

# 黄瓜 MADS-box 基因家族生物信息学分析及 *CsMADS-box AGL62* 的克隆

赵阳 周秀艳 辛明 秦智伟\*

(农业农村部东北地区园艺作物生物学与种质资源创制重点实验室, 东北农业大学园艺园林学院, 黑龙江哈尔滨 150030)

**摘要:** 分别以全雌系 D1104-2-4、强雌系 D1158 ♀ -2、雌雄异花同株 D0528-2 为母本, 与 1 个雌雄异花同株父本 D0708-2 杂交获得杂种一代, 调查开花初期、根瓜期、盛果期 3 个时期雌性相关农艺性状, 筛选出差异显著的组合进行转录组测序分析进而分析黄瓜杂交组合间 MADS-box 家族的差异基因。结果表明: 以全雌系 D1104-2-4 为母本的杂种一代 C15-114 在雌花数、雌花节率上均具有超亲优势, 3 个杂交组合  $F_1$  均具有早熟的中亲优势。转录组分析结果发现, 以全雌系 D1104-2-4 为母本的杂种一代基因表达情况显著的偏向母本, 挑选的 37 个黄瓜 MADS-box 家族基因只有 6 个与母本有表达差异, 17 个与父本有表达差异。生物信息学分析并命名了 1 个差异最显著的基因 *CsMADS-box AGL62*, 该基因全长 745 bp, 编码 187 个氨基酸, 编码的蛋白属于线粒体中不稳定的亲水性跨膜蛋白, NCBI 比对结果表明该蛋白是 1 个转录因子, 系统进化分析发现其与甜瓜中 AGL62 同源性最高, 在黄瓜中该蛋白和其他黄瓜 AGL62 亚型属于同一分支并且属于 I 型。

**关键词:** 雌性分化; MADS-box 家族; *CsMADS-box AGL62* 基因; 生物信息学分析

黄瓜是我国保护地栽培的第一大蔬菜作物, 2015 年全国栽培面积 125.8 万  $hm^2$  (沈辰等, 2017)。黄瓜的性别分化、花期的早晚对黄瓜的产量及上市的早晚有着不可忽视的影响, 这些特点决定了黄瓜品种的经济价值。目前, 黄瓜性别分化、花期及果实发育成为了研究的热点。MADS-box 基因是一类广泛存在于植物中序列特异的同源异型基因, 参与调控花和果实发育及成熟的多个过程。MADS-box 基因作为转录因子含有一段高度保守的 DNA 结合结构域, 通过该结构域 MADS-box 基因识别并结合在下游目标基因的特定靶位点, 从而调控下游目标基因的表达 (张则婷和李学宝, 2007)。植物含有众多 MADS-box 基因, 如拟南芥有 100 多个 MADS-box 基因 (De Bodt et al., 2003;

Parenicová et al., 2003), 它们形成一个庞大的基因家族。根据系统进化关系、基因结构和蛋白结构的不同, MADS-box 基因分为 I 型 (Type I) 和 II 型 (Type II) 两大类型 (Becker & Theissen, 2003)。Type I 可进一步分为  $M\alpha$ 、 $M\beta$ 、 $M\gamma$  和  $M\delta$ ; Type II 可进一步分为 MIKCC 和 MIKC\* (Henschel et al., 2002)。植物 I 型 MADS-box 基因研究较浅, 已有的研究表明, 该类型的基因主要参与雌配子体、胚胎和种子的发育进程, 并与不同物种间生殖器官分界的决定有关 (Masiero et al., 2011)。植物 II 型 MADS-box 基因研究比较深入, 该类型的基因大多与花形态形成有关, 如已知的 ABCDE 模型。黄瓜中共发现 43 种 MADS-box 基因。已有的研究表明黄瓜中不含有 FLC、AGL12 和 Bs 3 种亚型的 MADS-box 基因, 并且除 GL 亚型外, 黄瓜其他亚型的 MADS-box 基因与拟南芥等其他植物中基因具有相似的功能 (Gan et al., 2014)。

本试验在对黄瓜几组不同花期的杂交组合田间花期调查的基础上, 由转录组测序并通过荧光定量验证的方法获得了 1 个黄瓜 MADS-box 家族的基

赵阳, 男, 硕士研究生, 专业方向: 黄瓜分子遗传育种, E-mail: 1254200748@qq.com

\* 通讯作者 (Corresponding author): 秦智伟, 男, 教授, 博士生导师, 专业方向: 黄瓜分子遗传育种, E-mail: qzw303@126.com

收稿日期: 2018-06-25; 接受日期: 2018-10-14

基金项目: 国家重点研发计划项目子课题 (2016YFD0101705)

因,通过生物信息学分析的手段来预测该基因的功能,以期为进一步基因功能验证奠定基础。

## 1 材料与方法

### 1.1 试验材料

以黄瓜自交系 D0708-2 为父本(雌雄异花同株),自交系 D1158 ♀-2(强雌系)、D1104-2-4(全雌系)、D0528-2(雌雄异花同株)为母本配制杂交组合。F<sub>1</sub> 分别为东农 812(D1158 ♀-2 × D0708-2)、C15-114(D1104-2-4 × D0708-2)、C17-31(D0528-2 × D0708-2)。父母本及 F<sub>1</sub> 均由东北农业大学园艺园林学院黄瓜课题组提供,采用穴盘育苗并且于 2017 年 7 月幼苗两叶一心时定植于东北农业大学园艺站塑料大棚。每小区 2 行,每行 14 株。随机区组设计,3 次重复。

逆转录酶、DNA 聚合酶、dNTP、T/A 克隆载体均购自 TaKaRa 公司,TRIzol 购自 Invitrogen 公司,质粒小量提取试剂盒、DNA 凝胶回收试剂盒和荧光定量染料均购自北京金全生物技术有限公司。

### 1.2 试验方法

#### 1.2.1 田间性状调查及转录组测序

于黄瓜开花初期、根瓜期、盛果期时每小区随机抽取 3 株进行定株调查,分别考察各组合不同时期的雌花数、雌花节率、第 1 雌花节位及最高发育节位,最高发育节位指最顶端开放的花(不分雌花雄花)所在位置距离茎尖的节位。开花初期指黄瓜单株至少 3 朵花开放,小区内至少 50% 植株开花的时期;根瓜期指有 50% 黄瓜植株开始采收商品瓜的时期;盛果期指黄瓜单株商品瓜产量达到最大的时期(路洪凤,2014)。试验数据采用 Excel 2010 进行整理,DPS 7.05 软件进行统计分析。

通过田间试验结果挑选 1 个组合进行转录组测序。测序及分析交由华大基因(深圳,中国)使用 Illumina HiSeq 平台完成,参照数据库为 NCBI 数据库(<https://www.ncbi.nlm.nih.gov/genome/?term=cucumber>)。1.2.2 引物设计与合成 使用软件 Primer Premier 5.0 (Premier Bio-soft International, Palo Alto, CA) 设计克隆基因全长引物 AGL62,并通过 <http://biodb.swu.sdu.cn/qprimerdb> 网站下载荧光定量引物 qAGL62,使用 *CsEF1α* 基因作为内参(Lu et al., 2018)。克隆引物及荧光定量引物见表 1。

表 1 引物序列

引物名称	退火温度/℃	序列	用途	
AGL62	55	F: 5'-ACAATGAAGAAATCATTGGCCGTC-3'	R: 5'-TCATGAAACAACATGTAATAAGAG-3'	克隆序列全长
qAGL62	55	F: 5'-GCTATTCTCGTCTTCTCTTCCA-3'	R: 5'-GACGGAGATAACAGAGACACA-3'	荧光定量基因引物
CsEF1α	55	F: 5'-CCAAGCAAGGTACGATGAAA-3'	R: 5'-AGAGATGGGAACGAAGGGGA-3'	荧光定量内参引物

1.2.3 *CsMADS-box AGL62* 的克隆及荧光定量 PCR 扩增体系为 20.0 μL,其中: 10 × PCR Buffer (含 20 mmol · L<sup>-1</sup> Mg<sup>2+</sup>) 2.0 μL, dNTPs (10 mmol · L<sup>-1</sup>) 2.0 μL, *Taq* 酶 (2U) 0.2 μL, cDNA 模板 2.0 μL, 上下游引物各 0.5 μL, 最后加 12.8 μL 的 ddH<sub>2</sub>O。PCR 反应程序为: 94 °C 预变性 5 min; 94 °C 变性 30 s, 55 °C 退火 1 min (表 1), 72 °C 延伸 1 min, 35 个循环; 72 °C 延伸 10 min; 4 °C 保存。PCR 产物经 1% 琼脂糖凝胶电泳分析后,回收目的片段,连接到 pGEM-T 克隆载体上,采用热激法转化到 Trans1-T1 感受态细胞中,在氨苄抗性平板上进行蓝白斑筛选,挑白色菌落摇菌后送至苏州金唯智生物科技有限公司测序。

荧光定量体系为 20.0 μL: 染料 SYBR Green PCR Master Mix 10.0 μL, 上下游引物各 0.5 μL,

cDNA 模板 2.0 μL, ddH<sub>2</sub>O 7.0 μL。荧光定量在 BIO-RAD iQ5 荧光定量 PCR 仪上进行 PCR 扩增,反应程序为: 95 °C 3 min; 95 °C 10 s, 55 °C 30 s, 72 °C 30 s, 40 个循环; 72 °C 延伸步骤收集荧光信号,每个样品设置 4 个技术性重复,3 个生物学重复。

采用 2<sup>-ΔΔCT</sup> 相对定量分析方法计算出基因的相对表达量,并用 DPS 7.05 数据处理系统软件进行方差及显著性分析。

1.2.4 *CsMADS-box AGL62* 的生物信息学分析 通过 NCBI 的 BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) 对基因进行保守结构域的分析;利用 MEGA 5.2 对基因进行建树分析。利用 ProtParam (<http://web.expasy.org/protparam/>) 计算蛋白质的理论分子量、理论等电点、原子组成、稳定系数和脂肪系数。利用蛋白质亚细胞定位工具 PSORT

(<http://psort.hgc.jp/form2.html>) 预测蛋白质的定位。通过 SignalP 4.1 (<http://www.cbs.dtu.dk/services/SignalP/>) 及 IPSORT (<http://ipsort.hgc.jp/>) 进行蛋白质信号肽预测。利用 TMHMM (<http://www.cbs.dtu.dk/services/TMHMM/>) 对蛋白质进行跨膜区域预测。通过 ProtScale (<http://web.expasy.org/protscale/>) 分析蛋白质的亲水性与疏水性。用 SOPMA (<http://sopma.expasy.org/>) 及 Phyre 2 (<http://sbj.bio.ic.ac.uk/phyre2/>) 对蛋白质的二级结构、三级结构进行预测。

## 2 结果与分析

### 2.1 田间调查结果

亲本及 F<sub>1</sub> 调查结果见表 2。从表中可以看出, 3 个时期的雌花数、雌花节率显示了 F<sub>1</sub> 能够很好地遗传亲本性状, 如 D1158 ♀ -2、D1104-2-4 为雌性系, 其杂种一代东农 812 和 C15-114 也为雌性系, 并且 C15-114 的雌性表现比母本 D1104-2-4 还要

突出(表 2)。D0528-2 为雌雄异花同株, 杂种一代 C17-31 也为雌雄异花同株。

根据杂种优势指数的计算方法:

$$OPH(\%) = \frac{F_1 - H_p}{H_p} \times 100$$

$$MPH(\%) = \frac{F_1 - M_p}{M_p} \times 100$$

$$M_p = \frac{P_1}{2} + \frac{P_2}{2}$$

其中 F<sub>1</sub> 表示杂种一代, P<sub>1</sub>、P<sub>2</sub> 表示亲本, H<sub>p</sub> 表示亲本中较优良的亲本, M<sub>p</sub> 表示双亲的中亲值, OPH 表示超亲优势, MPH 表示超中优势。

通过和非雌性系父本杂交, 以强雌系 D1158 ♀ -2 为母本的杂种一代东农 812 的雌花数和雌花节率均低于母本, 这在 3 个时期表现一致(图 1), 但由全雌系 D1104-2-4 作母本的 C15-114 在雌花数和雌花节率上产生了超亲优势(图 1-A),

表 2 3 个组合田间性状表现

材料	开花初期				根瓜期				盛果期			
	雌花数	雌花节率/%	第 1 雌花节位	最高发 育节位	雌花数	雌花节率/%	第 1 雌花节位	最高发 育节位	雌花数	雌花节率/%	第 1 雌花节位	最高发 育节位
D0708-2	1±1	8.70±9.24	3±3	4±1	3±1	13.95±4.35	10±5	3±1	5±2	18.07±7.01	10±5	3±1
东农 812	7±3	60.02±18.63	3±2	6±1	17±2	80.17±9.50	3±2	5±1	24±2	84.08±6.35	3±2	5±1
D1158 ♀ -2	9±1	74.53±6.86	4±1	8±1	20±1	87.05±2.28	4±1	9±1	27±1	90.04±1.64	4±1	9±1
C15-114	9±2	75.77±12.08	3±0	7±1	21±2	91.37±0.73	3±0	6±1	29±2	93.55±0.41	3±0	6±1
D1104-2-4	7±1	69.70±11.2	4±1	12±1	20±2	86.79±2.64	4±1	12±1	25±2	89.31±2.08	4±1	12±1
C17-31	1±1	7.00±5.04	3±4	6±1	3±1	13.14±6.84	10±3	5±1	5±1	18.79±4.52	10±3	5±1
D0528-2	1±1	10.19±6.77	5±4	6±1	1±1	5.95±2.98	5±4	5±1	3±1	14.69±5.88	7±4	5±1

表明利用全雌系可以获得在雌性上具有杂种优势的新品种。但是通过杂交笔者并没有获得具有早花超亲优势性状的杂种一代(图 1-A), 不过通过和具有早花性状的亲本杂交, F<sub>1</sub> 能够很好地综合亲本的优良性状兼具早熟和雌性的特点(图 1-B)。

### 2.2 黄瓜 MADS-box 基因家族测序结果分析

根据田间调查结果, 以 D1104-2-4 为母本, D0708-2 为父本配制的 C15-114 组合在雌花数、雌花节率上具有超亲优势, 这一特点有利于研究黄瓜 MADS-box 基因家族在雌性杂种优势形成中的作用, 因此挑选该组合进行转录组测序。

测序结果显示(表 3), 黄瓜 37 个 MADS-box 家族基因中大部分基因在父母本及子代中无差

异表达。C15-114 和母本 D1104-2-4 之间 MADS-box 家族基因差异倍数在 -3.20~7.04, 其中 31 个基因无差异表达, 只有 6 个有差异表达, 其中 1 个基因差异显著(差异倍数>4)。而 C15-114 和父本 D0708-2 之间 MADS-box 家族基因差异倍数在 -5.83~5.55, 其中有 17 个基因差异表达, 20 个无差异表达(表 3)。根据 C15-114、D1104-2-4 及 D0708-2 的性状表现可知, C15-114、D1104-2-4 属于全雌系而 D0708-2 属于雌雄异花同株, 转录组的测序结果进一步体现了杂种一代的雌性偏向。

对黄瓜 MADS-box 家族基因进行聚类分析发现, 差异表达的基因富集在相似的类别中, 上调差异表达的基因和下调差异表达的基因分别聚集在 2

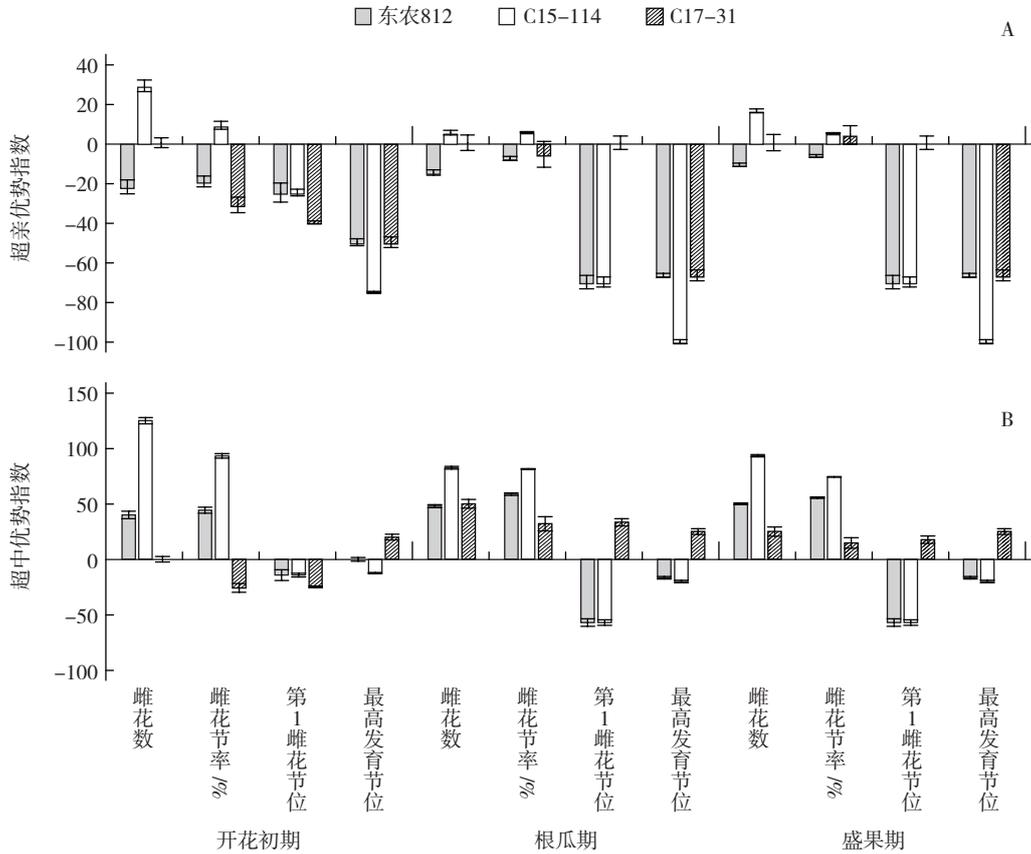


图1 3个F<sub>1</sub>的超亲及超中优势

A: 3个F<sub>1</sub>的超亲优势, 正值表示具有超亲优势, 负值表示没有超亲优势。B: 3个F<sub>1</sub>的超中优势, 正值表示具有超过亲本平均值的优势, 负值表示优势低于亲本的平均值。彩色图版见《中国蔬菜》网站: www.cnveg.org, 下图同。

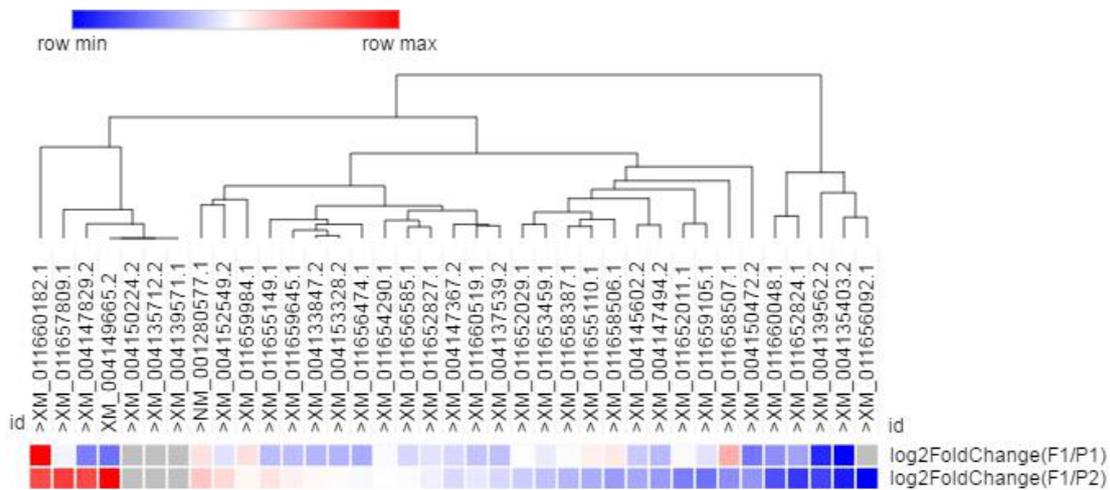


图2 黄瓜 MADS-box 家族基因转录组测序热图聚类分析

组不同的分支中 (图 2)。这表明杂种一代与亲本雌性产生差异的结果是由于部分基因差异表达。为了进一步分析差异基因特点, 对黄瓜 37 个 MADS-box 家族基因构建系统发育树并通过比对到黄瓜基因组数据库 (<http://www.icugi.org/>) 分析各个基因的

外显子分布 (图 3)。这些 MADS-box 家族基因的 2 种类型主要集中在 2 个区域 (图 4)。说明黄瓜中该家族基因进化上具有很大的保守性。通过转录组测序得到的差异较大且在 2 个亲本中差异表达相反的 2 个基因 ID: 101215235 和 ID: 101206073 均属于

Type I (图 3、4)。ID: 101215235 有 1 个外显子区域, ID: 101206073 有 2 个外显子区域, 二者都属于 AGL62 型。有研究表明, AGL62 的基因功能与 AGL61 功能类似, AGL61 可以与 AGL80 协作, 共同参与雌配子体内中心细胞的分化(Bemer et al., 2008), 并且 AGL62 同时还参与胚乳细胞的形成(Kang et al., 2008), 这些都与植物的雌性分化密切相关。综上分析表明: 引起杂交种 C15-114 与亲本间雌性差异及花期差异的原因很可能是这 2 个基因差异表达的结果。

为验证这一结论的准确性, 挑选 ID: 101215235 基因在 3 个组合中进行荧光定量验证分析(图 5)。将每个组合的杂种一代作为参照, C15-114 组合中父母本与子代均有差异, 母本表现为近 5 倍的差异表达量, 父本表现为近 3 倍的差异表达量, 荧光定量结果和转录组测序结果基本一致(表 3, 图 5)。分析发现 2 个雌性系组合父母本与杂种一代均有显著差异表达, 而非雌性系 C17-31 组合中父母本与杂种一代差异不显著。表明基因 ID: 101215235 在杂交组合中的差异表达可能参与黄瓜的雌性分化。

### 2.3 *CsMADS-box AGL62* 基因克隆及分析

从黄瓜茎尖提取总 RNA 后, 用琼脂糖凝胶电泳检测(图 6-A)。结果显示 RNA 提取良好。进行 PCR 克隆(图 6-B)后在 745 bp 左右出现条带, 回收后与 pGEM-T 克隆载体相连转化到 Trans1-T1 感受态细胞中, 挑白色菌落摇菌后测序。测序结果与 NCBI 上基因 ID: 101215235 序列一致, 表明克隆成功。

Protparam 分析结果显示基因 ID: 101215235 编码的蛋白含有 187 个氨基酸; 理论等电点为 9.18; 预测的分子量为 21.583 kDa; 分子式为  $C_{970}H_{1541}N_{267}O_{278}S_6$ ; 不稳定指数为 54.96, 属于不稳定蛋白(> 40 属于不稳定); 脂肪族指数为 84.49。

NCBI 比对结果表明该蛋白为 MADS-box 家族蛋白, 类别为 AGL62 (图 7), 并且具有 SRF-TF 结构域, 所以基因 ID: 101215235 还是 1 个转录因子, 把该基因命名为 *CsMADS-box AGL62*。PSORT 预测结果显示该蛋白可能为线粒体蛋白。

信号肽预测结果表明该蛋白没有明显的信号肽序列(图 8-A、B), 而 TMHMM 跨膜区域预测发现该蛋白在第

表 3 黄瓜 MADS-box 基因转录组测序结果

基因 ID	mRNA 名称	log2 值 (F <sub>1</sub> /P <sub>1</sub> )	上下调 Up/Down (F <sub>1</sub> /P <sub>1</sub> )	上下调 Up/Down (F <sub>1</sub> /P <sub>2</sub> )	log2 值 (F <sub>1</sub> /P <sub>2</sub> )
101215235	> XM_004149665.2	7.04	Up	Down	-1.36
105435626	> XM_011657809.1	0.80	*	*	5.55
101206073	> XM_004147829.2	-1.12	Down	Up	5.22
101208072	> NM_001280577.1	1.47	Up	Up	2.05
101213234	> XM_004152549.2	0.51	*	Up	1.60
101223132	> XM_011655149.1	-0.11	*	Up	1.29
101203152	> XM_011659645.1	-0.09	*	*	0.96
101215731	> XM_011659984.1	1.49	*	*	0.83
101209501	> XM_004133847.2	-0.20	*	*	0.77
101217704	> XM_004153328.2	-0.26	*	*	0.71
101203545	> XM_011654290.1	0.91	*	*	0.63
101212323	> XM_011656585.1	0.33	*	*	0.54
101220518	> XM_011656474.1	-0.38	*	*	0.42
101211542	> XM_011652827.1	0.53	*	*	0.24
101203217	> XM_011660519.1	0.03	*	*	0.03
101211025	> XM_004147367.2	0.35	*	*	-0.32
101220182	> XM_004137539.2	-0.05	*	*	-0.37
101207225	> XM_011652029.1	1.02	Up	*	-0.67
101218805	> XM_011653459.1	0.68	*	*	-0.95
101219421	> XM_011658387.1	0.94	*	Down	-1.25
101218674	> XM_011655110.1	1.25	Up	Down	-1.48
101210693	> XM_004145602.2	0.12	*	*	-1.58
101216798	> XM_011658506.1	1.34	Up	Down	-1.90
101203098	> XM_004147494.2	-0.15	*	Down	-1.91
101205930	> XM_004150472.2	-1.26	*	Down	-2.09
101216562	> XM_011658507.1	2.32	*	Down	-2.28
101204197	> XM_011652011.1	0.91	*	Down	-2.69
101220998	> XM_011659105.1	0.53	*	Down	-2.97
101203939	> XM_011660048.1	-0.87	*	Down	-3.84
101216951	> XM_004139562.2	-2.56	*	Down	-4.13
101212022	> XM_011652824.1	-0.52	*	Down	-4.46
101203859	> XM_004135403.2	-3.20	*	Down	-5.16
101222576	> XM_011656092.1	No	No	*	-5.83
101221957	> XM_004150224.2	No	No	No	No
101219796	> XM_004139571.1	No	No	No	No
101203059	> XM_004135712.2	No	No	No	No
101206453	> XM_011660182.1	No	No	No	No

注: P<sub>1</sub> 表示母本 D1104-2-4, P<sub>2</sub> 表示父本 D0708-2, F<sub>1</sub> 表示 C15-114。Up 表示上调, Down 表示下调, \* 表示有较大差异表达但不属于上调或下调表达, No 表示该基因在 F<sub>1</sub>/P<sub>1</sub> 和 F<sub>1</sub>/P<sub>2</sub> 中表达量极低。

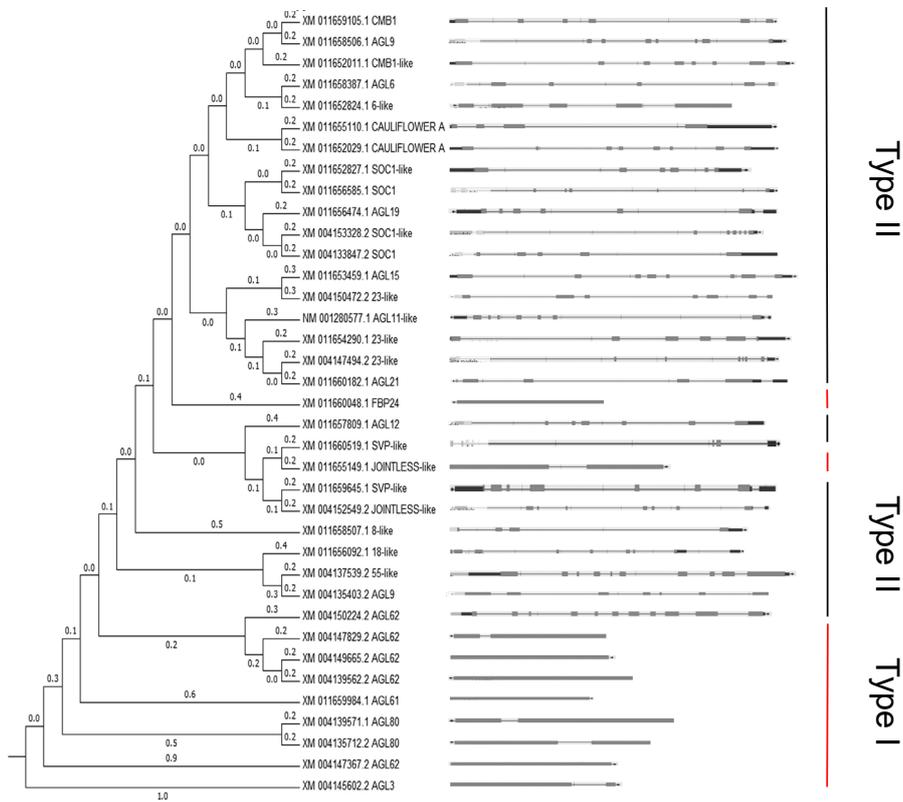


图3 黄瓜 MADS-box 家族基因发育树及外显子分布

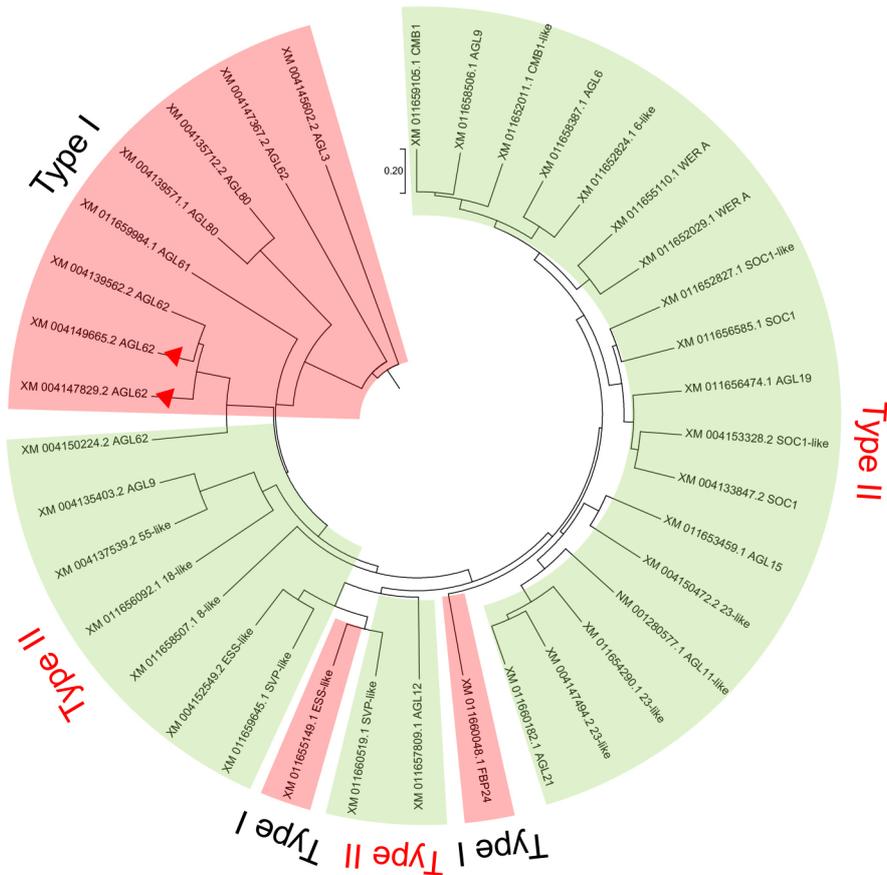


图4 黄瓜 MADS-box 家族基因分类

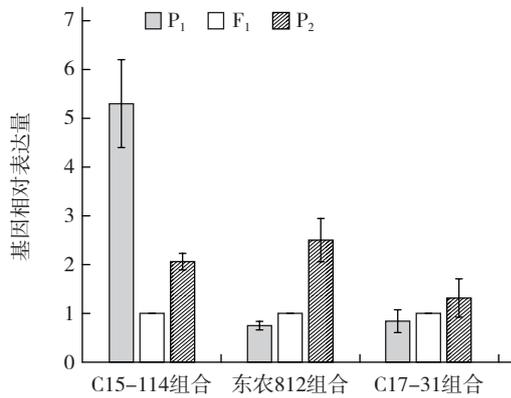


图5 ID: 101215235 基因在 3 个 F<sub>1</sub> 组合中的荧光定量结果  
P<sub>1</sub>, 母本; P<sub>2</sub>, 父本; F<sub>1</sub>, 杂种一代。

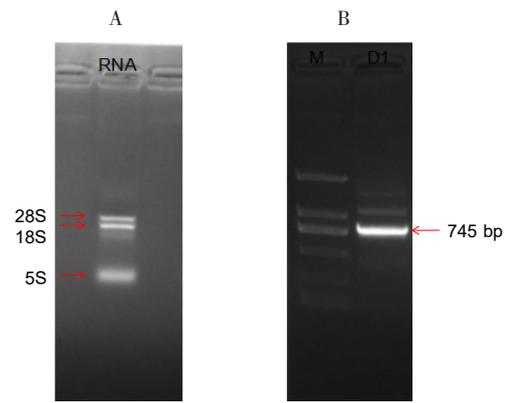


图6 RNA 提取及 PCR 电泳  
A: 黄瓜茎尖 RNA; B: *CsMADS-box AGL62* 基因克隆。

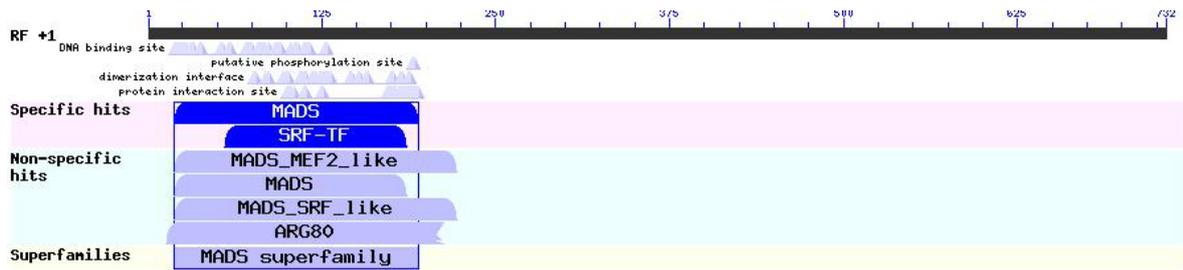


图7 *CsMADS-box AGL62* 基因保守结构域

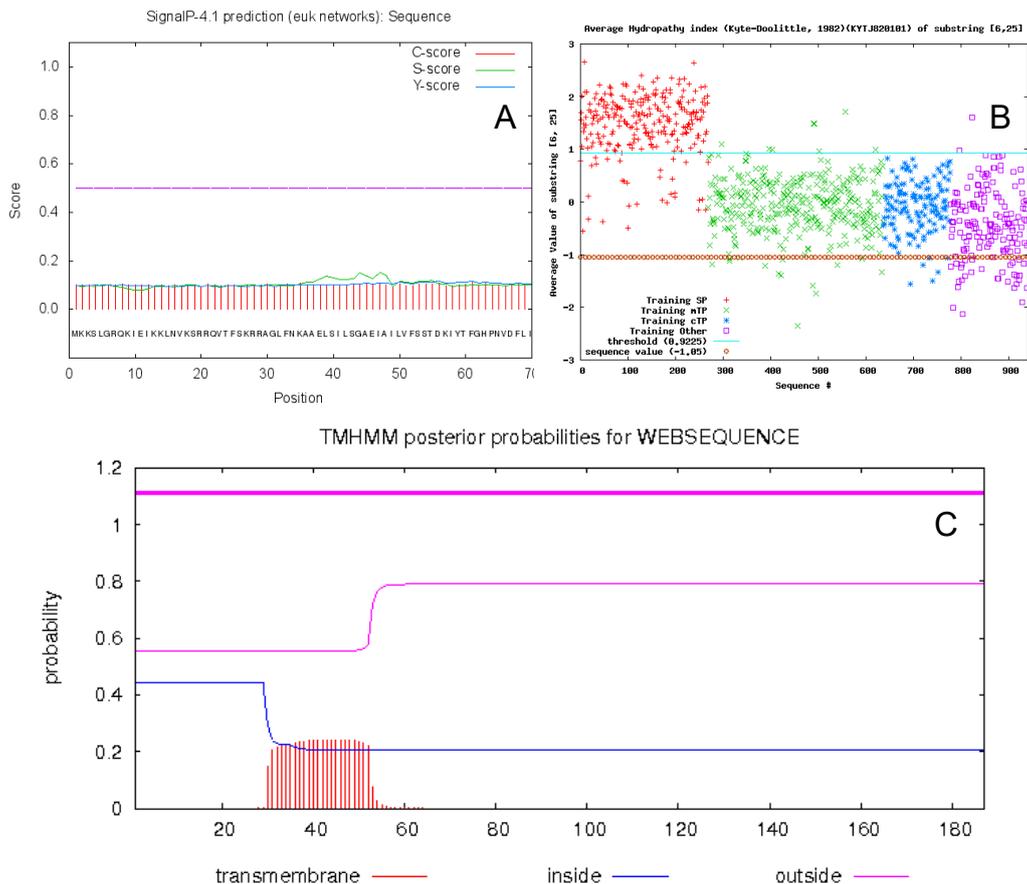


图8 *CsMADS-box AGL62* 蛋白信号肽及跨膜区域预测

30~50个氨基酸位置有跨膜，表明该蛋白是膜蛋白，且前30个氨基酸位于膜内，第50个氨基酸之后位于膜外（图8-C）。

对蛋白质的疏水性预测发现，主要的疏水位

点为第30~50个氨基酸（图9-A），该位点与跨膜预测结果中的跨膜位点一致。结合跨膜预测结果发现，氨基酸的膜外和膜内区域均为亲水性氨基酸，这些结果均与跨膜蛋白特征相似。蛋白质的二级

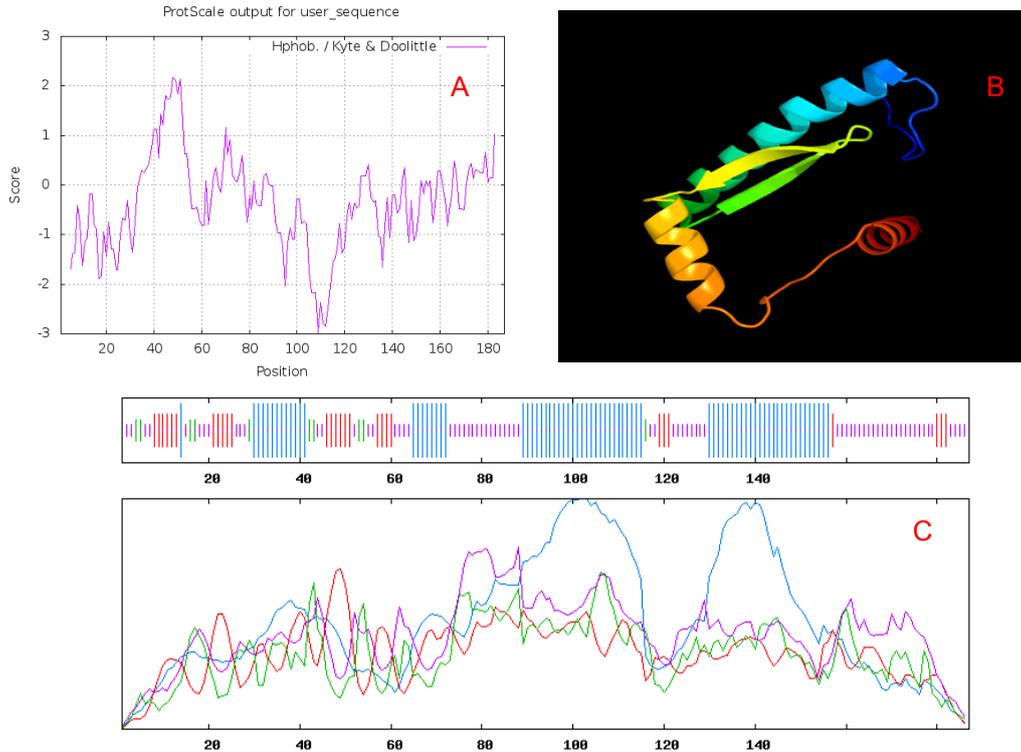


图9 CsMADS-box AGL62的疏水性及高级结构预测

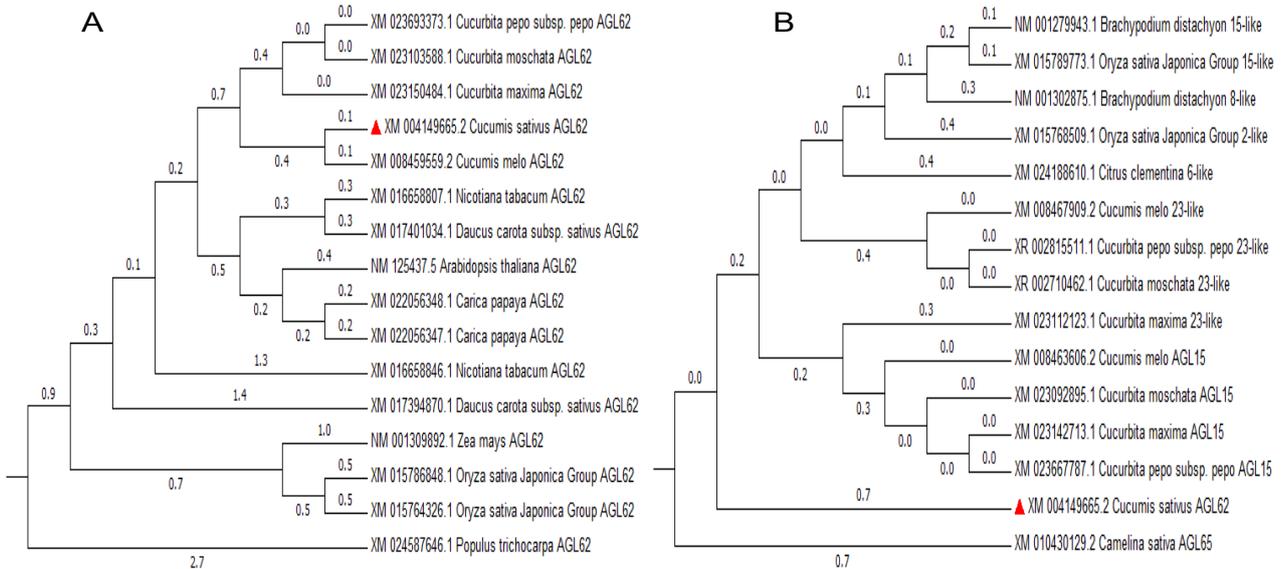


图10 CsMADS-box AGL62 系统进化分析

A: 不同物种 AGL62 基因发育树; B: 不同物种 MADS-box 家族基因发育树。

结构分析表明,该蛋白共有 76 个  $\alpha$  螺旋,占整个多肽链的 40.64%; 28 个延伸主链占整个多肽链的 14.97%; 10 个  $\beta$  折叠占整个多肽链的 5.35%; 73 个无规则卷曲占整个多肽链的 39.04% (图 9-C)。用 Phyre2 同源建模预测蛋白的三级结构如图 9-B 所示。

不同物种的 AGL62 亚型基因系统发育分析发现黄瓜与甜瓜的同源性最高 (图 10-A),不同物种不同亚型的 MADS-box 家族基因系统发育分析表明,同一个物种不同亚型的 MADS-box 家族基因进化程度更为相近 (图 10-B)。结合图 3 和图 4 分析, *CsMADS-box AGL62* 基因同一物种中相同的亚型具有最高的同源性,表明该基因在黄瓜进化中保守性很高。

### 3 讨论

本试验通过田间调查及转录组分析筛选出杂交种中与黄瓜雌花发育有关的差异表达基因,分析了 MADS-box 家族基因在黄瓜杂交种中的表达模式,并挑选 1 个该家族显著差异的基因 ID: 101215235 进行生物信息学分析,并命名为: *CsMADS-box AGL62*。

MADS-box 家族在植物中参与性别决定与花发育的机理已经有了深入的研究 (Lin et al., 2015; Dreni & Zhang, 2016; Zeng et al., 2018),但是该家族对植物雌性形成的杂种优势研究还很少。本试验发现 MADS-box 家族基因在杂交组合中表达模式显著偏向母本,这可能是  $F_1$  的雌性相关农艺性状与母本相似的原因。有研究发现 MADS-box 家族基因参与植物早花的形成 (de Oliveira et al., 2014),在本试验中发现所有的组合开花时间均介于母本与父本之间,并且更偏向父本,与雌花分化等性状偏向母本相反。MADS-box 家族基因在杂交种中的表达调控机理还有待进一步研究。

AGL62 基因属于 MADS-box 家族中 I 型基因,本试验在黄瓜中获得 1 个 *CsMADS-box AGL62* 基因。有研究表明,该基因调控胚乳的发育及细胞形成 (Hehenberger et al., 2012),生物信息学分析表明,黄瓜中该基因位于线粒体膜上,属于转录因子,这说明 *CsMADS-box AGL62* 基因可能在黄瓜中调控线粒体基因的表达。

### 参考文献

- 路洪凤. 2014. 黄瓜早花基因的 QTL 定位 [博士学位]. 北京: 中国农业科学院.
- 沈辰,熊露,韩书庆,马娟娟,吴建寨. 2017. 我国果菜类蔬菜生产与流通形势分析. 中国蔬菜, (9): 7-11.
- 张则婷,李学宝. 2007. Mads-box 基因在植物发育中的功能. 植物生理学报, 43 (2): 218-222.
- Becker A, Theissen G. 2003. The major clades of mads-box genes and their role in the development and evolution of flowering plants. *Molecular Phylogenetics & Evolution*, 29 (3): 464-489.
- Bemer M, Wolters-Arts M, Grossniklaus U, Angenent G C. 2008. The mads domain protein diana acts together with agamous-like80 to specify the central cell in *Arabidopsis ovules*. *Plant Cell*, 20 (8): 2088-2101.
- De Bodt S, Raes J, Florquin K, Rombauts S, Rouzé P, Theissen G, van de Peer Y. 2003. Genomewide structural annotation and evolutionary analysis of the type I MADS-box genes in plants. *Journal of Molecular Evolution*, 56 (5): 573-586.
- Dreni L, Zhang D. 2016. Flower development: the evolutionary history and functions of the agl6 subfamily MADS-box genes. *Journal of Experimental Botany*, 67 (6): 1625-1638.
- Gan D F, Jiao W U, Zhuang D, Liang D D, Zhan M D, Hao F L. 2014. Cloning and expression analysis of ap1-ful type mads-box gene in cucumber (*Cucumis sativus* L.). *Journal of Nuclear Agricultural Sciences*, 8: 1585-1590.
- Hehenberger E, Kradolfer D, Köhler C. 2012. Endosperm cellularization defines an important developmental transition for embryo development. *Development*, 139 (11): 2031-2039.
- Henschel K, Kofuji R, Hasebe M, Saedler H, Münster T, Theissen G. 2002. Two ancient classes of mick-type mads-box genes are present in the moss *Physcomitrella patens*. *Molecular Biology & Evolution*, 19 (6): 801-814.
- Kang I H, Steffen J G, Portereiko M F, Lloyd A, Drews G N. 2008. The agl62 mads domain protein regulates cellularization during endosperm development in *Arabidopsis*. *Plant Cell*, 20 (3): 635-647.
- Lin C S, Hsu C T, Liao D C, Chang W J, Chou M L, Huang Y T. 2015. Transcriptome-wide analysis of the MADS-box gene family in the orchid *Erycina pusilla*. *Plant Biotechnology Journal*, 14 (1): 284-298.
- Lu K, Li T, He J, Chang W, Zhang R, Liu M, Yu M, Fan Y, Ma J, Sun W, Qu C, Liu L, Li N, Liang Y, Wang R, Qian W, Tang Z, Xu X, Lei B, Zhang K, Li J. 2018. qprimerDB: a thermodynamics-based gene-specific qPCR primer database for 147 organisms. *Nucleic Acids Research*, 46 (Database issue): D1229 - D1236.
- Masiero S, Colombo L, Grini P E, Schnittger A, Kater M M. 2011. The emerging importance of type I MADS-box transcription factors

- for plant reproduction. *Plant Cell*, 23 (3): 865–872.
- Oliveira R R D, Cesarino I, Mazzafera P, Dornelas M C. 2014. Flower development in *Coffea arabica* L.: new insights into MADS-box genes. *Plant Reproduction*, 27 (2): 79–94.
- Parenicov  $\acute{a}$  L, de Folter S, Kieffer M, Horner D S, Favalli C, Busscher J, Cook H E, Ingram R M, Kater M M, Davies B, Angenent G C, Colombo L. 2003. Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in Arabidopsis: new openings to the MADS world. *Plant Cell*, 15: 1538–1551.
- Zeng X, Liu H, Du H, Wang S, Yang W, Chi Y. 2018. Soybean MADS-box gene *GmAGL1* promotes flowering via the photoperiod pathway. *Bmc Genomics*, 19 (1): 51–67.

## Bioinformatics Analysis of Cucumber MADS-box Gene Family and Cloning of *CsMADS-box AGL62*

ZHAO Yang, ZHOU Xiu-yan, XIN Ming, QIN Zhi-wei\*

[ Key Laboratory of Biology and Genetic Improvement of Horticultural Crops (Northeast Region), Ministry of Agriculture and Rural Affairs, School of Horticulture and Landscape Architecture, Northeast Agricultural University, Harbin 150030, Heilongjiang, China ]

**Abstract:** The gynoecious line ‘D1104–2–4’, pistillate line ‘D1158 ♀–2’ and monoecious line ‘D0528–2’ were used as female parents, and the hybrid generation was obtained by crossing with monoecious line ‘D0708–2’ as male parent. The paper investigated the female related agronomic characters at 3 periods including early blooming stage, fruiting early stage and fruiting period; then screened the significantly different combination to carry out trans-criptional sequence; and finally analyzed the differential genes of MADS-box family among cucumber hybrid combinations. The results showed that the  $F_1$  generation with ‘D1104–2–4’ as female parent had super parent advantage on the numbers of female flower and female flower node rate. And  $F_1$  of all 3 hybrid combinations had the early maturing mid-parent advantage. The results of transcriptional analysis showed that the  $F_1$  gene expression in the whole gynoecious line ‘D1104–2–4’ combination was significantly biased to the maternal parent, and only 6 of 37 cucumber MADS-box family genes were different from the female parent, and 17 of them were differ in expression with male parent. Bioinformatics analyzed and named a significantly differential express gene *CsMADS-box AGL62* that encodes 187 amino acids in the whole length of 745 bp. The encoded protein was belonged to an unstable hydrophilic transmembrane protein in the mitochondria. The NCBI alignment results showed that this protein was a transcription factor. Systematic analysis found this protein had the highest homology to AGL62 in muskmelon. In cucumber, this gene was belonged to the same branch with the other cucumber AGL62 subtypes and belonged to type I.

**Key words:** Female differentiation; MADS-box family; *CsMADS-box AGL62* gene; Bioinformatics analysis

## 本刊常用计量单位表示法

1. 时间:用 a (年)、d (天)、h (小时)、min (分)、s (秒)表示。
2. 面积:用  $\text{km}^2$  (平方千米)、 $\text{hm}^2$  (公顷)、 $\text{m}^2$  (平方米)、 $\text{cm}^2$  (平方厘米)表示,不用亩,可暂用  $667 \text{ m}^2$  代替。
3. 质量(原为重量):用 g (克)、kg (千克)、t (吨)表示。
4. 浓度:可用 % 表示质量分数和体积分数。质量浓度用  $\text{kg}\cdot\text{L}^{-1}$  (千克每升)、 $\text{g}\cdot\text{L}^{-1}$  (克每升)、 $\text{mg}\cdot\text{L}^{-1}$  (毫克每升)、 $\mu\text{g}\cdot\text{L}^{-1}$  (微克每升)表示。ppm 并非单位符号,不能使用,可根据具体情况改写成质量分数  $\text{mg}\cdot\text{kg}^{-1}$ 、体积分数  $\mu\text{L}\cdot\text{L}^{-1}$  或质量浓度  $\text{mg}\cdot\text{L}^{-1}$ ,数值保持不变。
5. 组合单位:
  - ① 组合单位中不能插入其他信息,如“VC 含量  $25 \text{ mg}/100 \text{ g}$  鲜重”,应为“VC 含量  $250 \text{ mg}\cdot\text{kg}^{-1}$  (鲜样质量)” ;“施肥量  $140 \text{ kg N}/\text{hm}^2$ ” 应为“施 N 肥量  $140 \text{ kg}\cdot\text{hm}^{-2}$ ”。
  - ② 组合单位书写错误,如“ $\text{mg}/\text{kg}\cdot\text{d}$ ”,应写为“ $\text{mg}\cdot\text{kg}^{-1}\cdot\text{d}^{-1}$ ”。