

DOI: 10.11913/PSJ.2095-0837.23363

CSTR: 32231.14.PSJ.2095-0837.23363

柳文, 马永鹏, 应支萍, 杨琳祥, 谢孟, 马宏. 朱红大杜鹃 MADS-box 基因家族的全基因组鉴定与特征分析[J]. 植物科学学报, 2024, 42 (5): 624-633

Liu W, Ma YP, Ying ZP, Yang LX, Xie M, Ma H. Genome-wide identification and characterization of MADS-box gene family of *Rhododendron griersonianum* Balf. f. et Forrest[J]. *Plant Science Journal*, 2024, 42 (5): 624-633

朱红大杜鹃 MADS-box 基因家族的全基因组鉴定与特征分析

柳文^{1,2}, 马永鹏³, 应支萍⁴, 杨琳祥^{1,2}, 谢孟^{1,2}, 马宏^{1,5*}

- (1. 中国林业科学研究院高原林业研究所, 昆明 650233; 2. 南京林业大学风景园林学院, 南京 210037; 3. 云南省极小种群野生植物综合保护重点实验室, 中国科学院昆明植物研究所, 昆明 650201; 4. 西南林业大学园林园艺学院, 昆明 650224; 5. 国家林业和草原局资源昆虫培育与利用重点实验室, 昆明 650233)

摘要: MADS-box 是在植物生长发育过程中扮演重要角色的一类转录因子, 特别是花器官形成、开花时间控制及果实发育与成熟等过程。本研究基于朱红大杜鹃 (*Rhododendron griersonianum* Balf. f. et Forrest) 全基因组测序数据, 利用生物信息学方法从中鉴定出 81 个 MADS-box 基因, 并进行了分析。根据系统发育关系和蛋白结构将 MADS-box 分为两类: Type-I 型包含 24 个基因, Type-II 型包含 57 个基因, 这些基因不均匀地分布于 12 条染色体上; 81 个 MADS-box 基因中, 存在 6 对片段复制和 1 对串联复制基因, 且它们经历了纯化选择作用; 同时, 基因启动子区域含有光响应、植物生长、激素反应和胁迫响应等顺式作用元件。综上, 本研究鉴定了朱红大杜鹃的 MADS-box 家族转录因子, 为深入研究其 MADS-box 蛋白的生物学功能提供了基础。

关键词: 朱红大杜鹃; MADS-box 基因家族; 生物信息学; 基因结构

中图分类号: Q943.2

文献标识码: A

文章编号: 2095-0837 (2024) 05-0624-10

Genome-wide identification and characterization of MADS-box gene family of *Rhododendron griersonianum* Balf. f. et Forrest

Liu Wen^{1,2}, Ma Yongpeng³, Ying Zhiping⁴, Yang Linxiang^{1,2}, Xie Meng^{1,2}, Ma Hong^{1,5*}

- (1. Institute of Highland Forest Science, Chinese Academy of Forestry, Kunming 650233, China; 2. College of Landscape Architecture, Nanjing Forestry University, Nanjing 210037, China; 3. Key Laboratory of Comprehensive Conservation for Extremely Small Populations of Wild Plants in Yunnan Province, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China; 4. College of Landscape Architecture and Horticulture, Southwest Forestry University, Kunming 650224, China; 5. Key Laboratory of Resource Insect Cultivation and Utilization, State Forestry and Grassland Administration, Kunming 650233, China)

Abstract: MADS-box genes play an important role in plant growth and development, especially in processes such as floral organ formation, flowering time regulation, and fruit development and ripening. Based on whole-genome sequencing data of *Rhododendron griersonianum* Balf. f. et Forrest, 81 MADS-box genes were identified and analyzed using bioinformatics methods. Phylogenetic analysis and protein structure classification divided these genes into two classes, including 24 genes in Type-I and 57 genes in Type-II. The MADS-box genes were unevenly distributed across 12 chromosomes, with no genes located on chromosome 2. Among the 81 MADS-box genes, six pairs showed segmental duplications and one pair showed tandem duplication, all of which have undergone purifying selection. The promoter regions of the MADS-box genes contained

收稿日期: 2023-12-04, 接受日期: 2024-03-12。

基金项目: 云南省重大科技计划专项 (202302AE090018, 202202AE090012); 云南省乡村振兴科技专项 (202404BI090014); 兴滇人才支持计划项目 (YNWRQNBK-2019-010); 极小种群野生植物保护项目 (2022SJ07X-03)。

作者简介: 柳文 (1996-), 男, 硕士研究生, 研究方向为杜鹃属植物保护生物学 (E-mail: 1585957360@qq.com)。

* 通信作者 (Author for correspondence. E-mail: hortscience@163.com)。

elements involved in light response, plant growth, hormone response, and stress response. Overall, the identification of MADS-box gene family members provides a reliable reference for further studies on the biological functions of MADS-box proteins in *R. griersonianum*

Key words: *Rhododendron griersonianum*; MADS-box gene family; Bioinformatics; Gene structure

朱红大杜鹃 (*Rhododendron griersonianum* Balf. f. et Forrest) 是杜鹃花属 (*Rhododendron*) 朱红大杜鹃亚族的唯一物种, 也是中国特有物种。全世界杜鹃花品种中超过 10% 的亲本都与其有着直接或间接的关系, 是名副其实的明星物种。然而, 该物种自 1924 年被发表后, 在野外几乎销声匿迹, 直到 1964 年再次被发现^[1]。由于人类活动频繁, 其生境遭到严重破坏。2020 年, 经本课题组多次实地调查发现, 其仅残存 322 株于云南腾冲市, 为典型的极小种群野生植物^[2]。该物种已被列入 *The Red List of Rhododendrons*^[3] 和中国高等植物受威胁物种名录^[4] 中, 均被评为极度濒危 (CR) 等级。朱红大杜鹃具有株形优美、花色鲜艳、花期长且易驯化等优点, 研究其花器官分化具有重要意义, 而目前与其相关的研究仅限于全基因组测序与组装等方面, 对其花发育方面的研究尚未见报道。通常, 花器官的分化主要由 MADS-box 基因家族控制, 且该基因家族在果实发育中也发挥作用, 但迄今为止, 尚未见朱红大杜鹃 MADS-box 基因家族的相关报道。

MADS-box 基因家族最初被鉴定为同源异型基因, 是植物中研究最广泛的转录因子家族之一^[5,6], 同时也广泛存在于动物和真菌中^[7], 其命名来自酵母 (*Saccharomyces cerevisiae*) *MCMI* 基因、拟南芥 (*Arabidopsis thaliana* (L.) Heynh.) *AGAMOUS* 基因、金鱼草 (*Antirrhinum majus* L.) *DEFTCI* 基因及人类 *SRF* 基因的首字母缩写^[8]。MADS-box 转录因子的主要特征为具有 58~60 个氨基酸的 DNA 结合结构域, 位于蛋白质的 N-端区域, 称为 MADS 结构域^[9]。根据其蛋白的结构域特征和系统发育关系, 可将植物 MADS-box 家族划分成两大类, 即 Type-I 型和 Type-II 型^[10]。基于 SRF-like 结构域的差异, Type-I 型可进一步划分成 $M\alpha$ 、 $M\beta$ 和 $M\gamma$ 等 3 个亚族^[11], 这些基因的 N 端都包含一个约 180 bp 的 MADS-box 保守区域, 且明确缺少 K-box 结构域 (Keratin-like domain);

Type-II 型不仅含 MEF2-like MADS 结构域, 还含有约 70 个氨基酸的 K-box 结构域、约 30 个氨基酸的 I 结构域 (Intervening domain) 和长度可变的 C 结构域 (Carboxy-terminal domain), 故又被称为 MIKC 型 MADS-box 基因^[12]; 基于 II 型结构域所对应的基因结构以及 K-box 结构域的不同, 可将 MIKC 划分为 $MIKC^+$ 型和 $MIKC^C$ 型^[13-15]。此外, K 结构域介导 MADS-box 蛋白的二聚化, 并已被证明参与高级复合物的形成; I 结构域在 DNA 结合二聚体的形成中提供特异性; C 结构域是一个高度多样的区域, 但已有研究证实 C 结构域在 MADS-box 蛋白相互作用和转录激活中具有关键作用^[16]。多项研究发现, 大多数植物中的 I 型 MADS-box 基因比 II 型基因更快地经历了基因复制过程, 这可能是由于节段基因复制的频率更高和纯化选择更弱所致^[9]。

MADS-box 基因家族在植物生长发育过程中承担多种功能, 在参与调控植物非生物胁迫响应、开花时间控制、分生组织和花器官确定、果实成熟、胚发育以及根和叶等营养器官发育中具有突出功能^[17-21], 且在植物生长发育过程中, I 型基因作用一般弱于 II 型^[22]。在植物中发现的第一个 MADS-box 基因是参与花发育的 II 型基因, II 型基因还控制开花时间^[23]。目前, 许多物种的全基因组测序已完成, 对 MADS-box 基因功能的理解已更加深刻。在拟南芥^[11]、茄子 (*Solanum melongena* L.)^[24]、小麦 (*Triticum aestivum* L.)^[25]、水稻 (*Oryza sativa* L.)^[26]、毛果杨 (*Populus trichocarpa* Torr. & Gray)^[27] 和菠萝 (*Ananas comosus* (L.) Merr.)^[28] 等植物中, MADS-box 基因已得到广泛研究, 如 Ruelens 等^[29] 研究发现, 二穗短柄草 (*Brachypodium distachyum* (L.) Beauv.) 的 MADS-box 在调控花发育的“ABCDE 模型”中扮演着非常重要的角色, E 类基因 (*SEP1/2*、*SEP3* 和 *SEP4*) 可以形成所有花器官发育所必需的 MADS-box 蛋白复合物; Ferrario 等^[30] 在圆叶

牵牛 (*Pharbitis purpurea* (L.) Voisgt) 的研究中发现, *FBP2* 基因能调控开花时间, 且可以调控花萼片的转化。拟南芥 *FLC* 亚族基因不仅可以调节其开花时间, 还可以改变或消除不同生态型的春化作用^[23]; 高玮林等^[31] 在对枣 (*Ziziphus jujuba* Mill.) E 类 MADS-box 基因的研究中发现, *SEP1/2*、*SEP3* 和 *SEP4* 同源基因在花发育的整个过程中均有表达, 但不同基因在功能和调控模式上均有略微差异, 表现出一定的功能分化。在拟南芥中, MADS-box 家族的 *AGAMOUS-LIKE6* (*AGL6*) 基因在调控其侧生器官发育和开花时间中起关键作用, 且在花发育过程中, 在所有花器官的近端区域和发育的胚珠中均能检测到 *AGL6*^[20]。孙福辉等^[32] 对马银花 (*Rhododendron ovatum* (Lindl.) Planch. ex Maxim.) MADS-box 基因家族的研究结果表明, *AP1*、*AP3/PI*、*AG*、*SEP* 和 *MIKC**分支基因均在花器官中特异表达, 并构建了花的发育模型。

本研究基于课题组前期获得的朱红大杜鹃全基因组数据^[2], 使用生物信息学方法, 对其 MADS-box 基因的理化性质、基因结构、染色体定位、共线性和系统进化关系等进行分析, 研究结果旨在为 MADS-box 基因家族成员的生物学功能分析奠定基础。同时, 理解朱红大杜鹃 MADS-box 基因分子的进化特点, 为杜鹃花科植物的花型遗传改良和育种工作提供参考。

1 材料与方法

1.1 朱红大杜鹃 MADS-box 基因的鉴定及理化性质分析

利用 BLAST 和 HMMER search 方法鉴定 MADS-box 基因。朱红大杜鹃基因组数据来源于本课题组的全基因组测序数据^[2], 拟南芥基因组数据下载自 TAIR 网站 (<https://www.arabidopsis.org/>)。从 Pfam 网站 (<http://pfam.xfam.org/>) 下载与 MADS 结构域 (PF00319) 对应的隐马尔可夫模型 (Hidden Markov model, HMM) 文件, 用 HMMER search 方法在朱红大杜鹃的蛋白序列中检索 MADS-box 蛋白序列, 得出一次结果, 然后以拟南芥的 MADS-box 蛋白序列为索引, 利用 TBtools^[33] 软件进行本地 BLAST 比对。将两种方

法鉴定的结果合并取交集, 将比对后的 MADS-box 基因家族蛋白序列, 利用 NCBI 的 CDD 数据库 (<https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi>) 和 SMART 数据库 (<https://www.ncbi.nlm.nih.gov/Structure/>), 对其保守结构域进行逐个鉴定分析, 剔除结构域不完整、重复的序列, 最终获得 81 个 MADS-box 转录因子序列。通过在线软件 ExPasy (<https://web.expasy.org/protparam/>) 分析其编码蛋白序列的相对分子质量、理论等电点等基本理化性质。

1.2 系统进化树构建、蛋白基序及保守结构域分析

根据文献获得拟南芥 MADS-box 蛋白序列的登录号^[11], 在 TAIR 网站下载拟南芥 107 个 MADS-box 蛋白序列文件。利用 MAGE 11 软件中的 Clustal W 程序, 对朱红大杜鹃和拟南芥的 MADS-box 蛋白序列进行多序列比对。采用邻近法 (Neighbor-joining method, NJ) 构建系统发育树, 参数设置为: 1 000 Bootstrap method、P-distance 模型和 Pairwise deletion, 使用 Itol 在线工具 (<https://itol.embl.de/upload.cgi>) 对进化树进行美化。将朱红大杜鹃的 MADS-box 基因家族全部进行提取, 采用 MEME 在线工具 (<http://meme-suite.org/meme/tools/meme>), 分析其蛋白基序, 基序的最大数目设置为 10。利用 NCBI 数据库的 Batch CD-Search 工具, 分析 MADS-box 蛋白序列的保守结构域, 最后使用 TBtools 软件的 Gene structure view (Advanced) 插件对蛋白基序及保守结构域进行可视化。

1.3 MADS-box 基因的结构分析

在朱红大杜鹃的 GFF3 文件中提取 MADS-box 基因序列的注释信息, 使用 TBtools 的 Amazing optional gene viewer 模块分析内含子与外显子的组成情况, 利用 TBtools 软件的 Gene structure view (Advanced) 插件进行可视化处理。

1.4 MADS-box 基因的染色体分布

在全基因组文件中提取 MADS-box 基因在染色体上的定位信息, 利用 TBtools 软件的 Gene density profile 插件对染色体颜色进行填充, 使用 Map Gene2 Chromosome v2 (http://mg2c.iask.in/mg2c_v2.0/) 在线工具绘制 MADS-box 基因在染色体上的位置。

1.5 MADS-box 启动子顺式作用元件分析

使用 TBtools 软件, 在基因组文件中提取 MADS-box 基因转录起始位点上游 2 000 bp 的序列作为基因的启动子区, 利用 Plant CARE 在线工具 (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html>) 进行分析, 预测启动子区中的顺式作用元件, 并对分析结果进行统计, 将整理后的数据利用 TBtools 软件进行可视化。

1.6 MADS-box 的共线性分析

利用基因组数据库 TAIR 和植物数据库 Ensembl Plants 分别下载拟南芥和毛果杨的 MADS-box 基因组数据, 使用 TBtools 软件的多重共线性扫描工具包 (One step MCScanX) 对 3 种植物进行全基因组共线性分析, 获得 *ctl*、*gff* 和 *collinearity* 文件。使用 TBtools 软件获得朱红大杜鹃各染色体长度文件, 找出 MADS-box 基因中的高亮基因及所在染色体位置, 利用 Advanced Circos 软件绘制其种内共线性关系图; 根据朱红大杜鹃、拟南芥及毛果杨的 MADS-box 基因 ID, 从共线性结果中筛选出三者间的共线性基因对, 利用 TBtools 软件的双系统图程序绘制种间共线性图。

1.7 MADS-box 的基因复制事件

根据 MADS-box 基因两两之间的 BLASTP 比对结果, 获取基因复制的同源事件, 利用 MCScanX 进行全基因组共线性分析, 鉴定片段复制, 然后利用 TBtools 软件计算复制基因的非同义替换率 (*Ka*) 和同义替换率 (*Ks*), 并进一步计算 *Ka/Ks* 的比值, 分析环境选择压力。判断基因复制事件有两个条件: 一是两个基因序列相似性大于 75%, 二是相似序列长度大于较长序列的 75%^[34]。

2 结果与分析

2.1 MADS-box 的鉴定及理化性质分析

基于朱红大杜鹃的全基因组数据, 进行初步筛选, 获得 93 个 MADS-box 基因家族候选基因。利用 NCBI 的 CDD 库和 SMART 对保守结构域进行进一步验证, 剔除了 12 个不含 MADS-box 结构域或重复的基因, 最终获得 81 个 MADS-box 基因家族成员, 将其命名为 *RhMADS1~RhMADS81*

(附表 1¹⁾)。

对 81 个 MADS-box 的蛋白序列进行理化性质分析, 发现不同序列之间差异较大 (附表 1)。MADS-box 蛋白序列的长度为 67~575 aa, 最长的是 RhMADS66, 最短的是 RhMADS30, 平均为 225 aa。理论等电点在 4.76~10.66, 平均等电点为 8.12, 等电点最大的为 RhMADS35, 最小的为 RhMADS49, 其中 58 个蛋白的等电点大于 7。蛋白分子量在 7 769.95~64 848.69 Da, 最大的为 RhMADS66, 最小的为 RhMADS30。在蛋白稳定性方面, 不稳定系数小于 40, 则该蛋白具有稳定性。在 81 个 MADS-box 蛋白序列中, 只有 11 个蛋白的不稳定系数小于 40, 其余的均大于 40。脂肪系数最大的是 RhMADS2, 为 118.08, 最小的是 RhMADS26, 为 56.50。在亲水性方面, 平均亲水系数大于 0, 表示蛋白有一定的疏水性, 小于 0, 则表示具有亲水性; 本研究发现, 除了 3 个蛋白序列的平均亲水性大于 0, 其余 MADS-box 蛋白均小于 0, 为亲水性蛋白。

2.2 MADS-box 系统发育树的构建

对朱红大杜鹃和拟南芥的 MADS-box 蛋白序列进行比对, 采用邻接法, 构建系统进化树, 并用 Itol 工具进行美化。结果显示, 81 个基因中, 有 24 个是 I 型基因, 其中 15 个为 *Mα* 亚族, 3 个为 *Mβ* 亚族, 6 个为 *Mγ* 亚族; II 型 MADS-box 基因有 57 个, 其中 MICK⁺ 型亚族有 5 个, 而 MICK^c 型又可划分为 13 个亚族, 除了 OsMADS32、AGL15 和 FLC 外, 其他亚族均含有 MADS-box 基因, 其中 SVP 亚族含有的成员最多, 为 13 个, AGL12 和 SEP 最少, 均为 2 个 (图 1)。由于 OsMADS32 亚族为单子叶植物所特有, 本研究在朱红大杜鹃中未发现。

2.3 MADS-box 蛋白基序及保守结构域分析

本研究从朱红大杜鹃 MADS-box 蛋白共鉴定出 10 个保守基序 (附图 1²⁾), 命名为 Motif 1~10。Motif 1 是 MADS 域最明显的特征, 仅 RhMADS63 不含该基序。II 型中仅 RhMADS11 和 RhMADS77 不含 Motif 2; Motif 3 和 Motif 4 是 II 型 MADS-box 的典型结构; 而 Motif 5~7 和 Motif 10 仅存在于 II 型 MADS-box 蛋白序列中, Mo-

1~2) 如需查阅附件内容请登录《植物科学学报》网站 (<http://www.plantscience.cn>) 查看本期文章。

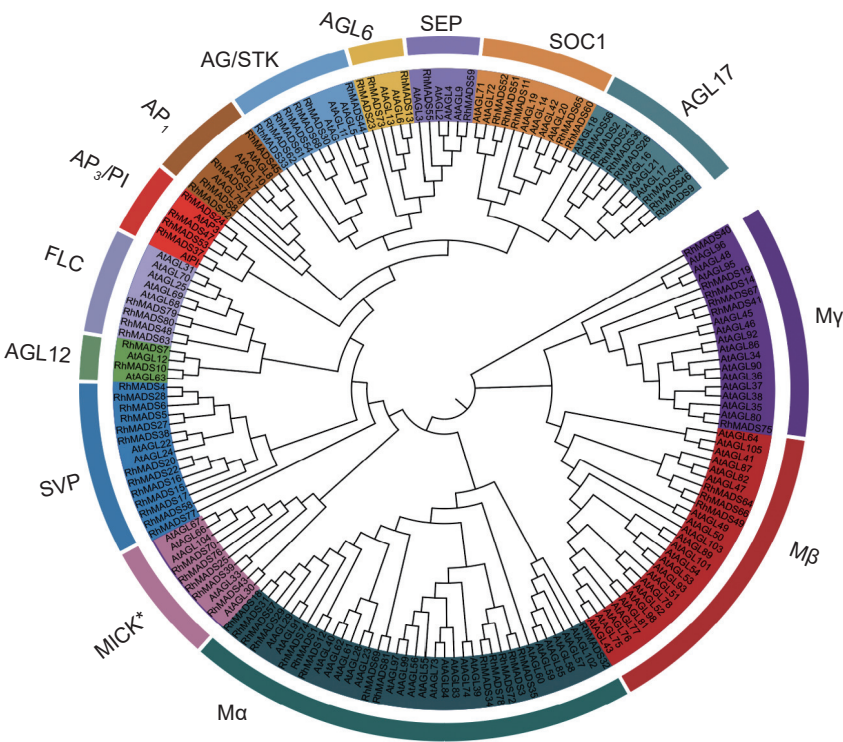


图 1 拟南芥与朱红大杜鹃 **MADS-box** 基因家族的系统进化分析
Fig. 1 Phylogenetic analysis of **MADS-box** family in *Arabidopsis thaliana* and *Rhododendron griersonianum*

tif 8 和 Motif 9 仅存在于 I 型蛋白序列中。同一亚族的 **MADS-box** 具有相似的 Motif 结构，而 II 型的结构组成更加多样化，且存在一些亚家族特异性的保守基序，如 SVP 蛋白序列特有的 Motif 5（附图 1）。对 **MADS-box** 蛋白的保守结构域进行分析发现，所有蛋白序列均含有 **MADS**、**MADS_SRF** 或 **MADS_MEF2** 结构域（附图 1），而 **K-box** 结构域仅存在于 II 型蛋白序列中。此外，朱红大杜鹃 **MADS-box** 蛋白还含有其他少数几个结构域。

2.4 **MADS-box** 基因结构分析

对朱红大杜鹃 **MADS-box** 基因的 CDS 序列和基因组 DNA 序列进行分析发现（附图 2¹⁾），不同基因的长度和外显子的数目及长度差异较大。外显子数目为 1~12 个，其中，*RhMADS39* 的外显子最多，为 12 个；I 型 **MADS-box** 基因有 23 个不含内含子，仅 *RhMADS14* 含有 1 个内含子；II 型 **MADS-box** 基因中有 21 个具有内含子，其中 *RhMADS43* 内含子最多，为 11 个；其余基因均没有内含子。

2.5 **MADS-box** 基因的染色体分布

染色体分布分析结果表明（附图 3²⁾），朱红大杜鹃 **MADS-box** 基因家族 81 个成员不均匀地分布在 12 条染色体上。其中，第 7 号染色体上分布数量最多，为 14 个，而 2 号染色体上没有分布。染色体长度与 **MADS-box** 基因的分布没有相关性，最长的第 4 号染色体上仅分布有 5 个基因，而最短的第 12 号染色体上有 8 个。

2.6 **MADS-box** 启动子顺式作用元件分析

本研究提取 **MADS-box** 转录因子起始位点上游 2 000 bp 的序列作为基因的启动子区，共鉴定出 48 种顺式作用元件（附图 4³⁾ 和附表 2⁴⁾）。根据其功能注释，可分为：光响应作用元件（23 种）、激素反应作用元件（11 种）、胁迫响应作用元件（6 种）以及植物生长作用元件（8 种）。在 **MADS-box** 基因启动子中共鉴定出 876 个光响应元件，主要包括 G-box、Box4、3-AF1 binding site、Chs-Unit 1 m1 等元件，其中，G-box 和 Box4 最多，且每个 **MADS-box** 至少含有 161 个与光响应相关的元件；同时，在 **MADS-box** 基因启

1~4) 如需查阅附件内容请登录《植物科学学报》网站 (<http://www.plantscience.cn>) 查看本期文章。

3 讨论

3.1 MADS-box 基因家族的鉴定与特征

本研究基于朱红大杜鹃的全基因组测序数据, 利用生物信息学方法鉴定出 81 个 MADS-box 基因。根据拟南芥中的分类方法, 将其分为 Type-I (24 个) 和 Type-II 型 (57 个) 两类。两种类型的成员数目显著差异, 可能 II 型基因较 I 型功能复杂, 且进化程度更高。目前, 在拟南芥^[11]、大豆 (*Glycine max* (L.) Merr.)^[35]、烟草 (*Nicotiana tabacum* L.)^[36]、番茄 (*Lycopersicon esculentum* Mill.)^[37]、马银花^[32]、萝卜 (*Raphanus sativus* L.)^[38]、菊花 (*Dendranthema morifolium* (Ramat.) Tzvel.)^[39]、中华猕猴桃 (*Actinidia chinensis* Planch.)^[40] 和芝麻 (*Sesamum indicum* L.)^[41] 等植物中分别鉴定出了 107、95、183、131、77、144、98、68 和 37 个 MADS-box 家族成员, 各物种 MADS-box 基因的数量有差异, 可能与其基因组大小及全基因组复制水平不同有关。理化性质分析结果显示, 不同家族成员的蛋白序列长度、等电点、亲水性等差异均比较大, 可能与其相对应的功能有关。本研究对朱红大杜鹃 MADS-box 的理化性质、系统进化、基因结构、Motif、保守结构域及顺式作用元件等方面进行了分析, 研究结果可为进一步探讨该基因家族的功能提供参考。

MADS-box 是植物转录因子中研究最为广泛的家族之一, 在植物生长发育、胁迫反应以及次生代谢调控等方面起着重要作用^[42], 尤其是在植物花序、花发育和果实成熟过程中表达最为明显。在植物花器官发育的“ABCDE”模型中, 所涉及的基因多为 MADS-box 基因家族, 这一家族在调控花器官的形成与分化过程中扮演着至关重要的角色, 其中不同花序中的特定表达过程受 II 型基因调节^[28]。譬如拟南芥的 SVP 亚族基因是花完整发育的网络调控中心, SOC1 亚族基因可以决定开花时间, SOC1 亚族中的 MADS-box 基因为植物开花信号通路的整合因子, AGL17 亚族在调控花发育过程中也发挥重要作用, 但拟南芥 SVP 亚族仅有两个成员^[43]。本研究发现, 朱红大杜鹃的 SVP 亚族成员多达 13 个, AGL17 和 SOC1 亚族成员分

别为 9 个和 5 个。因此, 朱红大杜鹃的 SVP 亚族可能涉及更加复杂的花发育调控, 而 AGL17 和 SOC1 亚族则可能协同调控花发育和开花时间。拟南芥的 AGL6 亚族基因在其侧生器官发育和开花时间的调控中同样不可或缺, 其在茎生叶原基的远轴和近端区域以及花苞片中均有表达^[20]。朱红大杜鹃的 AGL6 亚族仅有 3 个成员, 可能在进化过程中发生了基因丢失。由于 MADS-box 转录因子在控制植物开花过程中非常重要, 研究该基因家族可能有助于理解朱红大杜鹃濒危的原因。此外, MADS-box 基因启动子存在较多的光反应、激素反应及胁迫适应相关的顺式作用元件, 进一步表明该家族基因在植物的整个生命周期中, 特别是在其生长发育阶段以及面对各种环境胁迫时, 发挥着不可或缺的关键作用。

3.2 MADS-box 家族的基因结构及保守基序

内含子的数量越多、长度越大, 基因剪接的方式就越多样化, 因而更易影响基因表达, 从而调控蛋白质的活性。同时, 内含子丢失的基因具有较高的同义替换率, 且丢失的内含子比现存的内含子具有更高的突变率^[44, 45]。朱红大杜鹃 I 型 MADS-box 的结构较为简单, 其中 23 个成员没有内含子, 仅有 *RhMADS14* 含有 1 个内含子, 这可能是 MADS-box 基因家族多样化过程中经历了多个内含子的丢失所致; 而 II 型具有较多的内含子。蛋白基序分析结果表明, 同一组中含有最保守的基序, 且这些保守基序在群体特异性功能中发挥着重要作用。MADS-box 基因均含有 1 个 MADS 保守结构域, 而 II 型成员还含有 1 个独有的 K-box 结构域, 意味着 II 型基因功能更复杂, 与拟南芥^[11]、茄子^[24] 和马铃薯 (*Solanum tuberosum* L.)^[46] 中的研究结果相似, 可为朱红大杜鹃 MADS-box 的进化关系提供更多的线索。顺式作用元件分析结果表明, 启动子区主要的作用元件可分为 4 大类型: 光响应元件、胁迫响应元件、激素反应和植物生长元件。

3.3 MADS-box 基因家族的进化分析

在进化过程中, 基因复制产生的新基因可能发挥着新的功能, 同时也提高植物适应新环境的能力^[47]。本研究中, 片段重复 (6 对) 和串联重复 (1 对) 基因对均属于 II 型和 Ma 亚家族, 属于 II

型亚家族的基因占60%，这与浙江红山茶 (*Camellia chekiangoleosa* Hu)^[48]相似，Ⅱ型 MADS-box 的片段重复比Ⅰ型频率高，可能是Ⅱ型基因出现和分化的时间更早，且Ⅱ型基因的结构也更加保守。MADS-box 在7号染色体上分布最多，共有14个基因，这可能是由串联重复所致。基因间的共线性是理解两者亲缘关系的重要途径，朱红大杜鹃与拟南芥之间的共线性基因为25对，而与毛果杨间的共线性基因多达69对，表明朱红大杜鹃与毛果杨的共线性和同源性更高，亲缘关系可能更近。Ka/Ks 值被认为是研究蛋白质编码基因的选择压力或强度的重要指标^[34]。本研究，朱红大杜鹃中存在共线性基因对的 MADS-box 基因的 Ka/Ks 值均小于1，表明这些基因在进化过程中受到纯化选择作用。

4 结论

本研究基于课题组前期建立的朱红大杜鹃全基因组数据库，对其 MADS-box 家族进行了全基因组鉴定，并对其理化性质、系统进化、基因结构、保守结构域、顺式作用元件以及种内与种间共线性等进行了分析，研究结果将有助于理解 MADS-box 家族在朱红大杜鹃花分化过程中的作用，并为其他植物的相关基因家族研究提供参考。

参考文献:

- [1] Liu DT, Sun WB, Ma YP, Fang ZD. Rediscovery and conservation of the critically endangered *Rhododendron griersonianum* in Yunnan, China[J]. *Oryx*, 2019, 53 (1): 14.
- [2] Ma H, Liu YB, Liu DT, Sun WB, Liu XF, et al. Chromosome-level genome assembly and population genetic analysis of a critically endangered *Rhododendron* provide insights into its conservation[J]. *Plant J*, 2021, 107 (5): 1533–1545.
- [3] Gibbs D, Chamberlain D, Argent G. The Red List of Rhododendrons[M]. Richmond: Botanic Gardens Conservation International, 2011: 51.
- [4] 覃海宁, 杨永, 董仕勇, 何强, 贾渝, 等. 中国高等植物受威胁物种名录[J]. *生物多样性*, 2017, 25 (7): 696–744.
Qin HN, Yang Y, Dong SY, He Q, Jia Y, et al. Threatened species list of China's higher plants[J]. *Biodiversity Science*, 2017, 25 (7): 696–744.
- [5] De Folter S, Angenent GC. Trans meets cis in MADS science[J]. *Trends Plant Sci*, 2006, 11 (5): 224–231.
- [6] Nam J, dePamphilis CW, Ma H, Nei M. Antiquity and evolution of the MADS-box gene family controlling flower development in plants[J]. *Mol Biol Evol*, 2003, 20 (9): 1435–1447.
- [7] Theissen G, Becker A, Di Rosa A, Kanno A, Kim JT, et al. A short history of MADS-box genes in plants[J]. *Plant Mol Biol*, 2000, 42 (1): 115–149.
- [8] Shan HY, Zhang N, Liu CJ, Xu GX, Zhang J, et al. Patterns of gene duplication and functional diversification during the evolution of the AP1/SQUA subfamily of plant MADS-box genes[J]. *Mol Phylogenet Evol*, 2007, 44 (1): 26–41.
- [9] Nam J, Kim J, Lee S, An G, Ma H, Nei M. Type I MADS-box genes have experienced faster birth-and-death evolution than Type II MADS-box genes in angiosperms[J]. *Proc Natl Acad Sci USA*, 2004, 101 (7): 1910–1915.
- [10] Alvarez-Buylla ER, Pelaz S, Liljgren SJ, Gold SE, Burgeff C, et al. An ancestral MADS-box gene duplication occurred before the divergence of plants and animals[J]. *Proc Natl Acad Sci USA*, 2000, 97 (10): 5328–5333.
- [11] Pařenicová L, de Folter S, Kieffer M, Horner DS, Favalli C, et al. Molecular and phylogenetic analyses of the complete MADS-Box transcription factor family in *Arabidopsis*: new openings to the MADS world[J]. *Plant Cell*, 2003, 15 (7): 1538–1551.
- [12] Qin Y, Zhu GP, Li FD, Wang L, Chen C, Zhao H. MIKC-Type MADS-box gene family discovery and evolutionary investigation in Rosaceae plants[J]. *Agronomy*, 2023, 13 (7): 1695.
- [13] Henschel K, Kofuji R, Hasebe M, Saedler H, Münster T, Theissen G. Two ancient classes of MIKC-Type MADS-box genes are present in the moss *Physcomitrella patens*[J]. *Mol Biol Evol*, 2002, 19 (6): 801–814.
- [14] Kwantes M, Liebsch D, Verelst W. How MIKC* MADS-box genes originated and evidence for their conserved function throughout the evolution of vascular plant gametophytes[J]. *Mol Biol Evol*, 2012, 29 (1): 293–302.
- [15] Kofuji R, Sumikawa N, Yamasaki M, Kondo K, Ueda, K, et al. Evolution and divergence of the MADS-box gene family based on genome-wide expression analyses[J]. *Mol Biol Evol*, 2003, 20 (12): 1963–1977.
- [16] Masiero S, Colombo L, Grini PE, Schnittger A, Kater MM. The emerging importance of Type I MADS box transcription factors for plant reproduction[J]. *Plant Cell*, 2011, 23 (3): 865–872.
- [17] Alvarez-Buylla ER, Liljgren SJ, Pelaz S, Gold SE, Burgeff C, et al. MADS-box gene evolution beyond flowers: expres-

- sion in pollen, endosperm, guard cells, roots and trichomes[J]. *Plant Journal*, 2000, 24 (4): 457–466.
- [18] Chanderbali AS, Berger BA, Howarth DG, Soltis DE, Soltis PS. Evolution of floral diversity: genomics, genes and gamma[J]. *Philos Trans R Soc B Biol Sci*, 2017, 372 (1713): 20150509.
- [19] Grimplet J, Martínez-Zapater JM, Carmona MJ. Structural and functional annotation of the MADS-box transcription factor family in grapevine[J]. *BMC Genomics*, 2016, 17 (1): 80.
- [20] Koo SC, Bracko O, Park MS, Schwab R, Chun HJ, et al. Control of lateral organ development and flowering time by the *Arabidopsis thaliana* MADS-box gene *AGAMOUS-LIKE6*[J]. *Plant J*, 2010, 62 (5): 807–816.
- [21] Qiu YC, Köhler C, Svergel L. Endosperm evolution by duplicated and neofunctionalized Type I MADS-box transcription factors[J]. *Mol Biol Evol*, 2022, 39 (1): msab355.
- [22] Schwarz-Sommer Z, Huijser P, Nacken W, Saedler H, Sommer H. Genetic control of flower development by homeotic genes in *Antirrhinum majus*[J]. *Science*, 1990, 250 (4983): 931–936.
- [23] Amasino RM, Michaels SD. The timing of flowering[J]. *Plant Physiol*, 2010, 154 (2): 516–520.
- [24] Chen Q, Li J, Yang FJ. Genome-wide analysis of the MADS-box transcription factor family in *Solanum melongena*[J]. *Int J Mol Sci*, 2023, 24 (1): 826.
- [25] Arora R, Agarwal P, Ray S, Singh AK, Singh VP, et al. MADS-box gene family in rice: genome-wide identification, organization and expression profiling during reproductive development and stress[J]. *BMC Genomics*, 2007, 8 (1): 242.
- [26] Wei B, Zhang RZ, Guo JJ, Liu DM, Li AL, et al. Genome-wide analysis of the MADS-box gene family in *Brachypodium distachyon*[J]. *PLoS One*, 2014, 9 (1): e84781.
- [27] Leseberg CH, Li AL, Kang H, Duvall M, Mao L. Genome-wide analysis of the MADS-box gene family in *Populus trichocarpa*[J]. *Gene*, 2006, 378: 84–94.
- [28] Zhang XD, Fatima M, Zhou P, Ma Q, Ming R. Analysis of MADS-box genes revealed modified flowering gene network and diurnal expression in pineapple[J]. *BMC Genomics*, 2020, 21 (1): 8.
- [29] Ruelens P, Zhang ZC, van Mourik H, Maere S, Kaufmann K, Geuten K. The origin of floral organ identity quartets[J]. *Plant Cell*, 2017, 29 (2): 229–242.
- [30] Ferrario S, Immink RGH, Shchennikova A, Busscher-Lange J, Angenent GC. The MADS box gene *FBP2* is required for SEPALLATA function in petunia[J]. *Plant Cell*, 2003, 15 (4): 914–925.
- [31] 高玮林, 张力曼, 薛超玲, 张焱, 刘孟军, 赵锦. 枣 E 类 MADS 基因在花和果中的表达及其蛋白互作研究 [J]. 园艺学报, 2022, 49 (4): 739–748.
- Gao WL, Zhang LM, Xue CL, Zhang Y, Liu MJ, Zhao J. Expression of E-Type MADS-box genes in flower and fruits and protein interaction analysis in Chinese jujube[J]. *Acta Horticulturae Sinica*, 2022, 49 (4): 739–748.
- [32] 孙福辉, 方慧仪, 温小惠, 张亮生. 马银花 MADS-box 基因家族系统进化与表达分析 [J]. 植物学报, 2023, 58 (3): 404–416.
- Sun FH, Fang HY, Wen XH, Zhang LS. Phylogenetic and expression analysis of MADS-box gene family in *Rhododendron ovatum*[J]. *Chinese Bulletin of Botany*, 2023, 58 (3): 404–416.
- [33] Chen CJ, Chen H, Zhang Y, Thomas HR, Frank MH, et al. TBtools: an integrative toolkit developed for interactive analyses of big biological data[J]. *Mol Plant*, 2020, 13 (8): 1194–1202.
- [34] Zhang Z, Li J, Zhao XQ, Wang J, Wong GKS, Yu J. KaKs_calculator: calculating Ka and Ks through model selection and model averaging[J]. *Genomics Proteomics Bioinformatics*, 2006, 4 (4): 259–263.
- [35] Shu YJ, Yu DS, Wang D, Guo DL, Guo CH. Genome-wide survey and expression analysis of the MADS-box gene family in soybean[J]. *Mol Biol Rep*, 2013, 40 (6): 3901–3911.
- [36] 潘志演, 唐博希, 田慧源, 范吴蔚, 彭剑涛, 刘国琴. 烟草 MADS-box 基因家族鉴定及其腋芽发育相关基因表达分析 [J]. 分子植物育种, 2022, 20 (7): 2217–2232.
- Pan ZY, Tang BX, Tian HY, Fan WW, Peng JT, Liu GQ. Identification of MADS-box gene family and expression analysis of genes related to axillary bud development in *Nicotiana tabacum*[J]. *Molecular Plant Breeding*, 2022, 20 (7): 2217–2232.
- [37] Wang YS, Zhang JL, Hu ZL, Guo XH, Tian SB, Chen GP. Genome-wide analysis of the MADS-box transcription factor family in *Solanum lycopersicum*[J]. *Int J Mol Sci*, 2019, 20 (12): 2961.
- [38] Li C, Wang Y, Xu L, Nie SS, Chen YL, et al. Genome-wide characterization of the MADS-box gene family in radish (*Raphanus sativus* L.) and assessment of its roles in flowering and floral organogenesis[J]. *Front Plant Sci*, 2016, 7: 1390.
- [39] Won SY, Jung JA, Kim JS. Genome-wide analysis of the MADS-box gene family in Chrysanthemum[J]. *Comput Biol*

- Chem*, 2021, 90: 107424.
- [40] 高欢, 郑珂昕, 廖光联, 王海令, 陈璐, 等. 中华猕猴桃全基因组 *MADS-box* 基因家族鉴定及表达分析 [J]. 果树学报, 2023, 40 (11): 2307–2324.
- Gao H, Zheng KX, Liao GL, Wang HL, Chen L, *et al.* Genome-wide identification and expression analysis of the *MADS-box* gene family in *Actinidia chinensis*[J]. *Journal of Fruit Science*, 2023, 40 (11): 2307–2324.
- [41] Wei X, Wang LH, Yu JY, Zhang YX, Li DH, Zhang XR. Genome-wide identification and analysis of the *MADS-box* gene family in sesame[J]. *Gene*, 2015, 569 (1): 66–76.
- [42] 王溪唯, 陈璨, 王大玮. 云南栎 [木衣]*MADS-box* 基因家族鉴定与表达分析 [J]. 生物工程学报, 2023, 39 (7): 2897–2913.
- Wang XW, Chen C, Wang DW. Identification and expression analysis of *MADS-box* gene family in *Docynia delavayi* (Franch.) Schneid[J]. *Chinese Journal of Biotechnology*, 2023, 39 (7): 2897–2913.
- [43] Li D, Liu C, Shen LS, Wu Y, Chen HY, *et al.* A repressor complex governs the integration of flowering signals in *Arabidopsis*[J]. *Dev Cell*, 2008, 15 (1): 110–120.
- [44] Yang YF, Zhu T, Niu DK. Association of intron loss with high mutation rate in *Arabidopsis*: implications for genome size evolution[J]. *Genome Biol Evol*, 2013, 5 (4): 723–733.
- [45] Wu JY, Xiao JF, Wang LP, Zhong J, Yin HY, *et al.* Systematic analysis of intron size and abundance parameters in diverse lineages[J]. *Sci China Life Sci*, 2013, 56 (10): 968–974.
- [46] Gao HH, Wang ZM, Li SL, Hou ML, Zhou Y, *et al.* Genome-wide survey of potato *MADS-box* genes reveals that StMADS1 and StMADS13 are putative downstream targets of tuberigen StSP6A[J]. *BMC Genomics*, 2018, 19 (1): 726.
- [47] Airoidi CA, Davies B. Gene duplication and the evolution of plant *MADS-box* transcription factors[J]. *J Genet Genomics*, 2012, 39 (4): 157–165.
- [48] Zhou PY, Qu YS, Wang ZW, Huang B, Wen Q, *et al.* Gene structural specificity and expression of *MADS-box* gene family in *Camellia chekiangoleosa*[J]. *Int J Mol Sci*, 2023, 24 (4): 3434.

(责任编辑: 李惠英)