

盾叶薯蓣 WRKY 转录因子的鉴定和生物信息学分析

王丽娜¹, 王丰青^{2*}, 智惊宇², 张 苗¹, 张重义³, 谢彩侠^{1*}

1. 河南中医药大学药学院, 河南 郑州 450046

2. 河南农业大学农学院, 河南 郑州 450002

3. 福建农林大学中药材 GAP 研究所, 福建 福州 350002

摘 要: 目的 鉴定盾叶薯蓣 WRKY (DzWRKY) 转录因子家族基因, 分析 DzWRKY 蛋白的序列特征, 检测 DzWRKY 在根和叶中的表达量。方法 以拟南芥 WRKY 基因的 cDNA 为查询序列, 应用 BLASTn 进行比对, 应用生物信息学方法进行全长开放阅读框 (ORF) 预测和蛋白序列特征分析, 根据转录组数据检测 DzWRKY 的表达量。结果 共检测到 27 个具有全长 ORF 的 DzWRKY 家族基因, 其中 6 个基因编码的蛋白具有 2 个 WRKY 结构域。DzWRKY 蛋白均为亲水性蛋白, 亚细胞预测为核蛋白, WRKY 结构域序列保守程度高。进化分析均把 27 个 DzWRKY 蛋白和 19 个 AtWRKY 蛋白分为 3 个大类群。27 个 DzWRKY 在叶片中的表达量均比根茎低, 高表达的 DzWRKY 基因主要集中在第 I 类和第 IIId 亚类。盾叶薯蓣 WRKY 与已知功能的 WRKY 序列相似性较低。结论 首次从盾叶薯蓣中鉴定出 DzWRKY 基因, 为进一步阐明 DzWRKY 在盾叶薯蓣生长发育和药效成分代谢中的作用奠定基础。

关键词: 盾叶薯蓣; WRKY; 生物信息学; 转录因子; 蛋白序列特征分析

中图分类号: R282.12

文献标志码: A

文章编号: 0253-2670(2016)22-4062-10

DOI: 10.7501/j.issn.0253-2670.2016.22.022

Identification and bioinformatic analysis on rhizomes of *Dioscorea zingiberensis* WRKY transcription factor

WANG Li-na¹, WANG Feng-qing², ZHI Jing-yu², ZHANG Miao¹, ZHANG Zhong-yi³, XIE Cai-xia¹

1. School of Pharmacy, Henan University of Traditional Chinese Medicine, Zhengzhou 450046, China

2. College of Agronomy, Henan Agricultural University, Zhengzhou 450002, China

3. Institute of Chinese Crude Drugs GAP, Fujian Agriculture and Forestry University, Fuzhou 350002, China

Abstract: Objective To identify WRKY genes from the rhizomes of *Dioscorea zingiberensis* and analyze the protein characteristics and expression level of these genes. **Methods** The transcriptional EST database of the rhizomes of *D. zingiberensis* was used to search the analogs of *AtWRKY* genes by BLASTn, the full-length open reading frames (ORF) of DzWRKY and its protein characteristics were studied using bioinformatic method, and the expression levels of DzWRKY genes in rhizomes and leaves were detected from transcriptional data of the rhizomes of *D. zingiberensis*. **Results** Twenty-seven DzWRKY transcription factors family genes with full length ORF were isolated from the rhizomes of *D. zingiberensis*, and two WRKY domains were confirmed in six WRKY genes. All the DzWRKY proteins were predicted as hydrophilic proteins and nucleoproteins with highly conserved WRKY domains. The 27 DzWRKY and 19 AtWRKY proteins were divided into three groups by phylogenetic analysis. All DzWRKY genes showed higher expression level in the leaves compared to the rhizomes, and highly expressed genes were mainly in groups I and IIId. DzWRKY proteins exhibited lower sequence identity. **Conclusion** The DzWRKY genes are successfully isolated for the first time, which would provide a reference for the study on the roles of WRKY in development and active components biosynthesis of the rhizomes of *D. zingiberensis*.

Key words: *Dioscorea zingiberensis* C. H. Wright; WRKY; bioinformatics; transcription factor; analysis of protein sequence characteristics

WRKY 是一类植物特有的转录因子家族, 包含有一段 WRKYGQY 的核心序列而得名^[1]。而且, 1~2 个约 60 个氨基酸残基的保守结构域, 因其含 WRKY 结构域的 C 端还有一个 C₂H₂ 或 C₂HC 的锌

收稿日期: 2016-04-19

基金项目: 国家自然科学基金资助项目 (81473299)

作者简介: 王丽娜, 硕士研究生。Tel: 15136257132 E-mail: 273420473@qq.com

*通信作者 王丰青, 副教授, 主要从事中药资源可持续利用研究。Tel: (0371)63558122 E-mail: heauzyxcw@126.com

谢彩侠, 教授, 研究方向为中药材质量评价。E-mail: nanyangxcx@126.com

指结构, 根据其锌指结构特征可分为 III 类, 即 I 类含有 2 个 WRKY 结构域, 锌指序列分别为 C-X₄-C-X₂₂-H-X₁-H 和 C-X₅-C-X₂₃-H-X₁-H; II 类含 1 个 WRKY 结构域, 锌指序列为 C-X_{4,5}-C-X₂₃-H-X₁-H; III 类也是 1 个 WRKY 结构域, 锌指序列为 C-X₇-C-X₂₃-H-X₁-C^[2]。WRKY 结构域能够特异地绑定到下游靶基因启动子上的 W-box[(C/T)TGAC(T/C)] 序列, 调控靶基因的转录^[3]。已有的研究表明, WRKY 不但参与植物的器官发育^[4-5]和逆境胁迫响应^[6-7], 而且在药用植物药效成分生物合成中发挥调控作用。CjWRKY1 RNA 干扰可明显抑制黄连 *Coptis japonica* Makino 参与小檗碱生物合成途径多个关键酶基因的表达量^[8], 过表达 CrWRKY1 能促进长春花 *Catharanthus roseus* (L.) G. Don 蛇根碱的合成^[9], 过表达 AaWRKY1 可显著提高黄花蒿 *Artemisia annua* L. 青蒿素的量^[10], TcWRKY1 以 DBAT 为靶基因参与中国红豆杉 *Taxus chinensis* (Rehder & E. H. Wilson) Rehder 紫杉醇的生物合成^[11]。

盾叶薯蓣是薯蓣科 (Dioscoreaceae) 植物盾叶薯蓣 *Dioscorea zingiberensis* C. H. Wright 的根茎, 含有丰富的甾体皂苷类成分, 是合成多种甾体激素和避孕类药的重要原料^[12]。盾叶薯蓣是我国特有种, 为世界上最好的甾体激素类药源植物之一^[13]。然而, 迄今为止关于甾体皂苷类成分生物合成的分子机制研究很少, 已有的报道也局限于单个酶基因的同源克隆, 限制了盾叶薯蓣的进一步开发利用。本研究利用课题组已有的盾叶薯蓣转录组数据, 鉴定含有全长开放阅读框 (ORF) 的 WRKY 基因, 通过生物信息学方法分析蛋白序列特征及表达特性, 为进一步研究其在甾体皂苷类成分生物合成的分子功能提供依据。

1 材料与方法

1.1 材料

研究用于鉴定 WRKY 的 cDNA 序列来源于课题组构建的转录组数据库。用于转录组测序的实验材料为盾叶薯蓣 *Dioscorea zingiberensis* C. H. Wright 1 年生根茎和叶片, 采用 Illumina Hiseq 2000 平台进行 de novo 测序, 序列拼接软件为 SOAPdenovo。

1.2 盾叶薯蓣 WRKY 基因的鉴定

以水稻 *Oryza sativa* L.、玉米 *Zea mays* L.、拟南芥 *Arabidopsis thaliana* (L.) Heynh. 等物种的 WRKY CDS 为查询序列, 以本地序列比对软件 BLASTn 进行同源搜索, E-value 设定为 0.01, 以

DNASar 软件去除冗余序列。开放阅读框以 ORF Finder (<http://www.ncbi.nlm.nih.gov/gorf/orfig.cgi>) 软件进行预测, 以 DNASar 软件预测编码的氨基酸序列。最后, 为了确认获得的序列是 WRKY 基因, 所有的非冗余氨基酸序列提交到 <http://pfam.sanger.ac.uk> 网站进行 WRKY 结构域预测, 仅保留具有 DBD 结构域的序列。

1.3 盾叶薯蓣 WRKY 蛋白序列特征分析

WRKY 一级蛋白的理化特征以在线软件 ProtParam (<http://web.expasy.org/protparam/>) 进行分析, 二级结构特征以在线软件 SOPMA (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html) 进行预测。WRKY 蛋白的亚细胞定位以在线软件 PSORT (<http://www.genscript.com/psort.html>) 进行预测, 信号肽以在线软件 seqNLS (<http://mleg.cse.sc.edu/seqNLS/>) 进行预测。

应用在线软件 MEME version 4.11.1 (<http://meme-suite.org/tools/meme>) 对盾叶薯蓣 WRKY 蛋白进行保守结构域预测, 设定 Motifs 最大值为 3, 宽度最小值为 6、最大值为 100。应用结构域在线分析软件 SMART (<http://smart.embl.de/>) 对保守结构域进行验证和功能注释。WRKY 保守结构域的氨基酸多序列联配用在线软件 Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>) 进行比对, GeneDoc 软件进行读取分析。三级结构域以在线软件 SWISS-MODEL (<http://swissmodel.expasy.org/>) 进行三维建模。

1.4 进化分析

拟南芥 WRKY 的蛋白序列从 NCBI 数据库中下载, 其完整的氨基酸序列用于进化分析。进化树构建分别利用在线分析软件 MAFFT (<http://www.ebi.ac.uk/Tools/msa/mafft/>) 和 Clustal Omega 进行多序列联配, 以离线软件 MEGA6.06 的 Neighbor-Joining 算法进行 bootstrapping 分析 (1 000 次重复)。

1.5 表达分析

盾叶薯蓣 WRKY 基因在叶和根茎中的表达量以 Reads Per Kb per Million reads (RPKM) 进行计算。

$$RPKM = \frac{10^6 C}{NL / 10^3}$$

RPKM (A) 为基因 A 的表达量, 则 C 为唯一比对到基因 A 的 reads 数, N 为唯一比对到参考基因的总 reads 数, L 为基因 A 的碱基数

1.6 同源性分析

以 27 个 DzWRKY 的氨基酸序列为查询序列, 利用 NCBI 在线软件 BLAST (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) 在拟南芥蛋白质数据库中进行搜索, 分别获得相似性最高的 AtWRKY, 同时在 PubMed 中检索文献获取 AtWRKY 的分子功能。根据已经报道的药用植物药效成分相关的 WRKY 在 NCBI 上的注册号, 获取其氨基酸序列, 应用软件 DNAMAN 进行序列相似性比对。

2 结果与分析

2.1 盾叶薯蓣 WRKY 基因家族序列

应用 BLASTn 进行同源比对, 结果从盾叶薯蓣转录组数据库中共筛选出 109 条 Unigene 序列, 片段大小为 204~3 574 bp。通过 ORF 预测, 发现 27 条序列具有完整的 ORF (表 1), 片段大小从 810 bp (CL5012.Contig2) ~ 3 230 bp (CL2497.Contig8), 编码的氨基酸从 174 aa (Unigene15016) 到 724 aa (CL2497.Contig8)。27 个基因均包含非编码区, 其 5' UTR 为 39~675 bp, 3' UTR 为 32~408 bp。

2.2 盾叶薯蓣 WRKY 家族蛋白的结构特征

DzWRKY 蛋白的理化特性分析结果 (表 1) 表明, 27 个蛋白的平均相对分子质量为 40 100, 最小值为 19 600 (Unigene15016), 最大值为 78 100 (CL2497.Contig8)。等电点 pI 平均值为 7.18, 最小值为 CL1835.Contig4 的 5.18, 最大值为 CL5012.Contig2 的 10.17, 其中有 16 个蛋白 pI 小于 7, 说明多数蛋白表现偏酸性。27 个 WRKY 蛋白的不稳定系数均大于 40, 最小的为 41.63 (CL4098.Contig1), 最大的为 66.95 (Unigene25512 和 CL697.Contig2), 说明盾叶薯蓣 WRKY 家族蛋白均为不稳定蛋白。WRKY 家族蛋白的脂溶性指数 (AI) 介于 47.01~72.79, 具有较低的 AI, 说明其不同环境中稳定性较差。DzWRKY 蛋白的亲水性平均系数 GRAVY 均小于 0, 说明 27 个蛋白均属于亲水蛋白。

DzWRKY 蛋白的二级结构预测结果 (表 1) 表明, 27 个 DzWRKY 蛋白主要由 α 螺旋、延伸链和无规卷曲组成, β 转角的比例均较小。无规卷曲所占比例较大, 18 个蛋白的无规卷曲比例超过 50%, 仅有 1 个蛋白 (CL7347.Contig1) 比例低于 40%。22 个 DzWRKY 蛋白表现为无规卷曲 > α 螺旋 > 延伸链 > β 转角, 其余 5 个蛋白 CL2497.Contig1、CL2497.Contig2、CL4098.Contig1、CL4098.Contig3、

Unigene15016 表现为无规则卷曲 > 延伸链 > α 螺旋 > β 转角。

PSORT 预测 DzWRKY 蛋白的亚细胞定位结果 (表 1) 表明, 27 个蛋白可能的分布主要有细胞核、细胞质、氧化物酶体、线粒体、细胞骨架等, 所有 DzWRKY 蛋白分布在细胞核中的概率均超过 65%, 其中 Unigene34566 和 CL4098.Contig1 分布在细胞核中的概率超过 90%, 说明 DzWRKY 蛋白多数应为核蛋白。

2.3 盾叶薯蓣 WRKY 家族蛋白的结构域

具有 WRKY 结构域为 WRKY 蛋白最典型的特征。为了分析盾叶薯蓣 WRKY 的保守结构域, 以 MEME 4.9.1 在线软件预测了假定的 Motif, 结果显示, 检测发现了 3 个 Motif (图 1-A)。分析 3 个 motif 在 DzWRKY 蛋白上的分布规律发现, 26 个蛋白具有 Motif 1 和 Motif 2, 此外 CL4098.Contig1、CL4098.Contig3、CL2497.Contig1、CL2497.Contig8、Unigene1393、Unigene14939 等 6 个蛋白还具有 1 个 Motif 3。只有 1 个蛋白 CL5012.Contig2 没有检测到 Motif 分布。进一步分析保守的 motif 序列特征发现 (图 1-B), Motif 1 与 Motif 2 合在一起为典型 WRKY 蛋白的保守结构域, 含有一段高度保守的核心序列 WRKYGQK, C' 端还有一个典型的 C_2H_2 或 C_2HC 锌指结构。Motif 3 为一段保守的 WRKY 结构域, 具有核心序列 WRKYGQK 和一段 C_2H_2 锌指结构。说明 CL4098.Contig1、CL4098.Contig3、CL2497.Contig1、CL2497.Contig8、Unigene1393、Unigene14939 等 6 个蛋白具有 2 个 WRKY 保守结构域, 属于第 I 类 WRKY 蛋白^[2]。

为了分析 DzWRKY 蛋白 N' 端 WRKY 结构域的序列特征, 以 Clustal Omega 对 CL4098.Contig1、CL4098.Contig3、CL2497.Contig1、CL2497.Contig8、Unigene1393、Unigene14939 等 6 条氨基酸序列进行多序列联配, 结果 (图 1-C) 表明, 6 个蛋白 N' 端 WRKY 结构域序列保守程度很高, 核心序列 WRKYGQK, 锌指结构为 C_2H_2 型, 序列为 C-X₄-C-X₂₂-H-X₁-H。进而以 Clustal Omega 所有 DzWRKY 蛋白的氨基酸序列进行多序列联配, 结果 (图 1-D) 表明, 27 个 DzWRKY 均包含核心序列 WRKYGQK, CL5012.Contig2 缺失 C' 末端的锌指结构域。CL7347.Contig1、Unigene34566、Unigene36980 的锌指结构为 C_2HC 型, 序列为 C-X₆₋₇-C-X₂₃₋₂₄-H-X₁-C,

表 1 盾叶薯蓣 WRKY 基因家族成员信息
Table 1 Information of WRKY genes in rhizomes of *D. zingiberensis*

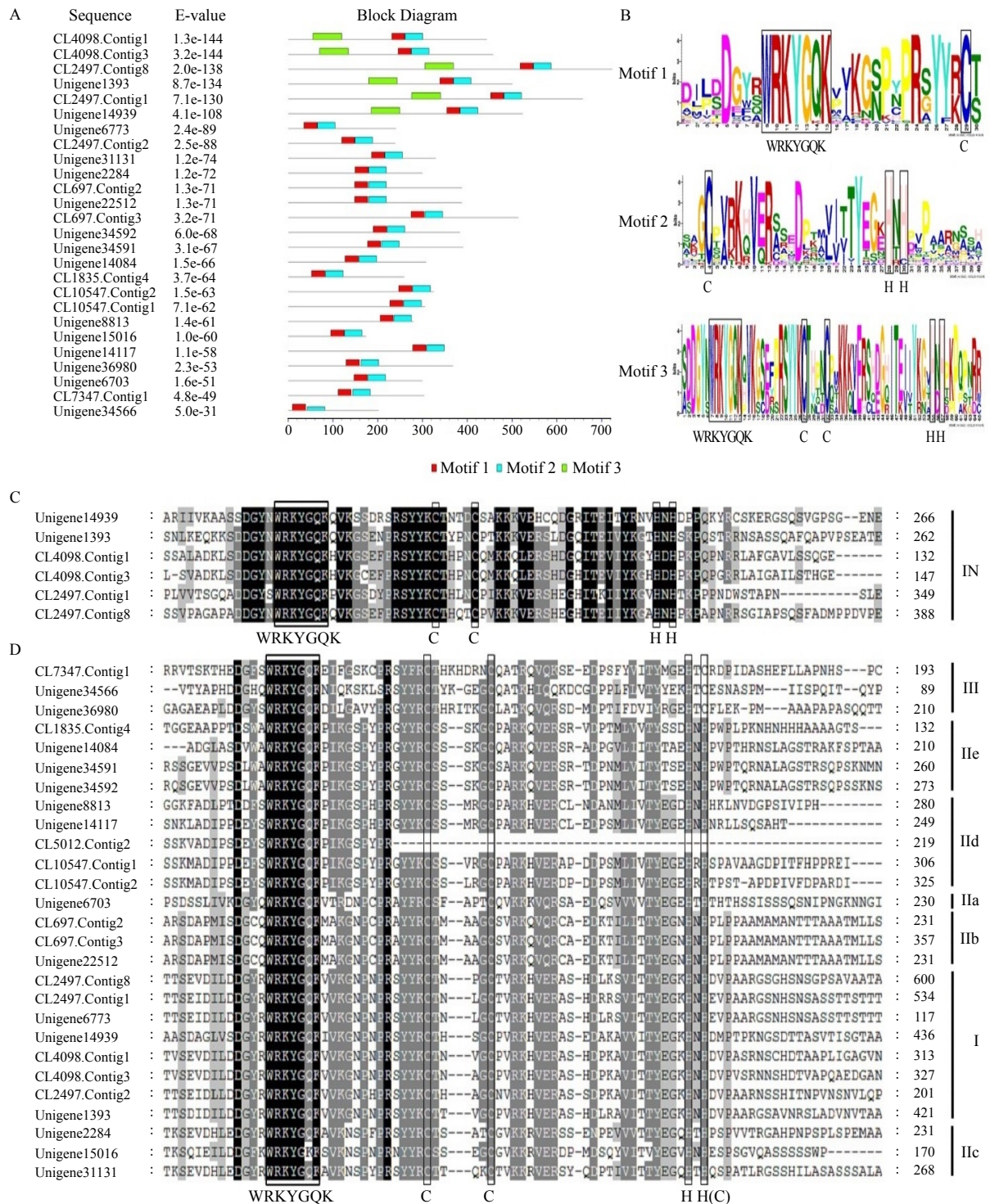
基因 ID	长度/bp	氨基酸数	理化参数					二级结构/%				亚细胞定位					
			相对分子质量	理论 pI	分子式	不稳定指数	AI	GRAVY	α螺旋	延伸链	β转角	无规卷曲	细胞核	细胞质	过氧化物酶	线粒体	细胞骨架
CL10547.Contig1	1 292	306	32 984.4	9.71	C ₁₄₃₅ H ₂₃₆₅ N ₄₂₅ O ₄₄₁ S ₁₃	54.20	64.12	-0.540	23.86	16.01	9.15	50.98	73.9	8.7		17.4	
CL10547.Contig2	1 242	325	35 079.1	9.54	C ₁₅₃₂ H ₂₄₆₅ N ₄₄₁ O ₄₆₇ S ₁₇	45.12	68.74	-0.473	19.69	19.08	6.77	54.46	69.6	13.0	4.3	13.0	
CL1835.Contig4	1 236	259	27 706.3	5.18	C ₁₁₉₈ H ₁₈₄₄ N ₃₄₄ O ₃₉₉ S ₈	57.67	48.65	-0.784	20.85	17.37	7.34	54.44	73.9	13.0	4.3	4.3	4.3
CL2497.Contig1	2 497	658	72 947.3	6.07	C ₃₁₅₁ H ₄₉₆₆ N ₉₀₈ O ₁₀₂₆ S ₃₀	49.05	63.39	-0.735	18.54	21.12	8.81	51.52	73.9	8.7		17.4	
CL2497.Contig2	1 091	239	26 465.1	6.15	C ₁₁₁₆ H ₁₇₉₂ N ₃₃₀ O ₃₈₁ S ₈	48.79	60.33	-1.000	14.64	20.92	8.37	56.07	82.6	8.7			4.3
CL2497.Contig8	3 230	724	78 111.1	5.67	C ₃₃₆₆ H ₅₂₉₉ N ₉₈₅ O ₁₁₀₆ S ₂₉	54.99	54.75	-0.736	20.58	15.61	7.18	56.63	78.3	13.0	4.3		4.3
CL4098.Contig1	1 792	444	49 095.9	6.12	C ₂₀₉₆ H ₃₂₇₄ N ₆₃₆ O ₆₉₆ S ₁₈	41.63	56.82	-0.922	12.39	22.07	8.33	57.21	91.3	4.3			4.3
CL4098.Contig3	1 800	458	500 58.1	6.55	C ₂₁₅₀ H ₃₃₇₇ N ₆₅₉ O ₇₀₅ S ₁₁	44.11	63.62	-0.870	19.21	23.14	10.26	47.38	65.2	13.0		17.4	4.3
CL5012.Contig2	810	219	23 529.8	10.17	C ₁₀₂₁ H ₁₆₇₃ N ₃₀₇ O ₃₁₄ S ₈	44.25	71.78	-0.454	23.74	14.16	8.22	53.88	73.9	8.7		17.4	
CL697.Contig2	1 716	388	41 983.5	8.36	C ₁₈₀₆ H ₂₈₈₀ N ₅₁₆ O ₅₇₉ S ₂₈	66.95	56.70	-0.572	32.47	11.60	4.90	51.03	69.6	13.0	13.0	4.3	
CL697.Contig3	1 704	514	55 581.4	7.09	C ₂₃₇₁ H ₃₇₈₉ N ₆₉₇ O ₇₇₉ S ₃₄	63.81	56.28	-0.688	32.10	11.48	4.28	52.14	73.9	13.0		13.0	
CL7347.Contig1	1 213	304	34 033.2	5.55	C ₁₄₇₇ H ₂₃₇₁ N ₄₁₅ O ₄₇₃ S ₁₈	64.40	63.85	-0.596	39.14	14.47	6.91	39.47	78.3	13.0	4.3		4.3
Unigene1393	1 769	501	55 470.3	6.77	C ₂₄₁₃ H ₃₇₃₅ N ₆₉₇ O ₇₈₀ S ₁₅	62.79	53.37	-0.864	25.95	15.97	6.59	51.50	78.3	13.0		4.3	4.3
Unigene14084	1 090	308	33 170.9	5.37	C ₁₄₃₄ H ₂₂₃₈ N ₄₁₂ O ₄₆₅ S ₁₅	59.80	57.69	-0.528	31.82	10.39	8.12	49.68	73.9	17.4	4.3	4.3	
Unigene14117	1 646	349	39 226.4	9.79	C ₁₆₈₈ H ₂₇₁₀ N ₅₃₂ O ₅₁₀ S ₁₉	63.14	64.81	-0.783	33.52	14.61	6.59	45.27	73.9	17.4	4.3	4.3	
Unigene15016	831	174	19 629.0	8.96	C ₈₅₀ H ₁₃₁₃ N ₂₄₁ O ₂₆₉ S ₁₃	48.95	47.01	-0.740	12.64	24.14	11.49	51.72	65.2	13.0		17.4	4.3
Unigene2284	1 176	300	32 849.6	6.09	C ₁₄₃₄ H ₂₂₄₃ N ₄₀₇ O ₄₅₉ S ₁₀	53.55	57.93	-0.743	25.00	15.00	7.67	52.33	78.3	13.0	4.3	4.3	
Unigene25512	1 863	388	41 983.5	8.36	C ₁₈₀₆ H ₂₈₈₀ N ₅₁₆ O ₅₇₉ S ₂₈	66.95	56.70	-0.572	32.47	11.60	4.90	51.03	69.6	13.0	13.0	4.3	
Unigene31131	1 160	330	36 145.9	6.09	C ₁₅₆₉ H ₂₄₃₈ N ₄₃₂ O ₅₁₂ S ₁₀	53.28	57.88	-0.758	30.61	16.67	11.52	41.21	78.3	13.0	4.3	4.3	
Unigene34566	908	202	22 994.8	6.42	C ₁₀₀₆ H ₁₅₄₇ N ₂₆₅ O ₃₁₉ S ₁₅	57.62	57.43	-0.650	25.74	19.80	7.92	46.53	95.7				4.3
Unigene34591	1 747	391	43 426.8	5.53	C ₁₈₈₂ H ₂₉₅₈ N ₅₃₈ O ₅₉₈ S ₂₃	49.67	60.41	-0.709	31.71	10.23	7.67	50.38	87.0	4.3	4.3	4.3	
Unigene34592	1 302	384	42 036.5	5.21	C ₁₈₂₁ H ₂₈₃₈ N ₅₁₈ O ₆₀₀ S ₁₄	61.69	61.51	-0.685	30.21	12.24	6.51	51.04	78.3	8.7		8.7	4.3
Unigene36980	1 364	368	41 080.6	5.20	C ₁₇₉₄ H ₂₇₆₉ N ₄₉₅ O ₅₈₁ S ₁₆	47.78	58.91	-0.615	26.36	16.58	6.52	50.54	69.6	17.4	4.3	4.3	4.3
Unigene6703	1 252	300	33 607.7	9.04	C ₁₄₃₈ H ₂₃₃₄ N ₄₂₆ O ₄₇₅ S ₁₃	49.59	64.30	-0.850	32.33	19.33	8.33	40.00	78.3	13.0	4.3	4.3	
Unigene6773	1 062	241	26 864.4	9.66	C ₁₁₅₁ H ₁₈₄₂ N ₃₅₆ O ₃₅₇ S ₁₆	45.89	63.07	-0.661	24.90	20.33	8.30	46.47	69.6	8.7	4.3	17.4	
Unigene8813	1 260	280	31 522.2	9.77	C ₁₃₆₉ H ₂₁₉₄ N ₄₁₆ O ₃₈₈ S ₂₁	43.96	72.79	-0.486	33.57	17.86	6.43	42.14	82.6	8.7	4.3	4.3	
Unigene14939	1 956	524	55 733.6	5.42	C ₂₃₉₂ H ₃₈₀₄ N ₆₉₀ O ₈₁₂ S ₁₆	50.45	61.83	-0.702	18.89	17.37	3.82	59.92	69.6	8.7			8.7

应为第 III 类 WRKY 蛋白。其余 23 个 DzWRKY 的锌指结构为 C₂H₂ 型，应为第 I 类和第 II 类 WRKY 蛋白。由于第 I 类 C'端 WRKY 蛋白和第 II 类 c 亚类 WRKY 蛋白的锌指结构序列均为 C-X₄-C-X₂₃-H-X₁-H，第 II 类 a、b、d、e 亚类的 WRKY 蛋白的锌指结构序列均为 C-X₅-C-X₂₃-H-X₁-H，应结合进化分析进一步区分。

2.4 盾叶薯蓣 WRKY 家族蛋白的三维结构

为了进一步了解盾叶薯蓣 WRKY 蛋白 N'端和 C'端 WRKY 结构域的三维结构，通过 Swiss-Model (<http://swissmodel.expasy.org>) 对盾叶薯蓣的 27 条氨基酸序列进行三维结构同源建

模，结果表明它们绝大多数具有十分相似的三维结构。以含有 N'端 WRKY 结构域的 CL2497.Contig8 N'端 580 aa 序列进行同源建模，结果（图 2）发现 CL2497.Contig8-N 包含 4 个 β 折叠，其中 β1 主要由核心序列 WRKYGQK 构成，CL2497.Contig8-N 与 AtWRKY1-C（2ayd.1.A）、AtWRKY4-C（2lex.1.A、1wj2.1.A）的序列一致性分别为 59.32%和 58.33%。以含有 C'端 WRKY 结构域的 CL2497.Contig8 C'端 240 aa 序列进行同源建模，结果发现 CL2497.Contig8-N 包含 5 个 β 折叠，其中 β2 主要由核心序列 WRKYGQK 构成，CL2497.Contig8-C 与 AtWRKY1-C、AtWRKY4-C



A-WRKY 蛋白保守结构域分布图 B-保守的结构域基序 C-N'端 WRKY 结构域氨基酸序列比对 D-C'端 WRKY 结构域氨基酸序列比对；一致序列显示黑色背景，高度保守的序列显示深灰色背景，类似的序列显示浅灰色背景

A-schematic representation of conserved motifs in WRKY proteins of *D. zingiberensis* B-conserved WRKY motifs of DzWRKY proteins C-multiple alignment of N-terminal WRKY domains among group I DzWRKY protein sequences D-multiple alignment of 27 C-terminal WRKY domains sequences; Highly conserved amino acid residues in each repeat are highlighted in dark gray

图 1 盾叶薯蓣 WRKY 蛋白结构域分析

Fig. 1 Analysis on WRKY domain sequences in rhizomes of *D. zingiberensis*

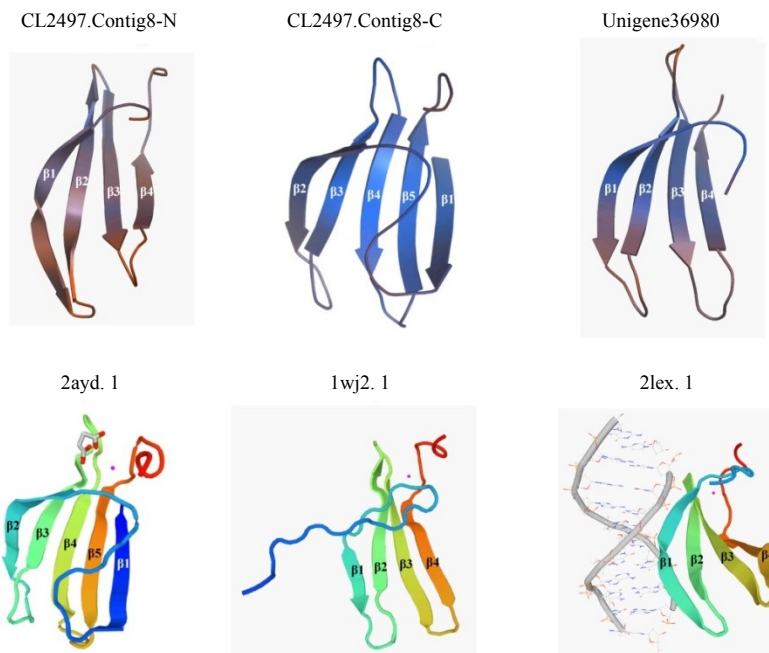


图 2 盾叶薯蓣和拟南芥 WRKY 结构域的三维结构图

Fig. 2 Ribbon representation of rhizomes of *D. zingiberensis* and *A. thaliana* WRKY domain

的序列一致性分别为 69.86%、85.92%。另外还选取第 III 类的一个蛋白 Unigene36980 进行同源建模,发现与 CL2497.Contig8-N 的三维结构类似,具有 4 个 β 折叠, Unigene36980 与 AtWRKY1-C、AtWRKY4-C 的序列一致性分别为 50.00%、55.36%。

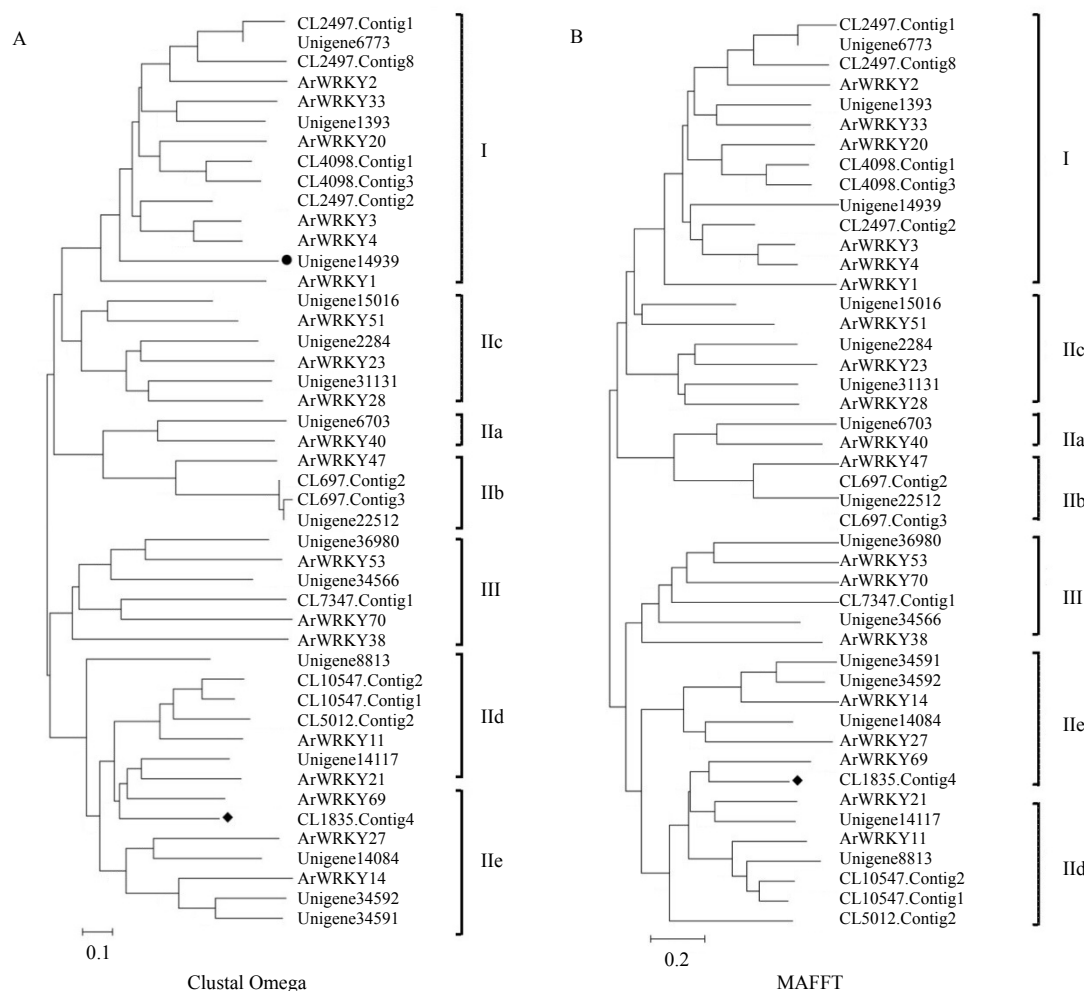
2.5 盾叶薯蓣 WRKY 蛋白的进化分析

为进一步分析盾叶薯蓣 WRKY 蛋白之间的进化关系,分别以 MAFFT 和 Clustal Omega 软件对 27 个 DzWRKY 及与之序列相似性较高的 19 个拟南芥 AtWRKY 蛋白的氨基酸序列进行多序列联配,再以 MEGA 6.06 软件邻接法构建系统进化树,结果表明,2 种联配方式获得的进化树类似,都能把 46 个蛋白分为 3 个大类群(图 3)。第 I 类群除包括 CL2497.Contig1 等含有 2 个 WRKY 结构域的 6 个盾叶薯蓣 WRKY 蛋白,此外此类群还包括 CL2497.Contig2 和 Unigene6773,与 6 个 I 类 AtWRKY 蛋白聚在一起,锌指结构序列为 C-X₄-C-X₂₃-H-X₁-H。第 II 类群包括 18 个 DzWRKY 蛋白,其中 1 个 IIa、3 个 IIb、3 个 IIc、5 个 IId、4 个 IIe DzWRKY 蛋白,分别与拟南芥同亚类群 WRKY 蛋白聚在一起,锌指结构序列为 C-X_{4.5}-C-X₂₃-H-X₁-H。第 III 类群包括 CL7347.Contig1、Unigene34566、Unigene36980 3

个 DzWRKY 和 3 个 III 类 AtWRKY 蛋白,锌指结构序列为 C-X_{6.7}-C-X₂₃₋₂₄-H-X₁-C。2 种联配方式获得的进化树也有一些差异,如 Unigene14939 在图 3-A 中单独为 1 分支,在图 3-B 中与 AtWRKY3、AtWRKY4、CL2497.Contig2 为同一分支。CL1385.Contig4 与 AtWRKY69 序列相似程度最高,在图 3B 中聚在同一个小分支,在图 3-A 中单独为 1 分支。相对而言,以 MAFFT 进行多序列联配构建的系统进化树较好。

2.6 盾叶薯蓣 WRKY 基因的表达特性

薯蓣皂苷(元)主要在根茎中积累,约莫皂苷(元)则主要在叶片中分布^[14]。应用盾叶薯蓣根茎、叶的转录组数据,分析了 27 个 DzWRKY 基因在根茎和叶中的表达量,结果发现不同 DzWRKY 基因的表达量有较大差异,而且所有基因在叶片中的表达量均较根茎低(图 4)。在叶中,只有 I 类的 Unigene14939、CL4098.Contig3 和 IId 亚类的 Unigene8813、Unigene14117 的 FPKM 值大于 10,IIb 亚类的 3 个基因 FPKM 值均小于 0.2。在根茎中有 13 个基因的 FPKM 值大于 10,其中 Unigene14117 的 FPKM 值高达 114.81,IIb 亚类的 3 个基因 FPKM 值均小于 8。说明在叶和根茎中高表达的基因主要属于 I 类 WRKY 基因和 IId 亚类 WRKY 基因。



A-Clustal Omega 联配 B- MAFFT 联配 进化树由 MEGA 6.06 软件构建, 在每个节点的数字为 1 000 次自引导值中该节点存在的百分数 The full-length sequences of the WRKY proteins were aligned using Clustal Omega (A) and MAFFT (B), and the phylogenetic tree was constructed using the neighbor-joining method in the MEGA6 software. Bootstrap values (shown at the corresponding nodes) were obtained from 1 000 replicates and are reported as percentages. The black arcs indicate different groups (or subgroups) of WRKY domains

图 3 盾叶薯蓣和拟南芥 WRKY 蛋白的系统进化树

Fig. 3 Phylogenetic tree of WRKY proteins from rhizomes of rhizomes of *D. zingiberensis* and *A. thaliana*

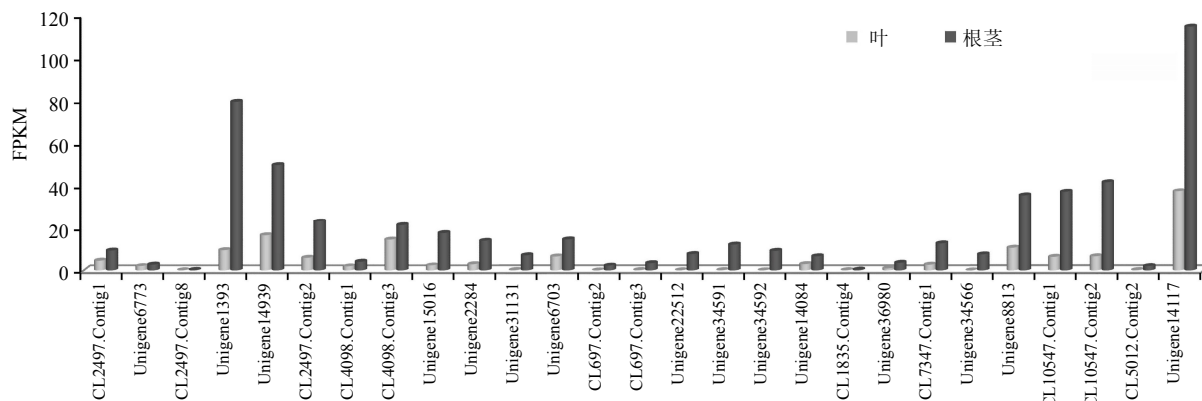


图 4 DzwRKY 基因在盾叶薯蓣叶和根茎中的表达

Fig. 4 Expression level of DzwRKY genes in leaves and rhizome of *D. zingiberensis*

2.7 盾叶薯蓣 WRKY 的功能注释

同源性较高的蛋白往往具有类似的功能^[15-16]。为了对获得的 27 个盾叶薯蓣 WRKY 基因进行功能注释,在 NCBI 上通过同源比对,检索获取与它们同源性最高的拟南芥 WRKY 序列和已知功能的药用植物 WRKY,根据已经发表的文章信息对 DzWRKY 的功能进行预测。结果(表 2)表明,27 个 DzWRKY 分别对应 18 个同源性较高的 AtWRKY,序列相似性从 39%到 61%,序列一致性较低。除 AtWRKY2 参与花粉和种子发育外,其他 AtWRKY 多为植物抗病性调控因子,参与水杨酸、茉莉酸等介导的植物抗性响应。Unigene2284、

Unigene31131 与黄连 CjWRKY1 的相似性分别为 52%、46%,CL7347.Contig1 与长春花 CrWRKY1 的相似性为 48%,均高于与拟南芥 WRKY 的同源性。Unigene36980 与黄花蒿 AaWRKY1 的相似性为 37%,同源性较低。

3 讨论

本研究鉴定出 27 个具有完整 ORF 的盾叶薯蓣 WRKY 基因,通过对预测的蛋白序列特征分析,发现所有 DzWRKY 均具有 WRKY 蛋白特有的结构域核心序列 WRKYGQK,6 个蛋白具有 2 个 WRKY 结构域,21 个只有 1 个 WRKY 结构域。三维建模表明预测的蛋白与拟南芥 AtDWKY 具有类似的 3D

表 2 拟南芥与部分药用植物 WRKY 的分子功能

Table 2 Function of WRKY transcription factors in *A. thaliana* and several medicinal plants

盾叶薯蓣基因 ID	同源蛋白	注册号	物种	同源性/%	功能	参考文献
CL10547.Contig1	AtWRKY11	NP_849559.1	拟南芥	55	细菌病害防御	17
CL10547.Contig2	AtWRKY11	NP_849559.1	拟南芥	54	细菌病害防御	17
CL1835.Contig4	AtWRKY69	XP_010916734.1	拟南芥	49		
CL2497.Contig1	AtWRKY2	NP_200438.1	拟南芥	40	花粉发育、种子发育	18-19
CL2497.Contig2	AtWRKY3	NP_178433.1	拟南芥	53	植物抗病	15
CL2497.Contig8	AtWRKY2	NP_200438.1	拟南芥	44	花粉发育、种子发育	18-19
CL4098.Contig1	AtWRKY20	NP_567752.1	拟南芥	50	调控病程相关基因	20
CL4098.Contig3	AtWRKY20	NP_567752.1	拟南芥	48	调控病程相关基因	20
CL5012.Contig2	AtWRKY11	NP_849559.1	拟南芥	49	细菌病害防御	17
CL697.Contig2	AtWRKY47	NP_192081.1	拟南芥	45		
CL697.Contig3	AtWRKY47	NP_192081.1	拟南芥	40		
CL7347.Contig1	AtWRKY70	NP_191199.1	拟南芥	39	植物抗病	21
	CrWRKY1	ADT82685.1	长春花	48	蛇根碱	9
Unigene1393	AtWRKY33	NP_181381.2	拟南芥	46	真菌抗性	22
Unigene14084	AtWRKY27	NP_568777.1	拟南芥	59	枯萎病抗性	23
Unigene14117	AtWRKY21	NP_565703.1	拟南芥	52		
Unigene15016	AtWRKY51	NP_568995.2	拟南芥	61	茉莉酸诱导的植物防御	24
Unigene2284	AtWRKY23	NP_182248.1	拟南芥	41	黄酮生物合成	25
	CjWRKY1	BAF41990.1	黄连	52	生物碱生物合成	8
Unigene25512	AtWRKY47	NP_192081.1	拟南芥	45		
Unigene31131	AtWRKY28	NP_193551.1	拟南芥	49	生物、非生物胁迫响应	26
	CjWRKY1	BAF41990.1	黄连	46	生物碱生物合成	8
Unigene34566	AtWRKY38	NP_197649.2	拟南芥	50	植物抗病的负调控因子	16
Unigene34591	AtWRKY14	NP_564359.1	拟南芥	46		
Unigene34592	AtWRKY14	NP_564359.1	拟南芥	42		
Unigene36980	AtWRKY53	NP_194112.1	拟南芥	39	叶片发育,干旱胁迫	27-28
	AaWRKY1	ACJ12926.1	黄花蒿	37	青蒿素生物合成	10
Unigene6703	AtWRKY40	NP_178199.1	拟南芥	43	调控胁迫响应基因的表达	29
Unigene6773	AtWRKY2	NP_200438.1	拟南芥	48	花粉发育、种子发育	18-19
Unigene8813	AtWRKY21	NP_565703.1	拟南芥	60		
Unigene14939	AtWRKY4	NP_172849.1	拟南芥	41	植物抗病	15

结构,具备绑定到靶基因启动子 W-box 序列调控靶基因转录的结构特征。并不是所有的 WRKY 家族蛋白都具有完整的锌指结构,如葡萄 *Vitis vinifera* L. 1 个 WRKY 结构域只有核心序列 WRKYGQK 而没有 C₂H₂ 锌指结构^[30],长春花 4 个 WRKY 结构域没有 C 端的锌指结构,1 个序列锌指结构不完全^[31]。本研究鉴定的 CL5012.Contig2 的 WRKY 结构域也只有核心序列 WRKYGQK,而没有 C 端的锌指结构。这可能就是 MEME 4.9.1 软件预测保守结构域时没有预测到 CL5012.Contig2 的 WRKY 结构域的主要原因。一般来说,Ⅰ类基因都含有 2 个 WRKY 结构域^[32],而进化分析把只含有 1 个 WRKY 结构域的 CL2497.Contig1 和 Unigene6773 与 6 个Ⅰ类基因聚在一起。这与长春花 WRKY 蛋白的进化分析结果类似,11 个Ⅰ类 WRKY 基因中有 4 个基因只有 1 个 WRKY 结构域^[13]。本研究中盾叶薯蓣第Ⅱ类 WRKY 蛋白分布不连续,Ⅱd 和Ⅱe 亚类群基因区分不够清晰,而最新的研究把 WRKY 家族分为类群Ⅰ、Ⅱa+Ⅱb、Ⅱc、Ⅱd+Ⅱe 和类群Ⅲ^[2],说明第Ⅱ类 WRKY 中亚类的进一步划分并不绝对。

表达分析发现在盾叶薯蓣叶和根茎中高表达的基因主要属于Ⅰ类 WRKY 基因和Ⅱd 亚类 WRKY 基因,而Ⅱb 亚类基因均为低表达基因,说明同类 WRKY 基因可能存在功能冗余,这与短柄草 *Brachypodium distachyon* (L.) P. Beauv. WRKY 家族基因的表达特征类似^[32]。研究表明,同一家族基因的表达模式及其编码的蛋白结构特征的相似性可能预示着功能的类似^[33],2 个结构类似的拟南芥Ⅲ类 WRKY 转录因子 AtWRKY38 和 AtWRKY62 均与组蛋白去乙酰化酶 HDA19 互作,是拟南芥病害防御的负向调控因子^[16]。不同物种 WRKY 同源蛋白也可能具有类似的分子功能,小麦 TaWRKY78 与同源的拟南芥 AtWRKY20 可交叉调控它们的 PR4 同源基因 wPR4e 和 AtHEL^[20]。已经解析的 WRKY 分子功能主要集中在调控植物的生物、非生物胁迫,个别参与植物的器官发育。植物的抗性响应主要通过激活次生代谢来实现,涉及多个次生代谢途径,其中就包括一些药效成分的合成代谢。这为解析分子生物学基础较薄弱的植物(特别是多数药用植物)基因分子功能提供了参考。

盾叶薯蓣作为世界上最好的甾体激素类药源植物之一,生产上现主要用于提取薯蓣皂苷。由于过度采挖,野生盾叶薯蓣资源不断减少,供需矛盾日

益突出^[14]。栽培盾叶薯蓣产量不稳定,薯蓣皂苷量偏低,均限制了进一步的开发利用。阐明薯蓣皂苷生物合成的分子途径,克隆参与薯蓣皂苷生物合成的关键酶和转录因子,应用分子修饰方法提高薯蓣皂苷的量,实现盾叶薯蓣的异源生产,能够在一定程度上缓解盾叶薯蓣的供需矛盾。本研究鉴定了 27 个盾叶薯蓣 WRKY 转录因子的全长 ORF,从推导的蛋白结构、结构域特征、进化分析和功能注释等多个方面深化对 DzWRKY 的认识,为进一步解析 DzWRKY 在薯蓣皂苷等药效成分合成中的分子功能奠定基础。

参考文献

- [1] 田云,卢向阳,彭丽莎,等.植物 WRKY 转录因子结构特点及其生物学功能[J].遗传,2006,28(12):1607-1612.
- [2] Rushton P J, Somssich I E, Ringler P, et al. WRKY transcription factors [J]. *Trends Plant Sci*, 2010, 15(5): 247-258.
- [3] Eulgem T, Rushton P J, Robatzek S, et al. The WRKY superfamily of plant transcription factors [J]. *Trends Plant Sci*, 2000, 5(5): 199-206.
- [4] Guo D, Zhang J, Wang X, et al. The WRKY transcription factor WRKY71/EXB1 controls shoot branching by transcriptionally regulating RAX genes in *Arabidopsis* [J]. *Plant Cell*, 2015, 27(11): 3112-3127.
- [5] Yang L, Zhao X, Yang F, et al. PtrWRKY19, a novel WRKY transcription factor, contributes to the regulation of pith secondary wall formation in *Populus trichocarpa* [J]. *Sci Rep*, 2016, 6: 1234-1240.
- [6] Li J, Brader G, Palva E T. The WRKY70 transcription factor: a node of convergence for jasmonate-mediated and salicylate-mediated signals in plant defense [J]. *Plant Cell*, 2004, 16(2): 319-331.
- [7] Jimmy J L, Babu S. Role of OsWRKY transcription factors in rice disease resistance [J]. *Trop Plant Pathol*, 2015, 40(6): 355-361.
- [8] Kato N, Dubouzet E, Kokabu Y, et al. Identification of a WRKY protein as a transcriptional regulator of benzylisoquinoline alkaloid biosynthesis in *Coptis japonica* [J]. *Plant Cell Physiol*, 2007, 48(1): 8-18.
- [9] Suttipanta N, Pattanaik S, Kulshrestha M, et al. The transcription factor CrWRKY1 positively regulates the terpenoid indole alkaloid biosynthesis in *Catharanthus roseus* [J]. *Plant Physiol*, 2011, 157(4): 2081-2093.
- [10] Han J, Wang H, Lundgren A, et al. Effects of overexpression of *AaWRKY1* on artemisinin biosynthesis

- in transgenic *Artemisia annua* plants [J]. *Phytochemistry*, 2014, 102: 89-96.
- [11] Li S, Zhang P, Zhang M, *et al.* Functional analysis of a WRKY transcription factor involved in transcriptional activation of the *DBAT* gene in *Taxus chinensis* [J]. *Plant Biol*, 2013, 15(1): 19-26.
- [12] 杨如同, 唐世蓉, 潘福生, 等. 药源植物盾叶薯蓣甾体皂苷及皂苷元的研究进展 [J]. 中国野生植物资源, 2007, 26(4): 1-5.
- [13] 谢彩侠, 高山林, 秦惠贞. 不同产地来源的盾叶薯蓣中薯蓣皂苷元含量分析 [J]. 河南中医学院学报, 2009, 24(1): 58-60.
- [14] 李朝阳, 杨朝霞, 陈功锡, 等. 不同居群盾叶薯蓣生长、叶形态及皂苷元含量的研究 [J]. 生命科学研究, 2003, 7(1): 58-64.
- [15] Lai Z, Vinod K M, Zheng Z, *et al.* Roles of *Arabidopsis* WRKY3 and WRKY4 transcription factors in plant responses to pathogens [J]. *BMC Plant Biol*, 2008, 8(1): 68-75.
- [16] Kim K C, Lai Z, Fan B, *et al.* *Arabidopsis* WRKY38 and WRKY62 transcription factors interact with histone deacetylase 19 in basal defense [J]. *Plant Cell*, 2008, 20(9): 2357-2371.
- [17] Journot-Catalino N, Somssich I E, Roby D, *et al.* The transcription factors WRKY11 and WRKY17 act as negative regulators of basal resistance in *Arabidopsis thaliana* [J]. *Plant Cell*, 2006, 18(11): 3289-3302.
- [18] Guan Y, Meng X, Khanna R, *et al.* Phosphorylation of a WRKY transcription factor by MAPKs is required for pollen development and function in *Arabidopsis* [J]. *PLoS Genet*, 2014, 10(5): e1004384.
- [19] Jiang W, Yu D. *Arabidopsis* WRKY2 transcription factor mediates seed germination and postgermination arrest of development by abscisic acid [J]. *BMC Plant Biol*, 2009, 9(1): 1.
- [20] Proietti S, Bertini L, Van der Ent S, *et al.* Cross activity of orthologous WRKY transcription factors in wheat and *Arabidopsis* [J]. *J Exp Bot*, 2011, 62(6): 1975-1990.
- [21] Li J, Brader G, Kariola T, *et al.* WRKY70 modulates the selection of signaling pathways in plant defense [J]. *Plant J*, 2006, 46(3): 477-491.
- [22] Zheng Z, Qamar S A, Chen Z, *et al.* *Arabidopsis* WRKY33 transcription factor is required for resistance to necrotrophic fungal pathogens [J]. *Plant J*, 2006, 48(4): 592-605.
- [23] Mukhtar M S, Deslandes L, Auriac M C, *et al.* The *Arabidopsis* transcription factor WRKY27 influences wilt disease symptom development caused by *Ralstonia solanacearum* [J]. *Plant J*, 2008, 56(6): 935-947.
- [24] Gao Q M, Venugopal S, Navarre D, *et al.* Low oleic acid-derived repression of jasmonic acid-inducible defense responses requires the WRKY50 and WRKY51 proteins [J]. *Plant Physiol*, 2011, 155(1): 464-476.
- [25] Grunewald W, De Smet I, Lewis D R, *et al.* Transcription factor WRKY23 assists auxin distribution patterns during *Arabidopsis* root development through local control on flavonol biosynthesis [J]. *Proc Natl Acad Sci USA*, 2012, 109(5): 1554-1559.
- [26] Chen X, Liu J, Lin G, *et al.* Overexpression of AtWRKY28 and AtWRKY75 in *Arabidopsis* enhances resistance to oxalic acid and *Sclerotinia sclerotiorum* [J]. *Plant Cell Rep*, 2013, 32(10): 1589-1599.
- [27] Xie Y, Huhn K, Brandt R, *et al.* REVOLUTA and WRKY53 connect early and late leaf development in *Arabidopsis* [J]. *Development*, 2014, 141(24): 4772-4783.
- [28] Sun Y, Yu D. Activated expression of AtWRKY53 negatively regulates drought tolerance by mediating stomatal movement [J]. *Plant Cell Rep*, 2015, 34(8): 1295-1306.
- [29] Van Aken O, Zhang B, Law S, *et al.* AtWRKY40 and AtWRKY63 modulate the expression of stress-responsive nuclear genes encoding mitochondrial and chloroplast proteins [J]. *Plant Physiol*, 2013, 162(1): 254-271.
- [30] Guo C, Guo R, Xu X, *et al.* Evolution and expression analysis of the grape (*Vitis vinifera* L.) WRKY gene family [J]. *J Exp Bot*, 2014, 65(6): 1513-1528.
- [31] 杨致荣, 王兴春, 薛金爱, 等. 药用植物长春花 WRKY 转录因子的鉴定及表达谱分析 [J]. 生物工程学报, 2013, 29(6): 785-802.
- [32] Wen F, Zhu H, Li P, *et al.* Genome-wide evolutionary characterization and expression analyses of WRKY family genes in *Brachypodium distachyon* [J]. *DNA Res*, 2014, 21(3): 327-339.
- [33] Wang F, Suo Y, Wei H, *et al.* Identification and characterization of 40 isolated *Rehmannia glutinosa* MYB family genes and their expression profiles in response to shading and continuous cropping [J]. *Int J Mol Sci*, 2015, 16(7): 15009-15030.