Advances in Bioresearch
Adv. Biores., Vol 15 (4) July 2024:223-236

Print ISSN 0976-4585; Online ISSN 2277-1573
Journal's URL:http://www.soeagra.com/abr.html
CODEN: ABRDC3
DOI: 10.15515/abr.0976-4585.15.4.213222

Advances in Bioresearch

ORIGINAL ARTICLE

In Silico Evolutionary Analysis, Identification of True Ortholog of Rice *SQUAMOSA* Promoter Binding Protein-Like 14 (SPL14) Gene between Monocots and Dicots

Mamta Sahani[1, 2] , R.K. Sharma[1], Arun Prasad Chopra[2]
[1]Department of Biotechnology and Life sciences Institute of Biomedical Education and Research, Mangalayatan University Beswan Aligarh Uttar Pradesh
[2]Department of Biotechnology, Hindustan College of Science and Technology, Sharda Group of Institutions (SGI) Farah Mathura Uttar Pradesh
***Author for correspondence's E-mail**: mamtasahanibt@gmail.com

ABSTRACT
SQUAMOSA Promoter Binding Protein-Like 14 (SPL14) is a plant-specific transcription factor (TF), which plays a vital regulatory role in plant growth and development. A higher expression of SPL14 at the reproductive stage in rice promotes shoot branching leading to higher grain yield. Using well characterized rice SPL14 gene as a reference, true orthologs were identified for the following ten monocots including wheat homeologs and six dicots: (i) Monocots; Oryza sativa japonica, Triticum aestivum, Zea mays, Triticum urartu, Aegilops tauschii, Hordeum vulgare, Brachypodium distachyon, Sorghum bicolor); (ii) Dicots: Arabidopsis thaliana, Glycine max, Prunus persica, Medicago trancatula, Vitis venifera and Populus trichocarpa. For each species, a single ortholog was available, except in wheat, where as expected three orthologues (one for each genome) were identified. The similarity of CDS sequence of each orthologs with reference gene ranged from 69.95 % to 74.21% in monocots and from 42.45% to 51.1% in dicots. Similarly, the similarity of each ortholog at gene sequence level ranged from 55.13% to 65.95% in monocots and from 36.37% to 50.18% in dicots. The gene from Arabidopsis thaliana was the smallest (1793 bp) and that from Hordeum vulgare was the longest (5188 bp). This difference in the gene size was largely due to differences in the size of the introns. Each ortholog had three exons, in monocots and dicots except Glycine max, where exon second got split in two exons by intrusion of intron were identified. All the three types of intron phases (0, 1, and 2) were found in all orthologs. The non-synonymous to synonymous ratio for the SPL14 genes was 0.5 in monocots and 0.11 in dicots suggesting purifying selection. The similarity in protein sequences ranged from 61.79% to 67.69% in monocots and from 34.24% to 45.98% in dicots. Proteins encoded by SPL14 gene contained a highly conserved SBP domain with specific conserved motifs. Several cis-acting regulatory elements were identified in the SPL14 promoters with their possible role in regulating spatial and temporal expression. Phylogenetic analysis reveals separate clusters between monocots and dicots. In Silico expression analysis revealed maximum expression at booting stage in all three genomes of wheat.
Keywords: *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 14, evolutionary analysis, gene structure, transcription factor, expression analysis, promoter analysis.*

Received 12.05.2024 Revised 09.06.2024 Accepted 09.07.2024

How to cite this article:
Mamta Sahani, R.K. Sharma, Arun Prasad Chopra. *In Silico* Evolutionary Analysis, Identification of True Ortholog of Rice *SQUAMOSA* Promoter Binding Protein-Like 14 (SPL14) Gene Between Monocots And Dicots: An Original Article Adv. Biores. Vol 15 [4] July 2024. 223-236

INTRODUCTION

Squamosa Promoter Binding Protein-like (SPL) genes encode plant specific transcription factors that play important role in plant growth, plant architecture, vegetative to reproductive phase transition, determining important agronomic traits and controlling stress responses in plants. These transcription factors are encoded by evolutionary conserved genes which have evolved and diversified with time. SPL genes have been extensively discovered in many terrestrial plants from green algae, *Chlamydomonas*, moss, to silver birch, *Arabidopsis*, rice, maize comprising of monocots to dicots. Orthologs and paralogs are present across the plant species found in different plant species from lower to higher organisms and plants have been shown to control wide variety of gene regulation by binding to the promoter region of

the genes determining specific plant function. Any perturbation in the levels of *Squamosa* TFs affects the plant architecture and final yield of the plant.

Transcription factors (TFs) are DNA-binding proteins that regulate gene expression at the level of mRNA transcription. They are capable of activating or repressing transcription of multiple target genes [36,52]. *Squamosa* Promoter Binding Protein-Like (SPL) box genes encode plant specific transcription factors (TFs) that play important role during the plant growth and development. Right from the germination, root shoot development, vegetative to reproductive phase transition, flowering and fruit development, *Squamosa* family of genes play very critical role in regulating overall plant architecture. Major important agronomic traits and stress responses are controlled by these transcription factors.

Squamosa Promoter-Binding Protein Like (SPL) genes encode plant specific transcription factors that play important role in plant growth and development. SBP box gene family belongs to the type of plant-specific zinc finger protein genes, which encode putative plant-specific transcription factor. SBP-box genes were first identified in *Antirrhinum majus* for the capacity of their protein products to bind to promoter region of the floral meristem identity gene *SQUAMOSA* [20]. Since then, SBP- box genes had been found in diverse plant species and their functions had been extensively investigated [3,10,39,24]. Birkenbihl et al. [7] found that there were 16 SBP-box genes in the model species Arabidopsis thaliana, named as SPL1 to SPL16 [6]. Their protein products could bind specifically to the related motifs in the promoter of AP1 (the orthologous gene in *Arabidopsis* of *SQUAMOSA*). SPL3 was found highly expressed in vegetative and inflorescence apex, floral meristem, leaf and floral primordial [10]. SPL3, SPL4 and SPL5 showed dramatically up-regulated in response to long day floral induction [37]. And SPL8 showed to affect pollen sac development [45]. One of the largest SBP-box genes, SPL14, was recently characterized as conferring resistance to the programmed cell death (PCD)-inducing fungal toxin fumonisin B1 (FB1) [42]. Another two SBP-box genes were isolated from maize, known as LIGULELESS1 (LG1) and TEOSINTE GLUME ARCHITECTURE (TGA1) [48]. Becraft et al [3]. reported that in the absence SQUAMOSA family of transcription factors are exclusively found in plants and play major role in plant growth, differentiation and development, determining important agronomic traits and stress responses in plants. These transcription factors are encoded by evolutionary conserved genes which have evolved and diversified with time. Orthologs and paralogs found in different plant species from lower to higher organisms and plants have been shown to control wide variety of gene regulation by binding to the promoter region of the genes determining specific plant function. *SQUAMOSA* transcription factors have been extensively characterized in crop plants.

Role of rice SPL14 have been recently implicated in ideal plant architecture including low tiller numbers with fewer unproductive tillers, stronger culms, higher panicle branching, more grains per panicle and higher grain productivity.

In-depth knowledge regarding the diversity and variation in the structure, function, and evolution of the underlying genes among different plant species is lacking, despite the abundance of information about SPL14 genes that is currently available. As a result, the current study identified and characterized the "true" orthologs of 16 plant species, including model and cultivated species as well as two wheat progenitors (*Triticum urartu* and *Aegilops tauschii*), using well-characterized rice SPL14 genes as a reference. The following findings from 16 species—including rice—are reported in this communication:

(i) finding the "true" orthologs of rice SPL14 genes in 16 species; (ii) evolution of gene structure for both subunits among the dicots and monocots; (iii) finding SPL14 protein motifs and domains that are essential for the enzyme's proper functioning and the subtle differences between and within SPL14 of monocots and dicots; (iv) promoter and expression analysis, revealing novel features that may have an impact on the temporal and spatial expression of SPL14 genes. Also covered is the potential application of this knowledge for crop improvement.

MATERIAL AND METHODS

Identification of true orthologs of SPL14

The rice *OsSPL14* transcription factor has been demonstrated to be important for plant architecture, phase transition, and grain productivity. For this reason, the rice full length SPL14 protein sequence was chosen as a reference in TBlastN to find "true" orthologs for sixteen plant species in eight monocots (*Oryza sativa japonica,* (*Triticum aestivum, Triticum urartu, Aegilops tauschii*), (*Zea mays*), (*Hordeum vulgare*), (*Sorghum bicolor*), and (*Brachypodium distachyon*) and six dicots (*Arabidopsis thaliana, Glycine max, Vitis venifera*), *Prunus persica, Medicago trancatula*, and *Populus trichocarpa* under investigation. The process outlined in [11] was followed in order to identify "true" orthologs. In a nutshell, a sequence could not be referred to as a "true" ortholog unless it satisfied the following requirements: (a) The highest level of query coverage and sequence identity along the cDNA sequences (b); the requirement that

orthologous sequences have the same domains and motifs as the protein used as a reference sequence (c); and (d) the equivalency of comparative size and distance among orthologous sequences. Full length gene sequences were identified using the recovered gene sequences of eight monocots and six dicots from a variety of sources, including Ensembl plants (http://plants.ensembl.org/index.html) and the NCBI (http://www.ncbi.nlm.nih.gov).

Comparative Analysis of Gene Structure

By aligning the cDNA with the corresponding genomic sequence using the pairwise alignment of Emboss Needle (https://www.embossneedle.com), the exons and introns as well as the exon-intron junction of the predicted sixteen orthologs were identified. Using rice exons as a reference, the exon/intron boundaries, translation start and stop sites, and predicted orthologs in other species were also marked (shown with the same color as of corresponding rice exon). The location of an intron insertion in relation to a codon was used to mark the intron phase distribution. An intron insertion between two codons is indicated by a phase distribution of 0; an insertion occurs after the first codon nucleotide in phase 1 and after the second codon nucleotide in phase 2. The MEGA 7.0 software was utilized to calculate the Ka/Ks values, which indicate the extent of non-synonymous to synonymous substitutions in different orthologous sequences. The Jukes Cantor substitution model [18].

Phylogenetic Analysis

Based on the predicted protein sequences for all sixteen orthologs, the MEGA 7 program was used to determine the evolutionary relationships between sixteen distinct SPL14 orthologous genes in the taxa under study [44]. A bootstrap involving 1000 iterations was utilized to build an unrooted phylogenetic tree for the phylogenetic analysis using the neighbor-joining method of the distance matrix. The lengths are measured in substitutions of amino acids per site.

Promoter Analysis

In Silico expression analysis by using the "Genevestigator" microarray database was performed for the *SPL14* gene orthologs in 10 monocots including wheat homeologs and dicots.

Using the Ensembl plant database (https://plants.ensembl.org/index.html), 1500bp sequence upstream of the translation start site was obtained for each of the species. These sequences were used to find the regulatory elements using the PlantCARE software with default parameters (http://bioinformatics.psb. ugent.be/webtools/plantcare/html/).

Protein Sequence Analysis

The consensus sequence was produced by aligning the predicted SPL14 amino acid sequences of the sixteen species using "Clustal Omega.". Sequence comparisons between monocots and dicots were conducted using the amino acid sequence of rice as a reference. The amino acid that appeared at the position most frequently across all species was kept, and this produced the consensus sequence. As a stand-in, rice SPL14 amino acid was utilized when there was no agreement. The consensus sequence contained the sequence insertions that were found in one or more of the species. Next, a similarity analysis was performed by comparing each of the sixteen species' protein sequences with the consensus sequence that was produced (Fig.3). Each amino acid's similarity to the consensus sequence was represented on a scale ranging from zero to six for dicots and from one to ten for monocots. Where as a monocot score of one means the residue is found in only one species, a score of ten for monocots indicates a conserved amino acid at a specific position in all ten monocots, including wheat homeologs. The conservation of the residue in every species is indicated by a score of six for dicots. The CDD analysis (http://www . ncbi . nlm . nih . gov/Structure/cdd/wrpsb.cgi) was used to identify domains and motifs in the consensus protein sequence. The predicted protein sequences of all orthologs were determined by three different translate tools, Expasy (https://prosite.expasy.org/prosite.html),

3D Structure analysis of SPL14

Using the corresponding amino acid sequences, the 3D structures of the orthologs of SPL14 were produced. This was accomplished by using the Swiss-Model in automated mode. Using the following servers, the 3D structures for every species that were produced were confirmed through geometric and energetic methods: (i) Structure Analysis and Verification Server (SAVES) (http://nihserve r.mbi.ucla.edu/ SAVES) using (a) PROCHECK to determine the percentage of amino acids that are found in the preferred region in comparison to other regions [25]; (b) VERIFY3D to ascertain whether a three-dimensional atomic model is compatible with its own aa sequence [13] as well as (c) ERRAT, which analyzes the statistics of non-bonded interactions between various atom types [10]; (ii) the Swiss-Model server, which makes use of a structure assessment tool. The 3D structures of sixteen SPL14 orthologs were verified using the FATCAT server (http://fatcat.sanfordburnham.org/) by superimposing the 3D structures of SPL14 for each plant species on the 3D structures of rice SPL14 as a reference. A total score

value as well as a number of parameters including percent identity and similarity, root mean square deviation (RMSD), and others were examined.

The 3D structures of SPL14 orthologs were generated using respective amino acid sequences. To do this, Swiss-Model was used in an automated mode. The 3D structures for all the species thus generated were verified by both geometric and energetic means using the following servers: (i) Structure Analysis and Verification Server (SAVES) (http://nihserver.mbi.ucla.edu/SAVES) employing (a) PROCHECK to find out the relative proportion of amino acids, which fall in favoured region, relative to other regions [25]; (b) VERIFY3D to determine the compatibility of an atomic model (3D) with its own aa sequence [13] and (c) ERRAT to analyse the statistics of non-bonded interactions between different atom types [10]; (ii) Swiss-Model server using structure assessment tool. FATCAT server (http://fatcat.sanfordburnham.org/) was used to confirm the 3D structures of sixteen SPL14 orthologs by superimposing the 3D structures of SPL14 for each of the plant species on the 3D structures of rice SPL14 as reference. Various parameters like root mean square deviation (RMSD), percent identity and similarity, and an overall score value was analyzed.

Expression Analysis

The "Genevestigator" microarray database (https://genevestigator.com/gv/index.jsp) was used to conduct in silico expression analysis for the SPL14 gene orthologs for a chosen monocot.

Ligand Binding Site Analysis

Ligand binding sites were predicted in all orthologs from 3D ligand site predictor (http://www. sbg.bio.ic.ac.uk / 3dligandsite [50].

RESULTS

Identification of "True" Orthologs

Rice full length *SPL14* cDNA sequence was used as reference to obtain true ortholog in ten monocots including wheat homeologs and six dicots identified using the criteria mentioned in the material and methods section. Only those sequences were selected that showed E-value of zero or less along with highest level of query coverage and maximum level of sequence identity throughout the sequence length corresponding to the reference cDNA. Table 1 shows the, gene length, cDNA, CDS, full length Protein. In all the species examined, a single ortholog of *SPL14* was identified except for wheat, in wheat as expected three orthologs (one for each, A, B, and D genome) were identified. The similarity of cDNA sequences of each orthologs of rice SPL14 in 16 species including wheat homeologs with reference to rice SPL14 ranged from 59.92 to 71.34% in monocots and from 38.09 to 46.74% in dicots. Similarity of CDS sequence ranged from 69.95 to 74.21% in monocots and from 42.45 to 51.1% in dicots. However, the similarity level of gene sequences ranged from 55.13 to 65.95% in monocots and from 36.37 to 50.18% in dicots.

Table: 1 Details of cDNA, CDS, proteins and gene sequences for SPL14 in different monocots and dicots with respect to genes for rice SPL14

Species	Chromosome Location	Gene sequence		cDNA		CDS		Protein		Conserved Domain
		Length	IDENTITY (%)	Length	IDENTITY	Length	IDENTITY	Length	IDENTITY (%)	IDENTITY (%)
Oryza sativa japonica	**8**	4156	100	1624	100	1254	100	417	100	100
Triticum aestivum 7AS	**7A**	4634	55.13	1470	66.28	1161	70.66	386	63.51	86.67
Triticum aestivum7BS	**7B**	4544	55.86	1885	65.04	1158	70.86	385	62.6	85.33
Triticum aestivum 7DS	**7D**	4451	56.15	1746	65.67	1158	70.4	385	61.79	85.33
Triticum urartu	**7**	4670	56.98	1631	68.01	1161	73.16	386	65.52	89.4
Aegilops tauschii	**7**	4258	64.83	1540	65.15	1155	70.41	384	61.79	85.33
Zea mays	**1**	4216	62.9	2029	59.92	1212	69.95	403	64.74	79.73

Sorghum bicolor	**7**	4213	65.95	1871	65.65	1227	74.21	408	67.69	83.78
Brachypodium distachyon	**3**	4524	65.92	1176	71.34	1176	72.14	391	63.83	86.67
Hordeum vulgare	**7H**	5188	65.11	1610	63.98	1185	70.69	394	63.66	88
Arabidopsis thaliana	**3**	1793	45.64	1455	38.09	1065	42.45	354	34.24	73.33
Prunus persica	**G6**	4557	50.18	1608	45.56	1155	50.72	384	45.98	82.67
Vitis venifera	**8**	4538	50.18	1859	46.74	1140	51.1	379	43.64	82.67
Medicago trancatula	**7**	3949	37.42	1557	41.21	1014	46.73	337	43.13	78.67
Glycine max	**3**	4477	36.37	1850	42.79	1104	48.32	367	43.2	82.67
Populus trichocarpa	**16**	4244	36.88	1952	42.38	1149	48.27	382	44.61	85.33

Comparison of Gene structure

The gene structures for different species were compared, using the reference rice gene. Major differences were observed in the sequence lengths of the genes 1793bp to 5188bp. The gene was smallest in *Arabidopsis thaliana* with 1793bp in dicots whereas that *Hordeum vulgare* was the longest with 5188bp in monocot (Table 1).

This variation in gene length is mainly attributed to the difference in size of the introns, as was apparent from comparison of cDNA sequences, which lack introns 1176bp to 2029bp. The length of cDNA for genes for both monocots and dicots almost similar except *Brachpodium distacyon.*

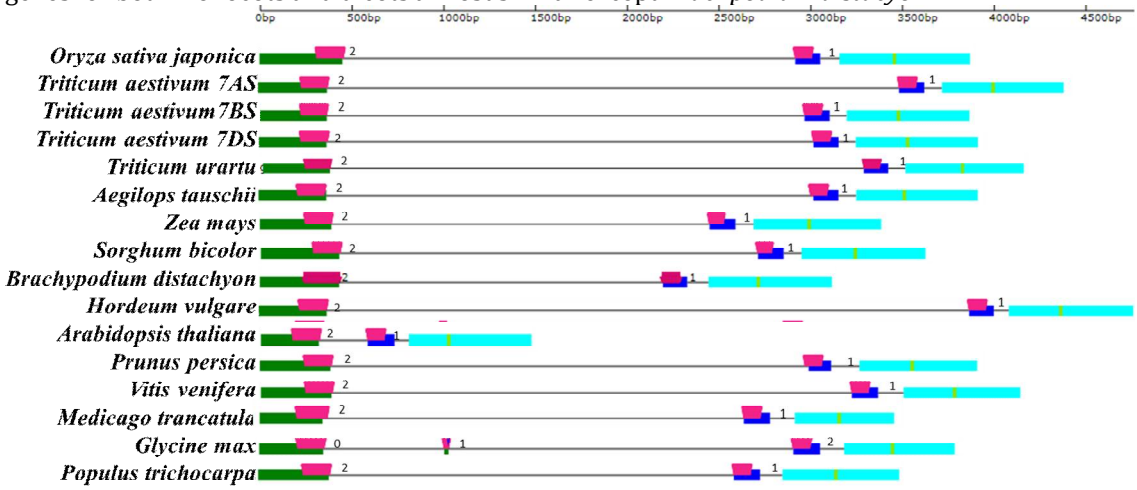

Fig.1 Structure of *SPL14* gene from eight monocots including wheat homeologs and six dicots species. Starting with the translational start site to stop site, the introns and exons were drawn to scale. The exons are represented by colored boxes and the introns as lines. Each of the three exons in rice was marked by different colors, and the corresponding sequences in other species were marked by the same color. Intron phases 0, 1 and 2 are marked above each intron. Pink colored wedges marked SBP domain which is highly conserved in selected monocots and dicots small green box in exon 3 mark miR156 targeted sites.

Each species had three exons and two introns except *Glycine max.* The length of intron 1 Comparison of introns and exons sequences (Fig.1) shows the presence and conservation of relative size and interval of each exon and intron among ten monocots including wheat homeologs and six dicots (exons depicted as color boxes with corresponding sequences in other species marked by same color boxes) in comparison to rice reference sequence. Each ortholog had three exons (exon 1, green color; exon 2, blue color; exon3 cyan color), except in *Glycine max* in dicots, where four exons were identified. Exon 2 got split in two exons by intrusion of intron in monocots. Length of exon 1 is significantly decrease by the deletion of some bases except *Brachypodium distacyon* and *Sorghum bicolor*, in monocots and in all six dicots in compared to reference refer rice gene. Exon 3 is conserved among in monocots and in dicot.

Two types of intron phases (, 1, and 2) were found in all orthologs, except Glycine max but in most of the orthologs intron phase 2 predominated between first and second exon and intronic phase 1

predominated between exon two and three. The gene for *Hordeum vulgare* was longest due the intron one was longest in monocots and the gene for Arabidopsis thaliana was smallest due to intron one was smallest in length comparatively to others. The SBP domain of 75 amino acids was highly conserved among monocots and dicots.

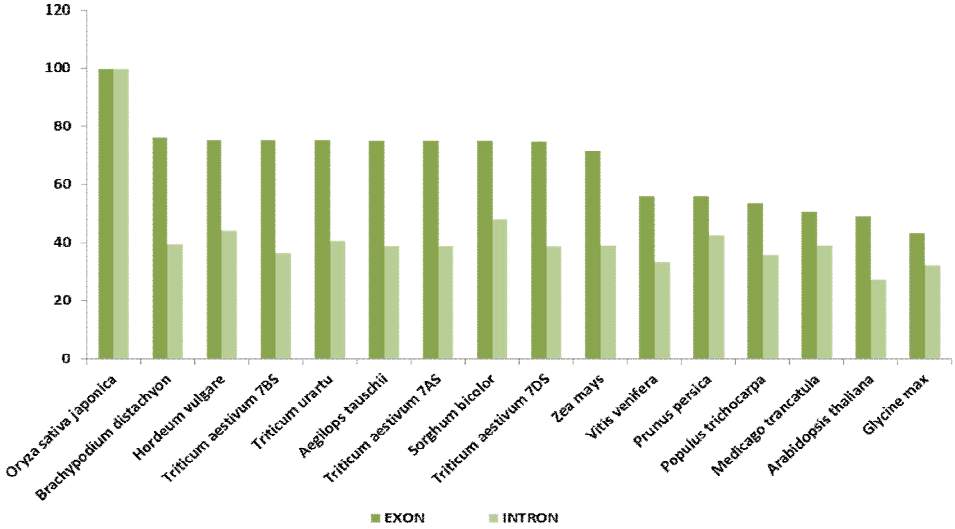

Fig. 2 Bar graph shows average percent similarity of exons and introns in monocots and dicots with reference to exons and introns of rice *SPL14* gene

The average percent sequence similarity for the SPL14 orthologous genes was low for both dicots 27.21 to 42.46% and monocots 36.62 to 47.69%, while it was higher for exons in monocots 71.68 to 76.56% than for introns 43.45 to 55.98%. %) in (Fig.2). For the SPL14 genes, the Ka/Ks values in monocots were higher (0.451) compared to the dicots (0.077).

Chromosome Assignment for SPL14 Gene

The chromosomes and the respective arms on which SPL14 gene are located are provided in Tables 1. *SPL14* gene were located on the short arms of wheat homoeologous group 7 (7AS, 7BS and 7DS) and the corresponding chromosome arms of other species

Protein Sequences and Structure

Comparison of Predicted Protein Sequences

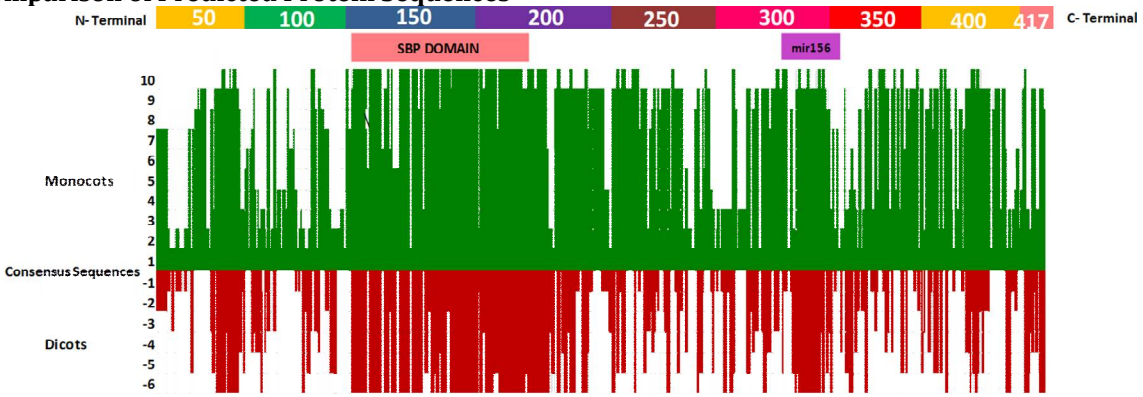

Fig.3 Amino acids sequence similarity of SPL14 among eight monocots (including wheat homologs on group 7 chromosomes) and six dicots with respect to consensus sequence. Position 0 (on y-axis) indicates amino acid consensus sequence. Presence of similar amino acids against consensus is plotted on a scale of 1–10 in monocots (green) and1–6 in dicots (red).

Th

e size of the predicted protein of SPL14 orthologs in16 species ranged from 354aa in *Arabidopsis* to 417aa in *Oryza sativa japonica*. In general, the average protein size among the monocots was 394 aa compared to 367 aa in the dicots (Table 1). The predicted protein sequence similarity of the dicots with the reference gene ranged from 61.79% for *Triticum urartu* to 67.69 % for *Sorgum bicolor* in monocots. The numbers for a similar comparison for the dicots were 34.24% for *Arabidopsis thaliana* to 67.69 % for

Prunus persica. Comparison of sequences among orthologs in highly conserved SBP domain reveled amino acid identity from 79.73 to 89.4% in monocots and 73.33 to 85.33% in dicots. Among all species examined the predicted SPL14 protein in *Sorghum bicolor* was longer than that of the other species. The SPL14 had one conserved domains (CDs), namely SBP domain with 75 aa is highly conserved in monocots and dicots. The N-terminal region (aa 1–97) was hypervariable, . Variation was mainly due to deletions, insertions and mismatches but position No. 34 to 46 (HGLKFGKKIYFED) was highly conserved in monocots shown in (Figure 3) position 79 to 84, at position 110 to 185 (75 amino acids) SBP domain was highly conserved.

In monocots, with respect to the consensus sequence, variations observed were classified in two categories (aa absent and additional aa present):

(a) Amino acids (aa) absent: (i) Glycine at position 6,7, 8,14, 15 16,18,19,21,22 and 23 Valine at position 17, alanine at position 9 to 13, serine at position 20 in *Zea mays*; (ii) alanine at position 9 to 13, valine at position 17 and 14,15,16, 18 in *Sorghum bicolor*; (iii) Glycine at position 15,16 and 18 and

(b) Valine at position 17 in monocots; (iv)alanine at position 9 to 13 and Glycine at postion 14 in *Triticum aestivum 7DS*, position 81 to 87, at position 92 to 98 in monocots

(c) Additional amino acids (aa) present: (i) alanine at position 5, serine at position 6 , lysine at position 33 in rice; (ii) alanine and glycine in rice, serine and cysteine in *Sorghum* and asparagine and valine in *Brachypodium* at positions 25 and 26, respectively.

(d) Mismatch aa at position 79 to 84 in *Zea mays* and *Sorghum* in monocots, 106 to 111 mismatches in monocots with reference to rice protein.

Promoter Analysis of the SPL14 Genes

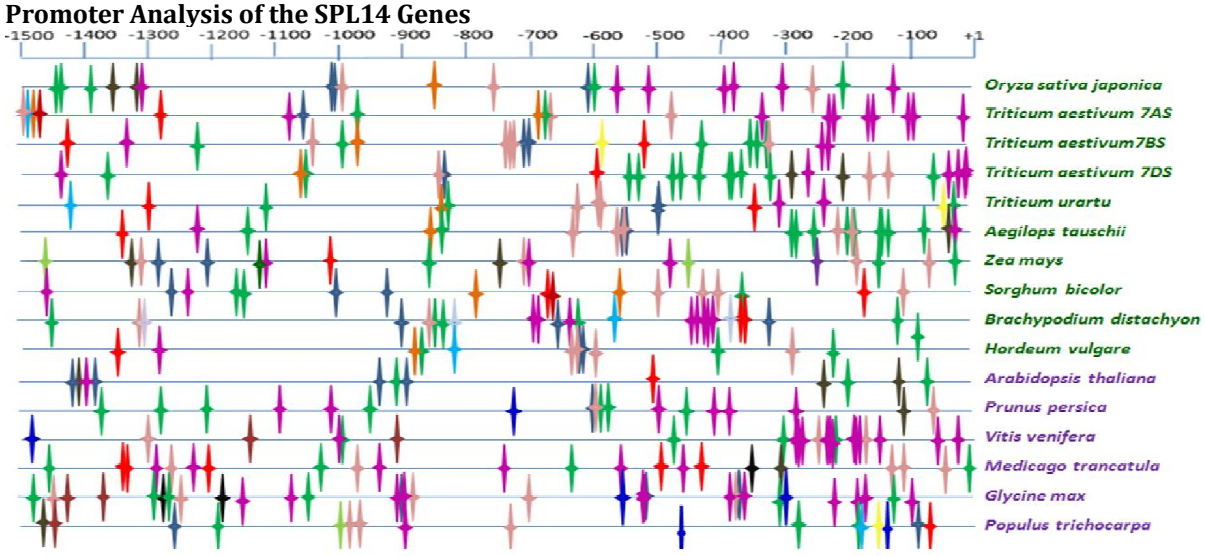

Fig.4 Representative figure showing regulatory elements identified in 1500 bp upstream region of SPL14 between monocots and dicots. Different colors motifs indicate major regulatory elements identified. TATA box, CAAT box , light responsive response elements , abiotic stresses responsive elements drought , cold , endosperm expression responsive elements , meristem expression responsive elements , auxin responsive element, gibberellin-responsive element, abscisic acid responsiveness , essential for the anaerobic induction , defense and stress responsiveness elements , element involved in differentiation of the palisade mesophyll cells , and MeJA-responsiveness

Promoter analysis allowed identification of cis-regulatory elements in the 1500bp upstream regions of *SPL14* genes. These regulatory elements (e.g., drought, cold) presumably respond to abiotic stresses and also to hormones like auxin, gibberellic acid and absicic acid. Regulatory elements responsible for tissue-specific expression (e.g., endosperm expression, meristem expression) and some other regulatory elements were also identified like MejA responsiveness, anaerobic induction, stress and defence responsiveness and also elements involved for the differentiation of palisade mesophyll cells those with unknown functions were also identified (Figure 4).For *SPL14* genes, regulatory elements including TATA box, CAAT box and light responsiveness elements were frequently present in all ten monocots including wheat homeologs and six dicots but drought related regulatory elements were present only *Triticum aestivum* 7AS, *Triticum urartu, Brachypodium distachyon* and *Hordeum vulgare* in monocots and *Populus trichocarpa* in dicots for cold stress related elements is only present in *Brachypodium distachyon*.

Phylogenetic analysis

Phylogenetic analysis was performed to establish evolutionary relationship among the orthologs and the results are given in (Fig.5). Phylogenetic analysis based on the aa sequences, the tree had reveals two

major separate clusters one for monocots and other for dicots, one with the ten monocots including wheat homoeologs, and the other with the six dicots. Among the monocots, *Triticum aestivum (7AS, 7BS,7DS)*, *Triticum urartu, Aegilops tauschii, Hordeum vulgare* and *Brachypodium distahyon* were grouped into the sub-cluster I, while *Zea mays, Sorghum bicolor* and *Oryza sativa japonica* were grouped in a separate sub-cluster. There were two sub-clusters in cluster II, with one containing *Glycine max, Medicago trancatula*, *Populus trichocarpa, Vitis venfera* and *Prunus persica* and the other containing *Arabidopsis thaliana*.

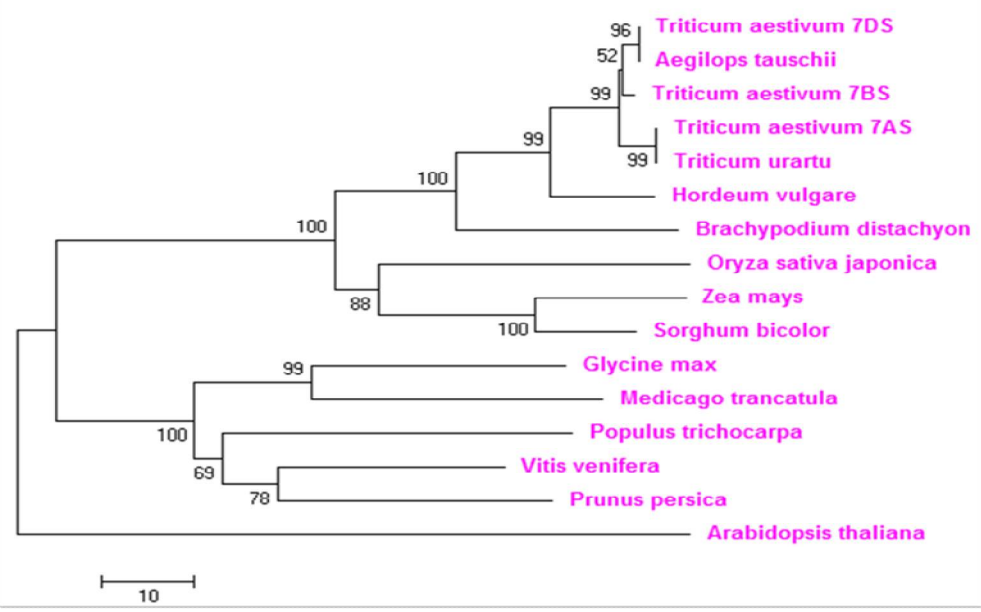

Fig. 5 Construction of phylogenetic tree by using neighbor-joining method using amino acid sequences of proteins encoded by genes for SPL14 to depict the relationship among monocots and dicots. The branch length represents magnitude of genetic change

Expression analysis of SPL14 genes

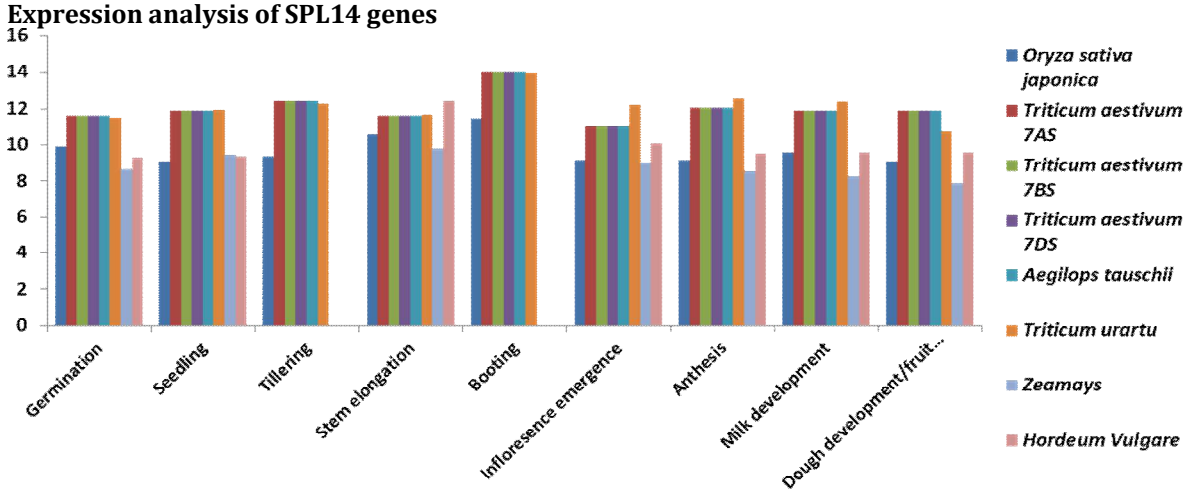

Fig.6A *In Silico* expression analysis of Rice *SPL14 gene* ortholog in seven monocots including Wheat homeologs (7AS. 7BS and 7DS) during different developmental stages.

Expression of SPL14 genes was examined using the microarray and transcriptome data, as mentioned in the section dealing with material and methods. The expression analysis based on microarray data indicated that the level of expression of genes was highest at booting in *Oryza sativa japonica, Triticum aestivum 7AS, 7BS, 7DS, Triticum urartu, Aegilops tauschii* and in case of *Zea mays* and *Hordeum vulgare* was highest at stem elongation stage, which declined during dough development and ripening stage in all of the eight cereals in monocots shown in (Fig. 6A), the expression was relatively low in vegetative stage.

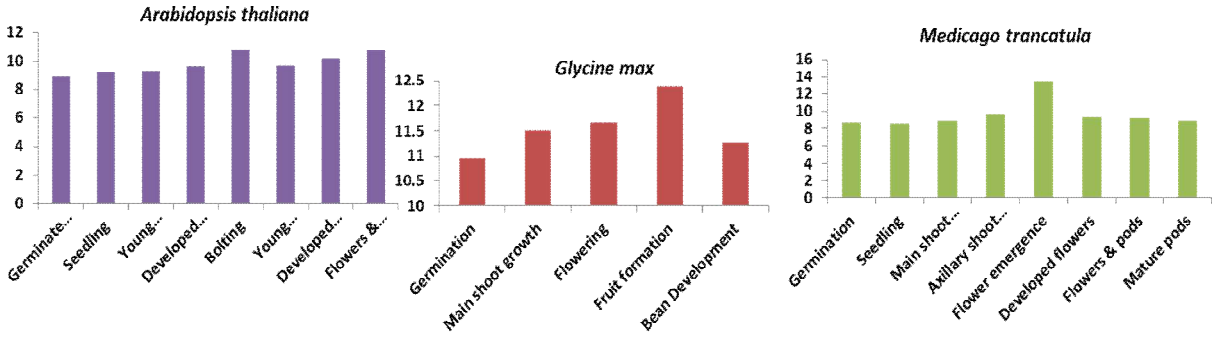

Fig.6B *In Silico* expression analysis of Rice *SPL14 gene* ortholog in three dicots during different developmental stages . A.*Arabidopsis thaliana* B. *Glycine max* C. *Medicago trancatula*

In *Arabidopsis thaliana,* maximum expression was observed during bolting stage, developed flowers and siliques formation stages which declined during senescence stage and in *Glycine max* the expression was observed at fruit formation stage and flower emergence stage in *Medicago trancatula* in dicots shown in (Fig.6B). Rice SPL14 orthologs genes among monocots and dicots showed developmental stages specific expression. In the booting stage, the expression of SPL14 ortholog in wheat homeologues on 7AS 7BS and 7DS was similar. Expression analysis was not done in *Sorghum bicolor* and *Brachypodium distachyon* in monocots and *Prunus persica, Vitis venifera* and *Populus trichocarpa* in dicots due to unavailability of data.

3D Protein Structure analysis

Based on the *Arabidopsis* SPL4 monomer template (PDB id: 1UL4), a high degree of confidence was found in the 3D structures that were generated for each of the sixteen species: (i) with respect to "allowed" and "disallowed" regions, a significant percentage of aa residues were in the preferred region of Ramachandran plots; (ii) the G-factor's overall value was within an acceptable range for all comparisons; (iii) the values of quality factors estimated by ERRAT and those of the 3D-1D score estimated by VERIFY3D were high; Qmean [4], which was based on the Swiss- Model, revealed negative energy values, indicating a favorable energy environment for the specific amino acids. This was further supported by the acceptable values of GMQE and QMEAN6.

Superimposition of 3D Protein Structures

The predicted values of different parameters obtained through superimposition of the 3D protein structure for each species was analysed. Pair-wise alignments of the 3D structures of rice SPL14 of each of the sixteen species with the corresponding structures in rice showed a higher level of similarity, which ranged from 75.3 to 85.2% in monocots,70.4 to 82.7% in dicots. Further, the 3D structures of SPL14 of each of the 16 species with the corresponding structures in rice showed a high level of similarity for ranging from 97 to 99.9% in monocots, 90 to 99.9% in dicots. The values for RMSD were 0.04 to 0.05 Å in monocots, 0.04 to 0.06 Å in dicots. The superimposition of predicted protein 3D structure of wheat over the rice SPL14 shown in (Fig. 9) with two different colors red color is rice protein structure and blue color is wheat protein structure with similarity of 99% rice to wheat and also wheat to rice protein structure both share same similarity.

Ligand Binding Site Analysis

The amino acids (aa) constituting the ligand binding sites were identified in all sixteen SPL14 orthologs (Fig.8). Most of the ligand binding sites were confined to the SBP domain . Generally, 1–2 clusters of ligands were predicted for the binding of Zn++ metallic heterogen. amino acid residues ranging from five in number but only four aa residues were involved in ligand binding, only one aa binding residue is different except for *Arabidopsis, Medicago trancatula* and *Populus trichocarpa* in dicots, where only four aa residues were involved in all 16 species that Cys-Cys-His-Cys. Among monocots and dicots, these four amino acid ligand binding residues are conserved and involved for ligand binding . These four ligand binding residues are highly conserved in all predicted rice SPL14 orthologs in sixteen species, ten for monocots including wheat homoeologs and six for dicots.

Conserved Motif Identification

A total of 9 conserved motifs for ten monocots including wheat homoeologs and six dicots were identified (Fig.7). The number of motifs varied in *Populus trichocarpa*, and most species rice SPL14 proteins shared similar motif profiles within the figure. The largest number of motifs found in SPL14 proteins was in monocots and dicots both, while there were a relatively small number of motifs in proteins belonging to *Populus trichocarpa*. Among these motifs, motif one was actually the SBP domain, which exists in all of the SPL14 proteins. The blue colour motif and cyan blue colour motif and majenta colour motif was not

existed in all of the Species four to five motifs less in *Populus trichocarpa.* These unique motifs might be tightly related to the specific roles of SPL14 orthologs in each species of monocots and dicots. There are maximum 13 motifs was conserved in *Triticum aestivum 7AS, 7BS, 7DS, Triticum urartu, Aegilops tauschii* in monocots and *Vitis venifera* in dicots, 12 motifs in Zea mays and 11 in *Sorghum bicolor* and *Brachypodium distachyon* and 10 in *Hordeum vulgare* in monocots 10 motifs conserved in dicots except *Populus trichocarpa* one different motif majenta colour motif was not present in monocots except *Hordeum vugare* but mostly present in dicots.

DISCUSSION

The present study demonstrated a high level of conservation of SPL14 gene not only at the protein level, but also at the gene level in both monocots and dicots. At the protein level, this conservation was evident through estimates of similarity, identity, coverage of a sequence, the presence of one conserved SBP domains and 3D structure analysis across all the sixteen species examined (ten monocots including wheat homeologs and six dicots). The data at the DNA level (gene sequence) suggested a higher level of conservation in exonic regions relative to that in introns, which is understandable, because introns are not involved in encoding the protein. The level of conservation was higher at the protein level comparatively gene levels in monocots and dicots both. The level of conservation of protein and gene level was higher in monocots than in dicots.

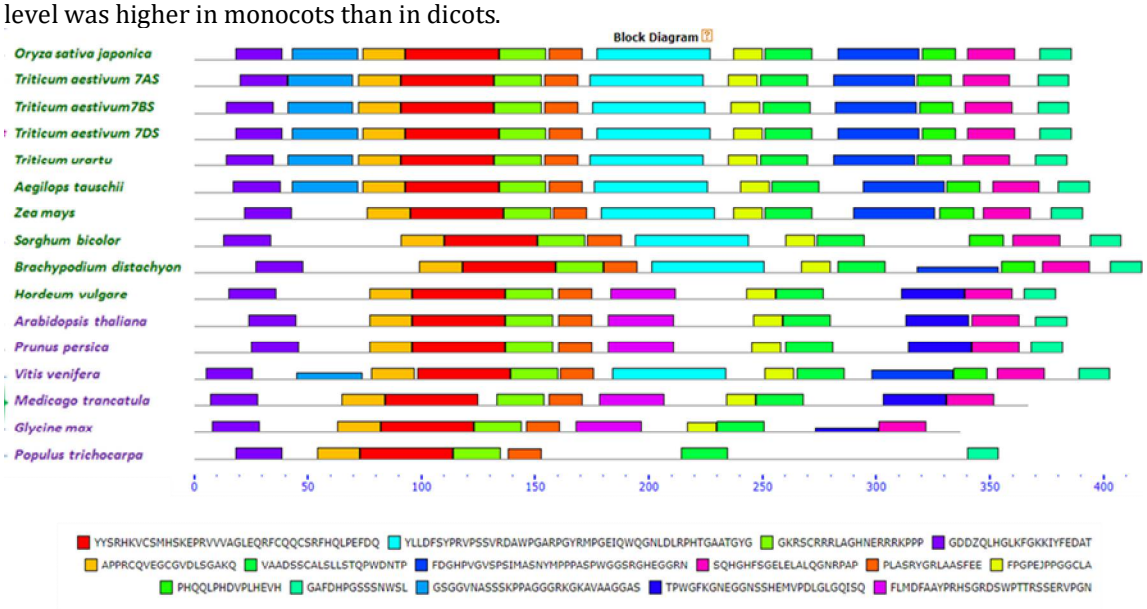

Fig. 7 Schematic representation of the 15 conserved motifs in SPL14 proteins orthologs. Motifs of the SPL 14 proteins were identified by MEME online tool. Each motif was represented by different coloured block. The different colors in boxes (1–15) represents motif 1 – motif 15, respectively. The position and length of each coloured box represents the actual motif size

Our analysis on intron phase gives further support to this conclusion. The prevalence of phase 2 and 1 in several genes was earlier attributed to primitive and conserved nature of the corresponding genes, since it allowed conservation of codons in the reading frame ; Figure 2. A comparison of intron length variation revealed that average introns one length was (3256 bp) generally longer in monocots than in dicots (2823 bp), intron 1 was longer than in intron 2 in monocot and dicot . Average length of intron 2 (75 bp) is slightly longer in monocots comparatively in dicots (62 bp) Available evidence also suggested that during the course of evolution, division of introns in SPL14 genes have given rise to longer introns in monocots. For, splitting of intron one in to two introns due to the exon two of (25 bp) in *Glycine max* of dicots and all other species exon 1, exon 2 and exon 3 is common with reference rice SPL14 gene.

The low GC content in the introns than in the exons of genes for SPL14, suggested that introns, intergenic regions and pseudogenes tend to have lower GC contents than ORFs. Differences in GC content allow discrimination between exons and introns and allow marking of exons for the splicing machinery [1]. A positive correlation between the size of introns/exons introns of SPL14 with the GC content, observed in the present study, was in agreement with similar correlation reported in case of plant genomes of rice and Arabidopsis and the genomes of fly, zebrafish and worms [55]

The sequences of genes for SPL14 of monocots and dicots differed, and seem to be more conserved in monocots at all four levels , gene, cDNA, CDS and protein but SBP domain is highly conserved in both

monocots and dicots 82.67 to 89.4% only *Zea mays* was 79.73% and in dicots 73.33% *Arabidopsis thaliana* and 78.67% *Medicago trancatula.*

The number of exons and introns in monocots largely remained constant during evolution except in case of *Glycine max* (4 exons). However, no variation seems to have occurred in monocots and dicots , there was intron loss in dicots (Roy and Gilbert, 2006), which receives support from the fact that the similarity of exons were more in monocots than in dicots (Fig.1). Higher variations in dicots also suggest a higher level of conservation in monocots than in dicots.

The superimposition of the structures of each species over the known *Arabidopsis* SPL4 (1UL4) [52] showed that the predicted structures are accurate and confirms identity of the true orthologs identified in the study. The state of all species 3D structure were monomer which is highly similar to monocots and dicots and also find two zinc finger motifs Cys-Cys-Cys-His and Cys- Cys-His-Cys and one Nuclear localization signal region(NLS) [6,19] in SBP domain which is highly conserved in selected species of monocots and dicots.

The N-terminal and C terminal region of the proteins of SPL14 was hypervariable in all species examined. The N-terminal regions of SPL14 in monocots and dicots. The SBP domain 75 amino acid is highly conserved in monocots and dicots and also two Zinc finger motifs one is Cys-Cys-Cys-His and Cys-Cys-His-Cys and one Nuclear localization signal region(NLS) was conserved in monocots and dicots. Arabidopsis and rice [52]. The DNA-binding domain, consisting of two a typical zinc-coordinating motifs, are thought to recognize a GTAC core motif [6]. The *miR156* region of six amino acid (ALSLLS) is also conserved in monocots and dicots. The SQUAMOSA PROMOTER BINDING PROTEIN-box direct binding core motif GTAC was highly enriched in *IPA1* binding peaks; interestingly, a previously uncharacterized indirect binding motif TGGGCC/T was found to be significantly enriched through the interaction of *IPA1* with proliferating cell nuclear antigen PROMOTER BINDING FACTOR 1 or PROMOTER BINDING FACTOR 2. Moreover, the results demonstrated that IPA1 could directly bind to the promoter of rice TEOSINTE BRANCHED1, a negative regulator of tiller bud outgrowth, to suppress rice tillering, and directly and positively regulate DENSE AND ERECT PANICLE1, an important gene regulating panicle architecture, to influence plant height and panicle length.

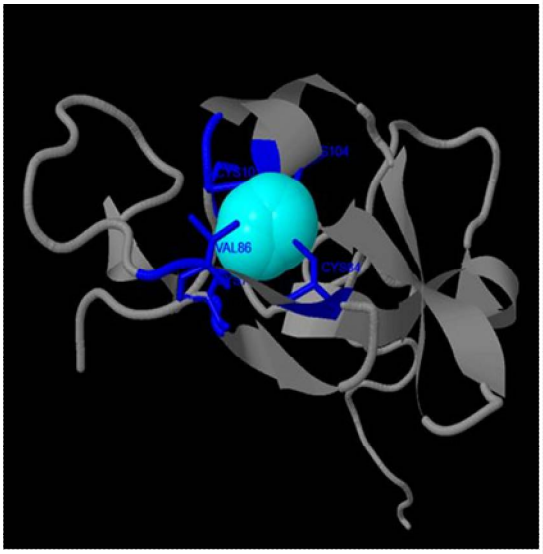

Fig.8 Predicted 3D structure of rice ortholog in wheat protein. The amino acids and their positions in the protein involved in ligand binding are shown in blue. cyan sphere represent metallic heterogen (Zn^{2+}) involved in ligand binding.

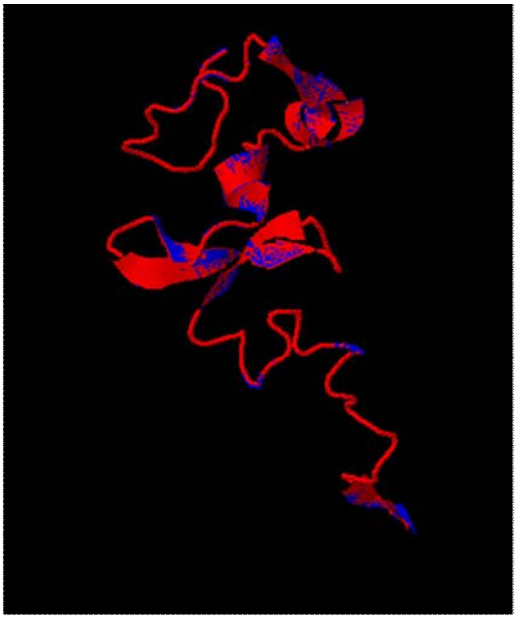

Fig.9 Superimposed 3D protein structure of predicted wheat ortholog over rice SPL14 .

The DNA-binding domain, consisting of two a typical zinc-coordinating motifs, are thought to recognize a GTAC core motif [6]. The *miR156* region of six amino acid (ALSLLS) is also conserved in monocots and dicots. The SQUAMOSA PROMOTER BINDING PROTEIN-box direct binding core motif GTAC was highly enriched in IPA1 binding peaks; interestingly, a previously uncharacterized indirect binding motif TGGGCC/T was found to be significantly enriched through the interaction of IPA1 with proliferating cell nuclear antigen PROMOTER BINDING FACTOR 1 or PROMOTER BINDING FACTOR 2. Moreover, the results demonstrated that IPA1 could directly bind to the promoter of rice *TEOSINTE BRANCHED1*, a

negative regulator of tiller bud outgrowth, to suppress rice tillering, and directly and positively regulate *DENSE AND ERECT PANICLE1*, an important gene regulating panicle architecture, to influence plant height and panicle length.

The elucidation of target genes of IPA1 genome-wide will contribute to understanding the molecular mechanisms underlying plant architecture and to facilitating the breeding of elite varieties with ideal plant architecture [17,31]. To meet the increasing demand for food, the concept of ideal plant architecture or new plant type has been proposed since the sixties of the last century [12,21]. Features of ideal plant architecture include low tiller number, few unproductive tillers, more grains per panicle, stronger culms, and robust roots [21,46]. *IDEAL PLANT ARCHITECTURE1 (IPA1)*, a pleiotropic gene isolated through a map-based cloning approach, has been shown to be one of the key regulators that determine plant architecture [16]. IPA1 encodes the protein *OsSPL14,* and in the ipa1 mutant, one nucleotide substitution located in the recognition site for microRNA156 (miRNA156) perturbs *IPA1* mRNA degradation, which results in accumulation of IPA1 and leads to the formation of ideal plant architecture with decreased tiller number and increased plant height and panicle branches [16]. *WEALTHY FARMER'S PANICLE*, another overexpression allele of *OsSPL14,* resulted from an epigenetic change in the OsSPL14 promoter and shows a similar phenotype [32]. Therefore, it has been suggested that OsSPL14 alleles have great potential for breeding [17,33]. However, how IPA1 affects plant architecture and whether any relationship exists between IPA1 and other plant architecture regulation genes remain largely unknown.

As one of the most important crops worldwide, rice has become a model plant of monocot species for functional genomics studies. Systematic analysis of SPL and *miR156* genes in rice will certainly improve our understanding of the complex regulatory networks in monocot species. Here we report on the analyses of SPL and *miR156* gene families in rice genome for their genomic organization, gene structures, motif composition, and expression levels in various tissues and organs of rice. Moreover, two *OsmiR156* precursors were overexpressed in rice to study the functional relationship of SPL and *miR156* genes.

CONCLUSION

This study determined the "true" orthologs of SPL14 genes and traced the evolution of the gene's structure and function in monocots and dicots. In summary, the comparative analysis demonstrated the structural conservation of rice SPL14, despite some variation in the length of the introns; the N-terminal regions exhibited comparatively greater variability. Of the sixteen species, only one SBP domain has been found to be highly conserved. This particular SBP domain has four ligand binding aa residues. Rice SPL14 gene: A comparative analysis reveals that the gene contains a conserved SBP domain that is crucial for better plant architecture and increased grain yield during the vegetative to reproductive phase transition. The primary sites of expression in cereals were found to be the endosperm and meristem, according to promoter and expression analyses.

REFERENCES.

1. Amit, M., Donyo, M., Hollander, D., Goren, A., Kim, E., Gelfman, S., Lev-Maor, G., Burstein, D., Schwartz, S., Postolsky, B. and Pupko, T., (2012). Differential GC content between exons and introns establishes distinct strategies of splice-site recognition. *Cell reports*, *1*(5), pp.543-556.
2. Batra, R., Saripalli, G., Mohan, A., Gupta, S., Gill, K.S., Varadwaj, P.K., Balyan, H.S. and Gupta, P.K., (2017). Comparative analysis of AGPase genes and encoded proteins in eight monocots and three dicots with emphasis on wheat. *Frontiers in Plant Science*, *8*, p.19.
3. Becraft, P.W., Bongard-Pierce, D.K., Sylvester, A.W., Poethig, R.S. and Freeling, M., (1990). The liguleless-1 gene acts tissue specifically in maize leaf development. *Developmental biology*, *141*(1), pp.220-232.
4. Benkert, P., Tosatto, S.C. and Schomburg, D., (2008). QMEAN: A comprehensive scoring function for model quality assessment. *Proteins: Structure, Function, and Bioinformatics*, *71*(1), pp.261-277.
5. Bennypaul, H.S., Mutti, J.S., Rustgi, S., Kumar, N., Okubara, P.A. and Gill, K.S., (2012). Virus-induced gene silencing (VIGS) of genes expressed in root, leaf, and meiotic tissues of wheat. *Functional & integrative genomics, 12*, pp.143-156.
6. Berkman, P.J., Skarshewski, A., Lorenc, M.T., Lai, K., Duran, C., Ling, E.Y., Stiller, J., Smits, L., Imelfort, M., Manoli, S. and McKenzie, M., (2011). Sequencing and assembly of low copy and genic regions of isolated *Triticum aestivum* chromosome arm 7DS. *Plant biotechnology journal,* 9(7), pp.768- 775.
7. Birkenbihl, R.P., Jach, G., Saedler, H. and Huijser, P., (2005). Functional dissection of the plant-specific SBP-domain: overlap of the DNA-binding and nuclear localization domains. *Journal of molecular biology*, *352*(3), pp.585-596.
8. Cao, R., Guo, L., Ma, M., Zhang, W., Liu, X. and Zhao, H., (2019). Identification and functional characterization of Squamosa promoter binding protein-like gene TaSPL16 in wheat (*Triticum aestivum* L.). *Frontiers in plant science*, *10*, p.212.

9. Cardon, G.H., Höhmann, S., Nettesheim, K., Saedler, H. and Huijser, P., (1997). Functional analysis of the Arabidopsis thaliana SBP-box gene SPL3: a novel gene involved in the floral transition. *The Plant Journal*, *12*(2), pp.367-377.
10. Colovos, C. and Yeates, T.O., (1993). Verification of protein structures: patterns of nonbonded atomic interactions. *Protein science*, *2*(9), pp.1511-1519.
11. Dhaliwal, A.K., Mohan, A. and Gill, K.S., (2014). Comparative analysis of ABCB1 reveals novel structural and functional conservation between monocots and dicots. *Frontiers in plant science*, *5*, p.657.
12. Donald, C.T., 1968. The breeding of crop ideotypes. *Euphytica*, *17*, pp.385-403.
13. Eisenberg, D., Lüthy, R. and Bowie, J.U., (1997). [20] VERIFY3D: assessment of protein models with three-dimensional profiles. In *Methods in enzymology* (Vol. 277, pp. 396-404). Academic Press.
14. Goff, S.A., Ricke, D., Lan, T.H., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H. and Hadley, D., (2002). A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). *Science, 296*(5565), pp.92-100.
15. Gupta, P.K., Langridge, P. and Mir, R.R., (2010). Marker-assisted wheat breeding: present status and future possibilities. *Molecular Breeding, 26*(2), pp.145-161.
16. Holzberg, S., Brosio, P., Gross, C. and Pogue, G.P., (2002). Barley stripe mosaic virus-induced gene silencing in a monocot plant. *The Plant Journal, 30*(3), pp.315-327.
17. Jiao, Y., Wang, Y., Xue, D., Wang, J., Yan, M., Liu, G., Dong, G., Zeng, D., Lu, Z., Zhu, X. and Qian, Q., (2010). Regulation of OsSPL14 by OsmiR156 defines ideal plant architecture in rice. *Nature genetics, 42*(6), p.541.
18. Jukes, T.H., 1969. Mammalian protein metabolism. *(No Title)*, p.21.
19. Kim, S.R., Ramos, J.M., Hizon, R.J.M., Ashikari, M., Virk, P.S., Torres, E.A., Nissila, E. and Jena, K.K., (2018). Introgression of a functional epigenetic OsSPL 1 4 WFP allele into elite indica rice genomes greatly improved panicle traits and grain yield. *Scientific reports*, *8*(1), p.3833.
20. Klein, J., Saedler, H. and Huijser, P., 1996. A new family of DNA binding proteins includes putative transcriptional regulators of the Antirrhinum majus floral meristem identity gene SQUAMOSA. *Molecular and General Genetics MGG*, *250*, pp.7-16.
21. Khush, G.S., (1995). Breaking the yield frontier of rice. *GeoJournal*, *35*, pp.329-332.
22. Kumar, J., Singh, S.P., Kumar, J. and Tuli, R., (2012). A novel mastrevirus infecting wheat in India. *Archives of virology, 157*(10), pp.2031-2034.
23. Lai, K., Berkman, P.J., Lorenc, M.T., Duran, C., Smits, L., Manoli, S., Stiller, J. and Edwards, D., (2012). Wheat Genome. info: an integrated database and portal for wheat genome information. *Plant and Cell Physiology, 53*(2), pp.e2-e2.
24. Lännenpää, M., Jänönen, I., Hölttä-Vuori, M., Gardemeister, M., Porali, I. and Sopanen, T., (2004). A new SBP-box gene BpSPL1 in silver birch (Betula pendula). *Physiologia plantarum*, *120*(3), pp.491-500.
25. Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M., (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of applied crystallography*, *26*(2), pp.283-291.
26. Lian, L., Xu, H., Zhang, H., He, W., Cai, Q., Lin, Y., Wei, L., Pan, L., Xie, X., Zheng, Y. and Wei, Y., (2020). Overexpression of OsSPL14 results in transcriptome and physiology changes in indica rice 'MH86'. *Plant Growth Regulation*, *90*, pp.265-278.
27. Liu, E. and Page, J.E., (2008). Optimized cDNA libraries for virus-induced gene silencing (VIGS) using tobacco rattle virus. *Plant methods, 4*(1), pp.1-13.
28. Liu, J., Cheng, X., Liu, P. and Sun, J., (2017). miR156-targeted SBP-box transcription factors interact with DWARF53 to regulate TEOSINTE BRANCHED1 and BARREN STALK1 expression in bread wheat. *Plant physiology, 174*(3), pp.1931-1948.
29. Li, L., Shi, F., Guan, Y., Wang, G., Zhang, Y., Chang, J., Chen, M., Yang, G., Wang, Y. and He, G., 2020. Genome-wide analysis of structures and characteristics of the SPL gene family in wheat (*Triticum aestivum* L.). Research Square: http://dx.doi.org/10.21203/rs.3.rs-88143/v1.
30. Luo, L., Li, W., Miura, K., Ashikari, M. and Kyozuka, J., (2012). Control of tiller growth of rice by OsSPL14 and strigolactones, which work in two independent pathways. *Plant and Cell Physiology*, *53*(10), pp.1793-1801.
31. Lu, Z., Yu, H., Xiong, G., Wang, J., Jiao, Y., Liu, G., Jing, Y., Meng, X., Hu, X., Qian, Q. and Fu, X., (2013). Genome-wide binding analysis of the transcription activator ideal plant architecture1 reveals a complex network regulating rice plant architecture. *The Plant Cell*, *25*(10), pp.3743-3759.
32. McSteen, P., (2009). Hormonal regulation of branching in grasses. *Plant Physiology, 149*(1), pp.46-55.
33. Miura, K., Ikeda, M., Matsubara, A., Song, X.J., Ito, M., Asano, K., Matsuoka, M., Kitano, H. and Ashikari, M., (2010). OsSPL14 promotes panicle branching and higher grain productivity in rice. *Nature genetics, 42*(6), pp.545-549.
34. Mohan, A., Dhaliwal, A.K., Nagarajan, R. and Gill, K.S., (2019). Molecular characterization of Auxin Efflux Carrier-ABCB1 in hexaploid wheat. *Scientific reports, 9*(1), pp.1-14.
35. Randhawa, H.S., Mutti, J.S., Kidwell, K., Morris, C.F., Chen, X. and Gill, K.S., (2009). Rapid and targeted introgression of genes into popular wheat cultivars using marker-assisted background selection. PloS one, 4(6), p.e5752.
36. Riechmann, J.L., Heard, J., Martin, G., Reuber, L., Jiang, C.Z., Keddie, J., Adam, L., Pineda, O., Ratcliffe, O.J., Samaha, R.R. and Creelman, 8., (2000). Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. *science*, *290*(5499), pp.2105-2110.

37. Schmid, M., Uhlenhaut, N.H., Godard, F., Demar, M., Bressan, R., Weigel, D. and Lohmann, J.U., (2003). Dissection of floral induction pathways using global expression analysis.
38. Sethi, S., Saini, J.S., Mohan, A., Brar, N.K., Verma, S., Sarao, N.K. and Gill, K.S., (2016). Comparative and evolutionary analysis of α-amylase gene across monocots and dicots. *Functional & integrative genomics*, *16*, pp.545-555.
39. Shao, C.X., Takeda, Y., Hatano, S., Matsuoka, M. and Hirano, H.Y., (1999). Rice genes encoding the SBP domain protein, which is a new type of transcription factor controlling plant development. *Rice Genet Newsl*, *16*, p.114.
40. Singh, K., Singh, J., Jindal, S., Sidhu, G., Dhaliwal, A. and Gill, K., (2019). Structural and functional evolution of an auxin efflux carrier PIN1 and its functional characterization in common wheat. *Functional & integrative genomics, 19*(1), pp.29-41.
41. Sorabh Sethi, S.S., Saini, J.S., Amita Mohan, A.M., Brar, N.K., Shabda Verma, S.V., Sarao, N.K. and Gill, K.S., (2016). Comparative and evolutionary analysis of α-amylase gene across monocots and dicots.
42. Stone, J.M., Liang, X., Nekl, E.R. and Stiers, J.J., (2005). Arabidopsis AtSPL14, a plant-specific SBP-domain transcription factor, participates in plant development and sensitivity to fumonisin B1. *The Plant Journal*, *41*(5), pp.744-754.
43. Takeda, S. and Matsuoka, M., (2008). Genetic approaches to crop improvement: responding to environmental and population changes. *Nature Reviews Genetics, 9*(6), pp.444-457.
44. Tamura, K., Battistuzzi, F.U., Billing-Ross, P., Murillo, O., Filipski, A. and Kumar, S., (2012). Estimating divergence times in large molecular phylogenies. *Proceedings of the National Academy of Sciences*, *109*(47), pp.19333-19338.
45. Unte, U.S., Sorensen, A.M., Pesaresi, P., Gandikota, M., Leister, D., Saedler, H. and Huijser, P., (2003). SPL8, an SBP-box gene that affects pollen sac development in Arabidopsis. The Plant Cell, 15(4), pp.1009-1019.
46. Virk, P.S., Khush, G.S. and Peng, S., (2004). Breeding to enhance yield potential of rice at IRRI: the ideotype approach.
47. Wang, H., Nussbaum-Wagler, T., Li, B., Zhao, Q., Vigouroux, Y., Faller, M., Bomblies, K., Lukens, L. and Doebley, J.F., (2005). The origin of the naked grains of maize. *Nature*, *436*(7051), pp.714-719.
48. Wang, L. and Zhang, Q., (2017). Boosting rice yield by fine-tuning SPL gene expression. *Trends in Plant Science*, *22*(8), pp.643-646.
49. Wang, Y. and Li, J., (2008). Molecular basis of plant architecture. Annu. Rev. *Plant Biol., 59*, pp.253-279
50. Wass, M.N. and Sternberg, M.J., (2009). Prediction of ligand binding sites using homologous structures and conservation at CASP8. *Proteins: Structure, Function, and Bioinformatics*, *77*(S9), pp.147-151.
51. Xie, K., Wu, C. and Xiong, L., (2006). Genomic organization, differential expression, and interaction of SQUAMOSA promoter-binding-like transcription factors and microRNA156 in rice. *Plant physiology, 142*(1), pp.280-293.
52. Xiong, Y., Contento, A.L. and Bassham, D.C., (2005). AtATG18a is required for the formation of autophagosomes during nutrient stress and senescence in Arabidopsis thaliana. *The Plant Journal*, *42*(4), pp.535-546.
53. Yamasaki, K., Kigawa, T., Inoue, M., Tateno, M., Yamasaki, T., Yabuki, T., Aoki, M., Seki, E., Matsuda, T., Nunokawa, E. and Ishizuka, Y., (2004). A novel zinc-binding motif revealed by solution structures of DNA-binding domains of Arabidopsis SBP-family transcription factors. *Journal of molecular biology*, *337*(1), pp.49-63.
54. Zhang, B., Liu, X., Zhao, G., Mao, X., Li, A. and Jing, R., (2014). Molecular characterization and expression analysis of Triticum aestivum squamosa-promoter binding protein-box genes involved in ear development. *Journal of integrative plant biology, 56*(6), pp.571-581.
55. Zhu, L., Zhang, Y., Zhang, W., Yang, S., Chen, J.Q. and Tian, D., (2009). Patterns of exon-intron architecture variation of genes in eukaryotic genomes. *BMC genomics*, *10*, pp.1-12.