植物学报 *Chinese Bulletin of Botany* 2015, **50** (5): 614–622, www.chinbullbotany.com
doi: 10.11983/CBB14169

·技术方法·

基于MADS-box诱饵与蛋白质相互作用的拟南芥花瓣发育分子网络拓展

杨黎[1, 2], 孙丛苇[1, 2], 代志军[1, 2], 何淼[3*], 袁哲明[1, 2*]

[1]湖南农业大学湖南省作物种质创新与资源利用重点实验室, 长沙 410128
[2]湖南农业大学湖南省植物病虫害生物学及防控重点实验室, 长沙 410128; [3]中山大学生命科学学院, 广州 510275

摘要 阐明花器官发育调控机理具重要的进化、发育和生态学意义。该文以拟南芥(*Arabidopsis thaliana*)花瓣发育为例, 整合蛋白质互作、亚细胞定位、基因芯片和基因功能注释等数据库, 通过组建蛋白质互作可信预测模型, 获得拟南芥花瓣蛋白质互作网络, 以含有MADS-box结构域蛋白为诱饵在网络中进行一级拓展, 得到含38个蛋白质和67对互作的拓展网络。基于拓展网络, DAVID基因功能注释表明, 多数蛋白质涉及的生物学过程与花发育调控相关; 提取到19个候选四元互作, 涉及ABCDE模型基因之外的8个基因, 其中含MADS-box结构域的*AGL16*可能是B类基因新成员或其冗余; *SEU*、*LUH*、*CHR4*、*CHR11*、*CHR17*和*AT3G04960*为拟南芥花瓣AP1-AP3-PI-SEP四聚体的候选靶标基因。研究结果为深入解析拟南芥花瓣发育分子调控网络奠定了基础。

关键词 拟南芥, MADS-box, 花瓣发育, 蛋白质相互作用

杨黎, 孙丛苇, 代志军, 何淼, 袁哲明 (2015). 基于MADS-box诱饵与蛋白质相互作用的拟南芥花瓣发育分子网络拓展. 植物学报 **50**, 614–622.

拟南芥(*Arabidopsis thaliana*)是国际上第1个完成全基因组序列测定的高等植物, 其全基因组包含约26 000个基因, 编码2万多个蛋白质(The Arabidopsis Genome Initiative, 2000)。花是研究植物器官发育过程的良好实验系统, 阐明花器官发育调控机理具重要意义。在基因层面, 控制拟南芥花器官发育的调控基因已从经典的ABC模型发展到目前广为接受的ABCDE模型(图1) (Kim et al., 2005), 其中花瓣发育受A、B和E类基因调控。在蛋白质层面, 四因子模型假设4种花同源异型基因(或基因产物)的不同组合决定不同花器官的特征, 其中AP1-AP1-SEP-SEP决定萼片形成(Pelaz et al., 2001), AP1-AP3-PI-SEP决定花瓣形成(Immink et al., 2010), AP3-PI-AG-SEP决定雄蕊形成(Ferrario et al., 2003; Immink et al., 2003), AG-AG-SEP-SEP决定心皮形成(Theissen and Saedler, 2001)。这些蛋白质复合物(可能是转录因子)通过黏着在特异目标基因的启动子上激活或抑制不同的器官特征基因发挥功能; 不同蛋白质复合物和DNA序列之间亲合力的不同和不同基因启动子区域的不同决定了蛋白质复合物和目标基因的相互选择(Theissen, 2001)。

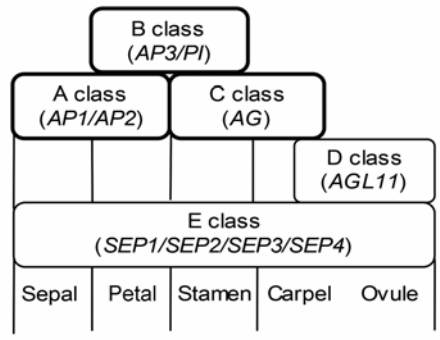

图1 花器官发育的ABCDE模型及相关基因(引自Kim et al., 2005)

Figure 1 The ABCDE model and related genes of floral organ development (from Kim et al., 2005)

Sánchez-Corrales等(2010)总结多个实验结果,

收稿日期: 2014-09-15; 接受日期: 2015-03-19
基金项目: 教育部博士点基金(No.20124320110002)、国家自然科学基金重大研究计划(No.91130009)和国家自然科学基金(No.11475273)
* 通讯作者。E-mail: lsshem@mail.sysu.edu.cn; zhmyuan@sina.com

构建了动态花特异性关键基因调控网络, 初步阐释了这些基因之间的调控关系, 但其涉及花器官发育的关键基因仅限于ABCDE模型中的基因。由于ABCDE模型中基因的产物大都被其它基因的转录产物所调控, 直接或间接参与花器官的形成和功能(Aukerman and Sakai, 2003)。因此, 拟南芥花器官发育分子调控网络有待拓展。

除*AP2*外, ABCDE模型中目前已发现的花器官发育功能基因都属于MADS-box转录因子家族, 含1个由56–58个氨基酸残基组成的高度保守的结构域——MADS-box(吕山花和孟征, 2007; Smaczniak et al., 2012)。该序列编码的转录因子黏着在相应DNA上调该基因转录(Theissen et al., 2000; Theissen, 2001)。含MADS-box结构域的转录因子是植物发育起始与组织特异性的主调节器(Pajoro et al., 2014)。目前, 拟南芥中有近6 000对蛋白质相互作用(protein-protein interaction, PPI)对被实验所验证; 这为以MADS-box蛋白为诱饵, 基于PPI预测拓展拟南芥花器官发育分子调控网络提供了契机。

理论上基于结构的PPI预测更为准确(Zhang et al., 2012), 但应用受限。仅基于序列的PPI预测方法近年来发展迅速, 其独立预测精度已达到可接受水平。如Shen等(2007)将20种天然氨基酸重划分为7类, 以343个三联体残基频率表征蛋白质序列, 采用支持向量机(support vector machine, SVM)建模在人类数据上获得了83%的独立预测精度。Guo等(2008)以自协方差编码蛋白质序列的同时考虑各残基间的距离, 在酵母数据上获得了88.09%的独立预测精度。

本文以拟南芥花瓣发育为例, 整合蛋白互作、亚细胞定位、基因芯片和基因功能注释等数据库, 通过组建PPI可信预测模型, 得到拟南芥花瓣PPI预测网络, 以含有MADS-box结构域的蛋白为诱饵在网络中进行一级拓展, 为深入解析拟南芥花瓣发育的分子调控网络提供依据。

1 数据与方法

1.1 拟南芥PPI预测模型的构建

样本正集: 实验验证的拟南芥(*Arabidopsis thaliana* L.) PPI对(正样本)来自IntAct、BioGRID、BIND与TAIR 4个数据库。自编python脚本剔除重复与自相关作用对后, 按如下规则进行筛选: (1) 序列中不含有未知氨基酸; (2) 序列长度大于50个氨基酸残基; (3) 相互作用的蛋白质相似性低于40%。最后得到6 720个PPI对, 命名为Pair-1-Pos; 含3 561个蛋白质, 命名为Single-1-Pos。

样本负集: 实验验证的拟南芥非相互作用蛋白对数据缺乏, 当前较可信的蛋白质相互作用负集构建法是亚细胞定位法, 它假定亚细胞定位不同的蛋白质之间不发生相互作用(Guo et al., 2008)。拟南芥蛋白质亚细胞定位包含细胞外、细胞质基质、细胞核和线粒体4类(http://suba.plantenergy.uwa.edu.au/), 两两组合后按如下规则进行筛选: (1) 序列中不含有未知氨基酸; (2) 序列长度大于50个氨基酸残基; (3) 配对的两个蛋白质相似性低于40%; (4) 不含有正样本中存在的蛋白对; (5) 随机抽取与正样本数相等的蛋白对。最后得到6 720个蛋白质非相互作用对, 命名为Pair-2-Neg; 含2 793个蛋白质, 命名为Single-2-Neg。

训练集与测试集划分: 从样本正集Pair-1-Pos与样本负集Pair-2-Neg中随机各取5 000对构成训练集Pair-3-Train, 剩下的3 440对构成独立测试集Pair-3-Test。

序列表征: 采用Guo等(2008)提出的自协方差编码表征序列, 每个氨基酸残基包含疏水性、亲水性、极性、极化率、侧链体积、溶剂可及表面积与侧链静电荷指数7种理化性质, 每对蛋白质含2×7×*lg*维特征, *lg*为序列中两个残基之间的最大距离。

SVM训练建模: SVM采用台湾林智仁的Libsvm, 核函数选用径向基核, 核函数c和g参数以网格遍历寻优(grid.py)经5次交叉测试优化获得。

1.2 拟南芥花瓣发育基因芯片数据及其共表达过滤

查询TAIR(http://www.arabidopsis.org/)的子数据库AtGenExpress, 获得6张拟南芥花瓣发育基因芯片ATGE_35(A、B、C)与ATGE_42(B、C、D)。花周期12的ATGE_35含13 024个基因, 花周期15的ATGE_42含12 451个基因。若某基因在6张芯片中表达值均非零, 则视该基因为拟南芥花瓣发育有表达基因。获得花瓣中有表达基因集, 含11 483个基因, 命名为Single-3。

11 483个基因两两组合数据庞大, 直接以前述

SVM模型预测PPI对，会有绝对数量较大的假阳性样本。基因共表达是指两个基因表达模式上的相似性，两个相互作用的蛋白质通常存在共表达(Zhang and Horvath, 2005)。拟南芥基因共表达数据库ATTED-II (http://atted.jp/)含22 574个基因两两间的皮尔逊相关系数(Pearson's correlation coefficient, PCC)矩阵。PCC一般默认阈值为0.6 (Srinivasasainagendra et al., 2008)。为进一步控制蛋白质互作预测的假阳性，设定阈值为0.7；若PCC绝对值大于或等于0.7，则认为两基因间存在共表达。花瓣中有表达基因集Single-3经共表达过滤，得到共表达明显的166 836对蛋白质，命名为Pair-4；含3 982个蛋白质，命名为Single-4。

1.3 含MADS-box诱饵蛋白在拟南芥花瓣PPI网络中的一级拓展

取拟南芥PPI样本正集中的3 561个蛋白质集合Single-1-Pos与花瓣中有表达的11 483个基因集合Single-3的交集，得花瓣中实验验证有相互作用的蛋白质1 998个，命名为Singe-5-Obs；包含3 133对实验验证相互作用蛋白质，命名为Pair-5-Obs。

合并花瓣中共表达明显且预测为阳性的PPI对以及花瓣中实验验证PPI对，构成拟南芥花瓣PPI网络；以拟南芥含MADS-box结构域的蛋白质为诱饵，在PPI网络中进行一级拓展。

2 结果与讨论

2.1 拟南芥PPI模型评估与预测

训练集Pair-3-Train经SVM建模，在*lg*为15、20、25、30、35、40和45时对含3 440个样本的独立测试集Pair-3-Test获得了76.24%–78.03%的预测精度，当*lg*为30时最高，达78.03%，显示该模型较为可信。

取*lg*=30，预测花瓣中共表达明显的166 836对蛋白质Pair-4，得15 621个阳性样本(预测有相互作用蛋白对)，命名为Pair-6-Pre；含2 235个蛋白质，命名为Single-6-Pre。

2.2 基于MADS-box诱饵与PPI的拟南芥花瓣发育分子网络

对实验验证有相互作用蛋白对Pair-5-Obs与预测有相互作用蛋白对Pair-6-Pre求并集，得花瓣中PPI对18 703，命名为Pair-7-Total；含3 692个蛋白质，命名为Single-7-Total。

拟南芥全基因组含MADS-box结构域的蛋白质有107个(Parenicová et al., 2003)，Single-7-Total覆盖其中的10个：AP1、AP3、PI、SEP1、SEP2、SEP3、CAL、AGL18、AGL16和FLM。将花瓣中PPI对Pair-7-Total导入网络可视化程序Cytoscape，以10个MADS-box蛋白质为诱饵进行一级拓展，得拓展的拟南芥花瓣发育分子网络(图2)，该网络含38个蛋白质，67对相互作用；不计与实验验证重合的PPI对，新增预测PPI为27对。其中，SEP3与SEP1之间的相互作用虽未被实验证实或证伪，但同源相互作用支持SEP3与SEP1之间存在互作(de Folter et al., 2005)。

由表1可知，网络中涉及的38个蛋白质中多数蛋白其DAVID (http://david.abcc.ncifcrf.gov/)功能注释涉及的生物学过程与花发育调控相关。此外，有报道显示，*AT3G04960*基因的表达在成花诱导中与花同源异型基因相关(Schimid et al., 2003)；*CYP77A6*参与花发育的角质层形成，属花发育ABCDE模型中A类基因亚家族的重要成员(Li-Beisson et al., 2009; Yang et al., 2010)；AGL18参与花发育的负调控、延迟花器官衰老和脱落(Chen et al., 2011)。

2.3 花瓣发育分子网络中候选的四元互作

花器官发育的四因子假说模型提示我们应关注网络中潜在的四元互作。以4个蛋白6对PPI为条件，从图2中可提取19个候选四元互作(图3)，涉及ABCDE模型基因之外的8个基因；特别是*AGL16*，其本身含MADS-box结构域，在系统发育上与B类基因*AP3*和*PI*较近，与A类基因*AP1*以及E类基因*SEP1*、*SEP2*和*SEP3*较远(Parenicová et al., 2003)，推测*AGL16*是B类基因新成员或其冗余。此外，AT1G02190蛋白参与角质层蜡质的合成与花粉发育(Gomez-Mena et al., 2005)；AT2G20870为细胞壁蛋白，在光周期通路中下调拟南芥的开花时间(Cai et al., 2007)。

2.4 花瓣四聚体的候选作用靶标

四因子假设模型中，AP1-AP3-PI-SEP决定花瓣形成，该四聚体通过黏着在特异目标基因的启动子上激活

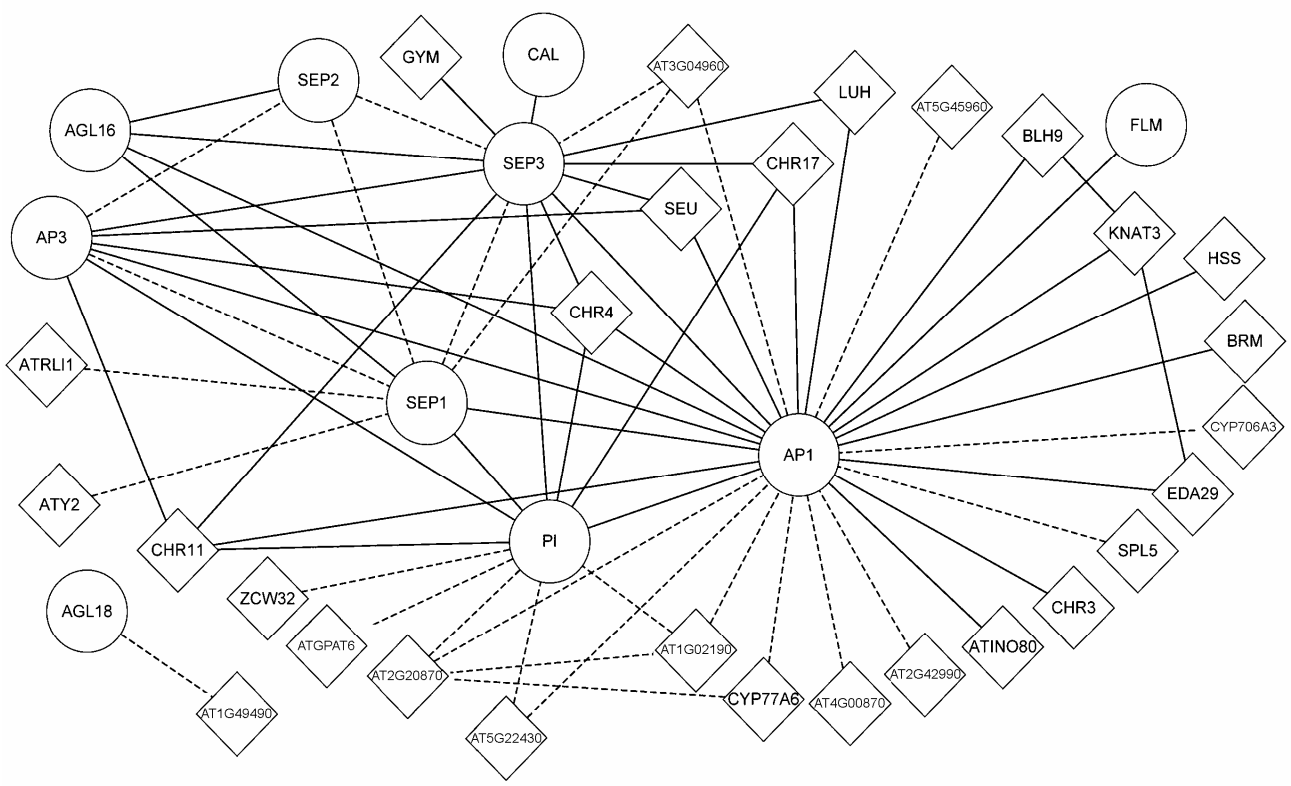

图2 基于MADS-box诱饵与蛋白质相互作用的拟南芥花瓣发育分子网络

○: 含MADS-box结构域的蛋白质; ◇: 不含MADS-box结构域的蛋白质; —: 实验验证的蛋白质互作(PPI); ---: 支持向量机预测的蛋白质互作(PPI)。

Figure 2 Molecular network related to petal development based on MADS-box bait protein and protein-protein interaction (PPI) in *Arabidopsis thaliana*

○: Proteins with MADS-box domain; ◇: Proteins without MADS-box domain; —: Verified PPI by experiment; ---: Predicted PPI by support vector machine (SVM).

或抑制不同的器官特征基因发挥功能(Theissen and Saedler, 2001)。基于PPI与靶基因的功能注释, 根据图2所示信息提取到花瓣AP1-AP3-PI-SEP四聚体的6个候选靶标基因: *SEU*、*LUH*、*CHR4*、*CHR11*、*CHR17*和*AT3G04960*。其中, SEU功能注释为花发育调控、胚珠发育、胚芽发育、雌蕊发育和多细胞组织发育等。LUH功能注释为花发育、胚芽发育和转录调控等。CHR4、CHR11和CHR17主要功能注释为染色体组装及染色质重塑。AP1-AP3-PI-SEP四聚体以MADS-box序列编码的转录因子通过黏着在CHR4、CHR11和CHR17的DNA上调其转录(Theissen and Saedler, 2001; Huanca-Mamani et al., 2005), 支持本文CHR4、CHR11和CHR17为花瓣四聚体候选作用靶标的推测。AT3G04960在成花诱导中与花同源异型基因相关(Schimid et al., 2003)。

2.5 讨论

影响PPI预测精度的因素众多。(1) 负集构建方法。常用的包括正样本中蛋白质随机配对法、亚细胞定位法和一侧蛋白质*k*-let重排人工构造序列法等。Guo等(2008)研究表明, 亚细胞定位法构造负集更为可信。然而, 假设A-B存在互作, 由于穿梭蛋白的存在, 若蛋白A本应同时定位于细胞质基质和细胞核, 蛋白B本应仅定位于细胞核; 由于当前亚细胞定位仅记录到蛋白A定位于细胞基质、蛋白B定位于细胞核, 且当前实验验证数据库中无A-B互作记录, 则A-B可能被错

表1 拓展网络中38个蛋白质涉及的生物学过程

Table 1 The related biological processes of 38 proteins in expansion network

Proteins	Biological processes
AP1	Flower development; floral meristem determinacy; meristem structural organization; positive regulation of transcription
AP3	Flower development; reproductive structure developmental; regulation of transcription
PI	Flower development; reproductive structure developmental; regulation of transcription
SEP1	Flower development; ovule development
SEP2	Flower development; ovule development
SEP3	Flower development; ovule development; specification of floral organ identity; regulation of transcription
LUH	Flower development; embryo development; negative regulation of transcription; regulation of transcription
SEU	Regulation of flower development; ovule development; embryo development; gynoecium development; multicellular organismal development
AGL16	Regulation of transcription
CHR4	Chromatin organization; chromatin assembly or disassembly; regulation of transcription; chromosome organization
CHR11	Embryo sac development; gametophyte development; chromatin organization; chromatin remodeling; chromosome organization
CHR17	Chromatin organization; chromatin remodeling; chromatin modification; chromosome organization
CHR3	Flower development; organ boundary specification between lateral organs and the meristem; response to wounding
AGL18	Negative regulation of flower development; negative regulation of short-day photoperiodism; flowering
CAL	Floral meristem determinacy; positive regulation of flower development
FLM	Negative regulation of flower development; photoperiodism; flowering; regulation of flower development
HSS	Ovule development; positive regulation of flower development; floral organ abscission; negative regulation of transcription
GYM	Regulation of lateral root development; negative regulation of transcription
BLH19	Fruit development; floral whorl morphogenesis; negative regulation of transcription; organ formation
SPL5	Regulation of transcription; regulation of vegetative phase change
ZCW32	Petal morphogenesis; regulation of transcription; DNA-dependent
CYP77A6	Cutin biosynthetic process; flower development
ATGPAT6	Cutin biosynthetic process; flower development; metabolic process
ATINO80	Positive regulation of DNA repair; regulation of transcription; somatic cell DNA recombination
AT4G00870	Regulation of transcription
AT1G02190	Fatty acid metabolic process; fatty acid biosynthetic process; lipid biosynthetic process; organic acid biosynthetic process
AT2G42990	Lipid catabolic process
AT5G45960	Lipid catabolic process
KNAT3	Regulation of transcription; detection of hormone stimulus; detection of cytokinin stimulus; response to organic substance
ATY2	Glycerol ether metabolic process; cellular homeostasis; cell redox homeostasis
EDA29	Polar nucleus fusion; response to abscisic acid stimulus
BRM	Regulation of gene expression
CYP706A3	Unknown
ATRLI1	Unknown
AT3G04960	Unknown
AT1G49490	Unknown
AT5G22430	Unknown
AT2G20870	Unknown

误地选择为负样本。物种间当前亚细胞定位信息准确性与全面性的差异是导致不同物种间负集构建合理性进而影响PPI预测准确性的重要原因。(2) 正样本代表性的覆盖度。PPI存在多种模式，若某物种已经

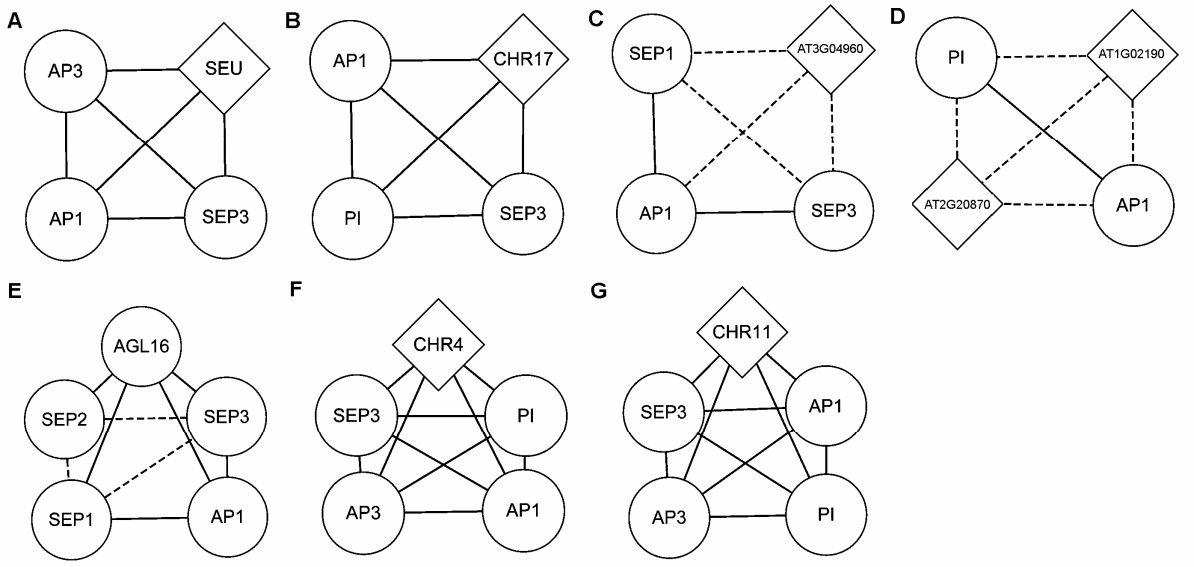

图3 拟南芥花瓣发育分子网络中候选的四元互作

○: 含MADS-box结构域的蛋白质; ◇: 不含MADS-box结构域的蛋白质; —: 实验验证的蛋白质互作(PPI); ---: 支持向量机预测的蛋白质互作(PPI)。

Figure 3 The candidate tetrameric interactions related to petal development molecular network in *Arabidopsis thaliana*

○: Proteins with MADS-box domain; ◇: Proteins without MADS-box domain; —: Verified protein-protein interaction by experiment; ---: Predicted protein-protein interaction by support vector machine (SVM).

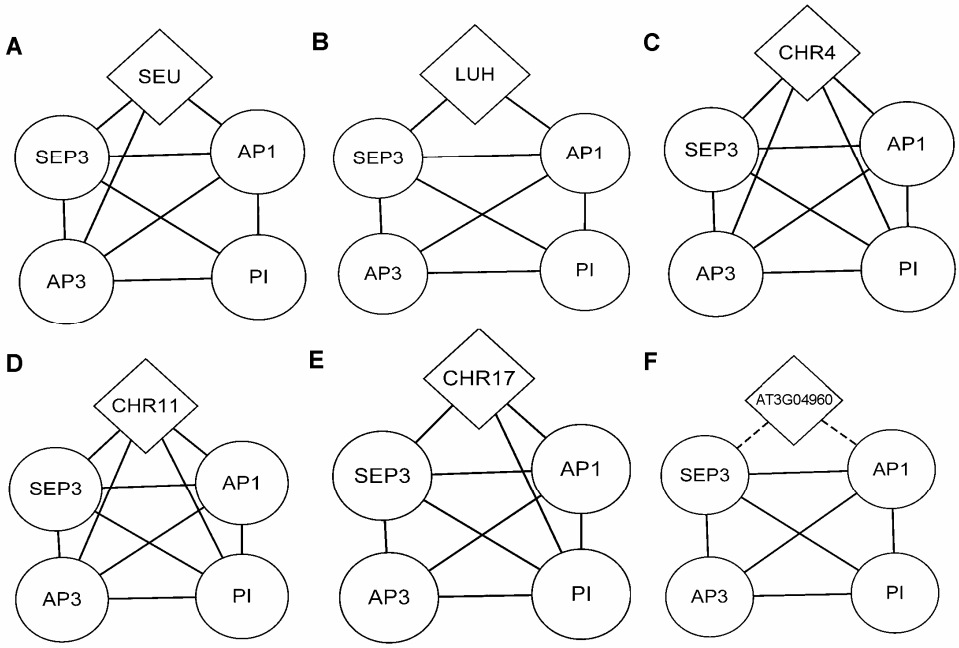

图4 拟南芥花瓣四聚体的候选作用靶标

○: 含MADS-box结构域的蛋白质; ◇: 不含MADS-box结构域的蛋白质; —: 实验验证的蛋白质互作(PPI); ---: 支持向量机预测的蛋白质互作(PPI)。

Figure 4 The candidate targets of petal tetramers in *Arabidopsis thaliana*

○: Proteins with MADS-box domain; ◇: Proteins without MADS-box domain; —: Verified protein-protein interaction by experiment; ---: Predicted protein-protein interaction by support vector machine (SVM).

过实验验证存在PPI的正样本代表性不够，仅覆盖少量PPI模式，必然导致全基因组范围内PPI预测精度偏低。(3) 序列表征与特征筛选。如氨基酸重分7类三联体残基频率表征(Shen et al., 2007)与残基7种理化性质自协方差编码表征(Guo et al., 2008)。7种理化性质可能不足以代表残基的复杂特性，AA531数据库包含天然氨基酸的531种理化性质，代表性虽充分，但又势必存在无关与冗余，此时特征筛选成为必要。(4) 样本A-B中各蛋白质序列特征的先后排列次序。可交换核函数是解决这一问题的有效手段(Shen et al., 2007)。相比Shen等(2007)在人类数据上获得的83%及Guo等(2008)在酵母数据上获得的88.09%的独立预测精度，本研究在拟南芥数据上获得的78.03%独立预测精度偏低。综合考虑上述改进因素，进一步较大幅度提升拟南芥PPI预测精度将是后续工作的重要内容。

本研究挖掘到花瓣发育分子网络中19个候选四元互作，含53对PPI，其中12对PPI有待分子生物学实验验证(图3)；推测花瓣AP1-AP3-PI-SEP四聚体的6个候选靶标基因为*SEU*、*LUH*、*CHR4*、*CHR11*、*CHR17*和*AT3G04960* (图4)，可供分子生物学实验优先验证。花器官发育包括萼片、花瓣、雄蕊、心皮和胚珠，本文仅涉及花瓣。随着其它花器官芯片数据的丰富，构建一个完整的拟南芥花器官发育分子调控网络是值得探讨的一个新课题。

参考文献

吕山花, 孟征 (2007). MADS-box基因家族基因重复及其功能的多样性. 植物学通报 **24**, 60–70.

Aukerman MJ, Sakai H (2003). Regulation of flowering time and floral organ identity by a microRNA and its *APETALA2*-like target genes. *Plant Cell* **15**, 2730–2741.

Cai XN, Ballif J, Endo S, Davis E, Liang MX, Chen D, DeWald D, Kreps J, Zhu T, Wu YJ (2007). A putative CCAAT-binding transcription factor is a regulator of flowering timing in Arabidopsis. *Plant Physiol* **145**, 98–105.

Chen MK, Hsu WH, Lee PF, Thiruvengadam M, Chen HL, Yang CH (2011). The MADS box gene, *FOREVER YOUNG FLOWER*, acts as a repressor controlling floral organ senescence and abscission in Arabidopsis. *Plant J* **68**, 168–185.

De Folter S, Immink RGH, Kieffer M, Pařenicová L, Henz SR, Weigel D, Busscher M, Kooiker M, Colombo L, Kater MM, Davies B, Angenent GC (2005). Comprehensive interaction map of the Arabidopsis MADS box transcription factors. *Plant Cell* **17**, 1424–1433.

Ferrario S, Immink RGH, Shchennikova A, Busscher-Lange J, Angenent GC (2003). The MADS box gene *FBP$_2$* is required for SEPALLATA function in petunia. *Plant Cell* **15**, 914–925.

Gómez-Mena C, de Folter S, Costa MMR, Angenent GC, Sablowski R (2005). Transcriptional program controlled by the floral homeotic gene *AGAMOUS* during early organogenesis. *Development* **132**, 429–438.

Guo YZ, Yu LZ, Wen ZN, Li ML (2008). Using support vector machine combined with auto covariance to predict protein-protein interactions from protein sequences. *Nucleic Acids Res* **36**, 3025–3030.

Huanca-Mamani W, Garcia-Aguilar M, León-Martínez G, Grossniklaus U, Vielle-Calzada JP (2005). CHR11, a chromatin-remodeling factor essential for nuclear proliferation during female gametogenesis in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* **102**, 17231–17236.

Immink RGH, Ferrario S, Busscher-Lange J, Kooiker M, Busscher M, Angenent GC (2003). Analysis of the petunia MADS-box transcription factor family. *Mol Genet Genomics* **268**, 598–606.

Immink RGH, Kaufmann K, Angenent GC (2010). The 'ABC' of MADS domain protein behaviour and interactions. *Semin Cell Dev Biol* **21**, 87–93.

Kim S, Koh J, Yoo MJ, Kong HZ, Hu Y, Ma H, Soltis PS, Soltis DE (2005). Expression of floral MADS-box genes in basal angiosperms: implications for the evolution of floral regulators. *Plant J* **43**, 724–744.

Li-Beisson Y, Pollard M, Sauveplane V, Pinot F, Ohlrogge J, Beisson F (2009). Nanoridges that characterize the surface morphology of flowers require the synthesis of cutin polyester. *Proc Natl Acad Sci USA* **106**, 22008–22013.

Pajoro A, Madrigal P, Muiño JM, Matus JT, Jin J, Mecchia MA, Debernardi JM, Palatnik JF, Balazadeh S, Arif M, Ó'Maoiléidigh DS, Wellmer F, Krajewski P, Riechmann JL, Angenent GC, Kaufmann K (2014). Dynamics of chromatin accessibility and gene regulation by MADS-domain transcription factors in flower development. *Genome Biol* **15**, R41.

Parenicová L, de Folter S, Kieffer M, Horner DS, Favalli C, Busscher J, Cook HE, Ingram RM, Kater MM, Davies B, Angenent GC, 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.

Pelaz S, Gustafson-Brown C, Kohalmi SE, Crosby WL, Yanofsky MF (2001). *APETALA1* and *SEPALLATA3* interact to promote flower development. *Plant J* **26**, 385–394.

Sánchez-Corrales YE, Álvarez-Buylla ER, Mendoza L (2010). The *Arabidopsis thaliana* flower organ specification gene regulatory network determines a robust differentiation process. *J Theor Biol* **264**, 971–983.

Schmid M, Uhlenhaut NH, Godard F, Demar M, Bressan R, Weigel D, Lohmann JU (2003). Dissection of floral induction pathways using global expression analysis. *Development* **130**, 6001–6012.

Shen JW, Zhang J, Luo XM, Zhu WL, Yu KQ, Chen KX, Li YX, Jiang HL (2007). Predicting protein-protein interactions based only on sequences information. *Proc Natl Acad Sci USA* **104**, 4337–4341.

Smaczniak C, Immink RGH, Muino M, Blanvillain R, Busscher M, Lange JB, Dinh QD, Liu SJ, Westphal AH, Boeren S, Parcy F, Xu L, Carles CC, Angenent GC, Kaufmann K (2012). Characterization of MADS-domain transcription factor complexes in Arabidopsis flower development. *Proc Natl Acad Sci USA* **109**, 1560–1565.

Srinivasasainagendra V, Page GP, Mehta T, Coulibaly I, Loraine AE (2008). CressExpress: a tool for large-scale mining of expression data from Arabidopsis. *Plant Physiol* **147**, 1004–1016.

The Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815.

Theissen G (2001). Development of floral organ identity: stories from the MADS house. *Curr Opin Plant Biol* **4**, 75–85.

Theissen G, Becker A, Di Rosa A, Kanno A, Kim JT, Münster T, Winter KU, Saedler H (2000). A short history of MADS-box genes in plants. *Plant Mol Biol* **42**, 115–149.

Theissen G, Saedler H (2001). Plant biology: floral quartets. *Nature* **409**, 469–471.

Yang WL, Pollard M, Li-Beisson Y, Beisson F, Feig M, Ohlrogge J (2010). A distinct type of glycerol-3-phosphate acyltransferase with *sn*-2 preference and phosphatase activity producing 2-monoacylglycerol. *Proc Natl Acad Sci USA* **107**, 12040–12045.

Zhang B, Horvath S (2005). A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol* **4**, Article 17.

Zhang QC, Petrey D, Deng L, Qiang L, Shi Y, Thu CA, Bisikirska B, Lefebvre C, Accili D, Hunter T, Maniatis T, Califano A, Honig B (2012). Structure-based prediction of protein-protein interactions on a genome-wide scale. *Nature* **490**, 556–560.

Expansion of the Molecular Network Related to Petal Development Based on MADS-box Proteins and Protein-Protein Interaction Network in *Arabidopsis thaliana*

Li Yang[1, 2], Congwei Sun[1, 2], Zhijun Dai[1, 2], Miao He[3*], Zheming Yuan[1, 2*]

[1]Hunan Provincial Key Laboratory of Crop Germplasm Innovation and Utilization, Hunan Agricultural University, Changsha 410128, China; [2]Hunan Provincial Key Laboratory for Biology and Control of Plant Diseases and Insect Pests, Hunan Agricultural University, Changsha 410128, China; [3]School of Life Sciences, Sun Yat-sen University, Guangzhou 510275, China

Abstract The regulatory mechanism of flower organ development has an important role in evolution, development and ecology. Here, we used petal development of *Arabidopsis thaliana* as an example and integrated protein-protein interaction, subcellular localization, gene-chip and gene functional annotation databases to reveal a protein-protein interaction network related to petals of *A. thaliana* and built a reliable predictive model of protein-protein interaction. By using proteins containing the MADS-box domain as bait, we could expand the network by one level and obtained an expanded network of 38 proteins and 67 protein-protein interactions. Gene functional annotation with the DAVID database suggested that most of the proteins were involved in regulation of flower development in the expanded network. We derived 19 candidate tetrameric interactions, involving 8 genes, from the expanded network. None of the 8 genes belonged to the ABCDE model genes: *AGL16,* with an MADS-box domain, may be a new member or a redundant gene of class B. *SEU*, *LUH*, *CHR4*, *CHR11*, *CHR17*, and *AT3G04960* were candidate targets of petal AP1-AP3-PI-SEP tetramers of *A. thaliana*. The results provide references for deeply analyzing the molecular regulatory network related to petal development of *A. thaliana*.

Key words *Arabidopsis thaliana*, MADS-box, petal development, protein-protein interaction

Yang L, Sun CW, Dai ZJ, He M, Yuan ZM (2015). Expansion of the molecular network related to petal development based on MADS-box proteins and protein-protein interaction network in *Arabidopsis thaliana*. *Chin Bull Bot* **50**, 614–622.

---

* Authors for correspondence. E-mail: lsshem@mail.sysu.edu.cn; zhmyuan@sina.com

(责任编辑: 白羽红)