

• 药材与资源 •

巴戟天 *AP2/ERF* 基因家族的全基因组鉴定及生物信息学分析

陈纪佟, 刘楚弦, 韩逸飞, 潘芸燕, 段礼新, 季爱加*

广州中医药大学中药学院国际中医药转化医学研究所, 教育部中医药防治肿瘤转化医学研究国际合作联合实验室, 广东省中医药防治肿瘤转化药学研究重点实验室 广东 广州 510000

摘要: **目的** 研究鉴定巴戟天 *Morinda officinalis* *AP2/ERF* 转录因子家族, 为巴戟天次生代谢调控网络提供理论基础。**方法** 通过对巴戟天 *AP2/ERF* 转录因子家族的探索, 利用 TBtools、MEGA 等生物信息分析软件对巴戟天基因组基因结构注释信息文件进行基本分析、进化分析, 得出巴戟天中 *AP2/ERF* 家族的基因, 以及这些基因中内含子排列情况、共线性分析的结果, 并且通过进化树的构建可视化 *AP2/ERF* 基因家族的分类。**结果** 初步筛选出可能调节巴戟天生长发育和代谢的转录因子。在巴戟天基因组中, 鉴定到有 131 条基因属于 *AP2/ERF* 基因家族成员, 通过基因结构注释文件, 得到 Mo*AP2/ERF* 在染色体定位分析、系统发育进化树、内含子外显子模式、motif、保守结构域分析和共线性分析等, 最终根据 Mo*AP2/ERF* 的表达量确认 13 条在巴戟天根部高表达的转录因子。**结论** 通过生物信息分析筛选出在根部特异性表达的基因 13 条巴戟天 *AP2/ERF* 转录因子, 为增加巴戟天主要有效成分产率奠定理论基础, 为后续研究提供信息。

关键词: 巴戟天; *AP2/ERF* 基因家族; 基因家族鉴定; 生物信息学; 表达分析**中图分类号:** R286.12 **文献标志码:** A **文章编号:** 0253-2670(2024)20-7062-10**DOI:** 10.7501/j.issn.0253-2670.2024.20.021Genome-wide identification and bioinformatics analyses of *AP2/ERF* gene family in *Morinda officinalis*

CHEN Jitong, LIU Chuxian, HAN Yifei, PAN Yunyan, DUAN Lixin, JI Aijia

Guangdong Provincial Key Laboratory of Translational Cancer Research of Chinese Medicines, Joint International Research Laboratory of Translational Cancer Research of Chinese Medicines, International Institute for Translational Chinese Medicine, School of Pharmaceutical Sciences, Guangzhou University of Chinese Medicine, Guangzhou 510000, China

Abstract: **Objective** To identify the Bajitian (*Morinda officinalis*) *AP2/ERF* transcription factor family for providing a theoretical basis for *M. officinalis* secondary metabolic regulatory network. **Methods** In this study, the genes of the *AP2/ERF* family in *M. officinalis*, as well as the intron arrangement and collinearity analysis results of these genes, were obtained by exploring the *AP2/ERF* transcription factor family in *M. officinalis*, and the classification of *AP2/ERF* gene family was visualized through the construction of phylogenetic tree. **Results** Transcription factors that may regulate the growth, development and metabolism of *M. officinalis* were preliminarily screened. In the genome of *M. officinalis*, 131 genes belonging to the *AP2/ERF* gene family were identified, and chromosome localization analysis, phylogenetic evolution tree, intron exon pattern, motif, conserve domain analysis and collinearity analysis of Mo*AP2/ERF* were obtained through gene structure annotation files, and finally 13 transcription factors with high expression in the root of *M. officinalis* were confirmed. **Conclusion** A total of 13 *AP2/ERF* transcription factors of *M. officinalis* were screened out by bioinformatics analysis, which laid a theoretical foundation for increasing the yield of the main active components of *M. officinalis* and provided information for follow-up research.

Key words: *Morinda officinalis*; *AP2/ERF* gene family; gene family identification; bioinformatics; expression analysis

收稿日期: 2024-05-06

基金项目: 科技部基础资源调查重点专项 (2018FY100700); 广东省基础与应用基础研究基金项目 (2022A151010646); 广东省中医药管理局面上项目 (20241085); 广东省级大学生创新训练项目 (202310572062)

作者简介: 陈纪佟 (2000—), 女, 硕士研究生, 药用植物功能基因组学。E-mail: 1363751260@qq.com

*通信作者: 季爱加 (1987—), 女, 副研究员, 药用植物功能基因组学。E-mail: ajji@gzucm.edu.cn

巴戟天 *Morinda officinalis* How 是一种藤状灌木, 根肉质肥厚, 整体呈圆柱形, 不规则膨大, 呈念珠状, 干燥根可作为中药应用到中医药治疗中。巴戟天生于山谷溪边、山地疏林下或人工栽培, 全年均可挖, 主产于福建、广东、海南、广西等地。巴戟天具有补肾阳、强筋骨、祛风湿的作用, 常用于阳痿遗精、宫冷不孕、月经不调、少腹冷痛、风湿痹痛、筋骨痠软等症状的治疗, 和益智仁、砂仁、槟榔合称为“四大南药”。其主要活性成分为蒽醌类、环烯醚萜类、糖类等^[1]。

AP2/ERF (APETALA2/乙烯反应元件结合因子) 转录因子是一大类主要在植物中发现的调控因子, 是植物中最大、最保守的基因家族之一^[2]。它的蛋白结构中存在着 1 个或 2 个特有的 AP2 结构域, 该结构域含有 60~70 个高度保守的氨基酸残基, 由 3 个 β -折叠和 1 个 α -螺旋组成^[3]。该基因家族主要分为 5 个主要亚家族: 脱水应答元件结合蛋白 (DREB)、乙烯应答转录因子 (ERF)、AP2、RAV 和 Soloist。AP2 亚家族有 2 个 AP2 结构域, 主要参与调节植物的发育过程; DREB 亚家族与 ERF 亚家族只含有 1 个 AP2 结构域, 主要参与响应植物生物胁迫和非生物胁迫; RAV 亚家族包含 1 个 AP2 结构域和 1 个 B3 结构域; Soloist 亚家族仅包含 1 个 AP2 结构域, 但其序列和其他亚家族基因的序列高度不同^[4]。AP2/ERF 转录因子最初是从模式植物拟南芥中分离, 随着测序技术的快速发展, 目前已有对蔓生花^[5]、光皮桦^[6]、阿拉伯岩芥^[7]、柳树^[8]、丹参^[9]、穿心莲^[10]以及小麦^[2]等植物中 AP2/ERF 转录因子家族的全基因组鉴定。

转录因子作为基因表达的分子开关, 主要控制转录网络并影响植物的生长发育, 如在穿心莲 *Andrographis paniculate* bZIP 和 WRKY 基因家族分析中, 通过比较 MeJA 或 ABA 处理前后的 ApbZIPs 和 ApWRKYs 表达量变化, 获得 7 个 ApbZIPs 和 7 个 ApWRKYs 可能参与调控穿心莲内酯类生物合成的候选转录因子^[11-12]; 大麻 *Cannabis sativa* ARF 基因家族分析预测了 12 个候选 CsARFs 可能参与大麻的生长发育调控^[13]。在植物基因家族分析预测的基础上, 有部分候选 AP2/ERF 转录因子已进行功能验证, 例如红豆杉的 AP2/ERF 转录因子 *TcERF12* 和 *TcERF15* 可调控该植物中紫杉醇的生物合成^[14]; 在喜树中异源性表达 *CaERF1*, 可以促进单喙喙生物碱喜树碱的积累^[15]; 在丹参中过表达 *SmERF128* 可

增加毛状根中丹参的含量^[16]。近年来, 巴戟天的药材需求量逐步增加, 通过探究参与调控主要活性成分合成通路的转录因子家族, 以提高巴戟天有效成分含量是巴戟天的研究热点之一。目前, 巴戟天 AP2/ERF 基因家族的全基因组鉴定尚未见报道。因此, 本研究通过生物信息分析对巴戟天 AP2/ERF 基因进行保守结构域、系统发育、染色体定位、启动子顺式作用元件和共线性等分析, 探究巴戟天 AP2/ERF 转录因子家族的序列特征和可能的生物学功能, 初步筛选与巴戟天次生代谢产物生物合成相关的转录因子, 为增加巴戟天主要有效成分含量奠定理论基础。

1 材料

本研究所使用的植物材料样本采于广东省肇庆市, 经广州中医药大学中药学院季爱加副研究员鉴定为巴戟天 *M. officinalis* How 三年生根、茎、叶, 采集后迅速冷冻于液氮中, 随后保存于 -80 °C 冰箱中。在拟南芥信息资源网站 (TAIR) (<https://www.arabidopsis.org/>) 下载和 NCBI 上下载的咖啡 AP2/ERF 家族的蛋白序列作为参考序列。巴戟天基因组、转录组数据由本课题组测序获得。

2 方法

2.1 巴戟天中 AP2/ERF 基因序列的搜索与鉴定

为了解基因在巴戟天染色体上的分布, 基于本研究团队前期测序获得的巴戟天基因组数据, 以 TAIR 上下载 AtAP2/ERF 蛋白序列 140 条作为查询序列, 使用 TBtools^[17]软件中的 BLASTP 功能, 对巴戟天转录组数据库进行本地检索初步筛选属于 AP2/ERF 家族的基因。去除不完整结构域和冗余序列后, 通过 NCBI 保守域数据库 (CDD)^[18]网站和 InterPro (<https://www.ebi.ac.uk/interpro/>) 确定获得序列的保守结构域^[19], 保留有 AP2/ERF 结构域的序列进行下一步分析。最后, 在 ExPASy 网站 (<https://www.expasy.org/>) 预测巴戟天 AP2/ERF 序列的蛋白理化性质^[20], 包括编码序列长度 (CDS)、相对分子质量 (M_w)、理论等电点 (PI) 和总平均水合性 (GRAVY) 等。

2.2 巴戟天 AP2/ERF 基因的染色体定位和命名、系统进化分析

从巴戟天基因组数据库的基因注释文件中获取巴戟天 AP2/ERF 序列的染色体位置信息, 如染色体号、染色体长度、基因起始和终止位置。运用 TBtools 软件对巴戟天中的 AP2/ERF 基因的染色体图谱进

行可视化,按照 *AP2/ERF* 基因在染色体上的位置顺序进行命名 (*MoAP2/ERF1-MoAP2/ERF131*)。使用 MEGA-X 软件将在 TAIR 网站上下载的拟南芥蛋白序列与巴戟天中获取的 *MoAP2/ERF* 一同构建系统发育树。

2.3 巴戟天 *AP2/ERF* 基因的保守结构域和基因结构分析、顺式元件分析、共线性分析

通过 MEME 5.5.4 在线分析程序预测 *MoAP2/ERF* 蛋白中的保守基序,数量设置为 10 个,其他参数设置为默认值^[21],使用 TBtools 软件对其进行可视化。从巴戟天基因组数据库中获得的 *MoAP2/ERF* 基因上游 2 kb 的序列作为启动子区域,将 *MoAP2/ERF* 蛋白序列上传至 PlantCare 网站 (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>),获得顺式作用元件^[22],使用 TBtools 对结果进行可视化。使用 TBtools 中的 MCSanX^[23]和

BLAST 方法,分析了 *MoAP2/ERF* 的串联重复事件。

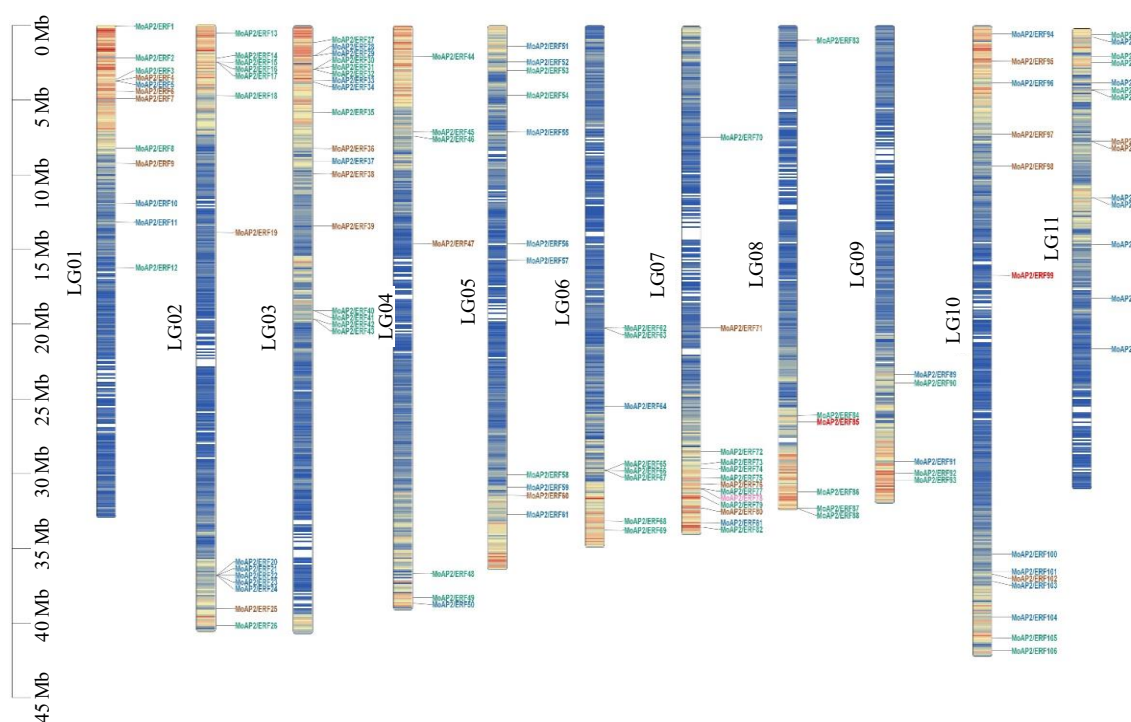
2.4 巴戟天 *AP2/ERF* 基因家族差异表达分析

基于前期本课题组于公司测序获得的巴戟天转录组数据,选择三年生巴戟天根茎叶的 RNA-seq,使用并用 Hisat2、Stringtie 及其自带 python 脚本 prepDE.py 分析巴戟天 *AP2/ERF* 基因的表达模式,并使用 TBtools 进行可视化。

3 结果与分析

3.1 巴戟天 *AP2/ERF* 基因的鉴定

根据巴戟天的基因注释文件获得 131 个蛋白序列在染色体上的位置,使用 TBtools 进行可视化。结果显示 (图 1),基因在染色体上的分布较分散, LG3、LG4 上分布的基因最多,均为 17 个。部分亚族基因在染色体上有聚集分布的现象,ERF 亚族在 LG2、LG3 和 LG6 上各有 7、9、8 条, DREB 亚族基因则在 LG2、LG5 和 LG11 上各有 5、8、7 条。



绿色-ERF 亚家族; 蓝色-DREB 亚家族; 褐色-AP2 亚家族; 红色-RAV 亚家族; 粉色-Soloist 亚家族。
green-ERF subfamily; blue-DREB subfamily; brown-AP2 subfamily; red-RAV subfamily; pink-Soloist subfamily.

图 1 *MoAP2/ERFs* 染色体定位

Fig. 1 Distribution of *MoAP2/ERF* across chromosomes

3.2 巴戟天 *AP2/ERF* 基因家族成员的理化性质分析

对所鉴定的 131 个 *MoAP2/ERF* 基因的理化性质进行分析 (表 1), 预测结果表示, 巴戟天 *AP2/ERF* 家族编码蛋白质的氨基酸数在 135 (*MoAP2/*

ERF53) ~ 717 (*MoAP2/ERF38*), M_w 在 15 526.45 (*MoAP2/ERF53*) ~ 78 216.34 (*MoAP2/ERF38*), PI 范围在 4.46 (*MoAP2/ERF2*) ~ 10.33 (*MoAP2/ERF56*)。

表 1 巴戟天 131 个 *AP2/ERF* 家族基因及其相关信息Table 1 131 *AP2/ERF* family genes and related information of *M. officinalis*

基因 ID	编号	亚家族	氨基酸	相对分子质量	等电点	基因 ID	编号	亚家族	氨基酸	相对分子质量	等电点
MoAP2/ERF1	EVM0008934	ERF	275	30 396.39	8.85	MoAP2/ERF67	EVM0002595	ERF	152	16 757.19	5.54
MoAP2/ERF2	EVM0021696	ERF	414	45 317.83	4.46	MoAP2/ERF68	EVM0020529	ERF	313	34 979.08	4.88
MoAP2/ERF3	EVM0002227	ERF	541	60 263.12	6.15	MoAP2/ERF69	EVM0020856	ERF	327	36 567.22	7.14
MoAP2/ERF4	EVM0008046	AP2	526	58 159.82	6.43	MoAP2/ERF70	EVM0023471	ERF	278	31 061.77	9.39
MoAP2/ERF5	EVM0021633	DREB	163	18 174.44	9.30	MoAP2/ERF71	EVM0016412	AP2	428	47 987.64	9.30
MoAP2/ERF6	EVM0006214	AP2	639	70 804.53	6.41	MoAP2/ERF72	EVM0013371	ERF	231	25 709.75	9.71
MoAP2/ERF7	EVM0019140	AP2	391	44 661.48	5.92	MoAP2/ERF73	EVM0017726	ERF	411	45 845.72	5.63
MoAP2/ERF8	EVM0004252	ERF	203	23 050.06	5.11	MoAP2/ERF74	EVM0019143	ERF	366	39 653.53	5.24
MoAP2/ERF9	EVM0006505	AP2	512	56 834.60	6.18	MoAP2/ERF75	EVM0024769	ERF	256	28 188.78	7.15
MoAP2/ERF10	EVM0004302	DREB	376	41 363.04	6.25	MoAP2/ERF76	EVM0007066	AP2	387	43 176.06	5.96
MoAP2/ERF11	EVM0020372	DREB	187	21 292.17	9.63	MoAP2/ERF77	EVM0005453	ERF	184	21 000.69	6.66
MoAP2/ERF12	EVM0016034	ERF	407	43 413.53	5.53	MoAP2/ERF78	EVM0023355	SOLOIST	281	31 037.31	9.63
MoAP2/ERF13	EVM0005385	ERF	351	39 727.76	6.18	MoAP2/ERF79	EVM0017053	ERF	212	23 449.05	8.84
MoAP2/ERF14	EVM0007893	ERF	177	19 538.44	9.64	MoAP2/ERF80	EVM0017069	AP2	652	72 083.32	6.01
MoAP2/ERF15	EVM0004982	ERF	153	17 543.32	5.51	MoAP2/ERF81	EVM0018057	DREB	260	27 884.60	5.62
MoAP2/ERF16	EVM0005334	ERF	227	25 260.28	6.54	MoAP2/ERF82	EVM0003710	ERF	322	36 397.68	8.18
MoAP2/ERF17	EVM0007130	ERF	235	26 303.99	4.69	MoAP2/ERF83	EVM0024072	ERF	241	26 272.34	10.02
MoAP2/ERF18	EVM0000207	ERF	270	30 593.91	5.27	MoAP2/ERF84	EVM0000849	ERF	386	43 034.87	4.50
MoAP2/ERF19	EVM0002027	AP2	518	57 042.53	6.37	MoAP2/ERF85	EVM0016541	RAV	350	39 102.31	8.96
MoAP2/ERF20	EVM0007560	DREB	242	26 808.91	8.88	MoAP2/ERF86	EVM0016173	ERF	394	43 038.50	7.62
MoAP2/ERF21	EVM0018875	DREB	249	27 627.14	9.51	MoAP2/ERF87	EVM0016083	ERF	211	23 581.83	6.44
MoAP2/ERF22	EVM0008010	DREB	242	26 809.85	7.77	MoAP2/ERF88	EVM0002467	ERF	335	37 176.62	6.15
MoAP2/ERF23	EVM0019647	DREB	241	26 640.94	9.30	MoAP2/ERF89	EVM0012689	DREB	497	56 004.46	5.26
MoAP2/ERF24	EVM0002494	DREB	241	26 482.67	9.07	MoAP2/ERF90	EVM0000270	ERF	217	23 038.59	7.86
MoAP2/ERF25	EVM0007596	AP2	511	56 896.87	6.26	MoAP2/ERF91	EVM0019308	DREB	354	40 087.12	8.41
MoAP2/ERF26	EVM0004204	ERF	278	31 104.62	6.59	MoAP2/ERF92	EVM0003417	ERF	310	34 175.90	6.73
MoAP2/ERF27	EVM0005023	ERF	392	43 848.46	4.61	MoAP2/ERF93	EVM0019030	ERF	238	26 236.08	8.72
MoAP2/ERF28	EVM0021892	DREB	269	28 952.12	5.14	MoAP2/ERF94	EVM0022683	DREB	401	45 270.00	5.16
MoAP2/ERF29	EVM0000636	DREB	235	25 669.69	4.97	MoAP2/ERF95	EVM0005825	AP2	472	52 358.80	6.19
MoAP2/ERF30	EVM0014698	ERF	217	24 172.87	9.21	MoAP2/ERF96	EVM0018678	DREB	237	25 105.26	5.74
MoAP2/ERF31	EVM0011511	ERF	280	30 943.68	7.64	MoAP2/ERF97	EVM0017399	AP2	552	59 796.53	5.98
MoAP2/ERF32	EVM0010189	ERF	256	28 035.60	6.24	MoAP2/ERF98	EVM0008438	AP2	386	44 086.01	6.71
MoAP2/ERF33	EVM0015527	DREB	346	38 750.84	4.90	MoAP2/ERF99	EVM0009707	RAV	318	35 987.70	9.08
MoAP2/ERF34	EVM0016850	DREB	351	38 365.06	6.59	MoAP2/ERF100	EVM0024583	DREB	162	18 134.66	9.32
MoAP2/ERF35	EVM0001311	ERF	259	29 319.28	5.07	MoAP2/ERF101	EVM0022966	DREB	281	32 274.94	5.41
MoAP2/ERF36	EVM0007620	AP2	492	54 691.67	6.16	MoAP2/ERF102	EVM0020395	AP2	346	38 058.28	9.41
MoAP2/ERF37	EVM0021323	DREB	227	25 284.01	5.45	MoAP2/ERF103	EVM0018133	DREB	229	25 186.60	4.66
MoAP2/ERF38	EVM0018400	AP2	717	78 216.34	6.33	MoAP2/ERF104	EVM0012314	DREB	289	31 341.25	5.48
MoAP2/ERF39	EVM0013742	AP2	438	48 336.83	8.62	MoAP2/ERF105	EVM0002985	ERF	261	28 875.28	7.06
MoAP2/ERF40	EVM0005669	AP2	161	17 798.08	8.05	MoAP2/ERF106	EVM0007125	ERF	312	34 905.68	5.87
MoAP2/ERF41	EVM0022609	ERF	259	27 980.21	7.15	MoAP2/ERF107	EVM0019664	ERF	391	43 158.40	4.87
MoAP2/ERF42	EVM0019409	ERF	182	19 884.15	9.12	MoAP2/ERF108	EVM0005356	DREB	309	34 577.12	9.19
MoAP2/ERF43	EVM0024626	ERF	284	32 187.85	5.47	MoAP2/ERF109	EVM0000937	ERF	203	23 189.22	9.07
MoAP2/ERF44	EVM0004813	SOLOIST	275	30 132.28	6.22	MoAP2/ERF110	EVM0003363	ERF	225	25 019.11	9.35
MoAP2/ERF45	EVM0003236	AP2	229	24 464.07	4.58	MoAP2/ERF111	EVM0000027	DREB	198	21 895.96	9.39
MoAP2/ERF46	EVM0023307	ERF	209	23 417.33	9.74	MoAP2/ERF112	EVM0004226	ERF	346	38 898.34	5.40
MoAP2/ERF47	EVM0010662	AP2	392	44 390.30	6.25	MoAP2/ERF113	EVM0005081	ERF	265	28 901.45	7.65
MoAP2/ERF48	EVM0008143	ERF	304	33 125.49	6.97	MoAP2/ERF114	EVM0010482	AP2	394	43 731.99	9.25
MoAP2/ERF49	EVM0007904	ERF	352	39 771.10	5.44	MoAP2/ERF115	EVM0004683	AP2	570	63 589.95	9.08
MoAP2/ERF50	EVM0017586	DREB	221	24 324.31	9.06	MoAP2/ERF116	EVM0022132	DREB	194	20 581.58	4.53
MoAP2/ERF51	EVM0013507	DREB	189	21 347.14	9.96	MoAP2/ERF117	EVM0022450	DREB	249	27 536.51	5.58
MoAP2/ERF52	EVM0007601	DREB	350	39 104.67	8.22	MoAP2/ERF118	EVM0003348	DREB	262	28 508.57	5.32
MoAP2/ERF53	EVM0013509	ERF	135	15 526.45	9.64	MoAP2/ERF119	EVM0015193	DREB	244	27 413.81	4.60
MoAP2/ERF54	EVM0010962	ERF	458	48 986.55	6.41	MoAP2/ERF120	EVM0005668	DREB	171	19 274.63	9.30
MoAP2/ERF55	EVM0016029	DREB	328	36 246.71	8.85	MoAP2/ERF121	EVM0008379	ERF	257	28 895.94	5.85
MoAP2/ERF56	EVM0005409	DREB	168	18 795.46	10.33	MoAP2/ERF122	EVM0021535	ERF	161	17 772.91	6.85
MoAP2/ERF57	EVM0006320	DREB	227	24 647.62	6.07	MoAP2/ERF123	EVM0023638	ERF	462	52 348.59	6.49
MoAP2/ERF58	EVM0004388	DREB	168	18 472.30	5.47	MoAP2/ERF124	EVM0021037	ERF	253	28 696.45	6.91
MoAP2/ERF59	EVM0019468	DREB	198	21 970.36	4.89	MoAP2/ERF125	EVM0002076	ERF	167	18 851.96	9.09
MoAP2/ERF60	EVM0024367	AP2	682	75 676.94	6.28	MoAP2/ERF126	EVM0004744	ERF	207	23 343.10	5.46
MoAP2/ERF61	EVM0018347	DREB	154	17 075.87	8.55	MoAP2/ERF127	EVM0021472	ERF	206	23 178.73	4.84
MoAP2/ERF62	EVM0022347	ERF	192	21 432.40	5.47	MoAP2/ERF128	EVM0016361	AP2	356	22 682.06	6.32
MoAP2/ERF63	EVM0022412	ERF	198	22 183.23	5.50	MoAP2/ERF129	EVM0016275	ERF	201	22 687.03	6.09
MoAP2/ERF64	EVM0006160	DREB	458	50 794.18	6.73	MoAP2/ERF130	EVM0007308	ERF	153	16 490.09	4.52
MoAP2/ERF65	EVM0006930	ERF	217	24 350.08	5.25	MoAP2/ERF131	EVM0013987	DREB	155	16 868.43	4.70
MoAP2/ERF66	EVM0001171	ERF	159	17 998.76	6.12						

3.3 巴戟天 *AP2/ERF* 基因家族成员进化树分析

将鉴定得到的 131 条巴戟天 *AP2/ERF* 基因与 *AtAP2/ERF* 和 *CaAP2/ERF* 蛋白序列使用 MEGAX 进行进化树分析, 根据建树中 *MoAP2/ERFs* 和 *AtAP2/ERFs* 基因家族成员的亲缘关系进行分类, 结

果表明 *MoAP2/ERFs* 基因共分成 5 个亚族, 包括 *AP2*、*ERF*、*DREB*、*RAV*、*Soloist* (图 2)。其中被归类到 *ERF* 亚族的 *MoAP2/ERF* 基因数量最多共有 70 个, 其次是 *DREB* 和 *AP2* 亚族, 分别有 38 和 20 个, 最少的为 *RAV* 和 *SOLOIST* 亚族。

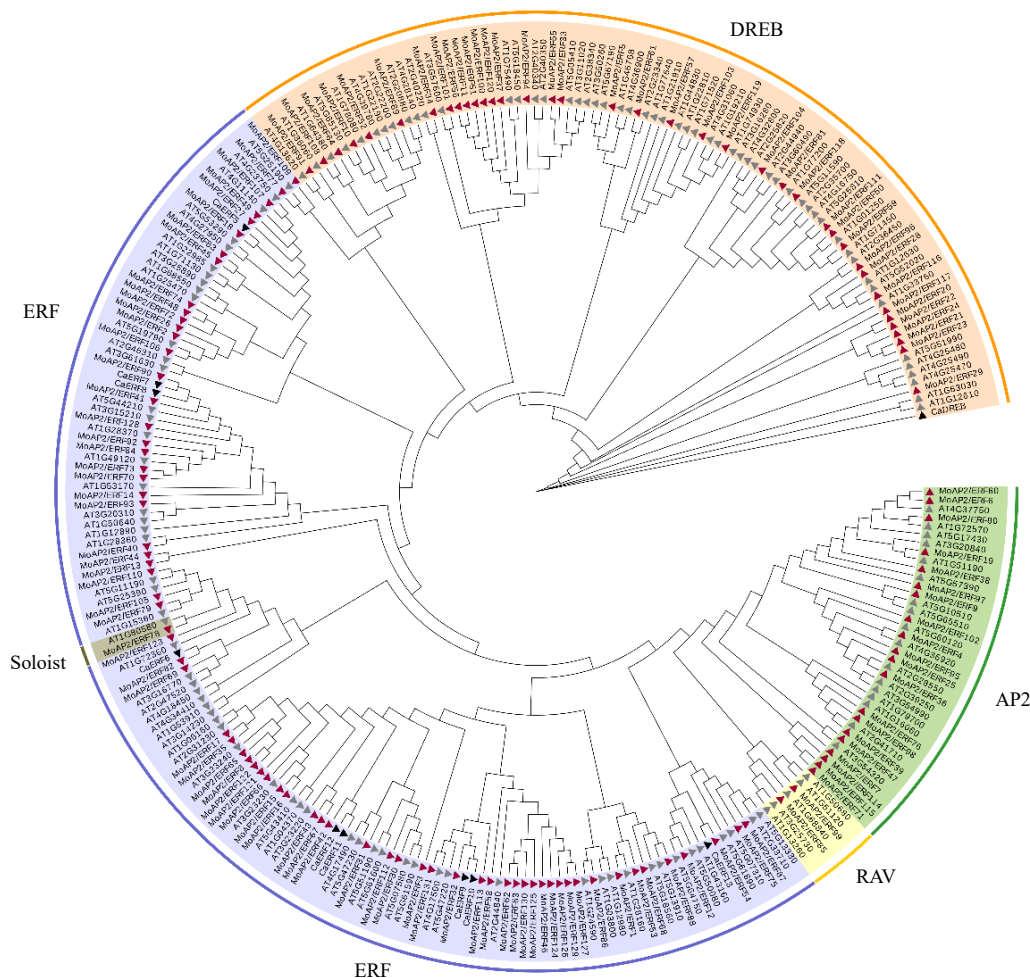


图 2 *MoAP2/ERFs* 家族基因发育进化树

Fig. 2 Family gene development evolutionary tree of *MoAP2/ERFs*

3.4 巴戟天 *AP2/ERF* 基因家族成员的结构域、motif 分布模式和基因结构

蛋白保守基序通常与特定功能相关, 本研究对巴戟天 *AP2/ERF* 蛋白进行保守基序鉴定 (图 3), 除了 *MoAP2/ERF128* 和 *MoAP2/ERF45*, 巴戟天 *AP2/ERF* 家族均含有 motif2、motif3, 为最保守的 motif。其次是 motif4, 在 *RAV* 以外的亚族中都有存在。部分 motif 特殊存在于一些亚家族中, 如只有 *AP2* 亚家族存在 motif5、6, motif2、7 只存在于 *DREB* 亚家族中。

为进一步分析 *AP2/ERF* 序列特征, 对 *AP2/ERF* 的基因结构及 *AP2* 结构域以及 *B3* 结构域进行分析, 结果显示巴戟天中 *DREB*、*ERF* 和 *SOLOIST* 亚族的基因都只包含 1 个 *AP2* 结构域, *B3* 结构域仅存在于 *RAV* 亚族中, *AP2* 亚族的基因都存在 2 个 *AP2* 结构域, 这与预期的结果一致^[4]。

3.5 巴戟天 *AP2/ERF* 基因的顺式作用元件

转录因子在植物对生物和非生物胁迫的响应中起着重要的作用, 为了探索巴戟天 *AP2/ERF* 基因的潜在功能, 对 *MoAP2/ERFs* 启动子的顺式调

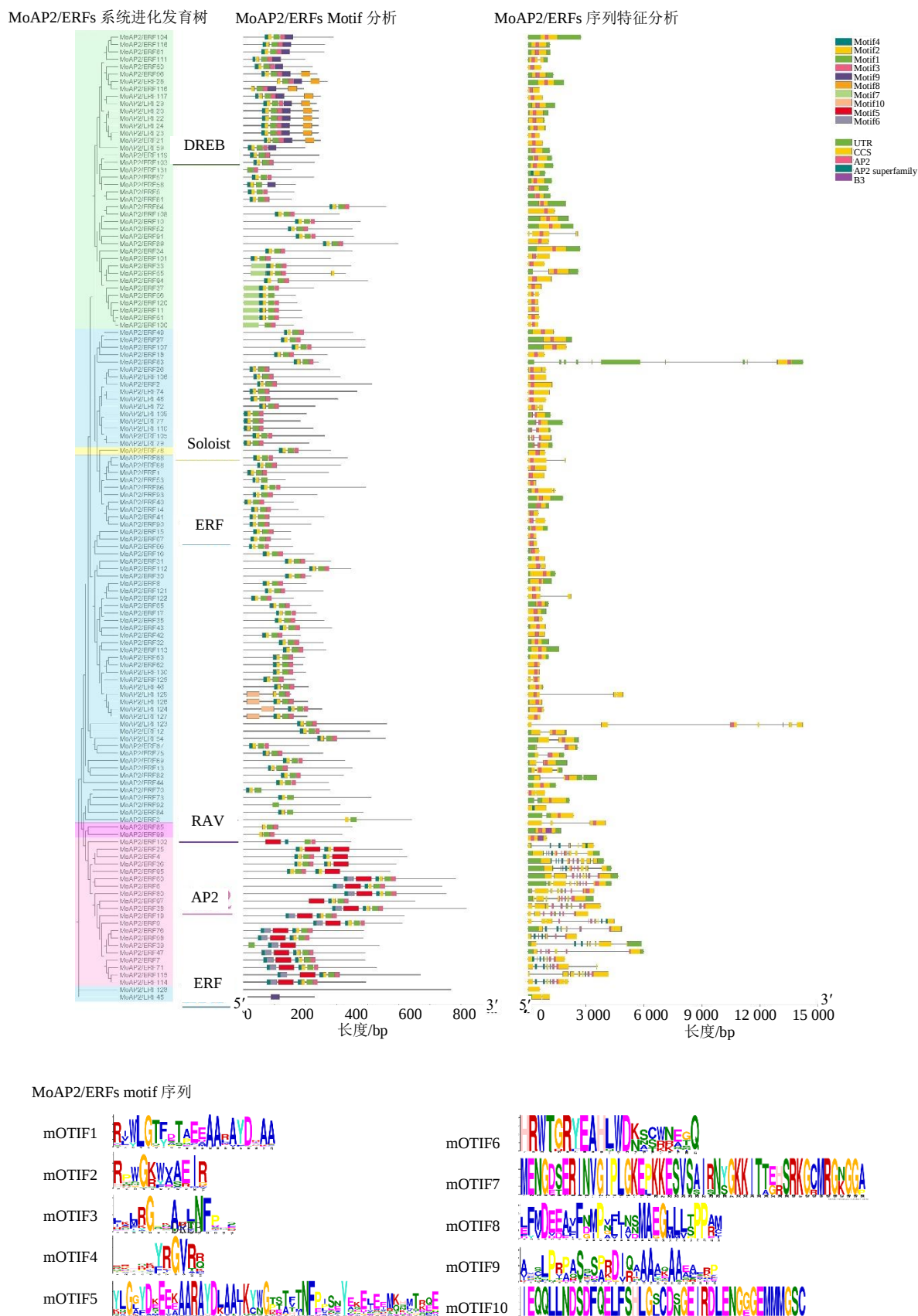


图 3 MoAP2/ERFs 家族基因保守结构域分析

Fig. 3 Analysis of family gene conserved structural domains of MoAP2/ERFs

节元件进行研究(图4),得到15种顺式作用调节元件,包括非生物胁迫、激素调节和植物生长发育调控元件。其中光响应元件占32.4%,是占比最

多的生长发育调节元件,其次是茉莉酸甲酯诱导的激素响应元件,占21.4%。这些发现表明,MoAP2/ERFs可能影响巴戟天激素响应、生长发

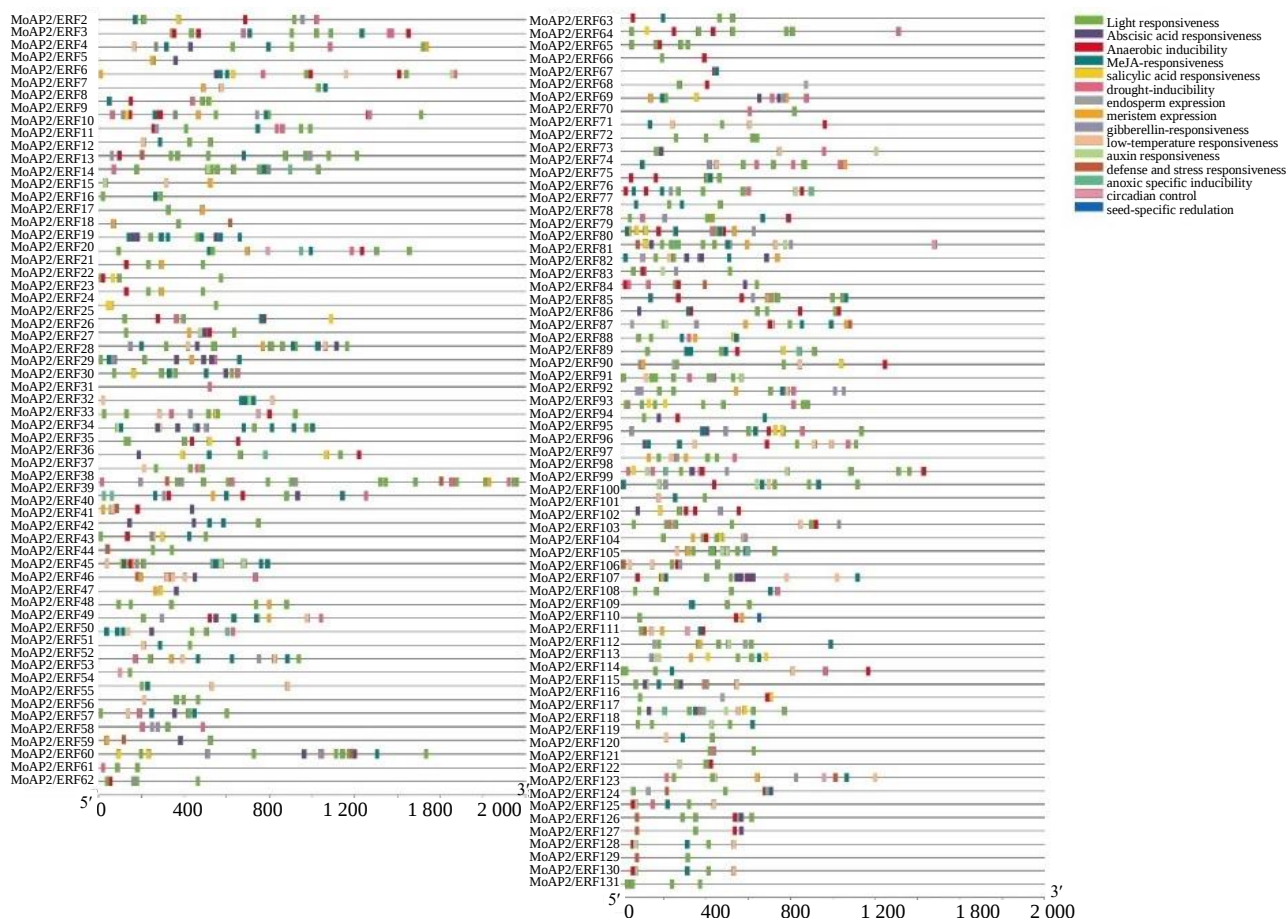


图4 MoAP2/ERFs启动子元件分析

Fig. 4 Promoter elements analysis of MoAP2/ERFs

育和非生物胁迫等。

3.6 巴戟天 AP2/ERF 基因共线性可视化分析

在 TBtools 进行自身蛋白序列 blast 等流程, 获得 link 等文件, 将得到的结果进行美化, 删掉没有共线性关联的部分, 可视化结果见图 5, 共有 30 对基因有共线性关联。其中, LG02、LG03、LG06、LG07、LG10、LG11 上都存在着串联重复, 并且在 LG07 上存在着大量的串联重复, LG09 上的基因对 MoAP2/ERF89 和 MoAP2/ERF91 存在片段重复。

3.6 巴戟天 AP2/ERF 基因表达模式分析

巴戟天的入药部位为根部, 其已被证实富含蒽醌、环烯醚萜苷和糖类成分^[24], 这些成分同时也是巴戟天的主要活性成分^[1]。为了探究 AP2/ERF 基因在巴戟天在根部有特异性表达的转录因子, 对巴戟天的根、叶、茎转录组测序数

据进行分析(图6)。结果表明 MoAP2/ERF 基因在根、叶、茎3个组织中的表达模式不同, 通过聚类分析发现 ERF 亚家族基因的 TPM 值普遍高于其他亚家族。并且, 在根部表达量远高于其他组织部位的基因有 10 条, 分别为 MoAP2/ERF82 (1 093.68)、MoAP2/ERF31 (2 280.33)、MoAP2/ERF84 (61.296 12) 等。其次是 AP2 亚家族, 分别为 MoAP2/ERF80 (13.25) 和 MoAP2/ERF39 (46.87); DREB 亚家族有一条为 MoAP2/ERF33 (69.29); Soloist 和 RAV 亚家族中没有根部表达较高的基因。

4 讨论

巴戟天作为大宗岭南道地药材, 是我国著名的四大南药之一, 临床可用于阳痿肾虚、腰酸背痛等症状, 在治疗骨质疏松、抑郁症、类风湿关节炎等

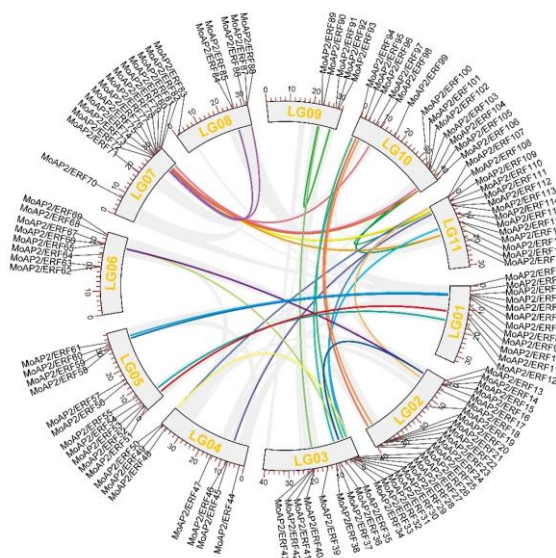


图 5 *MoAP2/ERFs* 共线性分析

Fig. 5 Synteny analysis of *MoAP2/ERFs*

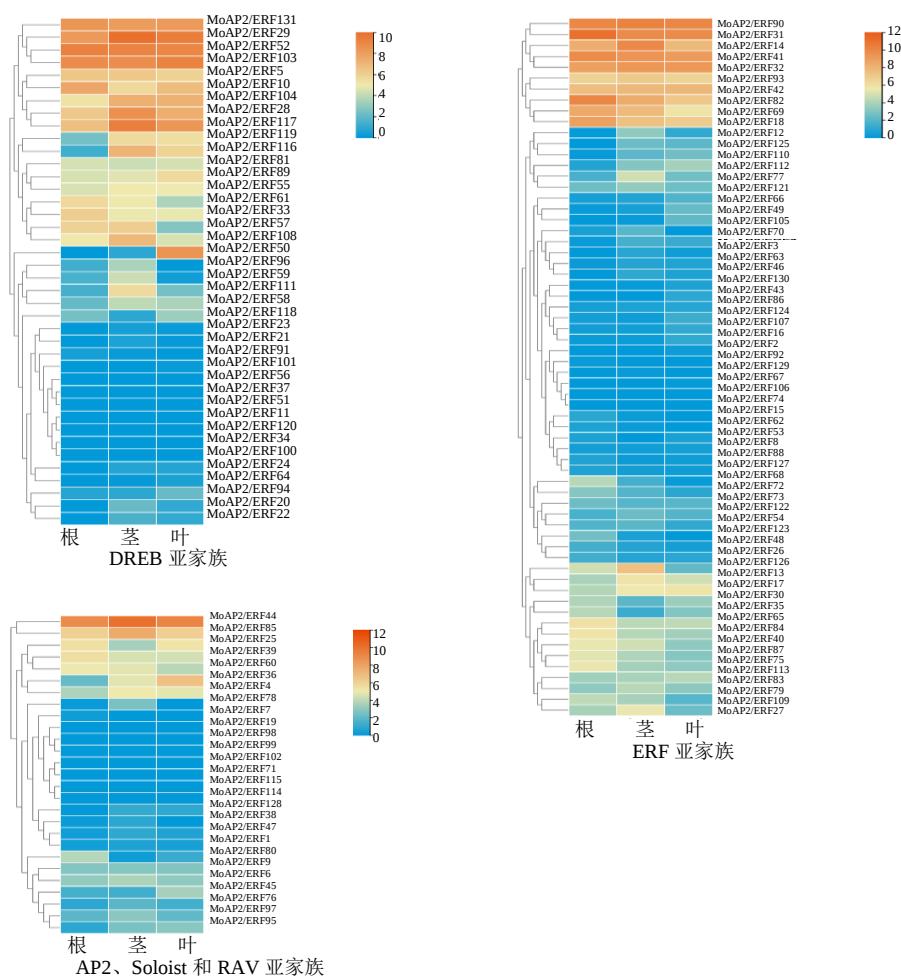


图 6 *MoAP2/ERFs* 在根、茎、叶各组织部位表达量

Fig. 6 Expression in each tissue site of roots, stems and leaves of *MoAP2/ERFs*

疾病中都有应用^[25-26]。因其丰富的药物疗效，近年来需求量日渐增长。根据调查研究，巴戟天野生种质资源已濒临灭绝，主要来源于人工栽培，目前最大的人工栽培产地在广东省德庆县、郁南县等地^[27]。通过研究挖掘能够有效提高巴戟天有效成分产率的转录因子是增加其产量的有效手段之一。*AP2/ERF* 基因家族转录因子在植物次生代谢中发挥着重要作用，如在植物甜橙 *Citrus sinensis* Osbeck 中，*CitERF71* 可激活萜烯合酶基因 *CitTPS16* 参与甜橙果实中 E-香叶醇的合成^[28]。因此，本研究选择探索巴戟天 *AP2/ERF* 转录因子家族的特征，寻找其与巴戟天有效次生代谢物间的关联。

本研究在巴戟天中共鉴定出 131 个 *AP2/ERF* 转录因子，染色体定位结果显示在每个染色体上都有各个亚家族基因的分布（图 1），LG6、LG8 上分布的 *ERF* 亚家族基因较多。根据系统发育树显示（图 2），相比拟南芥（147）、玉米（214）、丹参（170）的数量较少。其中，*ERF* 亚家族的基因数量最多，占比 48%，这和其他物种的研究结果相似^[29]。基因保守域分析显示（图 3），Motif1、Motif2、Motif3、Motif4 在 *MoAP2/ERFs* 中都存在，是该家族中最稳定的 Motif，该结果表明 *MoAP2/ERFs* 基因高度保守。根据 *AP2/ERF* 基因家族先前的结构研究，推测这 3 个保守基序应为 *AP2* 结构域。Motif7、8 和 Motif6 分别是 *DREB* 亚家族和 *AP2* 亚家族的特有 Motif。根据文献研究表明只有 *RAV* 亚家族含有 B3 结构域，这也与本研究结果相同。

在植物应对各种胁迫的反应以及不同的发育阶段中，一些激素可以诱导多种次生代谢产物的生物合成^[30]。例如，提高外源或内源的茉莉酸都可增加番茄的次生代谢产物咖啡酰丁二胺的含量^[31]。在本研究中，*MoAP2/ERF* 的顺式作用元件可分为生长发育调控、非生物胁迫反应和植物激素 3 种作用元件。激素响应元件中，茉莉酸甲酯响应元件最多占 56%，其次是脱落酸响应元件，最少的赤霉素和水杨酸响应元件。

本研究还对 *MoAP2/ERF* 基因的扩展模式和基因复制事件进行分析，结果表明巴戟天组内大量基因存在共线性关系（图 5），该基因家族内部也存在一定的同源进化关系，各个染色体上都存在着大量的串联重复，一些染色体（LG2、3、6、7、11）还存在着大量的片段重复，这表明 *MoAP2/ERF* 存在着一定的同源进化关系。最后，本研究可视化了

MoAP2/ERFs 的表达谱（图 6），通过基因表达谱筛选出在巴戟天根部高表达的基因 13 条，其中表达量最高的是 *MoAP2/ERF82*、*MoAP2/ERF31*、*MoAP2/ERF84*，后续需要进一步相关实验来验证这些转录因子在巴戟天生物合成。

综上所述，本研究从已有的巴戟天基因组获得了 131 条 *AP2/ERF* 基因家族转录因子，分为 *AP2*、*ERF*、*DREB*、*RAV*、*Soloist* 5 个亚家族。以表达模式、保守结构域分析、共线性分析等推断出 13 条在根中有特异性表达的基因，为深入探究 *MoAP2/ERFs* 基因功能奠定理论基础。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] 沈巍, 孙艺琦, 张奇, 等. 不同产地和生长年限对巴戟天中环烯醚萜苷类成分形成和积累的影响 [J]. 中草药, 2022, 53(17): 5484-5490.
- [2] 卞云迪, 张驰, 王雪晴, 等. 小麦 *AP2/ERF* 转录因子家族生物信息学分析 [J]. 天津师范大学学报: 自然科学版, 2022, 42(4): 39-45.
- [3] Zhao M H, Li Y, Zhang X X, et al. Genome-wide identification of *AP2/ERF* superfamily genes in *Juglans mandshurica* and expression analysis under cold stress [J]. *Int J Mol Sci*, 2022, 23(23): 15225.
- [4] Nakano T, Suzuki K, Fujimura T, et al. Genome-wide analysis of the *ERF* gene family in *Arabidopsis* and rice [J]. *Plant Physiol*, 2006, 140(2): 411-432.
- [5] 王玲, 刘晓伟, 江纳, 等. 蔓花生 *AP2* 基因家族的生物信息学分析 [J]. 江苏农业科学, 2020, 48(14): 65-77.
- [6] 黄奕孜, 钱旺, 邱姗, 等. 光皮桦 *AP2/ERF* 基因家族鉴定与表达分析 [J]. 浙江农林大学学报, 2022, 39(6): 1183-1193.
- [7] 付春, 刘晓伟, 王玲, 等. 阿拉伯岩芥 *AP2* 基因家族的生物信息学分析 [J]. 安徽农业科学, 2020, 48(8): 114-123.
- [8] 戴晓港, 李淑娟. 柳树 *AP2/ERF* 基因家族全基因组鉴定和表达分析 [J]. 江苏林业科技, 2021, 48(5): 1-12.
- [9] Ji A J, Luo H M, Xu Z C, et al. Genome-wide identification of the *AP2/ERF* gene family involved in active constituent biosynthesis in [J]. *Plant Genome*, 2016, 9(2): 23-36.
- [10] Yao W, An T Y, Xu Z C, et al. Genomic-wide identification and expression analysis of *AP2/ERF* transcription factors related to andrographolide biosynthesis in *Andrographis paniculata* [J]. *Ind Crops Prod*, 2020, 157: 112878.
- [11] Guan R, Xu S, Lu Z M, et al. Genomic characterization of bZIP transcription factors related to andrographolide biosynthesis in *Andrographis paniculata* [J]. *Int J Biol*

- Macromol*, 2022, 223(Pt A): 1619-1631.
- [12] Zhang R R, Chen Z Z, Zhang L B, *et al.* Genomic characterization of WRKY transcription factors related to andrographolide biosynthesis in *Andrographis paniculata* [J]. *Front Genet*, 2020, 11: 601689.
- [13] Zhang L B, Guan R, Li G L, *et al.* Genomic identification of ARF transcription factors and expression analysis in *Cannabis sativa* L. [J]. *Ind Crops Prod*, 2022, 186: 115118.
- [14] Zhang M, Li S T, Nie L, *et al.* Two jasmonate-responsive factors, TcERF12 and TcERF15, respectively act as repressor and activator of tasy gene of taxol biosynthesis in *Taxus chinensis* [J]. *Plant Mol Biol*, 2015, 89(4/5): 463-473.
- [15] 崔馨文, 王燕燕. 喜树转录因子 CaERF1 原核表达及蛋白纯化 [J]. *分子植物育种*, 2024, 10: 1-6.
- [16] Zhang Y, Ji A J, Xu Z C, *et al.* The AP2/ERF transcription factor SmERF128 positively regulates diterpenoid biosynthesis in *Salvia miltiorrhiza* [J]. *Plant Mol Biol*, 2019, 100(1/2): 83-93.
- [17] Chen C J, Wu Y, Li J W, *et al.* TBtools-II: A “one for all, all for one” bioinformatics platform for biological big-data mining [J]. *Mol Plant*, 2023, 16(11): 1733-1742.
- [18] Lu S N, Wang J Y, Chitsaz F, *et al.* CDD/SPARCLE: The conserved domain database in 2020 [J]. *Nucleic Acids Res*, 2020, 48(D1): D265-D268.
- [19] Paysan-Lafosse T, Blum M, Chuguransky S, *et al.* InterPro in 2022 [J]. *Nucleic Acids Res*, 2023, 51(D1): D418-D427.
- [20] Gasteiger E, Hoogland C, Gattiker A, *et al.* Protein identification and analysis tools on the ExPASy server [M] // Totowa: *Humana Press*, 2005: 571-607.
- [21] Bailey T L, Johnson J, Grant C E, *et al.* The MEME suite [J]. *Nucleic Acids Res*, 2015, 43(W1): W39-W49.
- [22] Rombauts S, Déhais P, van Montagu M, *et al.* PlantCARE, a plant cis-acting regulatory element database [J]. *Nucleic Acids Res*, 1999, 27(1): 295-296.
- [23] Wang Y P, Tang H B, Debarry J D, *et al.* MCSscanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity [J]. *Nucleic Acids Res*, 2012, 40(7): e49.
- [24] Yip K M, Xu J, Zhou S S, *et al.* Characterization of chemical component variations in different growth years and tissues of morinda officinalis *Radix* by integrating metabolomics and glycomics [J]. *J Agric Food Chem*, 2019, 67(26): 7304-7314.
- [25] 杨艳妮, 袁培培, 冯敖梓. 基于网络药理学和分子对接技术探讨巴戟天治疗类风湿性关节炎的作用机制 [J]. *中草药*, 2022, 53(5): 1463-1470.
- [26] 郎霞, 张海虹, 李会芳. 巴戟天治疗抑郁症在整合药理学下的作用机制探索 [J]. *药物评价研究*, 2019, 42(1): 64-71.
- [27] 伍秀珠, 罗登花, 赖宏武, 等. 广东韶关地区栽培巴戟天药材质量及其影响因素研究 [J]. *中药材*, 2022, 45(9): 2177-2182.
- [28] Li X, Xu Y Y, Shen S L, *et al.* Transcription factor CitERF71 activates the terpene synthase gene CitTPS16 involved in the synthesis of *E*-geraniol in sweet orange fruit [J]. *J Exp Bot*, 2017, 68(17): 4929-4938.
- [29] 悦曼芳, 张春, 吴忠义. 植物转录因子 AP2/ERF 家族蛋白结构和功能的研究进展 [J]. *生物技术通报*, 2022, 38(12): 11-26.
- [30] 蒋科技, 皮妍, 侯嵘, 等. 植物内源茉莉酸类物质的生物合成途径及其生物学意义 [J]. *植物学报*, 2010, 45(2): 137-148.
- [31] Chen H, Jones A D, Howe G A. Constitutive activation of the jasmonate signaling pathway enhances the production of secondary metabolites in tomato [J]. *FEBS letters*, 2006, 580(11): 2540-2546.

[责任编辑 时圣明]