

# 高粱淀粉合成酶新基因 *SSV* 的鉴定与特征分析

于桂玲<sup>1</sup> 肖前林<sup>2</sup> 魏彬<sup>2</sup> 杨进<sup>1</sup> 王丹<sup>3</sup> 刘汉梅<sup>1</sup>

(<sup>1</sup> 四川农业大学生命科学学院, 四川 雅安 625014; <sup>2</sup> 四川农业大学农学院,

四川 成都 611130; <sup>3</sup> 濮阳市农业科学院, 河南 濮阳 457000)

**摘要:** 淀粉合成酶(SS)是植物淀粉生物合成过程中的关键酶之一,目前已鉴定并揭示功能的淀粉合成酶有5个亚型。为利用丰富的基因组数据鉴定高粱基因组里的淀粉合成酶新亚型基因,本文利用生物信息学和分子生物学方法,首次鉴定并分离了编码高粱淀粉合成酶的新亚型基因 *SbSSV*,并对此基因的内含子-外显子结构、表达特性等进行了分析。结果显示,该基因的全长 ORF 为 2 097 bp, 其外显子数量、长度及内含子的位置分布等与玉米和水稻的直系同源基因基本一致,推导的蛋白具有细菌糖原合成酶和植物淀粉合成酶特有的糖基催化和糖基转移保守结构域。系统进化分析表明,*SbSSV* 与已知的 *SSIV* 亚型亲缘关系最近,推测本研究鉴定的 *SbSSV* 基因编码高粱淀粉合成酶新亚型。定量 PCR 分析表达特性结果显示 *SbSSV* 主要在高粱叶片中表达,其表达受光诱导,具有昼夜节律特性。研究结果可为深入揭示和完善植物淀粉合成代谢机制,以及改良植物产量和品质提供理论基础。

**关键词:** 高粱; 淀粉; 淀粉合成酶 V; 新基因鉴定

DOI: 10. 11869/j. issn. 100-8551. 2016. 01. 0001

植物淀粉是自然界重要的贮藏多糖,是人类食物和动物饲料中最主要的能量物质。植物贮藏淀粉主要在胚乳细胞的造粉体里合成,占成熟籽粒重量的 70% 以上<sup>[1]</sup>。根据结构的差异,可将淀粉分为直链淀粉和支链淀粉。直链淀粉主要是以  $\alpha$ -1,4-糖苷键连接的多糖链,其分子基本呈线形。支链淀粉是以  $\alpha$ -1,6-糖苷键和  $\alpha$ -1,4-糖苷键共同形成长度不一的分支,并以一定的次序相结合的簇状结构,其分子量远大于直链淀粉<sup>[2]</sup>。

植物体内的淀粉合成是一系列复杂的生化反应过程,涉及多个酶的协调作用。催化酶类主要包括 ADPG 焦磷酸化酶 (ADP-Glc pyrophosphorylase, AGPase)、淀粉合成酶 (starch synthase, SS)、淀粉分支酶 (starch branching enzyme, SBE)、去分支酶 (de-branching enzyme, DBE) 和淀粉磷酸化酶 (starch phosphorylase, SP)<sup>[3]</sup> 等。AGPase 催化淀粉合成所需的前体底物 ADPG 焦磷酸葡萄糖的形成; SS 催化  $\alpha$ -1,4-糖苷键的形成,延长淀粉链; SBE 催化形成支链淀粉分支处的  $\alpha$ -1,6-

-糖苷键,产生分支; DBE 的作用是水解  $\alpha$ -1,6-糖苷键,去除支链淀粉中过多的分支<sup>[2,4]</sup>。

目前,已鉴定并揭示功能的淀粉合成酶有5个亚型,即负责直链淀粉和超长支链淀粉合成的颗粒结合型淀粉合成酶 (granule-bound starch synthase, GBSS)、可溶性淀粉合成酶 (soluble starch synthases, SSI-III), 及负责淀粉颗粒起始形成的淀粉合成酶 IV (*SSIV*)<sup>[5]</sup>。其中 SSI-III 分别负责短链 (聚合度, degree of polymerization or chain length, DP of 8~12)、中等长度支链 (DP of 13~25) 和长链 (DP of >30) 的合成<sup>[2]</sup>。编码不同淀粉合成酶亚型的基因都来源于绿色植物分化早期发生的基因复制,复制基因随后经功能分化,形成具有不同底物特异性的亚型<sup>[3]</sup>。近年来,对一些淀粉合成酶基因突变体的研究发现,有的亚型间存在功能冗余,如 *SSIII* 能形成淀粉起始颗粒, *SSIV* 能合成短的淀粉链<sup>[5]</sup>。

目前尽管分离了参与植物淀粉合成的主要酶类,但淀粉生物合成的复杂机制还未能透彻解析,如合成

收稿日期: 2015-01-08 接受日期: 2015-06-03

基金项目: 国家重点基础研究发展计划 (973) (2011CB100100、2014CB138202), 四川省教育厅项目 (12ZA70)

作者简介: 于桂玲,女,主要从事生物信息学研究。E-mail: 1779719265@qq.com

通讯作者: 刘汉梅,女,副教授,主要从事植物育种与分子生物学研究。E-mail: hanmeil@163.com

过程是否存在其他酶类的参与;哪些分子参与合成代谢调控,使多种酶协同有序地行使功能等。随着测序技术的发展,许多生物的基因组序列被测定,为分离鉴定新基因开辟了新途径。本研究利用生物信息学方法,获得完整的 *SbSSV* 基因序列,同时采用 RT-PCR 方法,从高粱叶片中成功分离该基因的全长 CDS 序列。本文首次对 *SbSSV* 的鉴定、特征分析及与其他淀粉合成酶的系统发生关系进行报道,以期揭示和完善植物淀粉合成代谢机制奠定基础。

## 1 材料与方法

### 1.1 材料与试剂

试验材料:四川高粱农家种宜宾矮花陀。将高粱播种于含有蛭石的营养土中,在光照培养箱内培养(16 h 光照,25 ℃;8 h 黑暗,22 ℃)。待幼苗长至五叶一心,取光照 8 h 的根、茎、叶 3 部分组织,检测基因的组织表达特异性。同时取正常光照培养条件下,光照 0、4、8、16 h 和黑暗 2、4、8 h 的叶片,检测基因表达的昼夜节律特性。为进一步验证基因表达是否受光诱导,将材料进行更长时间的光照和黑暗处理,一部分材料从光照 16 h 后再持续光照 24 h,每隔 8 h 取样;另一部分材料黑暗 8 h 后再持续黑暗 24 h,每隔 8 h 取样。将所有材料在液氮中速冻后于 -80 ℃ 保存备用。

主要试剂:RNA 提取试剂盒,购自北京 Tiandz 公司;琼脂糖凝胶电泳 DNA 回收试剂盒,购自广州 Omega 公司;Trizol、氯仿、异丙醇均购自上海 Invitrogen 公司;反转录试剂盒(PrimeScript RT reagent Kit)、DL marker 5000、DNA 聚合酶、pMD19 - T Vector, SYBR PrimeScript RT-PCR Kit Perfect Real Time 试剂盒、高保真酶 KOD 试剂盒均购自大连 Takara 公司。

### 1.2 试验方法

1.2.1 总 RNA 的提取和 cDNA 第 1 链的合成 使用 Trizol 试剂提取高粱五叶期叶片的总 RNA,经 DNaseI 处理后抽提纯化,方法参照 RNA 提取试剂盒的操作说明。RNA 经 1.0% 琼脂糖凝胶电泳检测,并用核酸蛋白仪(NanoDrop NDY2000,上海创萌生物公司)测定 OD<sub>260</sub>/OD<sub>280</sub> 值,判断 RNA 的纯度和完整性。检测合格的 RNA 样本,用反转录试剂盒 PrimeScript RT reagent Kit,参照说明书的方法进行反转录。获得的 cDNA 产物存放于 -20 ℃ 保存备用。

1.2.2 RT-PCR 扩增高粱 SSV 基因 用玉米 SSV 的 CDS 序列作为查询序列,搜索高粱基因组获得同源区段。根据同源区段的序列信息设计引物,上游引物 F

为 TAGCAGGCAGCCATGGAGAC,下游引物 R 为 CTAGCATGCCAGTTCACAGC。以反转录获得的 cDNA 为模板进行 PCR 扩增,反应体系为稀释 10 倍的 cDNA 1 uL,上下游引物各 0.5 uL, KOD Buffer 12.5 uL, dNTP 5 uL, KOD 聚合酶 0.5 uL, ddH<sub>2</sub>O 5 uL, 共 25 uL 反应体系。PCR 反应程序为:94 ℃ 预变性 4 min, 98 ℃ 变性 10 s, 60 ℃ 退火 30 s, 68 ℃ 延伸 2 min 30 s, 共 34 个循环;68 ℃ 延伸 7 min, 4 ℃ 保存。取 5 uL PCR 产物于 1% 琼脂糖凝胶进行电泳。PCR 产物回收参照广州飞扬生物公司回收试剂盒说明书进行。

1.2.3 pMD19 - T 载体连接和重组质粒的鉴定与筛选 将纯化后的 PCR 产物与 pMD19 - T 载体连接,并将产物转化入大肠杆菌感受态细胞。将转化细胞涂布于含 25 μg · mL<sup>-1</sup> IPTG, 15 μg · mL<sup>-1</sup> X-gal, 100 μg · mL<sup>-1</sup> Amp<sup>r</sup> 的 LB 平板上进行蓝白斑筛选,挑选白色菌斑,将检测正确的阳性克隆送公司测序。

1.2.4 实时荧光定量 PCR (qRT-PCR) 根据 RT-PCR 获得的 SSV 基因全长序列,应用 Primer premier 5.0 软件跨内含子设计定量引物(F: TTCACCGCCCTG ATCGTCTG, R: ATTGCAGCCCATGTCTTGAACC),扩增长度为 150 bp。选择高粱组成型表达基因  $\beta$  - *actin* (Accession number: GU138757.1) 作为内参基因(F: TGGCATCTCTCAGCACATTCC, R: AATGGCTCTCTCG GCTTGC),扩增片段长度为 130 bp。qRT-PCR 采用 SYBR PrimeScript RT-PCR Kit (Perfect Real Time) 试剂盒,反应体系为 10 μL,包括 5 μL SYBR GreenII, 0.3 μL 上下游引物(10 μmol · L<sup>-1</sup>), 3.4 μL ddH<sub>2</sub>O 和 1 μL cDNA。实时荧光定量使用 BIO-RAD 定量分析仪(Connect Real-time PCR detection system,成都百乐科技有限公司)。反应程序为:95 ℃ 预变性 30 s, 95 ℃ 变性 5 s, 59 ℃ 延伸 30 s, 共 40 个循环;最后添加溶解曲线。本试验使用参照基因的  $\Delta\Delta C_t$  法,表达差异等于  $2^{-\Delta\Delta C_t}$ ,  $\Delta C_t = C_{t_{\text{目标基因}}} - C_{t_{\text{actin}}}$ 。相对定量是基于目标基因相对内参基因表达量的比较<sup>[6]</sup>。

1.2.5 生物信息学分析 根据 NCBI (<http://www.ncbi.nlm.nih.gov/>) 公布的玉米 SSV (*ZmSSV*) 的 CDS 序列(NM\_001130131)为查询序列,在 Gramene (<http://www.gramene.org/>) 数据库中用 BLASTn 程序搜索高粱基因组序列,参数设置如下:显示的最大结果数,100;期望值(expect threshold),10;其余参数默认。下载搜索到的同源区段序列,与 *ZmSSV* 的 CDS 进行精细比对,并结合内含子的剪切规则识别基因的外显子和内含子,拼接 *SbSSV* 外显子序列,获得其全长 CDS。使用 DNAMAN 软件对 RT-PCR 克隆并测序获得

的 *SbSSV* 序列与基因组同源序列搜索鉴定的 *SbSSV* 进行比对,验证 *SbSSV* 序列拼接的正确性。

通过 Gene Structure Display Server (<http://gsds.cbi.pku.edu.cn/>)<sup>[7]</sup> 在线分析工具对鉴定出的 *SbSSV* 进行基因结构分析;利用 TargetP (<http://www.cbs.dtu.dk/services/TargetP/>) 进行信号肽预测;SMART (<http://smart.embl-heidelberg.de/>)<sup>[8]</sup> 和 CD-search service (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>)<sup>[9]</sup> 用于分析蛋白序列的保守结构域。收集单子叶植物高粱、玉米和水稻,双子叶植物拟南芥的所有淀粉合成酶的基因和蛋白序列,并利用 MEGA5.0<sup>[10]</sup> 构建系统发育树,分别采用 neighbor-joining (NJ) 法和似然法建树。参数设置如下:距离模型, p-distance; bootstraps, 重复 1 000 次;替代模型, Jones-Taylor-Thornton (JTT) model;空位缺失数据的处理,两两间删除 (pairwise deletion)<sup>[11]</sup>。

## 2 结果与分析

### 2.1 *SbSSV* 的鉴定

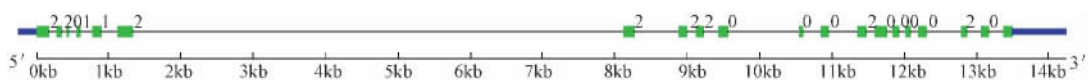
以 *ZmSSV* 的 CDS 序列为查询序列,应用 BLASTn 程序搜索高粱基因组序列,发现在 4 号染色体的 66 376 126 - 66 389 621 区域内存在多个高度相似的同源片段。因此下载该区段的基因组序列,约 13.5 kb,并与 *ZmSSV* 的 CDS 进行精细比对,结合内含子的 GT (C) - AG 剪切规则,识别基因外显子和内含子,拼接所有外显子序列,获得 *SbSSV* 的全长 CDS。分析结果表明,*SbSSV* 包含 20 个外显子和 19 个内含子,全长 CDS 为 2 097 bp,基因结构如图 1 所示,各外显子和内含子的长度见表 1。将新注释的 *SbSSV* 基因结构与高粱基因组在该区段注释的 2 个基因 (*Sb04g036580* 和 *Sb04g036590*) 结构比较(表 2),发现基因组注释的 *Sb04g036580* 的外显子 1、3 和 4 分别对应 *SbSSV* 的外显子 1、5 和 6, *Sb04g036590* 的外显子 2~4、5~8 分别对应 *SbSSV* 的外显子 13~15 和 17~20。 *SbSSV* 的外

显子 2~4 长度较短,且内含子 3 的剪切碱基为 GC-AG,因此在基因组注释的 *Sb04g036580* 基因中未能正确识别。同样, *SbSSV* 的外显子 16,及两侧的内含子 15 和 16 均较短,在 *Sb04g036590* 中也未能正确识别。根据高粱参考基因组序列获得的 *SbSSV* 结构显示,其内含子 6 长度为 6 790 bp,为长内含子,远大于高粱基因内含子的平均长度。该内含子包含了注释基因 *Sb04g036580* 的长外显子 5 (2 719 bp) 和 2 个注释基因的部分间隔区域。高粱 *SbSSV* 的外显子 7~11 位于 2 个注释基因的间隔区,外显子 12 的部分区域与 *Sb04g036590* 的外显子 1 一致。

为揭示 SSV 基因的结构特征,比较了高粱、玉米 (*Gmzm2g130043*) 和水稻 (水稻基因组注释的 *Os02g0807100* 基因结构不完整,因此利用数据库中的全长 CDS 序列 *EU621837*, 比对 *OS02G0807100* 区段的基因组序列,获得 SSV 完整基因结构的外显子长度) SSV 的外显子、内含子数量和长度。由表 1 可知,3 种植物都含有 20 个外显子,除外显子 1 和 4 外,其他外显子的长度完全一致。从这 3 个基因的内含子长度可知,部分内含子长度相近,如 intron1、3、5、8~11、13~16 等,部分内含子长度差异较大。高粱 intron 6 和玉米 intron4、17 均为巨大内含子。这些内含子的序列和长度差异都是在 SSV 基因随各物种分化后,在各物种内部积累突变形成的。由此可知,玉米、高粱和水稻 3 种作物的直系同源基因 SSV 来源于同一祖先基因,且在进化中基因结构保守。

### 2.2 RT-PCR 克隆 *SbSSV* 全长 CDS

为验证通过基因组同源序列搜索鉴定的高粱 *SbSSV* 的正确性,同时检测该基因是否能转录,根据高粱参考基因组序列,在该基因起始和终止密码子附近区域设计引物,以四川高粱农家种宜宾矮花陀为材料,提取幼苗叶片 RNA,扩增获得 *SbSSV* 的全长 CDS (图 2),大小约 2.1 kb。测序结果 (GenBank accession: KP192926) 显示,来源于宜宾矮花陀的 SSV 与根据基因组序列鉴定推导的编码区序列长度一致,仅有 6 个



注:绿框表示外显子,黑色线条表示内含子,蓝色表示 5'-UTR 和 3'-UTR。数字(0,1,2)表示内含子相位。

Note: Green boxes indicated exons, black lines indicated introns, blue lines represented 5'-UTR and 3'-UTR.

The numbers of 0, 1 and 2 showed the introns phase.

图 1 *SbSSV* 基因结构

Fig. 1 Structure organization of *SbSSV* gene

表 1 SSV 基因外显子和内含子的长度

Table 1 Lengths of exons and introns in the SSV genes

外显子 Exon	外显子长度 Exon length/bp			内含子 Intron	内含子长度 Intron length/bp		
	高粱 Sorghum	玉米 Maize	水稻 Rice		高粱 Sorghum	玉米 Maize	水稻 Rice
1	170	170	164	1	118	141	138
2	57	57	57	2	79	95	99
3	28	28	28	3	99	101	110
4	58	67	70	4	172	7349	181
5	114	114	114	5	220	220	204
6	223	223	223	6	6 790	874	301
7	135	135	135	7	628	608	328
8	105	105	105	8	132	130	226
9	96	96	96	9	202	200	385
10	130	130	130	10	993	1253	416
11	51	51	51	11	248	254	554
12	111	111	111	12	403	1 882	618
13	110	110	110	13	123	121	113
14	166	166	166	14	87	87	84
15	81	81	81	15	92	92	86
16	69	69	69	16	105	106	123
17	114	114	114	17	495	54 336	486
18	62	62	62	18	199	210	133
19	97	97	97	19	214	213	161
20	120	120	120				

注: 第 1 个外显子和最后 1 个外显子长度仅表示了编码区部分。

Note: The first and last exons are represented as the encoding sequence.

表 2 新鉴定 *SbSSV* 与基因组注释的 *Sb04g036580* 和 *Sb04g036590* 在染色体上的对应位置Table 2 The corresponding chromosome location of *SbSSV*, *Sb04g036580* and *Sb04g036590*

高粱淀粉合成酶 <i>V</i> <i>SbSSV</i>	位置 Chromosome location	长度 Length/bp	基因组注释基因 Annotation genes	位置 Chromosome location	长度 Length/bp
exon1	66 376 126 – 66 376 295	170	<i>Sb04g036580</i> exon1	66 376 126 – 66 376 295	170
intron1	66 376 296 – 66 376 413	118	intron1	66 376 296 – 66 376 413	118
exon2	66 376 414 – 66 376 470	57	exon2	66 376 414 – 66 376 436	23
intron2	66 376 471 – 66 376 549	79	intron2	66 376 437 – 66 376 906	470
exon3	66 376 550 – 66 376 577	28			
intron3	66 376 578 – 66 376 676	99			
exon4	66 376 677 – 66 376 734	58			
intron4	66 376 735 – 66 376 906	172			
exon5	66 376 907 – 66 377 020	114	exon3	66 376 907 – 66 377 020	114
intron5	66 377 021 – 66 377 240	220	intron3	66 377 021 – 66 377 240	220

表 2(续)

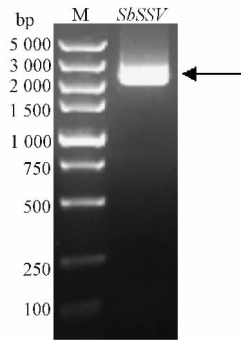
高粱淀粉合成酶 V <i>SbSSV</i>	位置 Chromosome location	长度 Length/bp	基因组注释基因 Annotation genes	位置 Chromosome location	长度 Length/bp
exon6	66 377 241 – 66 377 463	223		exon4	66 377 241 – 66 377 463
intron6	66 377 464 – 66 384 253	6790		intron4	66 377 464 – 66 377 634
				exon5	66 377 635 – 66 380 353
exon7	66 384 254 – 66 384 388	135	基因间隔区		66 380 354 – 66 387 000
intron7	66 384 389 – 66 385 016	628			
exon8	66 385 017 – 66 385 121	105			
intron8	66 385 122 – 66 385 253	132			
exon9	66 385 254 – 66 385 349	96			
intron9	66 385 350 – 66 385 551	202			
exon10	66 385 552 – 66 385 681	130			
intron10	66 385 682 – 66 386 674	993			
exon11	66 386 675 – 66 386 725	51			
intron11	66 386 726 – 66 386 963	248			
exon12	66 386 974 – 66 387 084	111	Sb04g036590	exon1	66 387 001 – 66 387 084
intron12	66 387 085 – 66 387 487	403		intron1	66 387 085 – 66 387 487
exon13	66 387 488 – 66 387 597	110		exon2	66 387 488 – 66 387 597
intron13	66 387 598 – 66 387 720	123		intron 2	66 387 598 – 66 387 720
exon14	66 387 721 – 66 387 886	166		exon 3	66 387 721 – 66 387 886
intron14	66 387 887 – 66 387 973	87		intron 3	66 387 887 – 66 387 973
exon15	66 387 974 – 66 388 054	81		exon 4	66 387 974 – 66 388 054
intron15	66 388 055 – 66 388 146	92		intron 4	66 388 055 – 66 388 320
exon16	66 388 147 – 66 388 215	69			
intron16	66 388 216 – 66 388 320	105			
exon17	66 388 321 – 66 388 434	114		exon 5	66 388 321 – 66 388 434
intron17	66 388 435 – 66 388 929	495		intron 5	66 388 435 – 66 388 929
exon18	66 388 930 – 66 388 991	62		exon 6	66 388 930 – 66 388 991
intron18	66 388 992 – 66 389 190	199		intron 6	66 388 992 – 66 389 190
exon19	66 389 191 – 66 389 287	97		exon 7	66 389 191 – 66 389 287
intron19	66 389 288 – 66 389 501	214		intron 7	66 389 288 – 66 389 501
exon20	66 389 502 – 66 389 621	120		exon 8	66 389 502 – 66 389 621

碱基差异,其中 2 个为同义突变,4 个为非同义突变。结果表明,该基因编码区序列在不同材料中也相对保守,同时也表明,通过基因组序列搜索进行同源新基因鉴定的方法是准确可靠的。

### 2.3 *SbSSV* 的序列特征分析

*SbSSV* 编码的蛋白质长度为 698 个氨基酸,理论分子量为 78.42 kD,等电点为 6.28。TargetP 预测显示,该蛋白定位于叶绿体,叶绿体信号肽的剪切位点位于 49 位和 50 位氨基酸之间。用 SMART 和 CD-search service 分析蛋白质的保守结构域,在 N 端发现 1 个

coiled coil 结构域,位于 109 ~ 144 位氨基酸,C 端发现 1 个 GT1 \_ Glycogen \_ synthase \_ DULL1 \_ like 和 glgA (glycogen/starch synthase, ADP-glucose type) 结构域,位于 221 ~ 681 位氨基酸(图 3)。目前已知的淀粉合成酶亚型 SSIV 的 N 端也存在 coiled coil 结构域。C 端的 2 个结构域是细菌糖原合成酶和植物淀粉合成酶,两类均以 ADP 葡萄糖为底物的催化酶所共有的保守区域。从这两类酶形成的空间结构可知,该区域又可分为糖基催化(starch catalytic domain, GT-5)和糖基转移(glycosyltransferase 1 domain, GT-1) 2 个相对独



注: M 为 DNA 分子量标准( DL 5 000), 箭头标注的是目的片段。  
Note: M indicated the DNA marker of DL 5 000. Arrow represented the amplified fragment.

图2 *SbSSV* 全长 CDS 的 RT-PCR 扩增

Fig.2 RT-PCR amplification product of *SbSSV*

立的单元(或亚结构域),每个单元由不同数量的 $\alpha$ - $\beta$ - $\alpha$ 二级结构原件折叠形成<sup>[12]</sup>。ADP 葡萄糖和多糖底物结合于 GT-5 和 GT-1 形成的夹缝内。因此,根据蛋白保守结构域的分析结果也可推测,新鉴定的 *SbSSV* 为植物淀粉合成酶家族成员。

#### 2.4 *SbSSV* 与其他淀粉合成酶的系统发生关系分析

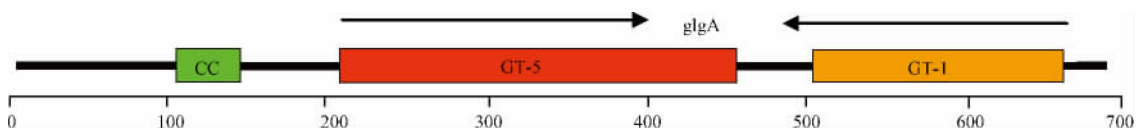
为揭示高粱淀粉合成酶 V 与高粱及其他植物的不同淀粉合成酶亚型的系统发生关系,本研究收集了单子叶植物高粱、玉米和水稻,及双子叶植物拟南芥的所有淀粉合成酶的基因和蛋白序列,并采用 NJ 法和似然法构建系统发育树,2 种方法构建的系统树拓扑结构一致(图 4)。由编码这些酶基因的进化起源可知,淀粉合成酶可分为 2 个大类群,分别命名为 A 和 B。A 类群包含 GBSS、SSI 和 SSII, B 类群包含 SSIII、SSIV 和 SSV。从系统发育树的分支关系可知, A 类群里, 2 次基因复制产生了 GBSS、SSI 和 SSII 3 个分支,第 1 次复制产生了 GBSS 和 SSI/SSII 的共同祖先,第 2 次复制产生了 SSI 和 SSII 2 个分支。类似的, B 类群的 3

个分支也来源于 2 次基因复制,第 1 次复制产生了 SSIII 和 SSIV/SSV 的共同祖先,第 2 次复制产生了 SSIV 和 SSV 2 个分支。由此可知,新鉴定的 SSV 亚型与已知的 SSIV 亚型亲缘关系最近。经多序列比对发现高粱、玉米和水稻 3 种作物 SSIV、SSV 的蛋白序列在 N 端的 coiled coil 结构域和 C 端的 glgA 结构域内,氨基酸序列一致性高于其它区域。

#### 2.5 *SbSSV* 的表达特性分析

qRT-PCR 检测 *SbSSV* 在不同组织中的表达(图 5-A),结果显示 SSV 在根、茎、叶 3 种组织中均有表达,叶片中的表达量最高,根中最低。在正常光照条件下(16 h 光照,8 h 黑暗),该基因在叶片中的表达具有昼夜节律特性(图 5-B),随着光照时间延长,表达量增高;光照结束后表达量迅速降低。为进一步验证 *SbSSV* 的表达受光诱导,对材料分别进行延长光照时间和黑暗时间的处理(图 5-C),结果显示,随着光照时间的进一步增长,基因表达量缓慢增加,延长光照时间至 24 h(累积光照 40 h)时表达量达到最高,以后几乎保持恒定。黑暗处理 8 h 后,尽管表达量较低,但仍呈现缓慢降低的趋势。综上,*SbSSV* 受光诱导表达,在正常光照培养下,其表达具有昼夜节律特性。

为进一步揭示高粱 *SbSSV* 受光诱导表达的分子机制,根据高粱参考基因组提供的序列信息,取基因转录起始上游的 1 500 bp 作为启动子区域,应用 PlantCARE (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) 在线分析工具预测其启动子中可能存在的顺式作用元件。结果发现,*SbSSV* 启动子区不仅含有真核生物启动子结构普遍具有的保守序列 TATA-box、CAAT-box,还具有脱落酸响应元件(ABRE),光响应元件(G-box、TCCC-motif、GATA-motif),发育相关元件(CCGTCC-box),真菌诱导响应元件(Box-W1)和其他功能调控元件。由此推断 *SbSSV* 基因可能参与叶片淀粉合成,其转录与其他淀

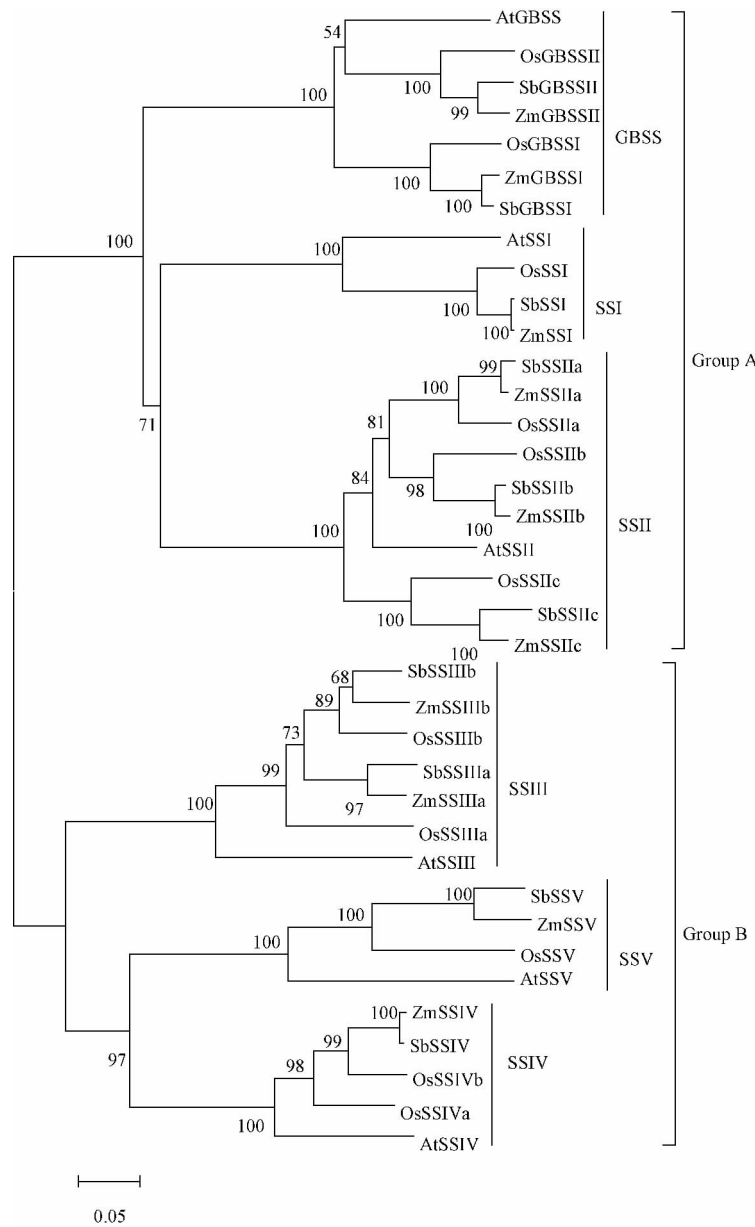


注: 卷曲螺旋结构域(Coiled coil, CC), 淀粉糖基催化结构域(GT-5) 和糖基转移结构域(GT-1) 分别以绿色、橘色和黄色表示, glgA 结构域包含了 GT-5 和 GT-1 结构域。

Note: The green, orange and yellow rectangles represent the coiled coil (CC) domain, starch catalytic domain (GT-5) and glycosyltransferase 1 domain (GT-1), respectively.

图3 *SbSSV* 蛋白的保守结构域

Fig.3 The conserved domains of *SbSSV*



注: 各个物种名称用属名和种名的第 1 个字母表示。Zm: 玉米; Sb: 高粱; Os: 水稻; At: 拟南芥。各节点处数值表示 bootstrap 值(重复 1 000 次)。

Note: Each node was labeled with the prefix of the initials of the genus and species. Zm: Zea mays, Sb: Sorghum bicolor, Os: Oryza sativa, At: Arabidopsis thaliana. The number of each node indicates the bootstrap values of 1 000 replicates.

图 4 高粱、玉米、水稻和拟南芥中的淀粉合成酶家族的系统进化关系

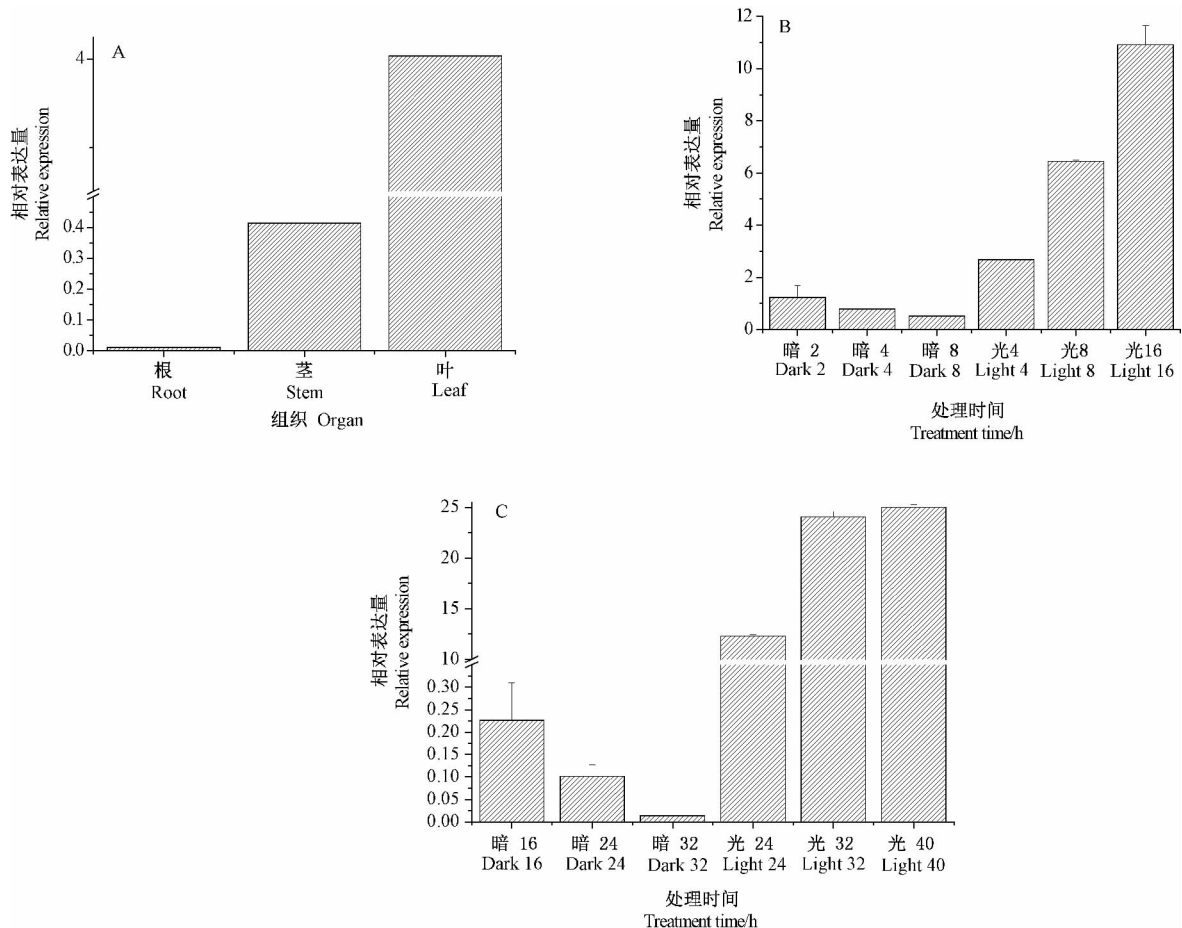
Fig. 4 Phylogenetic relationship of SSs in Zea mays, Sorghum bicolor, Oryza sativa and Arabidopsis thaliana

粉合成酶一样,受光、激素等多种信号调控。

### 3 讨论

淀粉是植物光合作用的最终产物,主要贮藏在禾本科作物的种子里,是人类食品能量的主要来源。近年来的大量研究已明确了水稻、玉米、拟南芥等模式植

物淀粉合成代谢的基本过程,分离了参与淀粉合成代谢的酶,也揭示了其编码基因的表达特征及功能<sup>[12-15]</sup>。为进一步阐明和完善植物淀粉合成代谢的分子机制,本研究利用玉米同源基因搜索比对高粱基因组序列,首次鉴定并克隆分离了编码高粱淀粉合成酶的新亚型基因 *SbSSV*,该基因结构保守,与玉米和水稻的直系同源基因具有相同的外显子和内含子数量,

图5 *SbSSV* 的表达特性分析Fig. 5 Expression patterns of *SbSSV*

且在 20 个外显子中,与 *ZmSSV* 有 19 个外显子长度一致,与 *OsSSV* 有 18 个外显子长度一致。RT-PCR 克隆测序结果与通过基因组鉴定的 *SbSSV* 基因的 CDS 序列基本一致,表明同源序列搜索基因组鉴定的新基因是准确的。

根据该基因编码区序列推导的蛋白质序列具有细菌糖原合成酶和植物淀粉合成酶共有的 *glgA* 结构域,该区域含糖基催化 (GT-5) 和糖基转移 (GT-1) 2 个亚结构域。GT-5 仅以 ADP-glucose 为底物,将其葡萄糖基转移到淀粉链或糖原链上,此结构域存在于细菌的糖原合成酶和植物的淀粉合成酶中。GT-1 结构域能够转移 UDP、ADP、GDP 或者 CMP-glucose 中的糖基到不同的底物上,这些底物包括糖原、果糖六磷酸等,因此 GT-1 结构域作用的单糖底物类型较多,它不是细菌糖原合成酶和植物淀粉合成酶独有的结构域<sup>[16]</sup>。系统发育分析显示,SS 包含 2 类不同进化起源的基因,分为 6 个亚家族。除 SSV 外,图 4 中其他淀粉合成酶亚型的系统发生关系与以前的报道一

致<sup>[13]</sup>。新鉴定的高粱及其他几种植物的 SSV 与已知的 SSIV 亲缘关系最近,来源于祖先基因的复制。因此,综合蛋白结构域和进化分析结果,推测本研究新鉴定的 *SbSSV* 为植物淀粉合成酶家族的新亚型。

定量分析表明,*SbSSV* 具有组织表达特异性,主要在叶片中表达,且其表达受光诱导,具有昼夜节律特性。这些表达特性与其他淀粉合成酶基因如 *GBSSII*<sup>[26]</sup>、*SSI*<sup>[1,17]</sup>、*SSIb*<sup>[27]</sup>、*SSIV*<sup>[5]</sup> 相似。叶片是合成淀粉的重要组织,在光照条件下,光合途径中基因表达,产生多种蛋白酶,经过复杂的代谢形成蔗糖,随后,在淀粉合成代谢途径中多种酶的催化下,将蔗糖转化为由众多葡萄糖残基构成的淀粉来储备能量和 C 源,供给植物体在夜间的生长发育代谢<sup>[19-21]</sup>。因此许多淀粉合成代谢途径中的基因,其表达都具有光诱导的昼夜节律特性。进一步分析 *SbSSV* 的启动子,发现该基因启动子内也具有与光诱导相关的保守基序。高等植物基因受光诱导表达的机制可能是利用光敏色素等受体,对生长环境中光强、光周期等条件的感应来调节



自身生长发育<sup>[22]</sup>。与光敏色素作用的转录因子能够与受光调控基因的启动子区的特定基序结合<sup>[23-24]</sup>,从而实现光照调控基因的表达。因此推测 SbSSV 与其他淀粉合成酶一样也参与了淀粉合成,且可能在叶片淀粉合成中扮演重要角色。已报道的植物淀粉合成酶有 5 种亚型,其在淀粉合成中的功能也基本明确<sup>[28]</sup>。本研究新鉴定的淀粉合成酶新亚型 SSV 具有保守的基因结构、核酸、蛋白序列及蛋白结构域等,且在高粱、玉米和水稻 3 种作物中均能正确转录,据此推测,高等植物中的淀粉合成酶亚型至少有 6 种,而非以往报道的 5 种,且 SSV 亚型在淀粉合成代谢中可能行使着重要功能,如负责延伸特定类型的糖链,具有底物特异性,或者与其他淀粉合成酶形成复合体,对淀粉合成起调节作用等,但其确切功能目前尚未知,还需要进一步研究探讨。

目前,从低等的微生物到人类已有数百种生物的全基因组测序工作已完成,基因组注释 (genome annotation) 已成为后基因组时代的主要研究内容<sup>[18, 25]</sup>。但由于基因预测算法本身不够完善,可利用的参考基因或蛋白序列有限,以及生物转录、翻译等的复杂性等因素,导致目前基因组内部分基因未能注释,或出现错误注释,如本研究中发现的高粱 *Sb04g036580* 和 *Sb04g036590*, 其外显子、内含子识别不准确,尤其是对于一些长度较短或较长的外显子和内含子,其预测准确性更低。因此基因组注释仅依靠现有的算法和软件验证难以实现准确全面的阅读框鉴定,应与常规的分子生物学、生物化学等实验方法相结合,进行准确高效的基因鉴定,完善基因组注释信息。

## 4 结论

本研究首次鉴定并分离了编码高粱淀粉合成酶的新亚型基因 *SbSSV*, 并对其序列特征和组织表达特性进行了分析。结果显示该基因外显子-内含子结构保守,且推导的蛋白序列也很保守,具有细菌糖原合成酶和植物淀粉合成酶特有的糖基催化和糖基转移结构域。系统发育分析表明,高等植物的 SSV 形成独立分支,与已知的 SSIV 亲缘关系最近。*SbSSV* 的组织表达特性与其他淀粉合成酶类似,主要在高粱叶片中表达,其表达受光诱导,具有昼夜节律特性,但确切功能还有待深入研究。

## 参考文献:

[1] Zeeman S, Kossmann J, Smith A. Starch: its metabolism,

evolution, and biotechnological modification in plants [J]. Annual Review of Plant Biology, 2010, 61(1): 209-234

- [2] Jeon J, Ryoo N, Hahn T, Walia H, Nakamura Y. Starch biosynthesis in cereal endosperm [J]. Plant Physiology and Biochemistry, 2010, 48(6): 383-392
- [3] Yan H, Pan XX, Jiang H, Wu G. Comparison of the starch synthesis genes between maize and rice: copies, chromosome location and expression divergence [J]. Theoretical and Applied Genetics, 2009, 119(5): 815-825
- [4] Zhang X, Szydlowski N, Delvallé D, D'Hulst C, James M, Myers A. Overlapping functions of the starch synthases SSII and SSIII in amylopectin biosynthesis in Arabidopsis [J]. BMC Plant Biology, 2008, 8(1): 96
- [5] Crumpton - Taylor M, Pike M, Lu K J, Hylton C, Feil R, Eicke S, Lunn J, Zeeman S, Smith A. Starch synthase 4 is essential for coordination of starch granule formation with chloroplast division during Arabidopsis leaf expansion [J]. New Phytologist, 2013, 200(4): 1064-1075
- [6] Livak K, Schmittgen T. Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta CT}$  method [J]. Methods, 2001, 25(4): 402-408
- [7] Guo A, Zhu Q, Chen X, Luo J. GSDS: a gene structure display server [J]. Hereditas, 2007, 29(8): 1023-1026
- [8] Letunic I, Copley R, Pils B, Pinkert S, Schultz J, Bork P. SMART 5: domains in the context of genomes and networks [J]. Nucleic Acids Research, 2006, 34(suppl 1): D257-D260
- [9] 关鹰, 许在恩, 郭小勤. 毛竹 IDD 基因家族的生物信息学分析 [J]. 核农学报, 2014, 28(6): 998-1005
- [10] Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods [J]. Molecular Biology and Evolution, 2011, 28(10): 2731-2739
- [11] 陈勋基, 李建平, 郝晓燕, 足目热木, 陈果, 黄全生. 玉米 ZmCIPK21 基因的克隆与分析 [J]. 核农学报, 2012, 26(6): 862-867
- [12] Leterrier M, Holappa L, Broglie K, Beckles D. Cloning, characterisation and comparative analysis of a starch synthase IV gene in wheat: functional and evolutionary implications [J]. BMC Plant Biology, 2008, 8(1): 98
- [13] Li C, Li Q, Dunwell J, Zhang Y. Divergent evolutionary pattern of starch biosynthetic pathway genes in grasses and dicots [J]. Molecular Biology and Evolution, 2012, 29(10): 3227-3236
- [14] 张海艳, 董树亭, 高荣岐, 李玉全. 玉米籽粒淀粉积累及相关酶活性分析 [J]. 中国农业科学, 2008, 41(7): 2174-2181
- [15] Comparot-Moss S, Denyer K. The evolution of the starch biosynthetic pathway in cereals and other grasses [J]. Journal of Experimental Botany, 2009, 60(9): 2481-2492
- [16] Furukawa K, Tagaya M, Tanizawa K, Fukui T. Role of the conserved Lys-X-Gly-Gly sequence at the ADP-glucose-binding site in Escherichia coli glycogen synthase [J]. Journal of Biological Chemistry, 1993, 268(32): 23837-23842

- [17] James M G, Denyer K, Myers A M. Starch synthesis in the cereal endosperm [J]. *Current Opinion in Plant Biology*, 2003, 6(3): 215 - 222
- [18] Bailey L, Searls D, Overton G C. Analysis of EST-driven gene annotation in human genomic sequence [J]. *Genome Research*, 1998, 8(4): 362 - 376
- [19] Streb S, Zeeman S C. Starch metabolism in Arabidopsis [J]. *Arabidopsis Book*, 2012, 10: e0160
- [20] Hirose T, Terao T. A comprehensive expression analysis of the starch synthase gene family in rice (*Oryza sativa* L.) [J]. *Planta*, 2004, 220(1): 9 - 16
- [21] 孟祥春, 彭建宗, 王小菁. 光和糖对非洲菊花色素苷积累及CHS, DFR基因表达的影响 [J]. *园艺学报*, 2007, 34(1): 227 - 230
- [22] Ral J, Colleoni C, Wattedled F, Dauvillée D, Nempont C, Deschamps P, Li Z, Morell M, Chibbar R, Purton S. Circadian clock regulation of starch metabolism establishes GBSSI as a major contributor to amylopectin synthesis in *Chlamydomonas reinhardtii* [J]. *Plant Physiology*, 2006, 142(1): 305 - 317
- [23] Kaiser T, Emmeler K, Kretsch T, Weisshaar B, Schafer E, Batchauer A. Promoter elements of the mustard CHS1 gene are sufficient for light regulation in transgenic plants [J]. *Plant Molecular Biology*, 1995, 28(2): 219 - 229
- [24] 黄海群, 林拥军. 水稻 rbcS 基因启动子的克隆及结构功能分析 [J]. *农业生物技术学报*, 2007, 15(3): 451 - 458
- [25] 吴飞珍, 马文丽, 王旺迪, 陈启龙, 郑文岭. 一种新的基因注释语义相似度计算方法 [J]. *生物信息学*, 2010, 8(1): 23 - 29
- [26] Momma M, Fujimoto Z. Interdomain disulfide bridge in the rice granule bound starch synthase I catalytic domain as elucidated by X-ray structure analysis [J]. *Bioscience, Biotechnology, and Biochemistry*, 2012, 76(8): 1591 - 1595
- [27] Ball S G, Morell M K. From bacterial glycogen to starch: understanding the biogenesis of the plant starch granule [J]. *Annual Review of Plant Biology*, 2003, 54(1): 207 - 233
- [28] Edwards A, Fulton D C, Hylton C M, Jobling S A, Gidley M, Rossner U, Martin C, Smith A M. A combined reduction in activity of starch synthases II and III of potato has novel effects on the starch of tubers [J]. *The Plant Journal*, 1999, 17(3): 251 - 261

## Identification and Characterization of a Novel Gene Encoding Starch Synthase V in *Sorghum bicolor*

YU Guiling<sup>1</sup> XIAO Qianlin<sup>2</sup> WEI Bin<sup>2</sup> YANG Jin<sup>1</sup> WANG Dan<sup>3</sup> LIU Hanmei<sup>1</sup>

<sup>1</sup>College of Life Sciences, Sichuan Agricultural University, Ya'an, Sichuan 625014;

<sup>2</sup>College of Agronomy, Sichuan Agricultural University, Chengdu, Sichuan 611130;

<sup>3</sup>Puyang Academy of Agricultural Sciences, Puyang, Henan 457000)

**Abstract:** Starch synthases (SS) are key enzymes involved in starch biosynthesis pathways. Five isoforms of starch synthase have been identified and characterized functionally. To identify genes encoding new SS isoforms based on rich sorghum genome data, *SbSSV*, a new isoform of sorghum SSs have been identified and isolated by employing bioinformatics and molecular biology methods. The exon-intron structure and expression patterns of *SbSSV* have been analyzed. Results show that the ORF of *SbSSV* is 2 097 bp, with its exon numbers, lengths and intron distributions were consistent with those of its homologues in maize and rice. The predicted protein has conserved starch catalytic domain and glycosyltransferase domain, which are unique characteristics of glycogen synthases in bacteria and starch synthases in plants. A phylogenetic analysis of SSs in higher plants shows that SSV is closely related to the known SSIV. Therefore, we infer that the *SbSSV* encodes a new isoform of SSs in sorghum. qRT-PCR analysis indicates that *SbSSV* is mainly expressed in leaf with its expression induced by light and follows circadian rhythm. The results provide a theoretical basis for further revealing the mechanism of starch biosynthesis metabolism and improving yield and quality in crops.

**Keywords:** sorghum bicolor, starch, starch synthase V, identification of a novel gene