Molecular characterization of the SPL gene family in *Populus trichocarpa*

Caili Li and Shanfa Lu*

**Abstract**

**Background:** SPLs, a family of transcription factors specific to plants, play vital roles in plant growth and development through regulation of various physiological and biochemical processes. Although *Populus trichocarpa* is a model forest tree, the *PtSPL* gene family has not been systematically studied.

**Results:** Here we report the identification of 28 full-length *PtSPLs*, which distribute on 14 *P. trichocarpa* chromosomes. Based on the phylogenetic relationships of SPLs in *P. trichocarpa* and *Arabidopsis*, plant SPLs can be classified into 6 groups. Each group contains at least a PtSPL and an AtSPL. The N-terminal zinc finger 1 (Zn1) of SBP domain in group 6 SPLs has four cysteine residues (CCCC-type), while Zn1 of SPLs in the other groups mainly contains three cysteine and one histidine residues (C2HC-type). Comparative analyses of gene structures, conserved motifs and expression patterns of PtSPLs and AtSPLs revealed the conservation of plant SPLs within a group, whereas among groups, the *P. trichocarpa* and *Arabidopsis* SPLs were significantly different. Various conserved motifs were identified in PtSPLs but not found in AtSPLs, suggesting the diversity of plant SPLs. A total of 11 pairs of intrachromosome-duplicated PtSPLs were identified, suggesting the importance of gene duplication in SPL gene expansion in *P. trichocarpa*. In addition, 18 of the 28 PtSPLs, belonging to G1, G2 and G5, were found to be targets of miR156. Consistently, all of the AtSPLs in these groups are regulated by miR156. It suggests the conservation of miR156-mediated posttranscriptional regulation in plants.

**Conclusions:** A total of 28 full-length SPLs were identified from the whole genome sequence of *P. trichocarpa*. Through comprehensive analyses of gene structures, phylogenetic relationships, chromosomal locations, conserved motifs, expression patterns and miR156-mediated posttranscriptional regulation, the *PtSPL* gene family was characterized. Our results provide useful information for evolution and biological function of plant SPLs.

**Background**

SPL proteins constitute a diverse family of transcription factors playing vital roles in plant growth and development. SPLs are specific to plants and have a highly conserved SBP (SQUAMOSA PROMOTER BINDING PROTEIN) domain with approximately 78 amino acid residues. The domain contains three functionally important motifs, including zinc finger 1 (Zn1), zinc finger 2 (Zn2), and nuclear location signal (NLS) [1,2]. Genes encoding SPLs were first identified for SBP1 and SBP2 in *Antirrhinum majus* [3]. Lately, it has been found in *Physcomitrella patens*, maize and tomato includes 16, 19, 13, 31 and 15 members, respectively [4-9].

For instance, the SPL gene family in *Arabidopsis*, rice, *Physcomitrella patens*, maize and tomato includes 16, 19, 13, 31 and 15 members, respectively [4-9].

The 16 *Arabidopsis* SPLs are termed as AtSPL1 to AtSPL16 [2], respectively, of which AtSPL1, AtSPL7, AtSPL12, AtSPL14 and AtSPL16 are relatively large and expressed constitutively, while the others are relatively small and highly expressed in flowers [4,10]. Ten of the 16 AtSPLs, including AtSPL2–AtSPL6, AtSPL9–AtSPL11, AtSPL13 and AtSPL15, are regulated by miRNAs belonging to the MIR156 family [11-17]. AtSPL3, AtSPL4 and AtSPL5 contain complementary sequences of miR156 in 3’ UTR, and all of them promote vegetative phase change and flowering [10,14,18]. AtSPL2, AtSPL10 and AtSPL11 regulate morphological traits of cauline leaves and flowers [19]. Overexpression of miR156b reduces the accumulation of AtSPL2, AtSPL10 and AtSPL11 mRNA [12,14,20]. AtSPL9 and AtSPL15 act redundantly...
in controlling the juvenile-to-adult growth phase transition and leaf initiation rate in *Arabidopsis* [21]. Six *AtSPLs*, including *AtSPL1*, *AtSPL7*, *AtSPL8*, *AtSPL12*, *AtSPL14* and *AtSPL16*, are not targets of miR156 in *Arabidopsis*. Among them, *AtSPL7* can bind directly to the Cu-response element (CuRE) containing a core sequence of GTAC and is a regulator of Cu homeostasis in *Arabidopsis* [22]. *AtSPL8* regulates pollen sac development [23], male fertility [24], GA biosynthesis and signaling [25]. *AtSPL14* plays significant roles in plant development and sensitivity to fumonisin B1 [26]. Among the 19 rice *SPLs*, half are predominantly expressed in various young organs [27]. OsSPLs targeted by miR156 are involved in the development of flowers in rice. OsSPL14 regulated by miR156 also controls shoot branching in the vegetative stage [8,28,29]. In maize, liguleless1 containing the SBP domain regulates ligule and auricle formation [30,31].

*Populus trichocarpa* is a model plant with whole genome sequence available [32]. A total of 352 miRNA precursors, including 12 for miR156, have been identified [33-39]. However, the regulation of miR156 in *P. trichocarpa* PtSPLs has not been analyzed. In our previous studies [40], 17 PtSPLs, which appeared to be full-length or partial sequence with at least 300 amino acids, were identified from the *Populus* genome assembly v1.1 (http://genome.jgi-psf.org/Poptr1_1/Poptr1_1.home.html). They were named PtSPL1–PtSPL17, respectively, of which PtSPL3 and PtSPL4 had the highest similarities with AtSPL7 involved in Cu homeostasis [40]. In order to investigate the evolutionary relationship between *P. trichocarpa* and *A. thaliana* SPL proteins, a neighbor-joining (NJ) phylogenetic tree was constructed for 28 PtSPLs and 16 AtSPLs using MEGA5.1. The reliability of branching was assessed by the bootstrap resampling method using 1,000 bootstrap replicates. Only nodes supported by bootstrap values >50% are used for further analysis. The results showed that the 44 SPL proteins clustered into 6 groups (named G1–G6), each of which contained at least one AtSPL and one PtSPL (Figure 2). It is consistent with the results from *SmSPLs* in *Salvia miltiorrhiza* [41]. To further confirm that there are 6 groups of SPLs, we also constructed a phylogenetic tree for 28 PtSPLs, 16 AtSPLs, 18 rice OsSPLs and 15 SmSPLs. As shown in Additional file 3, the 77 SPLs also clustered into 6 groups. The difference between the two trees constructed (Figure 2, Additional file 3) is that PtSPL12, PtSPL13, PtSPL28 and AtSPL6 belonging to G1 in Figure 2 are included in G2 in Additional file 3. An intron was found in the SBP domain-encoding region of all SPL genes from *P. trichocarpa* and *Arabidopsis* (Figure 3); however, sequence feature analysis showed that the SBP domain of SPLs in G6 (AtSPL7, PtSPL3 and PtSPL4) were divergent from the other groups. The N-terminal zinc finger of G6 SPLs has four cysteine residues in the SBP domain, while SPLs in the other groups mainly contain three cysteines and one histidine, indicating the diversification of plant SPL evolution. On the other hand, SPLs within a group have similar intron number, exon-intron structure, and coding sequence length. Consistently, the length, Mw and theoretical pI of deduced SPL proteins within a group are also similar, although they are divergent among groups. It suggests the conservation of plant SPLs in a group. Phylogenetic analysis showed that PtSPL3 and PtSPL4 had high homology with AtSPL7, an *Arabidopsis* SPL with the capability of binding CuREs in the MIR398 promoter in vitro.

**Results**

**Identification of 28 SPL genes in *P. trichocarpa* genome**

Analysis of the *Populus* genome assembly v1.1, v2.2 and v3.0 showed the existence of 28 full-length SPL genes in the *P. trichocarpa* genome (Table 1). All of the deduced PtSPL proteins contained the conserved SBP domain. The theoretical pI of deduced PtSPL proteins ranged from 5.87 to 9.49. The length varied between 148 and 1044 amino acids. The molecular weight (Mw) varied from 16.2 to 116.1 kDa (Additional file 1). The distribution of pI is similar to AtSPLs (Additional file 2); however, the length and Mw of PtSPLs are larger than AtSPLs. Mapping PtSPLs to the *P. trichocarpa* genome showed that 28 PtSPLs were unevenly distributed on 14 chromosomes with four on Chr2, 3 on each of Chr1, Chr8, Chr10 and Chr14, 2 on each of Chr3, Chr11 and Chr15, and one on each of Chr4, Chr5, Chr7, Chr12, Chr16 and Chr18 (Figure 1). Relatively high densities of PtSPLs were observed in the top and bottom regions of Chr8, Chr10, Chr11 and Chr14, the top of Chr1, Chr4, and Chr16, and the bottom of Chr3, Chr5, Chr7, Chr12 and Chr18. Few are in the central regions of chromosomes. Moreover, 11 pair of PtSPLs (Ks < 1.0) were evolved from intrachromosomal duplication (Table 2), indicating the importance of gene duplication for PtSPL gene expansion.

**Phylogenetic analysis of SPLs in *P. trichocarpa* and *Arabidopsis***

In order to investigate the evolutionary relationship between *P. trichocarpa* and *A. thaliana* SPL proteins, a neighbor-joining (NJ) phylogenetic tree was constructed for 28 PtSPLs and 16 AtSPLs using MEGA5.1. The reliability of branching was assessed by the bootstrap resampling method using 1,000 bootstrap replicates. Only nodes supported by bootstrap values >50% are used for further analysis. The results showed that the 44 SPL proteins clustered into 6 groups (named G1–G6), each of which contained at least one AtSPL and one PtSPL (Figure 2). It is consistent with the results from *SmSPLs* in *Salvia miltiorrhiza* [41]. To further confirm that there are 6 groups of SPLs, we also constructed a phylogenetic tree for 28 PtSPLs, 16 AtSPLs, 18 rice OsSPLs and 15 SmSPLs. As shown in Additional file 3, the 77 SPLs also clustered into 6 groups. The difference between the two trees constructed (Figure 2, Additional file 3) is that PtSPL12, PtSPL13, PtSPL28 and AtSPL6 belonging to G1 in Figure 2 are included in G2 in Additional file 3. An intron was found in the SBP domain-encoding region of all SPL genes from *P. trichocarpa* and *Arabidopsis* (Figure 3); however, sequence feature analysis showed that the SBP domain of SPLs in G6 (AtSPL7, PtSPL3 and PtSPL4) were divergent from the other groups. The N-terminal zinc finger of G6 SPLs has four cysteine residues in the SBP domain, while SPLs in the other groups mainly contain three cysteines and one histidine, indicating the diversification of plant SPL evolution. On the other hand, SPLs within a group have similar intron number, exon-intron structure, and coding sequence length. Consistently, the length, Mw and theoretical pI of deduced SPL proteins within a group are also similar, although they are divergent among groups. It suggests the conservation of plant SPLs in a group. Phylogenetic analysis showed that PtSPL3 and PtSPL4 had high homology with AtSPL7, an *Arabidopsis* SPL with the capability of binding CuREs in the MIR398 promoter in vitro.
and involved in response to copper deficiency in *Arabidopsis* [22]. It is consistent with our previous results for PtSPLs [40]. Based on the phylogenetic tree, PtSPL3 and AtSPL7 are very likely to be orthologous proteins (Figure 2). Additionally, 5 pairs of AtSPLs and 11 pairs of PtSPLs seem to be paralogous proteins (Figure 2). It includes AtSPL9/15, AtSPL10/11, PtSPL8/27, PtSPL12/13 and PtSPL11/19 belonging to G1, PtSPL18/22 and PtSPL14/15 from G2, PtSPL21/26 belonging to G3, AtSPL14/16, AtSPL11/12, PtSPL2/9, PtSPL1/5 and PtSPL6/7 included in G4, and AtSPL3/4, PtSPL16/23 and PtSPL20/25 clustering in G5. About 62.5% of the 16 AtSPLs and 78.5% of the 28 PtSPLs exist as paralogous pairs. It suggested that the expansion of SPL genes occurred after separation of paralogous genes. The results from paralogous pair identification were consistent with segmental duplications in the *P. trichocarpa* genome (http://chibba.agtec.uga.edu/duplication/) [32], suggesting the origination of paralogous PtSPLs from segmental duplication. Prediction of potential age of tandem duplication events using synonymous substitutions (Ks) values showed that the segmental duplication events for PtSPLs appeared to occur in 9–21 mya (Table 2). It is consistent with the age of *P. trichocarpa* genome duplication events [32].

| Gene name | Gene ID | V1.1 | V2.2 | V3.0 |
|-----------|---------|------|------|------|
| PtSPL1    | GW1.X.791.1 | POPTR_0010s16370 | Potri010G154000 |
| PtSPL2    | FGENESH4_PMC.LG_II000008 | POPTR_0002900440 | Potri020G024000 |
| PtSPL3    | ESTEXT_FGENESH4_PMC.LG_X0096 | POPTR_0010s02710 | Potri010G026200 |
| PtSPL4    | ESTEXT_FGENESH4_PMC.LG_VIII0830 | POPTR_000820160 | Potri008G197000 |
| PtSPL5a   | GRAIL3.010027501 + GRAIL3.010027301b | POPTR_000890810 | Potri008G098600 |
| PtSPL6a   | GRAIL3.0050015101b + GW1.8978.5.1b + GW1.II.489.1b | POPTR_0014s10960 | Potri014G114300 |
| PtSPL7a   | GRAIL3.0050015101b + GRAIL3.010027501b | POPTR_0002s18970 | Potri002G188700 |
| PtSPL8    | FGENESH4_PLC.LG_II001303 | POPTR_0002s14330 | Potri002G142400 |
| PtSPL9    | GRAIL3.0050015101b + EUGENE3.00051638b | POPTR_0005s28010 | Potri005G258700 |
| PtSPL10   | GRAIL3.0047015901b | POPTR_0003s17120 | Potri003G172600 |
| PtSPL11   | GRAIL3.0010026901 | POPTR_0008s09750 | Potri008G097900 |
| PtSPL12   | GRAIL3.010027501 | POPTR_0010s16400 | Potri010G154300 |
| PtSPL13   | FGENESH4_PLC.LG_X001404 | POPTR_0010s16400 | Potri010G154300 |
| PtSPL14   | ESTEXT_GENEWISE1_V1.C.LG_XV2187 | POPTR_0015s11100 | Potri015G098900 |
| PtSPL15   | EUGENE3.001002942 | POPTR_0012s10260 | Potri012G010700 |
| PtSPL16   | ESTEXT_GENEWISE1_V1.C.LG_1240186 | POPTR_0011s05480 | Potri011G055900 |
| PtSPL17a  | EUGENE3.00160416 | POPTR_0016s04880b + POPTR_0016s04890b | Potri016G048500 |
| PtSPL18   | GW1.I.7783.1b | POPTR_0001s13630 | Potri001G058600 |
| PtSPL19   | GW1.I.7693.1b | POPTR_0001s13890 | Potri001G055900 |
| PtSPL20   | GW1.107.39.1b | POPTR_0001s140870 | Potri001G098200 |
| PtSPL21   | GW1.II.3778.1b | POPTR_0002s14320 | Potri002G142200 |
| PtSPL22   | GW1.III.2396.1b | POPTR_0003s16780 | Potri003G169400 |
| PtSPL23   | GW1.IV.3037.1b | POPTR_0004s04630 | Potri004G046700 |
| PtSPL24   | GW1.IV.3778.1b | POPTR_0007s01030 | Potri007G138800 |
| PtSPL25   | GW1.III.3794.1b | POPTR_0011s11770 | Potri011G116800 |
| PtSPL26   | GW1.IV.80.1b | POPTR_0014s05680 | Potri014G057700 |
| PtSPL27   | GW1.IV.76.1b | POPTR_0014s05690 | Potri014G057800 |
| PtSPL28   | GW1.129.152.1b | POPTR_0015s07140 | Potri015G060400 |
| PtSPL29   | GW1.164.76.1b | POPTR_0018s14680 | Potri018G149900 |

*Genes are split into 2 or 3 gene models in v1.1 or v2.2;* 
*Gene models with partial sequence;* 
*The gene model includes 7 additional amino acids at the N-terminal compared with the model in v1.1.* 
*The gene model includes 49 additional amino acids at the N-terminal compared with the model in v2.2.*
Comparative analysis of PtSPL and AtSPL gene structures

Gene structure analysis showed that the number of introns in the coding region of 28 PtSPL genes varied from 1 to 10. The number of PtSPLs with 1, 2, 3, 4, 9, and 10 introns is five, ten, four, one, six, and two, respectively (Figure 3, Additional file 1). Similarly, the intron number of 16 AtSPLs varies between 1 and 9 (Additional file 2). The pattern of intron distribution in PtSPLs is quite similar to AtSPLs with the majority to be 2 and 9 introns, followed by 1 and 3 (Figure 3, Additional files 1 and 2) [41]. In addition, the position of intron in the SBP domain is highly conserved. It locates in the codon for the 48th amino acid of SBP domain (Additional file 4). These results suggest the conservation of exon-intron structures between PtSPLs and AtSPLs.

The length of introns varies significantly among PtSPL genes, such as those in G1, G2 and G5 (Figure 3). We analyzed the internal exons and introns of PtSPLs and AtSPLs. The results showed that the exons of PtSPLs had a size from 43 to 884 bp with an average of 314 bp, which is slightly greater than 293 bp of the average length of AtSPL exons. Approximately 59% of PtSPL exons and 63% of AtSPL exons have a size below 300 bp and 71% and 70% of exons are between 60 and 160 bp in PtSPLs and AtSPLs, respectively (Figure 4). Although the size distribution of PtSPL exons is similarity with AtSPL exons, intron size distribution is more variable, ranging from 30 bp to 3.0 kb. There are 6 PtSPL introns (5%) with sizes >1.5 kb; however, no such introns exist in AtSPLs. About 55% of PtSPLs have sizes below 300 bp and 56% of introns are between 60 and 160 bp; however, the majority of AtSPLs (94%) have sizes below 300 bp. The average size of PtSPL introns is 476 bp, which is much greater than 120 bp of AtSPLs. These results suggest the difference of exon and intron size distribution between PtSPLs and AtSPLs.

Identification of 25 conserved motifs

Conserved domains of PtSPLs were analyzed using Pfam (http://pfam.sanger.ac.uk) and by BLAST analysis of protein sequences against the Conserved Domain Database (CDD, http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). The results showed that all of the 28 PtSPLs and 16 AtSPLs contained a SBP domain with about 78 amino acid residues
Figure 2 Neighbor-joining (NJ) phylogenetic tree for 44 SPLs in *P. trichocarpa* and *Arabidopsis*. The groups of homologous genes identified and bootstrap values are shown. The reliability of branching was assessed by the bootstrap re-sampling method using 1,000 bootstrap replicates. Bootstrap values are shown below nodes.

Figure 3 Exon-intron structures of PtSPLs. Introns are represented by lines. Exons are indicated by green boxes. The SBP domains are shown in red boxes. Intron phases are shown by 0, 1 and 2.
in length (Figure 5). It is not surprising given that the SBP domain was used for PtSPL identification. Sequence analysis of SBP domains revealed that the conserved zinc-binding sites, Zn1 and Zn2, also existed in the SBP domain of PtSPLs (Figure 5). Zn1 is Cys3His-type (CCCH-type) in G1–G5 SPLs (Figure 5A); however, the His residue in Zn1 is replaced by a Cys residue in G6, which results in the signature sequence of G6 SPLs to be CCCC (Figure 5B). Unlike Zn1, the signature sequence (C2HC) of Zn2 is highly conserved in all SPLs analyzed. In addition to Zn1 and Zn2, the SBP domain contains a conserved nuclear location signal (NLS) in the C-terminus of SBP domains (Figure 5). The conservation of SBP domains between PtSPLs and AtSPLs indicates that the domain organization has been established in ancient plants. Moreover, six PtSPLs (PtSPL1, PtSPL2, PtSPL5, PtSPL6, PtSPL7 and PtSPL9) belonging to G4 contain an ANK or Ank-2 domain with three or four ankyrin repeats (Additional file 5), which are involved in protein-protein interaction [42]. It is consistent with previous results from AtSPLs and SmSPLs [41].

In addition to the conserved domains, other conserved motifs could also be important for the function of SPLs [27,43]. We searched conserved motifs using MEME and applied an e-value cut off of $10^{-10}$ to the recognition. It resulted in the identification of 25 motifs for 28 PtSPLs (Figure 6, Table 3). The majority of motifs identified are conserved between PtSPLs and AtSPLs [41], while three, including motifs 11, 19 and 23, are specific to PtSPLs. It indicates the conservation and diversity of PtSPLs and AtSPLs. The number of motifs in each SPL varies from 1 to 16 (Figure 6). Motif 1 is actually the SBP domain. Consistently, it exists in all SPLs analyzed. Motif 14 existed in G1 and G2 SPLs contains the target gene sequence of miR156, indicating the posttranscriptional regulation of G1 and G2 SPLs by miR156. In addition to motifs 1 and 14, several motifs widely exist in two SPL groups, such as motif 12 found in G1 and G2, motifs 2, 4, 5, 6, 15 and 16 existing in G4 and G6 (Figure 6), indicating the importance of these motifs. We also found several motifs to be group-unique, such as motif 24 specifically existing in G6 SPLs and motifs 7, 9, 10 and 18 specific to G4 (Figure 6). These group-unique motifs could be important for specific roles of SPLs in the group. Moreover, PtSPLs and AtSPLs [41] within a group share similar motif(s), indicating they probably play similar roles in plant growth and development.
Expression patterns of SPLs in P. trichocarpa
The expression pattern of a gene is often correlated with its function. In order to preliminarily elucidate the roles of PtSPLs in P. trichocarpa development, we first searched PopGenIE for gene expression data from microarray analysis [44]. Except for PtSPL17, the expression levels of 27 PtSPLs in roots, stems, young leaves and mature leaves were obtained (Figure 7). Next, we examined the relative expression levels of 28 PtSPLs in young leaves, mature leaves, young stems, young roots and tissues from developing xylem and phloem from the 4th–6th and 12th–25th internodes of one-year-old P. trichocarpa plants using the quantitative real-time RT-PCR method (Figure 8). The results showed that qRT-PCR data was generally consistent with microarray data for relative expression of PtSPLs in roots, stems, young leaves and mature leaves (Figures 7 and 8). Although all PtSPLs were expressed in at least one of the tissues examined, differential expression was observed. Many putative paralogous genes, such as PtSPL18/22 in G2, PtSPL21/26 in G3, PtSPL2/9, PtSPL1/5 and PtSPL6/7 in G4 and PtSPL16/23 belonging to G5, show similar expression patterns, suggesting redundant roles of these PtSPL gene pairs. However, the expression patterns of few gene pairs, including PtSPL12/13 in G1, and PtSPL14/15 belonging to G2 are distinct. It indicates these PtSPLs may play different roles in P. trichocarpa development, although they are paralogous genes.

MiR156-mediated posttranscriptional regulation of PtSPLs
It has been shown that 10 AtSPLs are regulated by miR156 [11]. The complementary sites of miR156 are in the coding regions or 3’ UTRs of AtSPLs. In order to know miR156-mediated posttranscriptional regulation of PtSPLs, we searched coding regions and 3’ UTRs of all PtSPLs for targets of P. trichocarpa miR156a–miR156j on the psRNATarget server using default parameters [45]. The results showed that 18 PtSPLs were potential targets of miR156 (Figures 9 and 10). MiR156-targeting sites in 13 PtSPLs belonging to G1 and G2 locate in the last exon and encode the conserved peptide ALSLLS. The target sites for other 5 PtSPLs belonging to G5 locate in the 3’ UTRs close to the stop codons (Figure 10). Consistently, AtPSLs clustering in G1, G2 and G5 are targets of miR156 in Arabidopsis. It suggests that miR156-mediated posttranscriptional regulation of SPLs is conserved in P. trichocarpa and Arabidopsis.

Discussion
SPLs are plant-specific transcription factors containing a highly conserved SBP (SQUAMOSA PROMOTER BINDING PROTEIN) domain. It can specifically bind to the promoters of floral meristem identity gene SQUAMOSA and its orthologous genes and plays important regulatory roles in plant growth and development [46-49]. The genes encoding SPLs have been identified from various plant species, such as Arabidopsis [2,10,23,26], maize [30], Antirrhinum majus [3], rice [50], silver birch [51], and S. miltiorrhiza [41]. SPL genes exist as a large gene family in plants. The number of SPLs in Arabidopsis, rice, P. patens, maize and tomato is 16, 19, 13, 31 and 15, respectively [4-9]. Availability of the whole genome sequence allows us to perform genome-wide identification of SPLs in P. trichocarpa. Analysis of three versions of the annotated P. trichocarpa genome showed the existence of 28 full-length
PtSPLs, which distribute on 14 chromosomes. It is the first attempt to analyze the PtSPL gene family. The results provide a basis for elucidating the functions of SPLs in *P. trichocarpa*, a model forest tree.

The number of SPL genes in *P. trichocarpa* is much greater than that in *Arabidopsis*, rice, *P. patens* and tomato, although it is similar to the number of maize SPLs [4-9]. Sequence homologous analysis suggests that gene duplication plays an important role in SPL gene expansion in *P. trichocarpa*. A total of 11 pairs of intrachromosome-duplicated PtSPLs were identified in this study. All of them clustered together in the phylogenetic tree (Figure 2). It is consistent with previous findings for generation and maintenance of gene families in other organisms, such as mouse, human and *Arabidopsis* [52,53]. Actually, gene duplication has been reported for many plant transcription factor gene families, such as *MYB*, *AP2*, *MADS* and so on [54-56] and 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 OsSPL6/18) [57-61]. However, the number of homologous PtSPL gene pairs is obviously greater than that in *Arabidopsis* and rice, indicating that more segment duplication events happened in *Populus* and most SPL genes in *Arabidopsis* and *Populus* expanded in a species-specific manner [62-64].

Comparative analysis of *P. trichocarpa* PtSPLs and *Arabidopsis* AtSPLs revealed many conserved sequence features. For instance, all of the deduced proteins contain the highly conserved SBP domain with about 78 amino acid residues. The intron position and intron phase in the SBP-domain-encoding regions are also conserved among all SPL genes in *P. trichocarpa* and *Arabidopsis*, indicating that plant SPL genes originate from a common ancestor. Based on the neighbor-joining (NJ) phylogenetic tree constructed using MEGA 5.1, 44 SPL proteins from *P. trichocarpa* and *Arabidopsis* were found to cluster into 6 groups. Each group includes at least a PtSPL and one AtSPL. The intron number and intron phase are similar for PtSPLs and AtSPLs within a group. The results suggest the conservation between *P. trichocarpa* PtSPLs and *Arabidopsis* AtSPLs.

It has been shown that AtSPLs play significant regulatory roles in a variety of developmental processes in *Arabidopsis*. For instance, morphological traits of cauline leaves and flowers are regulated by AtSPL2, AtSPL10 and AtSPL11 [19]. Juvenile-to-adult growth phase transition and leaf initiation rate are controlled by the
| Motif | E-value     | Consensus sequence                                                                 |
|-------|-------------|-----------------------------------------------------------------------------------|
| 1     | 1.0e-2549   | CQVEGCNADLSASKYDHRRRLHVSEVSKVIPVWAFLGEOICQCSRFHLSEDEGRKSCRRRLAGHNERRRKPOFD      |
| 2     | 3.3e-516    | SDQPS555S5GSQQCRTGKVRKFLKDFPNGTPTRGLQIDWLSHDMSYRPGCIIIITLAMPEAAWELCCDLG         |
| 3     | 6.2e-387    | LRPDPVAGPALLPHTLIAACKDGSEDVLALTEDPGEV SASWKNARDATGFTPYARLGHSHYHVLQRKLADKRRNGQVSM |
| 4     | 3.6e-174    | ASRSLYRPMALSMVAAAVCVALALKSCPeLYV                                              |
| 5     | 9.0e-136    | VEAGEEETFVKGRNLYQPGTIRLCVAGKLYTQETQALMD                                       |
| 6     | 7.4e-135    | FLPRLKFLFLEFSMORDWCAKIRLLDMLVEGNCRD                                             |
| 7     | 5.6e-100    | FWRTGFYVQRQLAFHKGQVWLDTSL                                                       |
| 8     | 3.0e-085    | GGSMNDDOQGGYTSLSTLSHLSNRSDQTQDQDLSSHLLRSLASHAGEHNGRNLFGLQGPRGL                 |
| 9     | 1.8e-095    | EGMPSKEQALDFLNEJQWLHRSDLKSR                                                   |
| 10    | 4.8e-088    | SSLEALSEMGHLHRAVRNRSRMVELLNR                                                   |
| 11    | 7.4e-085    | MAREFGEKESKHFYAPVPSLAVGKRGLEWLDNLKDGDLPASLPVPSDCRSRQFPTGPGLGEEKOOGNNSSSCS      |
| 12    | 3.2e-077    | STSLGAXSS55SGELGLKLRKRYFDAXGaNXK                                               |
| 13    | 3.5e-070    | FSPNNFAAKSEEPATAGQKQLNFLNDYDSDGGEDIEHSAPVNAWGMSFDCPLMVQOSHKSSPPHTSGNDS         |
| 14    | 1.3e-052    | ASDSADCALSSLSSQS                                                                |
| 15    | 8.9e-057    | NFFCSSXPLNLGRGFEVED                                                              |
| 16    | 9.8e-064    | PFIADAVCSEIRLQEFDF                                                              |
| 17    | 8.7e-053    | GERSSCNESPSEDOSQDSQDRPNPLQLFSSPENESRPKVASSKYFSSASSNPEDRSPSSSP                  |
| 18    | 1.8e-039    | PFRWELLGYTG                                                                    |
| 19    | 5.1e-035    | QHDGEMHILPPITTDWGDGLDFAVDOQFPLSFDPGLQPIDNTPPEIESQQLgenceAPVDRVRKPRPLTCPNSNLGIV|
| 20    | 2.8e-033    | DDWWLAKWDWGDGEFEA                                                               |
| 21    | 8.5e-032    | MDCNGKPHLQWDWENLIMFNATTENSK                                                    |
| 22    | 2.7e-030    | DEDNLGDEIKGRKREKRRRRVFFDDLDNNND                                              |
| 23    | 3.3e-028    | VNSARFSSNSQTYRHLFGSSQFSTSAMNAAWGAARERDPMNLTSSQMNDFDRKRNLFGSLPSYNEKQFQFPQLGS|
| 24    | 3.4e-027    | KGRMRVLYIUNMFIVKNXKGHAPHRLYHYHPTC                                                |
| 25    | 5.9e-025    | DERRQMSHAWDKAPVLHPANANLTWEGTSISKFTKTQDYIAKPABIGGNDQFHLPGDLPNTGATQHHKSN       |
The redundant action of AtSPL9 and AtSPL15 [21]. Pollen sac development, male fertility and GA biosynthesis and signaling are regulated by AtSPL8, a member of G3 [23-25]. Cu homeostasis in Arabidopsis is regulated by the member of group 6, AtSPL7 [22]. In this study, we found that many motifs were unique to or mainly existed in a group of SPLs. It is consistent with the redundant roles of AtSPLs in a group and indicates that the members of PtSPLs in the same group may play similar roles as their Arabidopsis counterparts. The function of SPLs in different groups could be functionally distinct. On the other hand, three PtSPL-specific motifs, including motifs 11, 19 and 23, were identified, suggesting that some PtSPLs may play species-specific roles. Consistently, most of paralogous PtSPL gene pairs in the same group show similar expression patterns, whereas a few of them exhibit differential patterns. The results indicate subfunctionalisation and neofunctionalisation of SPLs within a plant species and among different species.

MiR156-mediated posttranscriptional regulation is important for the function of a subset of SPLs [11,41,65]. Target prediction showed that all PtSPLs in groups 1, 2 and 5 were regulated by miR156. The complementary sites of miR156 locate in the coding region of G1 and G2 SPLs, whereas it locates in 3' UTR of G5 SPLs. It is consistent with the results from Arabidopsis SPLs and suggests the conservation of miR156-mediated posttranscriptional regulation in plants.

**Conclusion**

In this study, a total of 28 full-length SPLs were identified from the whole genome sequence of P. trichocarpa. Through a comprehensive analysis of gene structures, phylogenetic relationships, chromosomal locations, conserved motifs, expression patterns and miR156-mediated posttranscriptional regulation, the PtSPL gene family was characterized and compared with SPLs in Arabidopsis. The results showed that 28 PtSPLs and 16 AtSPLs clustered into 6 groups. Many PtSPLs and AtSPLs within a group are highly conserved in sequence features, gene structures, motifs, expression patterns and posttranscriptional regulation, suggesting the conservation of plant SPLs within a group. However, significant differences were observed for SPLs among groups. In addition, various motifs were identified in PtSPLs but not in AtSPLs. It suggests the diversity of plant SPLs. The results provide useful information for elucidating the functions of SPLs in P. trichocarpa.

**Methods**

**Identification of PtSPL genes**

The nucleotide sequences and deduced amino acid sequences of 16 known SPL genes in Arabidopsis [24] were obtained from the TAIR database (http://www.arabidopsis.org) (Additional file 2). The SBP domain of AtSPLs was identified using Pfam (http://pfam.sanger.ac.uk). BLAST search of PtSPLs against Populus trichocarpa v1.1, v2.2 and v3.0 was carried out using AtSPL SBP as the query sequences [32] (http://genome.jgi-psf.org/Poptr1_1/Poptr1_1.home.html,http://www.phytozome.net/poplar.php#B). An e-value cut off of 1e-5 was applied to the recognition. We also searched the databases for SBP using the keywords search tool on the web servers. Protein
Figure 8 Analysis of PtSPL gene expression in eight tissues of *Populus trichocarpa* using the qRT-PCR method. Fold changes of transcript levels in young root (Rt), young stems (St), young leaves (L1), mature leaves (L2), developing secondary xylem (X1), developing secondary phloem (P1), developed secondary xylem (X2) and developed secondary phloem (P2) of *Populus* plants are shown. Transcript levels in roots were arbitrarily set to 1 and the levels in other tissues were given relative to this. Error bars represent standard deviations of mean value from three biological replicates. ANOVA (analysis of variance) was calculated using SPSS. P < 0.05 was considered statistically significant.
sequences retrieved from *Populus trichocarpa* v1.1, v2.2 and v3.0 were then aligned and combined based on sequence identities.

**Chromosome location and sequence feature analyses**

Chromosome locations of *PtSPL* genes were determined by BLAST analysis of *PtSPLs* against *Populus trichocarpa* v3.0 (http://www.phytozome.net/poplar.php#B). Paralogous gene pairs were analyzed on the Plant Genome Duplication Database (PGDD) server (http://chibba.agtec.uga.edu/duplication/index/locus) with display range for 100 kb. The approximate date of the duplication events was calculated using $T = \frac{Ks}{2\lambda}$ by assuming clock-like rates ($\lambda$) in *Populus* for $9.0 \times 10^{-9}$ [32,57,66]. Synonymous substitutions (Ks) values of paralogous gene pairs were calculated using DnaSP [67]. The theoretical isoelectric point (pI) and molecular weight (Mw) were predicted using the Compute pI/Mw tool on the ExPASy server (http://web.expasy.org/compute_pi/) [68]. The intron/exon structure of *SPL* genes was predicted with the Gene Structure Display Server (http://gsds.cbi.pku.edu.cn/chinese.php) [69].

**Figure 9** Sequence alignment of *P. trichocarpa* miR156a–miR156j with their complementary sequence in coding regions and 3’ UTRs of 18 *PtSPLs*.

**Figure 10** *PtSPLs* targeted by miR156. Heavy grey lines represent ORFs. The lines flanking ORFs represent 3’-UTR. The blue lines represent SBP domain. miRNA complementary sites (green) with the nucleotide positions of *PtSPL* cDNAs are indicated. The RNA sequence of each complementary site from 5’ to 3’ and the predicted miRNA sequence from 3’ to 5’ are shown in the expanded regions.
**Phylogenetic construction and motif analysis**

Phylogenetic trees were constructed using the neighbor-joining (NJ) method in MEGA5.1. Branching reliability was assessed by the bootstrap re-sampling method using 1,000 bootstrap replicates. Only nodes supported by bootstrap values greater than 50% were analyzed. Conserved domains of PtSPLs were identified using Pfam (http://pfam.sanger.ac.uk) and by BLAST analysis of protein sequences against the Conserved Domain Database (CDD, http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) with the expected e-value threshold of 1.0 and the maximum size of hits to be 500 amino acids [70]. The 78 amino acids of SBP domain were aligned using clustalW. Sequence logos were generated using the weblogo platform (http://weblogo.berkeley.edu/). Potential protein motifs were predicted using the MEME package (http://meme.sdsc.edu/meme/) with the following parameters applied. It includes the distribution of motifs: zero and one per sequence, maximum number of motifs to find: 25, minimum width of motif: 8, and maximum width of motif: 150. An e-value cut off of 1e^{-10} was applied to the recognition.

**Quantitative real-time reverse transcription-PCR (qRT-PCR)**

*P. trichocarpa* plants were grown in an artificial climate chamber for about one year. Young leaves (2nd–3rd from the top), mature leaves (12th from the top), young stems (1st–3rd from the top), young roots, tissues of developing secondary xylem and phloem from the 4th–6th and 12th–25th internodes from the top of *P. trichocarpa* plants were collected. Three biological repeats were carried out. Total RNA was extracted using the plant total RNA extraction kit (Aidlab, China). Genomic DNA contamination was eliminated by pre-treating total RNA with RNase-free DNase (Promega, USA). RNA integrity was analyzed on a 1.2% agarose gel and its quantity was determined using a NanoDrop 2000C Spectrophotometer (Thermo Scientific, USA). Total RNA was reverse-transcribed by Superscript III Reverse Transcriptase (Invitrogen, USA). qRT-PCRs were carried out in triplicate for each tissue sample using gene-specific primers (Additional file 6) as described previously [71]. The program used for qRT-PCR is as follows: predenaturation at 95°C for 30s, 40 cycles of amplification at 95°C for 5 s, 60°C for 18 s and 72°C for 15 s. The length of amplicons was between 80 bp and 250 bp. *Actin* was used as a reference gene as described previously [72]. Dissociation curve was used to assess amplification specificity. Relative abundance of transcripts was analyzed using the comparative Ct method [73]. The arithmetic formula, 2-ΔΔCq, was used to achieve results for relative quantification. Cq represents the threshold cycle. Standardization of gene expression data from three biological replicates was performed as described [74]. For statistical analysis, ANOVA (analysis of variance) was calculated using SPSS (Version 19.0, IBM, USA). P < 0.05 was considered statistically significant.

**Microarray data analysis**

Microarray data of PtSPLs was obtained by the ePlant-tissue expression tool at PopGenIE (http://www.popgenie.org/). The data was gene-wise normalized and then analyzed using the average linkage clustering technique in Cluster 3.0 [75].

**Prediction of PtSPLs targeted by miR156**

The sequences of *P. trichocarpa* miR156a–miR156j were obtained from mirBase [36] (http://www.mirbase.org/). PtSPLs targeted by miR156 were predicted by searching the coding regions and 3’ UTRs of all PtSPLs for complementary sequences of *P. trichocarpa* miR156a–miR156j on the psRNATarget server using default parameters [45] (http://plantgrn.noble.org/psRNATarget?function=3).

**Availability of supporting data**

The data sets supporting the results of this article are included within the article and its additional files.

**Additional files**

- Additional file 1: Sequence features of PtSPLs in *P. trichocarpa*. Protein length, intron number, pI and molecular weight of SPLs in *P. trichocarpa* are shown.
- Additional file 2: Sequence features of AtSPLs in *A. thaliana*. Gene IDs, protein length, intron number, pI and molecular weight of SPLs in *A. thaliana* are shown.
- Additional file 3: Neighbor-joining (NJ) phylogenetic tree constructed for 77 SPLs from *P. trichocarpa*, *Arabidopsis*, rice and *S. miltiorrhiza*. The groups of homologous genes identified and bootstrap values are shown. The reliability of branching was assessed by the bootstrap re-sampling method using 1,000 bootstrap replicates. Bootstrap values are shown below nodes.
- Additional file 4: Intron distribution on SBP domains of *Populus* and *Arabidopsis*. Intron distribution on SBP domains of *Populus* and *Arabidopsis* are shown.
- Additional file 5: Alignment of the ANK/ANK-2 domain. The ANK/ANK-2 domain is indicated by solid lines.
- Additional file 6: Primers used for qRT-PCR analysis of PtSPL genes. Complete set of primers used for qRT-PCR.

**Abbreviations**

CuRE: Cu-response element; Mw: The molecular weight; NJ: Neighbor-joining; NLS: Nuclear location signal; pI: Isoelectric point; qRT-PCR: Quantitative realtime reverse transcription-PCR; SBP: SQUAMOSA PROMOTER BINDING PROTEIN; Zn1: Zinc finger 1; Zn2: Zinc finger 2.

**Competing interests**

The authors declare that they have no competing interests.

**Authors’ contributions**

CL contributed to bioinformatics and qRT-PCR analyses and participated in writing the manuscript. SL designed the experiment, performed bioinformatics analysis and wrote the manuscript. Both authors have read and approved the version of manuscript.
References

1. Yamasaki K, Kigawa T, Inoue M, Tateno M, Yamasaki T, Yabuki T, Aoki M, Seki M, Matsuoka T, Nonomura E, Itoh Y, Terada T, Shirouzu M, Osanai T, Tanaka A, Seimiya K, Yokoyama S. A novel zinc-binding motif revealed by solution structures of DNA-binding domains of Arabidopsis SBP-family transcription factors. J Mol Biol 2004, 337:99–63.

2. Birkenbihl RP, Jach G, Saedler H, Huijser P. Molecular characterisation of the Arabidopsis SBP-box genes. Gene 1999, 237:91–104.

3. Klein J, Saedler H, Huijser P. A new family of DNA binding proteins involved in the floral transition. Plant J 2012, 73:377–388.

4. Hultquist JF, Dorweiler JE. Feminized tassels of maize. Plant Biol 2002, 4:213–214.

5. Cardon G, Hohmann S, Klein J, Nettekens K, Saedler H, Huijser P. miR156-targeted and non-targeted SBP-box transcription factors act in concert to secure male fertility in Arabidopsis. Plant Cell 2010, 22:3935–3950.

6. Schwab R, Palatnik JF, Crété P, Vaucheret H. Specific effects of microRNAs on the plant transcriptome. Dev Biol 1990, 141:220–232.

7. Moreno MA, Harper LC, Krüger RW, Dellaporta SL, Freeling M. The SPL11 gene acts tissue specifically in maize leaf development. Dev Biol 2003, 269:217–226.

8. Tsukada A, Seki M, Shinozaki K, Yokoyama S. miR156 recognizes the ABRE motif in the 5′-flanking region of Arabidopsis SPL gene promoters. Plant Cell Physiol 2008, 49:1316–1317.

9. Zhang L, Wu B, Zhao D, Li C, Shao F, Lu S. miR397a is a four-node microRNA regulating the expression of OsSPL14 and triggering rice flowering. Nucleic Acids Res 2011, 39:152–153.

10. Puzyr J, Karger A, Astelli M, Kramer EH. Deep annotation of Populus trichocarpa miRNAs using diverse tissue sets. PloS One 2012, 7(10):e48304.

11. Lu S, Li Q, Wei H, Liu J, Song J, Sun Y, Yuan L, Yeh TF, Peszlen I, Ralph J, Sederoff RR, Chang VL. Ptrl-miR397a is a negative regulator of laccase genes affecting lignin content in Populus trichocarpa. Proc Natl Acad Sci USA 2013, 110(18):7489–7494.

12. Lu S, Yang C, Chiang VL. miR397a is a negative regulator of laccase genes affecting lignin content in Populus trichocarpa. Proc Natl Acad Sci USA 2013, 110(18):7489–7494.
