

# 大豆 DNA 去甲基化酶 ROS1 的生物信息学分析

梁喜龙<sup>1</sup>, 李国兰<sup>1</sup>, 崔洪秋<sup>1</sup>, 鞠世杰<sup>1</sup>, 洪艳华<sup>1</sup>, 方淑梅<sup>2</sup>, 郑殿峰<sup>1</sup>

(1. 黑龙江八一农垦大学 农学院, 黑龙江 大庆 163319; 2. 黑龙江八一农垦大学 生命学院, 黑龙江 大庆 163319)

**摘要:** 利用 NCBI、Phytozome、ExPASy 等网站及其数据库, 初步确定了大豆 DNA 去甲基化酶 ROS1 的蛋白及基因序列、基因拷贝数、理化特性等, 并进一步预测分析了二级结构及结构域, 同时结合 Clustal X2.0 和 MEGA4.0 等软件进行多重序列比对和分子系统进化关系研究。结果显示, 大豆 ROS1 包括 6 个拷贝, ROS1 各拷贝的分子量相近, 等电点酸性, 具有相似的潜在磷酸化位点, 属于不稳定的疏水性蛋白, 都位于细胞核内, 且不含信号序列。α 螺旋是主要的二级结构, 均含有 DNA\_glycosylase, HhH-GPD\_domain 和 HTH\_base\_excis\_C 三个结构域。

**关键词:** 大豆; DNA 去甲基化酶; ROS1; 生物信息学分析

**中图分类号:** S565.1

**文献标识码:** A

**DOI:** 10.11861/j.issn.1000-9841.2014.03.0321

## Bioinformatics Analysis of DNA Demethylase ROS1 in Soybean

LIANG Xi-long<sup>1</sup>, LI Guo-lan<sup>1</sup>, CUI Hong-qiu<sup>1</sup>, JU Shi-jie<sup>1</sup>, HONG Yan-hua<sup>1</sup>, FANG Shu-mei<sup>2</sup>, ZHENG Dian-feng<sup>1</sup>

(1. Agronomy College, Heilongjiang Bayi Agricultural University, Daqing 163319, China; 2. College of Life Science and Technology, Heilongjiang Bayi Agricultural University, Daqing 163319, China)

**Abstract:** Websites and data bases were employed such as NCBI, Phytozome and ExPASy in this paper to identify the sequences of proteins and genes, gene copies and some physicochemical properties of ROS1 in soybean. Its secondary structure and domain were predicted and analyzed. Meanwhile, phylogenetic tree was constructed based on the sequences of amino acids to analyze the evolutionary relations among different species. The results showed that ROS1 included six copies with similar molecular weight in soybean. There were acid isoelectric points and similar potential phosphorylation sites in the ROS1 proteins but signal sequences were lacked. All copies were all unstable and hydrophobic and were expressed in nucleus. α helix was the main secondary structure. DNA\_glycosylase, HhH-GPD\_domain and HTH\_base\_excis\_C domains were included in the ROS1 proteins of soybean.

**Key words:** Soybean; DNA demethylase; ROS1; Bioinformatics analysis

大豆是人类优质蛋白质和植物油的重要来源, 对人类生存和健康极为重要。但是随着全球环境和气候的变化, 各种逆境因素正严重威胁着大豆的产量和品质, 因此深入研究大豆抗逆机制, 进而培育高产优质的抗逆品种成为研究热点。

已有大量研究表明, 逆境胁迫能够诱导植物 DNA 甲基化发生变化, 进而影响基因表达调控和转座子活性, 改变遗传信息量并赋予特定的遗传属性, 从而为植物适应不良环境提供可塑性<sup>[1-4]</sup>。植物 DNA 甲基化以 5-甲基胞嘧啶 (5-methylcytosine) 为主, 而甲基化模式和水平是动态变化的过程, 由 DNA 甲基化酶和去甲基化酶的协同作用来维系, 在植物不同生长发育阶段和不同生长环境条件下, 会做出适应性改变<sup>[5-6]</sup>。

ROS1 (repressor of silencing 1) 是一种 DNA 糖基化酶/裂解酶, 它能够去除 DNA 上的 5-甲基胞嘧啶核苷酸而发动 DNA 的去甲基化过程。目前, 该酶的特性和功能仅在拟南芥中进行了较为详细的研究,

其在大豆中的研究尚未见报道。因此, 本文通过生物信息学方法对大豆 ROS1 基因及蛋白进行了分析和预测, 为进一步阐明大豆中 ROS1 的结构和功能提供有价值的参考。

## 1 研究方法

首先, 利用 NCBI 网站查找拟南芥的 ROS1 蛋白序列, 在 Phytozome 网站中通过同源比对获取候选蛋白 (S 值排在前 6, E 值为 0), 然后将候选蛋白序列输入拟南芥基因组数据库中进行同源比对, 同时进入 NCBI 数据库对获得的蛋白序列进行 BLAST 比对, 从而进一步确认所获得序列的正确性。利用 ExPASy 网站的 ProtParam 在线工具分析蛋白质分子量、等电点、亲水性和稳定性, 信号序列分析采用 SignalP 4.0 程序和 TargetP 1.1 Server。磷酸化位点预测利用 DISPHOS 1.3 软件, 亚细胞定位利用 Psort Prediction 和 ProtComp v. 9.0 在线软件。二级结构

收稿日期: 2013-11-05

基金项目: 黑龙江省教育厅科学技术研究项目 (12521381)。

第一作者简介: 梁喜龙 (1976-), 男, 在读博士, 副教授, 主要从事植物化控及生物逆境方面研究。E-mail: xilongliang@126.com。

通讯作者: 郑殿峰 (1969-), 男, 教授, 博士生导师, 主要从事大豆化控及生物逆境方面研究。E-mail: zdfnj@263.net。

预测分析利用 SOPMA 软件进行。结构域分析利用 InterPro 在线软件获得相关信息,然后利用 Domain-Draw 软件按比例绘制。系统进化树构建利用 Clustal X2.0 软件对蛋白质全序列进行多重匹配分析,生成两种格式文件(.aln 和 .dnd),再通过 MEGA 4.0 软件中的 Neighbor-Joining 绘制<sup>[7-8]</sup>。

## 2 结果与分析

### 2.1 大豆去甲基化酶基因 *ROS1* 的获取与定位

在 NCBI 数据库查找获得拟南芥 *ROS1* 基因 AT2G36490, 根据同源性比对获得大豆基因组中

表 1 大豆去甲基化酶基因 *ROS1* 的获取与定位

Table 1 Obtaining and location of DNA demethylase genes *ROS1* in soybean

转录子名称 Transcript name	基因长度 Length of gene/kb	染色体定位 Chromosome location	蛋白质长度 Length of protein/aa	同源性 Homology/%	S 值 S value
Glyma13g21460.2	11.970	Gm13;25010135-25022104	1993	72.3	771.2
Glyma03g34860.3	11.840	Gm03;42158730-42170567	1741	70.6	751.9
Glyma10g07601.1	17.280	Gm10;6336194-6353470	2014	71.0	756.9
Glyma03g34860.4	9.625	Gm03;42158730-42168354	1675	70.1	701.0
Glyma20g32960.2	12.040	Gm20;41566064-41578105	1848	65.8	689.9
Glyma10g34595.1	9.444	Gm10;42769468-42778911	1798	63.0	677.6

*ROS1* 基因的存在情况如表 1 所示。*ROS1* 基因存在 6 个拷贝,分别定位于第 3、13、10 和 20 号染色体上,其中 3 号和 10 号染色体上各存在 2 个拷贝。

### 2.2 大豆去甲基化酶 *ROS1* 的理化特性、信号序列及亚细胞定位分析

由表 2 可知,*ROS1* 各拷贝的分子量相近,等电点都在酸性范围内,为不稳定的疏水性蛋白,都位于细胞核内,且不含信号序列。丝氨酸(Ser)发生磷

酸化的数目最多,苏氨酸(Thr)次之。

### 2.3 大豆去甲基化酶 *ROS1* 的二级结构及结构域分析

$\alpha$  螺旋和  $\beta$  折叠是蛋白质常见的二级结构,可进一步形成超二级结构及结构域而发挥生物学功能。*ROS1* 各拷贝的二级结构如表 3 所示, $\alpha$  螺旋和  $\beta$  折叠之和所占比例均在 40% 左右,而其中又以  $\alpha$  螺旋为主要结构。

表 2 大豆去甲基化酶蛋白 *ROS1* 的基本特性及亚细胞定位

Table 2 The physicochemical properties and subcellular localization of DNA demethylase *ROS1* in soybean

转录子名称 Transcript name	分子量 Molecular weight/Da	等电点 Isoelectric point	亲水性 Hydrophilicity	稳定性 Stability	信号序列 Signal sequence	磷酸化位点/个 Phosphorylation sites/number	亚细胞定位 Subcellular localization
Glyma13g21460.2	221824.4	6.28	疏水 Hydrophobic	不稳定 Unstable	NO	Ser/22, Thr/3, Tyr/0	细胞核 Nucleus
Glyma03g34860.3	195503.1	6.48	疏水 Hydrophobic	不稳定 Unstable	NO	Ser/10, Thr/2, Tyr/0	细胞核 Nucleus
Glyma10g07601.1	224111.3	6.66	疏水 Hydrophobic	不稳定 Unstable	NO	Ser/15, Thr/0, Tyr/0	细胞核 Nucleus
Glyma03g34860.4	188028.7	6.30	疏水 Hydrophobic	不稳定 Unstable	NO	Ser/10, Thr/2, Tyr/0	细胞核 Nucleus
Glyma20g32960.2	209255.9	6.45	疏水 Hydrophobic	不稳定 Unstable	NO	Ser/16, Thr/2, Tyr/1	细胞核 Nucleus
Glyma10g34595.1	203069.1	6.48	疏水 Hydrophobic	不稳定 Unstable	NO	Ser/6, Thr/2, Tyr/2	细胞核 Nucleus

表 3 大豆去甲基化酶 *ROS1* 的二级结构分布

Table 3 The percentage of predicted secondary structure in DNA demethylase *ROS1* from soybean

转录子名称 Transcript name	$\alpha$ 螺旋 $\alpha$ helix/%	$\beta$ 折叠 $\beta$ -pleated sheet/%	$\beta$ 转角 Beta turn/%	无规卷曲 Random coil/%
Glyma13g21460.2	29.25	9.78	4.26	56.70
Glyma03g34860.3	29.75	10.40	4.14	55.72
Glyma10g07601.1	29.69	12.31	5.11	52.88
Glyma03g34860.4	29.13	10.69	4.48	55.70
Glyma20g32960.2	30.09	11.53	4.22	54.17
Glyma10g34595.1	30.53	9.79	3.73	55.95

结构域是构成蛋白质三级结构的基本单元,具有一定的生物学功能。为了进一步确定 ROS1 的功能与特征,本文分析和预测了 ROS1 各拷贝结构域形成区在肽链中的位置,并进行了多序列比对分析及功能结构域绘制。如图 1 所示,ROS1 的 6 个拷贝都具有相同的 3 个功能结构域(DNA\_glycosylase, HhH-GPD\_domain, HTH\_base\_excis\_C)。而且各拷贝结构域区的氨基酸序列非常相似(图 2),在肽链中的位置也基本相同(表 4),这些信息提示在大豆中这些不同的 ROS1 拷贝可能具有相似的三级结构及生物学功能。

2.4 大豆去甲基化酶 ROS1 的分子进化分析

为了分析 ROS1 在不同物种中的进化和变异情况,于大豆数据库中搜索到与 ROS1 同源的不同物种的氨基酸序列(表 5)。

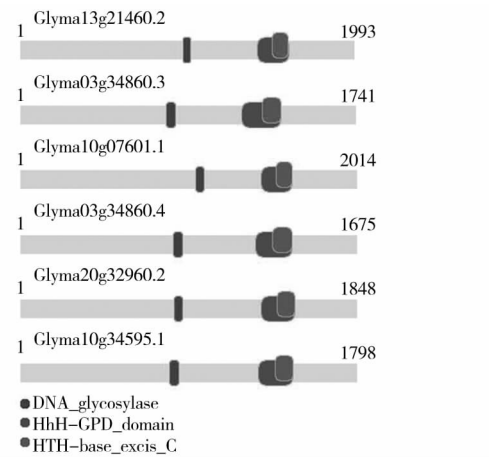
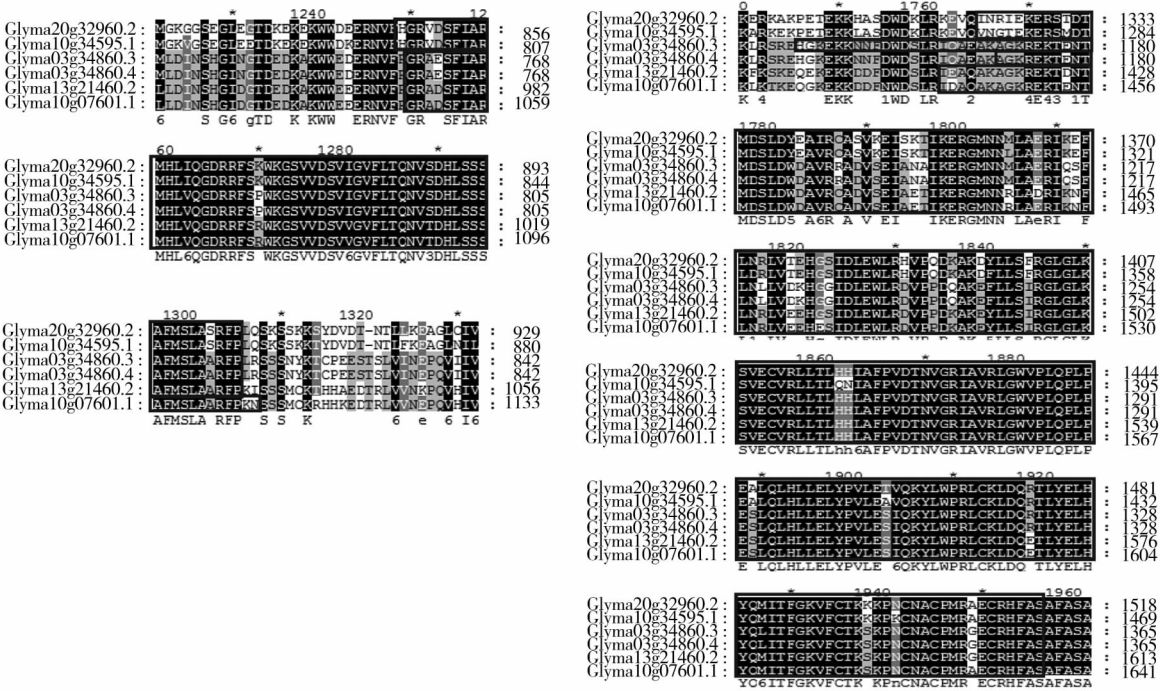


图 1 大豆去甲基化酶 ROS1 各结构域分布  
Fig. 1 The domain diagram of demethylase ROS1 in soybean

表 4 大豆去甲基化酶 ROS1 各结构域在肽链中的位置

转录子名称 Transcript name	DNA_glycosylase	HhH-GPD domain	HTH_base_excis_C
Glyma13g21460.2	974 ~ 1029	1413 ~ 1608	1505 ~ 1608
Glyma03g34860.3	760 ~ 815	1149 ~ 1360	1257 ~ 1360
Glyma10g07601.1	1051 ~ 1108	1441 ~ 1636	1533 ~ 1636
Glyma03g34860.4	760 ~ 815	1171 ~ 1360	1257 ~ 1360
Glyma20g32960.2	847 ~ 903	1322 ~ 1513	1409 ~ 1513
Glyma10g34595.1	798 ~ 854	1274 ~ 1464	1360 ~ 1464



方框内为形成结构域的序列。  
The amino acids in box are the sequences of forming domains.

图 2 大豆去甲基化酶 ROS1 结构域区氨基酸序列比对分析

Fig. 2 The comparison of amino acid sequence in DNA demethylase ROS1 domain region from soybean

表 5 构建进化树相关物种基本信息

Table 5 The basic information of related species in the phylogenetic tree

物种名称 Species	简写 Abbr.	分类 Classification	物种名称 Species	简写 Abbr.	分类 Classification
高粱 <i>Sorghum bicolor</i>	Sbi	单子叶植物 Monocotyledons	黄瓜 <i>Cucumis sativus</i>	Csa	双子叶植物 Dicotyledons
玉米 <i>Zea mays</i>	Zma	单子叶植物 Monocotyledons	番茄 <i>Solanum lycopersicum</i>	Sly	双子叶植物 Dicotyledons
谷子 <i>Setaria italica</i>	Sit	单子叶植物 Monocotyledons	盐芥 <i>Thellungiella halophila</i>	Tha	双子叶植物 Dicotyledons
水稻 <i>Oryza sativa</i>	Osa	单子叶植物 Monocotyledons	荠菜 <i>Capsella rubella</i>	Cru	双子叶植物 Dicotyledons
拟南芥 <i>Arabidopsis thaliana</i>	Ath	双子叶植物 Dicotyledons	芜菁 <i>Brassica rapa</i>	Bra	双子叶植物 Dicotyledons
大豆 <i>Glycine max</i>	Gma	双子叶植物 Dicotyledons	桃 <i>Prunus persica</i>	Ppe	双子叶植物 Dicotyledons
葡萄 <i>Vitis vinifera</i>	Vvi	双子叶植物 Dicotyledons	毛果杨 <i>Populus trichocarpa</i>	Ptr	双子叶植物 Dicotyledons
菜豆 <i>Phaseolus vulgaris</i>	Pvu	双子叶植物 Dicotyledons	金钱橘 <i>Citrus clementina</i>	Ccl	双子叶植物 Dicotyledons
苹果 <i>Malus domestica</i>	Mdo	双子叶植物 Dicotyledons	雷蒙德氏棉 <i>Gossypium raimondii</i>	Gra	双子叶植物 Dicotyledons

利用 MEGA4.0 软件进行 ROS1 系统进化树的构建,由图 3 可见,在大豆 ROS1 的 6 个拷贝中,Glyma13g21460.2、Glyma03g34860.3、Glyma10g07601.1 和 Glyma03g34860.4 位于双子叶植物组内,而且与菜豆的距离最近,显示了二者进化上的亲缘关系。而 Glyma20g32960.2 和 Glyma10g34595.1 却位于一个独立的分支上,位于双子叶植物和单子叶植物之间。

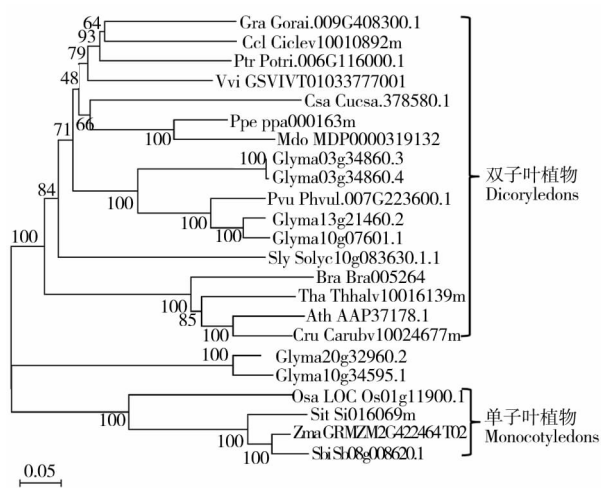


图 3 大豆去甲基化酶 ROS1 系统进化树

Fig. 3 Phylogenetic tree of DNA demethylases in soybean

### 3 结论与讨论

生物体内 DNA 的甲基化并不是一成不变的,总是处于动态平衡状态,其建立与维持机制比较复杂,与 DNA 甲基化酶、RNA 介导及 DNA 去甲基化酶有密切关系。目前研究表明:DNA 糖基化酶/裂

解酶(DNA glycosylase/lyase)ROS1 在植物 DNA 去甲基化过程中发挥重要作用,它能够识别甲基化的胞嘧啶,并将甲基化的胞嘧啶碱基从 DNA 骨架上移除,随后它的 AP 裂解酶活性使 DNA 骨架在原甲基化位点处裂开并移除脱氧核糖,最后 DNA 被未甲基化的胞嘧啶修复,完成脱甲基化过程<sup>[1,9-10]</sup>。对拟南芥研究表明,ROS1 能够使目的基因启动子去甲基化,因此 ROS1 基因突变必将致其超甲基化而导致基因沉默<sup>[11]</sup>。体外实验也显示重组的 ROS1 蛋白能够断裂甲基化的质粒 DNA,而对非甲基化的 DNA 却没有作用<sup>[11]</sup>。

本研究利用同源比对的方法,获得 6 个 ROS1 拷贝,其蛋白质分子量均大于拟南芥 ROS1 蛋白(1 393 AA),各同源蛋白的理化特性(等电点、亲水性、稳定性)、磷酸化位点、亚细胞定位、二级结构及结构域等均较相似,这些特点说明了其蛋白产物在功能上具有相似性。系统进化树构建使用的是各物种的氨基酸序列,是在分子水平上探究它们的亲缘关系。Glyma20g32960.2 和 Glyma10g34595.1 并未位于双子叶植物分支内,而是形成一个独立的分支,位于双子叶植物和单子叶植物之间,表明其蛋白质一级结构与双子叶植物存在较大差异,然而通过多序列比对分析发现导致这种差异的氨基酸并不位于结构域区域,各 ROS1 拷贝的二级结构及结构域均很相似,因此不会导致蛋白质功能改变。同时,对结构域的分析发现,大豆 ROS1 各拷贝均含有 DNA\_glycosylase 和 HhH\_GDP\_domain 结构域,其结构域在拟南芥的 ROS1 中也存在,研究表明该结构域在 ROS1 中是保守的<sup>[9,11]</sup>。

(下转第 333 页)

- naling in *Arabidopsis* seed germination[J]. Plant Journal, 2008, 55 (1): 77-88.
- [14] Kato H, Motomura T, Komeda Y, et al. Overexpression of the NAC transcription factor family gene ANAC036 results in a dwarf phenotype in *Arabidopsis thaliana*[J]. Journal of Plant Physiology, 2010, 167(7): 571-577.
- [15] Shahnejat-Bushehri S, Mueller-Roeber B, Balazadeh S. *Arabidopsis* NAC transcription factor JUNGBRUNNEN1 affects thermomemory-associated genes and enhances heat stress tolerance in primed and unprimed conditions [J]. Plant Signaling & Behavior, 2012, 7 (12): 1518-1521.
- [16] Distelfeld A, Pearce S P, Avni R, et al. Divergent functions of orthologous NAC transcription factors in wheat and rice[J]. Plant Molecular Biology, 2012, 78(4-5): 515-524.
- [17] Nuruzzaman M, Manimekalai R, Sharoni A M, et al. Genome-wide analysis of NAC transcription factor family in rice[J]. Gene, 2010, 465(1-2): 30-44.
- [18] 刘洋, 张慧, 辛大伟, 等. 大豆 TCP 转录因子家族结构域分析及功能预测[J]. 大豆科学, 2012, 31(5): 707-717. (Liu Y, Zhang H, Xin D W, et al. Domain analysis and function prediction of TCP transcription factors family in soybean[J]. Soybean Science, 2012, 31(5): 707-717.)
- [19] Ooka H, Satoh K, Doi K, et al. Comprehensive analysis of NAC family genes in *Oryza sativa* and *Arabidopsis thaliana*[J]. DNA Research, 2003, 10(6): 239-247.
- [20] 李乐, 许红亮, 杨兴露, 等. 大豆 LEA 基因家族全基因组鉴定、分类和表达[J]. 中国农业科学, 2011, 44(19): 3945-3954. (Li L, Xu H L, Yang X L, et al. Genome-wide identification, classification and expression analysis of LEA gene family in soybean[J]. Scientia Agricultura Sinica, 2011, 44(19): 3945-3954.)
- [21] 陈莹, 孙霞, 胡尚连, 等. 拟南芥次生长相关 NAC 转录因子保守结构域分析[J]. 西北农林科技大学学报, 2009, 37(5): 185-194. (Chen Y, Sun X, Hu S L, et al. Analysis of NAC transcription factor and its domain with secondary growth in *Arabidopsis thaliana*[J]. Journal of Northwest A&F University, 2009, 37(5): 185-194.)
- [22] 江董丽, 才华, 端木慧子, 等. 大豆 GST 基因家族全基因组筛选、分类和表达[J]. 分子植物育种, 2013, 11(5): 465-475. (Jiang D L, Cai H, Duanmu H Z, et al. Genome-wide filter, classification and expression analysis of GST gene family in soybean [J]. Molecular Plant Breeding, 2013, 11(5): 465-475.)
- [23] Zhang C H, Liu H, Yu M L, et al. Bioinformatics analysis for the NAC gene family in strawberry[J]. Genomics and Applied Biology, 2011, 30(41): 1261-1271.
- [24] 蒋瑶, 陈其兵. 植物 CBF1 转录因子的生物信息学分析[J]. 林业科学, 2010, 46(6): 43-50. (Jiang Y, Chen Q B. Bioinformatic analysis of CBF1 transcription factors from the plants[J]. Scientia Silvae Sinicae, 2010, 46(6): 43-50.)
- [25] Olsen A N, Ernst H A, Leggio L L, et al. NAC transcription factors: structurally distinct, functionally diverse[J]. Trends in Plant Science, 2005, 10(2): 79-87.
- [26] Strauch M A, Ballar P, Rowshan A J, et al. The DNA-binding specificity of the *Bacillus anthracis* AbrB protein [J]. Microbiology, 2005, 151(6): 1751-1759.
- [27] Peng H, Yu X W, Cheng H Y, et al. A survey of functional studies of the plant-specific NAC transcription factor family[J]. Chinese Bulletin of Botany, 2010, 45(2): 236-248.
- [28] Wang N, Zheng Y, Xin H, et al. Comprehensive analysis of NAC domain transcription factor gene family in *Vitis vinifera*[J]. Plant Cell Reports, 2013, 32(1): 61-75.
- [29] Nuruzzaman M, Sharoni A M, Satoh K, et al. Comprehensive gene expression analysis of the NAC gene family under normal growth conditions, hormone treatment, and drought stress conditions in rice using near-isogenic lines (NILs) generated from crossing Aday Selection (drought tolerant) and IR64 [J]. Molecular Genetics Genomics, 2012, 287(5): 389-410.

(上接第 324 页)

## 参考文献

- [1] Zhu J K. Active DNA demethylation mediated by DNA glycosylases [J]. Annual Review of Genetics, 2009, 43: 143-166.
- [2] Finnegan E J, Genger R K, Peacock W J, et al. DNA methylation in plants [J]. Annual Review of Plant Biology, 1998, 49(1): 223-247.
- [3] Wada Y, Miyamoto K, Kusano T, et al. Association between up-regulation of stress-responsive genes and hypomethylation of genomic DNA in tobacco plants [J]. Molecular Genetics and Genomics, 2004, 271: 658-666.
- [4] Pan Y J, Fu B Y, Wang D, et al. Spatial and temporal profiling of DNA methylation induced by drought stress in rice [J]. Scientia Agricultura Sinica, 2009, 42: 3009-3018.
- [5] Finnegan E J, Peacock W J, Dennis E S. DNA methylation, a key regulator of plant development and other processes [J]. Current Opinion in Genetics & Development, 2000, 10(2): 217-223.
- [6] Jost J P, Saluz H P. DNA methylation: Molecular biology and biological significance [M]. Basel: Birkhauser Verlag, 1993: 218-261.
- [7] 方淑梅, 梁喜龙, 纪伟波, 等. 大豆 DNA 甲基化酶生物信息学分析[J]. 中国油料作物学报, 2013, 35(2): 225-230. (Fang S M, Liang X L, Ji W B, et al. Bioinformatics analysis of DNA methylase from soybean[J]. Chinese Journal of Oil Crop Sciences, 2013, 35(2): 225-230.)
- [8] 范晓军, 宋志芳, 仙笑笑, 等. 金纹细蛾几丁质酶基因生物信息学分析[J]. 基因组学与应用生物学, 2013, 32(1): 91-96. (Fan X J, Song Z F, Xian X X, et al. Bioinformatic analysis of chitinase gene from *Lithocolletica ringoniella* [J]. Genomics and Applied Biology, 2013, 32(1): 91-96.)
- [9] Agius F, Kapoor A, Zhu J K. Role of the *Arabidopsis* DNA glycosylase/lyase ROS1 in active DNA demethylation [J]. Proceedings of the National Academy of Sciences, 2006, 103(31): 11796-11801.
- [10] Gehring M, Reik W, Henikoff S. DNA demethylation by DNA repair [J]. Trends in Genetics, 2009, 25(2): 82-90.
- [11] Gong Z, Morales-Ruiz, Ariza R R, et al. ROS1, a repressor of transcriptional gene silencing in *Arabidopsis*, encodes a DNA Glycosylase/Lyase [J]. Cell, 2002, 111(6): 803-814.