

doi: 10.3969/j.issn.0490-6756.2019.05.030

簸箕柳 AP2/ERF 家族基因的全基因组分析

田应明, 王皆恒, 沈方圆, 李校
(四川大学生命科学学院 四川省分子生物与生物技术重点实验室, 成都 610064)

摘要:为了研究簸箕柳 AP2/ERF 转录因子家族的成员和分类,及后续深入探究其在簸箕柳生长发育过程中的功能,本文基于簸箕柳基因组数据,利用生物信息学技术和方法,对簸箕柳的 AP2/ERF 家族基因进行鉴定、分类及染色体定位,并对其理化性质、基因结构、保守结构域、表达模式等进行分析。结果表明,簸箕柳含有 178 个 AP2/ERF 家族基因,并根据 AP2/ERF 保守域的数量与拟南芥 AP2/ERF 家族基因的进化关系将其划分成 AP2、ERF、RAV 和 Soloist 亚家族 4 类。此外组织表达模式分析结果显示,簸箕柳 AP2/ERF 家族基因的表达具有明显的组织特异性,且在 leaf、root、shoot、flower 中表达较强,在其他组织中表达较弱。

关键词:簸箕柳;AP2/ERF 转录因子;生物信息学;表达模式

中图分类号: Q943.2 **文献标识码:** A **文章编号:** 0490-6756(2019)05-0983-06

Genome-wide analysis of the AP2/ERF family genes in *Salix suchowensis*

TIAN Ying-Ming, WANG Jie-Heng, SHEN Fang-Yuan, LI Xiao
(Sichuan Key Laboratory of Molecular Biology and Biotechnology,
College of Life Sciences, Sichuan University, Chengdu 610064, China)

Abstract: In order to study the members and classification of the AP2/ERF transcription factor family of *Salix suchowensis*, and to further explore its function in the growth and development, the AP2/ERF family genes were identified, classified, chromosomed, and their physical and chemical properties, gene structure, conserved domains, and expression patterns were also analyzed using bioinformatics techniques and methods. The results showed that *Salix suchowensis* contains 178 AP2/ERF genes, which were classified into four subfamilies: AP2, ERF, RAV, and Soloist according to the evolutionary relationship between the AP2/ERF conserved domain and the AP2/ERF genes of *Arabidopsis thaliana*. In addition, the results of tissue expression pattern analysis showed that the AP2/ERF family genes had obvious tissue specificity, it was strongly expressed in leaf, root, shoot, and flower, and weakly expressed in other tissues.

Keywords: *Salix suchowensis*; AP2/ERF transcription factors; Bioinformatics; Expression pattern

1 引言

AP2/ERF 家族是植物中最大的转录因子家族之一,具有众多的家族成员,占植物转录因子

的 6%~7%^[1]。研究表明,AP2/ERF 转录因子家族在植物次生代谢、花器官发育、细胞增殖、植物激素应答及胁迫响应等生命活动中发挥重要作用。AP2/ERF 家族基因至少含有 1 个 AP2 结合域,

收稿日期: 2018-09-25

基金项目: 国家自然科学基金(61001149)

作者简介: 田应明(1994—),男,硕士生,研究方向为生物信息学。E-mail: yingmtian@gmail.com

通讯作者: 李校。E-mail: lix@scu.edu.cn

包含 60~70 个氨基酸, 且高度保守。根据 AP2/ERF 家族基因所含 AP2 结构域数量的不同及序列的相似程度将 AP2/ERF 家族基因分为 AP2、ERF、RAV 和 Soloist 亚家族^[2-3], 其中 AP2 亚家族含有 2 个 AP2 结构域, 该家族主要通过结合靶序列 GCAC(A/G)n(A/T)TCCC(A/G) 和(C/T) 来调节花、胚珠、种子等植物器官特异性的生长发育^[4-6]; RAV 亚家族含有 1 个 AP2 和 B3 结构域, 该家族参与对乙烯、油菜素内酯的刺激反应^[7-8]及对于非生物胁迫的响应^[9]; ERF 亚家族含 1 个 AP2 结构域, Sakuma 等^[10] 将拟南芥的 147 个 AP2/ERF 家族基因做进一步分类, 将 ERF 亚家族根据其保守氨基酸序列的不同分为 ERF、DREB 和 Soloist 亚家族基因, 其中 ERF 亚家族基因可以特异性结合 GCC-box 中的核心序列 AGCCGCC 并调控抗病相关基因的表达^[11], 而 DREB 亚家族基因则大多通过识别 DRE/CRT 中的 A/GC-CGAC 序列来调控干旱、低温、高盐等非生物胁迫应答基因的表达^[12]; Soloist 亚家族虽然也只含 1 个 AP2 结构域, 但与其他 AP2/ERF 家族基因的同源性较低。

簸箕柳 (*Salix suchowensis*) 属杨柳科 (Salicaceae) 柳属 (*Salix*), 是中国特有的树种, 分布在山东、河南、浙江、淮河中下游冲积平原、江苏等地, 其枝条强韧, 可用来编柳条箱、筐篮、农具等, 又可以作防风固沙树种, 具有很高的生态与经济价值。随着簸箕柳全基因组测序完成^[13], 可以在分子水平对簸箕柳的一些非生物胁迫因子和基因家族的功能与表达分析进行深层次研究, 这给科学工作者们带来了一个崭新的研究簸箕柳乃至柳属、杨柳科的平台。本研究通过生物信息学分析方法对簸箕柳 AP2/ERF 家族基因进行挖掘鉴定, 系统全面地分析簸箕柳 AP2/ERF 家族基因的染色体定位、理化性质、基因结构和表达模式等, 为进一步研究簸箕柳 AP2/ERF 家族基因的抗逆性表达及功能基因的挖掘奠定基础, 为簸箕柳的人工栽培及开发利用提供了方法和理论依据。

2 材料与方法

2.1 材 料

从簸箕柳数据库 (<http://115.29.234.170/node/3>) 下载了簸箕柳的基因组、蛋白序列及注释文件, 其中蛋白序列共 26599 条; 从 TAIR (<http://www.arabidopsis.org/>) 中下载了拟南芥蛋

白序列; 毛果杨作为第一个发布全基因组序列信息的杨柳科物种, 对于杨柳科其他相关物种的研究具有极为重要的参考意义^[14], 所以从 JGI (<https://phytozome.jgi.doe.gov/>) 下载了毛果杨蛋白序列; 从 PlantTFDB (<http://planttfdb.cbi.pku.edu.cn>) 数据库下载了拟南芥和毛果杨的 AP2/ERF 家族的基因信息。

2.2 方 法

2.2.1 簸箕柳 AP2/ERF 家族基因的鉴定 首先, 分别以拟南芥和毛果杨的 AP2/ERF 家族基因的蛋白序列为查询序列, 用 BLASTP 对簸箕柳蛋白数据库进行本地 blast 检索; 其次, 从 Pfam 数据库 (<http://pfam.janelia.org/>) 下载 AP2/ERF 保守结构域的 HMM (Hidden Markov Model) 文件: PF00084, 对簸箕柳蛋白数据库进行 HMM 本地搜索。将两种方法得到的候选基因整合, 去除冗余, 然后通过搜索 SMART 数据库 (<http://smart.embl-heidelberg.de/>) 和 Pfam 数据库, 剔除候选基因蛋白序列中没有完整 AP2/ERF 保守结构域的序列, 再根据 AP2/ERF 保守结构域的数量及结构特点, 对 AP2/ERF 家族基因进行初步分类。

2.2.2 簸箕柳 AP2/ERF 家族基因进化分析及进一步分类 将簸箕柳与拟南芥的 AP2/ERF 家族基因序列保存为 FASTA 格式, 利用 ClustalW 进行序列比对, 然后用 MEGA7.0, 采用邻接法 (neighbor-joining, NJ), 检验参数 (bootstrap) 设置为 1000 次, 进行系统发育分析。然后按照 Nakano 等人^[2] 对拟南芥的分组方法, 又将 ERF 亚家族进一步进行分类。

2.2.3 簸箕柳 AP2/ERF 家族基因染色体定位及理化性质分析 使用 MapInspect 工具对这 178 个 AP2/ERF 家族基因进行基因定位, 并制做定位图。然后利用在线服务器 PROTPARAM (<http://web.expy.org/protparam/>), 对簸箕柳 AP2/ERF 家族基因序列的理论等电点 (pI)、分子量 (MW) 等信息进行分析, 并将其统计整理。

2.2.4 簸箕柳 AP2/ERF 家族基因基因结构、保守结构域分析 利用在线软件 GSDS2.0 (<http://gds.cbi.pku.edu.cn/>) 对簸箕柳 AP2/ERF 家族基因序列的基因结构进行分析; 利用在线软件 MEME V4.11.1 (<http://meme-suite.org/tools/meme>) 对簸箕柳 AP2/ERF 家族基因序列进行 motif 分析, 预期搜寻 motif 数量设置为 6。

2.2.5 篦箕柳 AP2/ERF 家族基因表达分析 首先根据篦箕柳 AP2/ERF 家族基因的 ID 提取其对应的 CDS 序列, 然后从 NCBI 上下载杨柳科的 EST 数据, 共 423238 条, 使用 blastn 搜索比对, 选取结果中 score>90 和 identity>95% 的, 并删除其中的冗余序列。然后根据最终获得的 EST 序列从 EST 数据信息中找出每条序列来自哪个组织, 进行归类总结和组织表达分析。

3 结果与分析

3.1 篦箕柳 AP2/ERF 家族基因的鉴定

最终从篦箕柳基因组中鉴定出 178 个 AP2/ERF 家族基因, 其中 23 个基因含有 2 个 AP2 结构域, 被归为 AP2 亚家族; 4 个基因含有 1 个 AP2 和 B3 结构域, 被归为 RAV 亚家族; 151 个基因含有 1 个 AP2 结构域, 暂时被归为 ERF 亚家族(表 1)。

表 1 篦箕柳、毛果杨和拟南芥中 AP2/ERF 家族基因概况

Tab. 1 Summary of the AP2/ERF genes in *S. suchowensis*, *P. trichocarpa* and *A. thaliana*

Plant			<i>S. suchowensis</i>		<i>P. trichocarpa</i>		<i>A. thaliana</i>	
Family	Subfamily	Group	Number	Percent/%	Number	Percent/%	Number	Percent/%
AP2			23	12.92	26	12.87	18	12.24
ERF		Total	150	84.27	169	83.66	122	82.99
	DREB	Total	61	34.27	66	32.67	57	38.78
		I	10	5.62	5	2.48	10	6.80
		II	12	6.74	20	9.90	15	10.20
		III	30	16.85	35	17.33	23	15.65
		IV	9	5.06	6	2.97	9	6.12
	ERF	Total	89	50.00	103	50.99	66	44.90
		V	14	7.87	10	4.95	6	4.08
		VI	8	4.49	11	5.45	8	5.44
		VII	6	3.37	6	2.97	5	3.40
		VIII	19	10.67	17	8.42	15	10.20
		IX	28	15.73	42	20.79	17	11.56
		X	9	5.06	9	4.46	8	5.44
		VI-L	2	1.12	4	1.98	4	2.72
		Xb-L	3	1.69	4	1.98	3	2.04
RAV			4	2.25	6	2.97	6	4.08
Soloist			1	0.56	1	0.50	1	0.68
Total			178		202		147	

3.2 篦箕柳 AP2/ERF 家族基因进化分析及进一步分类

为了比较篦箕柳与拟南芥 AP2/ERF 基因的进化关系, 及更好地对篦箕柳 AP2/ERF 家族基因进行分类, 把篦箕柳 178 条 AP2/ERF 家族基因序列与拟南芥 147 条 AP2/ERF 家族基因序列作系谱发生树, 结果显示, 进化树有 15 个分支, 包括 AP2、RAV、ERF 和 Soloist 亚家族, 按照 Nakano 等人^[2]对拟南芥的分组方法, 又将其中的 ERF 亚家族分为 12 个亚组, 其中 DREB 亚家族分为 4 组(I、II、III、IV), 每组分别包括 10、12、30、9 个成员, ERF 亚家族分为 8 组(V、VI、VII、VIII、IX、X、VI-L、Xb-L), 每组分别包括 14、8、6、

19、28、9、2、3 个成员。其中 willow_GLEAN_10020148 基因虽然含有一个 AP2/ERF 结构域, 但与其他基因的同源性较低, 且与拟南芥的 Soloist 家族基因(AT4G13040.3)同源, 所以也被归为 Soloist 家族(图 1)。

3.3 篦箕柳 AP2/ERF 家族基因染色体定位及理化性质分析

按照毛果杨中 AP2/ERF、WRKY 和 C2H2 等转录因子的命名方法^[15-17]对这篦箕柳 178 个 AP2/ERF 家族基因进行命名。然后使用 MapInspect 工具对篦箕柳 178 个 AP2/ERF 家族基因序列进行基因定位发现, 有 4 个基因无法定位到篦箕柳 19 条染色体上, 而其余 174 个基因则不均

匀、无规律地分布于 19 条染色体上，其中 16 号染色体上 AP2/ERF 家族基因数目最多，有 22 个；12 号染色体上最少，只有两条(图 2)。对 AP2/ERF 家族基因序列进行 ExPASY 分析，结果表明 178 个 AP2/ERF 家族基因编码的蛋白全长为 117~1226 aa，且其中 AP2 家族普遍较长，平均长度达 540 左右，这或许与 AP2 家族含有 2 个 AP2 结构域有关，其余几个亚家族，平均长度在 200~300 之间；此外它们的相关分子量为 13.3~138.2 kDa，等电点从 4.05 到 10.75，其中簸箕柳 AP2/ERF 家族基因氨基酸理论等电点大多数在酸性范围内，表明该家族蛋白富含酸性氨基酸。

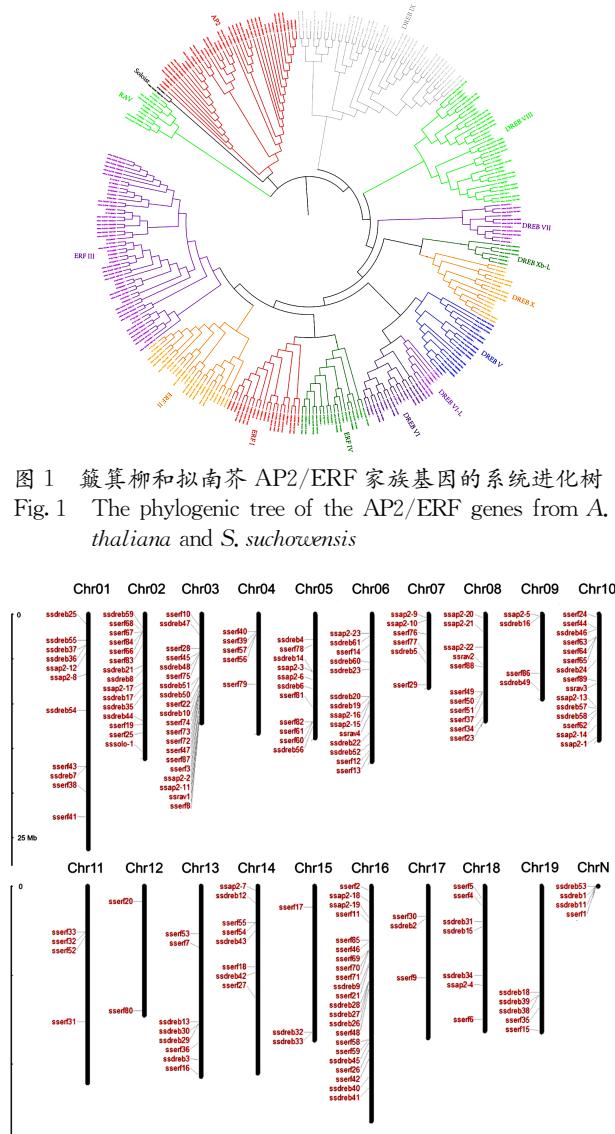


图 3 篓箕柳 AB2/FBF 家族基因染色体定位

图2 蚕豆AP2/ERF家族基因染色体定位
Fig. 2 The localization of AP2/ERF genes in *S. sucho-*

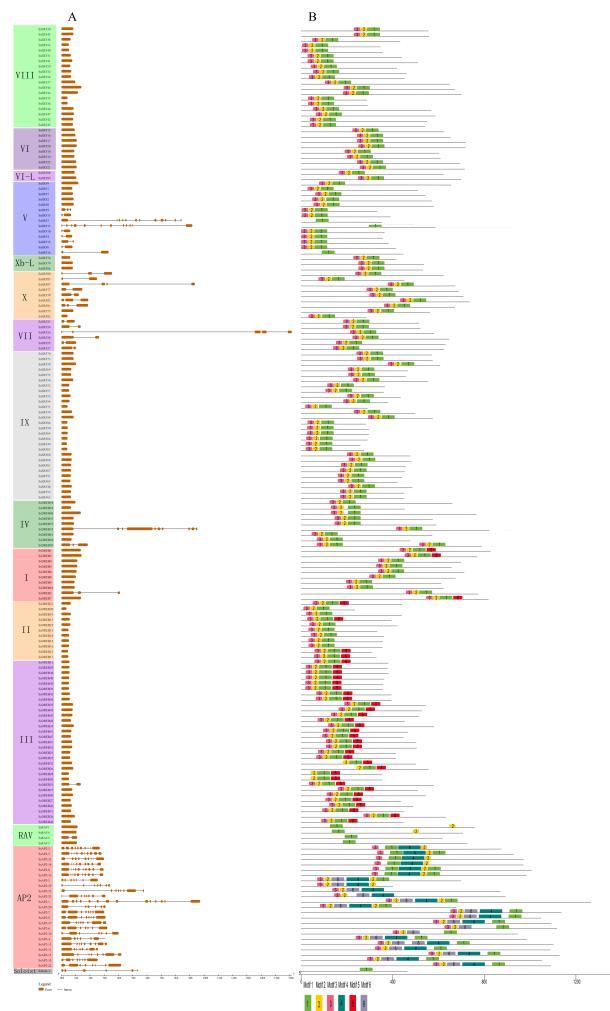


图3 篦箕柳AP2/ERF家族基因保守结构域及基因结构分析

A. AP2/ERF 家族基因外显子-内含子结构，框代表外显子，线条代表内含子

B. AP2/ERF 家族基因保守结构域的分布，带不同数字的框代表不同的保守结构域

Fig. 3 Analyses of gene structure and conserved motifs of AP2/FRE genes in *S. suchowensis*.

A. The graphic exon-intron structure of AP2/ERF genes, box indicates exons, and lines indicate introns.

B. The distribution of conserved motifs in each AP2/ERF genes. Boxes in different numbers represent different conserved motifs.

3.4 簕箕柳 AP2/ERF 家族基因基因结构、保守结构域分析

通过在线软件 GSDS2.0 对簾簾柳 AP2/ERF 家族基因序列的基因结构进行分析,结果显示大多数 DREB 亚家族、ERF 亚家族基因只包含 1 个外显子,没有内含子; RAV 亚家族中有 3 个基因只包含 1 个外显子,有 1 个基因含有两个外显子;而 AP2 亚家族和 Soloist 亚家族中的基因则含有多个外显子和内含子(图 3). 每个亚组中相似的外显子、内含子数目说明了他们之间较近的进化关系和相似的功能,这一结果和图 2 的进化结果一致。

进一步证实了 AP2/ERF 家族基因分类的准确性。

利用在线软件 MEME V4.11.1 对簸箕柳 AP2/ERF 家族基因序列进行 motif 分析, 预期搜寻 motif 数量设置为 6, 结果显示每个亚家族、亚组中基本含有相同的保守的 motif(图 3), 这也更进一步证明簸箕柳 AP2/ERF 家族基因进化和分类关系的准确性。

3.5 簸箕柳 AP2/ERF 家族基因表达分析

通过比对杨柳科的 EST 数据库, 共鉴定出 405 条 AP2/ERF ESTs 用来进行表达分析。如图 4 结果显示, AP2/ERF 家族基因在 9 个组织器官检测到表达。其中, 在 leaf 中表达最高, 达到 26.73%, 其次是 root, 为 18.81%, 接着是 shoot, 为 14.85%。然后分别是 flower(13.86%), seed(10.89%), petioles(5.94%), wood(3.96%)、stem(2.97%) 及 bud(1.98%)。其中 SsERF6、SsERF51 及 SsDREB2 只在 leaf 中表达; SsRAV4、SsERF88 及 SsERF89 只在 root 中表达; SsERF22、SsERF85 及 SsDREB55 只在 flower 中表达; SsERF46、SsERF47、SsERF54、SsDREB28 和 SsDREB43 只在 wood 中表达; SsAP2-22 及 SsDREB50 只在 shoot 中表达; Ss-DREB13 只在 bud 中表达。从簸箕柳 AP2/ERF 家族基因表达模式结果来看, 大部分 AP2/ERF 家族基因在 leaf、root、shoot 和 flower 中表达较强, 在其他组织中表达较弱。

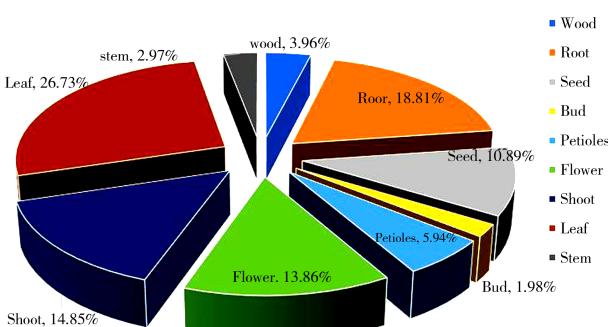


图 4 簸箕柳不同组织中 AP2/ERF 家族基因的表达分析
Fig. 4 Expression of AP2/ERF genes in different *S. suowensis* tissues in silico

4 讨 论

AP2/ERF 家族基因在植物的生长发育、胁迫应答等过程中具有重要作用。近年来随着拟南芥、水稻、毛果杨等物种全基因组测序的完成, 其 AP2/ERF 家族基因的全基因组分析也已被完成, 该领域的研究也越来越受到关注。

本研究对簸箕柳的 AP2/ERF 家族基因进行了分析, 共鉴定出 178 个 AP2/ERF 家族基因, 将其与拟南芥 AP2/ERF 家族基因进行进化分析, 并根据其序列的相似程度和结构域特征, 将其划分为 AP2(23 个)、ERF(150 个)、RAV(4 个)和 Soloist(1 个)4 个亚家族。将簸箕柳 AP2/ERF 家族基因与毛果杨、拟南芥的 AP2/ERF 家族基因进行对比发现, 毛果杨有 26 个 AP2、6 个 RAV 亚家族基因, 拟南芥有 18 个 AP2、6 个 RAV 亚家族基因, 物种间的 AP2、RAV 家族基因的数目比较接近, 这说明 AP2、RAV 亚家族基因高度保守。按照 Nakano 等人^[2]对拟南芥的分组方法, ERF 亚家族进一步被划分为 DREB 亚家族和 ERF 亚家族, DREB 亚家族又分为 4 组(I、II、III、IV), ERF 亚家族分为 8 组(V、VI、VII、VIII、IX、X、VI-L、Xb-L)。分别对其染色体分布、理化性质、基因结构及保守结构域等进行分析, 发现簸箕柳 AP2/ERF 家族基因在染色体上分布不均匀, 其中有 4 个基因无法定位到簸箕柳 19 条染色体上, 这可能是因为簸箕柳全基因组组装不够完善; Hu 等人^[18]认为, 外显子-内含子位置模式为进化关系提供了重要线索, Sakuma 等人^[12]的研究发现拟南芥的 ERF 家族基因大多数都没有内含子, 我们的研究结果发现簸箕柳 ERF 家族基因也具有类似特征, 而且每个亚家族基因都具有相似的基因结构, 进一步证明了 AP2/ERF 家族基因进化和分类的准确性; 此外, 每个亚家族基本含有相同的 motif, 这也预示着每个亚家族可能具有相似的功能。在对簸箕柳 AP2/ERF 家族基因进行表达分析时发现, 大部分 AP2/ERF 家族基因在 leaf、root、shoot 和 flower 中表达较强, 在其他组织中表达较弱, 这与白菜 (*Brassica rapa* ssp. *pekinensis*)^[3]、二穗短柄草 (*Brachypodium distachyon*)^[19] 及辣椒 (*Capsicum annuum* L.)^[20] 的研究结果一致, 这些 AP2/ERF 家族基因参与了植物生长发育的调节, 且在不同组织中差异表达。

综上所述, 通过对簸箕柳 AP2/ERF 家族基因进行分析, 为进一步挖掘和鉴定参与簸箕柳生长发育和胁迫应答调控的 AP2/ERF 家族功能基因提供了一定的理论基础, 更有助于后期簸箕柳的人工栽培及开发利用。

参考文献:

- [1] Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K.

- AP2/ERF family transcription factors in plant abiotic stress responses [J]. *Biochim Biophys Acta Gene Regul Mech*, 2012, 1819: 86.
- [2] 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: 411.
- [3] Song X, Li Y, Hou X. Genome-wide analysis of the AP2/ERF transcription factor superfamily in Chinese cabbage (*Brassica rapa* ssp. *pekinensis*) [J]. *BMC Genomics*, 2013, 14: 573.
- [4] Aukerman M J, Sakai H. Regulation of flowering time and floral organ identity by a microRNA and its APETALA2-like target genes [J]. *Plant Cell*, 2003, 15: 2730.
- [5] Elliott R C, Betzner A S, Huttner E, et al. AIN-TEGUMENTA, an APETALA2-like gene of *Arabidopsis* with pleiotropic roles in ovule development and floral organ growth [J]. *Plant Cell*, 1996, 8:155.
- [6] Jofuku K D, Omidyar P K, Gee Z, et al. Control of seed mass and seed yield by the floral homeotic gene APETALA2 [J]. *Proc Natl Acad Sci*, 2005, 102: 3117.
- [7] Alonso J M, Stepanova A N, Leisse T J, et al. Genome-wide insertional mutagenesis of *Arabidopsis thaliana* [J]. *Science*, 2003, 301: 653.
- [8] Hu Y X, Wang Y H, Liu X F, et al. *Arabidopsis* RAV1 is down-regulated by brassinosteroid and may act as a negative regulator during plant development [J]. *Cell Res*, 2004, 14: 8.
- [9] Sohn K H, Lee S C, Jung H W, et al. Expression and functional roles of the pepper pathogen-induced transcription factor RAV1 in bacterial disease resistance, and drought and salt stress tolerance [J]. *Plant Mol Biol*, 2006, 61: 897.
- [10] Sakuma Y, Maruyama K, Osakabe Y, et al. Functional analysis of an *Arabidopsis* transcription factor, DREB2A, involved in drought-responsive gene expression [J]. *Plant Cell*, 2006, 18: 1292.
- [11] Ohme-Takagi M, Shinshi H. Ethylene-inducible DNA binding proteins that interact with an ethylene-responsive element [J]. *Plant Cell*, 1995, 7: 173.
- [12] Sakuma Y, Liu Q, Dubouzet J G, et al. DNA-binding specificity of the ERF/AP2 domain of *Arabidopsis* DREBs, transcription factors involved in dehydration-and cold-inducible gene expression [J]. *Biochem Biophys Res Commun*, 2002, 290: 998.
- [13] Dai X, Hu Q, Cai Q, et al. The willow genome and divergent evolution from poplar after the common genome duplication [J]. *Cell Res*, 2014, 24: 1274.
- [14] 范李强, 胡欢, 郑洪蕾, 等. 响叶杨(杨属)叶绿体基因组测序与比较分析[J]. 四川大学学报: 自然科学版, 2018, 55: 165.
- [15] He H, Dong Q, Shao Y, et al. Genome-wide survey and characterization of the WRKY gene family in *Populus trichocarpa* [J]. *Plant Cell Rep*, 2012, 31: 1199.
- [16] Liu Q, Wang Z, Xu X, et al. Genome-wide analysis of C2H2 zinc-finger family transcription factors and their responses to abiotic stresses in poplar (*Populus trichocarpa*) [J]. *PLoS One*, 2015, 10: e0134753.
- [17] Zhuang J, Cai B, Peng R H, et al. Genome-wide analysis of the AP2/ERF gene family in *Populus trichocarpa* [J]. *Biochem Biophys Res Commun*, 2008, 371: 468.
- [18] Hu L, Liu S. Genome-wide identification and phylogenetic analysis of the ERF gene family in cucumbers [J]. *Genet Mol Biol*, 2011, 34: 624.
- [19] Chen L, Han J, Deng X, et al. Expansion and stress responses of AP2/EREBP superfamily in *Brachypodium distachyon* [J]. *Sci Rep*, 2016, 6: 21623.
- [20] Jin J H, Wang M, Zhang H X, et al. Genome-wide identification of the AP2/ERF transcription factor family in pepper (*Capsicum annuum* L.) [J]. *Genome*, 2018, 61: 663.

引用本文格式:

- 中 文: 田应明, 王皆恒, 沈方圆, 等. 篦箕柳 AP2/ERF 家族基因的全基因组分析[J]. 四川大学学报: 自然科学版, 2019, 56: 983.
- 英 文: Tian Y M, Wang J H, Shen F Y, et al. Genome-wide analysis of the AP2/ERF gene family in *Salix suchowensis* [J]. *J Sichuan Univ: Nat Sci Ed*, 2019, 56: 983.