Genome-wide analysis of the SPL family transcription factors and their responses to abiotic stresses in maize

Hu-De Mao a,b,⁎, Li-Juan Yu b, Zhan-Jie Li b, Yan Yan b, Ran Han b, Hui Liu b, Meng Ma b

a State Key Laboratory of Crop Stress Biology for Arid Areas and College of Plant Protection, Northwest A&F University, Yangling, Shaanxi 712100, China
b College of Life Sciences, Northwest A&F University, Yangling, Shaanxi 712100, China

A R T I C L E   I N F O

Article history:
Received 9 January 2016
Received in revised form 20 February 2016
Accepted 1 March 2016
Available online 4 March 2016

Keywords:
Maize
Genome-wide analysis
SPL genes
Posttranscriptional regulation
Expression pattern
Abiotic stress response

A B S T R A C T

SQUAMOSA promoter binding protein-like (SPL) transcription factors (TFs) are plant-specific and play vital regulatory roles in plant growth and development. Even though they are one of the unique groups of TFs in plants, their characteristics, evolutionary relationships and expression patterns are largely unknown in maize, an important food crop worldwide. In this study, we identified 31 SPL genes (ZmSPLs) in the maize B73 genome. A phylogenetic analysis showed that these genes were divided into six groups (Groups 1–6) and members within the same group shared conserved exon/intron distributions and motif compositions, implying their functional redundancy. The 31 ZmSPL genes were distributed unevenly on 9 of the 10 chromosomes, with 10 segmental duplication events, suggesting that the expansion of the ZmSPL genes occurred due to segmental duplication. Analysis of the Ka/Ks ratios showed that the duplicated ZmSPL genes had primarily undergone strong purifying selection. In addition, 19 of the 31 ZmSPLs, belonging to Groups 1, 2 and 3, were targets of microRNA miR156, indicating of the miR156-mediated posttranscriptional regulation of these ZmSPL genes. Expression analysis of the ZmSPLs in various tissues at different development stages revealed distinct spatiotemporal patterns. Moreover, quantitative real-time PCR analysis identified several ZmSPL genes that were potentially involved in response to abiotic stresses. Our results present a comprehensive overview of the maize SPL family and provide an important foundation for further uncovering the biological functions of ZmSPLs in the growth and development of maize.

© 2016 The Authors. Published by Elsevier B.V.

1. Introduction

Transcription factors (TFs) are a large class of regulators that control gene expression at the transcriptional level and often serve as on/off switches in the developmental processes of eukaryotic organisms (Sun and Oberley, 1996). SQUAMOSA promoter binding protein-like (SPL) TFs are specific to plants and have a highly conserved SQUAMOSA promoter binding protein (SBP) domain, with approximately 78 amino acid residues. This domain contains three functionally important motifs, including two zinc-binding sites, Cys–Cys–Cys–His (Zn1) and Cys–Cys–His–Cys (Zn2), and a nuclear localization signal (NLS) that partially overlaps with the second Zn-finger, located at the C-terminal of the SBP domain (Yamasaki et al., 2004; Birkenbihl et al., 2005). Genes encoding SPLs were first identified for SBP1 and SBP2 in Antirrhinum majus (Klein et al., 1996). These genes have recently been identified in single-celled green algae, mosses and gymnosperms, as well as angiosperms (Jin et al., 2014). SPLs are encoded by a large gene family in plants. For instance, there are 16 SPL genes in Arabidopsis thaliana (Cardon et al., 1999), 19 in rice (Xie et al., 2006), and 28 in Populus trichocarpa (Li and Lu, 2014).

Recent studies in various species have suggested that SPL genes affect a broad range of developmental processes. To date, sixteen SPL genes have been identified in the Arabidopsis genome (Cardon et al., 1999), and many of them have been found responsible for a diverse number of developmental process events, including embryogenesis (Unte et al., 2003), shoot and leaf development (Wu and Poethig, 2006; Schwarz et al., 2008), flowering (Gandikota et al., 2007), vegetative and reproductive phase transitions (Jung et al., 2011), plastochron formation (Wang et al., 2008), fertility (Xing et al., 2010), copper homeostasis (Yamasaki et al., 2009) and plant hormone signaling (Zhang et al., 2007). Besides Arabidopsis, knowledge of the functions of SPL genes in other plant species, especially important agricultural and economical crops, has also begun to accumulate, highlighting the diverse roles of the SPL proteins in plant development. For example, OsSPL14, also known as ideal plant architecture1 (IPA1), is related to plant architecture and substantially enhances grain yield (Jiao et al., 2010; Miura et al., 2010; Lu et al., 2013); OsSPL16 promotes grain quality and yield (Wang et al., 2012); and OsLGI regulates a closed panicle trait in domesticated rice (Ishii et al., 2013). Epigenetic mutation analysis of the tomato colorless non-ripening (cnr) mutant has demonstrated that one of the...
tomato SPL genes is vital for tomato fruit ripening (Manning et al., 2006). In maize, liguleless1, which contains the SBP domain, regulates ligule and auricle formation (Moreno et al., 1997), and tasselsheath4 (tsh4) is an SPL gene that regulates bract development, a necessity in branch meristem initiation and maintenance (Chuck et al., 2010).

Along with the deepening of research, an increasing number of microRNAs (miRNAs) have been found to play a crucial role in the regulation of gene function in plants (Chen and Rajewsky, 2007; Lausserserques et al., 2015). miRNAs are small RNA molecules (20–24 nucleotides in length) that can cause the degradation of mRNA, or repress translation by binding to the transcripts of their target genes and forming an RNA-induced silencing complex; approximately half of their target genes encode transcription factors (Rhoades et al., 2002; Bartel, 2004). As a gene family that encodes transcription factors, more than half of the SPL genes identified to date have been found to be targeted by miR156/157. In Arabidopsis, for instance, 10 of the 16 SPL genes are putative targets of AtmiR156 (Wu and Poethig, 2006; Gandikota et al., 2007; Addo-Quaye et al., 2008), while 11 of the 19 SPL genes in rice have been revealed to be putative targets of OsmiR156 (Xie et al., 2006). The target sites are located both in the coding region and in the 3′-untranslated region (3′ UTR) (Xie et al., 2006).

Unfortunately, despite an increasing body of physiological and biochemical data, the biological role of the SPL gene family remains elusive. Although regulatory and functional data of the SPL genes are increasing, the biological role of many members in this family still needs to be clarified. Maize as one of most important crop worldwide, it is therefore intriguing for us to do a genome wide analysis of SPL family in this species after its full genome sequence released (Schnable et al., 2009), even though some early work has been done (Hultquist and Dorweiler, 2008). In order to characterize the whole SPL gene family in maize, we searched the maize genome assembly which identified 31 full-length ZmSPLs. This current study aimed to systematically analyze phylogenetic relationships, gene structures, conserved protein motifs, chromosomal locations, gene duplications and expression patterns of all the identified ZmSPLs. Additionally, we investigated the miR156-mediated posttranscriptional regulation of ZmSPLs, providing useful information to elucidate further the biological functions of SPL genes in maize.

2. Materials and methods

2.1. Identification of maize SPL genes

Different approaches were applied to identify the putative SPL proteins from Zea mays L. Initially, amino acid sequences encoding SPL proteins from Zea mays were retrieved from the Plant Transcription Factor Database 3.0 (http://planttfdb.cbi.pku.edu.cn/index.php) (Jin et al., 2014). Then, genes in the maize genome annotated with a Pfam SBP domain (PF03110) were retrieved from Phytozome v10.0 (http://www.phytozone.net). In addition, BLAST searches were performed against the maize genome using 16 known Arabidopsis SPL proteins, to identify any additional SPL members. All but the longest splice variants were removed and redundant sequences were also removed using the decrease redundancy tool (web.expasy.org/ decrease_redundancy). The presence of the SBP domain in the proteins was evaluated using the Pfam (http://pfam.sanger.ac.uk/search) (Fin et al., 2006) and SMART (http://smart.embl-heidelberg.de/) tools (Letunic et al., 2004). Arabidopsis SPL proteins sequences were downloaded from the Arabidopsis Information Resource (TAIR, http://www.arabidopsis.org/). Rice and sorghum SPL sequences were downloaded from the Plant Transcription Factor Database 3.0 (http://planttfdb.cbi.pku.edu.cn/index.php) (Jin et al., 2014).

Information on the ZmSPL genes, including the number of amino acids, coding sequence (CDS), open reading frame (ORF) lengths and location coordinates were acquired from the Phytozone database. Physical parameters, including the molecular mass and isoelectric point (pl) of the deduced proteins were generated by ExPASy (http://web.expasy.org/protparam/).

2.2. Multiple sequence alignments and phylogenetic analysis

To reveal the phylogenetic relationships among the ZmSPL proteins and their orthologs in rice, sorghum and Arabidopsis, a neighbor-joining (NJ) phylogenetic tree was constructed for 31 ZmSPLs, 16 AtSPLs, 19 OsSPLs and 18 SbSPLs using MEGA5.0 (Tamura et al., 2011). The bootstrap value was set to 1000 replications and the pairwise deletion option was used. Protein sequences used for the phylogenetic tree are listed in Supplementary File A1.

2.3. Gene structure and conserved motif analysis

To illustrate the exon–intron structures of the individual ZmSPL genes, a comparison of the genomic sequences and their corresponding coding sequences (CDS) was operated using GSDS (Gene Structure Display Server) (http://gds.cbi.pku.edu.cn/) (Guo et al., 2007). Conserved motif of ZmSPL proteins was identified using online MEME (Multiple Expectation Maximization for Motif Elicitation) (http://meme.nbcr.net/meme/cgi/meme.cgi) (Bailey and Elkan, 1995). The parameters were adopted as follows: optimum motif width was set to between 6 and 150 residues; maximum number of motifs was set to 15. Each structural motif annotation was performed using the Pfam (http://pfam.sanger.ac.uk/search) and SMART (http://smart.embl-heidelberg.de/) tools. The 78 amino acids of the SBP domain were aligned using clustalW. Sequence logos were generated using the weblogo platform (http://weblogo.berkeley.edu/).

2.4. Chromosome location and gene duplication

Chromosome locations of the ZmSPL genes were determined by a BLAST analysis of the ZmSPLs against the maize genome from Phytozone v10.0 (http://www.phytozone.net/eucalyptus.php). The tandem duplicates were identified, according to Hanada et al. (2008), as pairs of ZmSPL genes within 100 kb of each other that had 10 or fewer nonhomologous genes between them. Segmental duplicate gene pairs were analyzed on the Plant Genome Duplication Database server (http://chibba.agtec.uga.edu/duplication/index/locus), with a display range of 100 kb. The nonsynonymous rates (Ka), synonymous rates (Ks) and evolutionary constraints (Ka/Ks) between the duplicated pairs of ZmSPLs were calculated using the CODEML program in PAML (Yang, 2007). The approximate date of the duplication events was calculated using T = Ks / 2λ × 10−6 million years ago (Mya), based on the clock-like rates (λ) in grasses of 6.5 × 10−9 (Gaut et al., 1996).

2.5. Prediction of ZmSPLs targeted by miR156

The sequences of maize miR156a–miR156l were obtained from the miRBase (Kozomara and Griffiths-Jones, 2011) (http://www.mirbase.org/). ZmSPLs targeted by miR156 were predicted by searching the coding regions and 3′ UTRs of all the ZmSPL genes for complementary sequences of maize miR156a–miR156l on the psRNATarget server (http://plantgrn.noble.org/psRNATarget/?function=3), using default parameters (Dai and Zhao, 2011).

2.6. Microarray analysis

The published transcriptome data of the genome-wide gene expression atlas, of the maize inbred line B73, provided a useful complement to understand the expression patterns of the ZmSPLs during different developmental stages (Sekhon et al., 2011). Normalized gene expression values, expressed as the number of fragments per kilobase of exon per million fragments mapped (FPKM), were transformed using log2 (FPKM + 1), and then used for further expression analysis.
2.7. Plant materials and growth conditions

Seeds of the maize inbred line B73 were surface-sterilized in 1% (v/v) Topsin-M (Rotam Crop Sciences Ltd.) for 10 min. Then they were washed in deionized water and germinated on wet filter paper at 28 °C for 3 days. The germinated seeds were placed in a solution of 0.75 mM K2SO4, 0.1 mM KCl, 0.25 mM KH2PO4, 0.65 mM MgSO4, 0.1 mM EDTA–Fe, 2 mM Ca(NO3)2, 1.0 mM MnSO4, 1.0 mM ZnSO4, 0.1 mM CuSO4, 0.005 mM (NH4)6Mo7O24) for hydroponic cultivation with a 16-h light/8-h dark cycle at 28 °C. Four kinds of treatments: 10 min at 95 °C, 40 cycles of 15 s at 95 °C, and 30 s at 60 °C. The specificity of the amplicon for each primer pair was verified by melting curve analysis. The expression of ZmUbi-2 (UniProtKB/TrEMBL; ACC:Q42415) was used as an internal control. The quantification method (2−ΔΔCt) (Livak and Schmittgen, 2001) was used and the variation in expression was estimated from the three biological replicates. The primer pairs used for the qRT-PCR analysis of the ZmSPL genes are listed in Supplementary File A2.

3. Results

3.1. Identification of SPL family genes in maize

Gene models in the maize genome containing a SBP domain were identified using a superfamily search (see the Materials and Methods section). The candidates were then examined by Pfam and SMART to confirm the presence of the SBP domain. After removing the redundant sequences, 31 non-redundant SPL genes were identified. The 31 SPL genes were designated as ZmSPL, followed by the Arabic numbers 1–31, according to their position (from top to bottom) on chromosomes 1–10 (Table 1).

The physical parameters of each SPL protein were calculated using ExPASy server. As shown in Table 1, the full-length coding sequences (CDS) of the ZmSPL genes ranged from 300 bp (ZmSPL7) to 3339 bp (ZmSPL14) with the deduced proteins of 99–1112 amino acids. Furthermore, the computed molecular weights of these SPL proteins ranged from 11.2 to 121.9 kDa. The theoretical pl of the deduced ZmSPL proteins ranged from 5.41 to 10.05.

3.2. Phylogenetic analysis of SPLs

To better understand the molecular evolution and phylogenetic relationship among SPLs in plants, we constructed a phylogenetic tree based on multiple sequence alignments of maize, sorghum, rice, and

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Sequenced ID</th>
<th>Protein Length (aa)</th>
<th>Mass (Da)</th>
<th>pI</th>
<th>Chr.</th>
<th>ORF length (bp)</th>
<th>Exons</th>
<th>GRASSIUS TF name</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZmSPL1</td>
<td>GRMZM2G160917</td>
<td>403</td>
<td>41,667.07</td>
<td>8.73</td>
<td>1</td>
<td>1213</td>
<td>3</td>
<td>ZmSBP8</td>
</tr>
<tr>
<td>ZmSPL2</td>
<td>GRMZM2G169270</td>
<td>1106</td>
<td>121,171.61</td>
<td>6.97</td>
<td>1</td>
<td>3321</td>
<td>10</td>
<td>–</td>
</tr>
<tr>
<td>ZmSPL3</td>
<td>GRMZM2G156756</td>
<td>972</td>
<td>105,912.24</td>
<td>5.43</td>
<td>1</td>
<td>2919</td>
<td>11</td>
<td>–</td>
</tr>
<tr>
<td>ZmSPL4</td>
<td>GRMZM2G036297</td>
<td>399</td>
<td>43,359.10</td>
<td>7.45</td>
<td>2</td>
<td>1200</td>
<td>3</td>
<td>ZmSBP15</td>
</tr>
<tr>
<td>ZmSPL5</td>
<td>GRMZM2G126018</td>
<td>383</td>
<td>40,195.40</td>
<td>8.64</td>
<td>2</td>
<td>1152</td>
<td>3</td>
<td>ZmSBP23</td>
</tr>
<tr>
<td>ZmSPL6</td>
<td>GRMZM2G061734</td>
<td>429</td>
<td>44,959.34</td>
<td>6.73</td>
<td>2</td>
<td>1290</td>
<td>3</td>
<td>–</td>
</tr>
<tr>
<td>ZmSPL7</td>
<td>GRMZM2G113779</td>
<td>99</td>
<td>11,294.57</td>
<td>10.05</td>
<td>2</td>
<td>300</td>
<td>2</td>
<td>ZmSBP13</td>
</tr>
<tr>
<td>ZmSPL8</td>
<td>GRMZM2G133646</td>
<td>881</td>
<td>98,614.88</td>
<td>7.27</td>
<td>3</td>
<td>2646</td>
<td>11</td>
<td>–</td>
</tr>
<tr>
<td>ZmSPL9</td>
<td>GRMZM2G078761</td>
<td>447</td>
<td>46,723.30</td>
<td>9.21</td>
<td>3</td>
<td>1344</td>
<td>4</td>
<td>–</td>
</tr>
<tr>
<td>ZmSPL10</td>
<td>GRMZM2G106798</td>
<td>125</td>
<td>34,384.04</td>
<td>8.70</td>
<td>4</td>
<td>978</td>
<td>3</td>
<td>ZmSBP1</td>
</tr>
<tr>
<td>ZmSPL11</td>
<td>GRMZM2G101511</td>
<td>464</td>
<td>45,168.34</td>
<td>7.18</td>
<td>4</td>
<td>1393</td>
<td>3</td>
<td>–</td>
</tr>
<tr>
<td>ZmSPL12</td>
<td>AC233751.1_FG002</td>
<td>588</td>
<td>62,871.73</td>
<td>8.64</td>
<td>4</td>
<td>1767</td>
<td>6</td>
<td>–</td>
</tr>
<tr>
<td>ZmSPL13</td>
<td>GRMZM2G460544</td>
<td>408</td>
<td>41,901.31</td>
<td>8.73</td>
<td>4</td>
<td>1227</td>
<td>3</td>
<td>ZmSBP30</td>
</tr>
<tr>
<td>ZmSPL14</td>
<td>GRMZM2G098557</td>
<td>1112</td>
<td>121,807.20</td>
<td>7.54</td>
<td>4</td>
<td>3339</td>
<td>10</td>
<td>–</td>
</tr>
<tr>
<td>ZmSPL15</td>
<td>GRMZM2G168229</td>
<td>406</td>
<td>42,601.38</td>
<td>8.88</td>
<td>4</td>
<td>1221</td>
<td>3</td>
<td>–</td>
</tr>
<tr>
<td>ZmSPL16</td>
<td>GRMZM2G163813</td>
<td>331</td>
<td>36,182.26</td>
<td>8.88</td>
<td>4</td>
<td>996</td>
<td>4</td>
<td>ZmSBP19</td>
</tr>
<tr>
<td>ZmSPL17</td>
<td>GRMZM2G055451</td>
<td>482</td>
<td>50,723.66</td>
<td>9.01</td>
<td>4</td>
<td>1449</td>
<td>4</td>
<td>ZmSBP20</td>
</tr>
<tr>
<td>ZmSPL18</td>
<td>GRMZM2G138421</td>
<td>961</td>
<td>105,083.27</td>
<td>5.78</td>
<td>5</td>
<td>2886</td>
<td>11</td>
<td>ZmSBP6</td>
</tr>
<tr>
<td>ZmSPL19</td>
<td>GRMZM2G414805</td>
<td>440</td>
<td>47,112.40</td>
<td>9.10</td>
<td>5</td>
<td>1323</td>
<td>4</td>
<td>ZmSBP25</td>
</tr>
<tr>
<td>ZmSPL20</td>
<td>GRMZM2G072725</td>
<td>479</td>
<td>50,856.67</td>
<td>9.15</td>
<td>5</td>
<td>1440</td>
<td>4</td>
<td>–</td>
</tr>
<tr>
<td>ZmSPL21</td>
<td>GRMZM2G126827</td>
<td>332</td>
<td>36,514.70</td>
<td>8.96</td>
<td>5</td>
<td>999</td>
<td>4</td>
<td>ZmSBP12</td>
</tr>
<tr>
<td>ZmSPL22</td>
<td>GRMZM2G111136</td>
<td>441</td>
<td>46,001.58</td>
<td>6.70</td>
<td>5</td>
<td>1326</td>
<td>3</td>
<td>ZmSBP10</td>
</tr>
<tr>
<td>ZmSPL23</td>
<td>GRMZM2G150621</td>
<td>332</td>
<td>36,514.70</td>
<td>8.96</td>
<td>5</td>
<td>999</td>
<td>4</td>
<td>ZmSBP11</td>
</tr>
<tr>
<td>ZmSPL24</td>
<td>GRMZM2G101499</td>
<td>434</td>
<td>46,035.69</td>
<td>9.14</td>
<td>6</td>
<td>1305</td>
<td>3</td>
<td>–</td>
</tr>
<tr>
<td>ZmSPL25</td>
<td>GRMZM2G073588</td>
<td>378</td>
<td>39,590.11</td>
<td>9.33</td>
<td>7</td>
<td>1137</td>
<td>3</td>
<td>–</td>
</tr>
<tr>
<td>ZmSPL26</td>
<td>GRMZM2G067624</td>
<td>206</td>
<td>21,040.19</td>
<td>9.75</td>
<td>7</td>
<td>621</td>
<td>2</td>
<td>ZmSBP29</td>
</tr>
<tr>
<td>ZmSPL27</td>
<td>GRMZM2G081127</td>
<td>861</td>
<td>96,563.72</td>
<td>7.48</td>
<td>8</td>
<td>2586</td>
<td>12</td>
<td>–</td>
</tr>
<tr>
<td>ZmSPL28</td>
<td>GRMZM2G109354</td>
<td>850</td>
<td>94,022.87</td>
<td>5.41</td>
<td>8</td>
<td>2553</td>
<td>10</td>
<td>–</td>
</tr>
<tr>
<td>ZmSPL29</td>
<td>GRMZM2G371033</td>
<td>309</td>
<td>33,053.13</td>
<td>8.84</td>
<td>8</td>
<td>930</td>
<td>3</td>
<td>–</td>
</tr>
<tr>
<td>ZmSPL30</td>
<td>GRMZM2G148467</td>
<td>450</td>
<td>47,007.47</td>
<td>9.31</td>
<td>10</td>
<td>1353</td>
<td>3</td>
<td>ZmSBP21</td>
</tr>
<tr>
<td>ZmSPL31</td>
<td>GRMZM2G058588</td>
<td>396</td>
<td>43,268.96</td>
<td>7.81</td>
<td>10</td>
<td>1191</td>
<td>3</td>
<td>ZmSBP28</td>
</tr>
</tbody>
</table>
Arabidopsis SPL proteins (Fig. 1). The 84 SPL proteins among the different plants were clustered into six groups (named Groups 1–6). The ZmSPL genes were distributed in all six groups. Groups 1–4 include short SPLs with no more than 588 amino acid residues, whereas the members of Groups 5 and 6 are longer and vary from 850 to 1112 amino acids (Table 1). As expected, the SPL proteins from maize generally exhibited closer relationships to the SPL proteins from monocotyledonous angiosperms, than to those from dicotyledonous angiosperms. The highest number of monocot ortholog pairs (18) was observed in maize/sorghum. Every sorghum SPL gene (SbSPLs) except for SbSPL8 and -13 had an orthologous gene in maize; both SbSPL8 and -13 had two orthologous genes in maize. This phenomenon suggested that most SPL genes existed before the divergence of maize and sorghum and the orthologous genes of SbSPL8 and -13 duplicated after their divergence to give rise to two new paralogs. Some rice SPL genes had two or more maize paralogs, indicating that maize SPL genes duplicated and diversified after the divergence of rice and maize (Fig. 1).

3.3. Expansion of the SPL gene family in maize

The number of SPL genes identified in maize is larger than that in rice and sorghum (Fig. 1). To investigate the extent of lineage-specific expansion of the SPL genes in maize and rice, and maize and sorghum, we performed a joint phylogenetic analysis of all maize, rice and sorghum SPLs. We identified the nodes that led to maize- and rice-specific, or maize- and sorghum-specific clades (rectangles in Fig. 2). The nodes indicate the divergence point between maize and rice, or maize and sorghum, and thus represent the most recent common ancestral genes before the split. Some SPL genes might have been present in the most recent common ancestor of maize and rice, or maize and sorghum, but were later lost in both species. As shown in Fig. 2a, we found that three clades contained only maize SPL genes (red arrows) and three clades contained only rice SPL genes (blue arrows); this indicates that gene loss might have occurred in these clades. Similarly, in Fig. 2b, we found 11 clades that contained only maize SPL genes (red arrows), but no clades that contained only sorghum SPL genes, indicates that SPL gene loss might only occurred in sorghum.

The number of clades indicated that there were at least 22 and 29 ancestral SPL genes before the maize–rice and maize–sorghum splits. After the maize–rice split, maize and rice gained twelve and zero genes, and lost three and three genes, respectively. This has resulted in the rapid expansion of the SPL genes in maize. Clearly, the number of genes gained in the maize lineage was greater than that in the rice lineage (Fig. 2c). Similarly, after the maize–sorghum split, maize and sorghum gained two and zero genes, and lost zero and eleven genes, respectively. That is, the number of genes lost in the sorghum lineage was greater than that in the maize lineage (Fig. 2d).

3.4. Duplication mechanisms accounting for the maize SPL gene family expansion

The 31 ZmSPL genes were localized across all of the maize chromosomes, except for chromosome 9 (Fig. 3). The distribution of the ZmSPL genes on the chromosomes appears to be nonrandom. Three ZmSPL genes were present on each of chromosomes 1 and 8, two on chromosomes 3, 7 and 10, four on chromosome 2, eight on chromosome 4, six on chromosome 5, and only one on chromosome 6 (Fig. 3).

We surveyed gene duplication events to explore the expansion patterns of the maize SPL gene family. In this study, we identified 10 segmental duplication pairs, including ZmSPL1/25, ZmSPL2/14, ZmSPL3/18,
Fig. 2. Phylogenetic tree and copy number changes of the maize and rice (a), and maize and sorghum (b) SPL genes. The numbers on the branches indicate the bootstrap percentage values calculated from 1000 replicates; only values higher than 50% are shown. The nodes that represent the most recent common ancestral genes before the maize–rice (a) and maize–sorghum (b) splits are indicated by the red rectangles (bootstrap support N 50%). Clades that contain only maize or rice SPLs are indicated by the red and blue arrows, respectively. The copy number changes in the maize and rice (c), and maize and sorghum (d) SPL genes are shown; the numbers in the circles and rectangles represent the numbers of SPL genes in the extant and ancestral species, respectively, while the numbers on the branches (with plus and minus symbols, respectively) represent the numbers of gene gains and losses.
ZmSPL4/31, ZmSPL5/25, ZmSPL7/26, ZmSPL8/27, ZmSPL9/29, ZmSPL16/21 and ZmSPL22/24, and each of them were located in a pair of paralogous blocks (Fig. 3; Table 2). In addition, no ZmSPL genes were arranged in tandem repeats. The results provide strong evidence that segmental duplication has made an important contribution to maize SPL gene family expansion.

Further, the selective constraints of duplicated ZmSPL genes were explored by calculating the ratio of nonsynonymous substitutions (Ka) to synonymous substitutions (Ks) of each duplicated pairs. Generally, duplicated genes with a high Ka/Ks ratio (>1) are deemed to be evolving under positive selection, Ka/Ks = 1 indicates neutral selection, while Ka/Ks < 1 indicates negative or purifying selection (Juretic et al., 2005). A summary of Ka/Ks for 10 ZmSPL duplicated pairs is shown in Table 2. The ratios of duplicated ZmSPL gene pairs were all less than 0.7, which implied that the ZmSPL genes have mainly undergone strong purifying selection after the duplication events with limited functional divergence. We also estimated the evolutionary timescale based on a substitution rate of $6.5 \times 10^{-9}$ substitutions per synonymous site per year. As shown in Table 2, the divergence time of duplicated ZmSPL gene pairs ranged from 15.57 to 60.66 million years.

3.5. Gene structure and conserved motif analysis of ZmSPLs

To better examine the structural diversity of the ZmSPL genes, we performed an exon/intron analysis (Fig. 4). The schematic structures suggest that introns existed in the coding sequences of all the ZmSPL genes and the number of exons ranged from 2 to 12. In addition, the majority of ZmSPL genes in the same group (Fig. 1) bore similar numbers of exons. For example, all genes within Group 2 were made up of four exons. However, genes in Group 5 contained either ten or eleven exons.

Analysis the conserved domains of ZmSPL proteins showed that all of the 31 ZmSPLs contained an SBP domain, which was located in a region close to the N-terminus and was encoded by the first two exons of the ZmSPL genes (Fig. 4). Sequence alignment of SBP domains revealed that the conserved zinc-binding sites, Zn1 and Zn2, also existed in the SBP domain of the ZmSPL proteins. Zn1 contains the CX$_4$CX$_{16}$CX$_2$H (CCCH) or CX$_4$CX$_{16}$CX$_2$C (CCCC) signature sequence, whereas Zn2 contains the CX$_2$CX$_{11}$HX$_1$C (CCHC) signature (Supplementary File A3). In addition to Zn1 and Zn2, the SBP domain contains a conserved nuclear location signal (NLS) in the C-terminus of the SBP domains, with the consensus sequence of KRX$_{11}$RRRK (Supplementary File A3). Moreover, six ZmSPL proteins (ZmSPL2, ZmSPL3, ZmSPL8, ZmSPL14, ZmSPL18 and ZmSPL27) belonging to Group 5 contained an ANK or Ank-2 domain, with three or four ankyrin repeats (Supplementary File A4), indicating that these ZmSPL proteins may function by interacting with other proteins in plant cells.

### Table 2

<table>
<thead>
<tr>
<th>Paralogous pairs</th>
<th>Ka</th>
<th>Ks</th>
<th>Ka/Ks</th>
<th>Duplication date (MY)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZmSPL1-ZmSPL25</td>
<td>0.33</td>
<td>0.74</td>
<td>0.45</td>
<td>60.66</td>
</tr>
<tr>
<td>ZmSPL2-ZmSPL14</td>
<td>0.05</td>
<td>0.24</td>
<td>0.21</td>
<td>19.67</td>
</tr>
<tr>
<td>ZmSPL3-ZmSPL18</td>
<td>0.05</td>
<td>0.19</td>
<td>0.26</td>
<td>15.57</td>
</tr>
<tr>
<td>ZmSPL4-ZmSPL31</td>
<td>0.08</td>
<td>0.25</td>
<td>0.32</td>
<td>20.49</td>
</tr>
<tr>
<td>ZmSPL5-ZmSPL25</td>
<td>0.12</td>
<td>0.32</td>
<td>0.38</td>
<td>26.23</td>
</tr>
<tr>
<td>ZmSPL7-ZmSPL26</td>
<td>0.13</td>
<td>0.21</td>
<td>0.62</td>
<td>17.21</td>
</tr>
<tr>
<td>ZmSPL8-ZmSPL27</td>
<td>0.09</td>
<td>0.23</td>
<td>0.39</td>
<td>18.85</td>
</tr>
<tr>
<td>ZmSPL9-ZmSPL29</td>
<td>0.09</td>
<td>0.23</td>
<td>0.39</td>
<td>18.85</td>
</tr>
<tr>
<td>ZmSPL16-ZmSPL21</td>
<td>0.07</td>
<td>0.19</td>
<td>0.37</td>
<td>15.57</td>
</tr>
<tr>
<td>ZmSPL22-ZmSPL24</td>
<td>0.44</td>
<td>0.74</td>
<td>0.59</td>
<td>60.66</td>
</tr>
</tbody>
</table>
In addition to the conserved domains, other conserved motifs could also be important for the function of SPLs (Xie et al., 2006; Guo et al., 2008). We identified 15 motifs for the 31 ZmSPLs (Fig. 5; Supplementary File A5). Motifs 1, 2, 5 and 6 make up the SBP domain (Supplementary File A3), and Motif 4 corresponds to the ANK domain. However, the biological annotation of the other putative motifs remain unclear. Moreover, the number of motifs in each SPL varied from 3 to 12, and the ZmSPLs within a same subfamily shared common motif compositions (Fig. 5), indicating functional similarities among these proteins.

3.6. MiR156-mediated posttranscriptional regulation of ZmSPLs

According to the available annotation information (Zhang et al., 2009), the twelve maize miR156 genes (Zm-miR156a–Zm-miR156l) were distributed on 7 of the ten maize chromosomes: two were present on each of chromosomes 2, 3, 4, 5 and 7, and one was present on each of chromosomes 6 and 10 (Fig. 3). In order to understand the miR156-mediated posttranscriptional regulation of the ZmSPL genes, we searched the coding regions and 3’ UTRs of all ZmSPLs for the targets of maize miR156a–miR156l. The results showed that 19 ZmSPL genes, belonged to Group 1–3 were targets of miR156. The miR156-target sites in 17 of these ZmSPL genes present in the last exon, and encoded the conserved peptide ALSLLS. The target sites for the other two ZmSPL genes were located in the 3’ UTRs, close to the stop codons (Fig. 4; Fig. 6). The miR156-targeted SPL genes consistently included sequences from rice and Arabidopsis and were distributed into only three of the subgroups (Groups 1, 2 and 3). This suggests that miR156-mediated posttranscriptional regulation of the SPLs is conserved in plants.

3.7. Expression profiles of ZmSPLs in different tissues

The global transcriptome data provide us valuable information to predict the gene functions. In this study, we focus on the temporal and spatial expression patterns of ZmSPL genes to reveal their roles in developmental regulation. For this purpose, we conducted a comprehensive expression analysis of ZmSPL genes using the publicly available microarray data (Sekhon et al., 2011). It can be seen from the heat map that the expression patterns of different ZmSPL genes varied greatly (Fig. 7). The transcripts of ZmSPL3, -8, -9, -12, -14, -18, -22, -24, -27, and -28 expressed at high levels in most of the developmental stages, while the transcripts of ZmSPL5, -9, -12, -15, -22, and -30 showed relatively low expression levels. In addition, some tissue- or organ-specific genes were discovered. For example, ZmSPL4 was mainly expressed in the husk, while ZmSPL6 and ZmSPL11 were expressed at high levels in the cob (Fig. 7). Moreover, there tended to be little or no variation in the expression (among the tissues tested) of the ZmSPL genes containing a miR156 target site, including ZmSPL2, -3, -4, -8, -14, -15, -18, -22, -24, -27, -28, and -31. In contrast, the expression levels of ZmSPL genes containing a miR156 target site varied among the different tissues, except for ZmSPL5, -9, -12, and -30 (Fig. 7). The results indicate that the ZmSPLs may play important roles in maize development.
3.8. Expression profiles of ZmSPLs in response to abiotic stresses

In order to decipher the role of the ZmSPL genes in response to abiotic stresses, the expression of 31 ZmSPLs was analyzed in response to acute dehydration, salinity, cold and ABA exposures (Fig. 8; Supplementary File A6). The heat map of the expression levels is shown in Fig. 8a. Overall, the qRT-PCR analysis demonstrates that all of the genes displayed variations in their expression behavior in response to one or more stresses. Among the four treatments, cold stress induced relatively more dramatic changes in the transcript abundances of the ZmSPL genes than dehydration, salinity and ABA. Some genes were differentially expressed in response to a specific stress. For example, ZmSPL11 and ZmSPL21 were only induced by ABA, while ZmSPL10 was only induced by the salinity treatment. Some genes were co-regulated by two or more stresses. For example, transcripts of ZmSPL2 and ZmSPL20 accumulated during the cold and drought treatments; ZmSPL4, ZmSPL19 and ZmSPL24 were induced by the drought, salinity and ABA treatments; and ZmSPL1, -9, -12, -13, -15, -22, -26, and -29 were induced by all four stresses (Fig. 8). The results suggest that these ZmSPL genes may involve in regulating abiotic stresses in maize.

4. Discussion

SPL genes are plant-specific transcription factors that are broadly exist in photosynthetic organisms (Cardon et al., 1999; Guo et al., 2008; Jin et al., 2014), and are reported to participate in many crucial biological processes in plants. Recently, SPL genes were identified in diverse plant species, and the number of SPL genes varies among land plants. In this study, three methods were used to identify all SPL genes in maize, and result the different numbers of SPLs, respectively. For example, there are 65 maize SPL members in Phytozone 10 (https://phytozone.jgi.doe.gov/pz/portal.html), while only 55 members in PlantTFDB 3.0 (http://planttfdb.cbi.pku.edu.cn/). After removed all redundant and splice variant sequences, 31 non-redundant ZmSPL genes, and distributed on nine chromosomes were identified in the maize genome (Table 1; Fig. 3). Phylogenetic analyzes demonstrated that all the SPL genes were clustered into six groups (Group 1–6) (Fig. 1), consistent with previous findings (Guo et al., 2008). In addition, this classification was consistent with gene structure and motif composition. Genes within a group shared a similar length, structure, motif distribution, and miR156 target site location (Fig. 4; Fig. 5). The results indicated that the classification and evolution of SPL genes might be closely related to their structural divergence and diversification.

Our analysis of the conserved protein motifs in the maize SPL family (using MEME) showed that the majority of the ZmSPL proteins in the same group shared similar motif distributions. However, between different groups, there was a high divergence in the motif patterns (Fig. 5). For example, motif 10 was more specific to Group 2, whereas motifs 4, 7, 8 and 13 were only found in Group 5; these results indicate the intricate nature of the function of SPL proteins in maize. In addition, we found some motifs, like motifs 1, 2, 5 and 6, were more conserved and appeared in many ZmSPL proteins (Fig. 5). These motifs could be essential components that determine the common molecular functions of...
the SBP domain among the different family members. The differences in motif patterns among the different subfamilies of the ZmSPL genes are the structural foundations for the diversity in the gene functions.

To illuminate the phylogenetic relationship of the SPL genes, a comparative genomic analysis of the plant SPL members from monocots (maize, rice and sorghum) and dicotyledons (Arabidopsis) was performed. Orthologs are generally defined as genes from different genomes that were derived from a single ancestral gene and may have the same function; in contrast, paralogs are genes that originated from a single gene within a genome, were created by gene duplication events,
pairs of close paralogs were found to be involved in segmental duplications among the 31 ZmSPL genes (Table 2; Fig. 3). These genes represented about 64% of the ZmSPL genes that evolved from duplicated chromosomal regions. The results indicate that segmental duplication played a predominant role in the expansion of the maize SPL genes. Gene duplication has been reported for many plant TF gene families, such as AP2, MADS, and DOF, among others (Zahn et al., 2005; Moreno-Risueno et al., 2007; Shigyo et al., 2007). Duplicated SPL gene pairs have been identified in Arabidopsis (AtSPL10/11, AtSPL4/5 and AtSPL1/12) and rice (OsSPL2/19, OsSPL3/12, OsSPL4/11, OsSPL5/10 and OsSPL16/18) (Yang et al., 2008). However, the number of duplicated ZmSPL gene pairs identified was obviously greater than that in Arabidopsis and rice, indicating that (1) more segment duplication events happened in maize, and (2) most SPL genes in Arabidopsis, rice and maize expanded in a species-specific manner.

To further investigate the possible functions of the ZmSPL genes in plant growth and development, expression patterns of the 31 ZmSPLs in 16 different tissues were detected, based on the transcriptome data (Fig. 7). The heat map of the ZmSPLs expression patterns showed that some genes, like ZmSPL9 and ZmSPL12, appeared to be consistently expressed at low levels among all tissues. In contrast, ZmSPL3 and ZmSPL18 exhibited high expression levels in all tissues (Fig. 7). The consistent gene expression levels across all tissues can be considered constitutive expression. The low-level expression of many ZmSPL genes observed indicates that these genes may work synergistically with other proteins during plant growth and development. In addition, the three groups of ZmSPL genes discussed above (Groups 1, 2 and 3) all contained a miR156 target site (Fig. 4; Fig. 6). The expression levels of these genes were varied in the different tissues. In contrast, the genes in Groups 4, 5 and 6 did not contain an miR156 target site and were all expressed ubiquitously and constitutively, with little or no variation in any of the tissues analyzed (Fig. 7). These results indicate that maize genes from these two groups may have functions that are distinct from the miR156-targeted SPL genes in Groups 1, 2 and 3.

To date, several important and divergent biological processes regulated by SPL genes have been reported; however, only a small number of these genes have been shown to play a role in response to stress. For example, in Arabidopsis, the expression of an SPL gene responded to various types of biotic and abiotic stresses, through interactions with genes involved in the defense response pathway (Wang et al., 2009). In addition, AtSPL14 was found to be involved in programmed cell death and plays a role in sensitivity to fumonisin B1 (Stone et al., 2005). In our study, a preliminary expression profiling of the 31 ZmSPL genes showed that they were influenced by several environmental stimuli, including drought, cold, salinity and ABA exposure (Fig. 8); this indicates their role in abiotic stress responses. The differential expression profiles of the ZmSPL genes observed in this study underscore the daunting task of comprehending the global milieu associated with a stress response. However, an important outcome could be the comparison of their expressions patterns to multiple environmental stimuli, for the accurate identification of prospective candidate genes.

In conclusion, a total of 31 full-length SPLs were identified in maize genome. These proteins harbor sufficient structural diversities, in terms of their exon/intron distributions and motif compositions that may be related to their diverse functions. The expression patterns of the 31 ZmSPLs in different tissues further prove the presence of functional diversification of this family genes. The different expression patterns of miR156-targeted and –nontargeted genes indicate that miR156 plays important roles in maize development. In addition, the change in expression of ZmSPL genes to various abiotic stresses indicated that these genes may play important roles in responding to abiotic stress. These preliminary results of the ZmSPL genes in maize provide a foundation for further studies on the physiological and biochemical functions of SPL proteins.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.plgene.2016.03.003.
Author contribution

Conceived and designed the experiments: HDM. Performed the experiments: HDM, LJY, ZJL. Analyzed the data: HDM, LJY, ZJL, YY, RH, HL, MM. Wrote the paper: HDM, LJY. All authors read and approved the final manuscript and have no conflicts of interest with regard to this research or its funding.

Conflicts of interest

The authors declare that there is no conflict of interest.

Acknowledgments

We thank all our collaborators who made the work on SPL transcription factors a rewarding experience. This work was supported by the National Natural Science Foundation of China (31471505).

References


