

王新亮, 彭玲, 王健, 等. 苹果 Dof 转录因子生物信息学及其表达分析[J]. 江苏农业学报, 2021, 37(2): 480-492.
doi:10.3969/j.issn.1000-4440.2021.02.026

苹果 Dof 转录因子生物信息学及其表达分析

王新亮^{1,2}, 彭玲², 王健¹, 贾晶晶¹, 唐立平¹

(1. 滨州学院《滨州学院学报》编辑部, 山东 滨州 256603; 2. 滨州学院山东省黄河三角洲生态环境重点实验室, 山东 滨州 256603)

摘要: 为了解苹果 Dof 转录因子家族的生物学信息与功能, 利用苹果基因组 GDDH13 v1.1 检索及 RNA-seq 转录本重构找到 51 个 Dof 基因, 并通过 Pfam 和 SMART 检测确认, 进一步对这些 Dof 基因进行全面分析。结果表明, 除 MD07G1265700 外, 其他 50 个 Dof 蛋白均含有一个明显的 Dof 结构域, 且有一个 CX2CX21CX2C 基序。这些 Dof 基因编码氨基酸数为 163~523, 相对分子质量为 18 210~55 800, 等电点为 5.01~10.30, 多数 Dof 成员定位于细胞核中, 少数定位于叶绿体或线粒体中。组织特异表达显示多数 Dof 基因在营养器官中的表达量高于生殖器官, 而 MD01G1084700、MD07G1153300、MD08G1040100、MD15G1034500 基因在未成熟的果肉中表达量最高。盐碱胁迫下, 除 MD05G1023800 基因没有检测到表达外, 其他 Dof 基因的表达均受盐碱胁迫影响, 只是响应强度和时间有差异, 基因表达显著上调的有 7 个, 显著下调的有 15 个。该研究结果为进一步揭示苹果 Dof 转录因子生物功能奠定了理论基础。

关键词: 苹果; Dof 基因; 转录因子; 表达分析; 生物信息学

中图分类号: S661.1 **文献标识码:** A **文章编号:** 1000-4440(2021)02-0480-13

Bioinformatics and expression analysis of the Dof transcription factors in apple

WANG Xin-liang^{1,2}, PENG Ling², WANG Jian¹, JIA Jing-jing¹, TANG Li-ping¹

(1. Editorial Department of Journal of Binzhou University, Binzhou University, Binzhou 256603, China; 2. Shandong Key Laboratory of Eco-Environmental Science for the Yellow River Delta, Binzhou University, Binzhou 256603, China)

Abstract: In order to understand the biological information and function of apple Dof transcription factor family, 51 Dof genes were found by searching apple genome GDDH13 v1.1 and RNA-seq transcripts reconstruction, and they were confirmed as members of Dof transcription factor family by Pfam and SMART detection. The Dof genes were generally analyzed furtherly. The results showed that, besides MD07G1265700, the other 50 Dof proteins all contained an obvious Dof domain and a CX2CX21CX2C motif. The number of amino acids encoded by Dof gene ranged from 163 to 523, and the relative molecular weight ranged from 18 210 to 55 800, while the isoelectric point ranged from 5.01 to 10.30. Most of the Dof proteins were located in the nucleus, and a few of them were located in the chloroplast or mitochondria. The results of tissue-specific expression showed that the expression level of most of the Dof genes expressed in vegetative organs were higher than in reproductive organs, while the expression levels of MD01G1084700, MD07G1153300, MD08G1040100 and MD15G1034500 were the highest in immature fruit flesh. Under saline-alkali stress, except for MD05G1023800 gene, the expression of other Dof

genes were affected by saline-alkali stress, but the intensity and time of response were different. There were seven Dof genes significantly up-regulated and 15 Dof genes significantly down regulated. The research results lay a theoretical foundation for further study on the biological function of Dof transcription factor in apple.

收稿日期: 2020-07-13

基金项目: 滨州学院博士学位人员及具有硕士学位的高级职称人员
科研启动费项目(2019Y36)

作者简介: 王新亮(1983-), 男, 山东宁津人, 博士, 高级工程师, 主要从事果树抗逆及氮素吸收利用研究。(E-mail) wiln1@163.com

Key words: apple; Dof gene; transcription factor; expression analysis; bioinformatics

苹果(*Malus pumila* Mill.)是蔷薇科中重要的果树之一,与其他果树相比,具有种植面积大、货架期长等特点,苹果在中国的种植面积和产量均为世界首位。土壤盐碱、干旱、倒春寒等非生物胁迫制约着苹果产业的发展^[1]。

Dof (DNA binding with one finger) 蛋白是一类植物特有的转录因子^[2],其 N 端含有一个由 50~52 个氨基酸组成的高度保守的锌指结构——Dof 结构域,此结构域中含有 1 个高度保守的 CX₂CX₂1CX₂C 基序。Dof 蛋白既能与特定的 DNA 序列结合调节基因表达,也能与某些蛋白质相互作用,参与植物生长发育及非生物胁迫响应的调控^[3-5]。Dof 蛋白识别的核心序列是 5'-AAAG-3'或 5'-CTTT-3'^[6];其 C 端氨基酸序列变异性较大,是 Dof 蛋白的转录调控结构域^[7]。在单子叶植物和双子叶植物均发现了许多 Dof 蛋白。目前,在小麦、大豆、水稻和拟南芥的基因组中,分别有 31、78、30 和 36 个 Dof 基因被发现^[8-10]。植物中存在如此多的 Dof 基因说明其参与植物多种复杂的生理功能控制。

本研究利用苹果基因组,通过 Dof 结构域序列检索和 RNA-seq 转录本重构,筛选并鉴定出 51 个苹果 Dof 基因家族成员,全面分析了它们编码的 Dof 转录因子的 Dof 结构域、保守基序、亚细胞定位、系统进化树及其特异性表达。以期为进一步研究 Dof 基因在调控苹果生长发育及非生物胁迫响应等过程中的作用奠定基础。

1 材料与方法

1.1 试验材料与处理

平邑甜茶种子用 0.2% KMnO₄ 浸泡 30 min,然后用流水冲洗干净,与洗净的细沙混匀置于 4℃ 层积 60 d。种子露白后播种在装有种植土(细砂、蛭石、土的体积比为 3:2:1)的育苗盘中,并置于滨州学院黄河三角洲生态环境重点实验室苗圃中露天管理,每隔 5 d 浇 1 次 1/2 Hoagland 营养液(pH=6.8±0.2)。当平邑甜茶苗长到 6 片真叶时,用加入了 100 mmol/L NaCl、50 mmol/L Na₂SO₄ 及 50 mmol/L NaHCO₃ 的 1/2 Hoagland 营养液(pH=8)处理,在第 0 d、1 d、3 d、6 d 分别采集叶片和新根(每

个样品 3 次重复)存放在液氮中备用。

1.2 RNA 提取和转录组测序

样品总 RNA 用 TRIzol 试剂盒提取,利用磁珠富集 mRNA,然后利用打断 Buffer 将获得的 RNA 片段化,使用随机 N6 引物进行反转录合成,再合成 cDNA 第二条链。通过特异引物进行 PCR 扩增。热变性使 PCR 产物形成单链,再利用桥式引物将单链 DNA 环化最终得到单链环状 DNA 文库。使用 BGISEQ-500 平台测序,所得原始数据经纯化去除低质量、接头及未知碱基 N 含量过高的 read 后得到可靠的高质量序列,使用 HISAT 将所得序列比对到参考基因组序列(*Malus x domestica* GDDH13 Whole Genome v1.1, https://www.rosaceae.org/species/malus/malus_x_domestica/genome_GDDH13_v1.1)^[11]。

1.3 GDDH13Dof 转录因子筛选与鉴定

利用 Dof 结构域序列,在苹果基因组 GDDH13 v1.1 mRNA 进行 BLAST 序列比对并搜索“Dof”关键词筛选苹果 Dof 候选基因。利用 StringTie (<http://ccb.jhu.edu/software/stringtie>) 对平邑甜茶转录组测序结果进行转录本重构,然后用 Cuffmerge (<http://cole-trapnell-lab.github.io/cufflinks>) 将所有样品的重构信息整合在一起,再使用 Cuffcompare (<http://cole-trapnell-lab.github.io/cufflinks>) 将整合后的转录本与参考注释信息比较,挑选出新转录本,然后使用 CPC2 (<http://cpc2.cbi.pku.edu.cn/>)^[12] 预测新转录本编码蛋白质的潜力,然后将具有蛋白质编码潜力的新转录本加入参考基因中,最后通过 Pfam (<http://pfam.xfam.org>)^[13] 和 SMART (<http://www.smart.embl-heidelberg.de/>) 对鉴定出的 Dof 蛋白进行确认。通过 ExPASy 网站 (<https://web.expasy.org/protparam/>) 对苹果 Dof 蛋白的理化性质进行分析^[14]。

1.4 Dof 蛋白系统进化树、保守域和保守基序分析

从拟南芥信息数据库(TAIR <https://www.arabidopsis.org/>) 下载获得拟南芥的 Dof 转录因子的氨基酸序列。利用 MAFFT version 7 (<https://mafft.cbrc.jp/alignment/server/>) 比对 Dof 结构域^[15]。通过 MAFFT online service 和 Evolview v3 (<https://www.evolgenius.info//evolview/#mytrees/>) 进行系统进化树构建^[15-16],用 Average linkage (UPGMA) 法构建进化树,参数均为默认值。利用 GSDS2 (

gsds.cbi.pku.edu.cn/) 绘制基因结构图^[17]。利用 MEME 5.1.1 版本 (<http://memesuite.org/tools/meme>) 分析 Dof 蛋白保守基序(数目设为 15, 宽度为 6~50 aa)^[18]。Dof 蛋白亚细胞定位用在线软件 WoLF PSORT (<https://wolfpsort.hgc.jp/>) 预测。

1.5 Dof 表达分析与功能注释

从 AppleMDO (<http://bioinformatics.cau.edu.cn/ApplesMDO/>) 网站的 RNA-seq 数据下载 Dof 基因在 7 个不同组织(茎尖、叶片、茎、花、未成熟果肉、成熟果肉、成熟果皮)中的表达数据^[19], 利用 Excel 绘制表达热图。利用 AppleMDO 数据库的 GO analysis 功能以 GDDH13 v1.1 homology with *Arabidopsis* 为参考进行 GO 分析。

2 结果与分析

2.1 Dof 基因的筛选与鉴定

利用 Dof 结构域序列, 在 GDR 数据库苹果基因组 (*Malus x domestica* GDDH13 Whole Genome v1.1)^[11] 进行 BLAST 序列比对, 筛选 Dof 候选基因。并对平邑甜茶转录组测序结果进行转录本重构, 挑选出新转录本, 然后使用 CPC2 预测新转录本的蛋白质编码潜力, 将新转录本中具有蛋白质编码潜力的添加到参考基因中。然后通过 Pfam 和 SMART 在线工具对筛选出的 Dof 蛋白进行确认, 共筛选鉴定出 51 个 Dof 家族成员。但是通过 MAFFT version 7 对这些 Dof 转录因子的 Dof 结构域进行比对, 发现只有 50 个基因具有典型的 Dof 结构域(图 1), 由图 1 可以看出, 有 50 个 Dof 转录因子均含有 1 个明显的高度保守的 Dof 结构域, 且结构域内均含有 CX2CX21CX2C 基序构成的单锌指结构。而 MD07G1265700 虽然通过 Pfam 和 SMART 在线工具被鉴定为 Dof 家族成员, 但是通过序列对比发现其并不存在典型的 CX2CX21CX2C 单锌指结构。对 Dof 基因的定位, 编码蛋白质的相对分子质量、等电点等理化性质进行分析(表 1)。这些 Dof 基因在苹果各个染色体上均有分布, 它们编码的蛋白质的长度为 163~523 aa; 相对分子质量为 18 210~55 800; 等电点为 5.01~10.30。亚细胞定位预测结果表明, 除 MD05G1023800、MD07G1265700、MD10G1280000 可能定位于叶绿体中, MD05G1301100 可能定位于叶绿体或/和细胞核中, MD15G1275300 可能定位于线粒体中, 其他 Dof 蛋白均主要定位于细胞核中。

2.2 Dof 家族蛋白系统进化树

利用鉴定的 Dof 蛋白与拟南芥的 Dof 蛋白氨基酸序列构建了一个系统进化树(图 2)。Dof 蛋白分成 11 组。在这 51 个 Dof 蛋白中, 8 个分到 A 组, 8 个分到 B 组, 2 个分到 C 组, 2 个分到 D 组, 6 个分到 E 组, 5 个分到 F 组, 4 个分到 G 组, 2 个分到 H 组, 9 个分到 I 组, 2 个分到 J 组, 3 个分到 K 组。

2.3 Dof 基因结构分析

利用 GDDH13 v1.1 基因组注释文件 gene_models_20170612.gff3.gz 绘制了 Dof 基因结构图(图 3)。图 3 显示, 进化树分析中亲缘关系近的基因, 它们的基因结构也相似。但是成员较多的 A、B 和 I 组中, 基因结构差异较大。

2.4 Dof 蛋白保守 motif 分析

利用 MEME 在线软件对 Dof 蛋白氨基酸序列进行 motif 分析, 可以看到 Dof 蛋白含有 15 个高度保守的 motif, 虽然它们的数目和位置差异较大, 但是根据进化树分析, 亲缘关系近的 Dof 蛋白的 motif 数量和序列相似度较高(图 4)。其中 50 个 Dof 蛋白均含有 1 个保守的 CX2CX21CX2C 基序(含有 motif 1 或 motif 13), 而 MD07G1265700 既不含有 motif 1, 也不含有 motif 13。

2.5 Dof 基因组织特异性表达

利用 AppleMDO 数据库的 RNA-seq 数据, 分析了 Dof 基因在茎尖、叶片、茎、花、未成熟果肉、成熟果肉和成熟果皮中的表达(图 5)。除 MD12G1041000 基因和 MD14G1040400 基因没在任何组织中检测到表达外, 其他 Dof 基因均至少在一个组织中检测到表达; 这些 Dof 基因在茎、叶片、茎尖等营养器官(组织)中的表达量大多高于在花、果肉、果皮等生殖器官(组织)中的表达量。而 MD01G1084700、MD07G1153300、MD08G1040100、MD15G1034500 基因在未成熟的果肉中表达量最高, 在成熟果肉中均未被检测到。

2.6 平邑甜茶 Dof 基因在盐碱胁迫下的表达分析

近年来, 越来越多的研究结果表明 Dof 基因参与植物对干旱、盐碱等非生物胁迫响应的调控。本研究利用 RNA-seq 技术检测了平邑甜茶在盐碱胁迫 0 d、1 d、3 d、6 d 后叶片和根中 Dof 基因的表达情况(图 6)。在平邑甜茶中, 除 MD05G1023800 基因没有检测到表达外, 其他 Dof 基因均检测到 1 次及以上的表达, 而且大多数 Dof 基因在根中的表达量

显著高于叶片。与对照(0 d)相比,在盐碱胁迫下平邑甜茶根系中表达显著上调的 *Dof* 基因有 MD13G1070600、MD00G1125100、*novelG001312*、MD16G1071400、MD05G1318600、MD10G1297900、MD05G1018200 基因。与对照(0 d)相比,表达显著下调的基因有 MD15G1309400、MD02G1161200、

MD14G1040400、MD01G1041500、MD14G1216800、MD07G1265700、MD15G1034500、MD15G1275300、MD06G1194100、MD06G1173600、MD06G1205900、MD05G1312400、MD14G1180100、MD07G1057700、MD05G1301100 基因。

```

C--C-----锌指结构-----C--C
|  |                               |  |
MD00G1115200  CPRCQSMNTKFCYNNYSLSQPRYFCKACRRYWTQGGTLRNVVGGGCRKGRKRAK--GGSS
MD10G1292200  CPRCHSMNTKFCYNNYSLSQPRYFCKACRRYWTQGGTLRNVVGGGCRKGRKRAK--AGSS
MD07G1265700  -----MNTKFCYNNYSLSQPRYFCKACRRYWTQGGTLRNVVGGGCRKMKRAK--GGSS
MD00G1125100  CPRCNSMDTKFCYNNYNNVQPRHFCKNCQRYWTAGGTMRNVVPGAGRRKNKNSA--SHYR
novelG001312  CPRCNSMDTKFCYNNYNNVQPRHFCKNCQRYWTAGGTMRNVVPGAGRRKNKNSA--SHYR
MD06G1194500  CPRCNSMDTKFCYNNYNNVQPRHFCKACQRYWTAGGTMRNVVPGAGRRKNKNSA--SHYR
MD14G1201600  CPRCNSMDTKFCYNNYNNVQPRHFCKACQRYWTAGGTMRNVVPGAGRRKNKNSA--SHYR
MD13G1070600  CPRCNSMDTKFCYNNYNNVQPRHFCKNCQRYWTAGGTMRNVVPGAGRRKNKNSASQYR
MD16G1071400  CPRCNSMDTKFCYNNYNNVQPRHFCKNCQRYWTAGGTMRNVVPGAGRRKNKNSASQFH
MD05G1301100  CPRCKSMETKFCYFNNYNNVQPRYFCKGCQRYWTAGGALRNVVPGAGRRKTKPPC--QGMA
MD10G1280000  CPRCKSMETKFCYFNNYNNVQPRHFCKGCQRYWTAGGALRNVVPGAGRRKTKPPC--RGLA
MD07G1057700  CPRCKSLETKFCYFNNYSVYQPRHFCKSCQRYWTAGGTIRNVPLGTRRRKNKHSS--SQYH
MD01G1126700  CPRCESTNTKFCYFNNYSLSQPRHFCKTCRRYWTGAGLRNVVGGGCRRNKRKSK--EGNS
MD07G1197500  CPRCESTNTKFCYFNNYSLSQPRHFCKTCRRYWTGAGLRNVVGGGCRRNKRKSK--GGNS
MD03G1096400  CPRCESTNTKFCYFNNYSLSQPRHFCKTCRRYWTGAGLRNVVGGGCRRNKRKSK--SNNS
MD11G1111100  CPRCESTNTKFCYFNNYSLTQPRHFCKTCRRYWTGAGLRNVVGGGCRRNKRKTK--SNSS
MD04G1126500  CPRCESTNTKFCYFNNYSLTQPRHFCKTCRRYWTGAGLRNVVGGGCRRNKRKSK--GTSS
MD12G1078400  CPRCESTNTKFCYFNNYSLTQPRHFCKTCRRYWTGAGLRNVVGGGCRRNKRKSK--GTSS
MD02G1161200  CPRCDSTNTKFCYNNYNNFSQPRHFCKSCRRYWTGAGTLRNPVGGGTRKNKTKRS--RTAT
MD15G1275300  CPRCDSTNTKFCYNNYNNFSQPRHFCKSCRRYWTGAGTLRNPVGGGTRKNKAKRS--RTAA
MD08G1040100  CPRCDSDNTKFCYNNYNNLAQPRHFCKSCRRYWTGAGTLRNVVGGGSRKNKTKRS--RPSC
MD15G1034500  CPRCDSDNTKFCYNNYNNLAQPRHFCKSCRRYWTGAGTLRNVVGGGSRKNKTKRS--RPSC
MD09G1195800  CPRCESSNTKFCYNNYNNLSQPRHFCKNCQRYWTGAGLRNPVGGGSRKNKTKKS--SSSA
MD17G1176100  CPRCESSNTKFCYNNYNNLSQPRHFCKNCQRYWTGAGLRNPVGGGSRKNKTKKS--SSSA
MD06G1205900  CPRCDSLNTKFCYNNYNNLSQPRHFCKACRRYWTGAGVLRNVVGGGCRKTKRKS--TKNS
MD14G1216800  CPRCDSLHTKFCYNNYNNLSQPRHFCKACRRYWTGAGVLRNVVGGGCRKTKRKS--PKNS
MD05G1312400  CPRCDSSNTKFCYNNYNNLSQPRHFCKACKRYWTGAGTLRNVVGGGCRKNKRVK--RPGS
MD10G1292100  CPRCDSSNTKFCYNNYNNLSQPRHFCKACKRYWTGAGTLRNVVGGGCRKNKRVK--RPGS
MD03G1110000  CPRCDSTNTKFCYNNYNNLSQPRYFCKSCRRYWTGAGTLRNVVGGGCRKNKRSS--SAAK
MD11G1123400  CPRCDSTNTKFCYNNYNNLSQPRYFCKSCRRYWTGAGTLRNVVGGGCRKNKRSS--SAKR
MD06G1173600  CPRCESTHTKFCYNNYNNLSQPRYFCKTCRRYWTGAGTLRNPVGGGCRKNKKS--SKKP
MD14G1180100  CPRCDSTHTKFCYNNYNNLSQPRYFCKTCRRYWTGAGTLRNPVGGGCRKNKKS--SKKP
MD01G1084700  CPRCNSTNTKFCYNNYNNLSQPRYFCKTCRRYWTGAGTLRNVVGGGSRKNKKS--SS--
MD07G1153300  CPRCSSTNTKFCYNNYNNLSQPRYFCKTCRRYWTGAGTLRNVVGGGSRKNKKS--SSSA
MD06G1114000  CPRCNSTNTKFCYNNYNNLSQPRYFCKTCRRYWTGAGTLRNPVGGGSRKNKRSS--SSNN
MD14G1135300  CPRCNSTNTKFCYNNYNNLSQPRYFCKTCRRYWTGAGTLRNPVGGGSRKNKRSS--SLNN
MD01G1224500  CPRCNSTTTKFCYNNYNNLSQPRYFCKTCRRYWTGAGTLRNVVGGGSRKNKRK--NNSD
MD07G1295200  CPRCNSTTTKFCYNNYNNLSQPRYFCKTCRRYWTGAGTLRNVVGGGSRKNKRK--NNSD
MD01G1041500  CPRCDSPNTKFCYNNYNNLSQPRHFCKTCRRYWTGAGLRNPVGGGCRKSKKIR--SSSS
MD15G1309400  CPRCDSPNTKFCYNNYNNLSQPRHFCKTCRRYWTGAGLRNPVGGGCRKSKKIR--SSSS
MD02G1163500  CPRCDSSNTKFCYNNYNNLTQPRHFCKTCRRYWTGAGLRNPVGGGCRKNKSVT--VSTA
MD15G1276300  CPRCDSSNTKFCYNNYNNLTQPRHFCKTCRRYWTGAGLRNPVGGGCRKNKSVT--GSTA
MD05G1318600  CPRCESTNTKFCYNNYNNKSPRHFCKRHWTKGAGTLRNVV--GGVRKNKRLK--KSSN
MD10G1297900  CPRCESTNTKFCYNNYNNKSPRHFCKRHWTKGAGTLRNVV--GGVRKNKRLK--KSSN
MD05G1018200  CPRCASSNTKFCYNNYNNLSQPRYFCKGCRRYWTGAGSLRNPVGGGCRKNRRGK--ATRL
MD10G1017800  CPRCASSNTKFCYNNYNNLSQPRYFCKGCRRYWTGAGSLRNPVGGGCRKNRRGK--AARL
MD12G1041000  CPRCGCSNTKFCYNNYNNLSQPRYFCKGCRRYWTGAGSLRNPVGGGCRKNRRGS--KSLR
MD14G1040400  CPRCGCSNTKFCYNNYNNLSQPRYFCKGCRRYWTGAGSLRNPVGGGCRKNRRGS--KSLR
MD05G1023800  CARCGGTNTTFRYFNQGLRKDTRKCHDCNRAWVVGAKLR-----
MD06G1194100  CPRCDSRNTTFDKFSRNNLSKPVRSCKDKRSWVARGNL-----GRG-----
MD14G1201300  CPRCDSRNTTFDKFSRNNPSKPVRSCKDKRSWVARGNL-----GRG-----
* * . . . : * : * . * . . :

```

图1 GDDH13 Dof 蛋白 Dof 保守结构域分析

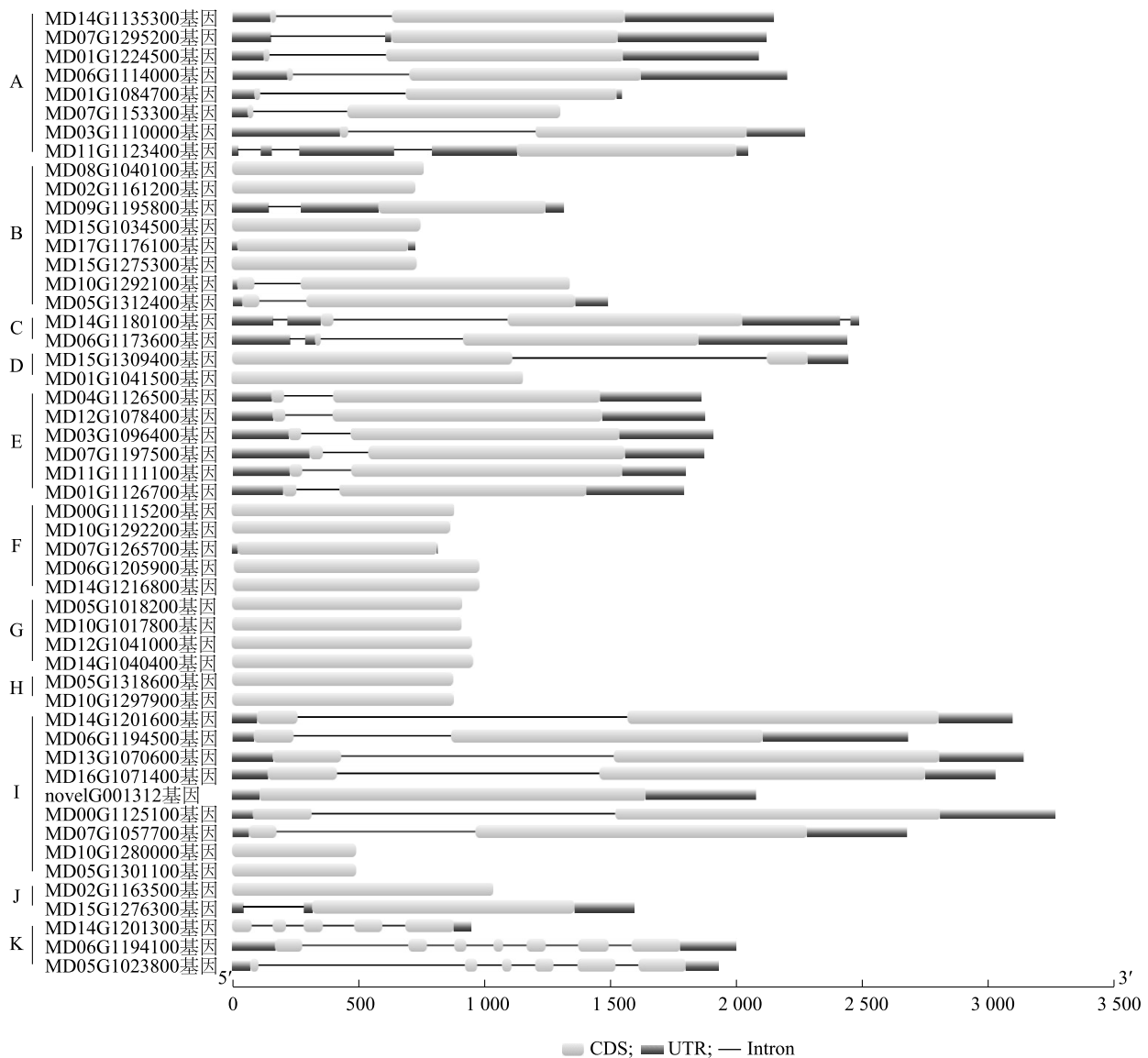
Fig.1 Analysis on the Dof conserved domain in GDDH13 Dof proteins

表 1 GDDH13 *Dof* 基因及 *Dof* 蛋白信息Table 1 Information of *Dof* genes and *Dof* proteins in GDDH13

基因序号	染色体定位	氨基酸	相对分子质量($\times 10^3$)	等电点	亚细胞定位	分组
MD00G1115200	染色体 00: 24529776-24530660	294	31.47	9.28	细胞核	F
MD00G1125100	染色体 00: 26696509-26699779	507	55.12	6.26	细胞核	I
MD01G1041500	染色体 01: 13899805-13900965	386	40.75	8.38	细胞核	D
MD01G1084700	染色体 01: 19126080-19127626	286	31.46	8.15	细胞核	A
MD01G1126700	染色体 01: 23844944-23846738	344	37.54	9.12	细胞核	E
MD01G1224500	染色体 01: 31562069-31564159	321	35.10	7.24	细胞核	A
MD02G1161200	染色体 02: 13514622-13515347	241	25.12	7.58	细胞核	B
MD02G1163500	染色体 02: 13947071-13948105	344	36.51	8.81	细胞核	J
MD03G1096400	染色体 03: 8185366-8187273	372	40.05	8.94	细胞核	E
MD03G1110000	染色体 03: 9569797-9572071	290	32.03	8.11	细胞核	A
MD04G1126500	染色体 04: 21254031-21255893	370	39.42	9.13	细胞核	E
MD05G1018200	染色体 05: 3198536-3199447	303	33.86	5.01	细胞核	G
MD05G1023800	染色体 05: 3917444-3919374	176	20.27	9.88	叶绿体	K
MD05G1301100	染色体 05: 43413218-43413709	163	18.34	9.17	叶绿体/细胞质	I
MD05G1312400	染色体 05: 44216735-44218220	378	41.04	8.58	细胞核	B
MD05G1318600	染色体 05: 44665208-44666083	291	32.33	7.11	细胞核	H
MD06G1114000	染色体 06: 25347209-25349408	314	34.45	9.00	细胞核	A
MD06G1173600	染色体 06: 31334478-31336919	319	35.57	6.61	细胞核	C
MD06G1194100	染色体 06: 32895163-32897161	219	25.07	9.43	细胞核	K
MD06G1194500	染色体 06: 32917350-32920033	465	50.89	5.80	细胞核	I
MD06G1205900	染色体 06: 34063347-34064321	324	34.68	7.21	细胞核	F
MD07G1057700	染色体 07: 5348996-5351671	475	52.06	6.60	细胞核	I
MD07G1153300	染色体 07: 22366665-22367967	289	31.81	7.25	细胞核	A
MD07G1197500	染色体 07: 27551969-27553842	357	38.64	8.99	细胞核	E
MD07G1265700	染色体 07: 33418803-33419618	263	28.53	10.30	叶绿体	F
MD07G1295200	染色体 07: 35586892-35589011	300	32.89	7.68	细胞核	A
MD08G1040100	染色体 08: 2966148-2966909	253	26.91	6.15	细胞核	B
novelG001312	染色体 09: 3499752-3502726	512	55.32	5.75	细胞核	I
MD09G1195800	染色体 09: 17769872-17771185	221	23.60	8.13	细胞核	B
MD10G1017800	染色体 10: 2245517-2246428	303	33.80	5.04	细胞核	G
MD10G1280000	染色体 10: 37109629-37110120	163	18.21	9.40	叶绿体	I
MD10G1292100	染色体 10: 38045551-38046887	378	41.07	7.64	细胞核	B
MD10G1292200	染色体 10: 38063329-38064195	288	30.86	8.90	细胞核	F
MD10G1297900	染色体 10: 38515004-38515882	292	32.52	8.23	细胞核	H
MD11G1111100	染色体 11: 9836381-9838175	375	40.16	8.87	细胞核	E
MD11G1123400	染色体 11: 11347736-11349782	290	32.19	8.35	细胞核	A
MD12G1041000	染色体 12: 4555425-4556378	317	35.09	6.26	细胞核	G
MD12G1078400	染色体 12: 9561904-9563780	373	40.04	9.32	细胞核	E
MD13G1070600	染色体 13: 4971087-4974231	522	55.80	6.05	细胞核	I
MD14G1040400	染色体 14: 3791701-3792657	318	34.92	6.38	细胞核	G
MD14G1135300	染色体 14: 21507409-21509559	315	34.59	8.86	细胞核	A
MD14G1180100	染色体 14: 27265541-27268028	327	36.24	6.72	细胞核	C
MD14G1201300	染色体 14: 29029745-29030691	170	19.58	9.55	细胞核	K

基因序号	染色体定位	氨基酸	相对分子质量($\times 10^3$)	等电点	亚细胞定位	分组
MD14G1201600	染色体 14: 29043548-29046646	466	50.98	5.75	细胞核	I
MD14G1216800	染色体 14: 30008141-30009118	325	34.86	7.16	细胞核	F
MD15G1034500	染色体 15: 2439144-2439890	248	26.31	5.77	细胞核	B
MD15G1275300	染色体 15: 24187449-24188183	244	25.15	8.18	线粒体	B
MD15G1276300	染色体 15: 24268103-24269696	346	36.76	8.83	细胞核	J
MD15G1309400	染色体 15: 31044507-31046953	425	45.03	8.60	细胞核	D
MD16G1071400	染色体 16: 5018920-5021950	523	55.49	5.71	细胞核	I
MD17G1176100	染色体 17: 19911181-19911903	225	24.08	8.50	细胞核	B





A、B、C、D、E、F、G、H、I、J、K 见图 2。

图 3 GDDH13 *Dof* 基因结构

Fig.3 Gene structure of the *Dof* in GDDH13

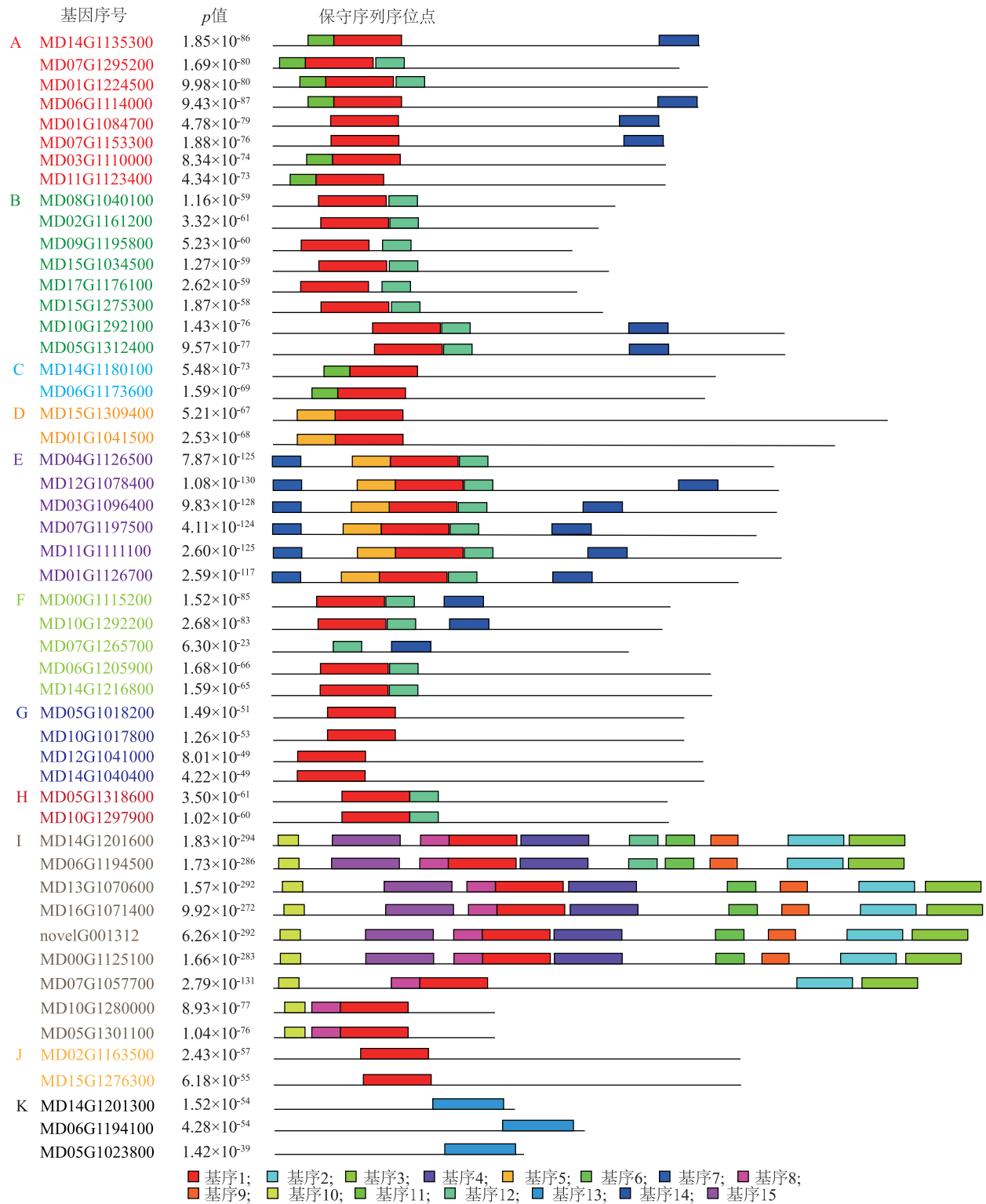
2.7 功能注释

利用 AppleMDO 数据库的 GO analysis 功能对已有的 50 个 *Dof* 基因进行了 GO 功能分析,结果显示,共涉及到 40 个 GO 条目 (图 7),其中细胞组成 (CC) 9 个、分子功能 (MF) 3 个、生物过程 (BP) 28 个。除分子功能中蛋白质结合和生物过程中细胞组分组织、对非生物胁迫反应、对刺激反应等 9 个条目参与的基因较少,其余 GO 条目均有 50 个 *Dof* 基因参与。

3 结论与讨论

Dof 转录因子在植物调节生长发育与响应非生

物胁迫中起着重要作用^[20-22]。目前,*Dof* 基因的报道主要集中在草本植物中^[8-10],在木本植物中报道较少。本研究对 GDDH13 v1.1 基因组预测基因进行了较全面的检索,从已有基因中共检索并鉴定了 50 个 *Dof* 转录因子基因。另外通过对平邑甜茶转录组测序结果进行转录本重构,挑选出新转录本,然后使用 CPC2 对新转录本进行蛋白质编码潜力预测,然后通过 Pfam 和 SMART 在线工具鉴定出 1 个新的 *Dof* 基因,并命名为 *novelG001312*。其中 MD07G1265700 虽然通过 Pfam 和 SMART 在线工具被鉴定为 *Dof* 转录因子家族成员,但是通过氨基酸

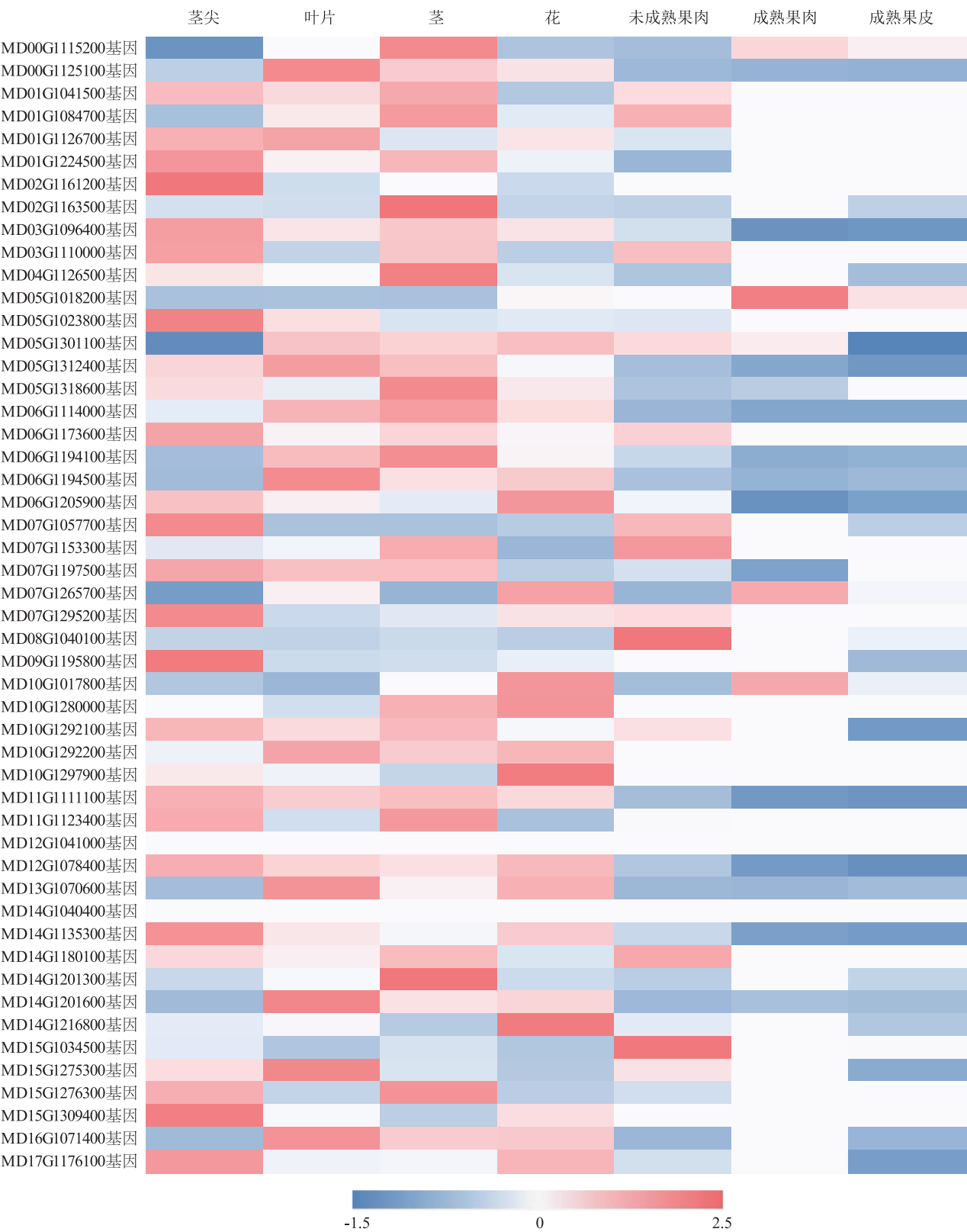


A、B、C、D、E、F、G、H、I、J、K 见图 2。 p 值为随机序列(长度相同且符合背景)与被测序列中的基序匹配概率的乘积小于或等于被测序列的计算值的概率。

图 4 Dof 蛋白的 motif 构成分析

Fig.4 Analysis on motif composing of Dof proteins

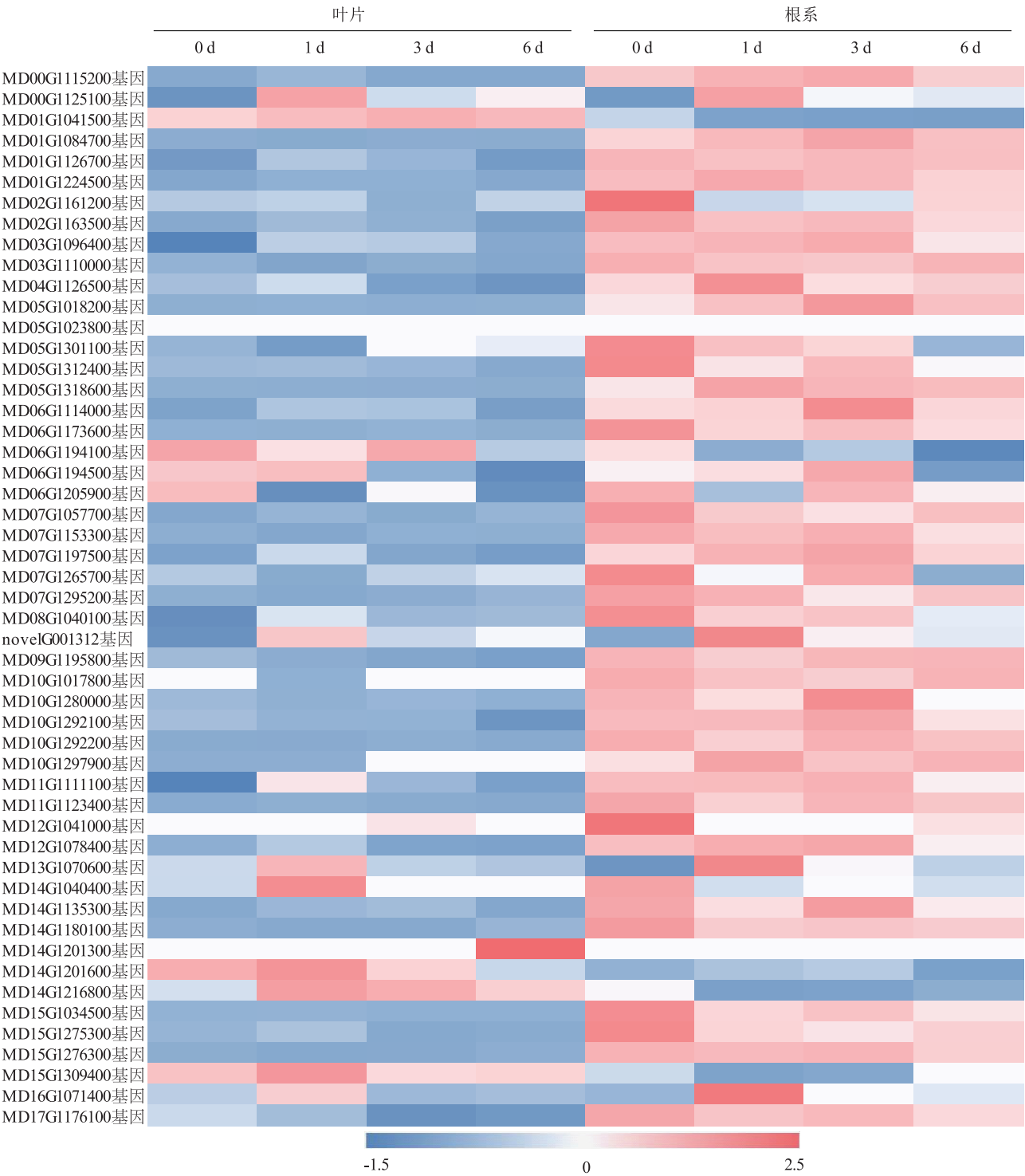
序列比较和 motif 分析发现其并不存在典型的 CX2CX21CX2C 单锌指结构,说明 MD07G1265700



蓝色表示相对表达量最低,红色表示相对表达量最高,0 为组织中没检测到该基因。

图 5 *Dof* 基因不同组织表达分析

Fig.5 Expression analysis on *Dof* genes in different tissues



蓝色表示相对表达量最低,红色表示相对表达量最高,0 为组织中没检测到该基因。

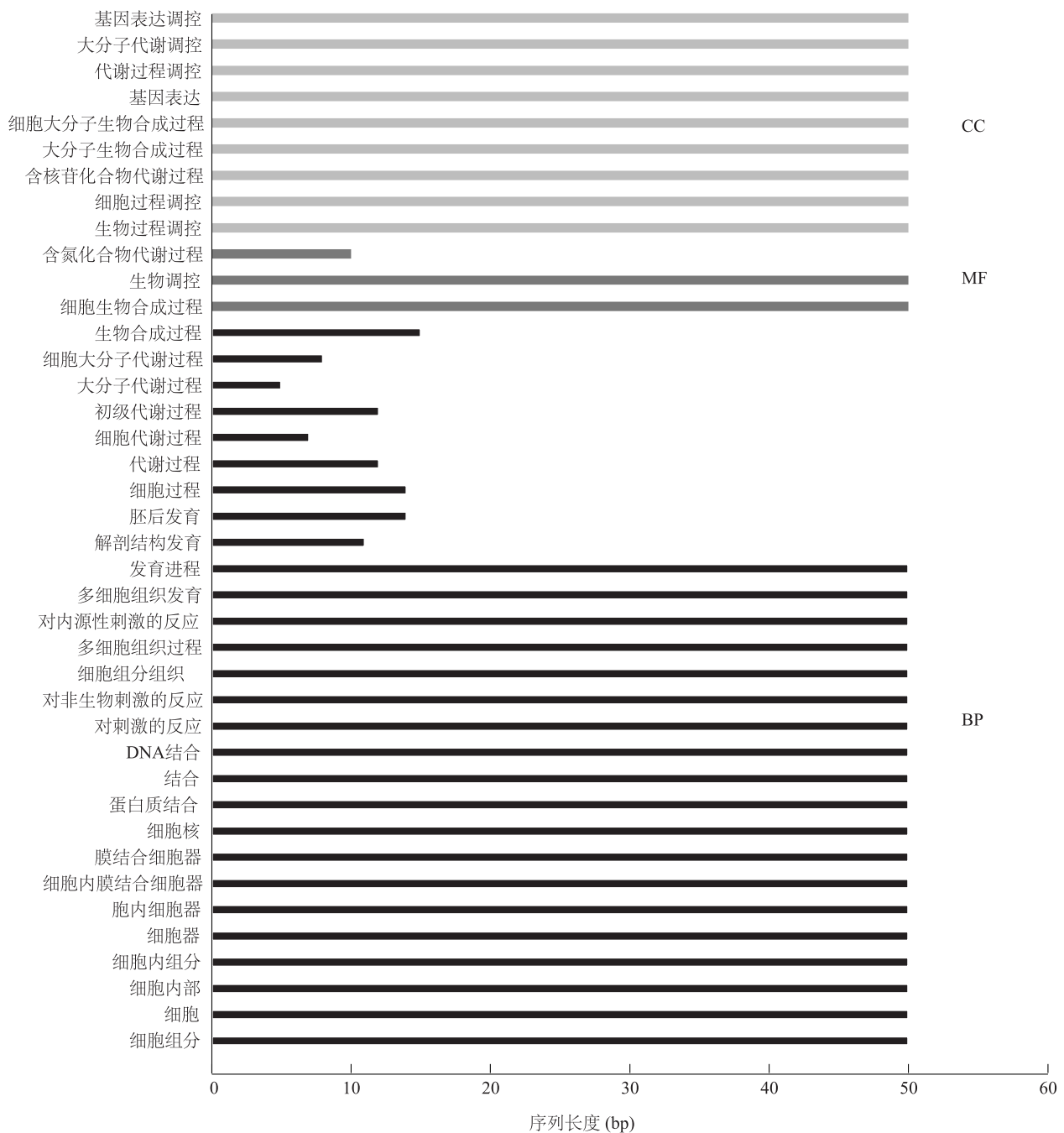
图 6 盐碱胁迫下平邑甜茶叶和根中 Dof 基因的表达分析

Fig.6 Expression analysis on Dof genes in leaves and roots of *Malus hupehensis* Rehd. under saline-alkali stress

可能不属于 Dof 转录因子家族。

有研究表明:不同植物 Dof 蛋白的理论等电点基本处于 5.41~6.97,且一般碱性氨基酸个数

多于酸性氨基酸^[23]。但本研究中苹果 Dof 蛋白的等电点主要集中在 5.01~10.30,多为碱性氨基酸,与小桐子和大豆 Dof 蛋白的等电点研究结果相

图 7 *Dof* 基因 GO 注释Fig.7 GO annotation of *Dof* genes

似^[24-25],说明不同植物间 *Dof* 家族成员的等电点差异较大^[26]。

系统进化树能为预测同源基因的功能提供一些依据。DAG1 (AtDof3.7) 能通过抑制 *GA3ox1* 基因,减少赤霉素含量,从而抑制种子萌发;而 DAG2 (AtDof2.5) 能促进种子萌发^[27-28], MD01G1084700、MD01G1224500、MD06G1114000、MD07G1153300、

MD07G1295200、MD14G1135300 与 DAG1、DAG2 位于 A 组的一个分支上,说明它们也可能参与种子的萌发调控。拟南芥 CDF1、CDF2 和 CDF3 通过抑制 *CONSTANS* 基因转录来控制开花期^[29-30], I 组中 MD00G1125100、MD06G1194500、MD13G1070600、MD14G1201600、MD16G1071400、novelG001312 与上述 3 个转录因子较近,它们也可能参与苹果花期调

控。B 组中, MD09G1195800、MD17G1176100 和 ADOF1 在一个分支, ADOF1 在胼胝质组织中表达量较高, ADOF1 可能影响酚类化合物的合成^[31]。因此, MD09G1195800 和 MD17G1176100 也可能与该功能有关。在氧化胁迫和水杨酸诱导下, OBP1 能调节防御基因表达, OBP1 还与细胞分裂周期有关^[32], 与其在同一分支上的 MD02G1161200、MD08G1040100、MD15G1034500、MD15G1275300 也可能具有类似的功能。

进化树分析中, 位于同一分支的基因亲缘关系近, 而亲缘关系越近^[33-36], Dof 基因的结构及其编码的氨基酸序列的 motif 数量和相似度越高。转录因子家族含有的 motif 在相同的组中是高度保守的^[37]。通过 motif 分析, 51 个 Dof 蛋白含有 15 个高度保守的 motif, 其中 motif 1 和 motif 13 含有 CX2CX21CX2C 单锌指结构。除 MD07G1265700 不含有 motif 1 和 motif 13, MD14G1201300、MD06G1194100、MD05G1023800 各含有 1 个 motif 13, 其他 Dof 蛋白各含有 1 个 motif 1。表明在相同组中含有相同保守基序的 Dof 蛋白也许具有相似的功能。

基因表达模式能为其功能预测提供一定的参考^[37-38]。OBP2 在拟南芥的根和叶片中表达量最高, 在茎和花中表达量低于根和叶^[39], OBP2 可以调控拟南芥硫代葡萄糖苷的生物合成^[40]。同在 E 组的 MD04G1126500、MD12G1078400、MD03G1096400、MD07G1197500、MD11G1111100、MD01G1126700 基因也是多数在根、叶中表达量较高。GO 分析结果表明, 它们都参与生物合成和代谢的调控, 因此推断这些基因可能参与苹果次生代谢的调控。OBP4 在拟南芥各个组织器官中均有表达, 并参与细胞周期和细胞扩增的负调控^[22]。同在 F 组的 MD00G1115200、MD10G1292200、MD07G1265700、MD06G1205900、MD14G1216800 基因也在各个组织均表达, 且多数在花中的表达量最高。说明它们可能主要参与花中细胞周期的调控。MD01G1084700、MD07G1197500、MD08G1040100 和 MD15G1034500 基因在未成熟的果肉中表达量最高, 在成熟果肉中均未被检测到, 它们可能与果实的生长发育有关。

本研究中, 除 MD05G1023800 基因没有检测到表达外, 其他 Dof 基因的表达均对盐碱处理有响应, 只是响应强度和时间有差异。在盐碱胁迫下, 显著上调的 Dof 基因有 7 个, 显著下调的有 15 个。总

之, 苹果大部分 Dof 家族基因可能在盐碱胁迫响应中起重要的调控作用, 但他们的功能还需要进一步验证。

参考文献:

- [1] 张婷婷, 康 慧, 付璐璐, 等. 苹果 *MdCYP707A* 家族基因表达分析和 *MdCYP707A1* 的功能鉴定[J]. 园艺学报, 2019, 46(8): 1429-1444.
- [2] UMEMURA Y, ISHIDUKA T, YAMAMOTO R, et al. The Dof domain, a zinc finger DNA-binding domain conserved only in higher plants, truly functions as a Cys2/Cys2 Zn finger domain[J]. The Plant Journal, 2004, 37(5): 741-749.
- [3] MARZABAL P, GAS E, FONTANET P, et al. The maize Dof protein PBF activates transcription of γ -zein during maize seed development[J]. Plant Molecular Biology, 2008, 67(5): 441-454.
- [4] TOKUNAGA S, SANDA S, URAGUCHI Y, et al. Overexpression of the DOF-Type transcription factor enhances lipid synthesis in *Chlorella vulgaris*[J]. Applied Biochemistry and Biotechnology, 2019, 189(1): 116-128.
- [5] KURAI T, WAKAYAMA M, ABIKO T, et al. Introduction of the *ZmDof1* gene into rice enhances carbon and nitrogen assimilation under low-nitrogen conditions[J]. Plant Biotechnology Journal, 2011, 9(8): 826-837.
- [6] 李 娅, 丁文杰, 江海燕, 等. Dof 基因家族调节植物生长发育功能的研究进展[J]. 西北植物学报, 2018, 38(9): 1758-1766.
- [7] YANAGISAWA S. The Dof family of plant transcription factors[J]. Trends in Plant Science, 2002, 7(12): 555-560.
- [8] SHAW L M, MCINTYRE C L, GRESSHOFF P M, et al. Members of the Dof transcription factor family in *Triticum aestivum* are associated with light-mediated gene regulation[J]. Functional & Integrative Genomics, 2009, 9(4): 485-498.
- [9] GUO Y, QIU L J. Genome-wide analysis of the Dof transcription factor gene family reveals soybean-specific duplicable and functional characteristics[J]. PLoS One, 2013, 8(9): e76809.
- [10] LIJAVETZKY D, CARBONERO P, VICENTE-CARBAJOSA J. Genome-wide comparative phylogenetic analysis of the rice and *Arabidopsis* Dof gene families[J]. BMC Evolutionary Biology, 2003, 3(1): 17.
- [11] JUNG S, LEE T, CHENG C H, et al. 15 years of GDR: New data and functionality in the Genome Database for Rosaceae[J]. Nucleic Acids Research, 2019, 47(D1): D1137-D1145.
- [12] KANG Y J, YANG D C, KONG L, et al. CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features[J]. Nucleic Acids Research, 2017, 45(W1): W12-W16.
- [13] EL-GEBALI S, MISTRY J, BATEMAN A, et al. The Pfam protein families database in 2019[J]. Nucleic Acids Research, 2019, 47(D1): D427-D432.
- [14] GASTEIGER E, HOOGLAND C, GATTIKER A, et al. Protein i-

- dentification and analysis tools on the ExPASy server[M]. New Jersey: Humana Press, 2005: 571-607.
- [15] KATO K, ROZEWICKI J, YAMADA K D. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization[J]. Briefings in Bioinformatics, 2019, 20(4): 1160-1166.
- [16] SUBRAMANIAN B, GAO S, LERCHER M J, et al. Evolvview v3: a webserver for visualization, annotation, and management of phylogenetic trees[J], Nucleic Acids Research, 2019, 47(W1): W270-W275.
- [17] HU B, JIN P, GUO A Y, et al. GSDB 2.0: an upgraded gene feature visualization server[J]. Bioinformatics, 2015, 31(8): 1296-1297.
- [18] BAILEY T L, ELKAN C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers[C] // ALTMAN R, BRUTLAG D, KARP P, et al. Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology. Menlo Park, California: AAAI Press, 1994: 28-36.
- [19] DA L L, LIU Y, YANG J T, et al. AppleMDO: a multi-dimensional omics database for apple co-expression networks and chromatin states[J]. Frontiers in Plant Science, 2019(10): 1333.
- [20] EWAS M, KHAMES E, ZIAF K, et al. The tomato DOF daily fluctuations 1, TDDF1 acts as flowering accelerator and protector against various stresses[J]. Scientific Reports, 2017(7): 10299.
- [21] CORRALES A R, CARRILLO L, LASIERRA P, et al. Multifaceted role of cycling DOF factor 3 (CDF3) in the regulation of flowering time and abiotic stress responses in *Arabidopsis* [J]. Plant Cell and Environment, 2017, 40(5): 748-764.
- [22] XU P P, CHEN H Y, YING L, et al. At DOF5.4/OBP4, a DOF transcription factor gene that negatively regulates cell cycle progression and cell expansion in *Arabidopsis thaliana* [J]. Scientific Reports, 2016(6): 27705.
- [23] 李辉, 黄蔚, 刘志薇, 等. 茶树两个 Dof 转录因子的分离及其在温度胁迫中的响应分析[J]. 茶叶科学, 2016, 36(3): 312-322.
- [24] 王海波, 唐利洲. 基于基因组鉴定小桐子 Dof 转录因子家族及其表达分析[J]. 分子植物育种, 2018, 16(3): 764-771.
- [25] 刘蓓, 邱爽, 何佳琦, 等. 8 个大豆 Dof 转录因子的生物信息学分析及干旱诱导表达[J]. 大豆科学, 2020, 39(3): 377-383.
- [26] 程冬梅, 邓志勇, 郭霁光. 小麦高分子量麦谷蛋白亚基等电点的特性分析[J]. 西北植物学报, 2006, 26(3): 532-536.
- [27] GABRIELE S, RIZZA A, MARTONE J, et al. The Dof protein DAG1 mediates PIL5 activity on seed germination by negatively regulating GA biosynthetic gene *AtGA3ox1* [J]. The Plant Journal, 2010, 61(2): 312-323.
- [28] GUALBERTI G, PAPI M, BELLUCCI L, et al. Mutations in the Dof zinc finger genes *DAG2* and *DAG1* influence with opposite effects the germination of *Arabidopsis* seeds [J]. The Plant Cell, 2002, 14(6): 1253-1263.
- [29] SONG Y H, SMITH R W, TO B J, et al. FKF1 conveys crucial timing information for CONSTANS stabilization in the photoperiodic flowering [J]. Science, 2012, 336(6084): 1045-1049.
- [30] FORNARA F, PANIGRAHI K C S, GISSOT L, et al. *Arabidopsis* DOF transcription factors act redundantly to reduce CONSTANS expression and are essential for a photoperiodic flowering response [J]. Developmental Cell, 2009, 17(1): 75-86.
- [31] TSUJIMOTO-INUI Y, NAITO Y, SAKURAI N, et al. Functional genomics of the Dof transcription factor family genes in suspension-cultured cells of *Arabidopsis thaliana* [J]. Plant Biotechnology, 2009(26): 15-28.
- [32] SKIRYCZ A, RADZIEJWOSKI A, BUSCH W, et al. The DOF transcription factor OBP1 is involved in cell cycle regulation in *Arabidopsis thaliana* [J]. The Plant Journal, 2008, 56(5): 779-792.
- [33] 裴徐梨, 荆赞革, 徐境, 等. 青花菜 *BoDof5.3* 基因的克隆及渍水胁迫表达特征分析[J]. 江苏农业学报, 2020, 36(6): 1498-1502.
- [34] 韩利红, 刘潮, 张维维, 等. 铁皮石斛热激转录因子 (*Hsf*) 基因家族鉴定及生物信息学分析[J]. 南方农业学报, 2019, 50(4): 677-684.
- [35] 庞文玉, 王安, 杨宝谊, 等. 大白菜 *ENT* 基因家族的鉴定与生物信息学分析[J]. 江苏农业科学, 2019, 47(12): 52-57.
- [36] 杨冬静, 孙厚俊, 谢逸萍, 等. 甘薯等 8 种植物 *JAZ1* 基因的生物信息学分析[J]. 江苏农业学报, 2019, 35(5): 1021-1027.
- [37] 唐跃辉, 包欣欣, 王健, 等. 小桐子 *Dof* 基因家族生物信息学与表达分析[J]. 江苏农业学报, 2019, 35(1): 15-25.
- [38] SADDEE A A, MALVANKAR M R, KUMAR K. Selection of reference genes for quantitative real-time PCR analysis in halophytic plant *Rhizophora apiculata* [J]. The Journal of Life and Environmental Sciences, 2018, 6(8): e5226.
- [39] KANG H G, SINGH K B. Characterization of salicylic acid-responsive, *Arabidopsis* Dof domain proteins: overexpression of *OBP3* leads to growth defects [J]. The Plant Journal, 2000, 21(4): 329-339.
- [40] SKIRYCZ A, REICHELT M, BUROW M, et al. DOF transcription factor AtDof1.1 (OBP2) is part of a regulatory network controlling glucosinolate biosynthesis in *Arabidopsis* [J]. The Plant Journal, 2006, 47(1): 10-24.

(责任编辑:陈海霞)