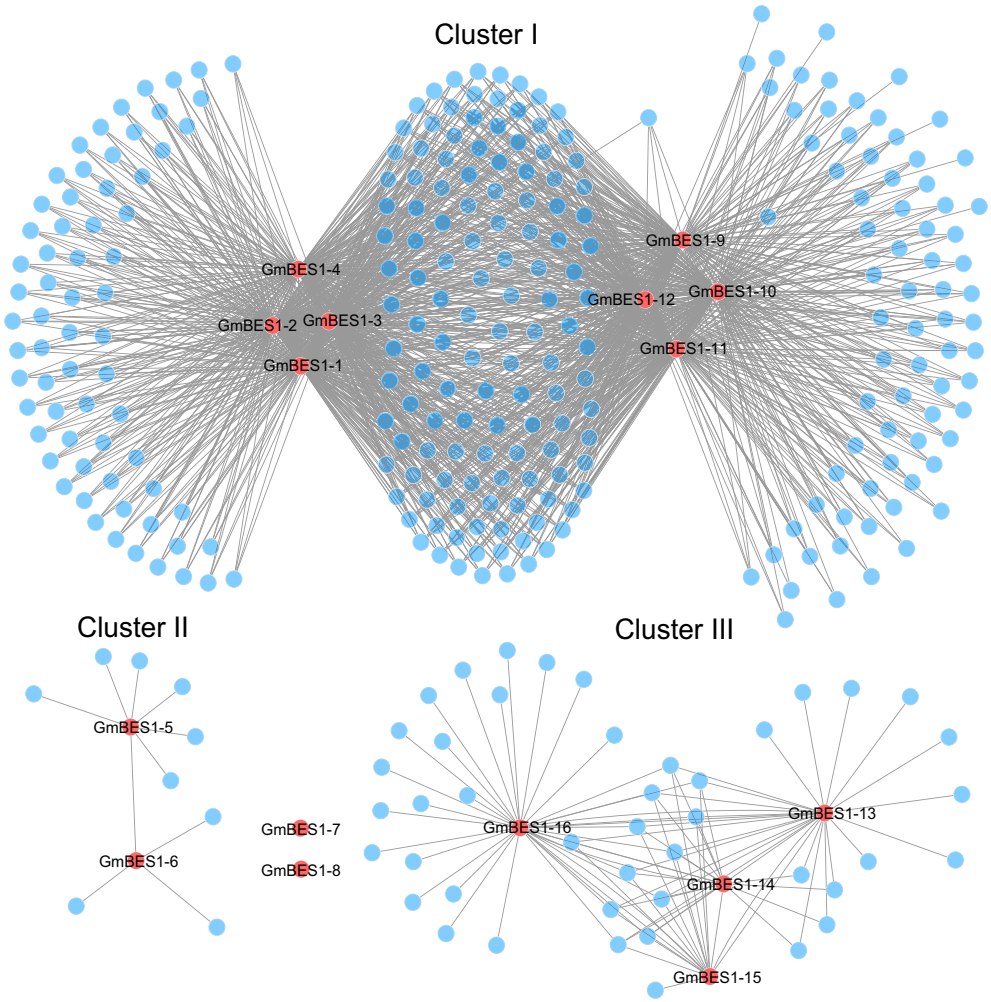

Fig. 4. The co-functional network of *GmBES1*-like genes in soybean.

GmBES1-6) and 270 other genes in soybean. These associated genes are involved in hormone signaling, stress responses, transcription regulation, protein modification, growth and development, and other metabolic processes in accordance with the functions of *GmBES1*-like genes in clade A1 as the master regulators. In addition to share the common interacting genes, *GmBES1-1* to *4* and *GmBES1-9* to *12* also had their own unique co-functional genes, implying that they have both the conserved and divergent functions. *GmBES1-5* and *GmBES1-6* formed the cluster II together with 9 other genes, suggesting that they may not have the roles as *AtBES1/AtBZR1*. In the cluster III, *GmBES1-13* to *16* from clade B were grouped with 44 other genes, many of which encode beta-amylases and modification proteases to regulate carbohydrate metabolism and cellular processes, suggesting that these *GmBES1* genes may function as *AtBMY2/AtBMY4*. Moreover, *GmBES1-16* had more of its specific associated genes, implicating that it has the functions other than *AtBMY2/AtBMY4*. It is also worthy to note that no co-functional gene of *GmBES1-7* and *GmBES1-8* was found, which is consistent with the prediction that they are pseudogenes. Taken together, the obtained results from the co-functional network analysis further support the above conclusions.

3.6. The genetic diversity of GmBES1-like genes in 302 resequenced soybean accessions

To investigate the allelic variation of *GmBES1*-like genes, we explored them in 302 resequenced soybean accessions from our previous study [32]. The result showed that these genes are very conserved except for *GmBES1-7*, *GmBES1-8* and *GmBES1-14* (Table 2). Because no or only few nonsynonymous SNPs at the conserved site were found, although the existences of some SNPs in them (Table 2). In contrast, *GmBES1-7*, *GmBES1-8* and *GmBES1-14* are not conserved due to the high number of nonsynonymous and alternative spliced SNPs in these genes (Table 2). This result is consistent with the fact that they are potential pseudogenes.

The cultivated soybean was domesticated from *Glycine. soja* (wild soybean) in China 5000 years ago [32]. The identification of genes contributing to domestication and improvement of crops could benefit for breeding elite varieties [32]. To identify potential selective signals during soybean domestication (wild soybeans versus landraces) and soybean improvement (landraces versus improved cultivars), we analyzed the SNP distribution frequency of these genes in 62 wild soybeans, 130 landraces and 110 improved cultivars. A total of 11 domestication-selective and one improvement-selective nonsynonymous SNPs were detected (Table S3). Surprisingly, most of the domestication-selective SNPs were found in *GmBES1-7*, *GmBES1-8* and *GmBES1-14*, considering that they may be non-functional in soybean. The remaining two domestication-selective SNPs were distributed in *GmBES1-9* and *GmBES1-10*. The only one improvement-selective nonsynonymous SNP was found in *GmBES1-6*. However, none of the 3 nonsynonymous SNPs was found at conserved site. The highly redundant functionality of these genes could explain why most *GmBES1*-like genes are neither domesticated nor improved.

Table 2
SNP summary of *GmBES1*-like genes within 302 resequenced soybean accessions.

Gene	Total SNP	SNP/kb[a]	Ns SNP[b]	Ns SNP/kb[c]	Ns SNP at conserved site	As SNP[d]
GmBES1-1	29	11.5	1	1.1	0	0
GmBES1-2	9	4.3	1	1	1 in clade A	0
GmBES1-3	26	7.4	3	3.2	0	0
GmBES1-4	6	2.8	1	1	0	0
GmBES1-5	6	3.8	1	1.1	0	0
GmBES1-6	13	6.9	1	1.1	0	0
GmBES1-7	32	24.1	10	18.6	1 in clade A	2
GmBES1-8	30	14.9	10	11.1	1 in clades A and B	1
GmBES1-9	29	7	1	1	0	0
GmBES1-10	36	8.7	2	2	0	0
GmBES1-11	44	8.6	1	1	0	0
GmBES1-12	38	8.2	3	3.1	0	0
GmBES1-13	28	4.6	2	1	0	0
GmBES1-14	54	7.8	13	6.6	1 in clades A and B; 6 in clade B	1
GmBES1-15	36	5.3	1	0.5	0	0
GmBES1-16	103	14.5	3	1.4	1 in clade B	0

[a] The mean number of SNPs per kb DNA sequence.
[b] The nonsynonymous SNP.
[c] The average number of nonsynonymous SNPs per kb CDS sequence.
[d] The alternative spliced SNP.

4. Conclusion

BES1 transcription factor family not only play a key role in the BR signaling pathway but also act as a hub that integrates diverse signals to regulate plant development and environment adaptability by activating or repressing thousands of genes including BR biosynthetic genes via a feedback loop [4]. Although the molecular mechanisms of these genes are well elucidated in Arabidopsis, the regulation mechanisms in crops, such as soybean, are remain poorly understood, and no genome-wide in-depth study of the *BES1* family in soybean has been previously reported. Thus, identification of *BES1* gene family members would be essential to elucidate the BR transcriptional networks in soybean. In this study, a systematic analysis was executed to study the *BES1* genes in soybean genome. A total of 16 *GmBES1*-like genes were identified in the soybean genome. We comprehensive analyzed their basic physical and chemical properties, phylogenetic relationships, gene structures, motif compositions, duplication status, evolutionary models, spatio-temporal and stressed gene expression patterns, co-functional network and genetic diversity. These results contribute for the further study on the functions of such *BES1*-like genes in soybean.

Declarations

Author contribution statement

Qing Li: Conceived and designed the experiments; Contributed reagents, materials, analysis tools or data; Analyzed and interpreted the data; Wrote the paper.

Luqin Guo, Hong Wang, Yu Zhang: Contributed reagents, materials, analysis tools or data.

Chengming Fan, Yanting Shen: Conceived and designed the experiments; Analyzed and interpreted the data.

Funding statement

This work was supported by China Postdoctoral Science Foundation (Grant No. 2017M622751), the PhD Start-up Fund of Guangdong Natural Science Foundation (Grant No. 2018A030310455) and Shenzhen Peacock Innovation Team Plan (Grant No. KQTD20170327151659 26).

Competing interest statement

The authors declare no conflict of interest.

Additional information

Supplementary content related to this article has been published online at https://doi.org/10.1016/j.heliyon.2019.e01868.

References

[1] H. Guo, L. Li, M. Aluru, et al., Mechanisms and networks for brassinosteroid regulated gene expression, Curr. Opin. Plant Biol. 16 (2013) 545–553.
[2] F. Nawaz, M. Naeem, B. Zulfiqar, et al., Understanding brassinosteroid-regulated mechanisms to improve stress tolerance in plants: a critical review, Environ. Sci. Pollut. Res. Int. 24 (2017) 15959–15975.
[3] S. Saini, I. Sharma, P.K. Pati, Versatile roles of brassinosteroid in plants in the context of its homoeostasis, signaling and crosstalks, Front. Plant Sci. 6 (2015) 950.
[4] Q.F. Li, J. Lu, J.W. Yu, et al., The brassinosteroid-regulated transcription factors BZR1/BES1 function as a coordinator in multisignal-regulated plant growth, Biochim. Biophys. Acta 1861 (2018) 561–571.
[5] T. Nolan, J. Chen, Y. Yin, Cross-talk of Brassinosteroid signaling in controlling growth and stress responses, Biochem. J. 474 (2017) 2641–2661.
[6] W. Tang, M. Yuan, R. Wang, et al., PP2A activates brassinosteroid-responsive gene expression and plant growth by dephosphorylating BZR1, Nat. Cell Biol. 13 (2011) 124–131.
[7] J.X. He, J.M. Gendron, Y. Sun, et al., BZR1 is a transcriptional repressor with dual roles in brassinosteroid homeostasis and growth responses, Science 307 (2005) 1634–1638.
[8] Y. Yin, D. Vafeados, Y. Tao, et al., A new class of transcription factors mediates brassinosteroid-regulated gene expression in Arabidopsis, Cell 120 (2005) 249–259.
[9] J.X. He, J.M. Gendron, Y.L. Yang, et al., The GSK3-like kinase BIN2 phosphorylates and destabilizes BZR1, a positive regulator of the brassinosteroid signaling pathway in Arabidopsis, Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 10185–10190.
[10] J.M. Li, K.H. Nam, D. Vafeados, et al., BIN2, a new brassinosteroid-insensitive locus in Arabidopsis, Plant Physiol. 127 (2001) 14–22.
[11] J. Jin, F. Tian, D.C. Yang, et al., PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants, Nucleic Acids Res. 45 (2017) D1040–D1045.
[12] Y.H. Yin, Z.Y. Wang, S. Mora-Garcia, et al., BES1 accumulates in the nucleus in response to brassinosteroids to regulate gene expression and promote stem elongation, Cell 109 (2002) 181–191.
[13] Z.Y. Wang, T. Nakano, J. Gendron, et al., Nuclear-localized BZR1 mediates brassinosteroid-induced growth and feedback suppression of brassinosteroid biosynthesis, Dev. Cell 2 (2002) 505–513.
[14] Y. Sun, X.Y. Fan, D.M. Cao, et al., Integration of brassinosteroid signal transduction with the transcription network for plant growth regulation in Arabidopsis, Dev. Cell 19 (2010) 765–777.
[15] X. Yu, L. Li, J. Zola, et al., A brassinosteroid transcriptional network revealed by genome-wide identification of BESI target genes in Arabidopsis thaliana, Plant J. 65 (2011) 634–646.
[16] H. Reinhold, S. Soyk, K. Simkova, et al., Beta-amylase-like proteins function as transcription factors in Arabidopsis, controlling shoot growth and development, Plant Cell 23 (2011) 1391–1403.
[17] R.D. Finn, P. Coggill, R.Y. Eberhardt, et al., The Pfam protein families database: towards a more sustainable future, Nucleic Acids Res. 44 (2016) D279–285.
[18] K.C. Chou, H.B. Shen, Cell-PLoc: a package of Web servers for predicting subcellular localization of proteins in various organisms, Nat. Protoc. 3 (2008) 153–162.
[19] K. Tamura, G. Stecher, D. Peterson, et al., MEGA6: molecular evolutionary genetics analysis version 6.0, Mol. Biol. Evol. 30 (2013) 2725–2729.
[20] T.L. Bailey, M. Boden, F.A. Buske, et al., MEME SUITE: tools for motif discovery and searching, Nucleic Acids Res. 37 (2009) W202–208.
[21] Y. Wang, H. Tang, J.D. Debarry, et al., MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity, Nucleic Acids Res. 40 (2012) e49.
[22] M. Krzywinski, J. Schein, I. Birol, et al., Circos: an information aesthetic for comparative genomics, Genome Res. 19 (2009) 1639–1645.
[23] A.J. Severin, S.B. Cannon, et al., Changes in twelve homoeologous genomic regions in soybean following three rounds of polyploidy, Plant Cell 23 (2011) 3129–3136.
[24] Y. Shen, Z. Zhou, Z. Wang, et al., Global dissection of alternative splicing in paleopolyploid soybean, Plant Cell 26 (2014) 996–1008.
[25] D. Kim, B. Landmead, S.L. Salzberg, HISAT: a fast spliced aligner with low memory requirements, Nat. Methods 12 (2015) 357–360.
[26] M. Pertea, G.M. Pertea, et al., StringTie enables improved reconstruction of a transcriptome from RNA-seq reads, Nat. Biotechnol. 33 (2015) 290–295.
[27] V. Belamkar, N.T. Weeks, A.K. Bharti, et al., Comprehensive characterization and RNA-Seq profiling of the HD-Zip transcription factor family in soybean (Glycine max) during dehydration and salt stress, BMC Genomics 15 (2014) 950.
[28] A. Liu, Z. Xiao, M.W. Li, et al., Transcriptomic reprogramming in soybean seedlings under salt stress, Plant Cell Environ. 42 (2019) 98–114.
[29] W. Deng, Y. Wang, Z. Liu, et al., HemI: a toolkit for illustrating heatmaps, PLoS One 9 (2014), e111988.

[30] E. Kim, S. Hwang, I. Lee, SoyNet: a database of co-functional networks for soybean Glycine max, Nucleic Acids Res. 45 (2017) D1082–D1089.
[31] M. Franz, C.T. Lopes, G. Huck, et al., Cytoscape.js: a graph theory library for visualisation and analysis, Bioinformatics 32 (2016) 309–311.
[32] Z. Zhou, Y. Jiang, Z. Wang, et al., Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean, Nat. Biotechnol. 33 (2015) 408–414.
[33] C. Fang, Y. Ma, S. Wu, et al., Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean, Genome Biol. 18 (2017) 161.
[34] J. Schmutz, S.B. Cannon, J. Schlueter, et al., Genome sequence of the palaeopolyploid soybean, Nature 463 (2010) 178–183.
[35] G. Xu, C. Guo, H. Shan, et al., Divergence of duplicate genes in exon-intron structure, Proc. Natl. Acad. Sci. U. S. A. 109 (2012) 1187–1192.
[36] Y. Zhang, Y.J. Zhang, B.J. Yang, et al., Functional characterization of GmBZL2 (AtBZR1 like gene) reveals the conserved BR signaling regulation in Glycine max, Sci. Rep. 6 (2016) 31134.
[37] Q. Yan, L. Wang, X. Li, GmBEHL1, a BES1/BZR1 family protein, negatively regulates soybean nodulation, Sci. Rep. 8 (2018) 7614.
[38] M. Lynch, M. O'Hely, B. Walsh, A. Force, The probability of preservation of a newly arisen gene duplicate, Genetics 159 (2001) 1789–1804.
[39] A. Force, M. Lynch, F.B. Pickett, et al., Preservation of duplicate genes by complementary, degenerative mutations, Genetics 151 (1999) 1531–1545.
[40] J.F. Wendel, Genome evolution in polyploids, Plant Mol. Biol. 42 (2000) 225–249.
[41] F.A. Kondrashov, I.B. Rogozin, Y.I. Wolf, et al., Selection in the evolution of gene duplications, Genome Biol. 3 (2002). RESEARCH0008.
[42] L.E. Flagel, J.F. Wendel, Gene duplication and evolutionary novelty in plants, New Phytol. 183 (2009) 557–564.
[43] B.P. Cusack, K.H. Wolfe, When gene marriages don't work out: divorce by subfunctionalization, Trends Genet. 23 (2007) 270–272.
[44] G. Blanc, K.H. Wolfe, Functional divergence of duplicated genes formed by polyploidy during Arabidopsis evolution, Plant Cell 16 (2004) 1679–1691.
[45] B.C. Thomas, B. Pedersen, M. Freeling, Following tetraploidy in an Arabidopsis ancestor, genes were removed preferentially from one homeolog leaving clusters enriched in dose-sensitive genes, Genome Res. 16 (2006) 934–946.
[46] A. Roulin, P.L. Auer, M. Libault, et al., The fate of duplicated genes in a polyploid plant genome, Plant J. 73 (2013) 143–153.
[47] J.J. Doyle, L.E. Flagel, A.H. Paterson, et al., Evolutionary genetics of genome merger and doubling in plants, Annu. Rev. Genet. 42 (2008) 443–461.
[48] C. Vriet, E. Russinova, C. Reuzeau, Boosting crop yields with plant steroids, Plant Cell 24 (2012) 842–857.
[49] W.W. Wasserman, A. Sandelin, Applied bioinformatics for the identification of regulatory elements, Nat. Rev. Genet. 5 (2004) 276–287.
[50] T. Lee, H. Kim, I. Lee, Network-assisted crop systems genetics: network inference and integrative analysis, Curr. Opin. Plant Biol. 24 (2015) 61–70.