

银杏MADS-box基因家族的表达及系统发育分析

董金金^{1,2}, 刘伟², 李萌^{1,2}, 王俊焱², 何崇单², 蔡雨蒙², 王启业², 王义强^{1,2,*}

¹中南林业科技大学经济林培育与保护教育部重点实验室, 长沙410004

²中南林业科技大学林业生物技术湖南省重点实验室, 长沙410004

摘要: MADS-box基因是真核生物中一类编码转录调控因子的基因家族, 在植物花器官发育中发挥重要的调控作用。为了探究MADS-box家族基因在银杏花发育中的功能, 本文对银杏花芽分化4个时期的样品进行转录组测序, 筛选其中的MADS-box家族基因, 利用生物信息学方法对筛选到的基因进行表达模式、蛋白结构、细胞定位和系统进化分析。结果显示: 共得到15个银杏MADS-box家族基因。表达分析表明, 目标基因根据表达模式可以分为3种类型。分析目标基因的编码序列显示, 15个银杏MADS-box蛋白的主要构件均为 α -螺旋和无规则卷曲; 亚细胞定位预测主要在细胞核内; 所有表达产物均包含MADS结构域。聚类分析显示, 银杏MADS-box基因家族可分为6个亚类。银杏MADS-box基因家族中的*newGb2734*、*Gb38883*、*Gb28587*和*Gb33168*可能在银杏开花调控中发挥重要作用; *Gb16301*可能是银杏花器官的发育过程中的关键基因。

关键词: 银杏; MADS-box基因家族; 开花调控; 花器官发育

MADS-box基因家族是一类含MADS结构域的同源异型基因, 编码的蛋白是一种转录因子, 在动物、植物、真菌等生物体中MADS-box基因主要通过调节其他基因的转录而发挥作用。Becker和Theissen (2003)根据基因结构、蛋白结构以及系统进化关系等将植物MADS-box家族基因分为I型和MIKC型两类。I型MADS-box基因一般含有1~2个外显子, 编码的蛋白主要包含1个高度保守的SRF-like域。MIKC型基因一般包括6个内含子和7个外显子, 该类型基因编码蛋白主要由4部分组成, 分别是: (1) MEF2-like MADS域是一段位于N端、由约56个氨基酸组成的非常保守的区域, 是结合特异DNA序列的基元; (2) 半保守的K域是一段包含约70个氨基酸残基的片段, 负责蛋白质之间的相互作用; (3) I域是位于MADS域和K域之间的氨基酸序列, 保守程度相对较低; (4) C域是位于K域下游、保守程度最低的区域, C域的差异导致了MADS-box蛋白具有不同的功能(Immink等2002)。分布于植物中的MADS-box基因家族参与了决定根结构、花芽分化、花发育以及果实成熟等植物生长发育和信号转导过程中的关键事件, 其中最重要的功能是在花和果实发育中的调控作用(刘菊花等2010)。

Becker和Theissen (2003)分离得到了大量的植物MIKC型MADS-box基因, 通过构建系统发育树的方法对得到的基因进行亚细胞分类, 将被子植物中的MADS-box基因分为*AGAMOUS (AG)*、*AGL2*

(*SEP*)、*AGL6*、*AGL12*、*AGL15*、*AGL17*、*AP3/PI*、*GGM13*、*STMADSII*、*TM3*、*API (SQUA)*和*FLC*共12个主要的进化分支。现代演化发育生物学认为, 植物花发育过程中的基因与调控植物花发育起始的基因在进化上是高度同源的。Melzer等(2010)对裸子植物和被子植物的MADS-box基因进一步分析归类, 将其划分为14个主要的进化分支: *AGAMOUS (AG)*、*AGL2 (SEP)*、*AGL6*、*AGL12*、*AGL15*、*AGL17*、*AP3/PI*、*GGM13*、*SVP*、*API (SQUA)*、*FLC*、*TM3*、*TM8*和*GpMADS4*。目前对于裸子植物MADS-box基因的研究对象主要集中在松柏类和买麻藤类植物中, 获得的MADS-box基因主要分为*DEF/GLO*、*GGM13*、*AG*、*AGL6*、*TM3*和*TM8*亚族。

银杏(*Ginkgo biloba*)是裸子植物中最古老的孑遗植物, 被称为植物界的活化石, 是我国特产的经济树种。银杏果鲜美且营养丰富, 是著名的保健干果; 其叶富含银杏黄酮与萜内酯, 是治疗心脑血管疾病的有效成分(郭明等2014); 其树势高雅、叶形独特优美, 是世界著名的庭园、城市绿化树种。然而, 银杏为单科单属单种, 遗传资源少, 且幼年期长达20年, 严重阻碍银杏育种的进程, 致使

收稿 2017-12-22 修定 2018-05-15

资助 国家自然科学基金(31570682)和国家重点研发计划(2017YFD0600701)。

* 通讯作者(wangyiqiang12@163.com)。

优良品种缺乏,限制了银杏的经济效益和社会价值。另外,银杏作为原始的裸子植物,可以作为研究裸子植物花器官发育调控的重要材料。研究银杏童期和开花调控的分子机制,对于促进银杏开花、加速优良品种改良具有十分重要的意义。

随着基因组测序技术的飞速发展,目前挪威云杉(*Picea abies*)、无油樟(*Amborella trichopoda*)、火炬松(*Pinus taeda*)、银杏等木本植物已完成全基因组测序(Chamala等2013; Guan等2016; Nystedt等2013; Zimin等2014)。木本植物的MADS-box基因的信息也已比较丰富,将MADS-box基因的进化与植物花器官的起源与进化相关联,可以为在全基因组水平揭示木本植物的开花调控机理奠定理论基础。本研究依据银杏转录组数据,鉴定了银杏的MADS-box基因家族成员,根据银杏MADS-box基因家族的染色体定位信息,初步了解了银杏MADS-box基因家族在基因组的分布特点。通过序列比对和系统进化分析,结合MADS-box家族基因在花发育3个时期的表达情况分析,完成了MADS-box基因家族成员的亚族分类。本研究为解析银杏MADS-box基因家族的重要作用以及发掘银杏开花调控关键基因提供参考。

1 材料与方法

1.1 植物材料

采用种植于中南林业科技大学校内树龄约为20年银杏(*Ginkgo biloba* L.)雌株作为实验材料,在该植株上采集花芽未分化期、花芽分化初期、花芽分化盛期、花芽分化末期的短枝顶芽(银杏没有完整的花结构,银杏花芽即短枝顶端孕育花的混合花芽),用于提取总RNA进行转录组测序。

1.2 转录组数据处理

使用Illumina HiSeq 2500平台进行转录组测序,将Clean Reads与银杏参考基因组(<http://gigadb.org/dataset/100209>)进行序列比对,获取基因在银杏参考基因组上的位置信息,以及银杏测序样品序列中特有的信息;使用BLAST软件将发掘的基因与Swiss-Prot、NR、KOG、GO、COG等数据库进行序列比对,获得基因的注释信息。

1.3 银杏MADS-box基因的蛋白结构和保守基序分析

使用ORF finder在线工具([https://www.ncbi.](https://www.ncbi.nlm.nih.gov/orffinder/)

[nlm.nih.gov/orffinder/](https://www.ncbi.nlm.nih.gov/orffinder/))分析银杏MADS-box家族基因的开放阅读框(ORF);使用ProtParam在线工具(<http://web.expasy.org/protparam/>)估计银杏MADS-box基因家族各成员的蛋白等电点、分子量等基本理化参数;使用SOPMA在线工具(<http://www.prabi.fr/>)预测银杏MADS-box基因家族的蛋白二级结构;使用PSORT Prediction在线分析工具(<http://www.genscript.com/psort.html>)对银杏MADS-box基因家族成员的亚细胞定位进行预测。利用MEME在线分析工具(<http://meme-suite.org/index.html>)分析银杏MADS-box蛋白的保守基序。运用SMART(<http://smart.embl-heidelberg.de/>)在线工具对得到的保守基序进行功能注释,并参照拟南芥和水稻的命名模式对银杏保守基序命名。

1.4 银杏MADS-box基因家族的系统进化分析

利用MEGA 7.0进行系统进化分析,系统进化分析参数使用neighbor-joining (NJ)法则的P-距离(P-distance)模型构建,并进行1 000次Bootstrap抽样。引入拟南芥、水稻、挪威云杉、油松等植物的MADS-box家族基因信息对银杏MADS-box蛋白进行亚组分类。拟南芥MADS-box蛋白序列下载于拟南芥数据库TAIR (<http://www.arabidopsis.org/>),挪威云杉、油松、日本柳杉等裸子植物的MADS-box蛋白序列信息下载于NCBI的蛋白数据库(<https://www.ncbi.nlm.nih.gov/>)。

1.5 银杏MADS-box基因家族的表达分析

转录组测序使用Cufflinks软件的Cuffquant和Cuffnorm组件,分析测序数据与参考基因组的匹配度,对转录本和基因的表达水平进行定量分析。Cuffquant和Cuffnorm组件采用FPKM (Fragments Per Kilobase of transcript per Million fragments mapped) (Florea等2013)作为衡量基因表达水平的指标,根据表达水平对银杏MADS-box基因进行聚类分析。

2 实验结果

2.1 转录组测序数据分析

对银杏短枝顶芽的转录组测序数据进行整理分析,结果显示,4组转录组测序数据的Clean Data均达到7.13 Gb以上, Q_{30} 碱基百分比在89.65%及以上;Clean Reads与银杏参考基因组的匹配率从78.24%~82.24%不等;参考转录组测序的注释信息,利用生物信息学方法,筛选到15个银杏MADS-box

家族基因, 基本信息见表1。真核生物同源蛋白簇数据库(KOG)、直系同源类群数据库(eggNOG)等数据库的功能注释为转录, 核酸序列数据库(NR)的注释信息为银杏、无油樟、毛果杨等木本植物的MADS-box基因, 蛋白序列数据库(Swiss Prot)的注释信息为藤本植物买麻藤以及草本植物拟南芥、水稻、矮牵牛、番茄的MADS-box基因。

染色体定位数据显示15个银杏MADS-box家族基因的分布不均匀, 其中*Gb05128*位于scaffold-12770、*Gb15398*位于scaffold23613、*Gb19178*位于scaffold29498, 其余12个基因位于scaffold1000~10000之间; scaffold1082包含2个基因*Gb01884*、*newGb184*, scaffold5068包含2个基因*Gb28587*和*newGb2734*, 其余scaffold仅有1个基因。

2.2 银杏MADS-box基因的蛋白结构和保守基序分析

筛选得到的15个银杏MADS-box家族基因的基本信息见表2。15个基因编码的蛋白长度在150~450 aa之间; 蛋白分子量在19 214.44~50 951.76 Da之间; 其中有10个蛋白的等电点大于7.5, 显碱性; 5个蛋白等电点小于6.5, 显酸性。蛋白二级结构预测结果显示, 15个蛋白的主要构件为 α -螺旋和无规则卷曲。绝大多数的蛋白定位于细胞核中, 只有少数的MADS-box蛋白定位到其他细胞器中, 定位于线粒体的是*Gb01884*和*Gb05128*, 定位于内质网的是*Gb28587*。

通过MEME软件预测到银杏MADS-box蛋白的3个保守基序, 并利用在线软件SMART对预测的保守基序进行命名。结果如图1所示, 15个银杏MADS-box家族成员中有13个包含保守结构域MADS域以及半保守结构域K域, 另有两个成员*Gb33168*和*Gb38883*只含有保守结构域MADS域。

2.3 银杏MADS-box基因系统进化分析

选取拟南芥、水稻与裸子植物MADS-box基因共计83条序列, 与所获得的银杏MADS-box基因利用MEGA 7.0构建I型和MIKC型MADS-box基因系统发育树。结果如图2和3所示, 黑色实心标注的为银杏的2条I型和13条MIKC型MADS-box基因。

系统发育学中一般将I型MADS-box基因分为Ma、Mb和M γ (Masiero等2011)。随机选取拟南芥 α 、 β 、 γ 三种类型的11条基因与*Gb33168*和*Gb38883*构建的系统发育进化树, 结果(图2)显示:*Gb33168*和*Gb38883*与拟南芥 γ 型基因聚为一支, 与拟南芥 β 型基因互为姐妹支。

MIKC型MADS-box基因系统发育进化树(图3)显示, 银杏13条MIKC型MADS-box基因被分到TM3、TM8、AG、GGM13、SVP亚族中。系统进化树中*Gb15398*、*Gb28587*、*newGb2734*与买麻藤GGM13以及挪威云杉PaDAL12、油松PtDAL22等基因分类到GGM13亚族, 分析系统进化关系可以

表1 银杏MADS-box基因的基本信息
Table 1 Basic information of the MADS-box gene in *G. biloba*

基因名称	转录组编号	序列长度/bp	NR注释	Swiss Prot 注释	染色体定位
<i>Gb01884</i>	<i>gene.Gb_01884</i>	1 117	银杏 <i>GbMADS6</i>	拟南芥MADS-box	scaffold1082:3,179,433..3,638,890
<i>Gb05128</i>	<i>gene.Gb_05128</i>	1 237	银杏MADS-box	水稻MADS-box	scaffold12770:72,950..237,867
<i>Gb15398</i>	<i>gene.Gb_15398</i>	834	银杏 <i>GbMADS4</i>	买麻藤MADS-box	scaffold23613:525,999..527,958
<i>Gb16301</i>	<i>gene.Gb_16301</i>	1 142	银杏 <i>GbMADS5</i>	矮牵牛MADS-box	scaffold2493:2,468,237..2,510,564
<i>Gb19178</i>	<i>gene.Gb_19178</i>	741	无油樟转录蛋白	水稻MADS-box	scaffold29498:65,158..231,954
<i>Gb28587</i>	<i>gene.Gb_28587</i>	774	毛果杨假定蛋白	买麻藤MADS-box	scaffold5068:319,234..321,439
<i>Gb30604</i>	<i>gene.Gb_30604</i>	501	银杏 <i>GbMADS9</i>	水稻MADS-box	scaffold5642:81,505..177,039
<i>Gb33168</i>	<i>gene.Gb_33168</i>	630	银杏 <i>GbMADS11</i>	水稻MADS-box	scaffold6370:264,907..265,536
<i>Gb34103</i>	<i>gene.Gb_34103</i>	504	北美云杉未知蛋白	拟南芥MADS-box	scaffold6722:2,142,749..2,290,910
<i>Gb36364</i>	<i>gene.Gb_36364</i>	1 214	银杏 <i>GbMADS1</i>	番茄MADS-box	scaffold7525:1,876,257..2,034,271
<i>Gb38883</i>	<i>gene.Gb_38883</i>	1 338	日本柳杉MADS-box	水稻MADS-box	scaffold8468:608,620..611,059
<i>Gb39109</i>	<i>gene.Gb_39109</i>	600	香蕉 MADS-box	拟南芥MADS-box	scaffold8562:79,971..235,840
<i>Gb41549</i>	<i>gene.Gb_41549</i>	738	银杏 <i>GbMADS8</i>	水稻MADS-box	scaffold9813:2,142,372..2,269,984
<i>newGb184</i>	<i>gingko_newGene_184</i>	1 250	银杏 <i>GbMADS7</i>	番茄MADS-box	scaffold1082:1,283,766..1,717,231
<i>newGb2734</i>	<i>gingko_newGene_2734</i>	875	银杏 <i>GbMADS9</i>	买麻藤MADS-box	scaffold5068:988,033..990,526

表2 银杏MADS-box蛋白的基本信息

Table 2 Basic information of the MADS-box proteins in *G. biloba*

基因名称	氨基酸/aa	分子量/Da	等电点	亚细胞定位	α-螺旋比例/%	β-折叠比例/%	β-转角比例/%	无规则卷曲比例/%
<i>Gb01884</i>	251	28 321.70	9.80	线粒体	44.29	21.17	8.36	26.18
<i>Gb05128</i>	244	28 054.61	4.96	线粒体	46.65	17.87	9.43	26.05
<i>Gb15398</i>	277	31 889.43	8.86	细胞核	51.62	17.69	7.58	23.10
<i>Gb16301</i>	221	25 382.94	8.96	细胞核	46.22	21.08	7.30	25.41
<i>Gb19178</i>	246	28 197.98	6.26	细胞核	44.31	18.70	18.70	27.64
<i>Gb28587</i>	257	29 385.11	9.90	内质网	52.53	15.95	5.06	26.46
<i>Gb30604</i>	166	19 373.46	10.20	细胞核	51.20	14.46	7.83	26.51
<i>Gb33168</i>	209	24 345.65	5.53	细胞核	49.76	6.22	8.13	35.89
<i>Gb34103</i>	167	19 214.44	9.58	细胞核	70.06	10.78	5.39	13.77
<i>Gb36364</i>	252	29 170.87	8.88	细胞核	43.15	14.97	6.60	35.28
<i>Gb38883</i>	445	50 951.76	5.93	细胞核	47.64	9.21	6.25	36.63
<i>Gb39109</i>	199	22 576.68	6.22	细胞核	52.26	15.58	4.52	27.64
<i>Gb41549</i>	245	28 348.38	9.61	细胞核	52.24	13.88	5.71	28.16
<i>newGb184</i>	239	27 476.62	9.90	细胞核	53.56	16.32	7.53	22.59
<i>newGb2734</i>	253	29 623.10	9.43	细胞核	54.33	12.99	11.42	21.26

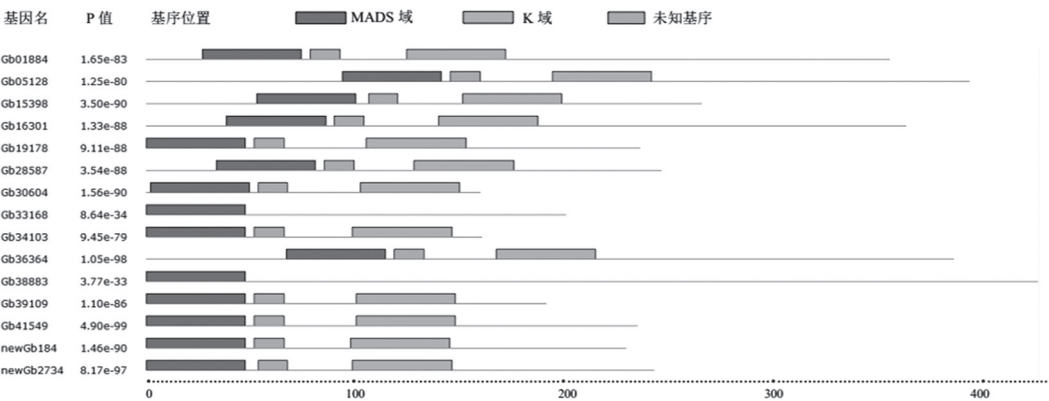


图1 银杏MADS-box基因家族保守基序
Fig.1 Conserved motifs of MADS-box genes in *G. biloba*

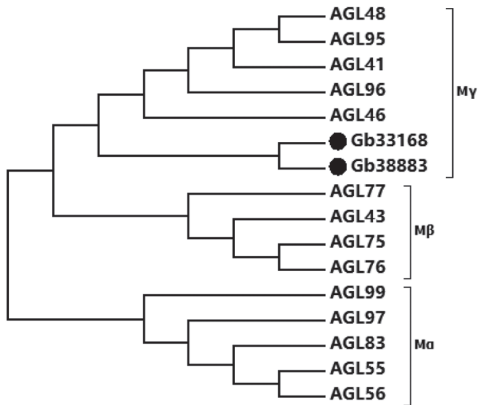


图2 银杏、拟南芥I型MADS-box基因的系统进化树
Fig.2 Phylogenetic relationships of I-type MADS-box genes
in *G. biloba* and *Arabidopsis thaliana*

发现GGM13亚族与SQUA、DEF/GLO亚族的进化关系较近; TM3亚族与SEP/AGL2亚族的进化关系较近, *Gb01884*、*Gb36364*、*Gb41549*与日本柳杉的TM3-like基因一起归类到TM3亚族, 分类到AG亚族的有为*Gb16301*和拟南芥*AtAGL1*、*AtAGL2*、*AtAGL5*等基因; 分类到TM8亚族的有*Gb19178*、*Gb30604*、*Gb39109*、*newGb184*、日本柳杉基因TM8-like以及油松基因*PtMADS4*、*PtMADS6*、*PtMADS7*; 分类到SVP亚族的基因有*Gb05128*、*Gb34103*、买麻藤GGM19、油松*PtMADS3*、*PtMADS5*以及拟南芥*AtAGL22*和*AtAGL24*; 未发现分类到SQUA亚族的A类功能基因和DEF/GLO亚族的B类功能基因。

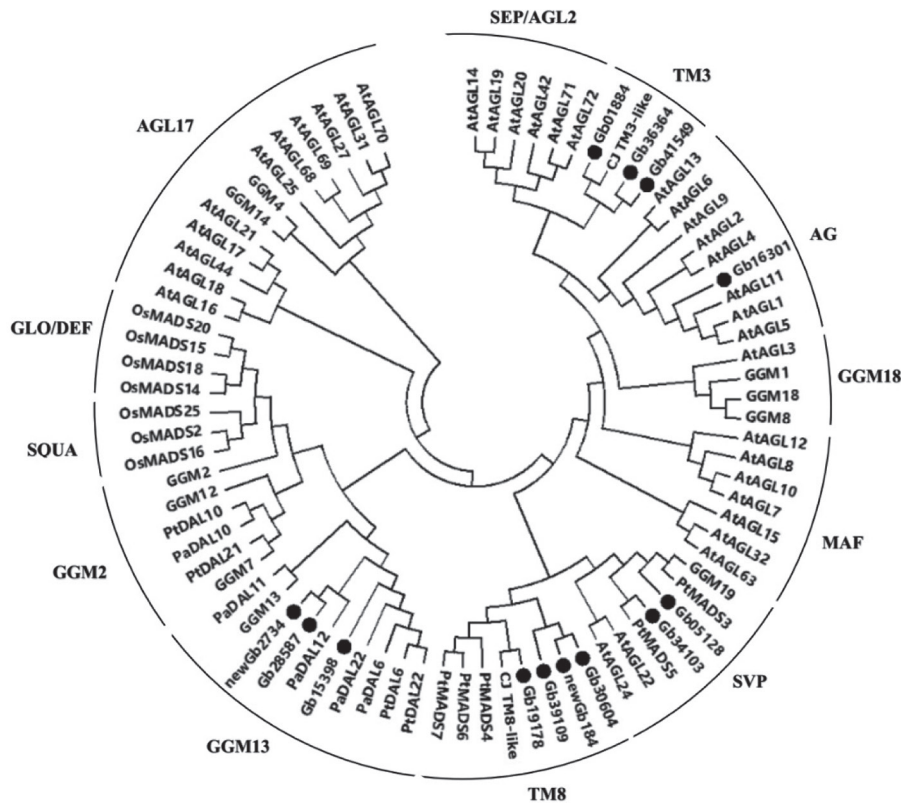


图3 银杏MIKC型MADS-box进化关系和亚族分类

Fig.3 Phylogenetic relationships and subgroup designations of MIKC-type MADS-box proteins in *G. biloba*

Gb and newGb: 银杏; At: 拟南芥; GGM: 买麻藤; Os: 水稻; Pa: 挪威云杉; Pt: 油松; CJ: 日本柳杉。

2.4 银杏MADS-box基因组织表达分析

将在花芽未分化期、花芽分化初期、花芽分化盛期、花芽分化末期的基因表达量数据进行整理, 绘制成表达模式聚类热图。根据表达模式聚类情况(图4), 将15个银杏MADS-box家族基因分为3类, I类包括*newGb184*、*Gb38883*和*newGb2734*, 其中*newGb184*在花芽分化初期和花芽分化末期表达, *Gb38883*和*newGb2734*只在花芽分化末期表达; II类包括在花芽未分化期、花芽分化初期、花芽分化盛期表达的*Gb28587*和*Gb33168*, III类包括在各个时期都表达的*Gb16301*、*Gb19178*、*Gb39109*等10个基因。

分析3类基因的均值聚类图可以发现, I类基因的平均表达量在花芽未分化期、花芽分化初期、花芽分化盛期较低, 在花芽分化末期达到最大值(图5-A); II类基因的平均表达量在花芽未分化期、花芽分化初期、花芽分化盛期、花芽分化末期中呈现出先增后减的趋势, 在花芽分化初期达到最

大值(图5-B); III类基因的平均表达量最高, 平均表达量在银杏花芽分化不同时期的变化不明显(图5-C)。

结合15个MADS-box家族基因在花芽未分化期、花芽分化初期、花芽分化盛期、花芽分化末期的表达聚类热图(图4)和均值聚类图(图5), 15个银杏MADS-box家族基因在银杏花发育不同阶段发挥着不同的作用, I类基因在花芽分化末期表达量较高, II类基因在花芽分化初期表达量较高, I类和II类基因可能在调控银杏开花过程中发挥作用; III类基因在各个时期的表达量稳定, 可能在银杏花发育整个过程发挥作用。

3 讨论

基因组测序技术的发展, 使得从全基因组层面, 更加系统和全面研究植物生长发育过程中的关键基因和家族基因成为现实。基于全基因组测序技术对被子植物MADS-box基因家族的研究工

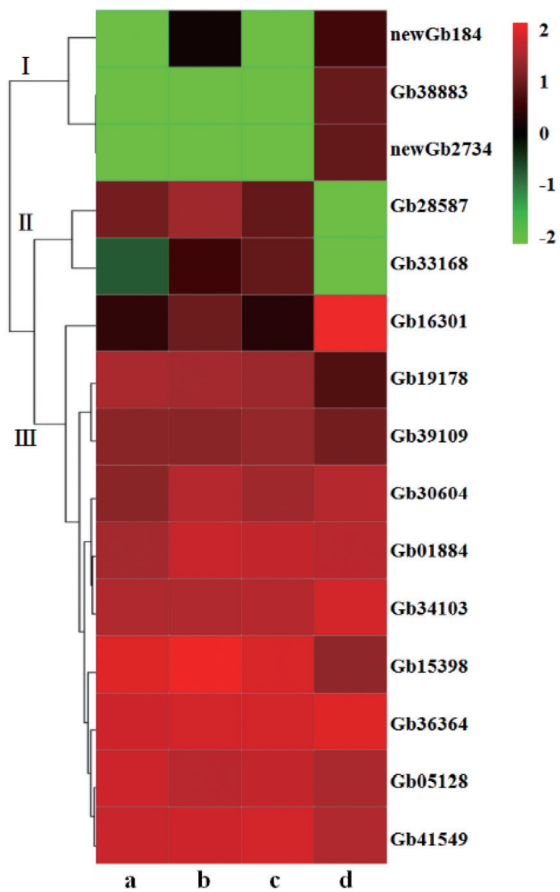


图4 银杏花发育MADS-box基因表达模式聚类热图
Fig.4 Heat map diagram of classification based on the expression patterns of the MADS-box genes in *G. biloba*

横坐标代表样品名称及样品的聚类结果, 纵坐标代表差异基因及基因的聚类结果。图中不同的列代表银杏短枝顶芽的不同分化时期, a: 花芽未分化期, b: 花芽分化初期, c: 花芽分化盛期, d: 花芽分化末期; 不同的行代表不同的基因; 颜色代表了基因在样品中的表达量水平 $\log_2(\text{fpkm}+1)$ 。

作相继完成, 拟南芥MADS-box家族有107个基因(Martinez-Castilla和Alvarez-Buylla 2003), 水稻MADS-box家族有75个基因(Arora等2007), 大豆MADS-box家族有106个基因(Shu等2013), 黄瓜MADS-box家族有41个基因(甘德芳等2012), 香蕉MADS-box家族有71个基因(张远森等2013), 毛果杨MADS-box家族有105个基因(Leseberg等2006)。通过转录组或者基因组数据鉴定的裸子植物中MADS-box基因的数量之间存在较大的差异: 挪威云杉253个、北美红杉16个、海岸松10个、罗汉松16个、买麻藤41个、越南篦齿苏铁12个(Gramzow等2014)。总体来说, MADS-box家族的功能研究在裸子植物中远没有被子植物(例如拟南芥)研究得清楚和深入。

在银杏的相关研究中, 银杏的ESTs数据库中的33条MADS-box相似序列以GBM1~33来命名, 其中的GBM5已被确定为拟南芥AGAMOUS亚族基因的同源基因, 并且在生殖器官和营养器官中都有表达(Jager等2003); 最近的研究发现, *GbMADS2*是AGAMOUS亚族同源基因, 并且与银杏花发育相关(Wang等2015), *GbMADS9*与银杏开花时间相关并且能够提高银杏非生物胁迫耐受性(Yang等2016), *GbSEP*是从银杏雌花中克隆出来并且与银杏花发育相关的E类功能的基因(Cheng等2016)。

本研究通过转录组测序得到的15个银杏MADS-box家族基因, 其中属于I型My亚族的有2个, 属于MIKC型基因有13个, 进一步分类到TM3、TM8、AG、GGM13、SVP亚族。MADS-box家族基因中I

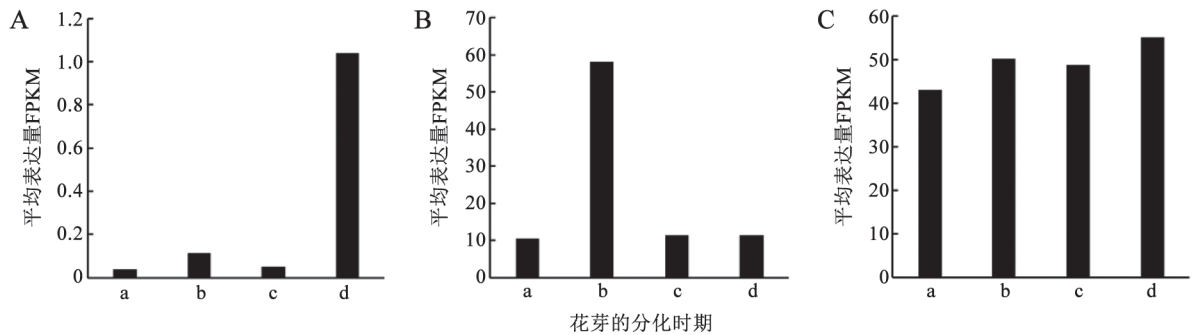


图5 银杏花发育MADS-box基因均值聚类图
Fig.5 Mean clustering graph of the MADS-box genes in *G. biloba*
A: I类基因的平均表达量; B: II类基因的平均表达量; C: III类基因的平均表达量; a: 花芽未分化期, b: 花芽分化初期, c: 花芽分化盛期, d: 花芽分化末期。

型基因的相关研究较少, Gramzow等(2014)在研究松柏类植物MADS-box家族基因时, 重建种子植物I型MADS-box基因系统进化树, 将I型MADS-box基因进一步分为M α 、M γ +M β 两个分支, 我们的结果与之一致, 说明在陆地植物中有两个古老的I型MADS-box基因的进化支, Gramzow等(2014)的转录组数据显示M α 进化支的基因在植物嫩芽组织中有表达, 而M γ +M β 进化支中的基因在球果和花芽中都有表达。银杏I型MADS-box基因中, *Gb38883*只在花芽分化末期表达, *Gb33168*在花芽分化初期表达, 2个基因的表达模式存在显著差异, 其具体功能还需进一步研究。

MADS-box家族基因中MIKC型基因的相关研究比较多, 植物中MIKC型基因主要分为以下几个亚族: *AG*、*AGL2 (SEP)*、*AGL6*、*AGL12*、*AGL15*、*AGL17*、*AP3/PI*、*GGM13*、*SVP*、*API (SQUA)*、*FLC*、*TM3*、*TM8*和*GpMADS4*。转录组测序中未发现与*SQUA*型A类功能基因和*DEF/GLO*型B类功能基因, 其原因主要是: *SQUA*型基因主要控制植物花瓣和萼片的发育, 目前为止还未能从任何的裸子植物中分离出来A类功能的基因(Zimin等2014); 裸子植物*DEF/GLO*型基因一般在雄花中表达(Sundstrom和Engstrom 2002)。转录组测序得到的*GGM13*亚族基因为*newGb2734*和*Gb28587*, 其中*newGb2734*在花芽分化末期表达, *Gb28587*在花芽分化末期不表达。有研究表明*GGM13*亚族基因主要参与调控雌性生殖器官的发育, 该亚族基因的具体表达模式和在花发育过程中的具体作用有待进一步研究; 非特异性表达的基因主要分类到*TM8*、*TM3*、*AG*、*SVP*亚族, 此类基因的功能还需要结合基因在不同组织中的表达情况进行进一步的分析。结合表达模式聚类热图和系统进化分析可以发现, *AG*亚族基因*Gb16301*在银杏花芽发育过程中的表达量较高, 在花芽分化末期的表达量尤为突出, 这与*AG*类功能基因在植物花发育过程中的作用一致。结合表达模式和系统进化分析结果可以推测: 在银杏15个MADS-box家族基因中, *Gb33168*、*Gb38883*、*Gb28587*和*newGb2734*可能在银杏开花调控中发挥重要作用, *Gb16301*在各时期都有较高的表达量, 可能在银杏花器官发育过程中发挥重要作用。

银杏作为单科单属的裸子植物, 在生长发育、

花发育、果实发育等方面都与被子植物以及其他裸子植物存在明显的不同, 本研究证明MADS-box家族基因在银杏生长发育、开花调控作用中表达量发生显著变化, 推测这类基因在银杏开花调控和花器官发育过程中也发挥了重要的作用。目前为止, 一些研究通过克隆已经得到了一定数量的银杏MADS-box家族基因, 但是对这些基因的功能和分类没有系统的认识。本研究通过对转录组测序得到的银杏MADS-box基因进行表达分析、生物信息学以及系统进化分析, 为进一步筛选银杏MADS-box家族基因, 并利用关键基因开展银杏早花品种的基因工程育种研究奠定理论基础。

参考文献(References)

- Arora R, Agarwal P, Ray S, et al (2007). MADS-box gene family in rice: genome-wide identification organization and expression profiling during reproductive development and stress. *BMC Genomics*, 8: 242
- Becker A, Theissen G (2003). The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Mol Phylogenet Evol*, 29 (3): 464–489
- Chamala S, Chanderbali AS, Der JP, et al (2013). Assembly and validation of the genome of the nonmodel basal angiosperm *Amborella*. *Science*, 342 (6165): 1516–1517
- Cheng S, Cheng J, Xu F, et al (2016). Molecular cloning and expression analysis of a putative E class MADS-box gene, *GbSEP*, from *Ginkgo biloba*. *J Anim Plant Sci*, 1 (26): 253–260
- Florea L, Song L, Salzberg SL (2013). Thousands of exon skipping events differentiate among splicing patterns in sixteen human tissues. *F1000Res*, 2: 188
- Gan DF, Ding F, Zhuang D, et al (2012). Genome-wide sequence characterization analysis of MADS-box transcription factor gene family in cucumber. *J Henan Agric Sci*, 26 (9): 1249–1256 (in Chinese with English abstract) [甘德芳, 丁飞, 庄丹等(2012). 黄瓜全基因组转录因MADS-box家族基因的序列特征分析. *核农学报*, (9): 1249–1256]
- Gramzow L, Weilandt L, Theissen G (2014). MADS goes genomic in conifers: towards determining the ancestral set of MADS-box genes in seed plants. *Ann Bot*, 114 (7): 1407–1429
- Guan R, Zhao Y, Zhang H, et al (2016). Draft genome of the living fossil *Ginkgo biloba*. *Giga Sci*, 5 (1): 49
- Guo M, Liu Y, Xu L, et al (2014). Cardiovascular pharmacological effects of traditional Chinese medicine *Ginkgo*: mechanisms and prospects. *Sci Sin Vitae*, 44 (6): 543–550 (in Chinese with English abstract) [郭明, 刘玥, 许琳

- 等(2014). 中药银杏制剂的心血管药理效应: 机制与展望. 中国科学:生命科学, 44 (6): 543–550]
- Immink RG, Gadella TJ, Ferrario S, et al (2002). Analysis of MADS box protein-protein interactions in living plant cells. *Proc Natl Acad Sci USA*, 99 (4): 2416–2421
- Jager M, Hassanin A, Manuel M, et al (2003). MADS-box genes in *Ginkgo biloba* and the evolution of the *AGA-MOVS* family. *Mol Biol Evol*, 20 (5): 842–854
- Leseberg CH, Li A, Kang H, et al (2006). Genome-wide analysis of the MADS-box gene family in *Populus trichocarpa*. *Gene*, 378 (15): 84–94
- Liu JH, Xu BH, Zhang J, et al (2010). The interaction of MADS-box transcription factors and manipulating fruit development and ripening. *Hereditas*, 32 (9): 893–902 (in Chinese with English abstract) [刘菊华, 徐碧玉, 张静等 (2010). MADS-box转录因子的相互作用及对果实发育和成熟的调控. 遗传, (09): 893–902]
- Martinez-Castilla LP, Alvarez-Buylla ER (2003). Adaptive evolution in the *Arabidopsis* MADS-box gene family inferred from its complete resolved phylogeny. *Proc Natl Acad Sci USA*, 100 (23): 13407–13412
- Masiero S, Colombo L, Grini PE, et al (2011). The emerging importance of type I MADS-box transcription factors for plant reproduction. *Plant Cell*, 23 (3): 865–872
- Melzer R, Wang YQ, Theissen G (2010). The naked and the dead: the ABCs of gymnosperm reproduction and the origin of the angiosperm flower. *Semin Cell Dev Biol*, 21 (1): 118–128
- Nystedt B, Street NR, Wetterbom A, et al (2013). The Norway spruce genome sequence and conifer genome evolution. *Nature*, 497 (7451): 579–584
- Shu Y, Yu D, Wang D, et al (2013). Genome-wide survey and expression analysis of the MADS-box gene family in soybean. *Mol Biol Rep*, 210 (4): 519–528
- Sundstrom J, Engstrom P (2002). Conifer reproductive development involves B-type MADS-box genes with distinct and different activities in male organ primordia. *Plant J*, 31 (2): 161–169
- Wang X, Chen J, Xu F, et al (2015). Molecular cloning and expression analysis of a MADS-box gene (*GbMADS2*) from *Ginkgo biloba*. *Not Bot Horti Agrobo*, 1 (43): 19–24
- Yang F, Xu F, Wang X, et al (2016). Characterization and functional analysis of a MADS-box transcription factor gene (*GbMADS9*) from *Ginkgo biloba*. *Sci Hortic*, 212 (2016): 104–114
- Zhang YS, Qin XM, Yan JP, et al (2013). Bioinformatics analysis of the mads-box gene family in banana. *Plant Physiol J*, 49 (10): 1030–1040 (in Chinese with English abstract) [张远森, 秦晓萌, 严金平等(2013). 香蕉MADS-box基因家族的生物信息学分析. 植物生理学报, 49 (10): 1030–1040]
- Zimin A, Stevens KA, Crepeau MW, et al (2014). Sequencing and assembly of the 22-gb loblolly pine genome. *Genetics*, 196 (3): 875–890

Gene expression and phylogenetic analysis of MADS-box family genes in *Ginkgo biloba*

DONG Jin-Jin^{1,2}, LIU Wei², LI Meng^{1,2}, WANG Jun-Yi², HE Chong-Dan², CAI Yu-Meng²,
WANG Qi-Ye², WANG Yi-Qiang^{1,2,*}

¹Key Lab of Non-wood Forest Nurturing and Protection of National Ministry of Education, Central South University of Forestry and Technology, Changsha 410004, China

²Hunan Provincial Key Laboratory for Forestry Biotechnology, Central South University of Forestry and Technology, Changsha 410004, China

Abstract: The MADS-box gene is a family of genes that encode transcriptional regulators in eukaryotes and plays an important regulatory role in floral development. In order to study the function of MADS-box gene in *G. biloba*, We sequenced the transcriptome of *G. biloba*, including four stages of flower bud differentiation, and screened the MADS-box family of *G. biloba*. Expression pattern, sequence signature, subcellular location and phylogenetic analysis were analyzed by the bioinformatics methods. The results showed that, 15 *G. biloba* MADS-box family genes were obtained. According to the different types of expression patterns of differentially expressed genes, the target gene could be divided into 3 types. Analysis of the coding sequence of the target gene showed that the main components of 15 *G. biloba* MADS-box proteins were α -helix and nonregular coil. The subcellular localization was mainly in the nucleus. All the expressed products contained MADS domain. The *newGb2734*, *Gb38883*, *Gb28587* and *Gb33168* in the MADS-box gene family of *G. biloba* might play an important role in the flowering regulation of *G. biloba*. *Gb16301* might be the key gene in the developmental process of *G. biloba*.

Key words: *Ginkgo biloba*; MADS-box gene family; expression analysis; phylogenetic analysis

Received 2017-12-22 Accepted 2018-05-15

This work was supported by the National Natural Science Foundation of China (31570682), and National Key Research and Development Programs in China (2017YFD0600701).

*Co-corresponding authors (wangyiqiang12@163.com).