

## 白菜型油菜 *BraSDG8* 基因的克隆与序列分析

谢青轩<sup>1,2</sup>, 彭琦<sup>2</sup>, 刘春林<sup>2</sup>, 阮颖<sup>1,2\*</sup>

(1.湖南农业大学 生物科学技术学院, 湖南 长沙 410128; 2.作物种质创新和资源利用国家重点实验室培育基地, 湖南 长沙 410128)

**摘要:**通过序列的同源性分析,在白菜型油菜中克隆得到拟南芥中的 1 个花期调控基因 *SDG8* 的全长 cDNA,命名为 *BraSDG8*。*BraSDG8* 全长 4 963 bp,编码 1 654 个氨基酸,其中第 867 位至第 1 060 位氨基酸连续构成 AWS domain、SET domain、Post SET domain 3 个结构域,与拟南芥中的 *SDG8* 包含的功能活性位点相同,生物信息学分析表明,*BraSDG8* 与 *SDG8* 同属于 ASH1 家族。

**关键词:**白菜型油菜; *BraSDG8* 基因; 组蛋白甲基转移酶; 序列分析

中图分类号: Q785 文献标志码: A 文章编号: 1007-1032(2011)04-0372-04

### Cloning and sequence analysis of *BraSDG8* gene in *Brassica rapa*

XIE Qing-xuan<sup>1,2</sup>, PENG Qi<sup>2</sup>, LIU Chun-lin<sup>2</sup>, RUAN Ying<sup>1,2\*</sup>

(1. College of Bioscience and Biotechnology, Hunan Agricultural University, Changsha 410128, China; 2. Pre-State Key Laboratory for Germplasm Innovation and Resource Utilization of Crop, Changsha 410128, China)

**Abstract:** *SDG8* in *Arabidopsis* (*AraSDG8*) was confirmed to be a gene involved in the regulation of flowering time, but the role of *SDG8* in *Brassica rapa* (*BraSDG8*) was unknown. In this work, full-length cDNA of *BraSDG8* from *Brassica rapa* was cloned and analyzed. The *BraSDG8* was 4 963 bp long encoding a polypeptide of 1 654 amino acids and three active domains which is similar to *AraSDG8* (AWS domain, SET domain and Post SET domain) were observed in this peptide from amino acid residue 867 to residue 1 060. High similarity of gene sequence and peptide structure between *BraSDG8* and *AraSDG8* indicated that *BraSDG8* may play an important role in regulation of flowering time in *Brassica rapa*.

**Key words:** *Brassica rapa*; *BraSDG8* gene; histone methyltransferase; sequence analysis

现有油菜品种普遍生育期较长,成熟期偏迟,在稻-稻-油三熟制中,油菜收割时间与水稻收割时间相冲突,造成稻田特别是双季稻田冬季闲置;因此,选育早熟油菜品种是解决双季稻田种植油菜季节突出矛盾的根本途径。

组蛋白的赖氨酸甲基化修饰对植物的开花有调控作用<sup>[1-2]</sup>。拟南芥 *SDG8* 基因编码的组蛋白甲基转移酶是 SET2 的同源物,能在体外对寡核苷酸酶

甲基化,从而调控拟南芥的开花时间<sup>[3]</sup>。油菜与拟南芥同属十字花科植物。*SDG8* 基因在拟南芥中的开花调控功能已得到确认,但是油菜中关于 *SDG8* 基因的研究报道还非常有限,GenBank 中尚无登录油菜 *SDG8* 基因的 cDNA 序列。笔者根据拟南芥 *SDG8* 基因与白菜基因组的同源序列设计引物,通过 RT-PCR 技术,克隆出白菜型油菜中的 *SDG8* 基因(*BraSDG8*)编码区 cDNA 序列,并对其序列的结

收稿日期: 2011-03-24

基金项目: 国家自然科学基金项目(31071129)

作者简介: 谢青轩(1986—),女,湖南长沙人,硕士研究生,主要从事植物分子生物学研究, zoe19227@sina.com; \*通信作者, yingruan@hotmail.com

构进行初步的生物信息学分析, 以期为研究 *BraSDG8* 基因在油菜开花中的调控功能和选育油菜早熟品种奠定基础。

## 1 材料与方法

### 1.1 材料

白菜型油菜(*Brassica rapa*)B3 和大肠杆菌 DH5 $\alpha$  均由湖南农业大学植物资源利用实验室提供; pMD19-T 载体购自 TaKaRa 公司; RNA 反转录试剂盒购自 Ferments 公司; DNA 凝胶回收 Kit 购自安比奥公司; LongAmp Taq DNA Polymerase 购自 BioLabs 公司; dNTPs、DNA marker 购自天根生物公司。

### 1.2 方法

以白菜型油菜 B3 抽薹后开花前的叶片为材料, 采用 TRIZOL 法提取总 RNA。参照 Ferments 公司 RNA 反转录试剂盒说明, 进行 cDNA 第一链的合成。

下载 GeneBank 中拟南芥的 *SDG8* 基因全长 mRNA 序列(DQ340869)5 280 bp, 于白菜基因组数据库 BRAD(<http://brassicadb.org/brad/blastPage.php>) 上进行 BLAST 比对, 获得 1 段同源性在 74% 以上、且长度为 7 223 bp 的白菜基因组序列。利用 GENSCAN 软件(<http://genes.mit.edu/GENSCAN.html>) 分析这段序列的 CDS, 经电子拼接发现白菜 *SDG8* 基因的全长 mRNA 序列为 5 010 bp。根据此同源序列 5' 端设计正向引物 ATGGATTGCAAGGAGAACGGTGT TA; 靠近 3' 端处设计反向引物 TTGATTGTTGATC TTCTTCTCTACAGTT。以白菜型油菜 B3 的 cDNA 为模板, 在 PCR 管中分别加入 LongAmp Taq Reaction Buffer 5  $\mu$ L、dNTP 0.75  $\mu$ L、LongAmp Taq DNA Polymerase 1  $\mu$ L、引物各 1  $\mu$ L、超纯水补齐至 25  $\mu$ L。反应程序: 95  $^{\circ}$ C 预变性 30 s, 95  $^{\circ}$ C 变性 10 s, 65  $^{\circ}$ C 退火 30 s, 65  $^{\circ}$ C 延伸 4.5 min, 65  $^{\circ}$ C 后延伸 10 min, 循环 30 次。将 PCR 产物在 1.2% 的琼脂糖凝胶上电泳分离回收后, 与 pMD19-T 载体连接成 pMD19-T+ *BraSDG8*, 热激法将此连接质粒转化到

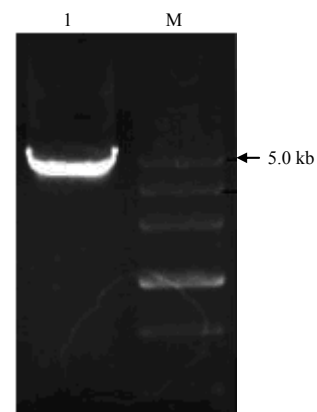
大肠杆菌 DH5 $\alpha$  感受态细胞中, 重组子经鉴定后送上海英骏生物技术有限公司测序。

测序结果用 DNA Star 软件进行拼接, 得到完整的 *BraSDG8* 的全长 cDNA。将此序列首先在 The *Brassica* database (BRAD) 进行比对, 确定其正确性, 再将此全长 cDNA 序列在 NCBI 的 Blast 上进行比对分析。运用 SMART (<http://smart.embl-heidelberg.de/>) 在线进行蛋白质结构预测及分析, 并运用 MEGA5.0 将 *BraSDG8* 与 GeneBank 中其他含有 SET 结构域的植物序列进行聚类分析。

## 2 结果与分析

### 2.1 *BraSDG8* 基因的克隆

以白菜型油菜 B3 提取总 RNA, 反转录得到的 cDNA 为模板进行 PCR 扩增, 得到 *BraSDG8* 基因。用 1.2% 的琼脂糖凝胶电泳。电泳检测结果如图 1 所示, 条带大小约为 5.0 kb, 与预期大小相符。测序结果表明, 扩增产物长度为 4 963 bp, 通过 DNA Star 软件比对分析发现, 扩增得到的片段序列与电子克隆产物的同源性达 98.8%, 且与拟南芥 *SDG8* 基因(DQ340869)序列的同源性达 87%。由此表明, 所克隆的片段为白菜型油菜中 *BraSDG8* 基因的 CDS 全长序列。



1 目的片段扩增结果; M 1 kb DNA 分子量标准

图 1 *BraSDG8* 基因电泳结果

Fig.1 Agarose gel electrophoresis patterns of *BraSDG8*

### 2.2 *BraSDG8* 蛋白质的结构分析

将 *BraSDG8* 基因序列在 NCBI 上进行翻译,

发现该基因编码 1 654 个氨基酸, 相对分子质量为 180 859.18, 等电点为 5.68。而拟南芥 *SDG8* 基因 (DQ340869) 编码 1 759 个氨基酸, 相对分子质量为 193 228.43, 等电点为 5.31。进一步将 *BraSDG8* 在 NCBI 上翻译得到的氨基酸序列用 SMART 在线进行蛋白质结构预测及分析, 结果表明, 在其第 867 至 918 位氨基酸构成了 1 个 AWS(associated with SET) 结构域, 第 919 至第 1 042 位氨基酸构成了 1

个 SET 结构域, 第 1 044 至第 1 060 位氨基酸构成了 1 个 Post SET 结构域。而目前已知的拟南芥中的 *SDG8* 的主要功能活性位点也是由 AWS、SET、Post SET 3 个结构域构成的 SET 结构域蛋白。*BraSDG8* 与拟南芥中的 *SDG8* 在 NCBI 上的氨基酸序列同源性达到 99%(图 2), 说明两者在序列结构上有高度的相似性。

<i>BraSDG8</i>	IWQDGIHGEAVMDLTIEDSSPTESAWVRCD DCFKWRRI PASV VES IDESSRWICMNN SDK	792
<i>SDG8</i>	I D HGE V+D+TIEDS TESAWVRCD DCFKWRRI PASV V IDESSRWICMNN SDK	899
<i>BraSDG8</i>	DFAHCSISQEMSN E EINEQLGIGQDEADAYDYEA AKR GK DKEQSKRLPVNKKACFRAIK	852
<i>SDG8</i>	FA CS SQEMSN E EINE+LGIGQDEADAYD +AAKRGK+KEQSKRL +KACF+AIK	959
	<b>AWS domain</b>	
<i>BraSDG8</i>	TNQFLHRNRKNTIDEIMVCHCKP PPDGRLGCGEECLNRMLNIECLHGTC PAGDLCSNQQ	912
<i>SDG8</i>	TNQFLHRNRK+QTIDEIMVCHCKP PDGRLGCGEECLNRMLNIECL GTC PAGDLCSNQQ	1 019
	<b>SET domain</b>	
<i>BraSDG8</i>	FQKRKYVKFERFQSGKKGYGLR LLEDVREGQFLIEYVGEV LDMQSYESRQKDYASMGQKH	972
<i>SDG8</i>	FQKRKYVKFERFQSGKKGYGLR LLEDVREGQFLIEYVGEV LDMQSYE+RQK+YA GQKH	1 079
<i>BraSDG8</i>	FYFMTLNGNEVIDAGAKGNLGRFINHSC EPNCRTEKMMVNGEICVGI FSMKDLKKGQELT	1 032
<i>SDG8</i>	FYFMTLNGNEVIDAGAKGNLGRFINHSC EPNCRTEKMMVNGEICVGI FSM+DLKKGQELT	1 139
	<b>post SET domain</b>	
<i>BraSDG8</i>	FDYNYVRVFGAARAKKCYCGSSHCRGYIGG DPLNGDVVQSDSDEEYPELVILDDDES GEG	1 092
<i>SDG8</i>	FDYNYVRVFGAARAKKCYCGSSHCRGYIGG DPLNGDV++QSDSDEEYPELVILDDDES GEG	1 199
<i>BraSDG8</i>	ILDATSKIFMDGADMQPQNSTKVDDSKDLA---SQSPSSVAVKLP ESEVLPS-FQPTEA	1 148
<i>SDG8</i>	IL ATS+ F D AD QMPQ+ KV+ KDLA +Q+ SSV+VKLPE E+ P QPTE	1 259
<i>BraSDG8</i>	SKELSTDMFVIDVQQA-LEKKTGKPS PASKSISRLSSDGANADKTVKHGSGEDVRILSR	1 207
<i>SDG8</i>	KELS+ + + VQQE EKKTK SP S S+SR+S G N+DKT KHGSGED +IL R	1 319
<i>BraSDG8</i>	PRPRTKTSRSGSSSQ---ALPGVNKAQTTPVKKLQQQPIKSKGSEEVSPSGRIETFEG	1 263
<i>SDG8</i>	PRPR KTSRSS SSK+ PGVNKAQ PV KLOQQPIKSKGSE+VSPS IETFEG	1 377
<i>BraSDG8</i>	KLNELLDAGGGISKRRDSAKGYLKL LLLTAASRGNANEGIQSNRDL SMILDALLKTKSRT	1 323
<i>SDG8</i>	KLNELLDA GGISKRRDSAKGYLKL LLLTAASRG EGI SNRDL SMILDALLKTKS++	1 437
<i>BraSDG8</i>	VLVDVINKNGLQMLHNIMKQYRDFK KPTILRKLKLVLEYLATRD ILALEHIVRPPPYAG	1 383
<i>SDG8</i>	VLVD+INKNGLQMLHNIMKQYR DFK+ PI+RKLKLVLEYLATR ILALEHI+R PP+AG	1 497

图 2 *BraSDG8* 与拟南芥中的 *SDG8* 氨基酸序列比对

Fig.2 Alignment of *BraSDG8* amino acid sequence with *SDG8* from *Arabidopsis thaliana*

### 2.3 聚类分析

植物和动物中存在许多可以编码含有 SET 结构域蛋白质的基因。目前, 通过对这些蛋白中 SET 结构域的系统进化分析, 可以将植物中的含有 SET 结构域的蛋白质划分为 I 至 VII 等 7 个家族, 它们分别是 E(Z) 家族、ASH1 家族、Trit horax 家族、Su (var) 家族、含有 SET 和 PHD 结构域的蛋白质家族及含

有被中断的 SET 结构域的蛋白质家族<sup>[4]</sup>。拟南芥中的 *SDG8* 属于 ASH1 家族, 能特异性地催化 H3 K36 的双甲基化和三甲基化。通过与 GeneBank 中水稻、玉米、烟草及拟南芥中含有 SET 结构域基因的蛋白质聚类分析后发现, 笔者所克隆的 *BraSDG8* 基因与拟南芥 *SDG8* 基因同源性最高, 且同属于 ASH1 家族(图 3)。由此推测 *BraSDG8* 基因的功能与拟南芥中的 *SDG8* 基因相似, 对开花时间起调控作用。

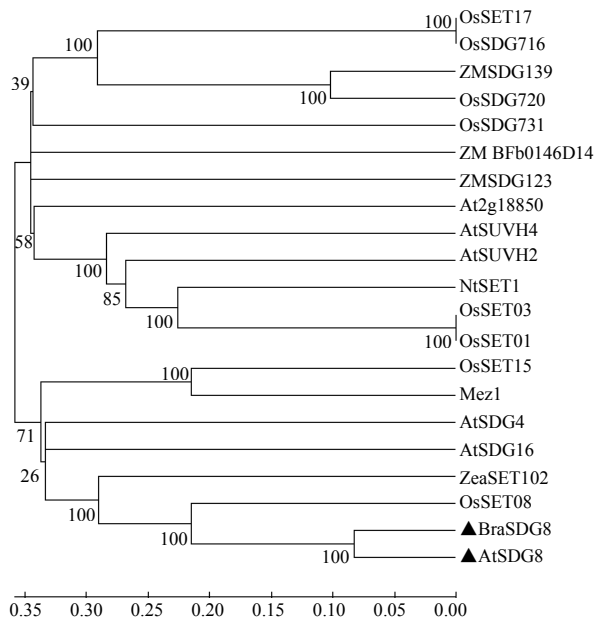


图 3 含编码 SET 结构域的进化树

Fig. 3 A phylogenetic tree of SET-domain-containing proteins

### 3 讨论

FLC 是拟南芥开花时间控制中重要的转录抑制物,当 *SDG8* 的功能缺失时,会导致 H3K36 的二甲基化降低,特别是当染色质缠绕在 FLC 的启动子和第 1 个内含子区域时。*SDG8* 介导的 H3K36 甲基化是在防止早花的 FLC 的表达中 1 个新的表观遗传记忆编码,所以 *SDG8* 的拟南芥突变体表现出 FLC 表达降低导致的早花现象。笔者在白菜型油菜中克隆得到 *SDG8* 的全长 cDNA。序列和蛋白质结构域的分析表明, *BraSDG8* 和拟南芥中的 *SDG8* 在序列上有较高的同源性;在 *SDG8* 的主要功能结构域 SET 结构域的构成上也是相同的。由此推测 *BraSDG8* 在白菜型油菜中对花期的调控可能有着重要作用。

#### 参考文献:

[1] Sims R J, Nishioka K, Reinberg D. Histone lysine methylation: A signature for chromatin function [J]. Trends in Genetics, 2003, 19(11): 629–639.

- [2] Xu L, Zhao Z, Dong A, et al. Di- and tri- but not monomethylation on histone H3 lysine 36 marks active transcription of genes involved in flowering time regulation and other processes in *Arabidopsis thaliana* [J]. Mol Cell Biol, 2008, 28(4): 1348–1360.
- [3] Zhao Z, Yu Y, Meyer D, et al. Prevention of early flowering by expression of *FLOWERING LOCUS C* requires methylation of histone H3 K36 [J]. Nat Cell Biol, 2005, 7(12): 1256–1260.
- [4] Ng D W, Wang T, Chandrasekharan M B, et al. Plant SET domain-containing proteins: Structure, function and regulation [J]. Biochim Biophys Acta, 2007, 1769(5/6): 316–329.
- [5] Strahl B D, Grant P A, Briggs S D, et al. Set2 is a nucleosomal histone H3-selective methyltransferase that mediates transcriptional repression[J]. Mol Cell Biol, 2002, 22(5): 1298–1306.
- [6] Yu Y, Dong A, Shen W H. Molecular characterization of the tobacco SET domain protein NtSET1 unravels its role in histone methylation, chromatin binding, and segregation [J]. Plant J, 2004, 40(5): 699–711.
- [7] 李想, 张飞雄. 组蛋白甲基化的研究进展[J]. 遗传, 2004, 26(2): 244–248.
- [8] Liu S, Yu Y, Ruan Y, et al. Plant SET- and RING-associated domain proteins in heterochromatinization [J]. Plant J, 2007, 52(5): 914–926.
- [9] Thorstensen T, Grini P E, Mercy I S, et al. The *Arabidopsis* SET-domain protein ASHR3 is involved in stamen development and interacts with the bHLH transcription factor ABORTED MICROSPORES (AMS) [J]. Plant Mol Biol, 2008, 66: 47–59.
- [10] Yu Y, Bu Z Y, Shen W H, et al. An update on histone lysine methylation in plants[J]. Progress in Natural Science, 2009, 19(4): 407–413.
- [11] Fischle W, Wang Y, Allis C D. Histone and chromatin cross-talk [J]. Curr Opin Cell Biol, 2003, 15: 172–183.
- [12] Goll M G, Bestor T H. Histone modification and replacement in chromatin activation[J]. Genes Dev, 2002, 16: 1739–1742.
- [13] Zinner R, Albiez H, Walter J, et al. Histone lysine methylation patterns in human cell types are arranged in distinct three-dimensional nuclear zones[J]. Histochem Cell Biol, 2006, 125: 3–19.

责任编辑: 罗慧敏