

丹参 *ERF* 基因组序列克隆与多态性分析

崔光红, 冯 华, 李文渊, 汪婉宜, 黄璐琦*

(中国中医科学院中药研究所, 北京 100700)

摘要: 乙烯应答因子结合蛋白 (ethylene responsive factor binding protein, ERF) 属于 AP2/ERF 转录因子超家族中的 ERF 家族, 是植物所特有的一类转录因子, 在植物生长、发育以及应对各种环境刺激等方面均发挥重要的作用。本文从丹参中克隆得到一个丹参 *SmERF* 基因, 该基因属于拟南芥和水稻 ERF 家族中的 VII 组, 含有 CMVII-1 的蛋白模体, 序列为 MCGGAI (I/L), 且在 AP2/ERF 功能域的 5' 端含有一个内含子。通过对丹参主产区山东、陕西、四川 181 个样本的检测, 发现 *SmERF* 第二外显子区域存在大量的变异位点, 264 bp 范围内共存在 21 个单核苷酸多态性 (single nucleotide polymorphism, SNP), 有 14 个为非同义突变, 7 个为同义突变。不同产地样品具有产地特异的基因型, 其中有 5 个位点在不同产地间均表现出显著差异, 聚类分析表明 *SmERF* 基因可按产地聚类。该结果显示 *SmERF* 基因有望作为丹参产地鉴别的标识基因, 并为 *SmERF* 基因功能的确定奠定了基础。

关键词: 丹参; 乙烯应答因子结合蛋白; 多态性分析

中图分类号: R931.5

文献标识码: A

文章编号: 0513-4870 (2010) 09-1188-06

Cloning and polymorphism analysis of *SmERF* in *Salvia miltiorrhiza*

CUI Guang-hong, FENG Hua, LI Wen-yuan, WANG Wan-yi, HUANG Lu-qi*

(Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, China)

Abstract: The transcription factor of ethylene responsive factor binding protein (ERF) is belonged to AP2/ERF superfamily, which is known to be unique in plants. AP2/ERF proteins have important functions in the transcriptional regulation of a variety of biological processes related to growth and development, as well as various responses to environmental stimuli. An ERF gene from *Salvia miltiorrhiza* is cloned and divided into ERF gene family group VII of Arabidopsis and Rice. It contains a MCGGAI (I/L) motif referred to as CMVII-1 and a single intron in the 5'-flanking region of the AP2/ERF domain. Sequence analysis reveals that the region of second exon has abundant polymorphism sites. There are 21 single nucleotide polymorphism sites (SNPs) in the 264 bp region, among them, 14 SNPs are synonymous substitutions and 7 SNPs are non-synonymous substitutions. Though analysis of 181 samples from Shandong, Shaanxi and Sichuan Provinces, it reveals that each production area has its own special genotypes, 5 SNPs show significant difference. Cluster based on UPGMA method reveals that different populations from specific province have clustered together. It shows that *SmERF* gene will be a candidate molecular marker for the identification of *Salvia miltiorrhiza* from different areas.

Key words: *Salvia miltiorrhiza*; ethylene responsive factor binding protein; polymorphism analysis

收稿日期: 2010-05-04.

基金项目: 国家重点基础研究发展计划资助项目 (2006CB504700); 国家高技术研究发展计划项目资助项目 (2003AA2Z2040).

*通讯作者 Tel: 86-10-84738625, Fax: 86-10-84027175, E-mail: huangluqi@263.net

乙烯应答因子结合蛋白 (ethylene responsive factor binding protein, ERF) 属于 AP2/ERF 转录因子超家族中的 ERF 家族, 是植物所特有的一类转录因子^[1], 最早作为烟草和 PR 基因启动子中顺式作用元件 GCC 盒的结合蛋白分离得到^[2]。AP2/ERF 超家族根据 AP2/ERF 结构域, 可分为 3 个家族, 其中 AP2 家族含有两个重复的 AP2/ERF 结构域, ERF 家族含有一个 AP2/ERF 结构域, RAV 家族除含有一个 AP2/ERF 结构域外, 还含有一个 B3 结构域。AP2/ERF 家族基因在植物生长和发育相关的转录调控以及应对各种环境胁迫方面发挥着重要的作用^[3]。其中, ERF 家族基因主要调控植物激素 (乙烯和 ABA 等)、病原和胁迫 (低温、干旱及高盐) 的应答反应^[4]。

通过基因芯片技术, 已从丹参中克隆得到 1 个丹参乙烯应答因子结合蛋白基因 (*SmERF*), 并显示它可能与丹参的生长发育有关^[5]。本文通过对丹参基因组 DNA 的扩增, 明确了 *SmERF* 的基因组结构, 并通过山东、陕西、四川的 181 个样品第二外显子区域的检测, 发现该区域存在大量的变异位点, 并可按不同产地聚类。该结果显示 *SmERF* 可用于不同产地丹参的鉴别, 同时, 将为 *SmERF* 变异是否影响丹参次生代谢产物的积累和其生物学功能的确定奠定基础。

材料与方法

材料 分别采集主产区陕西、四川及山东的丹参新鲜叶片, 用硅胶快速干燥后用于 DNA 提取 (表 1)。样品包含了丹参常见的变异种类, 如山东特有的白花丹参 (LWBH)、四川中江的大叶型 (HXDY、JQDY、SQDY) 和小叶型丹参 (JQXY、HXXY)。除

山东莱芜 (LWZHY) 居群为野生外, 其余均为栽培丹参。

DNA 提取、引物、PCR 扩增和测序 采用改进的 CTAB 法提取丹参总 DNA。DNA 样品经分光光度计检测 DNA 模板的浓度, 用含溴化乙锭的 1% 的琼脂糖凝胶电泳, DNA 条带清晰的用于 PCR 扩增。根据 *ERF* 第二外显子序列设计引物, ERF-322: 5'-ATGG GGAAAATGGGCAGCTGAAATC-3', ERF-683: 5'-GG GGGTCTGAGGCTCCTGAGTCGGT-3'。为保证 SNP 位点判读的准确性, PCR 扩增时选用 TaKaRa 公司的 PrimeSTAR HS DNA 聚合酶, 该酶是兼具高保真性和高扩增效率的 PCR 用 DNA 聚合酶。50 μ L PCR 反应体系包括 5xPrimeSTAR Buffer (Mg^{2+} plus) 10 μ L、dNTP mixture (各 2.5 mmol·L⁻¹) 4 μ L、引物 (各 10 mmol·L⁻¹) 1 μ L、DNA 模板 (100 ng· μ L⁻¹) 1 μ L、DNA Polymerase (2.5 U· μ L⁻¹) 0.5 μ L 和水 32.5 μ L。PCR 反应条件为 94 $^{\circ}$ C 预变性 5 min 后开始 40 个循环, 98 $^{\circ}$ C 变性 10 s, 68 $^{\circ}$ C 退火 30 s, 72 $^{\circ}$ C 后延伸 1 min。PCR 反应液配制均在冰水中进行。取 5 μ L 反应产物在含 EB 的 1.5% (w/v) 的琼脂糖凝胶上电泳, 取单一条带由北京鑫科公司代为测序。

SNP 确认及遗传分析方法 测序得到的序列用 Staden Package (<http://staden.sourceforge.net>) 软件包中的序列拼接程序进行序列比对和分析, 序列中同一位点的不同碱基及双峰中两峰之比大于 2:1 的位点被确定为 SNP 位点, 每个 SNP 位点看成一个信息位点, 分为纯合和杂合位点, 结果输入 Microsoft Excel 工作表中。利用 POPGENE (v.1.31) (<http://www.ualberta.ca/~fyeh/>) 进行遗传多样性、平均杂合度及

Table 1 Detailed sources of materials used in this study

Group	Total sample	Population	Sample	Source	Collected time
SX	33	SL1	20	Shanyang, Shangluo, Shaanxi	Sep, 2007
		SL2	13	Shangluo, Shaanxi	Sep, 2007
SC	64	HXDY	11	Hexing, Zhongjiang, Sichuan	Oct, 2006
		JQDY	10	Jiqing, Zhongjiang, Sichuan	Oct, 2006
		SQDY	19	Shiquan, Zhongjiang, Sichuan	Oct, 2006
		JQXY	10	Jiqing, Zhongjiang, Sichuan	Oct, 2006
		HXXY	14	Hexing, Zhongjiang, Sichuan	Oct, 2006
SD	84	LYZH	28	Mengyin, Linyi, Shandong	Oct, 2007
		LWZH	18	Laiwu, Shandong	Oct, 2007
		LWZHY	18	Laiwu, Shandong	Oct, 2007
		LWBH	20	Laiwu, Shandong	Oct, 2007

遗传距离分析。用 MEGA3.0 构建用 UPGMA 算法得到的 Nei's unbiased distance (D) 树系图^[6]。

基因型频率分析 对于基因型频率在 10% 以上的位点利用 SHEsis 软件^[7]分析各个居群的基因型频率和基因频率, 并进行卡方检验分析各个位点的基因型频率在不同产地的差异, 显著性水平为 $P = 0.05$ 。

结果

1 *SmERF* 基因组序列克隆和分析

在 *SmERF* cDNA 序列两侧设计引物, 对丹参基因组 DNA 进行扩增, PCR 产物经克隆测序后与其 cDNA 序列进行比对, 发现在开放阅读框 216 bp 处插入一个 935 bp 的内含子, 内含子位于 AP/ERF 功能域的 5' 端。根据拟南芥和水稻中 ERF 基因的分类原则, 丹参 *SmERF* 属于拟南芥和水稻 ERF 家族基因中的 VII 组基因, 与 At3g16770 相似度最高, VII 组中拟南芥有 5 个基因, 水稻有 15 个基因。该类基因均含有 CMVII-1 的蛋白模体, 序列为 MCGGAI (I/L), 且均

在 AP2/ERF 功能域的 5' 端含有一个内含子^[3]。拟南芥中的 AtERF (At3g16770) 显示与一个 bZIP 的转录因子 (OBF4) 在体外互相作用, 但其具体功能不清楚^[8], 水稻中的 OsERF89 (Os03g08460) 与一个 Myc 转录因子 (OsBP-5) 互相作用, 与一个 31 bp 的顺式作用因子共同调控水稻中颗粒结合型淀粉合成酶 waxy 基因的表达^[9]。

2 *SmERF* 第二外显子序列分析和 SNP 确定

经过对 *SmERF* 不同区域的筛选, 发现该基因第二外显子区域存在大量变异。对 181 个样品的该区域进行 PCR 扩增和测序, 取测序准确的 264 bp 片段用于 SNP 分析, 发现该片段位于第二外显子 122~385 bp, 共存在 21 个 SNPs, 平均 13 bp 就有一个 SNP 位点。从 11 个居群扩增得到的 *SmERF* 片段序列长度完全一致, 每个 SNP 位点只有两种碱基, 除 JQDY、SQDY 和 JQXY 的 SNP 位点完全一致外, 其他居群的 SNP 位点均存在差异。21 个位点中, 有 14 个为非同义突变, 7 个为同义突变 (表 2)。

Table 2 Distribution of SNPs in the 264 bp sequenced of *SmERF* from 11 populations. ^aThe exact position of the SNP is indicated. ^bHeterozygotes are indicated by K (G/T), R (G/A), S (G/C), W (A/T), Y (T/C), M (A/C) and were detected by double peak in the electropherograms. ^cConsensus sequence is indicated by Con. ^dSyn and Nsyn represented synonymous and non-synonymous substitution, respectively

No.	Position ^a /bp	Population ^b												Con. ^c	AA.	Syn/ Nsyn ^d
		SL1	SL2	HXDY	JQDY	SQDY	JQXY	HXXY	LYZH	LWZH	LWZHY	LWBH				
1	19	T/C/Y	T/C/Y	T/Y	Y	Y	Y	Y	T/C/Y	T/C/Y	T/Y	T/C/Y	T	F-F	Syn	
2	52	C/Y	C	C	C	C	C	C/Y	C/T/Y	C/T/Y	C/T/Y	C/Y	C	D-D	Syn	
3	58	G	G	G	G	G	G	G	G/C/S	G/C/S	G/S	G/S	G	A-A	Syn	
4	77	G	G	A/R	R	R	R	G/A/R	G/A/R	G/A/R	G/A/R	G/R	G	D-N	Nsyn	
5	117	G/S	G	G/S	G/S	G/S	G/S	G/C/S	G/S	G/C/S	G/S	G/S	G	R-P	Nsyn	
6	137	C	C	C/Y	C/Y	C/Y	C/Y	C/T/Y	C/T/Y	C/Y	C/Y	C/Y	C	P-S	Nsyn	
7	146	C	C	C-M	C-M	C-M	C-M	C/A/M	C/A/M	C/M	C/M	C/M	C	R-R	Syn	
8	155	G/R	G/R	G	G	G	G	G	G/R	G	G	G	G	V-I	Nsyn	
9	164	G/A/R	G/A/R	G	G/R	G/R	G/R	G/R	G/A/R	G/A/R	G/A/R	G/A/R	G	E-G	Nsyn	
10	168	C	C	C/Y	C/Y	C/Y	C/Y	C/Y	C/T/Y	C/Y	C/Y	C/Y	C	P-L	Nsyn	
11	187	G	G	G/R	G/R	G/R	G/R	G/R	G/A/R	G/R	G/R	G/R	G	P-P	Syn	
12	206	G/A/R	G/A/R	G/A/R	R	R	R	G/A/R	G/A/R	G/A/R	G/A/R	G/R	G	E-K	Nsyn	
13	222	G/A/R	G/A/R	G/A/R	R	R	R	G/A/R	G/A/R	G/A/R	G/A/R	A/R	G	C-F	Nsyn	
14	158	A	A/R	A	A	A	A	A	A	A	A	A	A	I-V	Nsyn	
15	159	C	C/Y	C	C	C	C	C	C/Y	C	C	C	C	I-T	Nsyn	
16	171	C	C	C	C	C	C	C	C/Y	C/Y	C/T/Y	C/Y	C	P-L	Nsyn	
17	185	C/Y	C	C	C	C	C	C	C	C	C	C	C	P-S	Nsyn	
18	195	C	C	C	C	C	C	C	C/Y	C/T/Y	C/Y	C	C	A-V	Nsyn	
19	201	G	G	G	G	G	G	G	G/S	G/S	G/S	G/S	G	S-T	Nsyn	
20	212	C/Y	C	C	C	C	C	C	C	C	C	C	C	L-L	Syn	
21	241	G/R	G	G	G	G	G	G	G	G	G	G	G	Q-Q	Syn	
N	10	7	9	10	10	10	10	11	17	15	15	14	21 (total)		Nsyn/syn	
Percentage (%)		47.62	33.33	42.86	47.62	47.62	47.6	52.38	80.95	71.43	71.4	66.67	53.81 (Ave.)		=14/7	

不同居群的多态位点百分率为 33.33%~80.95%, 平均多态位点百分率为 53.81% (表 2)。SNP 位点最多的居群是 LYZH, 其次是 LWZH 和 LWZHY, 分别为 17、15、15 个 SNP 位点, SNP 位点最少的是 SL2, 仅有 7 个 SNP 位点。21 个位点中有 8 个位点的频率小于 10%, 属于稀有等位基因, 其中有 3 个仅出现在陕西的 SL1 中, 1 个仅出现在陕西的 SL2 中, 3 个出现在山东的各个居群中, 另有 1 个出现在 SL2 和 LYZH 中, 可见不同产地存在着一些独特的基因位点。

3 居群间遗传关系

不同居群之间的遗传距离介于 0.000 1~0.078 7, 陕西商洛 SL1 和 SL2 的遗传距离为 0.003 6, 四川大叶和小叶型不同居群的遗传距离介于 0.000 1~0.012 3, 山东不同居群介于 0.006 9~0.022 6, 而陕西商洛与四川居群介于 0.037 3~0.078 7, 与山东居群介于 0.011 6~0.031 7, 四川居群与山东居群介于 0.020 6~0.059 1 (表 3)。可见, 同一产地居群间的遗传距离明显小于不同产地居群间的遗传距离。聚类结果亦显示四川大叶型 (SQDY、JQDY、HXDY) 和小叶型 (JQXY、HXXY) 单独聚为一类。陕西商洛的 SL1 和 SL2 聚为一类, 山东的 LYZH、LWZHY 和 LWZH 以及 LWBH 聚为一类, 陕西商洛与山东居群聚为一大类 (图 1)。

4 不同产地基因型分析

由于来自产区陕西 (SX)、四川 (SC) 和山东 (SD) 的不同居群能按产区分别聚类, 且同一产区内居群的基因型频率没有显著性差异, 因此, 按不同产区进行基因型和基因频率的计算和统计分析 (表 4)。经过卡方检验, SNP1-19、SNP4-77、SNP5-117、

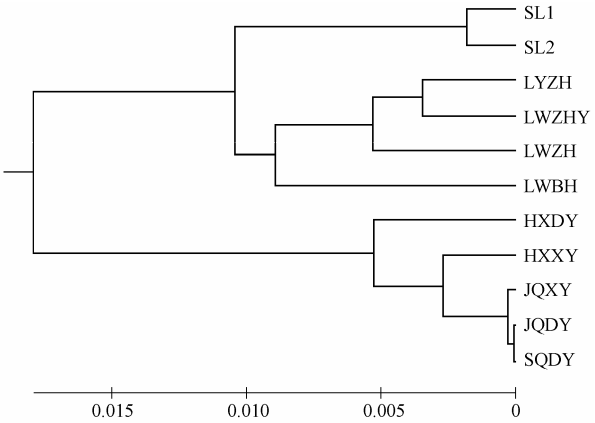


Figure 1 Dendrogram based on Nei's (1972) genetic distance method = UPGMA. Modified from NEIGHBOR procedure of PHYLIP Version 3.5

SNP6-137、SNP7-146 的基因型在 3 组之间均呈显著差异 ($P < 0.05$), SNP2-52 基因型在 SC 与 SD 间呈显著差异, SNP3-58 在 SD 与 SC 和 SX 间呈显著差异, SNP10-168、SNP11-187、SNP12-206、SNP13-222 在 SC 与 SX 和 SD 间呈显著差异, SNP9-164 在 SC 和 SX 间存在显著差异, SNP8-155 在 SX 与 SC 和 SD 间存在显著差异。通过上述分析可见, 13 个 SNP 所形成的基因型在不同产地间均表现出不同程度的差异, 由此可以确定各个产地的优势基因型。如 SNP4-77 中, SX 全部为 G/G 纯合型, 而 SC 以 A/G 杂合型为主 (85.9%), SD 则 G/G 纯合型为 53.6%, A/G 杂合型占 36.9%。在分析的 13 个位点中, SX 的 6 个位点为纯合型 (SNP3-58G/G、SNP4-77G/G、SNP6-137C/C、SNP7-146C/C、SNP10-168C/C 和 SNP11-187G/G), SC 的 2 个位点为纯合型 (SNP3-58G/G、SNP8-155G/G), 而 SD 所有位点均为杂合型。

Table 3 Nei's Unbiased measures of genetic identity and genetic distance. Estimates of mean genetic distance between 11 populations of *S. miltiorrhiza*

Populations	1	2	3	4	5	6	7	8	9	10	11
1. SL1	—										
2. SL2	0.003 6	—									
3. HXDY	0.070 9	0.078 7	—								
4. JQDY	0.043 7	0.045 9	0.011 3	—							
5. SQDY	0.043 6	0.046 0	0.010 9	0.000 1	—						
6. JQXY	0.037 3	0.039 5	0.012 3	0.000 6	0.000 5	—					
7. HXXY	0.039 4	0.042 5	0.007 6	0.006 1	0.005 8	0.004 2	—				
8. LYZH	0.014 5	0.013 5	0.042 8	0.024 6	0.024 5	0.020 6	0.020 6	—			
9. LWZH	0.027 9	0.028 8	0.033 7	0.023 1	0.022 9	0.019 3	0.015 9	0.008	—		
10. LWZHY	0.011 6	0.014 3	0.045 1	0.031 7	0.031 4	0.026 4	0.021 8	0.006 9	0.012 9	—	
11. LWBH	0.031 7	0.024 3	0.059 1	0.031 4	0.031 5	0.028 7	0.032 1	0.012 7	0.018 2	0.022 6	—

Table 4 Genotype and allele frequencies of SNPs among groups

Site/Group	Genotype			Allele	
SNP1-19	C/C (freq)	C/T (freq)	T/T (freq)	C (freq)	T (freq)
SX	2 (0.061)	22 (0.667)	9 (0.273)	26 (0.394)	40 (0.606)
SC		54 (0.844)	10 (0.156)	54 (0.422)	74 (0.578)
SD	8 (0.095)	33 (0.393)	43 (0.512)	49 (0.292)	119 (0.708)
SNP2-52	C/C (freq)	C/T (freq)	T/T (freq)	C (freq)	T (freq)
SL	31 (0.939)	2 (0.061)		64 (0.970)	2 (0.030)
SC	63 (0.984)	1 (0.016)		127 (0.992)	1 (0.008)
SD	63 (0.750)	17 (0.202)	4 (0.048)	143 (0.851)	25 (0.149)
SNP3-58	C/C (freq)	C/G (freq)	G/G (freq)	C (freq)	G (freq)
SX			33 (1.000)		66 (1.000)
SC			64 (1.000)		128 (1.000)
SD	4 (0.048)	21 (0.250)	59 (0.702)	29 (0.173)	139 (0.827)
SNP4-77	A/A (freq)	A/G (freq)	G/G (freq)	A (freq)	G (freq)
SX			33 (1.000)		66 (1.000)
SC	8 (0.125)	55 (0.859)	1 (0.016)	71 (0.555)	57 (0.445)
SD	8 (0.095)	31 (0.369)	45 (0.536)	47 (0.280)	121 (0.720)
SNP5-117	C/C (freq)	C/G (freq)	G/G (freq)	C (freq)	G (freq)
SX		1 (0.030)	32 (0.970)	1 (0.015)	65 (0.985)
SC	1 (0.016)	33 (0.516)	30 (0.469)	35 (0.273)	93 (0.727)
SD	2 (0.024)	20 (0.238)	62 (0.738)	24 (0.143)	144 (0.857)
SNP6-137	C/C (freq)	C/T (freq)	T/T (freq)	C (freq)	T (freq)
SX	33 (1.000)			66 (1.000)	
SC	29 (0.453)	34 (0.531)	1 (0.016)	92 (0.719)	36 (0.281)
SD	65 (0.774)	18 (0.214)	1 (0.012)	148 (0.881)	20 (0.119)
SNP7-146	A/A (freq)	A/C (freq)	C/C (freq)	A (freq)	C (freq)
SX			33 (1.000)		66 (1.000)
SC	1 (0.016)	33 (0.516)	30 (0.469)	35 (0.273)	93 (0.727)
SD	1 (0.012)	18 (0.214)	65 (0.774)	20 (0.119)	148 (0.881)
SNP8-155	A/A (freq)	A/G (freq)	G/G (freq)	A (freq)	G (freq)
SX		14 (0.424)	19 (0.576)	14 (0.212)	52 (0.788)
SC			64 (1.000)		128 (1.000)
SD		1 (0.012)	83 (0.988)	1 (0.006)	167 (0.994)
SNP9-164	A/A (freq)	A/G (freq)	G/G (freq)	A (freq)	G (freq)
SX	5 (0.152)	11 (0.333)	17 (0.515)	21 (0.318)	45 (0.682)
SC		24 (0.375)	40 (0.625)	24 (0.188)	104 (0.812)
SD	7 (0.083)	25 (0.298)	52 (0.619)	39 (0.232)	129 (0.768)
SNP10-168	C/C (freq)	C/T (freq)	T/T (freq)	C (freq)	T (freq)
SX	33 (1.000)			66 (1.000)	
SC	31 (0.484)	33 (0.516)		95 (0.742)	33 (0.258)
SD	71 (0.845)	12 (0.143)	1 (0.012)	154 (0.917)	14 (0.083)
SNP11-187	A/A (freq)	A/G (freq)	G/G (freq)	A (freq)	G (freq)
SX			33 (1.000)		66 (1.000)
SC	1 (0.016)	31 (0.484)	32 (0.500)	33 (0.258)	95 (0.742)
SD	1 (0.012)	15 (0.179)	68 (0.810)	17 (0.101)	151 (0.899)
SNP12-206	A/A (freq)	A/G (freq)	G/G (freq)	A (freq)	G (freq)
SX	3 (0.091)	12 (0.364)	18 (0.545)	18 (0.273)	48 (0.727)
SC	7 (0.109)	54 (0.844)	3 (0.047)	68 (0.531)	60 (0.469)
SD	8 (0.095)	32 (0.381)	44 (0.524)	48 (0.286)	120 (0.714)
SNP13-222	A/A (freq)	A/G (freq)	G/G (freq)	A (freq)	G (freq)
SX	6 (0.182)	20 (0.606)	7 (0.212)	32 (0.485)	34 (0.515)
SC	2 (0.031)	55 (0.859)	7 (0.109)	59 (0.461)	69 (0.539)
SD	30 (0.357)	40 (0.476)	14 (0.167)	100 (0.595)	68 (0.405)

讨论

1 ERF 基因家族的功能和分类

植物体内存在大量的转录因子^[10]。根据 DNA 结合域的不同, 其被划分成很多家族, 如参与植物防卫与胁迫反应的转录因子家族主要有 AP2/ERF、bZIP、MYB/MYC 和 WRKY 家族等^[2]。拟南芥中发现有 147 个 AP2/ERF 超家族基因, 其中有 122 个属于 ERF 家族, 水稻中有 139 个 ERF 基因。根据 *SmERF* 的基因结构, 推测其属于拟南芥和水稻 ERF 家族基因中的 VII 组基因, 该结构的确定为进一步研究 *SmERF* 的功能奠定了基础。

2 不同居群遗传关系

丹参不同居群的遗传关系按不同产地明显的分成不同类别, 其中四川中江丹参单独聚为一类, 与其他居群的遗传关系最远, 该结果与郭宝林等^[11]的 RAPD 结果一致。这可能与四川中江丹参长期的栽培以及品种选育有关, 四川大叶型丹参为产区传统大面积栽培品种, 分布海拔 400~800 m, 但退化严重, 产量较低; 小叶型丹参为新发现的优良栽培品种资源, 分布海拔 600~800 m, 属抗病型品种, 产量高。*SmERF* 基因并不能按大叶和小叶型单独分类, 表明其遗传分化并不明显。

白花丹参为紫花丹参 (原种) 的一个变型, 其花冠为白色或淡黄色, 为山东特产, 其有效成分中的脂溶性成分高于紫花丹参^[12]。聚类结果显示, 白花丹参与山东其他居群丹参聚为一类, 但与其他居群的遗传关系较远, 表明其有一定的遗传分化。山东莱芜野生居群 (LWZHY) 与栽培居群 LYZH 首先聚类后再与栽培居群 LWZH 聚类, 表明野生居群与栽培居群的遗传差异不大, 这可能与栽培基地多采用当地种源, 缺少品种选育有关。

3 产地独特基因型

在分析的 13 个位点中, 陕西商洛丹参具有 6 个纯合位点, 四川具有 2 个纯合位点, 而山东的均为杂合位点, 多态位点百分率结果表明山东丹参的遗传多样性最为丰富, 不同产地丹参具有差异显著的基因型, 即产地优势基因型。这与丹参酮生物合成关键酶乙酰 CoA 酰基转移酶 (acetyl-CoA C-acetyltransferase, AACT) 基因 SNPs 的研究结果一致^[13]。AP2/ERF 基因家族在植物应对生物和非生物胁迫中起着重要的作用^[6], 而这些独特的基因型是否是不同产地丹参适应环境而逐步进化的结果以及这

些基因型是否与药材的品质形成连锁将成为下一步研究的重点。

近年来, 利用 DNA 序列研究药用植物的分类、鉴定和进化逐渐成为研究的热点, 但利用功能基因序列进行研究的还未见报道。本研究表明 *SmERF* 在不同居群间具有丰富的多态性, 并且与产地呈现紧密的联系, 这些发现为利用 *SmERF* 基因研究不同产地丹参的鉴定、起源演化及与品质形成的关系提供了基础。

References

- [1] Riechmann JL, Ratcliffe OJ. A genomic perspective on plant transcription factors [J]. *Curr Opin Plant Biol*, 2000, 3: 423–434.
- [2] Ohme-Takagi M, Shinshi H. Ethylene-inducible DNA binding proteins that interact with ethylene-responsive element [J]. *Plant Cell*, 1995, 7: 173–182.
- [3] 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: 411–432.
- [4] Sakuma Y, Liu Q, Dubouzet JG, 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: 998–1009.
- [5] Xu B, Huang LQ, Cui GH, et al. Functional genomics of *Salvia miltiorrhiza* IV. Analysis of ethylene responsive element binding protein gene [J]. *China J Chin Mat Med (中国中药杂志)*, 2009, 34: 2564–2566.
- [6] Kumar S, Tamura K, Nei M. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment [J]. *Brief Bioinform*, 2004, 5: 150–163.
- [7] Shi YY, He L. SHEsis, a powerful software platform for analyses of linkage disequilibrium, haplotype construction, and genetic association at polymorphism loci [J]. *Cell Res*, 2005, 15: 97–98.
- [8] Buüttner M, Singh KB. *Arabidopsis thaliana* ethylene-responsive element binding protein (AtERF), an ethylene-inducible, GCC box DNA binding protein interacts with an ocs element binding protein [J]. *Proc Natl Acad Sci USA*, 1997, 94: 5961–5966.
- [9] Zhu Y, Cai XL, Wang ZY, et al. An interaction between a MYC protein and an ERF protein is involved in transcriptional regulation of the rice Wx gene [J]. *J Biol Chem*, 2003, 278: 47803–47811.
- [10] Eulgem T, Rushton PJ, Robatzek S, et al. The superfamily of plant transcription factors [J]. *Trends Plant Sci*, 2000, 5: 199–206.
- [11] Guo BL, Lin S, Feng YX, et al. Primary research on genetic relationship among main populations of *Salvia miltiorrhiza* and genuineness of herb [J]. *China J Chin Mat Med (中国中药杂志)*, 2002, 33: 1113–1116.
- [12] Guo BL, Feng YX, Zhao YJ. Review of germplasm resources studies on *Salvia miltiorrhiza* [J]. *China J Chin Mat Med (中国中药杂志)*, 2002, 27: 492–495.
- [13] Cui GH, Wang XY, Feng H, et al. Molecular cloning and SNP analysis of a acetyl-CoA C-acetyltransferase gene (*SmAACT*) from *Salvia miltiorrhiza* [J]. *Acta Pharm Sin (药学报)*, 2010, 45: 785–790.