

## · 基础研究 ·

基于丹参基因组的蛋白磷酸酶 2C 家族的系统分析<sup>△</sup>

徐志超, 浦香东, 宋经元\*

(中国医学科学院 北京协和医学院 药用植物研究所/国家中医药管理局 中药资源保护重点研究室, 北京 100193)

**[摘要]** **目的:** 基于全基因组策略系统解析药用植物丹参的蛋白磷酸酶 2C 家族成员, 为研究丹参逆境胁迫响应的分子机制奠定基础。**方法:** 采用 BLASTP 序列比对、系统发育树构建、基因表达分析方法解析丹参 PP2C 亚家族成员、基因结构以及差异表达图谱。**结果:** 丹参基因组共注释到 83 个 PP2C 家族成员, 在高等植物中保守存在; 系统进化分析将 PP2C 家族分为 13 个亚家族(PP2CA-PP2CL, A、B、C、D、E、F1、F2、G、H、I、J、K、L), 结构域高度保守; PP2CA 亚家族基因启动子含有多种逆境胁迫应答的顺式作用元件, 大部分 PP2CA 基因表达显著响应逆境胁迫信号。**结论:** 本实验系统研究丹参 PP2C 基因的结构、在不同组织部位的表达及逆境胁迫下的响应, 为揭示丹参响应生物或非生物胁迫的分子机制研究提供理论依据。

**[关键词]** 丹参; 蛋白磷酸酶; PP2C; 逆境胁迫

**Genome-wide Characterization of PP2C Gene Family in *Salvia miltiorrhiza***

XU Zhi-chao, PU Xiang-dong, SONG Jing-yuan\*

(Chinese Academy of Medical Sciences, Peking Union Medical College, Institute of Medicinal Plant Development; Key Lab of Chinese Medicine Resources Conservation, State Administration of Traditional Chinese Medicine of the People's Republic of China, Beijing 100193, China)

**[Abstract]** **Objective:** To systematically identify the protein phosphatases 2C (PP2C) family of *Salvia miltiorrhiza* based on genome-wide strategy. **Methods:** The family members of PP2C genes, structures, and expression profile were identified by BLSTP, phylogenetic tree, and different gene expression analysis. **Results:** In total, 83 PP2C genes were identified in *S. miltiorrhiza*, highly conserved in angiosperm. The PP2C family were phylogenetically divided into 13 subfamilies, including PP2CA, B, C, D, E, F1, F2, G, H, I, J, K, L subfamilies. The promoters of PP2CA subfamily included a series of cis-acting elements, and most of gene expression were significantly responded to stress. **Conclusion:** These results provided a basis for analyzing the evolutionary characterization of PP2C family, and for studying the molecular mechanism of PP2C genes under varies stress response.

**[Keywords]** *Salvia miltiorrhiza*; protein phosphatases; PP2C; stress response

doi:10.13313/j.issn.1673-4890.20180312001

蛋白激酶和蛋白磷酸酶所介导的可逆磷酸化反应在植物信号传递和逆境胁迫响应进程中发挥重要调控作用。蛋白磷酸酶按照底物特异性分为丝氨酸/苏氨酸蛋白磷酸酶类和酪氨酸蛋白磷酸酶类, PP2 蛋白磷酸酶作为丝氨酸/苏氨酸蛋白磷酸酶类的代表, 根据亚基结构等分为 PP2A、PP2B 和 PP2C 3 类<sup>[1]</sup>。拟南芥 *Arabidopsis thaliana* 的 PP2CA 亚家族已证实负调控 ABA 信号通路。化学遗传学和蛋白互作研究揭示 PYR/PYL/RCAR 复合体为 ABA 的受体,

当 ABA 信号缺失时, PP2C 通过去磷酸化使 SnRK2 激酶失去活性; 而响应 ABA 信号时, 受体 PYR/PYL/RCAR 与 ABA 结合后抑制 PP2C 的活性, SnRK2 激酶活性释放并磷酸化下游因子实现信号的传递及细胞对 ABA 胁迫的响应<sup>[2-3]</sup>。除了 ABA 信号途径, PP2C 还参与多种植物信号转导途径的调控, 如钾通道蛋白、MAPK 创伤信号途径等。拟南芥和水稻基因组分别注释到 80 和 78 个 PP2C 家族成员, 划分为 13 和 11 个亚家族, 负调控的 PP2CA 亚家族

<sup>△</sup> [基金项目] 国家自然科学基金青年科学基金(31700264)

\* [通信作者] 宋经元, 研究员, 研究方向: 本草基因组学; E-mail: jysong@implad.ac.cn

基因响应不同的胁迫刺激,如 ABA 胁迫响应,而 PP2CD 亚家族可能正调控 ABA 介导的信号通路<sup>[4]</sup>。基于基因组系统鉴定植物 PP2C 基因家族,有助于揭示其在植物逆境胁迫下的响应机制。

次生代谢产物是药用植物主要的药效成分,同时次生代谢产物的产生和积累也是植物对环境胁迫响应的一种适应。环境胁迫影响药用植物的品质,因此药用植物逆境胁迫的分子机制解析将为中药材种植环境及栽培技术的规范提供依据,推动中药材可持续发展。研究证实逆境胁迫虽抑制植物生长,但促进药用植物次生代谢产物的合成和积累,适当的逆境胁迫提升药用植物的品质<sup>[5-6]</sup>。如适度的干旱胁迫诱导甘草酸合成途径关键酶基因的上调,并促进甘草酸和甘草苷的积累<sup>[7]</sup>;轻度干旱可显著提高丹参活性成分丹参酮类和丹酚酸类化合物的合成与积累,同时增强植株对磷、钾、钙、等矿质元素的吸收与积累<sup>[8]</sup>。

丹参为常用大宗药材,遗传多样性丰富,是药用植物次生代谢研究理想的模式植物<sup>[9-12]</sup>。不同产地丹参品质参差不齐,研究逆境胁迫影响丹参品质的分子机制,将有助于丹参优良品种的种植农艺措施的改良。本研究基于丹参基因组注释 PP2C 基因家族,揭示 PP2C 在丹参不同组织部位的差异表达及对茉莉酸甲酯 MeJA 信号的响应,为丹参逆境胁迫分子机制的研究奠定基础。

## 1 材料和方法

### 1.1 丹参基因组、转录组数据

丹参基因组组装、注释及转录组测序工作由本课题组完成并已释放<sup>[10]</sup>。丹参基因组测序原始数据(Roche 454、PacBio RS、Illumina 平台)以及不同组织器官及处理的转录组数据(Illumina HiSeq2500 测序平台)已上传至 NCBI 的 SRA 数据库,登录号:SRX753381、SRR1640458、SRP028388、SRP051564<sup>[10-11,13]</sup>。

### 1.2 基于蛋白序列比对的丹参 PP2C 基因家族成员鉴定

下载拟南芥 PP2C 基因家族 A-L 亚家族的蛋白,采用 BLASTP 比对丹参基因组蛋白序列, Evalue 设置为 1E-5。从基因组数据库中提取所有 PP2C 序列,根据 NCBI 网站 BLASTP 比对的结果,利用 APOLLO 软件校正 PP2C 成员的 CDS、DNA、蛋白序列信息。在蛋白质数据库 PROSITE (<http://prosite.expasy.org/>)中分析所有蛋白的保守结构域序列。

### 1.3 PP2C 家族成员系统发育、基因结构和保守基序分析

将丹参和拟南芥 PP2C 家族成员的全长蛋白序列采用 MUSCLE 做多序列比对,构建 NJ (Neighbour-Joining) 树,选择 Jones-Taylor-Thornton (JTT) 模型,重复次数 1000 次。基于 CDS 序列及对应基因组序列,利用在线服务器 Gene Structure Display Server (GSDS 2.0) (<http://gsds.cbi.pku.edu.cn/index.php>)对基因结构进行分析。利用 MEME (Suite version 4.9.1) (<http://meme-suite.org/>)分析 PP2C 基因家族成员的保守基序,具体参数: E 值小于  $2 \times 10^{-30}$ ,基序重复次数不限,氨基酸个数 10~200 个。提取 PP2C 家族成员启动子区 1500 bp 通过 NEW PLACE 在线软件 (<https://sogo.dna.affrc.go.jp/cgi-bin/>)分析顺式作用元件。

### 1.4 PP2C 家族基因的差异表达分析

采用已报道<sup>[10-11,13]</sup>的丹参根组织(周皮、韧皮部、木质部)、根、茎、叶、花以及 MeJA 处理的转录组数据,通过 HISAT2 比对到丹参基因组,通过 cufflinks 计算丹参编码基因的 FPKM (Fragments Per Kilobase of exon model per Million mapped reads) 值,筛选出 PP2C 基因的表达量。FPKM 值通过 Z-score (Standard score) 标准化处理后,采用 R 语言的 gplots 下的 heatmap 来绘制热图。采用皮尔森相关系数 (Pearson's Correlation Tests,  $r$ ) 分析基因在不同组织部位的共表达情况。

## 2 结果和分析

### 2.1 丹参 PP2C 基因家族的注释及系统进化分析

在丹参基因组中共预测到 83 个 PP2C 基因成员,基因数量与高等植物中已报道的 PP2C 基因数量近乎一致,如拟南芥 (80)<sup>[4]</sup>、水稻 *Oryza sativa* (78)<sup>[4]</sup>、二穗短柄草 *Brachypodium distachyon* (86)<sup>[14]</sup>、谷子 *Setaria italica* (80)<sup>[15]</sup>。丹参 PP2C 的长度范围在 145 aa (SMil\_00028881) 到 1048 aa (SMil\_00011829),覆盖基因组的长度在 698 bp (SMil\_00028881) 到 8891 kb (SMil\_00011829)。

用丹参和拟南芥 PP2C 的氨基酸序列构建系统发育树,根据拟南芥的亚家族分类,丹参 PP2C 基因家族被分为 13 个亚家族 A~L (A、B、C、D、E、F1、

F2、G、H、I、J、K、L），各亚家族成员个数分别为11、7、6、8、9、6、4、9、5、2、1、7、2个，其中PP2CA亚家族成员数量最多，如图1所示。除此之外，6个成员聚为其他独立的分枝，暂未命名。丹参PP2C基因家族成员两两比对计算同源关系，发现PP2CF1亚家族的SMil\_00030307和SMil\_00028881的同源关系最近为98.57%，而SMil\_00011829(PP2CL)与SMil\_00021239(PP2CC)的同源关系最低为16.46%。

2.2 丹参PP2C基因家族的基因结构分析

PP2C基因家族的基因结构显示亚家族内和不同亚家族间基因结构差异较大，内含子数量变异较大，其中PP2CL亚家族的SMil\_00011829内含子数量最

多，为15个。丹参PP2CA亚家族的11个成员，命名为SmPP2CA1-SmPP2CA11，内含子数量为2~7个，多数为3个内含子，基因结构相似，如图2所示。

PP2CA亚家族各基因氨基酸大小在286~541 aa，等电点在4.78~6.84，分子量大小31~58 kD。对PP2CA亚家族的11个基序motif进行预测，Motif 1在11个亚家族成员中保守存在，大部分PP2CA成员保守存在motif 2、4、5、6、7。PP2A亚家族的系统进化分析显示3个同源分枝，其中同源分枝2和3的两对亚家族成员的基因结构和保守结构域基本相同。

2.3 PP2C基因的差异表达分析

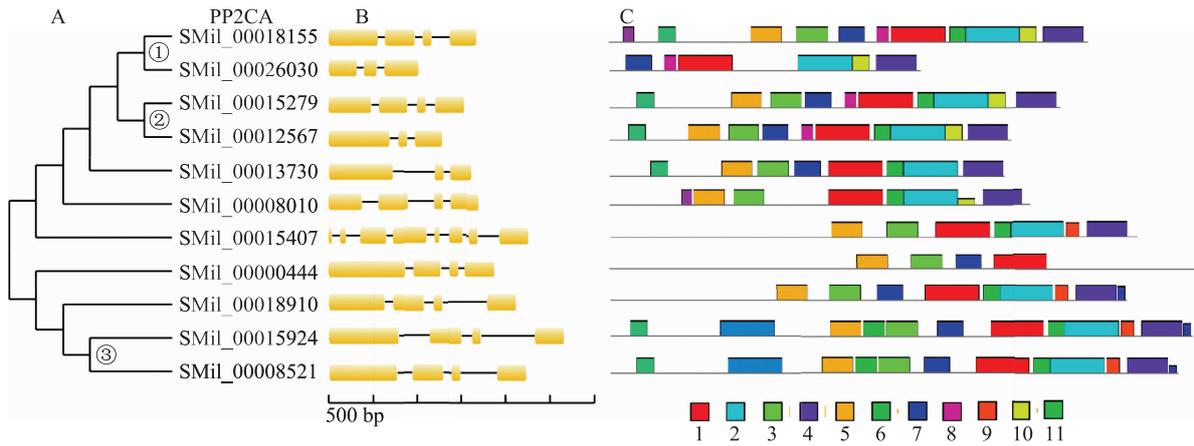
丹参酮是丹参重要的脂溶性活性成分，丹参根及根周皮是丹参酮的合成和积累部位<sup>[11]</sup>。本研究对



图1 丹参PP2C基因家族系统进化分析

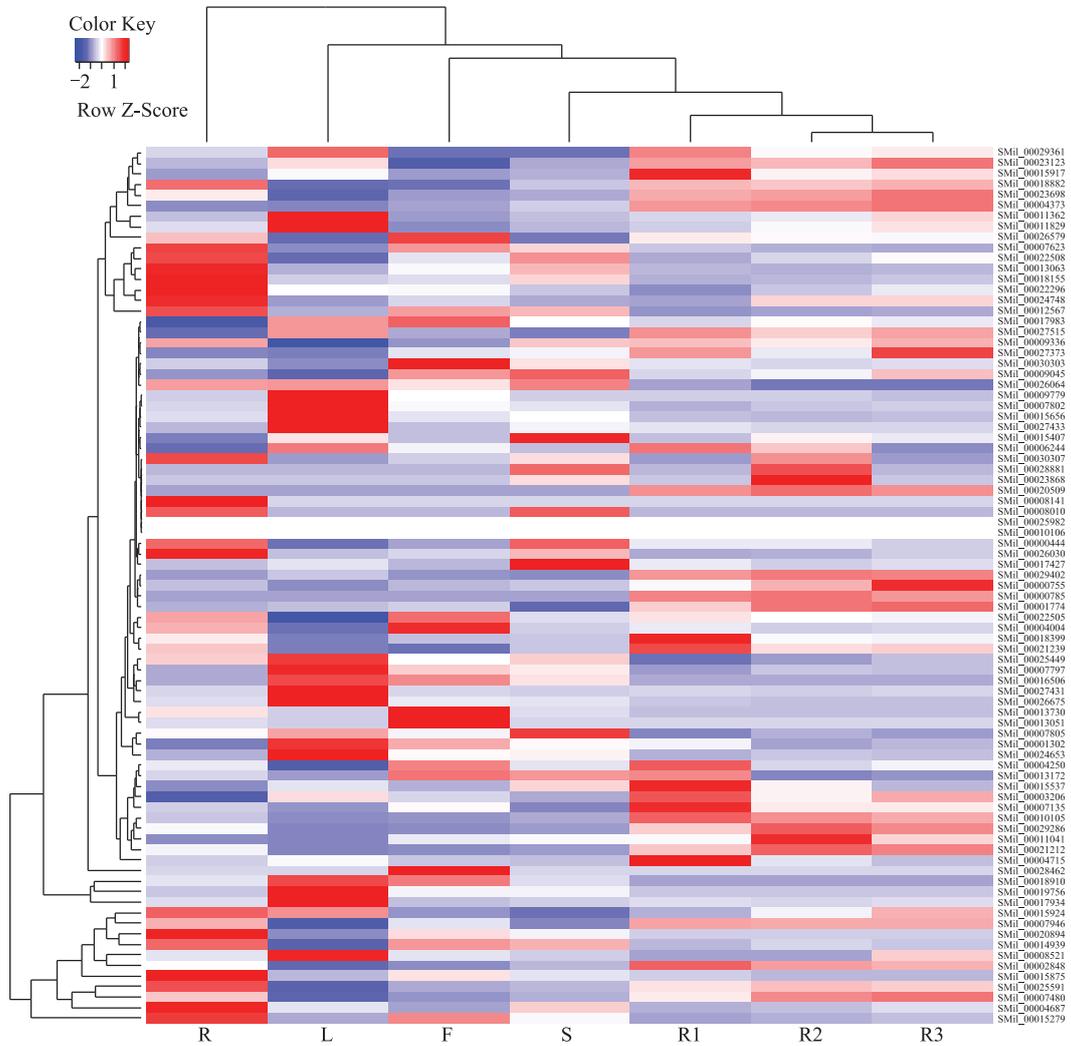
丹参不同组织部位, 如根、茎、叶、花、根周皮、根韧皮部和根木质部的转录组数据比对到丹参基因

组, 计算 FPKM 值, 将 FPKM 值用 scale 函数做 Z-score 标准化处理并聚类, 如图 3 所示。



注: A. PP2CA 系统进化树; B. PP2CA 基因结构; C. PP2CA 蛋白保守基序分布; ①~③代表序列同源分歧。

图 2 丹参 PP2CA 亚家族基因结构特征和保守基序分布



注: R. 根; L. 叶; F. 花; S. 茎; R1. 根周皮; R2. 根韧皮部; R3. 根木质部。

图 3 PP2C 基因家族成员在丹参不同组织部位中的差异表达图谱

差异表达结果显示 28.91% (24) 的基因低表达 (FPKM < 10), 其中 7 个基因显示沉默表达 (FPKM < 1)。丹参根 3 个组织部位 (周皮、韧皮部、木质部) 的 PP2C 基因表达聚在一起。组织特异性表达分析发现: PP2CE 亚家族的 SMil\_00028462 和 SMil\_00013051 两个基因在花中特异表达, 可能与花发育的调控相关; PP2CE 亚家族的 SMil\_00004715 在丹参根周皮特异高表达, 在茎、叶、花等器官表达极低, 推测间接调控丹参酮的生物合成。3 个 PP2CA 亚家族基因在丹参组织器官中表达低 (FPKM < 10), PP2CA 亚家族的同源分枝 3 的两个基因在丹参不同组织部位中表达最高, 且在根、茎、叶、花中表达差异不显著, 呈共表达 ( $r = 0.81$ )。茉莉酸甲酯 MeJA 作为重要的信号转导因子能够诱导植物次生代谢产物的产生, 在丹参中 MeJA 能够正调控丹参酮的生物合成。对于丹参叶片 MeJA 处理 12 h 后的转录组数据, 分析丹参 PP2C 家族成员编码基因的表达, 发现 13 个 PP2C 基因在 MeJA 处理下显著降低, 其中包含 6 个 PP2CA 和 3 个 PP2CK 亚家族的基因。PP2CA 亚家族同源分枝 2 的两个基因在丹参所有转录组 (不同组织器官及 MeJA 处理) 的表达中呈显著共表达 ( $r = 0.95$ )。

分析丹参 PP2CA 基因的启动子中的顺式作用元件发现: PP2CA 的启动子中 ABA 响应元件 (ABRE, ACGTGG) 均有分布 (见表 1), 大部分启动子区呈密集分布, 与响应 ABA 信号密切相关; 在 PP2CA 的启动子中富含 WRKY、MYB 和 MYC 3 类转录因子的作用元件 (见表 1), 推测在响应逆境胁迫时受转录因子的调控; 干旱和低温响应的元件部分分布。

表 1 丹参 PP2CA 亚家族成员启动子区顺式作用元件分析

PP2CA Gene IDs	ABRE	DRE	LTRE	WRKY	MYB	MYC	GCC
SMil_00026030	11	2	3	7	13	6	3
SMil_00012567	11	0	4	13	9	2	7
SMil_00013730	14	0	3	10	9	6	0
SMil_00015279	13	0	0	4	10	2	0
SMil_00015407	1	4	3	4	7	11	0
SMil_00018155	9	3	6	4	3	8	3
SMil_00018910	1	0	5	9	5	9	0
SMil_00015924	7	0	0	5	8	5	0
SMil_00008010	7	0	2	6	7	9	0
SMil_00008521	4	1	0	6	5	6	0
SMil_00000444	3	0	0	6	7	5	0

注: ABRE、DRE、LTRE 代表响应 ABA 信号、干旱、低温的作用元件; WRKY、MYB、MYC、GCC 代表转录因子 WRKY、MYB、MYC、AP2 的结合元件。

### 3 讨论

本研究基于丹参全基因组测序, 系统鉴定蛋白磷酸激酶 PP2C 家族成员, 丹参基因组注释的 PP2C 基因数量 (83) 与已报道的植物中相似; 丹参与拟南芥 PP2C 序列的系统进化显示丹参 PP2C 亚组的分类与拟南芥相似, 亚家族成员数量也大多相似, 保守结构域相似度高, 证明 PP2C 基因家族及其基因结构在植物中进化较为保守。低等维管植物 (石松纲) 卷柏 *Selaginella tamariscina*<sup>[16]</sup> 中的 PP2C 基因家族鉴定到 46 个成员 (未发表), 远少于已报道的高等植物, PP2C 在进化过程中应对复杂逆境环境时显著的基因扩张反映 PP2C 基因家族在植物逆境响应进程中重要的调控作用。

PP2CA 亚家族成员作为 ABA 的受体之一, 负调控 ABA 信号转导途径, 响应多种逆境胁迫, 如干旱、低温等<sup>[17-19]</sup>。丹参 PP2CA 基因启动子含有多种逆境胁迫应答的顺式作用元件如 ABRE、DRE、LTRE 等, 大多数的 PP2CA 基因响应 MeJA 信号转导因子。PP2CA 的同源分枝 2、3 的基因结构相似, motif 高度保守, 且显著共表达, 推测为基因组复制导致的相似或相同功能基因的分化。有报道指出植物中 ABA 信号与花发育相关, 如 ABA 信号途径中一个正调控因子 ABI4 (乙烯响应的转录因子) 正调控开花关键基因 FLC 的转录, 进而负调控开花时间<sup>[20]</sup>。丹参 PP2CE 亚家族的两个基因在花中特异表达, 一个基因在根中特异高表达, 预测 PP2CE 亚家族与植物组织器官发育相关, 如根和花的发育, 但需要进一步的实验验证。

丹参是最常用的大宗药材之一, 生态适应性较强, 种植面积广泛分布在华东、华北、西北、西南、中南等地区, 野生资源大幅减少, 主要来源于人工栽培品, 然而不同产地的丹参品质差异明显。研究指出轻度干旱可显著提高丹参活性成分丹参酮类和丹酚酸类化合物的合成与积累, 同时增强植株对矿质元素的吸收, 因此丹参对不同环境的适应导致活性成分合成和积累的差异, 揭示丹参对逆境胁迫的分子机制有助于指导丹参优良品种的栽培与繁育。丹参 PP2C 基因家族的系统分析将为丹参响应逆境胁迫分子机制的解析奠定基础, 同时丹参作为药用模式植物, 将对其他药用植物的栽培育种及道地性的理解具有指导意义。

## 参考文献

- [1] Shi Y. Serine/Threonine Phosphatases; Mechanism through Structure[J]. Cell, 2009, 139(3):468-484.
- [2] Zhu J K. Abiotic Stress Signaling and Responses in Plants[J]. Cell, 2016, 167(2):313.
- [3] Danquah A, Zelicourt A D, Colcombet J, et al. The role of ABA and MAPK signaling pathways in plant abiotic stress responses[J]. Biotechnol Adv, 2014, 32(1):40.
- [4] Xue T, Wang D, Zhang S, et al. Genome-wide and expression analysis of protein phosphatase 2C in rice and Arabidopsis[J]. BMC Genomics, 2008, 9(1):550.
- [5] 司灿, 张君毅, 徐护朝. 药用植物在干旱胁迫下生长代谢变化规律及应答机制的研究进展[J]. 中国中药杂志, 2014, 39(13):2432-2437.
- [6] 杨枝中, 马逾英, 卢晓琳, 等. 干旱和温度胁迫在药用植物研究中的应用现状与展望[J]. 时珍国医国药, 2010, 21(5):1223-1225.
- [7] 张春荣, 桑雪雨, 渠萌, 等. 基于转录组测序揭示适度干旱胁迫对甘草根基因表达的调控[J]. 中国中药杂志, 2015, 40(24):4817-4823.
- [8] 刘大会, 郭兰萍, 黄璐琦, 等. 土壤水分含量对丹参幼苗生长及有效成分的影响[J]. 中国中药杂志, 2011, 36(3):321-325.
- [9] 宋经元, 罗红梅, 李春芳, 等. 丹参药用模式植物研究探讨[J]. 药学学报, 2013(7):1099-1106.
- [10] Xu H, Song J, Luo H, et al. Analysis of the genome sequence of the medicinal plant *Salvia miltiorrhiza* [J]. Mol Plant, 2016, 9(6):949.
- [11] Xu Z, Peters R J, Weirather J, et al. Full-length transcriptome sequences and splice variants obtained by a combination of sequencing platforms applied to different root tissues of *Salvia miltiorrhiza* and tanshinone biosynthesis[J]. Plant J, 2015, 82(6):951.
- [12] Xu Z, Song J. The 2-oxoglutarate-dependent dioxygenase superfamily participates in tanshinone production in *Salvia miltiorrhiza* [J]. J Exp Bot, 2017, 68(9):2299-2308.
- [13] Luo H, Zhu Y, Song J, et al. Transcriptional data mining of *Salvia miltiorrhiza* in response to methyl jasmonate to examine the mechanism of bioactive compound biosynthesis and regulation[J]. Physiol Plantarum, 2014, 152(2):241.
- [14] Cao J, Jiang M, Li P, et al. Genome-wide identification and evolutionary analyses of the PP2C gene family with their expression profiling in response to multiple stresses in *Brachypodium distachyon* [J]. BMC Genomics, 2016, 17(1):175.
- [15] 闵东红, 薛飞洋, 马亚男, 等. 谷子 PP2C 基因家族的特性[J]. 作物学报, 2013, 39(12):2135-2144.
- [16] Banks J A. Selaginella and 400 million years of separation. [J]. Annu Rev Plant Biol, 2009, 60(60):223.
- [17] Zelicourt A D, Colcombet J, Hirt H. The Role of MAPK Modules and ABA during Abiotic Stress Signaling [J]. Trends Plant Sci, 2016, 21(8):677.
- [18] Zhu J K. Salt and Drought Stress Signal Transduction in Plants [J]. Annu Rev Plant Biol, 2002, 53(53):247.
- [19] Schweighofer A, Hirt H, Meskiene I. Plant PP2C phosphatases: emerging functions in stress signaling. [J]. Trends Plant Sci, 2004, 9(5):236-243.
- [20] Shu K, Chen Q, Wu Y, et al. ABSCISIC ACID-INSENSITIVE 4 negatively regulates flowering through directly promoting Arabidopsis FLOWERING LOCUS C transcription [J]. J Exp Bot, 2016, 67(1):195.

(收稿日期 2018-03-12)