

文章编号: 1674-7054(2012)04-0305-09

蓖麻 *WRKY14* 基因的鉴别与信息分析

黄启星^{1,2} 郭安平² 杨礼富¹ 邹智¹

(1. 中国热带农业科学院 橡胶研究所/农业部橡胶树生物学与遗传资源利用重点实验室, 海南 儋州 571737;

2. 中国热带农业科学院 热带生物技术研究所以农业部热带作物生物学与遗传资源利用重点实验室, 海南 海口 571101)

摘 要: 基于已公布的蓖麻基因组序列, 利用电子克隆技术纠正了拼接错误的 *WRKY* 基因序列, 获得蓖麻 *WRKY14* 基因信息。结果表明: 该基因含有 3 个内含子, 编码 318 个氨基酸的多肽, 理论相对分子质量为 35.49×10^3 , 等电点为 8.64, 总平均亲水性为 -0.803, 不稳定系数为 46.88, 核定位, 属于不稳定的亲水性核定位蛋白; 该蛋白含有 1 个 *WRKY* 结构域, 其基序为 *WRKYGQK_X13_C_X5_C_X23_H_X1_H*, 保守内含子位于 199 位的 K 和 200 位的 V 之间, 可归为 II 大类 a 亚类; *RcWRKY14* 的启动子区含有多种激素和胁迫应答相关元件 *RcWRKY14* 在叶片、花、胚乳和种子中都有表达, 其中在花中的表达量最高, 种子和叶片中次之, 胚乳中最低。

关键词: 蓖麻; *WRKY*; 转录因子; 电子克隆; 生物信息学分析

中图分类号: S 565.6

文献标志码: A

WRKY 蛋白是一类植物特异的转录因子, 其典型特征是含有 1~2 个由约 60 个高度保守的氨基酸组成的 *WRKY* 结构域, 该结构域包含 N 端高度保守的 *WRKYGQK* 核心序列和 C 端可变的锌指结构^[1-2], 它通过特异地与启动子区域的 W-box 结合进而调控靶基因的表达^[2]。根据 *WRKY* 结构域的数量及其锌指结构的特征, *WRKY* 转录因子可分为三大类: I 类含有 2 个 *WRKY* 结构域, 锌指结构为 C_2H_2 型; II 类只含 1 个 *WRKY* 结构域, 锌指结构为 C_2H_2 型, 拟南芥中该类又可进一步分成 II a, II b, II c, II d 和 II e 5 个亚类; III 类虽只含 1 个 *WRKY* 结构域, 但锌指结构为 C_2HC 型^[1-3]。自 20 世纪 90 年代初, *WRKY* 转录因子在甘薯 (*Ipomoea batatas*) 中被发现以来^[4], 又先后在燕麦 (*Avena fatua*)、拟南芥 (*Arabidopsis thaliana*)、水稻 (*Oryza sativa*) 等各种高低等植物中得到克隆和鉴定^[1,3,5], 研究结果显示, *WRKY* 广泛参与植物的生长、发育、代谢及各种生物与非生物胁迫应答^[6-12]。

蓖麻 (*Ricinus communis* L.) 是 1 种隶属于大戟科 (Euphorbiaceae) 的 1 年生或多年生灌木, 最早起源于非洲。因其具有种子含油量高、脂肪酸成分特异, 在热带和亚热带地区被作为 1 种特种油料作物广为种植^[13]。蓖麻的适应性特别强, 易于在不良环境下种植^[14], 但是, 其耐胁迫的分子机理至今还鲜为人知。最近, 蓖麻 4.6 倍基因组草图的绘制^[15] 及不同组织表达序列标签 (expressed sequence tags, ESTs) 序列^[16-19] 的释放在相关基因的克隆与鉴定打下了基础。2012 年, 在 phytozome (v8.0) 数据库中公布的蓖麻基因组大小约为 400 Mb, 包括 25 878 条 Scaffold, 31 221 条转录本, 其中包括 47 条 *WRKY* 基因的转录本。尽管如此, 目前尚未见有关蓖麻 *WRKY* 基因家族结构和功能的精细注释。

基于 ESTs 的电子克隆 (in silico cloning) 策略, 是近年来发展起来的一门快速克隆基因的新技术, 其技术核心是利用生物信息学技术组装延伸 ESTs 序列, 获得基因的部分乃至全长 cDNA 序列, 进一步利用 RT-PCR 的方法进行克隆分析、验证。这为深入分析和挖掘蓖麻基因组数据, 对蓖麻 *WRKY* 基因序列、结构和功能进行精细注释, 进而探讨 *WRKY* 转录因子在蓖麻耐胁迫中的作用提供了支持。笔者以保守的 *WRKY* 结构域作为种子, 从蓖麻基因组及 ESTs 序列中鉴别出 56 个潜在的 *WRKY* 基因 (见表 1), 其中, 基

收稿日期: 2012-09-26

基金项目: 转基因新品种培育重大专项 (008ZX08012-04); 中国热带农业科学院橡胶研究所基本科研业务费专项 (1630022011014); 海南省重点科技项目 (90107) 和海南省自然科学基金 (312026)

作者简介: 黄启星 (1981-), 男, 广东南雄人, 中国热带农业科学院橡胶研究所助理研究员, 博士。

通信作者: 邹智 (1982-), 男, 助理研究员, 硕士, 研究方向: 植物抗逆生物学, E-mail: zouzhi2008@126.com

因组注释号为 43951. m000016 的基因序列与 30131. m006850 及相关 ESTs 虽然存在差异,但具有很高的相似性,因而推测其可能是相似性非常高的同源基因或者是同一基因的错误拼接所致。为此,本研究基于已公布的基因组及 ESTs 序列,采用电子克隆技术对该基因进行克隆,获得了蓖麻 *WRKY14* 基因的全序列,在此基础上对其结构及可能的生物学功能进行预测,旨在为后续实验验证提供数据支持。

表 1 从蓖麻基因组及 ESTs 序列中发掘的 56 个 *WRKY* 基因

编号	<i>WRKY</i> 基因	登录号	编号	<i>WRKY</i> 基因	登录号	编号	<i>WRKY</i> 基因	登录号
1	<i>RcWRKY1</i>	29949. m000123	20	<i>RcWRKY20</i>	30076. m004548	39	<i>RcWRKY39</i>	29644. m000187 29644. m000188
2	<i>RcWRKY2</i>	29635. m000468	21	<i>RcWRKY21</i>	30010. m000675	40	<i>RcWRKY40</i>	30170. m013832
3	<i>RcWRKY3</i>	29687. m000562	22	<i>RcWRKY22</i>	29842. m003555	41	<i>RcWRKY41</i>	29598. m000445
4	<i>RcWRKY4</i>	30174. m009166	23	<i>RcWRKY23</i>	29929. m004624	42	<i>RcWRKY42</i>	29848. m004464
5	<i>RcWRKY5</i>	29929. m004587	24	<i>RcWRKY24</i>	30190. m010908	43	<i>RcWRKY43</i>	30076. m004648
6	<i>RcWRKY6</i>	28966. m000524	25	<i>RcWRKY25</i>	28040. m000035	44	<i>RcWRKY44</i>	27996. m000145
7	<i>RcWRKY7</i>	29820. m001029	26	<i>RcWRKY26</i>	29767. m000208	45	<i>RcWRKY45</i>	30026. m001461
8	<i>RcWRKY8</i>	29805. m001504	27	<i>RcWRKY27</i>	29771. m000072	46	<i>RcWRKY46</i>	30190. m011276
9	<i>RcWRKY9</i>	29717. m000222	28	<i>RcWRKY28</i>	29739. m003586	47	<i>RcWRKY47</i>	30174. m008663
10	<i>RcWRKY10</i>	29848. m004539	29	<i>RcWRKY29</i>	30147. m014474	48	<i>RcWRKY48</i>	30169. m006581
11	<i>RcWRKY11</i>	27613. m000639	30	<i>RcWRKY30</i>	30076. m004623	49	<i>RcWRKY49</i>	30174. m008651
12	<i>RcWRKY12</i>	29848. m004544	31	<i>RcWRKY31</i>	29693. m002060	50	<i>RcWRKY50</i>	30169. m006533
13	<i>RcWRKY13</i>	29848. m004545	32	<i>RcWRKY32</i>	30174. m009135	51	<i>RcWRKY51</i>	30190. m010814
14	<i>RcWRKY14</i>	43951. m000016 30131. m006850	33	<i>RcWRKY33</i>	30190. m011278	52	<i>RcWRKY52</i>	28690. m000025
15	<i>RcWRKY15</i>	29822. m003484	34	<i>RcWRKY34</i>	29889. m003321	53	<i>RcWRKY53</i>	29729. m002330
16	<i>RcWRKY16</i>	30147. m014087	35	<i>RcWRKY35</i>	30174. m008669	54	<i>RcWRKY54</i>	29729. m002370
17	<i>RcWRKY17</i>	29990. m000497	36	<i>RcWRKY36</i>	28644. m000915	55	<i>RcWRKY55</i>	29915. m000479
18	<i>RcWRKY18</i>	29736. m002023	37	<i>RcWRKY37</i>	28455. m000362	56	<i>RcWRKY56</i>	29729. m002369
19	<i>RcWRKY19</i>	30064. m000506	38	<i>RcWRKY38</i>	29883. m002008			

1 材料与方法

1.1 材料 笔者于 2012 年 8 月从 phytozome 数据库 (v8.0 <http://www.phytozome.net/ricinus.php>) 下载蓖麻的全基因组序列。蓖麻的基因(coding sequences ,CDS)、转录组和 ESTs 数据来源于 GenBank (<http://www.ncbi.nlm.nih.gov/>)。

1.2 方法

1.2.1 目标 *WRKY* 基因的电子克隆 分别以 43951. m000016 和 30131. m006850 的蓖麻基因序列作为查询序列,采用 BLASTn^[20] 对蓖麻 ESTs 数据进行同源搜索,编写 Perl 语言脚本,按照相似度(identity) > 95, 匹配长度 > 50 bp 筛选匹配的 EST 序列,获取相关 ESTs 后进行人工拼接,然后以拼接序列作为查询序列再次对 ESTs 数据进行 BLASTn 分析,直到序列不能再延长为止。

1.2.2 基因序列的生物信息学分析 同源分析采用 BLAST (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>); 基因编码区的预测采用 ORF Finder 软件(<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>); 转录起始位点的预测采用 BDGP 软件^[21]; 启动子区域的顺式作用元件分析采用 PLACE 软件^[22] 和 PlantCARE 软件^[23]; 多序列比对和进化树的构建采用 ClustalW2 软件^[24] 和 MEGA (v4.0) 软件^[25]; 蛋白的一级结构分析采用 Prot-Param 软件 (<http://web.expasy.org/protparam/>); 保守结构域和模体搜索采用 CDD 软件^[26]、SMART 软件^[27] 和 MEME 软件^[28]; 翻译后修饰采用 MotifScan 软件^[29]; 亚细胞定位分析采用 PSORT 软件^[30] 和 Nu-cPred 软件^[31]; 跨膜结构域的预测采用 TMHMM 软件 (<http://www.cbs.dtu.dk/services/TMHMM-2.0/>)。除 BLAST 分析按照相似度(identity) > 95, 匹配长度 > 50 bp 的条件过滤以外,其余软件均采用默认参数运行。

1.2.3 基因的组织特异性分析 从 GenBank 高通量测序数据库(<http://www.ncbi.nlm.nih.gov/sra/>) 下

载蓖麻转录组测序 SRA (sequence read archive) 文件。采用 Blat 软件^[32], 将叶片 (NCBI 登录号 ERX021378)、发育中的雄花 (NCBI 登录号 ERX021379)、发育中的胚乳 II/III 期 (NCBI 登录号 ERX021375)、发育中的胚乳 V/VII 期 (NCBI 登录号 ERX021376) 及萌发的种子 (NCBI 登录号 ERX021377) 的原始转录组测序数据比对到蓖麻基因组序列上, 按照相似度 (minIdentity) > 95, 匹配长度 > 50 bp 的条件过滤以后, 以 RPKM (Reads Per Kilo bases per Million reads, 每百万 reads 中来自于某基因每千碱基长度的 reads 数) 值比较基因在不同植物组织内的表达变化。RPKM 计算公式如下:

$$RPKM = (10^9 \times C) \div (N \times L),$$

式中: C 为基因比对上的 reads 数; N 为样本比对上的所有 reads 数; L 为基因长度。

2 结果与分析

2.1 目标 *WRKY* 基因的电子克隆与序列分析 以基因序列 43951. m000016 和 30131. m006850 作为查询序列, BLASTn 分析分别获得 6 条 (图 1A) 和 5 条 (图 1B) 相似度超过 95% 的 ESTs 序列。其中, 30131. m006850 获得的 5 条 (NCBI 登录号为 EG686466.1, EG688379.1, EG688380.1, EG686269.1 和 EG685091.1) 包含在 43951. m000016 获得的 6 条序列 (另 1 条为 EG686467.1) 中。序列比对表明, 43951. m000016 和 30131. m006850 为同一基因错误拼接的结果, 其相关 ESTs 经人工拼接后得到 1 条 1 371 bp 的重叠群, 以该序列作为种子进行 BLASTn 分析未能找到新的 ESTs 序列 (图 1C)。

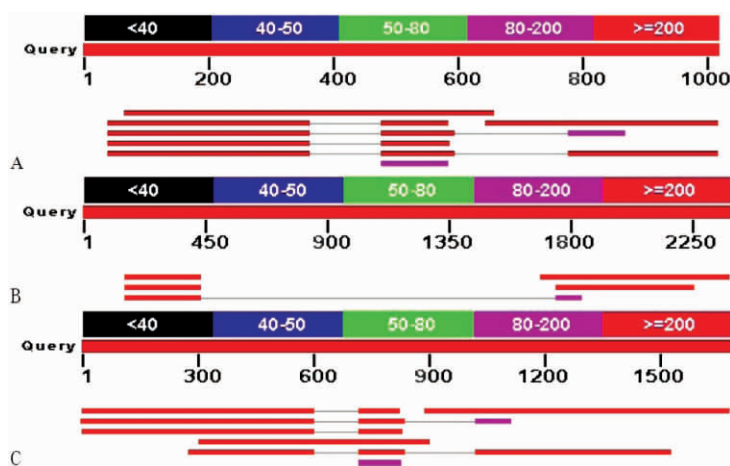


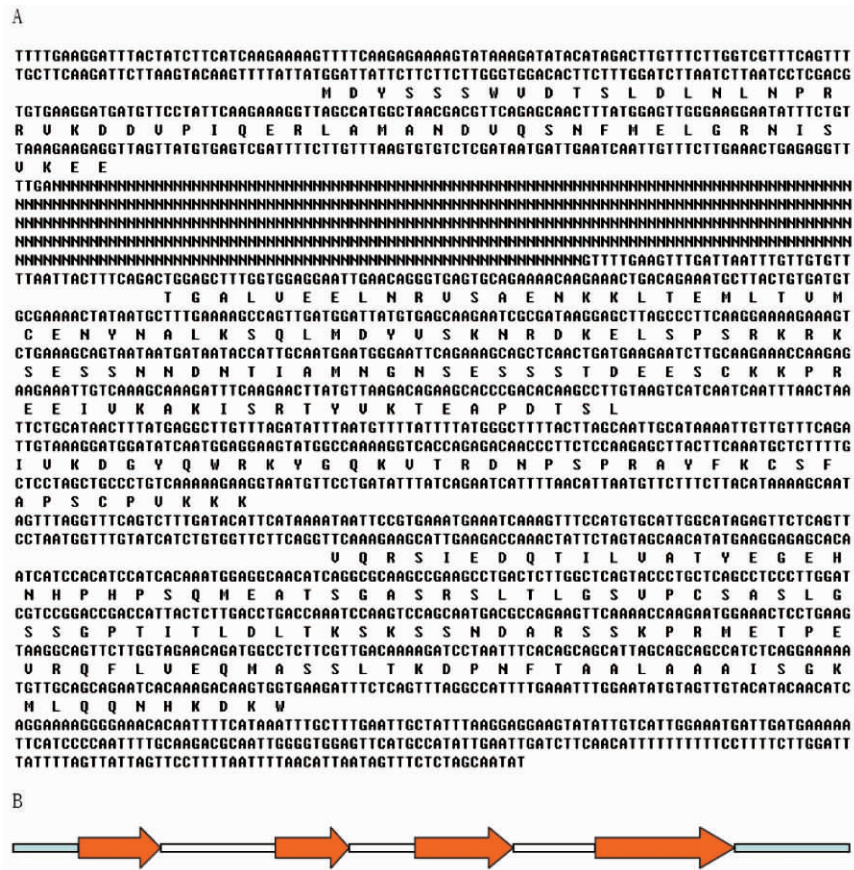
图1 目标 *WRKY* 基因的电子克隆

A, B, C 分别为 43951. m000016, 30131. m006850 和人工拼接重叠群的 BLASTn 分析结果

利用 ORF Finder 软件分析基因结构, 结果显示, 该序列包含 1 个 957 bp 的完整读码框以及 121 bp 的 5'UTR 和 293 bp 的 3'UTR。根据进化关系, 将该基因命名为 *RcWRKY14*。

BLASTn 分析表明, *RcWRKY14* 位于 scf_1107001315136 (NCBI 登录号 NW_003010092.1) 上。将 43951. m000016, 30131. m006850, 拼接得到的 contig, BLASTn 获得的所有 ESTs 分别比对到 NW_003010092.1 上, 发现 *RcWRKY14* 包含 3 个内含子。由于第 1 内含子未测通, 发现对 43951. m000016 和 30131. m006850 的原始拼接存在错误, 存在大约 400 bp 的 gap (图 2)。由于有 2 条 EST (NCBI 登录号 EG686466.1, EG686467.1) 比对到 *RcWRKY14* 的第 2 个和第 3 个内含子上, 表明 *RcWRKY14* 存在可变剪接 (图 1)。

2.2 *RcWRKY14* 编码蛋白的一级结构分析 序列分析表明, *RcWRKY14* 编码 1 条 318 个氨基酸的多肽 (图 2A), 理论相对分子质量 (mw) 为 35.49×10^3 , 等电点 (pI) 为 8.64, 带电荷数为 +4.458, 包含 46 个强碱性氨基酸 (K, R), 42 个强酸性氨基酸 (D, E), 81 个疏水氨基酸 (A, I, L, F, W, V) 和 107 个极性氨基酸 (N, C, Q, S, T, Y)。 *RcWRKY14* 蛋白总平均亲水性 (GRAVY) 为 -0.803, 脂肪族指数 (Aliphatic index) 为 64.40, 半衰期为 30 h, 不稳定系数 (II) 为 46.88, 表明 *RcWRKY14* 为不稳定的亲水性蛋白。

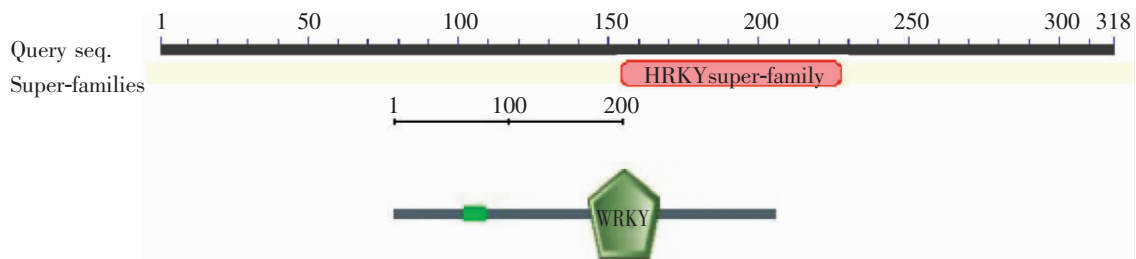
图2 *RcWRKY14* 的基因结构

A: *RcWRKY14* 的基因序列及预测的编码蛋白, N 代表未知序列;

B: *RcWRKY14* 的结构示意图, 橙色箭头表示编码区及转录方向,

绿色框表示转录非翻译区, 白色框表示内含子

2.3 *RcWRKY14* 蛋白保守结构域和模体分析 结构域分析显示, *RcWRKY14* 的 162 ~ 222 位为保守的 *WRKY* 结构域, 59 ~ 77 位为卷曲螺旋区 (coiled coil region) (图 3)。MEME 软件^[28] 分析显示, *RcWRKY14* 包含 3 个主要模体 (Motif)。模体 1 为 MDY [SV] S [KS] [NW] [RV] D (即 1 ~ 9 位的 MDYSSSWVD 和 90 ~ 98 位的 MDYVSKNRD); 模体 2 为 D [GN] [PY] [QS] [PW] R [AK] Y (即 164 ~ 171 位的 DGYQWRKY 和 178 ~ 185 位的 DNPSPRAY); 模体 3 为 [EQ] [HQ] NH [KP] [DH] (即 217 ~ 222 位的 EHNHPH 和 311 ~ 316 位的 QQNHKD)。

图3 *RcWRKY14* 蛋白的保守结构域分析

2.4 *RcWRKY14* 蛋白翻译后修饰、亚细胞定位和跨膜结构域分析 MotifScan 软件^[29] 分析结果表明, *RcWRKY14* 蛋白的 47 ~ 50 和 296 ~ 299 位为潜在的天冬酰胺 N 末端糖基化 (ASN_GLYCOSYLATION) 位点; 229, 237, 255 和 259 位的 T 为潜在的氧糖基化 (O-glycosylated) 位点; 69 ~ 72 和 106 ~ 109 位为潜在的 cAMP 和 cGMP 依赖的蛋白激酶磷酸化 (cAMP-and cGMP-dependent protein kinase phosphorylation) 位点; 6 ~ 9, 10 ~ 13, 49 ~ 52, 112 ~ 115, 127 ~ 130, 203 ~ 206, 224 ~ 227 和 263 ~ 266 位为潜在的酪氨酸蛋白激酶

II 磷酸化(Casein kinase II phosphorylation) 位点; 122 ~ 127 , 172 ~ 177 , 231 ~ 236 和 239 ~ 244 位为潜在的 N 末端豆蔻酰化(N-myristoylation) 位点; 49 ~ 51 , 104 ~ 106 , 133 ~ 135 , 181 ~ 183 , 269 ~ 271 和 306 ~ 308 位为潜在的蛋白激酶 C 磷酸化 (Protein kinase C phosphorylation) 位点。

亚细胞定位分析表明 , RcWRKY14 含有 1 个核定位信号序列。TMHMM 软件分析显示 , RcWRKY14 无跨膜结构域。

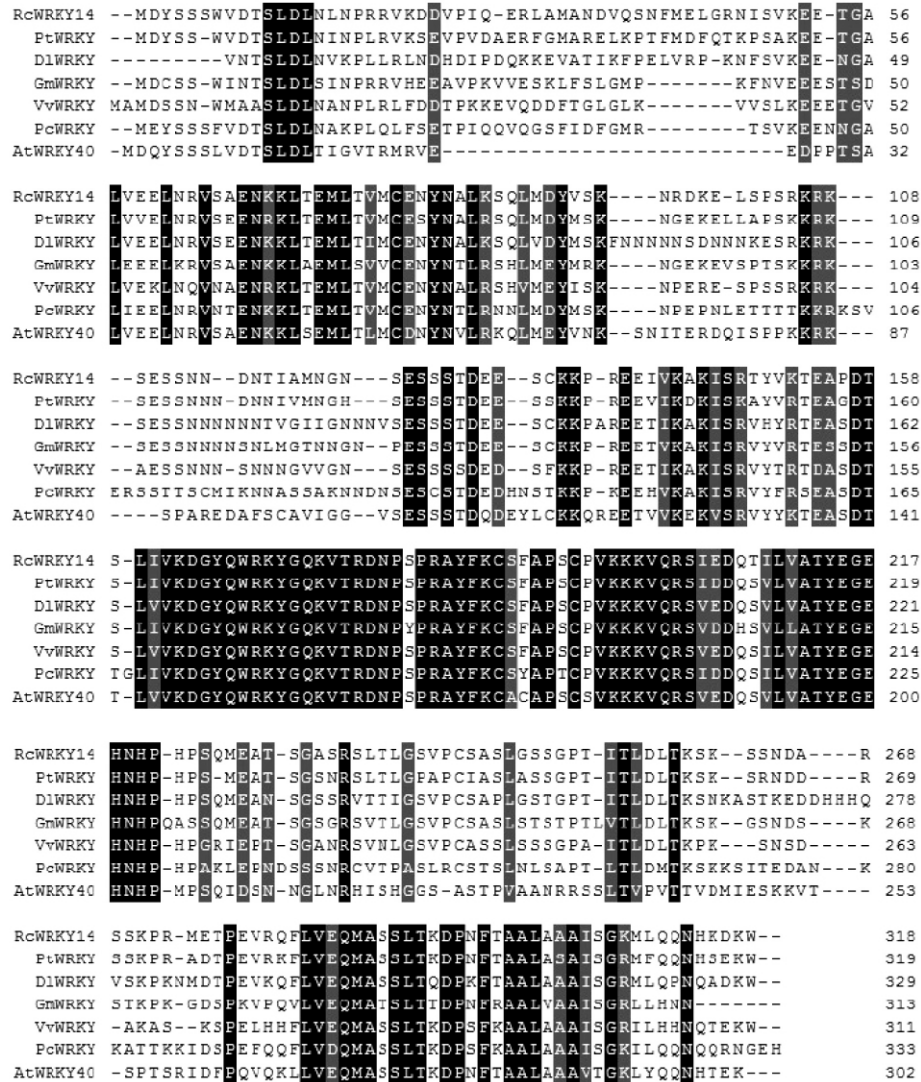


图 4 RcWRKY14 与其他物种中的同源蛋白的多重比对

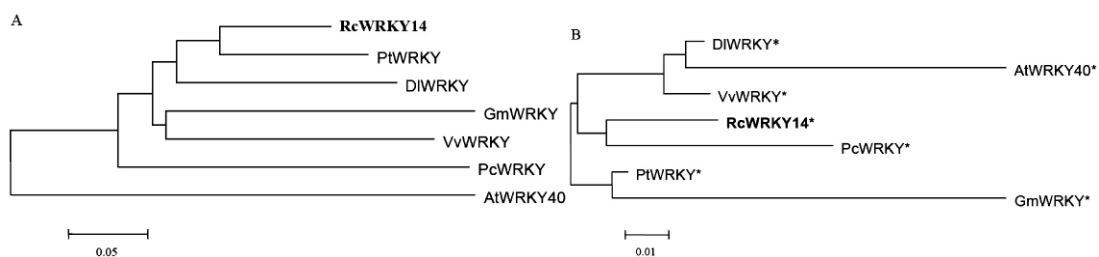


图 5 RcWRKY14 与其他物种同源蛋白的进化分析

A: 基于全蛋白序列构建的进化关系树; B: 基于 WRKY 结构域构建的进化关系树

2.5 RcWRKY14 蛋白同源比对与进化分析 序列比对表明 , RcWRKY14 与杨树 PtWRKY (*Populus trichocarpa* , XP_002303852. 1) 、大豆 GmWRKY (*Glycine max* , ACU18913. 1) 、龙眼 DIWRKY (*Dimocarpus lon-*

gan, AEO31523.1)、葡萄 VvWRKY (*Vitis vinifera*, XP_002265707.1)、荷兰芹 PcWRKY (*Petroselinum crispum*, AAG35658.1) 等物种的 WRKY 蛋白的相似性都在 70% 以上,与拟南芥 AtWRKY40 (*Arabidopsis thaliana*, NP_178199.1) 的相似性也接近 60%。

多重比对表明,不同物种中的 WRKY 蛋白在 WRKY 结构域附近的氨基酸非常保守,而其他区域变异相对较大(图 4)。分别采用全蛋白序列和 WRKY 结构域序列构建进化树,两者结果差异较大(图 5)。其中,以全蛋白序列构建的进化树更符合这些物种的真实进化关系。根据 Eulgem 等^[1]的分类标准, *RcWRKY14* 归属于 II 大类 α 亚类转录因子,其 WRKY 结构域基序为 WRKYGQK_X13_C_X5_C_X23_H_X1_H,保守内含子位于 199 位的 K 和 200 位的 V 之间,与 *AtWRKY40* 类似。

2.6 *RcWRKY14* 基因启动子区域的顺式作用元件分析 BDGP 软件^[21]分析结果表明,在 *RcWRKY14* 起始密码子上游 2033 bp 的序列中,含有的 7 个分值在 0.94 以上的基础启动子序列,其中,上游 142~195 位的序列得分最高。结合 ESTs 的界定结果(起始密码子上游 123 bp),推测 *RcWRKY14* 的转录起始点最大可能为起始密码子上游 155 位的 A 碱基(图 6)。

```

1      tatacatgtacggctagtaaagtcaagaaatgcaattatttttgacaaatgaatgtatggtacgaccatgga
71     ctagactagttgggaaaggtgtgcatcacaccttatataaaaatgcatggttaattgtccaaatcctttt
141    aacttaaaaaatagtaattaattgagggcggtgaataagcactaacttatatatatttttcathtagaaa
211    aggaaaaaggaagaaaaaaacaaaaggatgatcaacaagtcacatgcatgcttgggaaagaaaagaatgat
281    cgtaatttgaattgaaattgaatggctactacttttctactgtcttccatttctctgtacttggtctaaag
351    ttggattagaaataaataaataaataaataaattgggggaatttagtttgacatcctttgttaaagtaata
421    acttttcatatcattaagattataaaactagggtttaaatatttttattttcatgaagaagggtgagtgaat
491    gtggaaaacaaggattttaatatgttgttttaattataactaatagcttttatctcaagtaattaattagaa
561    gattaataatatgcaaataaaaataaataaaacactatatgaaatatgtgattcattctatagttcctcta
631    actagacatatgtatcatgtgtgaaaatgtaaactctattgtaatttgtgatattcacgattgattaaaca
701    aaaatttgggtatgctgtcaaatgtacagatttatggatcacatgtaaatattctaaacatttataaaaaa
771    aaaataaaaaataaattataagcatgtgcaaatcctttttctttattttgcataaacatttgcacatattta
841    ttagtaataatgaaatcataaaataggaaaacctttttcttccgtatacaaaaagaaaaagataataaaa
911    taaaataaaaaataaagaatgaggattcctttttaaaccataatataatgaaggcatagttctataacgtgg
981    aaaatatttggaaaacgtgactgttaaagaaagctttttatatacacgaggaattaaataagtaaaaaggta
1051   gaccccataataaaaacaaaagagtgtagcaaaaaaaagaaaaaaaagattttaaagatcggttagctac
1121   ccaataccataaaacccccaaataaaataaaaagcatttttggtcaatgttaaacttccaatccttctaa
1191   aatagatgccaaatgcatagtttgactattgtatatgaataccttcttaacttctaataatttttcaatta
1261   ctatttcaacttattatataatttcatacataaaaaacttagtacatataaaaaaatatataagtgaaaaat
1331   taaaattcataaaattaaaataataattgagaatgaatgctaaagaaaatctcaataagtaaaaaaattt
1401   aaaaatggaatttaggagtgtagcagtgaaattccttttctataaaataaaaaccagaccataaaattgg
1471   ttaaatcatatcgacacatgaacatccattgtttattgctcttttgggtgggcgttgactttgaaaa
1541   gcaaatgttgttacccaacccgaatgtctcttaccaataatcaatccacctttccaatgttgtgtgttg
1611   ttttttttttttttttttttttcttcttaccctatatagagagcatgcacatctcttccaatcacttcat
1681   aaattcatgtggcaaccataatcttatctctctctctatatatccccacactcgccactttccaccttgt
1751   ttcttcataaacgagtagagccttgattcttttatttctctctctatatatgcgtatatatatatatatcg
1821   tgttcttgcatttttgggttcAtcaggaggactagctagctacaagaatTTTGAAGGATTACTATCTTC
1891   ATCAAGAAAAGTTTCAAGAGAAAAGTATAAAGATATACATAGACTTGTTCCTTGGTCGTTTCAGTTTTCG
1961   CTTCAAGATTCTTAAGTACAAGTTTATTATG

```

图 6 *RcWRKY14* 启动子区域的顺式作用元件分析

灰色背景区代表预测的基础启动序列;灰色背景区的大写字母 A 代表预测的转录起始位点;白色背景区的大写字母代表 ESTs 界定的转录区;加黑的大写字母 ATG 代表起始密码子;单下划线标识的代表 TATA-box;双下划线标识的代表 CAAT-box;黄色加亮区代表 W-box

顺式作用元件分析显示, *RcWRKY14* 的启动子区域除含有多多个 TATA-box 和 CAAT-box 等基础启动子元件外(仅正向链就含有 8 个 TATA-box 和 13 个 CAAT-box,图 6),还包含 GATA-BOX, 10PEHVPSBD,

-300ELEMENT, -300CORE 和 CTRMCAMV35S 等多个增强子序列, 包含脱落酸、细胞分裂素、生长素、赤霉素、乙烯、水杨酸等多种激素的响应元件, 以及光、O₂、缺氧、CO₂、Ca²⁺、低温、干旱、脱水及病原响应相关的元件(表 2)。RcWRKY14 的启动子区含有 28 个 W-box, 其中仅正向链就含有 5 个(图 6, 表 2), 表明 RcWRKY14 具有复杂的转录调控网络, 可能同时受其他 WRKY 蛋白调控和自我调控, 这也预示该基因可能在很多生物学过程中发挥作用。

表 2 RcWRKY14 启动子区域中预测的主要顺式作用元件

核心序列	PLACE序列名称	启动子中出现 的次数	功能或特性	核心序列	PLACE序列名称	启动子中出现 的次数	功能或特性
TATAAAT	TATABOX2	2	core promoter	ACGTG	ABRELATERD1	3	ABA-responsive
TATTAAT	TATABOX3	1		ATACGTGT	ZDNAFORMINGATCAB	1	
TATATAA	TATABOX4	7		NGATT	ARR1AT	21	Cytokinin and light-responsive
TTATT	TATABOX5	23		TATTAG	CPBCSPOR	3	Cytokinin-responsive
CAAT	CAATBOX1	27	Enhancer	ACTTTA	NTBBFIARROLB	3	
GATA	GATABOX	9		CATATG	CATATGMSAUR	2	Auxin-responsive
TATTCT	10PEHVPSBD	1	Enhancer and tissue-specific	TGTCTC	ARFAT	1	
TGHAARK	300ELEMENT	4		CCWWWWWWWG	CARGNCAT	3	GA-responsive
TGTAAG	300CORE	1	Enhancer	TAACAAR	GAREAT	1	
TCTCTCTC	CTRMCAMV35S	2		TAACAAA	MYBGAVH	1	GA and sugar-responsive
YACT	CACTFTPPCA1	28	Mesophyll-specific	AWTTCAAA	ERELEE4	1	Ethylene-responsive
CANNTG	EBOXBNNAPA	20	Seed-specific	GTAC	CURECORECR	16	Oxygen-responsive
AGAAA	POLLENILELAT52	18	Pollen-specific	CATGTG	MYCATERD1	5	Dehydration-responsive
AAAGAT	NODCON1GM	5	Nodule-specific	GANTTNC	EECCRAH1	4	CO ₂ -responsive
CTCTT	NODCON2GM	5	Root-specific	AAACAAA	ANAERO1CONSENSUS	3	Anaerobic-responsive
ATATT	ROOTMOTIFTAPOX	16		TGTCA	BIHD1OS	3	Pathogen-responsive
KCACGW	RHERPATEXPA7	2	Root hair cell-specific	MACGYGB	ABRERATCAL	1	Ca ²⁺ -responsive
CANNTG	MYCCONSUSAT	20	ABA and Drought-responsive	AACCAA	REALPHALGLHCB21	1	Required for phytochrome regulation
CACATG	MYCATRD22	5		GATAAG	IBOX	1	light-responsive
ACACNNG	DPBFCOREDCDC3	3		RYCGAC	CBFHV	1	Low temperature-responsive
CCGAC	LTRECOREATCOR1	1		AAAG	DOFCOREZM	44	Dof transcription factor
WAACCA	MYBIAT	4		TGAC	WRKY71OS	9	WRKY transcription factor

2.7 RcWRKY14 基因的组织特异性分析 表达谱分析表明, RcWRKY14 基因在蓖麻的叶片、花、胚乳 II / III 期、胚乳 V / VI 期和种子中都有表达。其中, 在蓖麻花中的表达丰度最高(RPKM = 13.17), 在胚乳 V / VI 期的表达丰度最低(RPKM = 0.16) (图 7)。

3 讨 论

蓖麻是一种经济价值非常高的特种油料作物, 其种子含油量在 $w = 35\% \sim 55\%$ 之间, 而种子来源的蓖麻油约有 $w = 90\%$ 为蓖麻油酸, 相对单一的脂肪酸组成赋予了蓖麻油独特的理化特性, 并使其成为生产高品质润滑油、化妆品、医药和化工产品的重要原料^[13]。但人们对蓖麻油、蓖麻毒素生物合成的分子网络及蓖麻耐胁迫的分子机理知之甚少。虽然蓖麻的基因组草图已公布^[14], 但

由于测序深度有限及遗传图谱的缺乏, 至今 31 221 个预测的基因仍散布于 25 878 条 scaffolds 上, 很难避免有些基因被错误注释。ESTs 是指将表达基因的转录本反转录成 cDNA 后再进行测序而获得的短 DNA 片段(通常 300 ~ 500 bp)。自 Adams 等^[33]首次报道应用 ESTs 寻找新基因, 至今 ESTs 数据已成为新基因发现、基因克隆(包括电子克隆)、高通量基因表达谱分析、基因组组装与基因结构注释的重要资源。截至 2012 年 7 月, NCBI 中已公布的蓖麻 ESTs 序列达 73 360 923 条, 这为校正被错误注释的基因序列提供了宝贵的资源。

本研究中, 笔者将被错误组装和注释成 43951. m000016 和 30131. m006850 的 2 个基因成功拼接, 最终获得完整的 RcWRKY14 基因。该基因编码的蛋白含有 1 个保守的 WRKY 结构域, 锌指结构为 C₂H₂ 型, 与拟南芥中的 AtWRKY40 相似性最高, 可归为 II a 类转录因子。启动子分析表明, RcWRKY14 至少可以在叶肉细胞、种子、花粉、节、根、根毛等组织或细胞中表达, 这为后面的表达谱分析所证实。此外, Rc-

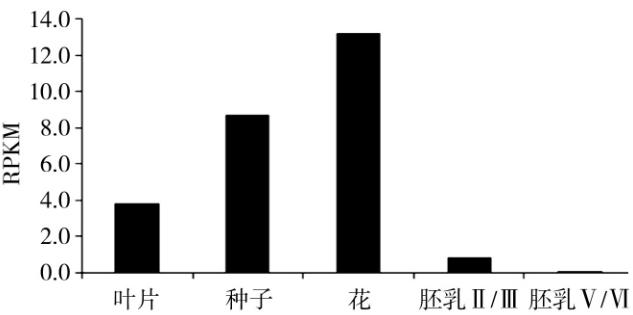


图 7 RcWRKY14 在蓖麻不同组织中的表达变化

WRKY14 可能还受脱落酸等多种激素和干旱等多种胁迫因子诱导,研究结果初步揭示出蓖麻 *RcWRKY14* 基因具有丰富的生物学功能。在众多转录因子结合元件中,预测的 Dof 结合位点最多(44 个)(见表 2),其确切作用还有待深入研究。值得注意的是,EST 比对显示 *RcWRKY14* 存在可变剪接(见图 1),然而,由于该转录本预测的蛋白并不包含 WRKY 结构域,因而该可变剪接转录本是否能翻译成功能性的蛋白还有待进一步的实验证实。需要指出的是,电子克隆技术只是基于已有的基因组序列进行的生物信息学分析,不能替代实验室实验。因此,电子克隆得到的 *RcWRKY14* 序列仍需后续 PCR 实验验证。总之,对蓖麻 *RcWRKY14* 基因的鉴别与生物信息学分析,能为验证蓖麻 *RcWRKY14* 基因和进一步研究该基因在蓖麻耐胁迫应答中的调控功能提供参考。

参考文献:

- [1] 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.
- [2] RUSHTON P J, SOMSSICH I E, RINGLER P, et al. WRKY transcription factors [J]. Trends Plant Sci, 2010, 15(5): 247–258.
- [3] ZHANG Y, WANG L. The WRKY transcription factor superfamily: its origin in eukaryotes and expansion in plants [J]. BMC Evol Biol, 2005(5): 1.
- [4] ISHIGURO S, NAKAMURA K. Characterization of a cDNA encoding a novel DNA-binding protein, SPF1, that recognizes SP8 sequences in the 5' upstream regions of genes coding for sporamin and beta-amylase from sweet potato [J]. Mol Gen Genet, 1994, 244(6): 563–571.
- [5] WU K L, GUO Z J, WANG H H, et al. The WRKY family of transcription factors in rice and Arabidopsis and their origins [J]. DNA Res, 2005, 12(1): 9–26.
- [6] LKER B, SOMSSICH I E. WRKY transcription factors: from DNA binding towards biological function [J]. Curr Opin Plant Biol, 2004, 7(5): 491–498.
- [7] EULGEM T, SOMSSICH I E. Networks of WRKY transcription factors in defense signaling [J]. Curr Opin Plant Biol, 2007, 10(4): 366–371.
- [8] PANDEY S P, SOMSSICH I E. The role of WRKY transcription factors in plant immunity [J]. Plant Physiol, 2009, 150(4): 1648–1655.
- [9] CHEN L, SONG Y, LI S, et al. The role of WRKY transcription factors in plant abiotic stresses [J]. Biochim Biophys Acta, 2012, 1819(2): 120–128.
- [10] 陈儒钢, 巩振辉, 逯明辉, 等. 植物抗逆反应中的转录因子网络研究进展 [J]. 农业生物技术学报, 2010, 18(1): 126–134.
- [11] 李冉, 姜永根. 植物中逆境反应相关的 WRKY 转录因子研究进展 [J]. 生态学报, 2011, 31(11): 3223–3231.
- [12] 王磊, 高晓清, 朱苓华, 等. 植物 WRKY 转录因子家族基因抗病相关功能的研究进展 [J]. 植物遗传资源学报, 2011, 12(1): 80–85.
- [13] DA SILVA NDE L, MACIEL M R, BATISTELLA C B, et al. Optimization of biodiesel production from castor oil [J]. Appl Biochem Biotechnol, 2006, 129–132: 405–414.
- [14] 曾小龙. 蓖麻在逆境胁迫下的抗性及耐性机制研究进展 [J]. 中国农学通报, 2010, 26(4): 123–125.
- [15] CHAN A P, CRABTREE J, ZHAO Q, et al. Draft genome sequence of the oilseed species *Ricinus communis* [J]. Nat Biotechnol, 2010, 28(9): 951–956.
- [16] VAN DE LOO F J, BROUN P, TURNER S, et al. An oleate 12-hydroxylase from *Ricinus communis* L. is a fatty acyl desaturase homolog [J]. Proc Natl Acad Sci USA, 1995, 92: 6743–6747.
- [17] DOERING-SAAD C, NEWBURY H J, COULDRIDGE C E, et al. A phloem-enriched cDNA library from *Ricinus*: insights into phloem function [J]. J Exp Bot, 2006, 57: 3183–3193.
- [18] LU C F, WALLIS J G, BROWSE J. An analysis of expressed sequence tags of developing castor endosperm using a full-length cDNA library [J]. BMC Plant Biol, 2007(7): 42.
- [19] BROWN A P, KROON J T, SWARBRECK D, et al. Tissue-specific whole transcriptome sequencing in castor, directed at understanding triacylglycerol lipid biosynthetic pathways [J]. PLoS One, 2012, 7(2): e30100.
- [20] ALTSCHUL S F, MADDEN T L, SCHÄFFER A A, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs [J]. Nucleic Acids Res, 1997, 25(17): 3389–3402.
- [21] REESE M G. Application of a time-delay neural network to promoter annotation in the *Drosophila melanogaster* genome [J]. Comput Chem, 2001, 26(1): 51–56.

- [22] HIGO K ,UGAWA Y ,IWAMOTO M ,et al. Plant cis-acting regulatory DNA elements (PLACE) database: 1999 [J]. Nucleic Acids Res ,1999 ,27(1) : 297 – 300.
- [23] LESCOT M ,DEHAIS P ,THIJS G ,et al. Plant CARE ,a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences [J]. Nucleic Acids Res ,2002 ,30(1) : 325 – 327.
- [24] THOMPSON J D ,HIGGINS D G ,GIBSON T J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting ,position-specific gap penalties and weight matrix choice [J]. Nucleic Acids Res ,1994 ,22(22) : 4673 – 4680.
- [25] TAMURA K ,DUDLEY J ,NEI M ,et al. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0 [J]. Mol Biol Evol ,2007 ,24(8) : 1596 – 1599.
- [26] MARCHLER-BAUER A ,LU S ,ANDERSON J B ,et al. CDD: a Conserved Domain Database for the functional annotation of proteins [J]. Nucleic Acids Res ,2011 ,39: D225 – 229.
- [27] LETUNIC I ,DOERKS T ,BORK P. SMART 7: recent updates to the protein domain annotation resource [J]. Nucleic Acids Res ,2012 ,40: D302 – 305.
- [28] BAILEY T L ,WILLIAMS N ,MISLEH C ,et al. MEME: discovering and analyzing DNA and protein sequence motifs [J]. Nucleic Acids Res ,2006 ,34: W369 – 373.
- [29] SIGRIST C J ,CERUTTI L ,DE CASTRO E ,et al. PROSITE ,a protein domain database for functional characterization and annotation [J]. Nucleic Acids Res ,2010 ,38: D161 – 166.
- [30] HORTON P ,PARK K J ,OBAYASHI T ,et al. WOLF PSORT: protein localization predictor [J]. Nucleic Acids Res ,2007 ,35: W585 – 587.
- [31] BRAMEIER M ,KRINGS A ,MACCALLUM R M. NucPred-predicting nuclear localization of proteins [J]. Bioinformatics ,2007 ,23(9) : 1159 – 1160.
- [32] KENT W J. BLAT—the BLAST-like alignment tool [J]. Genome Res ,2002 ,12(4) : 656 – 664.
- [33] ADAMS M D ,KELLEY J M ,GOCAYNE J D ,et al. Complementary DNA sequencing: expressed sequence tags and human genome project [J]. Science ,1991 ,252(5013) : 1651 – 1656.

In silico Identification and Bioinformatic Analysis of Castor Bean *WRKY14* Gene

HUANG Qi-xing^{1,2} , GUO An-ping² , YANG Li-fu¹ , ZOU Zhi¹

(1. Key Laboratory of Biology and Genetic Resources of Rubber Tree ,Ministry of Agriculture/Rubber Research Institute ,Chinese Academy of Tropical Agricultural Sciences ,Danzhou 571737 ,China; 2. Institute of Tropical Biosciences and Biotechnology ,Chinese Academy of Tropical Agricultural Sciences/ Ministry of Agriculture Key Laboratory of Biology and Genetic Resources of Tropical Crops ,Haikou 571101 ,China)

Abstract: *WRKY* is a plant-specific transcription factor that could be a key regulator in the stress response processes of castor bean (*Ricinus communis* L.). Consulting the published castor bean genome information ,we corrected some gene annotation errors and successfully obtained castor bean *RcWRKY14* gene by *in silico* cloning. Bioinformatic analysis showed that *RcWRKY14* contains three introns ,encoding a peptide of 318 amino acids ,and that it is nuclear localized. *RcWRKY14* encodes an unstable and hydrophilic protein with a theoretical molecular weight of 35.49 kDa ,isoelectric point of 8.64 ,total average hydrophilic of -0.803 ,instability coefficient (II) of 46.88. The protein contains one *WRKY* domain with a pattern of WRKYGQK_X13_C_X5_C_X23_H_X1_H. The conservative intron is located between site 199 (K amino acid residue) and site 200 (V amino acid residue) in the *WRKY* domain of *RcWRKY14*. The protein could be classified into subclass a under class II. Further analysis showed that a variety of hormone and stress-responsive cis-acting regulatory elements were found in the promoter region of *RcWRKY14*. Gene expression analysis showed that *RcWRKY14* was expressed in castor bean leaves ,flowers ,endosperms and seeds ,with the highest expression level in flowers ,the lower in seeds and leaves and the lowest in the endosperms.

Key words: *Ricinus communis* L. ; *WRKY*; transcription factor; *in silico* cloning; bioinformatic analysis