ELSEVIER

Gene 237 (1999) 91–104

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

www.elsevier.com/locate/gene

Molecular characterisation of the *Arabidopsis* SBP-box genes

Guillermo Cardon [1], Susanne Höhmann, Joachim Klein [2], Klaus Nettesheim, Heinz Saedler, Peter Huijser *

Max-Planck-Institut für Züchtungsforschung, Department of Molecular Plant Genetics, Carl-von-Linné-Weg 10, 50829 Cologne, Germany

Received 2 May 1999; received in revised form 28 June 1999; accepted 9 July 1999; Received by W. Martin

Abstract

The *Arabidopsis thaliana SPL* gene family represents a group of structurally diverse genes encoding putative transcription factors found apparently only in plants. The distinguishing characteristic of the *SPL* gene family is the SBP-box encoding a conserved protein domain of 76 amino acids in length, the SBP-domain, which is responsible for the interaction with DNA. We present here characterisation of 12 members of the *SPL* gene family. These genes show highly diverse genomic organisations and are found scattered over the *Arabidopsis* genome. Some *SPL* genes are constitutively expressed, while transcriptional activity of others is under developmental control. Based on phylogenetic reconstruction, gene structure and expression patterns, they can be divided into subfamilies. In addition to the *Arabidopsis SPL* genes, we isolated and determined the sequences of three SBP-box genes from *Antirrhinum majus* and seven from *Zea mays*.

Keywords: DNA-binding protein; Gene family; Phylogeny; Plants; *SPL* genes

1. Introduction

Precise and coordinate gene expression forms the basis of growth and development. Of the many mechanisms operating to control gene expression, regulation of gene transcription is the primary one. Transcriptional control relies on transcription factors, proteins that are imported to the nucleus. Through their interaction with gene regulatory sequences, often in complexes with other specific or ubiquitous proteins, they modulate transcription. Recent isolation and characterisation of several putative and bonafide transcription factors in plants have shown that these proteins possess a variety of DNA binding motifs. Some of them are related to motifs also present in yeast and animals, while others seem to be unique to plants.

SBP1 and SBP2, two putative transcription factors of a novel type, were recently isolated from *Antirrhinum majus* by their capacity to interact in vitro with a promoter sequence element of the floral meristem identity gene *SQUAMOSA* (*SQUA*; Klein et al., 1996; Huijser et al., 1992). Both proteins share a region of high sequence similarity, the SBP-domain encoded by the SBP-box. The SBP-domain was found to be necessary and sufficient for DNA binding. The presence of several cross-hybridising bands in *Antirrhinum* genomic Southern blots hybridised at low stringency with the corresponding *SBP1* and *SBP2* genes suggested that there should be other SBP-box genes in the *Antirrhinum* genome.

In order to study the role of SBP-box genes in plant development, we initiated a search for their homologues in *Arabidopsis thaliana*, a model system more suitable and highly developed for molecular genetic studies. This approach led to the isolation of several members of the *Arabidopsis* SBP-box gene family, the *SPL* genes (*SQUA promoter-binding protein-like*; Cardon et al., 1997). Among them, *SPL3* was the best candidate for being the *SBP1* orthologue. The early flowering phenotype of transgenic *Arabidopsis* plants constitutively expressing *SPL3* suggests a role for *SPL3* in the control of the floral transition (Cardon et al., 1997). On the other

Abbreviations: BAC, bacterial artificial chromosome; CAPS, cleaved amplified polymorphic sequence; EST, expressed sequence tag; PCR, polymerase chain reaction; RFLP, restriction fragment length polymorphism; UTR, untranslated region.

* Corresponding author. Tel.: +49-221-5062-170; fax: +49-221-5062-113.

E-mail address: huijser@mpiz-koeln.mpg.de (P. Huijser)

[1] Present address: Plant Genetic Systems N.V./AgrEvo, Gent, Belgium.

[2] Present address: University of Frankfurt, Frankfurt a.M., Germany.

0378-1119/99/$ – see front matter
PII: S0378-1119(99)00308-X

hand, genetic studies showed that this phenotype is not due to ectopic activation of *APETALA1* (*AP1*; Mandel and Yanofsky, 1995), the presumed *Arabidopsis SQUA* orthologue (Mandel et al., 1992).

The notion that SBP-box genes might play a role in the control of plant development was further strengthened by the discovery that the molecular basis for the Liguleless1 mutant phenotype in maize is a mutation of an SBP-box gene (Moreno et al., 1997). Mutations in the *LG1* gene lead to the development of maize leaves lacking ligules and auricules (Becraft et al., 1990). Moreover, the isolation of *LG1* showed that SBP-box genes are not limited to dicots such as *Antirrhinum* and *Arabidopsis*.

In this work we describe the characterisation of 12 members of the *Arabidopsis SPL* gene family of SBP-box genes. Members of the *SPL* gene family are diverse in expression and structure. They form subfamilies with sequence similarity outside the SBP-domain. In addition, we isolated further SBP-box genes from *Antirrhinum* and maize and present a first study of the phylogenetic relationship of the known members of this new family of putative plant transcription factors.

2. Materials and methods

2.1. Screening of cDNA and genomic libraries

SBP-box genes from *Arabidopsis thaliana* were isolated from a λNM1149 (Murray, 1983) cDNA library prepared from inflorescence mRNA from the ecotype Columbia (Col) (a gift of H. Sommer, MPI, Cologne). Screening was done at low stringency using the cDNAs of the *Antirrhinum SBP1* and *SBP2* genes as probes. Inserts of positive clones were subcloned as *Eco*RI fragments into pUC19.

Additional *SPL* genes were isolated from a λZAP II (Stratagene) cDNA library from the ecotype Landsberg erecta (Ler) prepared by D. Weigel and E. Meyerowitz (obtained from the *Arabidopsis* DNA Stock Centre at MPI, Cologne). Positive clones from the λZAP II library were rescued with ExAssist helper phage (Stratagene).

The Smart™ PCR cDNA Synthesis Kit from Clontech was used to generate cDNA from ecotype Col inflorescence poly(A)$^+$ RNA. This cDNA pool was used to isolate via PCR cDNA clones in ecotype Col background of *SPL* genes initially isolated from the ecotype Ler λZAP II library. cDNAS from *SPL* genes first identified as genomic clones were also isolated from the Smart™ PCR cDNA pool by PCR with different gene specific primers.

Genomic clones of the *SPL* genes were isolated from a genomic library of the ecotype Col constructed in λGem11 (Promega) from J. Mulligan and R. Davis (distributed by the *Arabidopsis* DNA Stock Centre at MPI, Cologne). The sequence of the *SPL* genomic clones was determined directly from phage DNA using primers designed from the cDNA sequences.

Additional SBP-box genes from *Antirrhinum majus* were isolated from a λNM1149 cDNA library synthesised from poly(A)$^+$ RNA from young wild type inflorescences (H. Sommer, MPI, Cologne). The cDNAs of *SBP1* and *SBP2* were used as hybridisation probes. For sequencing, inserts from positive clones were PCR amplified using λNM1149 specific primers.

Maize SBP-box genes were isolated from a λZAP II cDNA library generated from poly(A)$^+$ RNA from young female inflorescences from line T232 (kindly provided by Dr. J. Cacharron, MPI, Cologne). Since hybridisation at low stringency with *Arabidopsis* SBP-box sequences was not successful, a 270 bp fragment of the SBP-box of the maize *LG1* gene (Moreno et al., 1997) was used as probe. This fragment was generated by PCR amplification from maize genomic DNA with the *LG1* primers 5′-CCA CCC TCG TCA GCT CCG CCG AGG AG-3′ and 5′-TCT GCT GCA CTG CTG GCA GAA CCT CTG-3′, and subcloned in pUC19 (*Sma*I site).

2.2. Oligonucleotide synthesis, DNA sequencing and sequence analysis

All oligonucleotides used for this work were synthesised by MWG-Biotech (Ebersberg, Germany) and GibcoBRL Life Technologies (Eggenstein, Germany).

DNA sequencing was performed either using the fmol DNA sequencing system (Promega) or an automated DNA sequencer (Applied Biosystems). DNA sequence analysis was done using the GCG analysis tools (Wisconsin Package version 9.1, Genetics Computer Group, Madison, WI). For phylogenetic analysis the pileup/plot program of GCG and the Phylogeny Inference Package (J. Felsenstein, PHYLIP version 3.5c, 1993, distributed by the author, Department of Genetics, University of Washington, Seattle) were used. Protein sequences were aligned with the help of MacVector™ (Oxford Molecular Group PLC, 1996).

Putative functional protein domains, outside the SBP-domain, were identified by screening the PROSITE database of the ExPASy molecular biology WWW server (http://expasy.hcuge.ch/) of the Swiss Institute of Bioinformatics (SIB).

2.3. Northern blot analysis

For the isolation of RNA, *Arabidopsis* plants from ecotype Col were grown in a growth chamber (Conviron CMP 3244) at 22°C, 50% relative humidity and $\approx 150\ \mu E\ m^{-2}\ s^{-1}$ light (fluorescent Sylvania F72T12 cold white light, 75%, and incandescent Sylvania 100 W lamps, 25%) on a long day photoperiod (LD, 16 h light,

8 h dark) or a short day photoperiod (SD, 8 h light, 16 h dark). Plants were grown in plastic trays filled with ready-to-use commercial, pre-fertilised soil mixture (Type ED73, Werkverband EV). RNA isolation, poly(A)$^+$ RNA separation, gel fractionation, blotting and hybridisation were performed as described by Cardon et al. (1997). The different *SPL* gene probes lacked the SBP-box to avoid cross-hybridisation.

2.4. *In situ hybridisation*

Sequences 5′ of the SBP-box of both the *SPL4* and *SPL5* genes allowed the preparation of probes discriminating both genes. Such fragments were PCR amplified from cloned genomic DNA using the following pair of oligos: 5′-CGT AAT CAT GCA CTT ATT CCA CCA G-3′ and 5′-CCA AGC TTC TAA TAC GAC TCA CTA TAG GGA GAT GCC TTT GCA TGA ACT TCA C-3′. The latter oligo contains a T7-RNA polymerase promoter sequence allowing the PCR products to be used as templates to prepare antisense RNA probes (Logel et al., 1992) discriminating the *SPL4* and *SPL5* genes. RNA probes were digoxigening-labelled according to the Boehringer Mannheim nucleic acid labelling kit and used for in situ hybridisation on sections of inflorescences of *A. thaliana* Col according to Huijser et al. (1992) and as modified by Samach et al. (1997).

2.5. *Mapping of the* SPL *genes*

SPL2 to *5*, *SPL7* and *SPL12* were mapped using the recombinant inbred (RI) lines from a cross between the ecotypes Col × Ler from Lister and Dean (1993; kindly provided by the Nottingham *Arabidopsis* Stock Centre). Mapping was carried out by scoring 65 RI lines. *SPL2* to *5* were mapped using conventional RFLP. The restriction endonucleases used were *Taq*I (*SPL2*), *Bgl*II (*SPL3*), *Hin*dIII/*Taq*I (*SPL4*) and *Hin*dIII (*SPL5*). The corresponding cDNAs were used as hybridisation probes. For the mapping of *SPL7* and *SPL12*, sequence polymorphisms between ecotypes Col and Ler were uncovered by sequencing the genomic loci and exploited to generate ecotype specific CAPS markers. For *SPL7* the PCR primers 5′-GCC TCT GAT TCC GAC GCA AAC TCC G-3′ and 5′-TCC TTC ATC AAA GTC CGG GAG CAA ATG-3′ allowed the amplification of a 800 bp fragment. The fragment derived from ecotype Col was differentially cleaved with *Bbs*I. The PCR primers used for *SPL*12, 5′-TCA GAT TCA GCA TCT GAC CAG TCA CC-3′ and 5′-TGA TGC TGC AAG TCT TGA CTT GAG TTC-3′, yielded a 1.5 kb fragment which was differentially cut with *Aci*I if derived from ecotype Ler. The RI scoring data were kindly analysed by Clare Lister (John Innes Centre) or Mary Anderson (Nottingham *Arabidopsis* Stock Centre).

For *SPL6*, no RFLP between ecotypes Col and Ler could be found with any of the 18 different restriction enzymes tested (*SPL6* cDNA used as probe). Therefore, in order to map *SPL6*, the IGF *Arabidopsis* BAC library (Mozo et al., 1998) was screened with the cDNA of *SPL6*. A ready-to-use blot of the library was provided by the Resourcenzentrum at the MPI für Molekulare Genetik, Berlin, Germany. The positive clones were part of a contig anchored with the chromosome 2 marker m429 (list of contigs of IGF BAC library available at http://www.mpimp-golm.mpg.de/101/mpi_mp_map/bac.html).

The *SPL1* and *SPL8* to *11* loci are located in BAC clones that have already been fully sequenced within the *Arabidopsis* Genome Initiative. A mapped genomic survey sequence representing *SPL1R2* is also found in the electronic databases. Therefore, their map positions are known.

2.6. *Remaining techniques and methods*

Standard molecular biology techniques were performed according to Sambrook et al. (1989). Autoradiographs of Northern blots were digitised and cropped using Adobe PhotoshopTM 4.0 (Adobe Systems, Inc.).

3. Results

3.1. *Isolation of* Arabidopsis *SBP-box genes: the* SPL *genes*

In order to isolate *Arabidopsis* SBP-box genes we screened cDNA and genomic libraries at low stringency using initially the cDNAs of the *Antirrhinum SBP1* and *SBP2* genes and later *Arabidopsis* SBP-box sequences as probes. In addition, we searched the electronic sequence databases for the presence of additional *Arabidopsis* SBP-box sequences (ESTs and genomic sequences). The outcome was the isolation of 12 *Arabidopsis* SBP-box genes, which we named *SQUAMOSA promoter-binding protein-like1* to *12* (*SPL1* to *12*). It should be mentioned that the numbering of the *SPL* genes presented here merely reflects the chronology of their isolation. The sequence of full-size cDNAs of all 12 *SPL* genes as well as those of the genomic loci was determined. The EMBL nucleotide database accession numbers of these sequences are listed in Table 1 together with genomic and EST sequences of the *SPL* genes and SBP-box genes of other species present in the EMBL nucleotide databases (until February 1999).

The SBP-domains encoded by the 12 characterised *SPL* genes show a high level of sequence conservation (Fig. 1). The alignment also reveals the conserved bipartite nuclear localisation signal located within the SBP-

Table 1
Electronic database accession numbers of sequences representing SBP-box genes (state February 1999)

Species	Gene	Sequence origin	Accession numbers	References/ comments
A. majus	*SBP1*	cDNA	X92369	Klein et al. (1996)
	SBP2	cDNA	X92079	Klein et al. (1996)
	SBPH3	cDNA	AJ011621	This work
	SBPH4	cDNA	AJ011622	This work
	SBPH5	cDNA	AJ011623	This work
A. thaliana	*SPL1*	genomic (Col)	AJ011577	This work
	SPL1	cDNA (Col; Ler)	AJ011629; AJ011628	This work
	SPL1	EST	AA042466; H36838; T42259	
	SPL1	genomic	AC004411; AC007236	
	SPL2	genomic (Col)	AJ011624	This work
	SPL2	cDNA (Col; Ler)	AJ011626; AJ011625	This work
	SPL2	EST	AA395451; N37151; T44492	
	SPL2	genomic	AB017070	
	SPL3	genomic (Col)	AJ011627	Cardon et al. (1997)
	SPL3	cDNA (Col; Ler)	Y09727; AJ011633	Cardon et al. (1997); this work
	SPL3	EST	N37626; T22734	
	SPL3	cDNA	Y09427	
	SPL3	genomic	U78721	
	SPL4	genomic (Col)	AJ011630	This work
	SPL4	cDNA (Col; Ler)	AJ011631; AJ011632	This work
	SPL5	genomic (Col)	AJ011609	This work
	SPL5	cDNA (Col; Ler)	AJ242960; AJ011610	This work
	SPL6	genomic (Col)	AJ011644	This work
	SPL6	cDNA (Col)	AJ011643	This work
	SPL6	EST	F13583; F14174; Z24507	
	SPL6	genomic	B28493; B73706; B96309	
	SPL7	genomic (Col)	AJ011613	This work
	SPL7	cDNA (Col; Ler)	AJ011612; AJ011611	This work
	SPL7	EST	AA712846; F13566; F13585; H76647; T04465	
	SPL8	genomic (Col)	AJ011641	This work
	SPL8	cDNA (Col)	AJ011642	This work
	SPL8	genomic	U89959	
	SPL8	genomic	B08640; B09506; B12095	
	SPL9	genomic (Col)	AJ011640	This work
	SPL9	cDNA (Col; Ler)	AJ011638; AJ011639	This work
	SPL9	genomic	AC002561	
	SPL9	genomic	B27659	
	SPL10	genomic (Col)	AJ011636	This work
	SPL10	cDNA (Col)	AJ011637	This work
	SPL10	genomic	AC004557	
	SPL11	genomic (Col)	AJ011634	This work
	SPL11	cDNA (Col)	AJ011635	This work
	SPL11	genomic	AC004557	
	SPL12	genomic (Col)	AJ132097	This work
	SPL12	cDNA (Col)	AJ132096	This work
	SPL12	EST	R29934	
	SPL12	genomic	B25955; B29472	
	SPL1R2	genomic (Col)	AJ243094	This work
	SPL1R2	cDNA (Col)	AJ242957	This work
	SPL1R2	EST	AA585893; AA720290; N65045; W43420; Z37590	
	SPL1R2	EST	Z18223	
	SPL1R2	genomic	B08252; B09556; B10967; B22716	
	SPL1R3	cDNA (Ler)	AJ242958	This work
	SPL1R3	EST	F20063; F20064; N65430	
B. vulgaris	*COXII*	mitochondrial	X55297	Chimaeric clone?
	MITOCH. RIBO. PROT. S7	mitochondrial	AB008450	Chimaeric clone?
B. rapa		EST	L38193	

Table 1
Electronic database accession numbers of sequences representing SBP-box genes (state February 1999)

Species	Gene	Sequence origin	Accession numbers	References/ comments
O. sativa		EST	AF074808	Similarity with SPL1/7/12 C-term. of SBP domain
		genomic	AQ273505	
		genomic	AQ289632; AQ362713	3′ SBP box exon
		EST	C72011; C72359; C91611	
	RMSOD2	cDNA	L34039	Chimaeric clone?
P. balsamifera		EST	AI166843	Similarity with SPL1/12 C-term.
Z. mays	*LG1*	cDNA	U89496	Moreno et al. (1997)
	ZMSBP1	cDNA	AJ011614	This work
	ZMSBP2	cDNA	AJ011615	This work
	ZMSBP3	cDNA	AJ011616	This work
	ZMSBP4	cDNA	AJ011617	This work
	ZMSBP5	cDNA	AJ011618	This work
	ZMSBP6	cDNA	AJ011619	This work
	ZMSBP7	cDNA	AJ011620	This work

domain of all SPL proteins, which was shown experimentally to be functional in the maize SBP-box gene *LG1* (Moreno et al., 1997). The size of the deduced peptides is quite variable, ranging from 131 amino acids (SPL3) to 927 amino acids (SPL12), and amino acid sequence similarity remains largely restricted to the SBP-domain. However, pairs of *SPL* genes sharing high sequence homology throughout the whole predicted peptide were detected. These pairs are formed by SPL4/SPL5, SPL10/SPL11 and SPL1/SPL12 (Fig. 2). The percentage of sequence identity/similarity between members of each pair is 65%/73%, 78%/83% and 69%/78%, respectively. Among the remaining SPL proteins some were found to share only one region of sequence homology, in addition to the SBP-domain, with one of the previously described *SPL* gene pairs. These are SPL3 and SPL2 with homology in their N-termini to the SPL4/SPL5 and SPL10/SPL11 pairs, respectively, and SPL7 to the SPL1/SPL12 pair in a region C-terminal of the SBP-domain (Fig. 2).

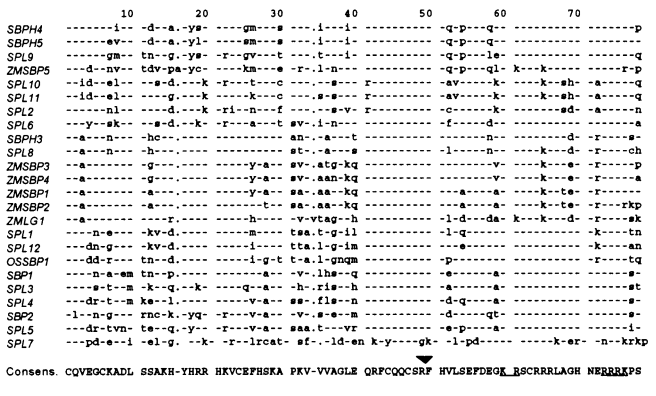

Fig. 1. A high level of sequence conservation is revealed by an alignment of the SBP-domains encoded by the 12 characterised *Arabidopsis SPL* genes and by SBP-box genes of other species. The conserved bipartite nuclear localisation signal located within the SBP-domain is underlined in the consensus sequence. The triangle above the consensus sequence refers to the position of an intron found to be conserved in all SBP-box gene genomic sequences analysed.

Screening the electronic DNA databases with the protein sequences of the *SPL* genes detected ESTs with sequence similarity to the C-terminus of SPL1 and SPL12. Sequencing of the clones from which the ESTs were derived showed that they represent transcripts of two different genes which were named *SPL1R2* and *SPL1R3* (*SPL1-related2* and *3*). Unfortunately, all these clones carry a 5′ truncation. Longer *SPL1R3* clones were isolated from a cDNA library but also exhibited 5′ truncations and an SBP-box could not be detected. Sequencing of the genomic region encoding the known portion of the *SPL1R2* coding sequence revealed conservation of most of the intron positions with *SPL*1/*SPL*12. At this point it is not possible to say whether *SPL1R2* and *SPL1R3* represent two members of the *SPL* gene family. Their SBP-box could be located in the 5′ end of the transcripts, which have not been isolated yet. Alternatively *SPL1R2* and *SPL1R3* could represent two novel genes that share some protein domains with *SPL*1/*SPL*12 but lack an SBP-box. Cloning of full-size cDNAS and genomic clones will clarify this matter.

3.2. *Organisation of the genomic loci of the* SPL *genes*

The alignment of the cDNA and genomic sequence of each *SPL* gene uncovered a diverse genomic organisation among the *SPL* loci. Not only are the sizes of introns and exons highly variable, but also the number of introns per locus varies from one to nine. Fig. 3 shows schematically the exon/intron structure of the 12 *SPL* loci. The SBP-box turned out to be interrupted by an intron in all 12 genes. Therefore the SBP-domain is encoded by portions of two exons. The position of this

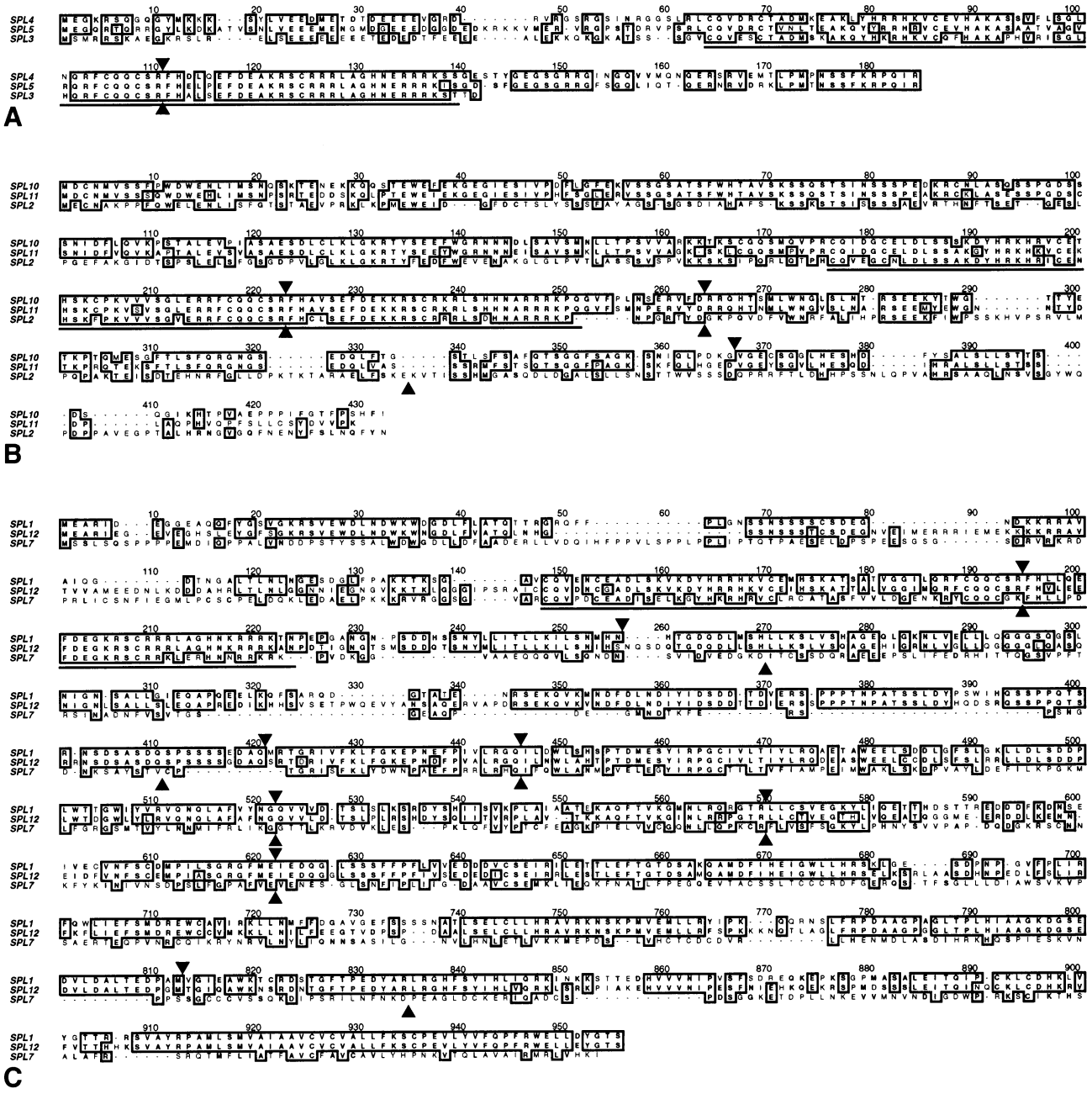

Fig. 2. Alignment of complete SPL peptides reveals sequence similarity outside the SBP-domain. (**A**) Alignment of SPL3, 4 and 5. (**B**) Alignment of SPL2, 10 and 11. (**C**) Alignment of SPL1, 12 and 7. Identical amino acids at corresponding positions are bold and boxed, chemically similar amino acids are bold. The SBP-domains are underlined. Intron positions within the first two sequences of each trio are marked by triangles above and intron positions within the third sequence by triangles below the sequences.

intron is conserved but its sequence and length are very variable (74 bp in *SPL1* to 646 bp in *SPL9*). An intron in the same position was also found in the *Antirrhinum SBP1* and *SBP2* genes (Klein et al., 1996). In all 12 *SPL* genes the SBP-box is spread over the first two exons. However, due to differences in length of these exons, the position of the SBP-box within the coding sequence is variable. Fig. 3 also shows that the variation in *SPL* transcript size is mainly due to an extension of the coding sequence downstream of the SBP-box. As an exception, an intron in the 5′ UTR was identified in *SPL2*.

Genes encoding peptides with sequence similarity outside the SBP-domain have similar numbers of introns, supporting the idea that these *SPL* genes are evolutionary more related to each other than to the other *SPL* genes. *SPL4*, *5* and *3* have only one intron, *SPL1*, *12* and *7* have nine introns and *SPL10*, *11* and *2*

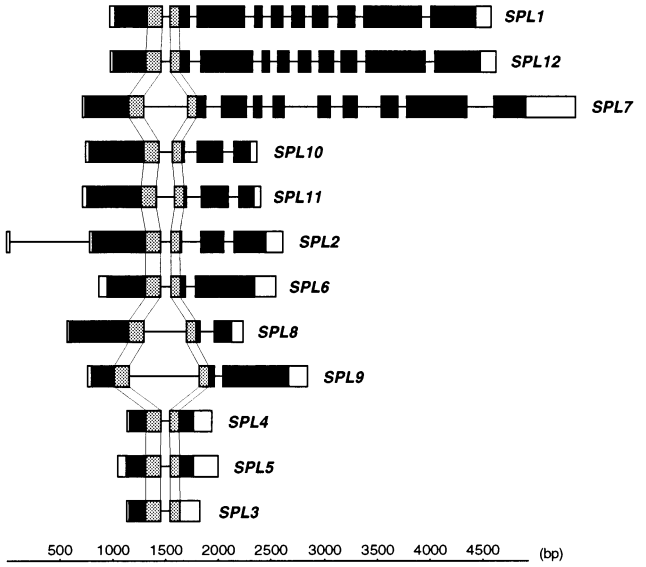

Fig. 3. Schematic representation of the exon/intron structure of the 12 *SPL* loci. The genes are ordered based on intron number and on their alignment as shown in Fig. 2. Introns are represented by horizontal lines, exons by boxes. Untranslated 5′ and 3′ regions are represented white. The different SBP-box coding sequences (grey) are connected through thin lines.

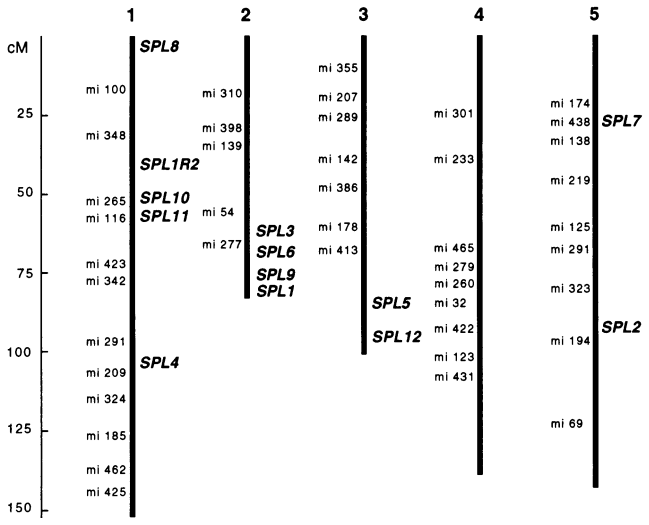

Fig. 4. Positions of the 12 *SPL* genes on the *Arabidopsis* chromosome map. Although so far no *SPL* gene was found to map to chromosome 4, there seems to be no obvious clustering within the genome. The reference mi RFLP markers are found at http://genome-www3.stanford.edu/cgi-bin/AtDB/miRFLPintromap.

have three introns within the coding sequence. As mentioned before, *SPL2* has an additional intron in the 5′ UTR. Between pairs of *SPL* genes with high sequence homology the positions of the introns are conserved, whereas for less related genes such as *SPL2* (related to *SPL10*/*SPL11*) and *SPL7* (related to *SPL1*/*SPL12*) only some of the intron positions are conserved (Fig. 2). However, the regions of *SPL7* with higher sequence homology to *SPL1*/*SPL12* are flanked by conserved introns.

Even though all the remaining characterised *SPL* genes *SPL6*, *SPL8* and *SPL9* have two introns, they lack significant sequence similarity outside the SBP-box/domain. Moreover, the exon sizes and position of the SBP-box within the coding sequences are also different. Either these genes are unique and no related *SPL* genes exist in the genome or these have not yet been detected.

3.3. *Determination of the map position of* SPL1 *to* SPL12

The chromosomal map position of the *SPL* genes was determined using different complementary approaches. Some genes were mapped with the Col × Ler RI lines from Lister and Dean (1993) using either conventional RFLP analysis (*SPL2* to *5*) or CAPS markers (*SPL7* and *12*) as described in Materials and methods. *SPL6* was mapped using hybridisation to IGF BAC library filters (Mozo et al., 1998), uncovering a contig that had been anchored with mapped markers. *SPL1* and *SPL8* to *11* are located within sequenced BAC clones of known map position. Fig. 4 displays schematically the five *Arabidopsis* chromosomes including the position of *SPL1* to *12* and reference mi RFLP markers (Liu et al., 1996). The Lister and Dean RI map, which can be viewed at http://nasc.nott.ac.uk/new_ri_map.html, allows a more detailed visualisation of the map positions of the 12 *SPL* loci described here. Whereas *SPL2* to *5*, *SPL7* and *SPL12* can be found directly on this map, the molecular markers close to *SPL1*, *SPL6* and *SPL8* to *11* are mi79a, m429, ve001, m336, mi15 and mi15, respectively. *SPL1R2*, one of the presumed SBP-box genes with sequence similarity to *SPL1*/*SPL12*, maps to chromosome 1, close to the molecular marker mi203.

The 12 characterised *SPL* genes seem to be scattered throughout the genome, although so far no *SPL* gene localises to chromosome 4. When we compare the map position of the pairs of *SPL* genes that show high sequence homology outside the SBP-box, we can see that *SPL4*/*SPL5* and *SPL1*/*SPL12* map to two different chromosomes. Therefore, when these gene pairs arose through duplications these events must have been followed by translocations. Since these genes, with the exception of *SPL1*, localise to not yet sequenced areas of the genome, the extensions of these duplications remain unknown. However, preliminary studies show that both *SPL4* and *SPL5* show considerable sequence similarity beyond their 5′ UTR.

SPL10 and *SPL11* are located on the same chromosome and are found in immediate proximity (convergently transcribed, with 1738 bp between the stop codons) on a sequenced BAC clone. Sequence compari-

son revealed that the duplication does not extend further than 900 bp upstream of the *SPL10*/*SPL11* start codons.

3.4. *Analysis of* SPL *gene expression during development*

The transcription pattern during *Arabidopsis* development of the 12 *SPL* genes was studied using Northern blot analysis. Poly(A)$^+$ RNA was isolated from all the aerial tissues of LD grown plants at 1, 2 and 3 weeks after sowing as well as from young inflorescences. The

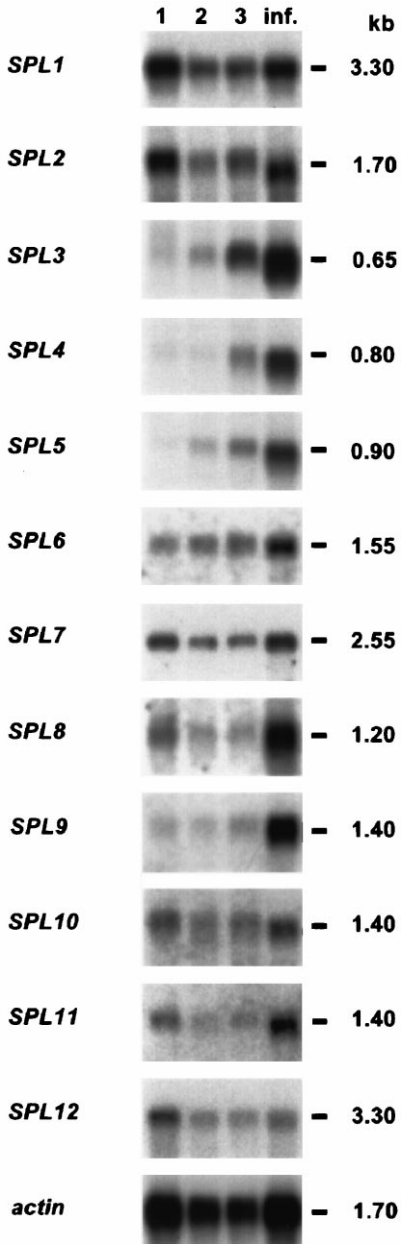

Fig. 5. Northern blot analysis of *SPL* gene expression during development (1, 2 and 3 weeks after germination) of LD grown *Arabidopsis* plants. Some *SPL* genes are constitutively expressed during development while others are up-regulated. Actin mRNA hybridisation has been used as a control for equal RNA loading. Inf., inflorescence.

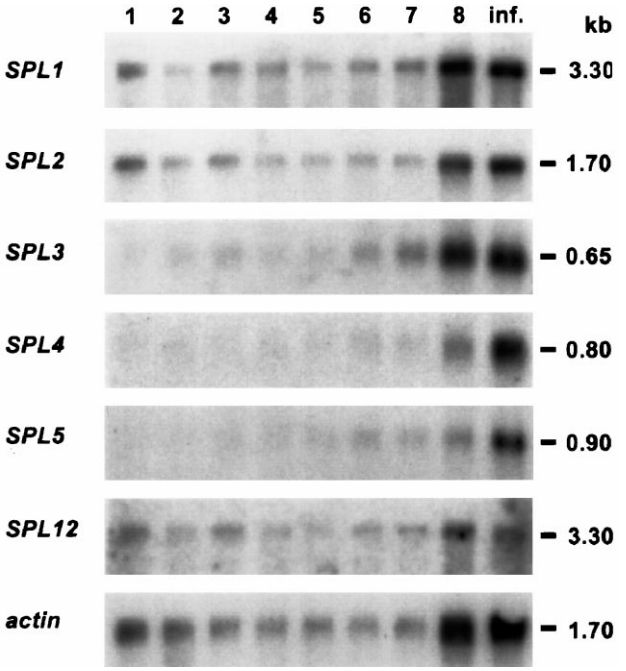

Fig. 6. Northern blot analysis of *SPL* gene expression during development of SD grown *Arabidopsis* plants. The temporal expression of the *SPL* genes (1 to 8 weeks after germination) reflects their behaviour under LD conditions. Actin has been used as a control for equal RNA loading. Inf., inflorescence.

RNA fraction from 3 week-old plants includes young inflorescence tissue since under our LD growth conditions bolting started around 3 weeks after germination. The abundance of transcripts of the different *SPL* genes varies considerably over the different developmental stages (Fig. 5). In general, based on their expression patterns the *SPL* genes can be divided into two groups. The transcript levels of *SPL1*, *2*, *6*, *7*, *8*, *10*, *11* and *12* remain rather unchanged during development. *SPL9* transcripts are also detected in all four RNA fractions (constitutive expression) but their amount is much higher in the inflorescence fraction. On the other hand, *SPL3*, *4* and *5* expression is hardly detectable before week 2 and then increases during further development (developmentally up-regulated). Remarkably, all genes having more than one intron are constitutively expressed. Even though we loaded equal amounts of RNA (based on OD measurements), the hybridisation pattern generated with an actin gene probe, which is generally used as control for equal loading, changes during development. If we assume that indeed every blotted fraction represents an equal amount of RNA, then the abundance of actin transcripts decreases with plant age and is higher again in the inflorescence fraction (Fig. 5). This could be due to the relative contribution of meristematic tissue to the samples.

The *SPL3*, *4* and *5* genes were found to be up-regulated under LD and were therefore analysed under SD as well (Fig. 6). Under these conditions their

expression is hardly, if at all, detectable in RNA fractions from plants in an early vegetative state. An increase in expression at later stages of vegetative development, however, seems to correlate with the floral phase transition. Genes such as *SPL1*, *2* and *12* showed constitutive expression under both LD and SD.

The differences as seen in Northern blots are likely to reflect not only temporal but also spatial differences in expression. Among the genes that are developmentally up-regulated, for example, there are only small differences in transcript abundance within the same developmental stage. However, a more detailed analysis of expression by in situ hybridisation showed that in spite of similar developmental expression patterns the transcripts of *SPL3*, *4* and *5* partly localise to different tissues (Cardon et al., 1997; Fig. 7).

3.5. Different SPL proteins bind similar DNA sequence motifs in vitro

We have shown previously that the *Antirrhinum* SBP-box genes *SBP1* and *2* and the *Arabidopsis SPL3* gene encode DNA-binding proteins and that their SBP-domain is required and sufficient for this interaction (Klein et al., 1996; Cardon et al., 1997). Therefore, the presence of the SBP-box suggests that other *SPL* genes encode DNA binding proteins as well.

To support this hypothesis we expressed three additional *SPL* genes, *SPL1*, *4* and *5*, in an in vitro rabbit reticulocyte system and assayed the proteins for binding activity. While the full-size SPL5 protein was expressed, SPL1 and 4 had short deletions at their N-termini (15 and 11 amino acids, respectively; the SBP-domain was however intact). As their target in electrophoretic mobility shift assays (EMSA) we chose an *AP1* promoter sequence motif, previously shown to interact in vitro with SPL3, SBP1 and SBP2 (Cardon et al., 1997). All three proteins interact in vitro with this *AP1* promoter element with apparently similar affinity (data not shown). It should be recalled that, whereas SPL4 and 5 are related to SPL3, SPL1 is completely unrelated to SPL3 outside the SBP-domain. Besides, the truncated SPL1 protein expressed has 881 amino acids while SPL3 only has 131.

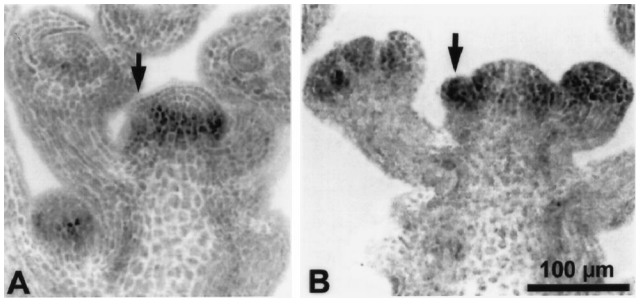

Fig. 7. In situ hybridisation analysis of spatial expression patterns of developmentally regulated *SPL* genes. (**A**) Specific expression of *SPL4* in the rib meristem and inter-primordial regions of the inflorescence apex. (**B**) Specific expression of *SPL5* in the inflorence apical meristem and young flowers. Note especially the difference in expression with respect to flower primordia (arrows).

Table 2
Comparison of DNA sequences experimentally tested for binding SBP-domain proteins in vitro. A consensus core sequence was derived for the binding site of SBP-domain proteins in different promoters

Sequence origin	Motif	Binding
SQUA promoter	GACG**TCCGTACAA**CAAGTC	+
AP1 promoter	GTGG**TCCGTACAA**TGTTAC	+
DEFH84 promoter	AATA**TTCGTACAA**CATATA	+
consensus	----**T-CGTACAA**-----	
FLO promoter	TTCT**TCTGTACA**GAGTACT	−
FLO promoter	TACT**TCAGTAAA**AATTAAA	−

The outcome of the EMSA suggests that all *SPL* genes are indeed likely to encode for DNA binding proteins.

In the promoter region of the *Antirrhinum* gene *DEFH84*, like *SQUA* and *AP1* a MADS-box gene specifically expressed during flower development (Huijser and Leitner, unpublished results), we could identify a sequence motif which was also recognised by SBP1 and SPL3 (data not shown). Two putative motifs were also detected in the promoter of the *Antirrhinum* gene *FLORICAULA* (*FLO*; Coen et al., 1990), a gene which does not belong to the MADS-box family but which controls, like *SQUA*/*AP1*, the identity of the inflorescence meristem. However, no interaction of these sequence motifs was detected with SBP1 or SPL3. Based on the DNA sequences experimentally tested for binding SBP-box genes in vitro, a consensus core sequence was derived (Table 2).

Random oligo site-selection experiments could further address the question concerning target site specificity but a definitive answer will only come from the isolation of in vivo targets.

3.6. Structural features of SPL proteins outside the SBP-domain

Except for their SBP-domain, the *SPL* gene products are quite diverse in size and sequence. In order to learn more about their function, a screen for known functional protein motifs outside the SBP-domain was conducted. Similarly to SBP1 and 2, the N-termini of SPL3, 4 and 5 are acidic in character and could therefore be involved in transcriptional activation (Hahn, 1993). Specific domains of other SPL proteins were found to match motifs known to be involved in protein/protein interaction such as a Myc type 'helix–loop–helix' dimerisation domain in SPL6 and 8 (at positions 369–377 respectively 214–222) and a leucine zipper in SPL7

(position 408–429). Therefore, some SPL proteins might require the formation of dimers to exert their function.

3.7. Presence of SBP-box genes in other plant species

Screening *Antirrhinum* cDNA libraries under low stringency with the *SBP1* and *2* cDNAs as probes resulted in the isolation of three additional SBP-box genes. These new members of the *Antirrhinum* SBP-box gene family were named *SBPH3*, *4* and *5* (for *SQUAMOSA promoter-binding protein homologue3* to *5*; their EMBL nucleotide database accession numbers are listed in Table 1). Northern analysis showed that *SBPH3*, *SBPH4* and *SBPH5* are constitutively expressed during *Antirrhinum* development (data not shown).

SBP-box genes, however, are not restricted to *Arabidopsis* and *Antirrhinum*. Screening of electronic sequence databases identified cDNAs, ESTs and genomic sequences in other plant species as well, both dicots (*Beta vulgaris*, *Brassica rapa*) and monocots (*Oryza sativa* and *Zea mays*; see Table 1). This suggests that SBP-domain proteins are ubiquitous in higher plants. However, due to the poor quality of the ESTs, chimeric cDNAs or genomic survey sequences, their use for sequence comparison or evolutionary studies is limited. For instance, the two *B. vulgaris* SBP-box sequences are found in the UTR of mitochondrial transcripts and are likely to be chimeras of mitochondrial and nuclear DNA. Another case of a chimeric clone, already pointed out by Klein et al. (1996), is the GenBank accession l34039 for the rice manganese superoxide dismutase. This clone includes a SBP-box related sequence in the 5′ UTR, in antisense orientation.

The only SBP-box gene for which a (loss-of-function) mutant phenotype is known is the maize *LG1* gene. As mentioned before, *LG1* is involved in the control of ligule and auricle development in monocot leaves. The assumption that the SBP-box genes code for transcription factors is further supported by experimental demonstration that LG1 is nuclear localised (Moreno et al., 1997). An alignment of the LG1 SBP-domain with those of the *Antirrhinum* and *Arabidopsis* SBP-box genes revealed the presence of an additional amino acid in LG1 (Fig. 1). The accuracy of the published sequence was corroborated by re-sequencing from three different maize lines the genomic region of the *LG1* locus where this additional amino acid is encoded (data not shown). In order to determine whether the presence of this additional amino acid is a peculiarity of LG1 or a common feature of all maize (and maybe monocot) SBP-box genes, additional members of the maize SBP-box gene family were isolated from a cDNA library (see Materials and methods). The seven novel maize SBP-box genes recovered were named *ZMSBP1* to *7* (for *Zea mays* SBP-box genes 1 to 7). Alignment of the SBP-domains of ZMSBP1 to 7 with LG1 showed that six of them lacked an additional amino acid like in LG1. However, the remaining gene, *ZMSBP5*, also encoded an additional amino acid but at a different position compared to *LG1*. The variability in the SBP-domain structure observed in maize is not found in any of the 17 dicot SBP-box genes we isolated (see alignment in Fig. 1).

Similarly as described for *Arabidopsis*, some of the ZMSBP proteins share extensive sequence homology outside the SBP-domain (ZMSBP1/ZMSBP2 and ZMSBP3/ZMSBP4 with sequence similarity/identity of 82%/80% and 72%/71%, respectively). One intriguing feature specific for the ZMSBP genes is the presence of a tandem of two or more in-frame AUG start codons at their putative translation start site (two methionine residues in LG1, ZMSBP3 and ZMSBP4; three in ZMSBP1 and four in ZMSBP2).

Screening of the proteins encoded by the newly isolated SBPH and ZMSBP for protein motifs or structural features already described for other proteins did not yield any significant match.

3.8. Phylogenetic relationship among known SBP-box genes

All full-size SBP-domain sequences available were aligned to compare their structures and estimate their phylogenetic relationships. This alignment, shown in Fig. 1, includes the SBP-domain from five *Antirrhinum*, 12 *Arabidopsis*, one rice and six maize SBP-domain proteins. The SBP-box intron, whose presence and position is apparently conserved in all SBP-box genes studied so far, determines that the N-portion of the SBP-domain (62% of the whole SBP-domain) is encoded by one exon and the C-portion (38%) by another. The overall sequence identity/similarity is higher in the C-portion of the SBP-domain. Within this region, there is the highly conserved bipartite nuclear localisation signal. A deletion analysis on SPL3, more detailed than that carried out by Klein et al. (1996) on SBP1, showed that the whole SBP-domain is required for interaction with DNA (Cardon and Adolphs, unpublished results). As mentioned before, a variation in the number of amino acid residues in the SBP-domain is observed in the maize LG1 and ZMSBP5 proteins. The additional amino acids in LG1 and ZMSBP5 are located within the N-portion of the SBP-domain.

Based on the SBP-domain sequences, a phylogenetic reconstruction of all available SBP-boxes was carried out using the program PHYLIP (Phylogeny Inference Package; Felsenstein, 1989). The resulting phylogenetic tree (Fig. 8) shows that the *Arabidopsis* genes that show sequence similarity outside the SBP-domain tend to cluster and allow, together with SBP-box genes of other species, the identification of subfamilies. The exception

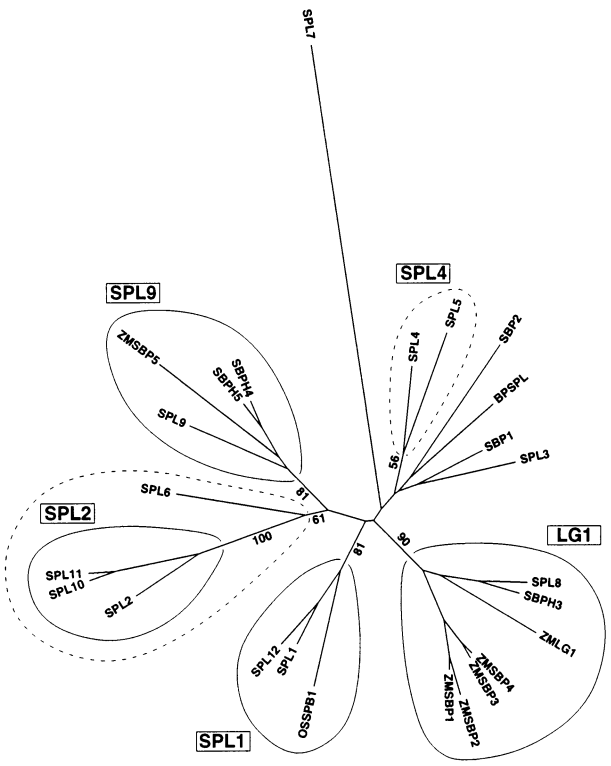

Fig. 8. Reconstruction of the phylogenetic relationships between the SBP-box genes based on their SBP-domains allows the recognition of subfamilies. Subfamilies based on bootstrap values above 80 are enclosed by a solid line, above 50 by a dashed line. The names of the subfamilies are boxed. BPSPL represents a SBP-box gene of *Betula pendula*, the sequence of which was kindly provided by Dr. M. Lännenpää (University of Joensuu, Finland).

is *SPL7* which, based on its encoded SBP-domain, seems to be quite unrelated to all other SBP-box genes identified so far. This is very likely due to the presence of a relatively high number of non-consensus amino acids at highly conserved positions in its SBP-domain. Close pairing is observed among *SPL* genes *11* and *10*, the maize genes *ZMSBP1*/*ZMSBP2* and *ZMSBP3*/*ZMSBP4*, and the *Antirrhinum* genes *SBPH4*/*SBPH5*. These gene pairs are likely to be due to recent gene duplication events, an assumption strongly supported in the case of the physically closely linked *SPL10* and *11* genes.

When dealing with a gene family as diverse as the SBP-box genes we could assume that orthologous genes in different species should have protein domain(s) in common in addition to the SBP-domain. To test this hypothesis we compared all available SBP-domain proteins over their entire length from different species. No significant sequence conservation outside the SBP-domain among different species could be detected. Only the SBP1 and 2 (*Antirrhinum*) share a common feature with the *Arabidopsis* SPL3, 4 and 5 proteins: the presence of an acidic region near the N-terminus. Remarkably, the expression of these five genes is developmentally regulated. Moreover, these genes have a similar genomic structure, i.e. only one intron.

The dicot genes closest to *LG1* are *SPL8* and *SBPH3*. However, these genes do not display obvious sequence similarity outside their SBP-domains. Since no complete set of *SPL* genes has been isolated yet, there might be other *SPL* genes more related to *LG1* in the *Arabidopsis* genome. Among the set of maize *ZMSBP* genes isolated, most belong to the *LG1* subfamily, only one to the *SPL9* subfamily and none to the other subfamilies. This could be due to the probe used for their isolation, which consisted of the 5′ portion of the SBP-box of *LG1*.

4. Discussion

4.1. SPL *is a diverse gene family*

We have described the *Arabidopsis SPL* gene family, a novel gene family encoding DNA binding proteins and putative transcription factors. The *SPL* genes are characterised by the presence of the SBP-box, which encodes a protein domain, the SBP-domain, required and sufficient for interaction with DNA. We initially isolated the first two SBP-domain proteins from *Antirrhinum majus* (Klein et al., 1996). However, to undertake a structural and functional characterisation of the SBP-box gene family, we chose the more favourable plant model system *Arabidopsis thaliana*. The 12 genes described here likely do not represent all members of the *Arabidopsis SPL* gene family. At present, seven of the 12 characterised *SPL* genes can be found in the genomic sequences from the *Arabidopsis* Genome Initiative (AGI; see Table 1). The amount of sequence generated by AGI and accessible for public screening accounts for 40% (status of February 1999) of the *Arabidopsis* genome (based on an estimated 120 Mb genome size; see http://genome-www.stanford.edu/*Arabidopsis*/agi.html). The number of *SPL* genes, therefore, in the *Arabidopsis* genome would be around 18 ($7 \times 100/40$), if we assume that there is a random distribution of *SPL* genes in the genome, an assumption supported by our findings. A similar number was estimated by hybridisation of replica filters of an *Arabidopsis* genomic library at low stringency with different *SPL* genes (data not shown). The rapid progress made in sequencing the *Arabidopsis* genome (http://genome-www.stanford.edu/*Arabidopsis*/progreport7.html) will soon provide a definitive answer.

Except for the presence of an SBP-box/domain, the *SPL* genes are very variable with respect to genomic organisation, transcript size and size and amino acid sequences of their proteins. The position of the SBP-domain within the peptide is also variable. However, some *SPL* genes share sequence homology in protein regions other than the SBP-domain. In addition, these

genes display a similar genomic structure as reflected in number and position of their introns. This supports the assumption that these *SPL* genes have arisen rather recently through gene duplication events. Moreover, in the case of the gene pair *SPL10*/*SPL11*, these genes are physically closely linked on chromosome 1.

The heterogeneity of the SBP-box gene family raises the question of the origin of the SBP-domain. Functional regions of proteins are often defined within single exons allowing domain swapping simply by exon shuffling. However, the SBP-domain is encoded by portions of two different exons. Matsuo et al. (1994) discussed the presence of introns as a protection mechanism against illigimate recombination events between family members.

The position of the SBP-box intron is conserved in all 12 *SPL* genes as well as in SBP-box genes of other species as far as their genomic sequences have been determined. Therefore, the formation of a functional SBP-domain relies on the correct splicing of this intron. Isolation of poly-adenylated transcripts with unspliced SBP-box introns (unpublished results) suggests that regulation at the level of splicing could play a role in the control of SBP-box gene expression.

Instead of exon shuffling, the distribution of the SBP-box over several non-related genes scattered over the genome could have involved transposons as shown for some isolated MADS-box sequences found in the genome of maize (Fischer et al., 1995; Montag et al., 1995). However, remnants of SBP-box related sequences flanked by transposon-like sequences have not been detected so far.

Comparison of the 12 *SPL* genes with all SBP-box sequences available to date from other plants revealed a high level of sequence conservation and colinearity within the SBP-domain. A variation is observed in the *LG1* and *ZMSBP5* genes of *Zea mays*, which at different positions carry an additional amino acid within their SBP-domain. However, this is not a feature common to all SBP-box genes of *Zea mays* or monocots. The insertion of single amino acids in these genes could be due to footprints left behind after insertion and excision of a transposon as, for instance, is known for the maize *En*/*Spm* transposable element system (Schwarz-Sommer et al., 1985).

4.2. SBP-domain proteins seem to be plant specific

In recent years, large-scale sequencing projects have yielded a wealth of genomic sequence data from a wide array of organisms. The genomes of several prokaryotes, including the bacterium *Escherichia coli* (Blattner et al., 1997), have been fully sequenced. The first eukaryote genome sequence completed was that of the unicellular budding yeast *Saccharomyces cereviseae* (Goffeau et al., 1996) and the first genome sequence of a multicellular organism, the nematode worm *Caenorhabditis elegans* has recently been declared 'essentially complete' (The *C. elegans* Sequencing Consortium, 1998). Screening of nucleotide sequence databases for SBP-box sequences has only yielded significant matches to sequences from higher plants. Therefore, the SBP-domain type of DNA binding domain seems to be plant specific.

The characterisation of hundreds of plant transcription factors over the last years revealed that in most cases plants have adopted pre-existing prototype functional motifs, which are common to other non-plant eukaryotes as well. It is thus interesting to find that plants have evolved new specific DNA-binding domains such as the SBP-domain. The SBP-domain might well be involved in the regulation of biological processes that are unique to plants.

4.3. SBP-box genes are likely to be involved in the control of diverse aspects of plant development

We still know very little about the function of SBP-box genes. Experimental work has shown that SBP-box genes encode nuclear proteins (shown for LG1; Moreno et al., 1997) which interact with DNA in a sequence specific manner (Klein et al., 1996; Cardon et al., 1997; this work). SBP-box genes are therefore likely to code for transcription factors, but this has not been shown experimentally yet.

Since SBP1 and SBP2 were isolated by their specific interaction in vitro with a sequence motif in the promoter of the floral meristem identity gene *SQUA* of *A. majus*, MADS-box genes could represent in vivo targets of SBP-box genes. The plant MADS-domain transcription factors are encoded by a multigene family. Although several members are known to be involved in flower development, the vegetative expression pattern of some MADS-box genes of unknown function suggests that not all are restricted to floral development (reviewed by Riechmann and Meyerowitz, 1997).

A DNA sequence motif related to that found in the *SQUA* promoter was identified in its presumed *Arabidopsis* orthologue, *AP1*, as well as in another *Antirrhinum* MADS-box gene, *DEFH84*. SBP1 and its presumed *Arabidopsis* orthologue, SPL3, can bind to these sequence motifs (Cardon et al., 1997; this work), further strengthening the hypothesis that some SBP-box genes could be transcriptional regulators of MADS-box genes. Moreover, constitutive expression of *SPL3* in transgenic plants causes early flowering, a phenotype often observed when MADS-box genes are ectopically expressed (Mandel and Yanofsky, 1995; Mizukami and Ma, 1997). However, no clear link between SPL3 and its putative target *AP1* could be shown in vivo because the early flowering phenotype of *35S::SPL3* plants is not altered by loss-of-function mutations of *AP1*

(Cardon et al., 1997) and hence seems not to be mediated via AP1.

Clues to the function of a transcription factor can be obtained by inducing changes in gene expression in transgenic plants expressing the appropriate constructs. Except for constitutive expression of *SPL3*, transgenic studies with *SPL* genes have not been very informative so far. Constitutive expression of *SPL3* in antisense also did not cause a noticeable phenotypic change (Cardon et al., 1997). Transgenic plants constitutively expressing *SPL1*, *SPL2*, *SPL4* and *SPL5* under the CaMV 35S promoter in sense or antisense orientation did not reveal clear phenotypical differences from the wild type phenotype (Cardon et al., unpublished results). The lack of a phenotypic effect when single SBP-box genes are constitutively expressed (with the exception being *SPL3*) could be due to the requirement of additional factors for the function of SBP-domain proteins. On the other hand, these transgenic experiments have shown that the early flowering phenotype following *SPL3* constitutive expression is specific, because constitutive expression of *SPL4* or *SPL5*, the *SPL* genes most closely related to *SPL3*, does not cause early flowering. Therefore, the in vivo targets of *SPL4* and *SPL5* seem to be different from those of *SPL3*. However, since all three genes interact in vitro with the same target DNA it cannot be excluded that the different behaviour observed in planta is due to differences in interacting with other factors and/or in post-transcriptional/translational modifications.

It is remarkable that, with the exception of the maize SBP-box gene *LG1* (Moreno et al., 1997), no other SBP-box gene has been identified based on a mutant phenotype. This is particularly striking in *Arabidopsis* considering the availability of numerous collections of mutagenised plants and the relative ease of isolating genes identified by mutations.

Only the maize Lg1 mutant phenotype provides a link between an SBP-box gene and the development of the monocot leaf. *Arabidopsis* leaves lacks ligules and auricules but Mooney and Freeling (1997) speculated that there could be parallels between auricules and stipules. From all *Arabidopsis* SBP-box genes we characterised so far, the *SPL8* gene is most similar to *LG1*. However, no clear sequence similarity outside the SBP-domain can be detected and thus the *Arabidopsis* orthologue of *LG1* is still missing if it exists at all.

One explanation for the present lack of characterised Spl mutants could be: either they are gametophytic lethal or have no (obvious) phenotype due to redundancy or lack of severe effects on primary plant functions. Numerous lethal mutations have been described in *Arabidopsis* but few genes have so far been isolated. Especially redundancy could explain the absence of obvious mutant phenotypes. The presence of redundant gene functions in plants is a well-documented phenomenon (Pickett and Meeks-Wagner, 1995) and the existence of *SPL* genes with sequence similarity outside the SBP-box gives this hypothesis further support. An obvious test for functional redundancy would be the construction of double mutants.

4.4. Outlook

Completion of the sequencing of the *Arabidopsis* genome in the near future will reveal the true size of the *SPL* gene family and provide structural information for the remaining *SPL* genes. It will also offer the possibility of studying the evolution of this gene family in *Arabidopsis*. In order to study the origin/evolution of the SBP-domain in plants it would also be interesting to determine whether SBP-box genes can be found in, for instance, gymnosperms, ferns and mosses.

However, the real challenge is to assign functions to the *SPL* genes and the isolation of loss-of-function mutants should be instrumental to address this issue. Our current approach for mutant isolation involves PCR-mediated reverse genetic screens of transposon mutagenised *Arabidopsis* populations (Baumann et al., 1998).

5. Addendum

After submission of this paper, the following novel SBP-box genes/related sequences were submitted to the GenBank. *A. thaliana* AB025619 and AB023037 representing the genomic (Col) sequence of *SPL13*, a new *SPL* gene located on chromosome 5 near marker nga129 (below *SPL2* in Fig. 4); AC007369, representing the genomic (Col) sequence of *SPL1R2*. This sequence confirms *SPL1R2* to be an SBP-box gene which we therefore suggest renaming *SPL14*. *G. max* EST AI443033; *Gossypium hirsutum* EST AI726550 (retained SBP-box intron); *O. sativa* genomic sequence AQ362713; *Z. mays* EST AI657269 (retained SBP-box intron). In addition, several sequences were deposited displaying similarity to the C-termini of SPL1, 7 and 12: *G. max* ESTs AI443713, AI494880, AI437733, *O. sativa* genomic sequence AQ574317 and *Z. mays* ESTs AI491353, AI461572, AI461494.

Acknowledgements

We are grateful to Dr. Jan Kim for his support in the phylogenetic analysis of the SPB-box gene family. We also thank Dr. Günter Theißen and all members of the laboratory for their helpful discussions on the paper and Dr. Mark Wilkinson for correcting the English. This work was funded in part by a grant from the BIOTECH program of the European Union, as part of the Project of Technical Priority 1993–1996 and by

the Deutsche Forschungsgemeinschaft (Schwerpunktprogramm *Arabidopsis*-294).

References

Baumann, E., Lewald, J., Saedler, H., Schulz, B., Wisman, E., 1998. Successful PCR-based reverse genetic screens using an En-1-mutagenised *Arabidopsis thaliana* population generated via single-seed descent. Theor. Appl. Genet. 97, 729–734.

Becraft, P.W., Bongard-Pierce, D.K., Sylvester, A.W., Poethig, R.S., Freeling, M., 1990. The *Liguleless-1* gene acts tissue specifically in maize leaf development. Dev. Biol. 141, 220–232.

Blattner, F.R., Plunkett III, G., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B., Shao, Y., 1997. The complete genome sequence of *Escherichia coli* K-12. Science 277, 1453–1462.

Cardon, G.H., Höhmann, S., Nettesheim, K., Saedler, H., Huijser, P., 1997. Functional analysis of the *Arabidopsis thaliana* SBP-box gene SPL3: a novel gene involved in the floral transition. Plant J. 12, 367–377.

Coen, E.S., Romero, J.M., Doyle, S., Elliott, R., Murphy, G., Carpenter, R., 1990. *Floricaula* a homeotic gene required for flower development in *Antirrhinum majus*. Cell 63, 1311–1322.

Fischer, A., Baum, N., Saedler, H., Theißen, G., 1995. Chromosomal mapping of the MADS-box multigene family in *Zea mays* reveals dispersed distribution of allelic genes as well as transposed copies. Nucl. Acids Res. 23, 1901–1911.

Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., Louis, E.J., Mewes, H.W., Murakami, Y., Philippsen, P., Tettelin, H., Oliver, S.G., 1996. Life with 6000 genes. Science 274, 563–567.

Hahn, S., 1993. Structure (?) and function of acidic transcription activators. Cell 72, 481–483.

Huijser, P., Klein, J., Lönnig, W.E., Meijer, H., Saedler, H., Sommer, H., 1992. Bracteomania, an inflorescence anomaly, is caused by the loss of function of the MADS-box gene *SQUAMOSA* in *Antirrhinum majus*. EMBO J. 11, 1239–1249.

Klein, J., Saedler, H., Huijser, P., 1996. A new family of DNA-binding proteins includes putative transcriptional regulators of the *Antirrhinum majus* floral meristem identity gene *SQUAMOSA*. Mol. Gen. Genet. 259, 7–16.

Lister, C., Dean, C., 1993. Recombinant inbred lines for mapping RFLP and phenotypic markers in *Arabidopsis thaliana*. Plant J. 4, 745–750.

Liu, Y.G., Mitsukawa, N., Lister, C., Dean, C., Whittier, R.F., 1996. Plant J. 10, 733–736.

Logel, J., Dill, D., Leonard, S., 1992. Synthesis of cRNA probes from PCR-generated DNA. Biotechniques 13, 604–610.

Mandel, M.A., Gustafson-Brown, C., Savidge, B., Yanofsky, M.F., 1992. Molecular characterisation of the *Arabidopsis* floral homeotic gene *APETALA1*. Nature 360, 273–277.

Mandel, M.A., Yanofsky, M.F., 1995. A gene triggering flower formation in *Arabidopsis*. Nature 377, 522–524.

Matsuo, K., Clay, O., Kuenzler, P., Georgiev, O., Urbanek, P., Schaffner, W., 1994. Short introns interrupting the Oct-2 POU domain may prevent recombination between POU family genes without interfering with potential POU domain 'shuffling' in evolution. Biol. Chem. Hoppe-Seyler 375, 675–683.

Mizukami, Y., Ma, H., 1997. Determination of *Arabidopsis* floral meristem identity by *AGAMOUS*. Plant Cell 9, 393–408.

Montag, K., Salamini, F., Thompson, R.D., 1995. *ZEMa*, a member of a novel group of MADS-box genes, is alternatively spliced in maize endosperm. Nucl. Acids Res. 23, 2168–2177.

Mooney, M., Freeling, M., 1997. Using regulatory genes to investigate the evolution of leaf form. Maydica 42, 173–184.

Moreno, M.A., Harper, L.C., Krueger, R.W., Dellaporta, S.L., Freeling, M., 1997. *Liguleless1* encodes a nuclear-localised protein required for induction of ligules and auricles during maize leaf organogenesis. Genes Devel. 11, 616–628.

Mozo, T., Fischer, S., Shizuya, H., Altmann, T., 1998. Construction and characterization of the IGF *Arabidopsis* BAC library. Mol. Gen. Genet. 258, 562–570.

Murray, N.E., 1983. Phage lambda, molecular cloning. Lambda II. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Pickett, F.B., Meeks-Wagner, D.R., 1995. Seeing double: appreciating genetic redundancy. Plant Cell 7, 1347–1356.

Riechmann, J.L., Meyerowitz, E.M., 1997. MADS-domain proteins in plant. Dev. Biol. Chem. 378, 1079–1101.

Samach, A., Kohalmi, S.E., Motte, P., Datta, R., Haughn, G.W., 1997. Divergence of function and regulation of class B floral organ identity genes. Plant Cell 9, 559–570.

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. In: Nolan, C. (Ed.), Molecular Cloning: a Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Schwarz-Sommer, Z., Gierl, A., Cuypers, H., Peterson, P.A., Saedler, H., 1985. Plant transposable elements generate the DNA sequence diversity needed in evolution. EMBO J. 4, 591–598.

The *C. elegans* Sequencing Consortium 1998. Genome sequence of the nematode *C. elegans*: A platform for investigating biology. Science 282, 2012–2018.