

大豆 SBP 转录因子家族的预测分析

朱命喜^{1,2}, 刘洋^{1,2}, 吴琼^{1,2}, 刘春燕², 徐晶³, 陈庆山¹, 胡国华²

(1. 东北农业大学 农学院, 黑龙江 哈尔滨 150030; 2. 黑龙江省农垦科研育种中心, 黑龙江 哈尔滨 150090; 3. 国家农业标准化监测与研究
中心, 黑龙江 哈尔滨 150036)

摘要: SBP 转录因子基因家族是一个植物的特异转录因子家族, SBP 基因都含有一段保守的核苷酸序列即 DNA 结合结构域, 又称为 SBP 盒 (SBP-box), SBP 盒编码的蛋白质序列称 SBP 结构域 (SBP-domain), 含 79 个氨基酸残基, 并具有高度保守性。该研究通过对拟南芥、水稻等植物已知的转录因子与大豆基因组数据比对, 并设置一系列严格的筛选标准, 共得到 44 条新的大豆 SBP 转录因子; 又通过基因分析注释以及启动子功能预测, 进一步发现这些转录因子参与生长发育、逆境胁迫响应、激素应答、抗霉菌应答、光合作用等调控过程。

关键词: SBP 家族; 转录因子; 大豆

中图分类号: S 565.1

文献标识码: A

文章编号: 1000-9841(2011)02-0177-07

Forecasting Analysis of SBP Transcription Factor Families in Soybean

ZHU Ming-xi^{1,2}, LIU Yang^{1,2}, WU Qiong^{1,2}, LIU Chun-yan², XU Jing³, CHEN Qing-shan¹, HU Guo-hua²

(1. College of Agriculture, Northeast Agricultural University, Harbin 150030, Heilongjiang; 2. The Crop Research and Breeding Center of Land-Reclamation, Harbin 150030, Heilongjiang; 3. National Agriculture Standardization Supervision and Research Center, Harbin 150036, Heilongjiang, China)

Abstract: In transcription factor family, SBP is a specific group only found in plant. The genes in this family contain a conservative nucleotide sequences namely DNA-binding domain, also called SBP-box, which encodes the SBP-domain including 79 amino acid residues and is highly conservative. With homology alignment, all previously known SBP from *Arabidopsis*, *Oryza sativa* and other plant species were blasted against the soybean genome sequences to mining novel SBP TFs in soybean by a series of strict filtering criteria. In total, 44 members in SBP TFs were identified and predicted. These transcription factors were predicted and verified to be involved in soybean growth and development, adversity response, hormone response, resistance to mold response and photosynthesis control process.

Key words: SBP family; Transcription factor; Soybean

转录调控是真核生物基因表达调控的重要组成部分, 通过顺式作用元件和反式作用因子相互作用实现。顺式作用元件 (cis-element) 主要包括启动子 (promoter)、增强子 (enhancer) 和抑制子 (suppressor) 三类。反式作用因子泛指与顺式作用元件直接或间接结合并参与靶基因转录过程的调控因子, 通称转录因子 (transcription factor)。转录因子种类繁多、功能复杂, 它们与靶基因上游各种特定 DNA 序列片段结合, 激活或抑制靶基因转录活性, 以调控靶基因在不同组织、细胞和环境条件下特异表达。

按照 DNA 结合结构域序列特征, 可以将转录因子分为不同种类, 即不同家族。相比较拟南芥的研究, 大豆转录因子研究还只局限于几个较大的转录因子家族如 DREB 和 MYB, 并主要集中在非生物胁迫抗逆性研究领域, 据北大生物信息中心预测的

大豆转录因子有 3 546 个, 大部分大豆转录因子尚未被研究。2008 年大豆基因组测序完成 (DOE/JGI, 2008), 为进一步大规模发掘研究大豆转录因子提供了重要的信息资源。

SBP 转录因子基因家族是一个植物特异转录因子家族。1996 年, 德国马普研究所 (Max Planck Institute for Plant Breeding Research) Huijser 领导的研究小组从金鱼草 (*Antirrhinum majus*) 中克隆得到 2 个基因, 它们能与花发育相关基因 *Squamosa* 的启动子结合并调控其表达^[1], 命名为 *Squamosa* 启动子结合蛋白 (*Squamosa promoter Binding Protein*, SBP), 简称 SBP1 和 SBP2。之后, 先后在拟南芥和苔藓中进行了 SBP 基因家族的 cDNA 克隆, 发现在这 2 个植物中有多个 SBP 转录因子基因拷贝^[2-3]。由于多数基因尚未进行功能鉴定, Huijser 等^[4]将它们称为

收稿日期: 2010-12-11

基金项目: 转基因专项资助项目 (2009ZX08004-009B)。

第一作者简介: 朱命喜 (1985-), 男, 在读硕士, 研究方向为大豆生物技术。E-mail: m-x-zhu@163.com。

通讯作者: 胡国华 (1950-), 男, 研究员, 博士生导师, 从事大豆栽培育种研究。E-mail: hugh757@vip.163.com。

类 SBP (SBP-like, SPL) 基因。因为这些基因都有编码 DNA 结合结构域的保守核苷酸序列, 又称为 SBP 盒 (SBP-box) 基因, SBP 盒编码的蛋白质序列称 SBP 结构域 (SBP-domain), 含有 79 个氨基酸残基。Yamasaki 等^[2] 利用核磁共振方法测定了拟南芥中 2 个 SBP 结构域的溶液三维构象, 均含有 2 个锌指结构 (Zinc finger)。其中一个 (SPL3) N-端锌指结构序列模体为 Cys3His, C-端锌指结构为 Cys2HisCys; 另一个 (SPL7) N-端锌指结构序列模体为 Cys4, C-端锌指结构与 SPL3 相同, 也是 Cys2HisCys。其与 SBP 结构域结合的顺式元件具有以 GTAC 为核心的回文结构。SBP 结构域 C 端富含碱性氨基酸, 可能为双向核定位信号 (Nuclear localization signal, NLS)^[5]。据文献报道, SBP 转录因子家族具有许多重要生物学功能, 包括花和果实发育、孢子发育、激素应答、抗霉菌侵蚀等。拟南芥 *SPL3* 基因的转录受发育调控, 主要集中在花序顶端分生组织、花分生组织和花器官原基, 可能与控制开花时间有关^[3]。拟南芥 *SPL8* 基因影响孢子囊发育, 进而影响植物生殖^[6], 并可能在赤霉素介导的发育过程中起局部调节作用^[7]。最近报道证实, 拟南芥 3 个 SBP 基因 (*SPL3*、*SPL4* 和 *SPL5*) 中具有 microRNA156 的调控位点^[4,8]。除拟南芥外, 其它植物 SBP 转录因子研究也屡有报道。玉米 SBP 转录因子 Liguleless (1LG1) 缺失突变体不能形成舌叶和叶耳^[9]。白桦 *BpSPL1* 基因能特异结合 BpMADS5 启动子, 参与调节花发育^[10]。衣藻 (*Chlamydomonas reinhardtii*) 中铜应答调控子 1 (Copper Response Regulator 1, *CRR1*) 基因含典型的 SBP 结构域^[7]。最近研究表明, SBP 基因在植物果实成熟和发育中起重要作用。野生玉米 (teosinte) 粒有硬壳包裹, 而栽培玉米粒则无外壳包裹。这种表型变化主要由玉米果实构造基因 (teosinte glume architecture, *tga1*) 控制^[11]。该基因为 SBP 转录因子家族成员, 野生玉米第 6 位氨基酸为赖氨酸 (Lys), 而栽培玉米中为天冬酰胺 (Asn)。Manning 等^[12] 发现, 控制西红柿果实成熟的关键基因 (Colorless non-ripening locus) 是一个 SBP 基因 (*LeSPL-CNR*), 启动子区域甲基化修饰突变体可抑制果实成熟。

PlantTFDB 收录的大豆 SBP 类转录因子是通过对比已有 EST 数据分析预测得到的, 至今尚未见实验方面的报道。大豆基因组测序完成, 根据基因组数据重新分析预测, 以期获得更为详尽的大豆 SBP 转录因子的信息, 丰富更新大豆转录因子数据库。

1 材料与方法

1.1 原始数据收集

大豆基因组序列、EST 序列均从 NCBI 上下载, 以及从 PlnTFDB, PlantTFDB, RiceTFDB 下载的拟南芥 (*Arabidopsis thaliana*)、水稻 (*Oryza sativa*)、大豆 (*Glycine max*) 的转录因子 (表 1)。

表 1 三种主要作物的转录因子信息

Table 1 The transcription factor information of three main crops

数据来源 Data source	物种 Species	转录因子个数 TFs	主页 Homepage
RiceTFDB	<i>Oryza sativa</i>	2856	http://plntfdb.bio.uni-potsdam.de/
PlnTFDB	<i>Arabidopsis</i>	2016	http://plntfdb.bio.uni-potsdam.de/
PlantTFDB	<i>Glycine max</i>	3546	http://plantfdb.cbi.pku.edu.cn/

1.2 SBP 基因家族隐马氏模型序列谱 (HMM profiles)

SBP 基因家族的隐马氏模型序列谱 (PF03110) 在蛋白质家族数据库 Pfam 上下载 (<http://pfam.sanger.ac.uk/family>)。并对其进行了修正修改, 包括数据扩充和精准修正, 最终产生两个不同文件, 一个是全序列比对文件, 一个是精准的 SBP 文件, 以扩大对预测结果的全面性。SBP 基因编码 DNA 结合结构域的保守核苷酸序列即 SBP 盒, 其编码的蛋白质序列称 SBP 结构域 (SBP-domain)。应用 HMMER3.0 程序包中的 hmmbuild 命令将其转换成隐马氏模型序列模式。

1.3 大豆转录因子的鉴定

首先运用 GENSCAN 将在 NCBI 上下载的大豆基因组数据预测分析得到大豆蛋白序列, 然后应用 HMMER 中的 hmmssearch 程序对这些蛋白序列进行转录因子预测。根据经验及其它物种数据库的统一标准, 采用 HMMER 默认内置 E-value 0.01。程序命令为 “#hmmssearch SBP. hmm SBP. out”。

1.4 转录因子注释筛选

对预测到的结果运用 InterProScan 做出了进一步的鉴定和注释, 最终确定出 44 个 SBP 家族基因, Interpro 注释号为 IPR004333, GO 注释号为 GO:0003677。

1.5 SBP 家族基因的命名

对预测基因统一命名, 冠以物种名缩写 Gm; 预测基因所在的不同染色体 01、02、03 予以区分, 以及

基因在染色体的位置编号,如 GmSPL01_1、GmSPL01_2、GmSPL01_3 等。

1.6 SBP 转录因子的生物信息学分析

利用 ClustalW (含 BOXSHADE) (<http://bio-web.pasteur.fr>) 对其氨基酸序列进行多重比对并进行人工校对。使用 PHYLIP (版本 3.5c) (<http://bioweb.pasteur.fr>) 构建系统发育树。利用生物信息资源 (<http://www.ch.embnet.org/software/COILS2-form.html>) 分析蛋白质卷曲螺旋可预测 $\alpha 2$ 螺旋的卷曲螺旋排列方式。运用 Phyre (<http://www.sbg.bio.ic.ac.uk/phyre/>) 的三级结构预测功能对 SBP 转录因子进行三级结构图化分析。

2 结果与分析

2.1 SBP 转录因子在大豆基因组中的分布

大豆中 44 个 SBP 基因均匀分布在除 14 号染色体之外的所有 19 条染色体上(图 1),其中在第 13 号染色体上有 5 个 SBP 基因。

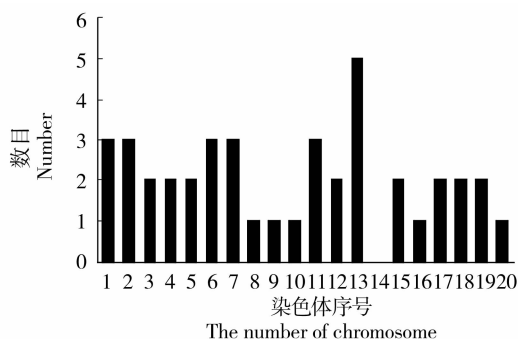


图 1 SBP 基因在大豆染色体上的分布

Fig. 1 Distribution of SBP genes in soybean chromosomes

2.2 大豆 SBP 转录因子 DNA 结合结构域序列和结构分析

对 44 个大豆 SBP 转录因子进行多序列比对,结果表明所有这些序列均有保守的 DNA 结合结构域(图 2),长度约为 80 个残基,大部分位于 N-端。序列图标分析(图 3)表明,SBP 结构域可分为 3 个区域:N-端保守区(3-30)、中间非保守区(31-41)和 C-端保守区(42-76)。C-端序列的保守程度高于 N-端,除形成锌指结构序列模体残基半胱氨酸和组氨酸外,另有一个保守序列模式“KRSCR[RK][RK]Lx2HNxRR[RK][KR]”,主要由碱性氨基酸残基组成。Birkenbihl^[1]等报道 SBP 转录因子在胞浆内合成后进入细胞核,需要核定位信号介导,而这一保守序列片段就是介导该转录因子以主动运输方式通过核膜进入细胞核的识别信号,并利用基因工程定位突变方法证明,该序列片段第 3 位保守残基丝

氨酸十分关键,将它突变成天门冬酰胺(Asp)后进入细胞核的能力降低。

2.3 SBP 转录因子 DNA 结合结构域三维结构与功能预测分析

转录因子与 DNA 序列顺式元件结合,是基因转录调控的关键步骤。除了与 DNA 结合外,蛋白质分子间的相互作用也是实现复杂调控机制的基础。在拟南芥中也只有 3 个基因 *SPL03*、*SPL08* 和 *SPL14* 的功能有文献报道^[13-15],其它成员的功能尚待研究。大豆中 SBP 转录因子的功能研究尚未见报道,为进一步搞清它们之间的关系,对 SBP 转录因子基因家族进行序列分析,了解该基因家族演化进程。

首先,用 ClustalW 程序以 SBP 结构域构建了系统发育树(图 4),大豆 SBP 基因家族 44 个转录因子的 SBP 结构域序列之间的异同,与基因结构、序列长度有一定关系,但并不完全一一对应。

为进一步分析这 44 个转录因子,根据它们之间的进化关系,将其分为 4 组(表 2)。

表 2 大豆 44 个 SBP 转录因子分组

Table 2 Groups of 44 SBP transcription factors of soybean

分组	名称	长度	分组	名称	长度
Group	Name	Length/bp	Group	Name	Length/bp
G1	GmSPL15_36	1153	G3	GmSPL13_30	1298
G1	GmSPL17_38	811	G3	GmSPL06_15	1683
G1	GmSPL09_22	1096	G3	GmSPL17_39	498
G1	GmSPL10_23	1507	G3	GmSPL03_8	277
G1	GmSPL07_20	954	G3	GmSPL01_2	238
G1	GmSPL02_4	1378	G3	GmSPL15_35	716
G1	GmSPL07_19	133	G3	GmSPL01_3	376
G1	GmSPL13_29	95	G3	GmSPL07_17	274
G1	GmSPL13_31	127	G3	GmSPL11_25	104
G1	GmSPL15_34	127	G3	GmSPL01_1	482
G2	GmSPL04_11	365	G4	GmSPL19_43	429
G2	GmSPL18_40	407	G4	GmSPL06_16	185
G2	GmSPL06_14	203	G4	GmSPL12_27	412
G2	GmSPL19_42	365	G4	GmSPL13_32	175
G2	GmSPL05_13	472	G4	GmSPL13_33	147
G2	GmSPL11_24	139	G4	GmSPL20_44	934
G2	GmSPL11_26	377	G4	GmSPL12_28	196
G2	GmSPL16_37	365	G4	GmSPL07_18	994
G2	GmSPL08_21	581	G4	GmSPL18_41	350
G3	GmSPL02_5	464	G4	GmSPL03_7	452
G3	GmSPL04_10	564	G4	GmSPL02_6	358
G3	GmSPL05_12	636	G4	GmSPL03_9	639

第 1 组有 10 个转录因子,最大的 GmSPL10_23 有 1507 bp,最小的 GmSPL13_29 有 95 bp,它们共同

```

GmSPL01_1 VQOVYGGN.MDLSS.....SKDYHRRHRVCDIHSKRTAKVIWNGIEQRFCCQGS.....RFHLLVEFDDCKRSCRRLAGHMRRRRK
GmSPL01_3 AAKNISFHQSGLR.....GRHIKDFEVLRSGLLCFRISQAQVPRREMEFGG.....NMFCLANRDNONGNTMNNGNTFAWDSMETP
GmSPL02_5 VQOVYGGN.MDLSS.....SKDYHRRHRVCDIHSKRTAKVIWNGIEQRFCCQGS.....RFHLLAEFDDCKRSCRRLAGHMRRRRK
GmSPL02_6 RQOVAGGK.VDLS.....GAKAVYSRHVCAHSKSPFTVIWAGLEQRFCCQGS.....RFHLLSEFDDCKRSCRRLAGHMRRRRK
GmSPL03_7 AATVDFTH.HHHFT.....PQPHHHNQHHMHRRHHAFDPRAFHGASNTSYP.....PPHSMLSLDP.LPHVNAVGPQHGCPG...S
GmSPL04_11 ACLVDGDN.SDLSN.....CRDYHRRHRVCEVHSKRTAPEVTIGGFKQRFCCQGS.....RFHLLAEFDDCKRSCRRLAGHMRRRRK
GmSPL05_13 SCLVDGCH.ADLSN.....CRDYHRRHRVCEVHSKTAQVSIIGGQKQRFCCQGS.....RFHLLSEFDDCKRSCRRLAGHMRRRRK
GmSPL11_26 SQOVVEGG.LDLS.....AKDYHRRHRVCEVHSKSPKVVITGLERRFCCQGS.....RFHLLSEFDDCKRSCRRLAGHMRRRRK
GmSPL12_27 SQOAEERG.ADLT.....DAKRYHRRHRVCEVHSKAPVVVWVAGLRQRFCCQGS.....RFHLLAEFDECKRSCRRLAGHMRRRRK
GmSPL16_37 QCSVDGDN.SDLS.....CRDYHRRHRVCEVHSKRTPVVWVGGKQRFCCQGS.....RFHLLGEFDEVKRSCRRLAGHMRRRRK
GmSPL17_39 VQOVYGGN.KDLS.....CKDYHRRHRVCEVHSKTAIVWNGIEQRFCCQGS.....RFHLLAEFDDCKRSCRRLAGHMRRRRK
GmSPL18_40 SQOVVEGG.LDLS.....AKDYHRRHRVCEVHSKSPKVVITGLERRFCCQGS.....RFHLLSEFDDCKRSCRRLAGHMRRRRK
GmSPL18_41 RQOVVEGG.VDLS.....GAKAVYSRHVCAHSKSPFTVIWAGLEQRFCCQGS.....RFHLLSEFDECKRSCRRLAGHMRRRRK
GmSPL19_42 QCSVDGDN.SDLS.....CRDYHRRHRVCEVHSKRTPVVWVGGKQRFCCQGS.....RFHLLGEFDEVKRSCRRLAGHMRRRRK
GmSPL19_43 DHAPMASD.AKLSSESLSNGLKFGQKIYFEDVGLATPATSLTSSSSSSAATVTSSSSSSRKGRVLRHFHCLPEFDDCKRSCRRLAGHMRRRRK
GmSPL15_36 VQOVYGGN.ADLSN.....VKDYHRRHRVCEVHSKTAQALVGNVMQRFCCQGS.....RFHLLGEFDEVKRSCRRLAGHMRRRRK
GmSPL17_38 VQOVYGGN.ADLSN.....AKDYHRRHRVCEVHSKTAQALVGNVMQRFCCQGS.....RFHLLGEFDEVKRSCRRLAGHMRRRRK
GmSPL20_44 MCOVDNR.EDLSK.....AKDYHRRHRVCEVHSKASKALLANQMCRFCQGS.....RFHLLSEFDECKRSCRRLAGHMRRRRK
GmSPL13_30 RQOVPAQE.VDISE.....LKGYHRRHRVCEVHSKANAATVMLNGEARRVCCQGG.....KFHWLSDFDECKRSCRRLAGHMRRRRK
GmSPL09_22 VQOVYGGN.ADLSN.....AKDYHRRHRVCEVHSKTAQALVGNVMQRFCCQGS.....RFHLLGEFDEVKRSCRRLAGHMRRRRK
GmSPL10_23 VQOVYGGN.ADLSK.....AKDYHRRHRVCEVHSKASRALVGNAMCRFCQGS.....RFHLLGEFDECKRSCRRLAGHMRRRRK
GmSPL07_20 ..KVQDGR.ADLSN.....AKDYHRRHRVCEVHSKATMALVGNVMQRFCCQGS.....RFHLLGEFDECKRSCRRLAGHMRRRRK
GmSPL07_18 MCOVDNR.EDLSK.....AKDYHRRHRVCEVHSKASKALLANQMCRFCQGS.....RFHLLSEFDECKRSCRRLAGHMRRRRK
GmSPL06_15 VQOVYGGN.KDLSG.....CKDYHRRHRVCEVHSKTAQALVGNVMQRFCCQGS.....ICIDYDIMALGYAGAKKLYSFLGPTCVS
GmSPL02_4 VQOVYGGN.ADLSK.....AKDYHRRHRVCEVHSKASRALVGNAMCRFCQGS.....RFHLLGEFDECKRSCRRLAGHMRRRRK
GmSPL01_2 LDPINQGGYSDLSN.....CRDYHRRHRVCEVHSKTAQVITGAQKQRFCCQGS.....RGTVFSQATINDALAAATK.....
GmSPL06_16 SQOAEERG.ADLT.....DAKRYHRRHRVCEVHSKAPVVVWVAGLRQRFCCQGS.....RFHLLAEFDECKRSCRRLAGHMRRRRK
GmSPL07_17 TLHTSTAAEGYAG.....GSHSHHHQCLYGGEG..CH..VQADPHLTLHLG.....RRHVFEDVTTGVDVNNATPLGNRRVCSGG
GmSPL11_24 FQADEEG.VNLSM.....AKSYHRRHRVCEVHSKAPVVVWVSSIRQRFCCQGS.....FDDTKRSCRRLAGHMRRRRK
GmSPL13_31 SQOVYGGN.ADLS.....AKPYHRRHRVCEVHSKAPVVVWVIGDQHQRFCCQGS.....RFHLLSEFDDCKRSCRRLAGHMRRRRK
GmSPL13_32 SQOAEIQQ.ADLT.....VAKRYHRRH.....KPKVMSRC.....RFHLLAEFDECKRSCRRLAGHMRRRRK
GmSPL15_34 SQOVYGGN.ADLS.....AKPYHRRHRVCEVHSKAPVVVWVIGDQHQRFCCQGS.....RFHLLSEFDDCKRSCRRLAGHMRRRRK
GmSPL12_28 SQOAEERG.ADLT.....VAKRYHRRH.....KPKVMSLC.....RFHLLAEFDECKRSCRRLAGHMRRRRK
GmSPL03_8 APHT.TTAAEGYAG.....GGHRLHHHCLYGGEG..CH..VQADPHLTLHLG.....RRHVFEDATNGDDANMVVSLGNRRVCSGG
GmSPL04_9 RQOVVEGG.VDLS.....NAKTYHRRHRVCEVHSKSPFTVIWAGLQRFCCQGSSELILAAPRFHCLPEFDDCKRSCRRLAGHMRRRRK
GmSPL04_10 VCLVYGGN.KDLSG.....CKDYHRRHRVCEVHSKTAIVWNGIEQRFCCQGS.....RFHLLAEFDDCKRSCRRLAGHMRRRRK
GmSPL05_12 VQOVYGGN.KDLS.....CKDYHRRHRVCEVHSKTAIVWNGIEQRFCCQGS.....RFHLLAEFDDCKRSCRRLAGHMRRRRK
GmSPL15_35 RQOVVEGG.ADISE.....LKGYHRRHRVCEVHSKANAATVMLNGEARRVCCQGG.....KFHWLSDFDECKRSCRRLAGHMRRRRK
GmSPL08_21 SCLVDGCH.SDLSN.....CRDYHRRHRVCEVHSKTAQVITIGGQKQRFCCQGS.....RFHLLSEFDECKRSCRRLAGHMRRRRK

```

图 2 大豆 SBP 转录因子 DNA 结合结构域多序列联配

Fig. 2 Multiple sequence alignment of DNA-binding structure domain in SBP transcription factors of soybean

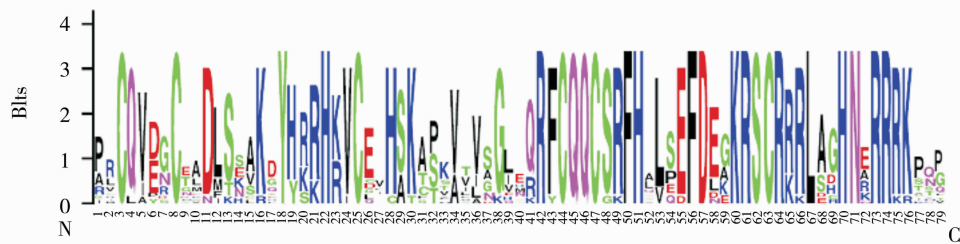


图 3 大豆 SBP 转录因子 DNA 结合结构域序列图标

Fig. 3 The sequence icon of DNA-binding structure domain in SBP transcription factors of soybean

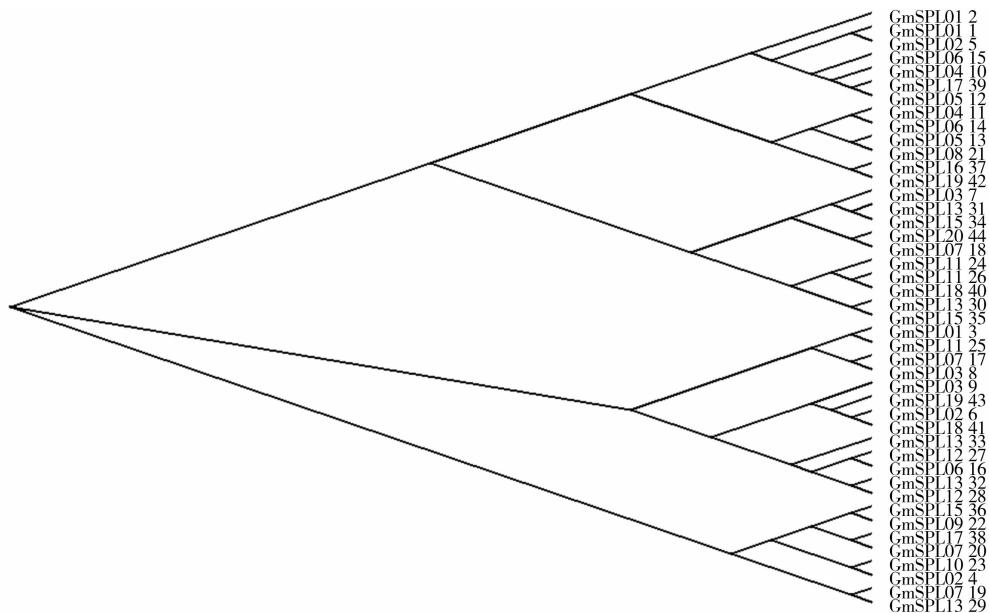


图 4 大豆 SBP 转录因子基因结构域系统发育树

Fig. 4 Phylogenetic tree of soybean SBP transcription factors genetic structure domain

的特征是含有高度保守的 [CR/SADLSK/EARDY-HRRHKV] 序列,第 2 组除了和第 1 组有共同的保守序列外,还在 SBP-box N 端有高度保守的 [SSPSGSSKPAP] 序列,第 3 组多数则含有 [DLS/GCKDYHKGHKVCCEVHSKT] 保守序列,而第 4 组则包含 [RFHQLPEFDQGRSCRRRLAGHNERRRK] 高度保守序列;推测可能与结合不同靶基因有关。

对这 4 组分别作了 SBP 结合域的三级结构分析,发现 4 组成员均都含有 33 个氢键,2 个 α -螺旋,3 个 β 转角,但无规卷曲的数目(图 5)。

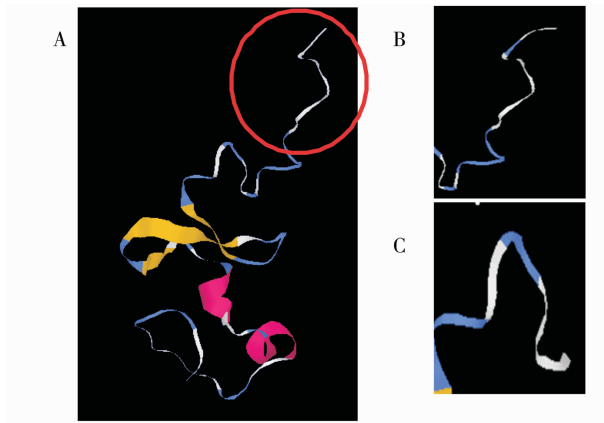


图 A 中的红圈所示即为各组中有差异的无规卷曲部分。A 为 G1 组;B 为 G2 组;C 为 G3、G4 组。

The red circle of Fig. 5-A marked the difference part among groups.

A is G1 group; B is G2 group; And C is the G3 and G4 groups.

图 5 转录因子 SBP-box 三级结构

Fig. 5 The SBP-box tertiary structure

为进一步预测其功能,通过预测它们的启动子并在 PlantCARE (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) 上预测其可能包含的顺式作用元件。表 3 为大豆 SBP 转录因子启动子区预测结果,44 个转录因子在启动子区的各种功能元件有所不同,除了最基本的 TATA-box、CAAT-box 等,有个别的转录因子如 GmSPL15_36、GmSPL19_42 等还含有 W-box、MBS、MBSII 等一些其它转录因子的结合位点,其中 W-box 是 AP2 类转录因子结合位点,涉及植物生长发育如糖酵解和脂肪酸循环途径中关键酶的合成^[15],MBS、MBSII 则是 MYB 转录因子家族特有的结合位点。推测其可能与其它转录因子互作共同完成植物生长发育相关的功能。此外还有大量光刺激应答元件,推测其参与光合作用以及开花等光反应相关的功能。其余是一些热激效应以及脱落酸、真菌侵染等响应外界环境刺激的功能元件,推测与应答生物与非生物胁迫反应相关。

表 3 大豆 SBP 转录因子启动子区预测结果

Table 3 The forecasting results of promoter area of SBP transcription factors of soybean

位点名称 Site name	来源 Origin	位置 Position	序列 Sequence	功能 Function
ABRE	<i>Hordeum vulgare</i>	954	GCAACGT-GTC	Cis-acting element involved in the abscisic acid responsiveness
AE-box	<i>Arabidopsis thaliana</i>	1142	AGAAA-CAA	Part of a module for light response
ARE	<i>Zea mays</i>	875	TGGTTT	Cis-acting regulatory element essential for the anaerobic induction
ATCT-motif	<i>Arabidopsis thaliana</i>	285	AATCTA-ATCT	Part of a conserved DNA module involved in light responsiveness
ATCT-motif	<i>Pisum sativum</i>	917	AATCTA-ATCC	Part of a conserved DNA module involved in light responsiveness
ATCT-motif	<i>Arabidopsis thaliana</i>	729	AATCTA-ATCT	Part of a conserved DNA module involved in light responsiveness
Box 4	<i>Petroselinum crispum</i>	550	ATTAAT	Part of a conserved DNA module involved in light responsiveness
Box I	<i>Pisum sativum</i>	2	TTT CAAA	Light responsive element
Box I	<i>Pisum sativum</i>	270	TTT CAAA	Light responsive element
Box III	<i>Pisum sativum</i>	336	CATTT ACACT	Protein binding site
Box-W1	<i>Petroselinum crispum</i>	436	TTGACC	Fungal elicitor responsive element
CGTCA-motif	<i>Hordeum vulgare</i>	70	CGTCA	Cis-acting regulatory element involved in the Me-JA-responsiveness
CGTCA-motif	<i>Hordeum vulgare</i>	1353	CGTCA	Cis-acting regulatory element involved in the Me-JA-responsiveness
CGTCA-motif	<i>Hordeum vulgare</i>	1193	CGTCA	Cis-acting regulatory element involved in the Me-JA-responsiveness
ERE	<i>Dianthus caryophyllus</i>	2	ATTT CAAA	Ethylene-responsive element

(续表3)

位点名称 Site name	来源 Origin	位置 Position	序列 Sequence	功能 Function	位点名称 Site name	来源 Origin	位置 Position	序列 Sequence	功能 Function
G-Box	<i>Pisum sativum</i>	956	CACGTT	Cis-acting regulatory element involved in light responsiveness	Skn ⁻¹ _motif	<i>Oryza sativa</i>	1354	GTCAT	Cis-acting regulatory element required for endosperm expression
GARE-motif	<i>Brassica oleracea</i>	935	TCTGTTG	Gibberellin-responsive element	TC-rich repeats	<i>Nicotiana tabacum</i>	572	AT-TCTCTA-AC	Cis-acting element involved in defense and stress responsiveness
GARE-motif	<i>Brassica oleracea</i>	1089	AAACA-GA	Gibberellin-responsive element	TC-rich repeats	<i>Nicotiana tabacum</i>	1255	ATTTTCTC-CA	Cis-acting element involved in defense and stress responsiveness
GT1-motif	<i>Solanum tuberosum</i>	476	AATCCA-CA	Light responsive element	TC-rich repeats	<i>Nicotiana tabacum</i>	724	AT-TCTCTA-AC	Cis-acting element involved in defense and stress responsiveness
HD-Zip 1	<i>Arabidopsis thaliana</i>	341	CAAT (A/T)ATTG	Element involved in differentiation of the palisade mesophyll cells	TCA-element	<i>Nicotiana tabacum</i>	1151	CCATC TTTTT	Cis-acting element involved in salicylic acid responsiveness
HD-Zip 1	<i>Arabidopsis thaliana</i>	833	CAAT (A/T)ATTG	Element involved in differentiation of the palisade mesophyll cells	TCA-element	<i>Brassica oleracea</i>	1179	GAGAA GAATA	Cis-acting element involved in salicylic acid responsiveness
HD-Zip 2	<i>Arabidopsis thaliana</i>	341	CAAT (G/C)ATTG	Element involved in the control of leaf morphology development	TCT-motif	<i>Arabidopsis thaliana</i>	292	TCTTAC	Part of a light responsive element
HD-Zip 2	<i>Arabidopsis thaliana</i>	833	CAAT (G/C)ATTG	Element involved in the control of leaf morphology development	TCT-motif	<i>Arabidopsis thaliana</i>	425	TCTTAC	Part of a light responsive element
HSE	<i>Brassica oleracea</i>	387	AAAAA ATTTC	Cis-acting element involved in heat stress responsiveness	TGACC-motif	<i>Hordeum vulgare</i>	70	TGACC	Cis-acting regulatory element involved in the MeJA-responsiveness
MBS	<i>Zea mays</i>	435	CGGTCA	MYB Binding Site	TGACC-motif	<i>Hordeum vulgare</i>	1353	TGACC	Cis-acting regulatory element involved in the MeJA-responsiveness
MBSII	<i>Petunia hybrida</i>	584	AAAAGT-TAGTTA	MYB binding site in flavonoid biosynthetic genes regulation	W box	<i>Arabidopsis thaliana</i>	436	TTGACC	Cis-acting regulatory element involved in auxin responsiveness
P-box	<i>Oryza sativa</i>	747	CCT TTG	Gibberellin-responsive element	AuxRR-core	<i>Nicotiana tabacum</i>	1367	GGTCCAT	Cis-acting regulatory element involved in auxin responsiveness
RY-element	<i>Helianthus annuus</i>	1098	CATG-CATG	Cis-acting regulatory element involved in seed-specific regulation	box S	<i>Arabidopsis thaliana</i>	873	AGCCACC	
Skn ⁻¹ _motif	<i>Oryza sativa</i>	218	GTCAT	Cis-acting regulatory element required for endosperm expression	rbcS-CMA7a	<i>Lemna gibba</i>	1224	GTCGATA-AGG	Part of a light responsive element
Skn ⁻¹ _motif	<i>Oryza sativa</i>	941	GTCAT	Cis-acting regulatory element required for endosperm expression					
Skn ⁻¹ _motif	<i>Oryza sativa</i>	221	GTCAT	Cis-acting regulatory element required for endosperm expression					

3 讨论

通过对大豆 SBP 转录因子家族基因的分析,得知该基因家族不同成员之间的相互关系和演化历程,以及它们的生物学功能。如逆境胁迫应答、生长发育、光合作用等。同时证明了预测数据的精确性和可靠性。

SBP 转录因子家族是植物特异的转录因子家族。最初发现它是调控花发育相关基因的一种转录因子,并因此得名 Squamosa 启动子结合蛋白(Squamosa promoter Binding Protein, SBP)。然后在拟南芥、苔藓、玉米等植物中陆续发现多个 SBP 转录因子,并具有许多重要生物学功能,包括花和果实发育、孢子发育、激素应答、抗霉菌侵蚀等。通过生物信息学的方法预测获得 44 条 SBP 转录因子,并通过注释鉴定以及对其进行功能分析发现 SBP 家族基因与花发育等植物生长发育、抗逆境胁迫、抗霉菌侵蚀以及热激效应等功能相关,证明该研究预测方法的可行性和可靠性。采用的方法是基因预测和数据库构建的经典方法,并在数据来源和分析中采用新的标准。预测的结果相对于 plantTFDB 少了 2 个,是由于所采用的数据不同,采用的是大豆基因组数据作为原始数据,并且种子文件更加精确,而 plantTFDB 采用的是早期 EST 数据聚类的结果,序列数据不全,导致基因断裂分散所致。

参考文献

- [1] Klein J, Saedler H, Huijser P, et al. A new family of DNA binding proteins includes putative transcriptional regulators of the *Antirrhinum majus* floral meristem identity gene *SQUAMOSA* [J]. *Molecular and General Genetics*, 1996, 250(1): 7-16.
- [2] Yamasaki K, Kigawa T, Inoue M, et al. *Arabidopsis* SBP-domain fragment with a disrupted C-terminal zinc-binding site retains its tertiary structure [J]. *FEBS Letters*, 2006, 580(8): 2109-2116.
- [3] Stone J M, Liang X, Neel E R, et al. *Arabidopsis* AtSPL14, a plant-specific SBP-domain transcription factor, participates in plant development and sensitivity to fumonisin B1 [J]. *The Plant Journal*, 2005, 41(5): 744-754.
- [4] Cardon G, Hohmann S, Huijser P, et al. Molecular characterization of the *Arabidopsis* SBP-box genes [J]. *Gene*, 1999, 237(1): 91-104.
- [5] Birkenbihl R P, Jach G, Saedler H, et al. Functional dissection of the plant-specific SBP-domain: overlap of the DNA-binding and nuclear localization domains [J]. *Journal Molecular Biology*, 2005, 352(3): 585-596.
- [6] Moreno M A, Harper L C, Dellaporta S L, et al. *Freeling* M liguleless1 encodes a nuclear-localized protein required for induction of ligules and auricles during maize leaf organogenesis [J]. *Genes Development*, 1999, 16(3): 620-628.
- [7] Cardon G, Hohmann S, Huijser P, et al. Functional analysis of the *Arabidopsis thaliana* SBP-box gene *SPL3*: a novel gene involved in the floral transition [J]. *The Plant Journal*, 1997, 12(2): 367-77.
- [8] Wang H, Nussbaum-Wagler T, Li B, et al. The origin of the naked grains of maize [J]. *Nature*, 2005, 436(10): 714-719.
- [9] Huijser P, Klein J, Lonnig W E, et al. Bracteomania, an inflorescence anomaly, is caused by the loss of function of the MADS-box gene *squamosa* in *Antirrhinum majus* [J]. *EMBO Journal*, 1992, 11(4): 1239-1249.
- [10] Kropat J, Tottey S, Birkenbihl R P, et al. A regulator of nutritional copper signaling in *Chlamydomonas* is an SBP domain protein that recognizes the GTAC core of copper response element [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2005, 102(51): 18730-18735.
- [11] Lännpää M, Jänänen I, Hölttä-Vuori M, et al. A new SBP-box gene *BpSPL1* in silver birch (*Betula pendula*) [J]. *Physiologia Plantarum*, 2004, 120(3): 491-500.
- [12] Manning K, Tör M, Poole M, et al. A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening [J]. *Nature Genetics*, 2006, 38: 948-952.
- [13] Unte U S, Sorensen A M, Pesaresi P, et al. *SPL8*, an SBP-box gene that affects pollen sac development in *Arabidopsis* [J]. *Plant Cell*, 2003, 15(4): 1009-1019.
- [14] Wu G, Poethig R S. Temporal regulation of shoot development in *Arabidopsis thaliana* by miR156 and its target *SPL3* [J]. *Development*, 2006, 133(18): 3539-3547.
- [15] Yamasaki K, Kigawa T, Inoue M, et al. A novel zinc-binding motif revealed by solution structures of DNA-binding domains of *Arabidopsis* SBP-family transcription factors [J]. *Journal Molecular Biology*, 2006, 337: 49-63.
- [16] Maeo K, Tokuda T, Ayame A, et al. An AP2-type transcription factor, WRINKLED1, of *Arabidopsis thaliana* binds to the AW-box sequence conserved among proximal upstream regions of genes involved in fatty acid synthesis [J]. *Plant Journal*, 2009, 60(3): 476-487.