



大豆 Whirly 基因家族的鉴定和表达分析

李清^{1,2}, 王洪¹, 郭禄芹^{1,3}, 罗秋兰^{1,4}

(1. 深圳大学 生命与海洋科学学院, 广东 深圳 518060; 2. 深圳大学 光电工程学院/光电子器件与系统(教育部/广东省重点实验室), 广东 深圳 518060; 3. 河南农业大学 园艺学院, 河南 郑州 450002; 4. 韩山师范学院 食品工程与生物科技学院, 广东 潮州 521000)

摘要: Whirly 家族是一类植物特异的转录因子, 它们能与单链 DNA 分子结合, 无论在细胞核还是在细胞器内都有着广泛而复杂的生物学功能。然而, 目前还没有关于大豆 Whirly 基因的研究报道。本研究首次在全基因组水平鉴定和综合描述大豆 Whirly 基因家族成员, 以期为其后续功能研究提供理论基础。该研究在大豆基因组中鉴定到了 7 个 Whirly 基因, 并对其理化性质、亚细胞定位、进化关系、基因结构、蛋白基序组成、三维结构和基因表达模式等进行综合分析。根据它们的进化关系、基因结构和蛋白基序组成, 将其分为两大类。蛋白三维结构分析结果表明, GmWHY5 蛋白由于 N 端的缺失, 缺少了 $\beta 1-4$ 和 $\alpha 1$, 与其它大豆 Whirly 蛋白结构差异很大。利用已公布的转录组数据进一步研究了大豆 Whirly 基因的表达谱, 结果显示其组织和胁迫表达模式差异很大。例如 GmWHY1、GmWHY2、GmWHY6 和 GmWHY7 的表达水平明显高于 GmWHY3、GmWHY4 和 GmWHY5。总之, 这些发现将为大豆 Whirly 基因家族的研究提供重要的资源, 并为进一步阐明其分子机制奠定基础。

关键词: 大豆; Whirly 基因家族; 基因表达; 干旱; 盐害

Identification and Expression Analysis of the Whirly Gene Family in Soybean

LI Qing^{1,2}, WANG Hong¹, GUO Lu-qin^{1,3}, LUO Qiu-lan^{1,4}

(1. College of Life Science and Oceanography, Shenzhen University, Shenzhen 518060, China; 2. College of Optoelectronic Engineering/Key Laboratory of Optoelectronic Devices and Systems of Ministry of Education and Guangdong Province, Shenzhen University, Shenzhen 518060, China; 3. College of Horticulture, Henan Agricultural University, Zhengzhou 450002, China; 4. College of Food Engineering and Biotechnology, Hanshan Normal University, Chaozhou 521000, China)

Abstract: Whirly family is a kind of plant-specific transcription factor, which can bind to single-stranded DNA and play myriad roles both in the nucleus and organelles. However, no research on the functions of the Whirly genes in soybean has been reported. To provide a theoretical basis for further functional study of soybean Whirly genes, the genome-wide identification and comprehensive description of this family were firstly carried out in this study. In total, 7 Whirly genes were identified in soybean, and then their physical and chemical properties, subcellular localizations, evolutionary relationships, gene structures, motif compositions, three-dimensional structures and gene expression patterns were comprehensively analyzed. The 7 GmWHY genes were divided into two clades based on their phylogenetic relationships, gene structures and the motif compositions. Protein three-dimensional structure analysis showed that GmWHY5 lacked $\beta 1-4$ and $\alpha 1$ due to the absence of N-terminal, which was quite different from other soybean Whirly proteins. We further studied the expression profiles of GmWHY genes using the released transcriptome data, and found their tissular and stressed expression patterns varied a lot. For example, the expression levels of GmWHY1, GmWHY2, GmWHY6 and GmWHY7 were significantly higher than those of GmWHY3, GmWHY4 and GmWHY5. Together, these findings will provide a valuable resource for characterizing the GmWHY gene family and lay a foundation for further elucidation of their molecular mechanisms.

Keywords: Soybean; Whirly gene family; Gene expression; Drought; Salt stress

Whirly 蛋白家族是近年来发现的一类广泛存在于植物中的高度保守的单链 DNA (ssDNA) 结合蛋白。该蛋白家族通常包含 3 个功能区域: N 端区域, 中间的 Whirly 结构域和 C 端多变区。N 端区域一般包含叶绿体(或线粒体)信号肽和转录激活结构域。中间的 Whirly 结构域非常保守, 是与单链 DNA

结合的区域, 其中还包含一段细胞核定位信号。C 端多变区具有自我调节结构域, 可以调控 Whirly 蛋白与 ssDNA 的结合活性^[1-2]。Desveaux 等^[3]于 2000 年从马铃薯中分离出首个 Whirly 家族成员 PBF-2 (PR-10a binding factor 2), 后来又称为 StWHY1^[1]。它可以与病原响应基因 PR-10a 启动子上的顺式应

收稿日期: 2018-11-28

基金项目: 中国博士后基金(2017M622751); 广东省自然科学基金(2018A030310455); 深圳孔雀创新团队基金(KQTD2017032715165926); 广东省生物学特色重点学科专项经费。

第一作者简介: 李清(1986-), 男, 博士, 助理研究员, 主要从事植物分子生物学研究。E-mail: liqing1986102@163.com。

通讯作者: 罗秋兰(1984-), 女, 博士, 副教授, 主要从事植物分子生物学研究。E-mail: luoqiulan_79@163.com。

答元件ERE(elicitor response element)以单链形式结合,从而诱导该基因的表达,调节病原菌引起的信号传导途径^[3]。随后,在拟南芥^[4]、水稻^[5]、玉米^[6-7]、大麦^[8]和番茄^[9]等众多植物中陆续鉴定到 Whirly 基因家族成员的存在。例如在模式植物拟南芥中发现了3个该家族成员 *AtWHY1*、*AtWHY2* 和 *AtWHY3*^[4]。*AtWHY1* 和 *AtWHY3* 的 N 端均含有叶绿体或质体定位信号,主要定位于叶绿体或质体中,少量定位于细胞核中^[4,10-11]。*AtWHY2* 的 N 端则含有线粒体定位信号,主要定位于线粒体中,是否也定位于细胞核中目前还有待证实^[4,12]。Whirly 蛋白的不同亚细胞定位与它们的功能存在着紧密的联系。

ssDNA 结合蛋白在核酸代谢过程中发挥着特别重要的调控作用,参与端粒合成、DNA 复制、转录、重组和修复等调控过程^[13]。Whirly 家族作为一类植物特异的 ssDNA 结合蛋白,其功能发挥依赖于它和 ssDNA 的结合。目前已知 Whirly 蛋白能与4种核酸序列相结合:包括 ERE 元件^[3,14-15],端粒末端序列^[10],*AtKPI* 基因启动子上的 KPRE 元件^[16],以及 *WRKY53* 基因启动子上同时包含的 ERE-like 基序和类似于端粒重复序列的 AT 富集区^[17]。除了同时包含 ERE-like 和 A/T 富集区的序列以外,其它的3种核酸序列之间没有明显的相似性。近年来的研究表明 Whirly 蛋白不仅能调控植物抗病信号转导(细胞核中)^[3,14],而且在保持端粒长度(细胞核中)^[10],维护质体和线粒体基因组稳定、损伤修复和基因表达调控(质体和线粒体中)^[6,12,18-20],调节叶片衰老(细胞核)^[17,21-22],以及控制角果、胚胎发育和种子萌发(质体和线粒体中)^[7,23]等方面都发挥着重要的功能。

大豆起源于中国,是一种重要的粮油兼用作物,也是人类优质蛋白和动物饲料蛋白的主要来源。自大豆基因组公布以来,大豆基因功能的研究进程明显加快,但是大豆中关于 Whirly 基因的研究报道还没有。为此,本研究对大豆 Whirly 基因家族进行了生物信息学分析,首次在全基因组水平鉴定大豆 Whirly 基因家族成员,并对其进行理化特征、系统发生树、基因结构、蛋白基序、蛋白三维结构、亚细胞定位、组织器官和胁迫表达模式分析,以期为进一步研究 Whirly 家族成员的功能提供理论依据。

1 材料与方法

1.1 材料

大豆(*Glycine max*, Taxonomic ID 3847)和莱茵

衣藻(*Chlamydomonas reinhardtii*, Taxonomic ID 3055)、团藻(*Volvox carteri*, Taxonomic ID 3067)、杜氏盐藻(*Dunaliella salina*, Taxonomic ID 3046)、小立碗藓(*Physcomitrella patens*, Taxonomic ID 3218)、地钱(*Marchantia polymorpha*, Taxonomic ID 3197)、卷柏(*Selaginella moellendorffii*, Taxonomic ID 88036)、无油樟(*Amborella trichopoda*, Taxonomic ID 13333)、水稻(*Oryza sativa*, Taxonomic ID 39947)、二穗短柄草(*Brachypodium distachyon*, Taxonomic ID 15368)、高粱(*Sorghum bicolor*, Taxonomic ID 4558)、玉米(*Zea mays*, Taxonomic ID 4577)、雷蒙德氏棉(*Gossypium raimondii*, Taxonomic ID 29730)、拟南芥(*Arabidopsis thaliana*, Taxonomic ID 3702)、甘蓝(*Brassica oleracea*, Taxonomic ID 3712)、苹果(*Malus domestica*, Taxonomic ID 3750)、毛果杨(*Populus trichocarpa*, Taxonomic ID 3694)、黄瓜(*Cucumis sativus*, Taxonomic ID 3659)、苜蓿(*Medicago truncatula*, Taxonomic ID 3880)等其它18个物种的基因组和 Whirly 基因信息分别来源于 NCBI (<https://www.ncbi.nlm.nih.gov/>) 和 phytozome (<https://phytozome.jgi.doe.gov/pz/portal.html>) 数据库。

1.2 试验设计

利用同源比对和蛋白结构域分析鉴定大豆 Whirly 基因家族成员。根据系统发生树、基因结构和蛋白保守基序对大豆 Whirly 基因进行分类。利用 Phytozome 数据库中的 RNA-seq 表达数据研究大豆 Whirly 基因在不同组织和器官中的表达模式。利用已公布的 RNA-seq 数据检测大豆 Whirly 基因在干旱或盐处理条件下的表达情况。利用 PlantCARE 数据库(<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>)分析大豆和拟南芥 Whirly 基因的启动子序列(ATG 上游1 500 bp 序列)上的顺式作用元件。

1.3 方法

1.3.1 大豆 Whirly 基因成员的鉴定 以拟南芥 *AtWHY1*、*AtWHY2* 和 *AtWHY3* 氨基酸序列作为诱饵在全基因组水平检索大豆基因组数据库(<https://phytozome.jgi.doe.gov/pz/portal.html>)。Whirly 成员具有典型的 Whirly 结构域。进一步使用 Pfam 工具^[24](<http://pfam.xfam.org/>)去除无典型 Whirly 结构域的同源序列。剔除同一基因的非主要剪接形式,最终得到所有大豆 Whirly 基因家族成员。其它18个代表物种中的 Whirly 基因家族成员采用类似的方法进行筛选。根据大豆和18个代表物种的基因组信息,利用 phyloT 工具(<https://phylot.biobyte.de/>)构建物种进化树。

1.3.2 进化和描述大豆 *Whirly* 成员 利用 ClustalW 程序对大豆和拟南芥 *Whirly* 成员的氨基酸序列进行比对。使用 MEGA 6.0 软件构建大豆和拟南芥的 *Whirly* 蛋白序列的邻接 (NJ) 进化树,bootstrap 值设置为1 000^[25]。利用 GSDS 2.0 在线软件^[26] (<http://gsds.cbi.pku.edu.cn/>) 绘制大豆和拟南芥 *Whirly* 基因的外显子-内含子结构图。利用 MEME 程序^[27] (<http://meme-suite.org/>) 鉴定保守的 *Whirly* 蛋白基序,蛋白基序数设置为 8。*Whirly* 蛋白的相对分子量和等电点使用 Protparam 程序 (<https://web.expasy.org/protparam/>) 进行分析。亚细胞定位使用 TargetP 1.1 软件^[28] (<http://www.cbs.dtu.dk/services/TargetP/>) 和 WoLF PSORT 软件 (<https://wolfpsort.hgc.jp/>) 进行预测。*Whirly* 蛋白的三维结构利用 Phyre2 软件^[29] (<http://www.sbg.bio.ic.ac.uk/phyre2/html/>) 预测。

1.3.3 转录组数据分析 大豆 *Whirly* 基因家族成员在不同组织和器官中的表达数据来源于 Phytozome 数据库,在干旱和盐胁迫条件下的原始转录组数据来源于文献^[30-31]。利用 HISAT 软件^[32] 将原始 reads 映射到大豆参考基因组 (Wm82. a2. V 1) 上,使用 StringTie 软件^[33] 进行转录本组装和表达值计算。基因的表达量经过标准化处理,用 FPKM (fragments per kilobase of exon per million fragments mapped) 值表示。差异表达用处理样本和对照样本 (0 h 处理)

的表达比 fold-change (FC) 表示。基因表达值和差异表达值以 log2 形式转化,并利用 Heml 工具以热图形式展示^[34]。

2 结果与分析

2.1 全基因组水平鉴定大豆 *Whirly* 基因家族成员

在大豆基因组中检索,去除无典型 *Whirly* 结构域的冗余序列和同一基因的非主要转录形式以后,一共发现 7 个大豆 *Whirly* 基因,分别命名为 *Gm-WHY1* ~ 7,依次分布在 18、8、1、2、2、3 和 19 号染色体上(表 1)。7 个 GmWHYs 蛋白序列长度为 126 ~ 269,相对分子量为 14 607. 87 ~ 29 946. 88 Da,等电点为 5. 54 ~ 9. 71(表 1)。由于 GmWHY5 蛋白的 N 端缺失了 100 多个氨基酸残基,导致 GmWHYs 蛋白的基本理化性质差异较大。另外,这些 GmWHYs 蛋白两两之间的序列一致性差异也很大,变化范围为 22. 5% ~ 94. 3%。由于拟南芥 *Whirly* 成员主要定位在叶绿体或线粒体中,利用 Target P 1.1 软件对 GmWHYs 蛋白的亚细胞定位情况进行预测,发现 *GmWHY1* ~ 4 蛋白定位在叶绿体中,*GmWHY6* 和 *GmWHY7* 则定位于线粒体中(表 1)。利用 WoLF PSORT 在线工具预测 *GmWHY5* 的亚细胞定位,发现其定位在胞质中。*GmWHYs* 蛋白在细胞内不同区域的定位暗示它们具有不同的功能。

表 1 大豆 *Whirly* 基因家族成员的基本信息
Table 1 The basic information of *Whirly* gene family members in soybean

基因 Gene name	ID	基因定位 Gene localization	蛋白长度 Length/aa	蛋白分子量 MW/Da	蛋白等电点 pI	亚细胞定位 Sub-cellular localization
<i>GmWHY1</i>	Glyma. 18G124800. 1	Chr18:16338250 ~ 16340740 -	229	25231. 81	9. 54	叶绿体
<i>GmWHY2</i>	Glyma. 08G297200. 1	Chr08:41340473 ~ 41343853 +	269	29946. 88	9. 28	叶绿体
<i>GmWHY3</i>	Glyma. 01G043000. 1	Chr01:4683289 ~ 4687883 +	263	29740. 20	9. 56	叶绿体
<i>GmWHY4</i>	Glyma. 02G019900. 1	Chr02:1816417 ~ 1821235 -	262	29790. 50	9. 71	叶绿体
<i>GmWHY5</i>	Glyma. 02G020400. 1	Chr02:1854656 ~ 1859613 -	126	14607. 87	5. 54	细胞质
<i>GmWHY6</i>	Glyma. 03G252100. 1	Chr03:44768483 ~ 44772056 +	242	26998. 73	9. 10	线粒体
<i>GmWHY7</i>	Glyma. 19G249500. 1	Chr19:49547385 ~ 49551367 +	235	26123. 84	9. 24	线粒体

2.2 *Whirly* 基因家族在不同物种中的比较分析

为了进行比较基因组分析,本研究也分析了其它 18 个物种的 *Whirly* 基因家族成员。这 18 个代表性物种,包括 7 个双子叶植物、4 个单子叶植物、1 个基础被子植物、1 个裸子植物、1 个地钱植物、1 个苔藓植物和 3 个绿藻植物(图 1)。经过筛选,在这些物种中一共鉴定到 36 个 *Whirly* 基因。大体上双子叶植物比其它植物具有更多的 *Whirly* 基因。另外,

Whirly 基因数目与基因组的大小并不相关。玉米具有最大的基因组,但是它的 *Whirly* 基因数目和平均基因密度却低于拟南芥、甘蓝 (*Brassica oleracea*)、苹果 (*Malus domestica*)、毛果杨 (*Populus trichocarpa*)、大豆 (*Glycine max*) 和苜蓿 (*Medicago truncatula*)。反之,物种的染色体倍数可能会影响 *Whirly* 基因数目。古四倍体的大豆拥有最多的 *Whirly* 基因数目,大约是其它二倍体被子植物的 2 ~ 3 倍。

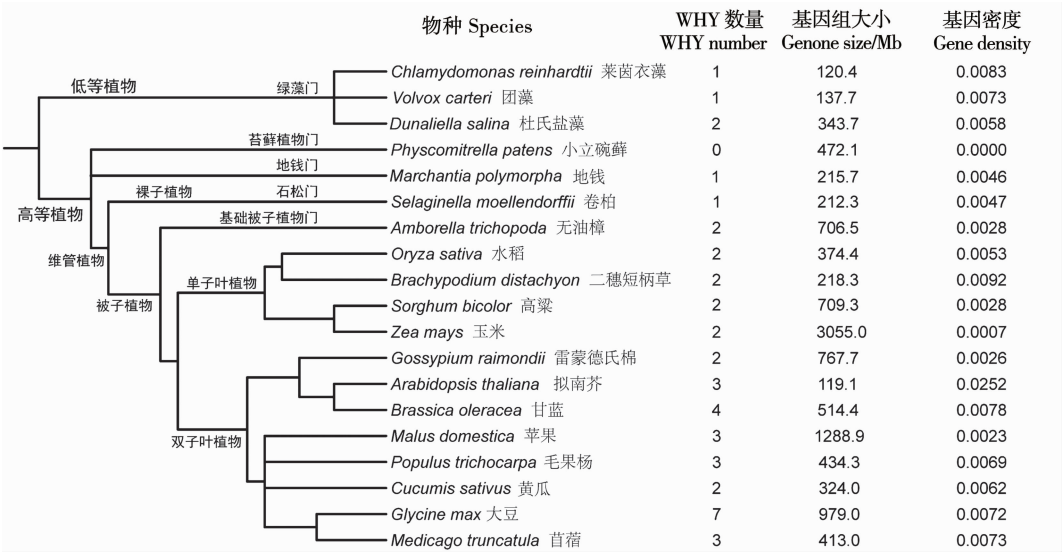


图 1 大豆和其它 18 个代表物种中 Whirly 家族成员概述

Fig. 1 Summary of the Whirly family in soybean and 18 other representative species

2.3 Whirly 基因家族的进化树和分类分析

对大豆 Whirly 基因家族成员的系统发生树分析结果如图 2 所示:根据进化树的拓扑结构和 bootstrap 值,发现这些 Whirly 蛋白可以分成两大类。分类 I 包含 AtWHY1、AtWHY3、GmWHY1 ~ GmWHY5, 分类 II 则包括 AtWHY2、GmWHY6 和 GmWHY7。基因家族成员的外显子/内含子多样化在多基因家族的进化过程中起着非常重要的作用^[35]。对每个 Whirly 基因的外显子和内含子的数量以及位置进一步分析,结果如图 3 所示:所有的 Whirly 基因都具有多个外显子;分类 I 除了 GmWHY5 只包含 3 个外显子外,其它 Whirly 基因均含有 6 ~ 7 个外显子;分类 II 均包含 8 个外显子。来自相同亚家族且具有相似蛋白基序组合的蛋白质很可能具有同样的功能^[27]。为此,本研究利用 MEME 程序来预测大豆和拟南芥 Whirly 蛋白中最保守的 8 个氨基酸基序(图 3)。结果与预期一致:在分类 I 中,除了 GmWHY5 少了 N 端部分的蛋白基序外,其它 Whirly 蛋白都具有蛋白

基序 6、3、1、4、7,大部分还包含蛋白基序 2 和 5,少部分含有蛋白基序 8;分类 II 中的 Whirly 蛋白均只包含蛋白基序 3、1、4、2,与分类 I 相比,少了蛋白基序 6 和 7。

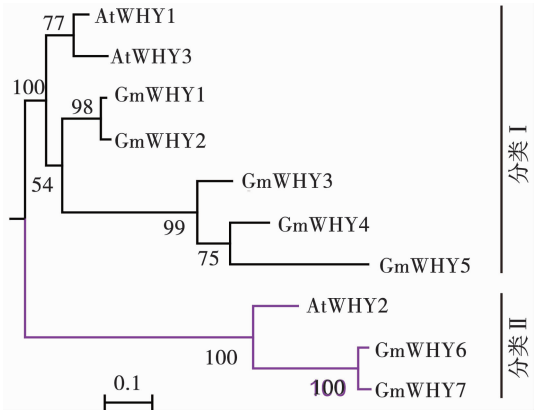


图 2 大豆和拟南芥 Whirly 基因家族的系統发生树

Fig. 2 Phylogenetic tree of Whirly gene family in Arabidopsis and soybean

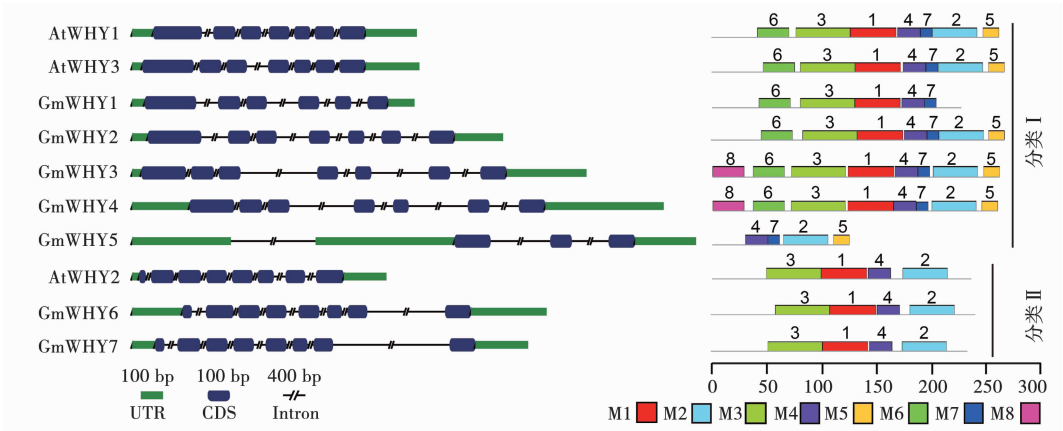


图 3 大豆和拟南芥 Whirly 基因家族的基因结构和保守基序

Fig. 3 Gene structures and conserved motifs of Whirly gene family in Arabidopsis and soybean

2.4 Whirly 蛋白的同源建模

对 Whirly 家族成员的三维蛋白结构预测结果如图 4 所示:除了 GmWHY5 和 GmWHY1 以外,其它的 Whirly 蛋白与之前报道的 StWHY1 蛋白结构类似^[36],都由 8 个 β -折叠($\beta 1 \sim 8$)和 3 个 α -螺旋($\alpha 1 \sim 3$)构成。4 个反向平行的 β -折叠组成 1 个 β -片层, $\beta 1 \sim 4$ 形成 β -片层 1, $\beta 5 \sim 8$ 形成 β -片层

2;两个 β -片层之间互相垂直,由 $\alpha 1$ 连接;C 端区域包含两个 α -螺旋($\alpha 2 \sim 3$)。GmWHY5 蛋白由于 N 端的缺失,缺少了 $\beta 1 \sim 4$ 和 $\alpha 1$ 。GmWHY1 蛋白的 C 端则缺少 $\alpha 3$ 。GmWHY5 和 GmWHY1 与其它 Whirly 蛋白不同的三维结构暗示它们可能具有不同的功能。

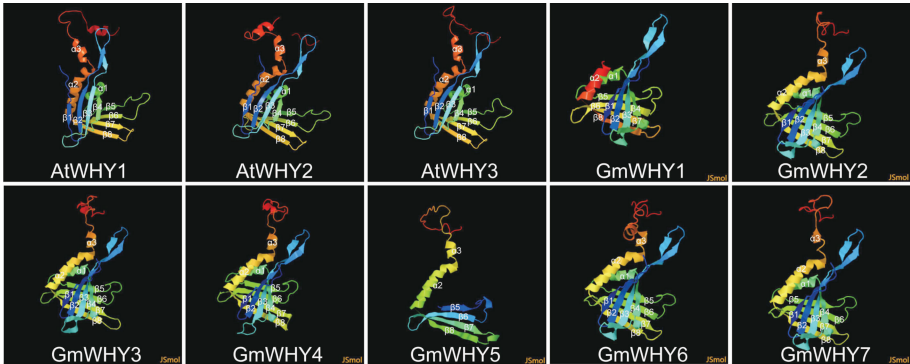


图 4 大豆和拟南芥 Whirly 蛋白的三维结构

Fig. 4 Three-dimensional structures of Whirly proteins from Arabidopsis and soybean

2.5 大豆 Whirly 基因的组织表达模式

基因表达模式可以为基因功能的研究提供重要的线索。对大豆 Whirly 基因在根、根毛、茎、茎顶端分生组织、叶片、结节、花、豆荚和种子的表达情况的研究结果图 5 所示:GmWHY1、GmWHY2、GmWHY6 和 GmWHY7 的表达水平明显高于 GmWHY3、GmWHY4 和 GmWHY5;GmWHY2、GmWHY3、GmWHY4

在叶片中的表达量最高,GmWHY1 在茎顶端、叶片和种子中特异地高表达,GmWHY5 在花和叶片中的表达量最高,GmWHY6 呈组成型表达,GmWHY7 在茎顶端、结节和种子中特异地高表达。大豆 Whirly 基因的不同组织表达模式暗示它们可能发挥不同的作用。

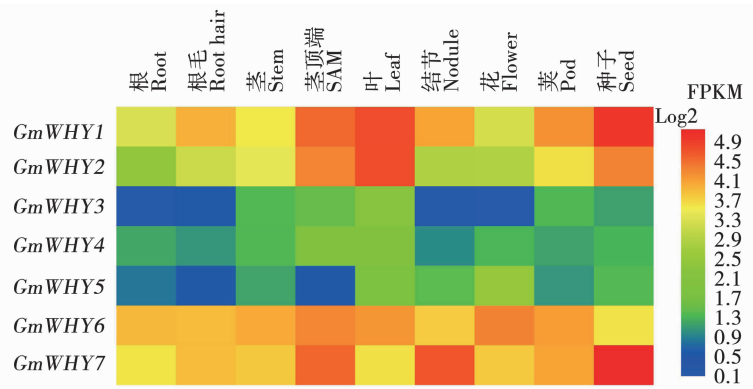


图 5 大豆 GmWHY 基因的组织表达模式

Fig. 5 The tissue expression patterns of GmWHY genes

2.6 大豆 Whirly 基因在干旱和盐胁迫条件下的表达模式

土壤干旱和盐碱化是限制植物生长和作物产量的两种最常见和最严重的非生物胁迫。为了探究大豆 Whirly 基因在干旱和盐胁迫下的潜在作用,本研究利用已发布的 RNA-seq 原始数据检测它们在干旱^[30]和盐 (NaCl)^[31]处理条件下的表达情况(图 6)。结果发现:在脱水处理的大豆幼苗根中,除了 GmWHY1、GmWHY3 和 GmWHY5 基因在某些时间

点略有上调(1.4 ~ 1.6 倍)以外,其它基因的表达基本不受干旱影响。这些结果暗示大豆 Whirly 基因可能不受干旱调控;在盐处理条件下,尤其是盐害早期 (NaCl 处理 1 h、2 h),大豆 Whirly 基因 (尤其 GmWHY1 ~ 5)在根中的表达水平明显下调。与之相反,大多数大豆 Whirly 基因 (尤其 GmWHY1)在盐害后期 (NaCl 处理 24 h、48 h)叶片中的表达量明显上调。这些结果表明大豆 Whirly 基因可能参与盐胁迫的调控。

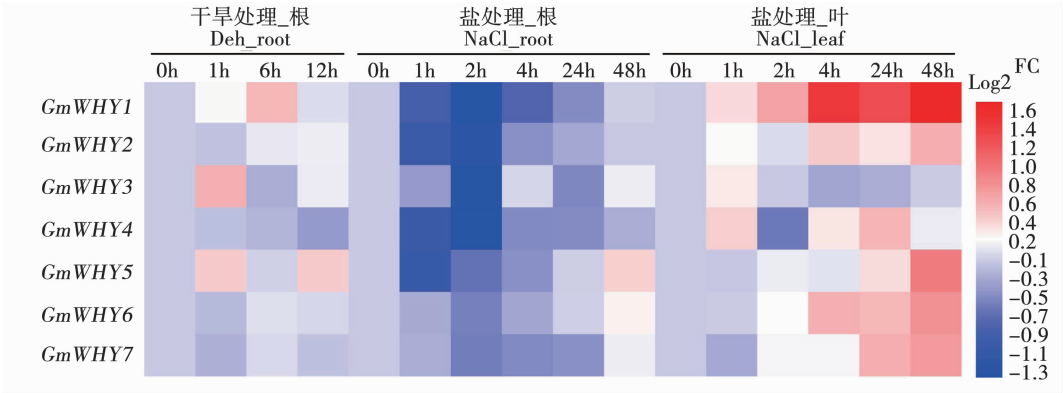


图 6 大豆 *GmWHY* 基因在干旱和盐胁迫条件下的表达模式

Fig. 6 The expression patterns of *GmWHY* genes against dehydration and salt stresses

2.7 大豆 Whirly 基因的启动子分析

基因的转录受与启动子上顺式元件相结合的 反式元件、辅因子和染色质状态的调控。对大豆和拟南芥 *Whirly* 基因的启动子序列 (ATG 上游1 500 bp) 的顺式作用元件的分析结果显示:在 *Whirly* 基因的启动子上存在多种顺式作用元件,包括激素、胁迫、光响应和发育相关的顺式作用元件,这暗示它们的基因表达受多种因素调控 (图 7)。这些 *Whirly* 基因的启动子上既有共有的顺式元件,也有各自特异的顺式元件。几乎所有的 *Whirly* 基因的

启动子上都包含顺式元件 ABRE、ARE、G-Box,尽管它们的数目并不完全相同。与之相反,与激素相关的顺式元件 AuxRR-core、TGA-box 分别只存在于 *AtWHY1* 和 *GmWHY7* 的启动子上;WUN-motif、ACE、I-box、chs-Unit 1 m1、AT1-motif、ATC-motif、Gap-box、CAT-box、HD-Zip 1、RY-element 和 MSA-like 等顺式元件也特异地存在于某一 *Whirly* 基因的启动子上。这些结果暗示这些 *Whirly* 基因的表达既受某些相同因素的调控,也特异地响应一些不同因素,从而导致功能的趋异化。

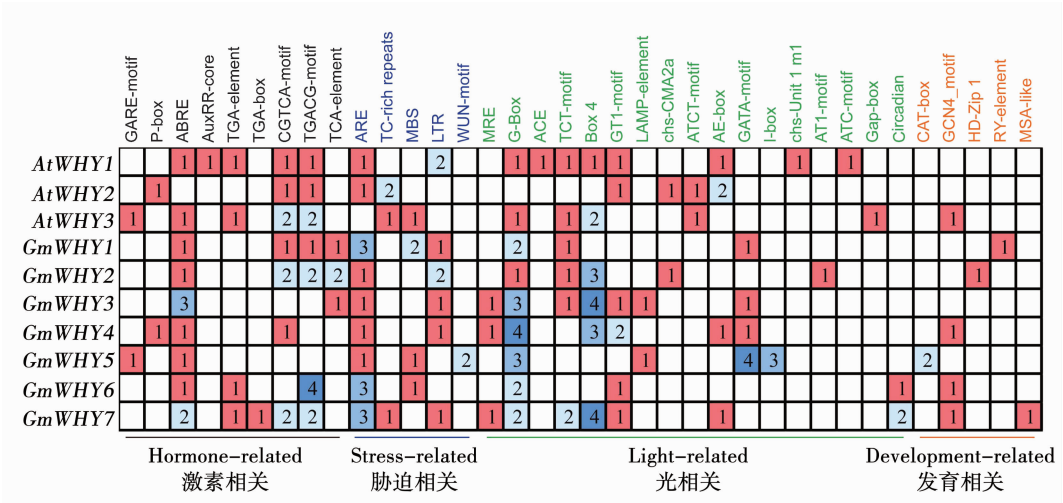


图 7 Whirly 基因启动子顺式作用元件统计

Fig. 7 The statistics of the cis-acting elements in each promoter region of *Whirly* genes

3 讨论

Whirly 蛋白是一类广泛存在于植物中的特异转录因子小家族,能与单链 DNA 分子结合,双定位于细胞核和细胞器(线粒体或叶绿体)中,在植物细胞核和细胞器中发挥多种功能,参与植物抗病信号转导、叶片衰老、角果发育、种子萌发、质体和线粒体基因组稳定等多个方面的调控^[2,37-38]。虽然这些基因的分子机制在模式植物拟南芥中得到了很好的阐明,但在其它植物中的调控机制仍不清楚。大豆

是一种古四倍体生物,经历了至少两轮的全基因组重复事件,导致大豆基因组中大约 75% 的基因存在多个拷贝^[39]。对不同物种的 *Whirly* 基因家族成员的比较结果表明大豆具有最多的 *Whirly* 基因数目,大约是其它二倍体被子植物的 2~3 倍。此外,根据 Soybase 数据库的基因注释信息可知 *GmWHY1* 和 *GmWHY2*,*GmWHY6* 和 *GmWHY7* 分别是一对旁系同源基因。基因复制通过亚功能化、新功能化、去功能化等途径为基因进化提供了重要的遗传原始材料^[40]。大豆相比拟南芥等其它物种,具有更多的

Whirly 基因家族成员,使得大豆 *Whirly* 基因的研究变得更为复杂。

比较拟南芥与大豆的直系同源有助于了解大豆同源基因的功能。综合分析和对比大豆与拟南芥 *Whirly* 基因家族成员的基因结构、蛋白相似性、蛋白保守基序、亚细胞定位、蛋白三维结构等,结果发现大豆 *Whirly* 家族成员分为两大类。*GmWHY1~4* 可能具有类似 *AtWHY1*、*AtWHY3* 的功能;*GmWHY5* 由于 N 端缺失,缺少 $\beta 1 \sim 4$ 和 $\alpha 1$,并且定位在细胞质中,因此丧失了 *Whirly* 蛋白的功能;*GmWHY6* 和 *GmWHY7* 可能与 *AtWHY2* 的功能相似。此外,通过比较大豆和拟南芥 *Whirly* 基因家族成员的基因表达水平发现大豆和拟南芥具有类似的 *Whirly* 基因表达模式。在拟南芥的大部分组织或器官中,*AtWHY1* 和 *AtWHY2* 的基因表达水平明显高于 *AtWHY3*(数据来源于 Tair 网站)。在大豆中,与 *AtWHY2* 同属于一个分类的 *GmWHY6*、*GmWHY7* 以及 *AtWHY1* 同源性更高的 *GmWHY1*、*GmWHY2* 的基因表达量明显高于 *GmWHY3~5*。大豆 *Whirly* 基因时空表达模式的不同可能影响它们的功能趋异化。利用过表达、反义表达或定点突变等潜在工具改变大豆 *Whirly* 基因或蛋白的表达水平,有助于进一步推断 *Whirly* 蛋白在大豆中的作用。

4 结 论

本研究在大豆基因组水平鉴定到了 7 个 *Whirly* 基因,并对其理化特征、亚细胞定位、系统进化关系、基因结构、蛋白基序组成、蛋白三维结构、组织和胁迫表达模式进行综合分析。本研究结果将为进一步研究大豆 *Whirly* 基因的功能提供非常有价值的线索。

参考文献

[1] Desveaux D, Marechal A, Brisson N. Whirly transcription factors: Defense gene regulation and beyond[J]. Trends in Plant Science, 2005, 10(2): 95-102.

[2] 孔凡英, 邓永胜, 周斌, 等. Whirly 转录因子研究进展[J]. 植物生理学报, 2012, 48(7): 643-653. (Kong F Y, Deng Y S, Zhou B, et al. Research advancement of Whirly transcription factors[J]. Plant Physiology Journal, 2012, 48(7): 643-653.)

[3] Desveaux D, Despres C, Joyeux A, et al. PBF-2 is a novel single-stranded DNA binding factor implicated in PR-10a gene activation in potato[J]. Plant Cell, 2000, 12(8): 1477-1489.

[4] Krause K, Kilbienski I, Mulisch M, et al. DNA-binding proteins of the Whirly family in *Arabidopsis thaliana* are targeted to the organelles[J]. FEBS Letters, 2005, 579(17): 3707-3712.

[5] 姚沁涛, 张文蔚, 刘莉, 等. Whirly 转录因子对非寄主菌诱导水稻 HR 反应的负调控作用[J]. 中国农业科技导报, 2008,

10(5): 53-58. (Yao Q T, Zhang W W, Liu L, et al. Negative regulation of rice Whirly transcription Factor for the hypersensitive response induced by non-host pathogen bacterium[J]. Journal of Agricultural Science and Technology, 2008, 10(5): 53-58.)

[6] Prikryl J, Watkins K P, Friso G, et al. A member of the Whirly family is a multifunctional RNA- and DNA-binding protein that is essential for chloroplast biogenesis[J]. Nucleic Acids Research, 2008, 36(16): 5152-5165.

[7] Zhang Y F, Hou M M, Tan B C. The requirement of WHIRLY1 for embryogenesis is dependent on genetic background in maize [J]. Plos One, 2013, 8(6): e67369.

[8] Grabowski E, Miao Y, Mulisch M, et al. Single-stranded DNA-binding protein Whirly1 in barley leaves is located in plastids and the nucleus of the same cell[J]. Plant Physiology, 2008, 147(4): 1800-1804.

[9] Zhao S Y, Wang G D, Zhao W Y, et al. Overexpression of tomato WHIRLY protein enhances tolerance to drought stress and resistance to *Pseudomonas solanacearum* in transgenic tobacco[J]. Biologia Plantarum, 2018, 62(1): 55-68.

[10] Yoo H H, Kwon C, Lee M M, et al. Single-stranded DNA binding factor AtWHY1 modulates telomere length homeostasis in *Arabidopsis*[J]. Plant Journal, 2007, 49(3): 442-451.

[11] Ren Y, Li Y, Jiang Y, et al. Phosphorylation of WHIRLY1 by CIPK14 shifts its localization and dual functions in *Arabidopsis* [J]. Molecular Plant, 2017, 10(5): 749-763.

[12] Cai Q, Guo L, Shen Z R, et al. Elevation of pollen mitochondrial DNA copy number by WHIRLY2: Altered respiration and pollen tube growth in *Arabidopsis*[J]. Plant Physiology, 2015, 169(1): 660-673.

[13] Dickey T H, Altschuler S E, Wuttke D S. Single-stranded DNA-binding proteins: Multiple domains for multiple functions [J]. Structure, 2013, 21(7): 1074-1084.

[14] Desveaux D, Subramaniam R, Despres C, et al. A 'whirly' transcription factor is required for salicylic acid-dependent disease resistance in *Arabidopsis* [J]. Developmental Cell, 2004, 6(2): 229-240.

[15] Krupinska K, Dahnhardt D, Fischer-Kilbienski I, et al. Identification of WHIRLY1 as a factor binding to the promoter of the stress- and senescence-associated gene HvS40 [J]. Journal of Plant Growth Regulation, 2014, 33(1): 91-105.

[16] Xiong J Y, Lai C X, Qu Z, et al. Recruitment of AtWHY1 and AtWHY3 by a distal element upstream of the kinesin gene AtKPI1 to mediate transcriptional repression[J]. Plant Molecular Biology, 2009, 71(4-5): 437-449.

[17] Miao Y, Jiang J, Ren Y, et al. The single-stranded DNA-binding protein WHIRLY1 represses WRKY53 expression and delays leaf senescence in a developmental stage-dependent manner in *Arabidopsis*[J]. Plant Physiology, 2013, 163(2): 746-756.

[18] Marechal A, Parent J S, Veronneau-Lafortune F, et al. Whirly proteins maintain plastid genome stability in *Arabidopsis* [J]. PNAS, 2009, 106(34): 14693-14698.

[19] Marechal A, Parent J S, Sabar M, et al. Overexpression of mtDNA-associated AtWhy2 compromises mitochondrial function [J]. BMC Plant Biology, 2008, 8(1): 1-15.

[20] Cappadocia L, Marechal A, Parent J S, et al. Crystal structures of

DNA-Whirly complexes and their role in *Arabidopsis* organelle genome repair[J]. *Plant Cell*, 2010, 22(6): 1849-1867.

[21] Janack B, Sosoi P, Krupinska K, et al. Knockdown of WHIRLY1 affects drought stress-induced leaf senescence and histone modifications of the senescence-associated gene HvS40 [J]. *Plants*, 2016, 5(3): 37.

[22] 黄晨星, 任育军, 李燕云, 等. 拟南芥 *WHIRLY2* 基因超表达引起拟南芥角果发育异常的分子细胞学分析[J]. *福建农林大学学报(自然科学版)*, 2016, 45(3): 277-281. (Huang C X, Ren Y J, Li Y Y, et al. Molecular and cytology analysis of *Arabidopsis* silique abnormal development induced by WHIRLY2 overexpression[J]. *Journal of Fujian Agricultural and Forestry University (Nature Science Edition)*, 2016, 45(3): 277-281.)

[23] Isemer R, Krause K, Grabe N, et al. Plastid located WHIRLY1 enhances the responsiveness of *Arabidopsis* seedlings toward abscisic acid[J]. *Frontiers in Plant Science*, 2012, 3(4): 283.

[24] Finn R D, Coggill P, Eberhardt R Y, et al. The Pfam protein families database: Towards a more sustainable future[J]. *Nucleic Acids Research*, 2016, 44(Database issue): D279-D285.

[25] Tamura K, Stecher G, Peterson D, et al. MEGA6: Molecular evolutionary genetics analysis version 6.0 [J]. *Molecular Biology and Evolution*, 2013, 30(12): 2725-2729.

[26] Hu B, Jin J, Guo A Y, et al. GSDS 2.0: An upgraded gene feature visualization server [J]. *Bioinformatics*, 2015, 31(8): 1296.

[27] Bailey T L, Boden M, Buske F A, et al. MEME SUITE: Tools for motif discovery and searching [J]. *Nucleic Acids Research*, 2009, 37(Web server issue): 202-208.

[28] Emanuelsson O, Brunak S, Von Heijne G, et al. Locating proteins in the cell using TargetP, SignalP and related tools[J]. *Nature Protocols*, 2007, 2(4): 953-971.

[29] Kelley L A, Mezulis S, Yates C M, et al. The Phyre2 web portal for protein modeling, prediction and analysis[J]. *Nature Protocols*, 2015, 10(6): 845-858.

[30] Belamkar V, Weeks N T, Bharti A K, et al. Comprehensive characterization and RNA-Seq profiling of the HD-Zip transcription factor family in soybean (*Glycine max*) during dehydration and salt stress[J]. *BMC Genomics*, 2014, 15: 950.

[31] Liu A, Xiao Z, Li M W, et al. Transcriptomic reprogramming in soybean seedlings under salt stress [J]. *Plant Cell & Environment*, 2018, 12(12): e0189159.

[32] Kim D, Langmead B, Salzberg S L. HISAT: A fast spliced aligner with low memory requirements [J]. *Nature Methods*, 2015, 12(4): 357-360.

[33] Pertea M, Pertea G M, Antonescu C M, et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads [J]. *Nature Biotechnology*, 2015, 33(3): 290-295.

[34] Deng W, Wang Y, Liu Z, et al. HemI: A toolkit for illustrating heatmaps[J]. *PLoS One*, 2014, 9(11): e111988.

[35] Xu G X, Guo C C, Shan H Y, et al. Divergence of duplicate genes in exon-intron structure [J]. *PNAS*, 2012, 109(4): 1187-1192.

[36] Desveaux D, Allard J, Brisson N, et al. A new family of plant transcription factors displays a novel ssDNA-binding surface[J]. *Nature Structural Biology*, 2002, 9(7): 512-517.

[37] 蔡倩, 任育军, 黄晨星, 等. 植物单链 DNA 结合蛋白 WHIRLY2 研究进展[J]. *福建农林大学学报(自然科学版)*, 2018, 47(3): 267-273. (Cai Q, Ren Y J, Huang C X, et al. Research progress of single-stranded DNA binding protein WHIRLY2 in plants[J]. *Journal of Fujian Agricultural and Forestry University (Nature Science Edition)*, 2018, 47(3): 267-273.)

[38] 林文芳, 任育军, 缪颖. 植物 Whirly 蛋白调控叶片衰老的研究进展[J]. *植物生理学报*, 2014, 50(9): 1274-1284. (Lin W F, Ren Y J, Miao Y. Research progress of Whirly proteins in regulation of leaf senescence[J]. *Plant Physiology Journal*, 2014, 50(9): 1274-1284.)

[39] Schmutz J, Cannon S B, Schlueter J, et al. Genome sequence of the palaeopolyploid soybean [J]. *Nature*, 2010, 463(7278): 178-183.

[40] Flagel L E, Wendel J F. Gene duplication and evolutionary novelty in plants[J]. *New Phytologist*, 2009, 183(3): 557-564.