

• 农业生物技术 •

## 羽衣甘蓝 *SEPALLATA*-like 基因的系统发育与表达分析

相元萍<sup>1</sup>, 黄云彤<sup>2</sup>, 贺洪军<sup>3</sup>, 徐启江<sup>3,4</sup>

1 青岛农业大学 园艺学院, 山东 青岛 266109

2 黑龙江护理高等专科学校, 黑龙江 哈尔滨 150086

3 德州市农业科学研究院, 山东 德州 253015

4 东北林业大学 生命科学学院, 黑龙江 哈尔滨 150040

相元萍, 黄云彤, 贺洪军, 等. 羽衣甘蓝 *SEPALLATA*-like 基因的系统发育与表达分析. 生物工程学报, 2020, 36(11): 2398–2412.

Xiang YP, Huang YT, He HJ, et al. Phylogenetic and expression analysis of *SEPALLATA*-like gene in *Brassica oleracea* L. var. *acephala*. Chin J Biotech, 2020, 36(11): 2398–2412.

**摘要:** E类 MADS-box 基因 *SEPALLATA* (*SEP*)-like 在被子植物生殖生长特别是花器官发育方面具有重要作用。为分析羽衣甘蓝 E 功能 MADS-box 基因 *SEP*-like 基因的序列特征及其在花发育过程中的时空表达模式, 以羽衣甘蓝品系‘14 line’为试材, 利用 cDNA 末端快速扩增 (Rapid amplification of cDNA ends, RACE) 技术克隆了 *SEP* 直系同源基因 *BroaSEP1/2/3* (GenBank 登录号: KC967957、KC967958、KC967960)。序列和系统进化树分析表明, 这 3 个基因分别与野甘蓝 (*Brassica oleracea* var. *oleracea*)、芜菁 *Brassica rapa*、萝卜 *Raphanus sativus*、甘蓝型油菜 *Brassica napus* 的 *SEP1*、*SEP2*、*SEP3* 基因具有很高的同源性。推导的氨基酸序列显示, 这些基因编码的蛋白质都包含高度保守的 MADS 结构域、I 结构域和 K 结构域, 每一基因都有其亚家族特异的 C-末端功能域 SEP I 和 SEP II 基序。*BroaSEP1*、*BroaSEP2*、*BroaSEP3* 基因的开放阅读框长度分别为 801 bp、759 bp、753 bp, 分别编码 266、252、250 个氨基酸残基。半定量 RT-PCR 和实时荧光定量 PCR 研究结果表明, *BroaSEP1*、*BroaSEP2*、*BroaSEP3* 在发育的花芽中特异性表达, 但是表达水平在不同发育时期以及野生型、多瓣型和少瓣型花芽中存在明显差异。

**关键词:** 羽衣甘蓝, 花发育, E 类 MADS-box 基因, *BroaSEP1/2/3* 基因

**Received:** July 11, 2020; **Accepted:** October 8, 2020

**Supported by:** Modern Agricultural Industrial Technology System Funding of Shandong Province, China (No. SDAIT-04-03), Agricultural Variety Improvement Project of Shandong Province, China (No. 662-2316109), The Fundamental Research Funds for the Central Universities (Nos. 2572014EA03, 2572020DY15), Natural Science Foundation of Heilongjiang Province, China (No. C2018002).

**Corresponding author:** Qijiang Xu. Tel: +86-451-82191783; E-mail: qijiangxu@nefu.edu.cn

山东省现代农业 (蔬菜) 产业技术体系项目 (No. SDAIT-04-03), 山东省良种工程项目 (No. 662-2316109), 中央高校基本科研业务费专项资金 (Nos. 2572014EA03, 2572020DY15), 黑龙江省自然科学基金 (No. C2018002) 资助。

网络出版时间: 2020-10-14

网络出版地址: <https://kns.cnki.net/kcms/detail/11.1998.q.20201013.0948.003.html>

# Phylogenetic and expression analysis of *SEPALLATA*-like gene in *Brassica oleracea* L. var. *acephala*

Yuanping Xiang<sup>1</sup>, Yuntong Huang<sup>2</sup>, Hongjun He<sup>3</sup>, and Qijiang Xu<sup>3,4</sup>

<sup>1</sup> College of Horticulture, Qingdao Agricultural University, Qingdao 266109, Shandong, China

<sup>2</sup> Heilongjiang Nursing College, Harbin 150086, Heilongjiang, China

<sup>3</sup> Dezhou Academy of Agricultural Sciences, Dezhou 253015, Shandong, China

<sup>4</sup> The College of Life Sciences, Northeast Forestry University, Harbin 150040, Heilongjiang, China

**Abstract:** The E class MADS-box genes *SEPALLATA* (*SEP*)-like play critical roles in angiosperm reproductive growth, especially in floral organ differentiation. To analyze the sequence characteristics and spatio-temporal expression patterns of E-function MADS-box *SEP*-like genes during kale (*Brassica oleracea* L. var. *acephala*) flower development, *BroaSEP1/2/3* (GenBank No. KC967957, KC967958, KC967960) homologues, three kale *SEP* MADS-box gene, were isolated from the kale variety 'Fourteen Line' using Rapid amplification of cDNA ends (RACE). Sequence and phylogenetic analysis indicated that these three *SEP* genes had a high degree of identity with *SEP1*, *SEP2*, *SEP3* from *Brassica oleracea* var. *oleracea*, *Brassica rapa*, *Raphanus sativus* and *Brassica napus*, respectively. Alignment of the predicted amino acid sequences from these genes, along with previously published subfamily members, demonstrated that these genes comprise four regions of the typical MIKC-type MADS-box proteins: the MADS domain, intervening (I) domain and keratin-like (K) domain, and the C-terminal domain SEP I and SEP II motif. The longest open reading frame deduced from the cDNA sequences of *BroaSEP1*, *BroaSEP2*, and *BroaSEP3* appeared to be 801 bp, 759 bp, 753 bp in length, respectively, which encoded proteins of 266, 252, and 250 amino acids respectively. Expression analyses using semi-quantitative RT-PCR and quantitative real-time PCR indicate that *BroaSEP1/2/3* are specifically expressed in floral buds of kale during flower development process. The expression levels of the three genes are very different at different developmental stages, also in wild type, mutant flower with increased petals, and mutant flower with decreased petals. These different patterns of gene expression maybe cause the flowers to increase or decrease the petal number.

**Keywords:** kale (*Brassica oleracea* var. *acephala*), flower development, E-class MADS-box genes, *BroaSEP1/2/3* gene

关键调控基因的复增及其功能趋异是生物形态多样性产生和进化的重要途径<sup>[1-2]</sup>。作为被子植物特有的创新性状和重要的生殖器官，决定花器官发育的 MADS-box 基因其数目与功能的改变是花形态多样性的基础<sup>[3-6]</sup>。例如拟南芥 A 类基因 *APETALA1* (*AP1*) 和 *APETALA2* (*AP2*)<sup>[7-8]</sup>、B 类基因 *APETALA3* (*AP3*) 和 *PISTILLATA* (*PI*)<sup>[9-11]</sup>、C 类基因 *AGAMOUS* (*AG*)<sup>[12]</sup>、D 类基因 *SEEDSTICK* (*STK*)<sup>[13-15]</sup>、E 类基因 *SEPALLATA1/2/3/4* (*SEP1/2/3/4*)<sup>[16-18]</sup>。这些编码调控花发育关键转录因子的 MADS-box 基因经历了大量基因复增事件，通过编码序列和 (或) 表达区域的改变而发生亚功能化和新功能化，进而导致被子植物花形态结构的多样性<sup>[1-2,19]</sup>。花器官发育的 ABCE 理论

模型阐释了 4 类花器官同源异型基因 (A、B、C 和 E) 决定各花器官特征属性的分子机制<sup>[20-27]</sup>。A+E 功能基因控制萼片发育；A+B+E 功能基因控制花瓣发育；B+C+E 功能基因控制雄蕊发育；C+E 功能基因控制雌蕊发育。除 A 类基因 *AP2* 属于 *APETALA2/EREBP* 基因家族外，所有的 A、B、C、E 基因均属于 MADS-box 转录因子基因家族，以高度有序的 MADS 蛋白质复合体形式<sup>[28]</sup>结合在 CArG 序列[一致性序列：5'-CC(A/T)<sub>6</sub>GG-3']上而激活靶基因的表达，如 *FD*<sup>[15]</sup>、*UNUSUAL FLORAL ORGANS CRABS CLAW*<sup>[29]</sup>。

*SEP* MADS-box 基因不仅调控花分生组织的确定性，而且与 A、B 和 C 类 MADS-box 基因共同决定花器官特征属性<sup>[18,30-31]</sup>，在被子植物花起

源和发育中发挥关键作用<sup>[32-34]</sup>。在现存被子植物多样性产生之前,发生第二次基因复增事件,*SEP* 亚家族产生 *LOSEP* 和 *SEP3* 进化系。随后, *SEP* 亚家族在有花植物中经历多次独立的物种特异性基因复增<sup>[35-39]</sup>。在双子叶植物类群中, *LOFSEP* 进化系经历 2 次复增产生 3 类进化支: *AGL3* (包括 *AtSEP4*)、*FLORAL BINDING PROTEIN9/23* (*FBP9/23*) 和 *AGL2/4* (包括 *AtSEP1* 和 *AtSEP2*)。在单子叶植物类群中同样经历 2 次复增而分别产生 *OsMADS34* (*PAP2*) 进化支和 *LEAFY HULL STERILE1* (*LHS1*)/*OsMADS1* 与 *OsMADS5* 进化支。与 *LOFSEP* 进化系相比较, *SEP3* (*AGL9*) 基因进化系所经历的复增事件较少,基因拷贝数少。目前,仅从双子叶植物中鉴定出了单一的 *SEP3* 拷贝。在禾本科植物中, *SEP3* 进化系发生 1 次基因复增事件而形成 *OsMADS7* 和 *OsMADS8* 进化枝<sup>[32-33]</sup>。*SEP* 同源基因间存在功能冗余现象。在花发育阶段 2,拟南芥 *SEPI/2/4* 基因在整个花分生组织中表达,而 *SEP3* 仅在花原基起始前于形成内三轮花器官的区域内表达<sup>[13,18,40-41]</sup>。随着花原基的发育, *SEPI/2* 在所有的花器官原基中表达, *SEP3* 在内三轮花器官原基内表达,而 *SEP4* 在萼片中的表达明显降低。*sep1/2/3/4* 突变体的所有花器官转变为叶状器官,表明 *SEP* 基因在促进花分生组织确定性和决定花器官特征属性的功能上具有冗余性<sup>[18]</sup>。已从多种模式植物中分离鉴定出了 E 类基因,例如金鱼草 *Antirrhinum majus* 的 *DEFH49/72/200* 基因、番茄 *Lycopersicon esculentum* 的 *TM5/29* 基因、白麦瓶草 *Silene latifolia* 的 *SISEP1/SISEP3* 基因等。

羽衣甘蓝 (*Brassica oleracea* var. *acephala*) 花的形态结构与拟南芥的相同,四轮花器官由外向内分别为 4 枚萼片、4 枚花瓣、6 枚雄蕊、1 枚雌蕊。但是,在同一品系群体内也存在花瓣数目减少及增多的纯合突变体 (图 1)。为分析 *SEP-like* 基因在羽衣甘蓝花发育中的作用及其与花瓣数目突变体表型的关系,本研究利用 cDNA 末端快速扩增 (Rapid amplification of cDNA ends, RACE) 技术克隆调控羽衣甘蓝花发育的 E 类 MADS-box 基因,在利用生物信息学方法分析该类基因的序列结构特征及其系统发育的基础上,基于半定量 RT-PCR 和 qRT-PCR 对该类基因在羽衣甘蓝花发育不同时期的表达水平,以期利用 E 类 MADS-box 进行羽衣甘蓝花型的遗传改良 (包括增加花型多样性特别是多瓣花)、提高观赏价值奠定前期研究基础。

## 1 材料与方法

### 1.1 试验材料

羽衣甘蓝自交系‘14 Line’ (花瓣数目正常, 4 枚)、在‘14 Line’群体中出现多瓣 (8 枚以上) 突变体、少瓣 (0-3 枚) 突变体 (图 1), 进一步单株自交 2 代, 突变性状稳定。由东北林业大学生命科学学院花卉生物工程研究所提供。野生型和两种突变体的遗传背景相同。2016 年 6 月中旬将羽衣甘蓝种植于温室内, 常规栽培管理, 第 2 年 2 月中旬开始现蕾。采集 5 个时期的花蕾 (花) (F1、F2、F3、F4、F5) (图 2), 同时采集开花期植株的根、茎、叶, 用铝箔纸包裹后立即液氮速冻, 保存于-80 °C 冰箱, 用于提取总 RNA。



图 1 羽衣甘蓝野生型及突变型花

Fig. 1 Kale wild-type and mutant flowers. (A) Wild-type flower. (B) Mutant flower with increased petals number. (C) Mutant flower with decreased petals number.



图2 羽衣甘蓝花发育的5个时期  
Fig. 2 Five developmental stages of kale flower.

## 1.2 RNA 提取与 cDNA 合成

Trizol 法 (Cat. No. 15596018, Thermo Fisher Scientific) 提取野生型花蕾的总 RNA, NanoDrop1000 微量紫外分光光度计与甲醛琼脂糖凝胶电泳检测提取 RNA 的浓度和质量。利用 TransScript<sup>®</sup> 第一链 cDNA 试剂盒 (Cat. No. AT301-02, 北京全式金生物技术有限公司) 进行反转录合成第一链 cDNA。在 20  $\mu$ L 反应体系中加入 5  $\mu$ g 总 RNA、1.0  $\mu$ g 引物 P18E (5'-GACTCGAGTGCACATCG(T)<sub>17</sub>-3') (表 1)、10  $\mu$ L 2 $\times$ TS 反应混合液、1  $\mu$ L SuperScript<sup>®</sup> RT/RI 酶混合液。充分混匀后于 42  $^{\circ}$ C 条件下温育 30 min, 然后 80  $^{\circ}$ C 温育 5 s 终止反应。反应结束后加入 1  $\mu$ L RNase H (2 U/ $\mu$ L), 37  $^{\circ}$ C 温育 20 min, 降解可能存在的 RNA。

## 1.3 羽衣甘蓝 *SEPALLATA*-like 基因的 cDNA 扩增

根据 *SEPALLATA*-like 转录因子 MADS 区的保守氨基酸序列 VLCDAEV 设计正义兼并引物 SEP-F (5'-GTTCTHTGYGATGCWGAGGT-3'), 以接头引物 18E (5'-GACTCGAGTGCACATCG-3') 为反义引物、第一链 cDNA 为模板克隆 *SEP*-like 基因的 3'-cDNA 序列。50  $\mu$ L 反应体系中包含 0.5  $\mu$ L *LA Taq* DNA 聚合酶 (5 U/ $\mu$ L)、5.0  $\mu$ L 10 $\times$ *LA* PCR

Buffer II、5.0  $\mu$ L dNTPs (2 mmol/L)、1.0  $\mu$ L *SEP*-F (20  $\mu$ mol/L)、1.0  $\mu$ L 18E (20  $\mu$ mol/L)、2.0  $\mu$ L cDNA。PCR 扩增条件: 94  $^{\circ}$ C 预变性 5 min; 94  $^{\circ}$ C 变性 30 s, 54  $^{\circ}$ C 退火扩增 30 s, 72  $^{\circ}$ C 延伸 90 s, 25 个循环; 72  $^{\circ}$ C 延伸 10 min。PCR 产物利用 EZNA<sup>™</sup> 胶回收试剂盒 (Cat. No. D2501-01, OMEGA) 回收约 950 bp 的 PCR 产物, 用 *pEASY*<sup>®</sup>-T5 Zero Cloning Kit (Cat. No. CT501-02, 北京全式金生物技术有限公司) 进行 PCR 纯化产物的克隆, 提取质粒 (TIANprep Rapid Mini Plasmid Kit, Cat. No. DP105, TIANGEN), 送北京六合华大基因科技有限公司测序。依据 3'-cDNA 序列设计 *SEP*-like 基因特异引物 (表 1), 用 5'/3'-RACE 试剂盒 (SMARTer<sup>®</sup> RACE 5'/3' Kit, Cat. No. 634858, TaKaRa), 以 *SEP*-like 基因特异引物和 oligo(dT) 锚定引物通过巢式 PCR 克隆 *SEP*-like 基因 5'-cDNA 序列。PCR 产物回收纯化、克隆、测序与 3'-RACE 实验操作程序一致。将 5'-和 3'-序列拼接全长 cDNA 序列, 设计特异引物克隆 *SEP*-like 基因全长 cDNA。

## 1.4 序列比对及系统进化树的构建

为分析羽衣甘蓝 E 类 *SEP*-like 基因 *BroaSEP1/2/3* 的系统发育, 在 NCBI (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) 上对该 cDNA 序列进行核苷酸 Blast 分析。依据核苷酸序列推导出氨基酸序列, 与已知的 *SEP* 蛋白利用 GenDoc 软件进行同源性比对, 用 MEGA7.0 软件以邻近相连法 (Neighbor joining, NJ) 构建系统发育树, 并进行 Bootstrap 检测。采用重复抽样分析系统发育树分支的置信度, 重复抽样次数为 1 000 次, 大于 50% 的 bootstrap 标注在树图上。

## 1.5 基于半定量 RT-PCR 和 qRT-PCR 的 *BroaSEP1/2/3* 基因表达分析

提取 F1、F2、F3、F4、F5 花蕾的总 RNA 作为模板, 利用 Oligo(dT)<sub>17</sub> 引物反转录合成第一链 cDNA。再以第一链 cDNA 为模板进行基因特异

性 RT-PCR, 以 *Broa18S* rRNA 基因作为内参。

在 PE-9700 型 PCR 仪上进行半定量 RT-PCR, *BroaSEPI/2/3* 与 *Broa18S* rRNA 基因特异引物 *BroaSEPI/2/3*-RT-F、*BroaSEPI/2/3*-RT-R 以及 *Broa18S*-RT-F、*Broa18S*-RT-R 见表 1。PCR 扩增程序为: 94 °C 预变性 5 min; 94 °C 变性 30 s, 51 °C 退火 30 s, 72 °C 延伸 30 s, 25 个循环; 72 °C 终延伸 5 min。1.0% 琼脂糖凝胶检测分析 PCR 扩增产物的质量和浓度。

以 5 个发育时期的羽衣甘蓝花蕾第一链 cDNA 为模板, 利用 SYBR<sup>TM</sup> Green PCR Master Mix 试剂盒 (Cat. No. 4344463, Thermo Fisher

Scientific) 配制 PCR 反应体系, 7500 fast 型实时荧光定量 PCR 仪进行 qRT-PCR, 分析 *BroaSEPI/2/3* 基因在野生型、多瓣突变体、少瓣突变体花发育不同时期的特异性表达模式。20 μL 反应体系: 10 μL 稀释的第一链 cDNA 模板 (2 ng cDNA)、浓度为 10 μmol/L 的正、反向引物各 0.8 μL、SYBR<sup>®</sup> Green PCR Master Mix (2×) 10 μL。反应程序如下: 95 °C 预变性 2 min; 随后 40 个扩增循环 (95 °C 10 s, 60 °C 45 s)。采用 2<sup>-ΔΔC<sub>t</sub></sup> 法分析 *BroaSEPI/2/3* 基因在不同发育时期的相对表达水平。每个样品设置 3 个生物学重复。

表 1 羽衣甘蓝 *BroaSEP* 基因克隆及表达分析所用引物

Table 1 List of primers used for cloning and expression analysis of *BroaSEP* from kale

No.	Primer name	Sequence (5'-3')
1	P19E	GACTCGAGTGCACATCG(T) <sub>17</sub>
2	P18E	GACTCGAGTGCACATCG
3	<i>BroaSEP</i> gene specific primer for 3'-RACE	GTTCTT(C/A)TGT(C)GATGCT(A)GAGGT
4	<i>BroaSEPI</i> GSP (gene specific primer)1	CCCTCATATCTACCCTTGAGC
5	<i>BroaSEPI</i> GSP2	GAGCTTGCCACGGTTGGAGA
6	<i>BroaSEP2</i> GSP1	CTCGAGTTCTTTGGCAGGTTTG
7	<i>BroaSEP2</i> GSP2	TCTTGAGCATGTTGGAGGTGCT
8	<i>BroaSEP3</i> GSP1	GTCCGGAGAGCTCTGATATGC
9	<i>BroaSEP3</i> GSP2	CATCTGATACCCATCAGCTAACC
10	Oligo(dT) <sub>17</sub> anchor primer	GGCCACGCGTCGACTAGTAC(T) <sub>17</sub>
11	Adaptor primer	GGCCACGCGTCGACTAGTAC
12	<i>BroaSEPI</i> -F	GGGCATATCTCCTTCTCAAGAC
13	<i>BroaSEPI</i> -R	CTGGAGATCCTCAGGAC
14	<i>BroaSEP2</i> -F	CCCTACTACAATCAAATCAA
15	<i>BroaSEP2</i> -R	GGCTATGACAGTTATAGCC
16	<i>BroaSEP3</i> -F	AAAGGATTCAACGGGAGAG
17	<i>BroaSEP3</i> -R	GCACACGAGACAGAGTATAG
18	<i>BroaSEPI</i> -RT-F	CAGAGCAGATAACTGCAACAAC
19	<i>BroaSEPI</i> -RT-R	GCGAGATTGATCACAGCATTAG
20	<i>BroaSEP2</i> -RT-F	AGTCATCCCAACCAGGAAA
21	<i>BroaSEP2</i> -RT-R	TCCTTATTGGGAGGGTAGAAGA
22	<i>BroaSEP3</i> -RT-F	GGTTACCGTATGACACCAACTC
23	<i>BroaSEP3</i> -RT-R	CGAGACAGAGTATAGAGAGAAGGG
24	<i>Broa18S</i> -RT-F	AGCCTGAGAAACGGCTAC
25	<i>Broa18S</i> -RT-R	CGAAGAGCCCGGTATTGTTATT



## 2 结果与分析

### 2.1 羽衣甘蓝 E 类 MADS-box 基因 *BroaSEP1/2/3* 克隆与序列结构分析

以野生型 F1 至 F5 发育期的花蕾 (花) 构成混合样品池提取的总 RNA (图 3A) 经 P19E 反转录合成的第一链 cDNA 为模板, 利用 *BroaSEP1/2/3* 基因 3'-RACE 特异引物和 P18E 接头引物分离羽衣甘蓝 *SEP*-like 基因的 3'-端序列, 3'-RACE 扩增产物的长度约为 950 bp (图 3B)。PCR 产物回收、克隆、测序、比对, 克隆得到了 3 条 cDNA 序列, 长度分别为 873 bp、974 bp 和 865 bp。核苷酸序列比对结果表明, 分别与甘蓝型油菜 *Brassica napus*、油菜 *Brassica rapa* 的 *SEPALLATA1*-like (XM\_013874090.2; XM\_009133252.3)、*SEPALLATA2*-like (XM\_013792022.2 ; XM\_009119650.3)、*SEPALLATA3*-like (XM\_022714754.1; XM\_013839134.2) 高度同源, 序列相似性达 97%–100%。其编码的氨基酸序列与芸薹属植物的 *SEP1/2/3* 蛋白高度同源。因此, 将克隆获得的序列分别命名为 *BroaSEP1*、*BroaSEP2*、*BroaSEP3* 基因。

混合花蕾 (花) 的总 RNA 经 Oligo(dT)<sub>17</sub> 锚定引物反转录合成第一链 cDNA, 用 *BroaSEP1*、*BroaSEP2*、*BroaSEP3* 基因特异引物和接头引物进行巢式 5'-RACE 扩增。PCR 产物回收、克隆、测序, 克隆获得了 *BroaSEP1*、*BroaSEP2*、*BroaSEP3*

基因的 5'-cDNA 序列, 长度分别为 312 bp、536 bp 和 580 bp。用 DNAMAN 软件分别将 *BroaSEP1/2/3* 基因 3'和 5'序列进行拼接, 得到 3 个基因的全长 cDNA 序列 (图 4)。最后用 3 个基因各自的特异引物 (*BroaSEP1*-F/*BroaSEP1*-R、*BroaSEP2*-F/*BroaSEP2*-R、*BroaSEP3*-F/*BroaSEP3*-R, 表 1) 扩增, 确定了 cDNA 序列的准确性。*BroaSEP1*、*BroaSEP2*、*BroaSEP3* 基因的 GenBank 登录号为分别为 KC967957、KC967958、KC967960。利用 NCBI 的开放阅读框预测工具 (Open reading frame finder) 进行最大开放阅读框推测。结果表明, *BroaSEP1* 的开放阅读框长 801 bp, 编码 266 个氨基酸, 包含 73 个氨基酸残基的 MADS 结构域、27 个氨基酸残基的 I 结构域、87 个氨基酸残基的 K 结构域、79 个氨基酸残基的 C 结构域 (图 4)。理论等电点 pI 为 9.05, 分子量为 30.6 kDa。*BroaSEP2* 的开放阅读框长 759 bp, 编码 252 个氨基酸, 包含 57 个氨基酸残基的 MADS 结构域、31 个氨基酸残基的 I 结构域、87 个氨基酸残基的 K 结构域、77 个氨基酸残基的 C 结构域 (图 4)。理论等电点 pI 为 8.65, 分子量为 28.8 kDa。*BroaSEP3* 的开放阅读框长 753 bp, 编码 250 个氨基酸, 包含 57 个氨基酸残基的 MADS 结构域、31 个氨基酸残基的 I 结构域、87 个氨基酸残基的 K 结构域、75 个氨基酸残基的 C 结构域 (图 4)。理论等电点 pI 为 7.68, 分子量为 28.9 kDa。

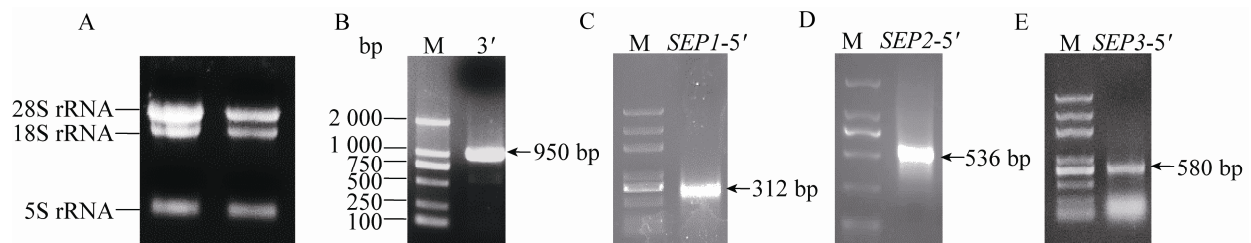


图 3 羽衣甘蓝花芽 *BroaSEP1/2/3* 基因全长 cDNA 序列克隆电泳图

Fig. 3 Electrophoresis of amplified products of *BroaSEP1/2/3* gene full-length cDNA sequence in floral buds of kale.

**BroaSEPI**

1	GGGCATATCTCGTTCTCAAGACCCCTAAAGAACAAAAAGATCAGATCTCAGATTTTTGTTGTCTGTATATAGATAAAACATTTACATACCAATAT	
100	TGTATGCATATATATATAAAAGAGAGAAACTACAGAGCAACAAAAATAAATGGGAAGGGGAAGAGTAGAGCTGAAGAGGATAGAGAACAAGATCAAC	
	<u>M H I Y I K E R K L Q S N K N K M G R G R V E L K R I E N K I N</u>	32
199	AGACAAGTAACGTTTGCAAAACGTAGGAACGGTTTGTGAAGAAAGCTTACGAATTTGTTGTCTGTCTGTATAGATAAAACATTTACATACCAATAT	
	<u>R Q V T F A K R R N G L L K K A Y E L S V L C D A E V A L I I F S</u>	65
298	AACCGTGGCAAGCTCTAGTCTCTGAGCTCTCTAATCATGATCAAGCACTGGAACCGTACCAGAAATGCAAGTTATGGTTCTATTGAAGTCAACAAC	
	<u>N R G K K L Y E F C S S S N M I K T L E R Y Q K C S Y G S I E V N N</u>	98
397	AAACCTGCCAAAGAACTTGAGAAATAGTCACAGAGATATCTGAAGCTCAAGGGTAGATATGAGGGCTTCAAGCTCAACAAGAAATCTCTTGGGGAG	
	<u>K P A K E L E N S Y R E Y L K L K G R Y E G L Q R Q Q R N L L G E</u>	131
496	GATTTAGACCTCTGAATTCAGAGGAGTTAGAGCAGATTGAGCGTCAACTAGATGGCTCTCTCAAGCAAGTTCGCTCCATCAAGACACAGTACATGCTT	
	<u>D L G P L N S K E L E Q I E R Q L D G S L K Q V R S I K T Q Y M L</u>	164
595	GACCAGCTCTGACCTTCAAACCAAGAGCAAATGTTGCTTGAACCAATAGAGCTTGGCTATGAAGCTGGATGATGATGGTGTGAGAGTCAAT	
	<u>D Q L S D L Q T K E Q M L L E T N R A L A M K L D D M I G V R S H</u>	197
694	CATATGGGTGGAGGAGGAGGATGGGAAGCAATGAACATAATGTTTCCTATGCGCATCATCAAGCTCAGTCTCAAGGACTATCCAGCCTTTGAA	
	<u>H M G G G G G W E G N E H N V S Y A H H Q A Q S Q G L F Q P L E</u>	230
793	TGCAATCCAACCTCTCAGATGGGGTATGACAATCCAGTATGCTCAGAGCAGATAACTGCAACAACAAGCTCAGGACACAGCCGGTACATTCAGAC	
	<u>C N P T L Q M G Y D N P V C S E Q I T A T T Q A Q A Q P G Y I P D</u>	263
892	TGGATGCTGAAAGTCATGTGATCATTCTAATGCTGTGATCAATCTGCCAACAAATAAAGACCTGTTGATATATAAAGAAAGTGAGACACAAG	
	<u>W M L *</u>	266
991	ACTTGAATTTAGACATAATATGTAATGCTCAGGATCTCCAGTACATTTGTGATTTTTGGGAACCTTGCTATATAAGGGTTCGAATATGTGGAA	
1 090	CTTGATTAATAAAAAAAAAAAAAA	
1	GATAGCTTTTAAAGATTGACAAAAGCTTCTCAGATTACAATCTCACACACACACAACCTTCAAAAAAGAAAGAAAGACAGACAAAAGATC	
100	TAAAGAATAAACAAGAACCTACTACAATCAAATCAAAACCAACAAGAAGCAAAAGTTTCTATTTATATCTCTCTCTCTCTCTCTCTTGTTGAAA	
199	AATAGGGTTTTGTATTTATGTCAAAAGAAAAGATCTTTCATCAGAAAAAGCAACCAAACTCATGTTGTGTGTGTGTAATATAGATAAACA	
299	TTACATCACATAATAAGTTACATATAAAGAAGAAAGAAAATGGGAAGGGGAAGGTAGAGCTAAAGAGGATCGAGAACAAGATTAACAGACAACTC	
	<u>M G R G R V E L K R I E N K I N R Q V</u>	19
398	ACGTTTGTCAAGCTAGGAACGGTTTGTGAAAAAGCCTATGAGCTTCAGTCTCTGCGATGCTGAGGTTTCTCTCATCGTCTCTCCAACCGTGGC	
	<u>T F A K R R N G L L K K A Y E L S V L C D A E V S L I V F S N R G</u>	52
497	AACTCTACGAATTCGAGCACTCCAACATGCTCAAGCACTGGAAGGTACCAGAAGTGTAGCTATGTTCAAGTCAAGCAACAACCTGCC	
	<u>K L Y E F C S T S N M L K T L E R Y Q K C S Y G S V E V N N K P A</u>	85
596	AAAGAATCGAATAGCTACAGAGATTTGAAGCTGAAGGTAGATAGAAAATCTGCAACGCCAGCAGAGAAATCTACTTGGAGAAGACTTGGGA	
	<u>K E L E N S Y R E Y L K L K G R Y E N L Q R Q Q R N L L G E I L G</u>	118
695	CCCTCAACTCAAAGGAGCTAGAGCAGCTTGAAGCCTCAACTAGAGCGCTCTCAAGCAAGTTCGTTGCATCAAGACACAGTATATGCTTACCAGCTC	
	<u>P L N S K E L E Q L E R Q L D G S L K Q V R C I K T Q Y M L D Q L</u>	151
794	ACTGACCTCCAAGGCAAGAGCATACTTGTCTGAAGCCAATCGTCTTGTCAATGAAGCTGGAAGATATGATAGGCTGAGAAGTCAACAAATAGGA	
	<u>A L L Q G K E H I L L E A N R A L S M K L E D M I G V R S H I L G</u>	184
893	GGGGCGTGGAAAGGTGATCAACAACATGTTGCTATGGACATCAGGCTCAATCTCAGGACTATCCAGTCTTTGAGTGTGATCCCACTTTG	
	<u>G A W E G G D Q Q H V A Y G H H Q A Q S Q G L F Q S L E C D P T L</u>	217
992	CAAATGGGATACAACCATCCAGTGTCTCAGAGCAATGTAGTACGGCACAAGGTGAGTATCCCAACAGGAAACAACGGCTACATCCCTGGCTGG	
	<u>Q M G Y N H P V C S E Q M V Y T A Q G Q S S Q P G N N G Y I P D W</u>	250
1 091	ATGCTGTGAGTGTACTTCTTACCCTCCCAATAAGGATCTTAAAGATGCAAGTGGTGGTCTTGTATGATGATGAATGAGAATGGATCTACGA	
	<u>M L *</u>	252
1 190	ATTAAGTATGGTTTGTGTTATATATGCTGCTGCATAAGACTCTGATTTCTAGACATAAAGTGGCTATAACTGTCATAGCCCTCAATATCTCTCTCT	
1 289	GTTTAGTAGATGTTTGTGCTTTTGGAGCCTTGTCTATATATAAATGGATGTAATGAGTGAATGTAACATTAACCATGTCTGATCTGGTTAACT	
1 388	AAGTAAAAAAAAAAAAA	
1	AAGGAGAGAGAGGCTTCATTCATATATAAAGATTCAACAAGGGAGAGAGAGGAAAAAATAAAGAAAATGGGAAGAGGAGAGTGGAGTTGAAGAGG	
100	<u>M G R G R V E L K R</u>	10
199	ATAGAGAACAAGATCAATAGGCAAGTGACGTTTCAAGAAGAAGAAATGGCCTTTTGAAGAAAGCATACAGCTTTCCGTTCTATGTATGATGAGCAGGTT	
298	<u>I E N K I N R Q V T F A K R R N G L L K K A Y E L S V L C D A E V</u>	43
397	GCTCTCATCTCTCTAGTAGAGGAAAACAGTACGAGTTTTCAGTGTGAGCATGCTTCGGACACTAGACAGTACCAGAAGTCAATACCGGA	
	<u>A L I I F S S R G K Q Y E F C S S S S M L R T L D R Y Q K C N Y G</u>	76
496	GCTCCAGGCCAATGTGCTTCAAGAGAGGCTTAGCAGAAGTAAAGCAGCAGGATATCTCAAGCTTAAGGAGCGTTACGACGCTTACAGAGA	
	<u>A P E P N V P S R E A L A E L N S Q Q E Y L K L K E R Y D A L Q R</u>	109
595	ACTCAAAGGAATCTATTGGGAGAAGATCTGGACCTCTTAGCACAAGAGCTTGAAGTCACTTGAAGACAGCTTGATTTCTTGAAGCATATCAGA	
	<u>T Q R N L L G E D L G P L S T K E L E S L E R Q L D S S L K H I R</u>	142
694	GCTCTCCGGACACAATTCATGCTCGACCAGCTCAATGATCTCCAGAGTAAAGAACGCATGCTGGCTGAGACAAAATAAAGCTTAAAGGTAAGGTTAGCT	
	<u>A L R T Q F M L D Q L N D L Q S K E R M L A E T N K T L R L R L A</u>	175
793	GATGGGTATCAGATGCCACTCAACTCAACCCGAACCCAGAAGTATGACTACGGTGTGATCAACAACATGAACACTCTCATCATCAAGCTTTCTTC	
	<u>D G Y Q M P L Q L N P N P E D H D Y G R H Q Q H E H S H H F</u>	208
892	CAGCCTTGGAAATGCGAACCATTCTTCAATGGGGTGACGGGCAGCAAGATCATGGAATGGGAGCAGGACCCAGTGAATAATACATGTTGGGT	

图 4 羽衣甘蓝 *BroaSEP* cDNA 的核苷酸序列及其推导的氨基酸序列  
 Fig. 4 Nucleotide and deduced amino acid sequences of *BroaSEP* cDNA from kale. MADS, I, K and C domains are underlined and defined by single, double, wave and dash line, respectively. The positions of the nucleotides and amino acids are shown on the left and the right, respectively.

## 2.2 羽衣甘蓝 *BroaSEPI/2/3* 基因的同源性分析及系统发育分析

多重氨基酸序列比对结果表明, *BroaSEPI/2/3* 蛋白具有被子植物 MADS-box 蛋白质典型的 MIKC 结构<sup>[42-43]</sup> (图 3-4)。高度保守的 M 结构域大约由 55-60 个氨基酸残基构成<sup>[44]</sup>, 包含保守的核定位序列<sup>[45]</sup>, 在转录因子二聚体化及核定位的过程中发挥功能; 位于 M 与 K 结构域之间的 I (Intervening) 结构域序列保守性较差, 由 25-30 个氨基酸残基构成, 主要参与二聚体的形成<sup>[46]</sup>; K (Keratin-like) 结构域含有 3 个  $\alpha$ -螺旋, 主要作用是介导 MADS 转录因子形成二聚体和四聚体, 参与花器官属性决定<sup>[47]</sup>; C 末端结构 (C-terminal domain) 序列保守性差, 但是也含有短序列保守基序 (Motif), 具有靶基因转录激活作用或促进多聚蛋白复合体形成<sup>[43-44]</sup>。

羽衣甘蓝的 *BroaSEPI/2/3* 蛋白分别与甘蓝型油菜 (*Brassica napus*) 的 *SEP1* (XM\_013874090.1)、*SEP2* (XP\_013647476.1)、*SPE3* (XP\_013647476.1) 蛋白, 野甘蓝 (*Brassica oleracea* var. *oleracea*)<sup>[48]</sup> 的 *SEP1* (XP\_013625167.1)、*SEP2* (XP\_013639591)、*SPE3* (XP\_013638955.1), 萝卜 (*Raphanus sativus*) 的 *SEP1* (XP\_018475126.1)、*SEP2* (XP\_018437242.1)、*SPE3* (XP\_018480838.1), 芜菁 (*Brassica rapa*) 的 *SEP1* (XP\_009131500.1)、*SEP2* (XP\_009117898.1) 高度同源, 氨基酸序列的相似性大于 96%, 说明 *SEP*-like 基因在十字花科芸薹属植物中是高度保守的。

E 类 MADS-box 基因家族可分为两个进化系 *LOFSEP* (包括 *SEPI/2/4*) 和 *SEP3*, 每个进化系都具有高度保守的 *SEP I* 和 *SEP II* 基序<sup>[32-33,49]</sup>。*BroaSEPI/2/3* 蛋白的 C 末端区域具有保守的 *SEP I* 和 *SEP II* 基序 (图 5), 证明克隆获得 *BroaSEPI/2/3* 基因属于 B 类 MADS-box 基因中的 *SEP* 进化系, C 末端序列尽管保守性较差, 然而, 同属植物还是显示出较高的保守性。

为深入分析羽衣甘蓝 *BroaSEPI/2/3* 基因与其

他被子植物 E 类 MADS-box 基因间的系统发育关系, 利用 MEGA7.0 软件, 以拟南芥 *Arabidopsis thaliana* 的 *AP1* (Z16421.1)、非洲菊 *Gerbera hybrida* 的 *GSQUA2* (CAX65661.1)、芜菁 *Brassica rapa* 的 *BrraAP1* (XP\_009105460.1) 等 A 类 MADS-box 基因为外部类群, 从 DDBJ/EMBL/GenBank 数据库中选取 3 个 A 类、9 个 AGL6 类、75 个 *SEP* 类 MADS 蛋白, 对 MIKC 区氨基酸序列进行 NJ (Neighbor-Joining) 分析, 构建系统发育树 (图 6)。E 功能 MADS-box 蛋白被划分为 *LOFSEP* 和 *SEP3* 进化系<sup>[32]</sup>, 其中 *LOFSEP* 进化系又细分为 *SEPI/2*、*FBP9* 和 *SEP4* 3 个单元组<sup>[50]</sup>。同属物种的 *SEP* 同源蛋白具有较高的自展支持率。羽衣甘蓝 *BroaSEPI/2/3* 蛋白分别与甘蓝型油菜、野甘蓝、萝卜、芜菁的 *SEP* 蛋白聚为一支, 分别属于 *SEPI/2* 和 *SEP3* 进化系。

## 2.3 *BroaSEPI/2/3* 基因在羽衣甘蓝花发育不同时期的表达模式

半定量 RT-PCR 和 qRT-PCR 分析 *BroaSEPI/2/3* 在野生型、多瓣型和少瓣型花发育的 5 个时期及营养器官根、茎、叶中的特异性表达 (图 7)。结果表明, *BroaSEPI/2/3* 基因在花发育的 5 个时期均有表达, 但是表达水平在野生型、多瓣型和单瓣型花中存在差异。*BroaSEPI* 基因在野生型花发育的前 4 个时期高丰度表达, 而在花发育的第 5 时期则弱表达; 在多瓣型花的整个发育过程均高丰度表达; 在少瓣型花发育的第 2、3 时期弱表达。*BroaSEP2* 基因在野生型花发育的 1、4 时期表达水平较高, 在第 5 时期几乎无表达; 在多瓣型花的整个发育过程均有表达, 但第 1 时期弱表达; 而在少瓣型花发育的第 4 时期表达水平最高, 而在发育初期表达水平较低。*BroaSEP3* 基因在所有类型花的发育过程中均高丰度表达, 特别是在花发育的第 1、2 时期, 相比较而言, 在多瓣型花中的表达水平最高。但是这 3 个基因在根、茎中几乎无表达, 在叶中微量表达, 说明羽衣甘蓝 *SEP* 类是花发育特异性基因。



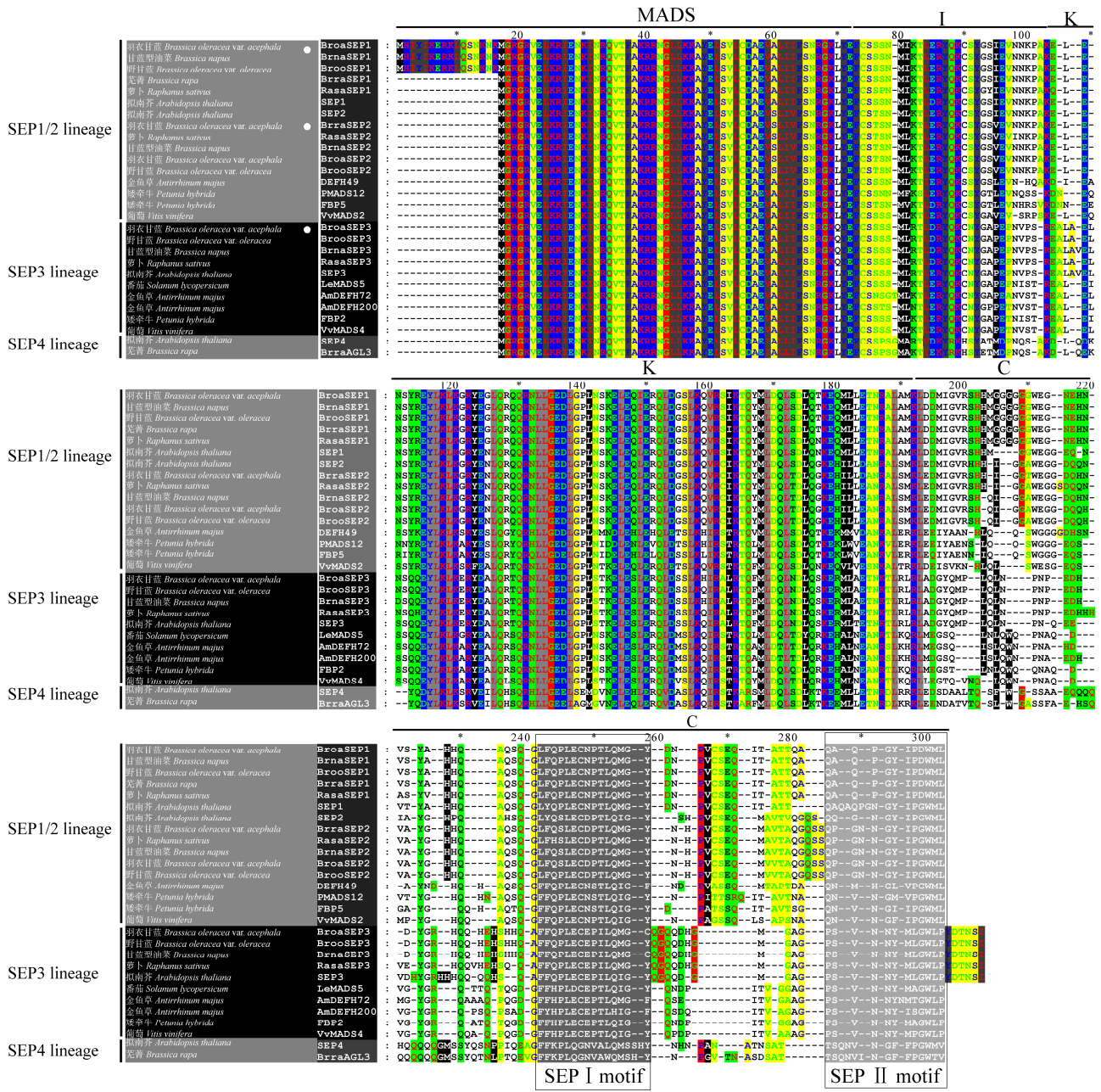


图5 部分 SEP-like 蛋白质氨基酸序列比对

Fig. 5 Sequence alignment of amino acid sequences among several SEP-like proteins. SEP I motif and SEP II motif are boxed. M: MADS DNA-binding domain. I: intervening domain. K: keratin-like domain. C: C-terminal domain.

### 3 讨论

#### 3.1 羽衣甘蓝 *BroaSEP1/2* 和 *BroaSEP3* 基因分别属于 E 功能 MADS-box 基因家族中的 *LOSEP* 和 *SEP3* 进化系

MADS-box 转录因子在被子植物花发育过程

中具有关键调控作用，而且在被子植物进化史中这种重要的同源功能是高度保守的。系统发育分析表明，调控花发育的 E 类 MADS-box 基因在被子植物进化过程中，发生了多次基因重复事件。第一次重复事件发生在现存被子植物起源之前，

产生了 *SEP3* (以前命名为 *AGL9*) 和 *LOFSEP* (以前命名为 *AGL2/3/4*) 进化系<sup>[32]</sup>; 在基部核心真双子叶植物出现之前, *LOFSEP* 进化系经历两次基因重复事件, 产生 *SEP1/2*、*FBP9/23* 和 *SEP4* 进化系<sup>[51]</sup>。 *SEP1* 和 *SEP2* 则是通过最近的基因组复

增而产生的<sup>[2]</sup>。这些旁系同源基因通过编码序列和 (或) 调控元件的变化而发生新功能化和亚功能化, 引发花形态结构的多样性<sup>[2-3]</sup>。E 类功能基因每个进化系编码的 MADS 蛋白在其 C 末端具有特异保守基序 SEP I 和 SEP II 基序 (图 5)。*SEP*

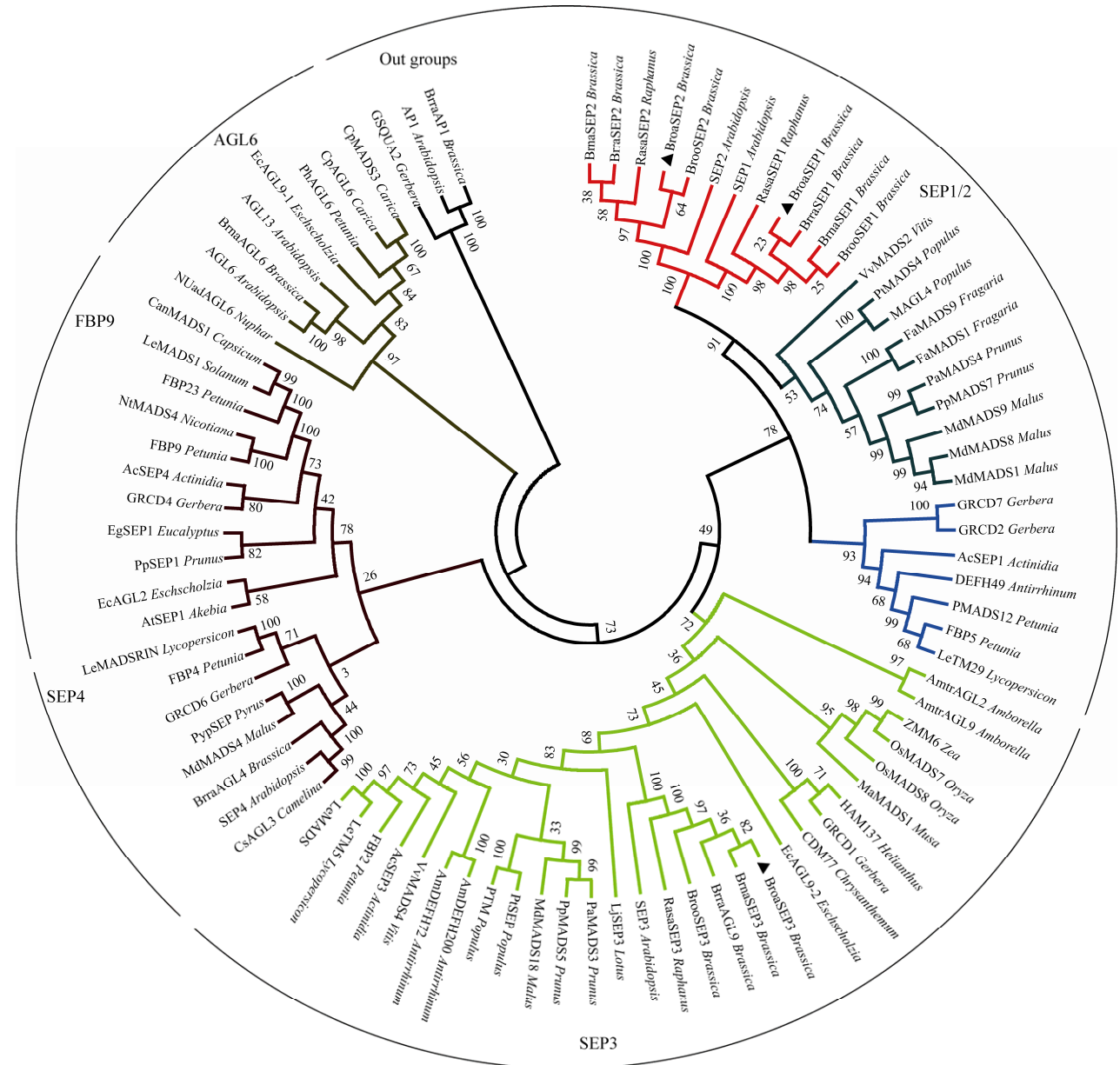


图 6 以邻位相连法构建的羽衣甘蓝 E 功能 MADS-box 蛋白系统发育树

Fig. 6 The phylogenetic tree of kale MADS-box proteins of class E generated using the Neighbor-joining method. The numbers next to the nodes indicate bootstrap values of 50% or more support from 1 000 replicates. Three SEP genes in kale is shown by black triangles.

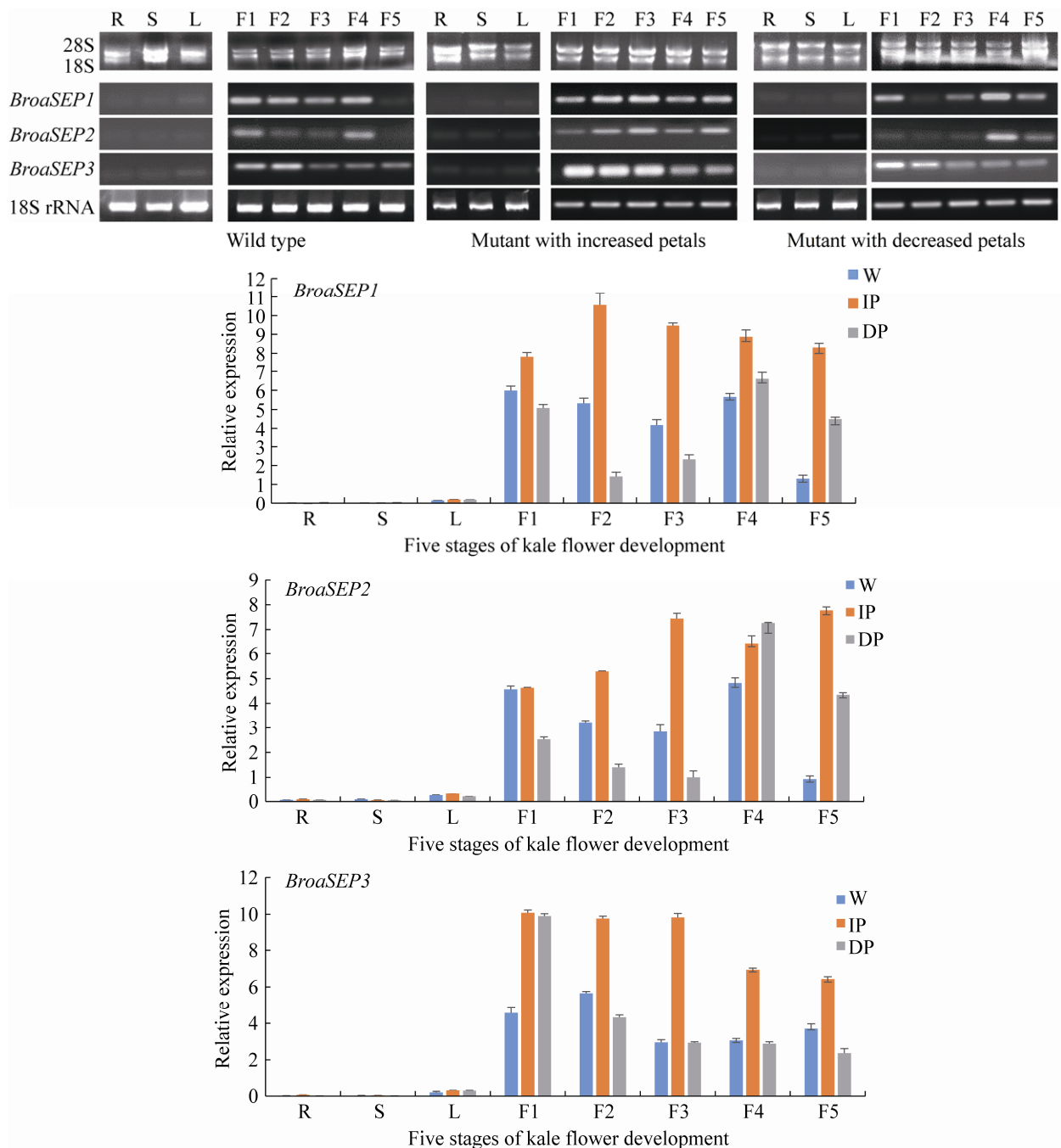


图7 羽衣甘蓝 *BroaSEP1/2/3* 基因半定量 RT-PCR 及 qRT-PCR 结果

Fig. 7 Quantification of expression levels of the *BroaSEP1/2/3* gene in different developmental stages of kale flower as determined by gene-specific semi-quantitative RT-PCR and qRT-PCR.

家族基因的数量、表达模式和功能因物种不同而有所差异,但其编码的蛋白质可以与其他 MADS-box 基因蛋白质形成复合体,与靶基因的 CArG-box (CCArichGG, 保守序列为 5'-CC(A/T)<sub>6</sub>GG-3')

结合而调控靶基因表达<sup>[52-53]</sup>。

本研究分离鉴定的羽衣甘蓝 *BroaSEP1/2/3* 属于 E 类 MADS-box 基因,其中 *BroaSEP1/2* 属于 *SEP1/2* 亚家族、*BroaSEP3* 属于 *SEP3* 亚家族。编

码的氨基酸序列在 MADS 区具有保守的钙调蛋白依赖的蛋白激酶磷酸化位点 RQVTF<sup>[49]</sup>, *SEP I* 和 *SEP II* 位于 C-末端区。*BroaSEP1/2/3* 蛋白与被子植物其他 E 类 MADS 转录因子具有很高的相似性, 其中与芸薹属植物 E 类 MADS 蛋白亲缘关系最近, 特别是与野甘蓝、芜菁、萝卜、甘蓝型油菜 E 类 MADS 蛋白序列的同源性较高 (96% 以上)、亲缘关系最近 (图 5-6)。研究结果证实羽衣甘蓝 *BroaSEP1/2*、*BroaSEP3* 基因分别是 *SEP1/2* 和 *SEP3* 的直系同源基因。

*SEP* 类基因家族在被子植物中具有丰富的多样性, 通过基因复制、可变剪辑而形成功能趋异、新功能、失功能、亚功能的旁系同源基因, 羽衣甘蓝的 *SEP* 类基因是否存在旁系同源基因? 还需通过花分化过程的特异组学的深入分析。

### 3.2 *BroaSEP1/2/3* 基因在花发育期的特异性表达模式

半定量和实时定量 RT-PCR 结果证实, *BroaSEP1/2/3* 在野生型、多瓣型和少瓣型花中特异性表达。但是 3 个基因在不同发育时期的表达水平存在差异, *BroaSEP1/2/3* 多瓣型花的 5 个发育时期表现为高表达, *BroaSEP1/2* 在少瓣型花的 2、3 发育期表现为弱表达, 可能与花瓣减少表型有关。*BroaSEP1/2/3* 在花发育的初期均高水平表达, 表明 E 类基因是花分生组织分化和花器官形成所必需的<sup>[54]</sup>。研究表明, 拟南芥<sup>[55]</sup>、矮牵牛<sup>[56]</sup>、番茄<sup>[57]</sup>、水稻<sup>[58]</sup>的 *SEP* 类基因异位表达会引起早花, 抑制 *SEP* 类基因如 *SEP1/2/3/4*、*FBP2/5* 和 *TM5/29*, 则产生多层萼片或叶状结构的花, 花序分生组织处于无限生长模式。除了调控开花时间和花器官形成, *SEP* 类基因如番茄的 *LeMADS-RIN* 和 *TAGL2* 基因<sup>[57]</sup>、苹果的 *SEP1/2* 同源基因<sup>[59]</sup> 在果实发育、烟草 *NsMADS3* 和 *NtMADS4*<sup>[60]</sup> 在营养生长等过程中发挥作用。拟南芥中的 *SEP1/SEP2/SEP4* 在花分生组织中起始表达, *SEP3* 主要在内三轮花器官原基中表达。随着花原基的

发育, *SEP1/SEP2* 基因在四轮器官中均表达, *SEP3* 仅在花瓣、雄蕊和雌蕊中表达, *SEP4* 主要在萼片中表达<sup>[18]</sup>。与拟南芥 *SEP* 基因在调控花器官特征属性方面存在功能冗余<sup>[18]</sup>不同, 蝴蝶兰 *Phalaenopsis amabilis* 的 *PeSEP3* 主要调控萼片和花瓣的形成<sup>[61]</sup>, 唐松草属 *Thalictrum thalictroides* 的 *SEP* 基因主要调控心皮属性、器官边界以及萼片花瓣化<sup>[39]</sup>, 非洲菊 *Gerbera hybrida* 的 *SEP1/2/4* 类基因 *GRCD1* 和 *GRCD2/7* 分别决定雄蕊和心皮的属性, 而 *SEP1* 类基因 *GRCD4* 和 *SEP3* 类基因 *GRCD5* 在决定花瓣属性方面存在功能冗余<sup>[62]</sup>, E 类基因还参与调控花分生组织属性<sup>[18,62]</sup>。

本研究结果表明, 羽衣甘蓝 *BroaSEP1/2/3* 在野生型、多瓣型、少瓣型花发育进程中的表达水平存在差异, 具有花型特异性。这种表达差异性可能是由于基因表观突变或者基因启动子区突变造成的, 需要进一步分析 3 种花型植株中这 3 个基因的 DNA 序列特别是启动子的表观修饰, 同时还需要深入分析是否存在 *BroaSEP1/2/3* 的旁系同源基因以及在萼片、花瓣、雄蕊和心皮中特异性表达模式, 并分析 E 类转录因子与其他 MADS-box 蛋白的互作关系。此外, 也需要解析 E 类 MADS 转录因子调控的靶基因, 以便综合阐释多瓣型、少瓣型花表型产生的分子基础。

## REFERENCES

- [1] Force A, Lynch M, Pickett FB, et al. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics*, 1999, 151(4): 1531–1545.
- [2] Moore RC, Grant SR, Purugganan MD. Molecular population genetics of redundant floral-regulatory genes in *Arabidopsis thaliana*. *Mol Biol Evol*, 2005, 22(1): 91–103.
- [3] Irish VF, Litt A. Flower development and evolution: gene duplication, diversification and redeployment. *Curr Opin Genet Dev*, 2005, 15(4): 454–460.
- [4] Kaufmann K, Melzer R, Theissen G. MIKC-type MADS-domain proteins: structural modularity,

- protein interactions and network evolution in land plants. *Gene*, 2005, 347(2): 183–198.
- [5] Airoidi CA, Davies B. Gene duplication and the evolution of plant MADS-box transcription factors. *J Genet Genomics*, 2012, 39(4): 157–165.
- [6] Pabón-Mora N, Di Stilio VS, Becker A. Editorial: genetic regulatory mechanisms underlying developmental shifts in plant evolution. *Front Plant Sci*, 2019, 10: 710.
- [7] Mandel MA, Gustafson-Brown C, Savidge B, et al. Molecular characterization of the *Arabidopsis* floral homeotic gene *APETALA1*. *Nature*, 1992, 360(6401): 273–277.
- [8] Jofuku KD, Den Boer BGW, Van Montagu M, et al. Control of *Arabidopsis* flower and seed development by the homeotic gene *APETALA2*. *Plant Cell*, 1994, 6(9): 1211–1225.
- [9] Jack T, Brockman LL, Meyerowitz EM. The homeotic gene *APETALA3* of *Arabidopsis thaliana* encodes a MADS box and is expressed in petals and stamens. *Cell*, 1992, 68(4): 683–697.
- [10] Goto K, Meyerowitz EM. Function and regulation of the *Arabidopsis* floral homeotic gene *PISTILLATA*. *Genes Dev*, 1994, 8(13): 1548–1560.
- [11] Yang YZ, Fanning L, Jack T. The K domain mediates heterodimerization of the *Arabidopsis* floral organ identity proteins, *APETALA3* and *PISTILLATA*. *Plant J*, 2003, 33(1): 47–59.
- [12] Yanofsky MF, Ma H, Bowman JL, et al. The protein encoded by the *Arabidopsis* homeotic gene *agamous* resembles transcription factors. *Nature*, 1990, 346(6279): 35–39.
- [13] Savidge B, Rounsley SD, Yanofsky MF. Temporal relationship between the transcription of two *Arabidopsis* MADS box genes and the floral organ identity genes. *Plant Cell*, 1995, 7(6): 721–733.
- [14] Pinyopich A, Ditta GS, Savidge B, et al. Assessing the redundancy of MADS-box genes during carpel and ovule development. *Nature*, 2003, 424(6944): 85–88.
- [15] Kaufmann K, Wellmer F, Muiño JM, et al. Orchestration of floral initiation by *APETALA1*. *Science*, 2010, 328(5974): 85–89.
- [16] Pelaz S, Ditta GS, Baumann E, et al. B and C floral organ identity functions require *SEPALLATA* MADS-box genes. *Nature*, 2000, 405(6783): 200–203.
- [17] Honma T, Goto K. Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature*, 2001, 409(6819): 525–529.
- [18] Ditta G, Pinyopich A, Robles P, et al. The *SEP4* gene of *Arabidopsis thaliana* functions in floral organ and meristem identity. *Curr Biol*, 2004, 14(21): 1935–1940.
- [19] Becker A, Theißen G. The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Mol Phylogenet Evol*, 2003, 29(3): 464–489.
- [20] Coen ES, Meyerowitz EM. The war of the whorls: genetic interactions controlling flower development. *Nature*, 1991, 353(6339): 31–37.
- [21] Purugganan MD, Rounsley SD, Schmidt RJ, et al. Molecular evolution of flower development: Diversification of the plant MADS-box regulatory gene family. *Genetics*, 1995, 140(1): 345–356.
- [22] Rounsley SD, Ditta GS, Yanofsky MF. Diverse roles for MADS box genes in *Arabidopsis* development. *Plant Cell*, 1995, 7(8): 1259–1269.
- [23] Theißen G, Saedler H. MADS-box genes in plant ontogeny and phylogeny: Haeckel's 'biogenetic law' revisited. *Curr Opin Genet Dev*, 1995, 5(5): 628–639.
- [24] Theissen G, Becker A, Di Rosa A, et al. A short history of MADS-box genes in plants. *Plant Mol Biol*, 2000, 42(1): 115–149.
- [25] Theißen G, Saedler H. Plant biology: Floral quartets. *Nature*, 2001, 409(6819): 469–471.
- [26] Litt A, Kramer EM. The ABC model and the diversification of floral organ identity. *Semin Cell Dev Biol*, 2010, 21(1): 129–137.
- [27] Otani M, Sharifi A, Kubota S, et al. Suppression of B function strongly supports the modified ABCE model in *Tricyrtis* sp. (Liliaceae). *Sci Rep*, 2016, 6: 24549.
- [28] Smaczniak C, Immink RG, Muiño JM, et al. Characterization of MADS-domain transcription factor complexes in *Arabidopsis* flower development. *Proc Natl Acad Sci USA*, 2012, 109(5): 1560–1565.
- [29] Wuest SE, O'Maoileidigh DS, Rae L, et al.



- Molecular basis for the specification of floral organs by APETALA3 and PISTILLATA. *Proc Natl Acad Sci USA*, 2012, 109(33): 13452–13457.
- [30] Rümpler F, Theißen G, Melzer R. A conserved leucine zipper-like motif accounts for strong tetramerization capabilities of *SEPALLATA*-like MADS-domain transcription factors. *J Exp Bot*, 2018, 69(8): 1943–1954.
- [31] Wellmer F, Graciet E, Riechmann JL. Specification of floral organs in *Arabidopsis*. *J Exp Bot*, 2014, 65(1): 1–9.
- [32] Zahn LM, Kong HZ, Leebens-Mack JH, et al. The evolution of the *SEPALLATA* subfamily of MADS-box genes: a preangiosperm origin with multiple duplications throughout angiosperm history. *Genetics*, 2005, 169(4): 2209–2223.
- [33] Malcomber ST, Kellogg EA. *SEPALLATA* gene diversification: brave new whorls. *Trends Plant Sci*, 2005, 10(9): 427–435.
- [34] Chen F, Zhang XT, Liu X, et al. Evolutionary analysis of MIKC<sup>c</sup>-type MADS-box genes in gymnosperms and angiosperms. *Front Plant Sci*, 2017, 8: 895.
- [35] Agrawal GK, Abe K, Yamazaki M, et al. Conservation of the E-function for floral organ identity in rice revealed by the analysis of tissue culture-induced loss-of-function mutants of the *OsMADS1* gene. *Plant Mol Biol*, 2005, 59(1): 125–135.
- [36] Cui RF, Han JK, Zhao SZ, et al. Functional conservation and diversification of class E floral homeotic genes in rice (*Oryza sativa*). *Plant J*, 2010, 61(5): 767–781.
- [37] Kobayashi K, Maekawa M, Miyao A, et al. *PANICLE PHYTOMER2* (*PAP2*), encoding a *SEPALLATA* subfamily MADS-box protein, positively controls spikelet meristem identity in rice. *Plant Cell Physiol*, 2010, 51(1): 47–57.
- [38] Yockteng R, Almeida AMR, Morioka K, et al. Molecular evolution and patterns of duplication in the *SEP/AGL6*-Like lineage of the zingiberales: a proposed mechanism for floral diversification. *Mol Biol Evol*, 2013, 30(11): 2401–2422.
- [39] Soza VL, Snelson CD, Hazelton KDH, et al. Partial redundancy and functional specialization of E-class *SEPALLATA* genes in an early-diverging eudicot. *Dev Biol*, 2016, 419(1): 143–155.
- [40] Flanagan CA, Ma H. Spatially and temporally regulated expression of the MADS-box gene *AGL2* in wild-type and mutant arabidopsis flowers. *Plant Mol Biol*, 1994, 26(2): 581–595.
- [41] Mandel MA, Yanofsky MF. The *Arabidopsis* *AGL9* MADS box gene is expressed in young flower primordia. *Sex Plant Reprod*, 1998, 11(1): 22–28.
- [42] Smaczniak C, Immink RGH, Angenent GC, et al. Developmental and evolutionary diversity of plant MADS-domain factors: insights from recent studies. *Development*, 2012, 139(17): 3081–3098.
- [43] Theißen G, Melzer R, Rümpler F. MADS-domain transcription factors and the floral quartet model of flower development: linking plant development and evolution. *Development*, 2016, 143(18): 3259–3271.
- [44] Lai XL, Daher H, Galien A, et al. Structural basis for plant MADS transcription factor oligomerization. *Comput Struct Biotechnol J*, 2019, 17: 946–953.
- [45] Krizek BA, Meyerowitz EM. Mapping the protein regions responsible for the functional specificities of the *Arabidopsis* MADS domain organ-identity proteins. *Proc Natl Acad Sci USA*, 1996, 93(9): 4063–4070.
- [46] van Dijk ADJ, Morabito G, Fiers M, et al. Sequence motifs in MADS transcription factors responsible for specificity and diversification of protein-protein interaction. *PLoS Comput Biol*, 2010, 6(11): e1001017.
- [47] Vandereyken K, van Leene J, de Coninck B, et al. Hub protein controversy: taking a closer look at plant stress response hubs. *Front Plant Sci*, 2018, 9: 694.
- [48] Zhang XL, Shan XZ, Jiang HM, et al. Research progress and prospect of Wild Cabbage. *Acta Horticult Sin*, 2018, 45(9): 1715–1726 (in Chinese). 张小丽, 单晓政, 江汉民, 等. 野生甘蓝的研究进展及展望. *园艺学报*, 2018, 45(9): 1715–1726.
- [49] Kanno A, Hienuki H, Ito T, et al. The structure and expression of *SEPALLATA*-like genes in *Asparagus* species (Asparagaceae). *Sex Plant Reprod*, 2006, 19(3): 133–144.
- [50] Zhou YZ, Xu ZD, Yong X, et al. *SEP*-class genes in *Prunus mume* and their likely role in floral organ

- development. *BMC Plant Biol*, 2017, 17: 10.
- [51] Morel P, Chambrier P, Boltz V, et al. Divergent functional diversification patterns in the SEP/AGL6/AP1 MADS-box transcription factor superclade. *Plant Cell*, 2019, 31(12): 3033–3056.
- [52] Melzer R, Verelst W, Theißen G. The class E floral homeotic protein SEPALLATA3 is sufficient to loop DNA in ‘floral quartet’-like complexes *in vitro*. *Nucleic Acids Res*, 2009, 37(1): 144–157.
- [53] Jetha K, Theißen G, Melzer R. *Arabidopsis* SEPALLATA proteins differ in cooperative DNA-binding during the formation of floral quartet-like complexes. *Nucleic Acids Res*, 2014, 42(17): 10927–10942.
- [54] Kobayashi K, Yasuno N, Sato Y, et al. Inflorescence meristem identity in rice is specified by overlapping functions of three *API/FUL*-Like MADS box genes and *PAP2*, a *SEPALLATA* MADS box gene. *Plant Cell*, 2012, 24(5): 1848–1859.
- [55] Tzeng TY, Hsiao CC, Chi PJ, et al. Two lily *SEPALLATA*-like genes cause different effects on floral formation and floral transition in *Arabidopsis*. *Plant Physiol*, 2003, 133(3): 1091–1101.
- [56] Ferrario S, Immink RGH, Shchennikova A, et al. The MADS box gene *FBP2* is required for *SEPALLATA* function in petunia. *Plant Cell*, 2003, 15(4): 914–925.
- [57] Ampomah-Dwamena C, Morris BA, Sutherland P, et al. Down-regulation of *TM29*, a tomato *SEPALLATA* homolog, causes parthenocarpic fruit development and floral reversion. *Plant Physiol*, 2002, 130(2): 605–617.
- [58] Chung YY, Kim SR, Finkel D, et al. Early flowering and reduced apical dominance result from ectopic expression of a rice MADS box gene. *Plant Mol Biol*, 1994, 26(2): 657–665.
- [59] Vrebalov J, Ruezinsky D, Padmanabhan V, et al. A MADS-box gene necessary for fruit ripening at the tomato *ripening-inhibitor (rin)* locus. *Science*, 2001, 296(5566): 343–346.
- [60] Jang S, An K, Lee S, et al. Characterization of tobacco MADS-box genes involved in floral initiation. *Plant Cell Physiol*, 2002, 43(2): 230–238.
- [61] Pan ZJ, Chen YY, Du JS, et al. Flower development of Phalaenopsis orchid involves functionally divergent *SEPALLATA*-like genes. *New Phytol*, 2014, 202(3): 1024–1042.
- [62] Zhang T, Zhao YF, Juntheikki I, et al. Dissecting functions of *SEPALLATA*-like MADS box genes in patterning of the pseudanthial inflorescence of *Gerbera hybrida*. *New Phytol*, 2017, 216(3): 939–954.

(本文责编 郝丽芳)