

赤桉 *EcWRKY75* 基因的电子克隆 和生物信息学分析

王鹏良¹, 杨利平², 李铁英³, 王华宇², 高磊⁴, 覃雯霞³, 甘四明^{5,6}

(1. 广西林业科学研究院 广西优良用材林资源培育重点实验室, 广西南宁 530002; 2. 广西钦州市林业科学研究所 钦州市植物生物工程技术研究中心, 广西钦州 535000; 3. 广西国有三门江林场, 广西柳州 545006; 4. 中国科学院武汉植物园, 湖北武汉 430074; 5. 中国林业科学研究院 热带林业研究所, 广东广州 510520; 6. 林木遗传育种国家重点实验室, 北京 100091)

摘要: *WRKY75* 基因在植物的生长发育以及生物和非生物胁迫过程中发挥着重要作用, 对其遗传改良进行研究极为必要。以赤桉的 EST 序列为材料, 利用电子克隆技术克隆 1 个与拟南芥中 *WRKY75* 高度同源的 *WRKY* 转录因子, 命名为 *EcWRKY75*, 并对其开展生物信息学分析。结果表明: *EcWRKY75* 开放阅读框长为 579 bp, 编码 192 个氨基酸, 分子量为 21.395 KD, 等电点为 9.72; 该基因编码的蛋白包含 1 个 *WRKY* 保守结构域和 1 个 *C₂H₂* 的结构域, 属 Group II 成员; 该蛋白很可能在细胞核中起转录、转录调控或信号转导等作用。该基因的电子克隆和生物信息学分析为其实验克隆和功能分析提供可靠的参考依据。

关键词: 赤桉; *EcWRKY75*; 电子克隆; 生物信息学分析

中图分类号: S792.39

文献标识码: A

文章编号: 1002-2767(2014)12-0028-05

赤桉 (*Eucalyptus camaldulensis*) 是分布最广的桉树树种之一, 其木材颜色从淡红色至深红色, 结构细致, 极耐腐蚀, 是澳大利亚著名的枕木用材^[1]; 同时赤桉具有较好的抗寒性^[2]、抗旱性^[3]。因其对环境极强的适应能力, 被引种到世界各地, 成为全世界最广泛种植的桉树树种之一。

WRKY 是植物特有的最大超级基因家族之一^[5], 拥有众多的基因家族成员^[6-7], 参与众多生理代谢通路的调控^[8], 并起重要作用。*WRKY75* 是 *WRKY* 超级基因家族中的重要一员。Devaiah 等研究表明, 转录因子 *WRKY75* 在磷吸收中起至关重要的正向调控作用^[9-10], 同时也调控根的发育^[9,11]; Encinas-Villarejo 等^[12] 研究发现, *WRKY75* 作为重要元件参与草莓抗尖孢炭疽菌的防卫反应。Li 等^[13] 发现, *WRKY75* 对叶片衰老起正向调控作用。Chen 等^[14] 研究表明, *WRKY75* 的过量表达可以提高对草酸和核盘菌的抵抗能力。Ranjan 等^[15] 在研究草棉时发现 *WRKY75* 受干旱胁迫响应, 说明该基因也参与干旱调控的相关通路。这些研究均说明 *WRKY75* 基因在植物的生物和非

收稿日期: 2014-05-12

基金项目: 广西优良用材林资源培育重点实验室开放课题基金资助项目(12A0301); 国家高技术研究发展计划(863 计划)资助项目(2013AA102705); 钦州市科学研究和技术开发资助项目(20137003)

第一作者简介: 王鹏良(1978-), 男, 浙江省新昌县人, 博士, 工程师, 从事林木分子育种研究。E-mail: pengliang_wang@163.com。

通讯作者: 甘四明(1970-), 博士, 研究员, 从事林木分子育种研究。E-mail: siming_gan@126.com。

Study on Tissue Culture and Fast Propagation of *Lorpetalum chindense*

ZHANG Yan¹, LI Zhi-fang², HAN Xiang-feng², ZHOU Yong-qing²

(1. Department of Horticulture, Xinyang College of Agriculture and Forestry, Xinyang, Henan 464000; 2. Agricultural Science Institution of Foshan, Foshan, Guangdong 528145)

Abstract: In order to satisfy the demand for *Lorpetalum chindense* on production, the optimum medium of tissue culture and fast propagation were studied. The results showed that the optimum primary medium was MS+6-BA 1.0 mg·L⁻¹+NAA 0.5 mg·L⁻¹. The optimum multiplication medium was MS+6-BA 0.1 mg·L⁻¹+NAA 0.1 mg·L⁻¹. The optimum root medium was 1/2MS+IAA 2.5 mg·L⁻¹+0.1% AC(active carbon). The tissue culture bottle with air holes was conducive to the growth of seedlings and roots.

Key words: *Lorpetalum chindense*; tissue culture; fast propagation

生物逆境、生长发育和衰老等过程中起重要调控作用。

电子克隆(In silico cloning)是利用 EST 为探针在相应物种的 EST 数据库中进行搜索,对搜索到的 EST 序列进行拼接,再以拼接好的 contig 为探针,继续搜索 EST 文库,直到没有新的可供拼接的 EST 为止,利用 Blast 对完整的 contig 进行比对和开放阅读框(ORF)的搜索,并得到基因 cDNA 全长的过程^[16]。电子克隆是目前获得基因全长的常用手段^[17],然而,关于赤桉 *EcWRKY75* 基因的克隆和功能分析相关研究的报道较少见。因此,该研究利用生物信息学工具和赤桉的 EST 信息,对 *EcWRKY75* 基因进行电子克隆及基因序列和结构分析,并对该基因编码蛋白从氨基酸组成、理化性质、亚细胞定位、高级结构和功能分类等方面进行分析和预测,为 *EcWRKY75* 基因的实验克隆和功能分析奠定基础。

1 材料与方法

1.1 材料

58 584 条赤桉 EST 序列于 2013 年 12 月 5 日从 dbEST/GenBank(NCBI)中以 FASTA 格式下载。

1.2 方法

1.2.1 赤桉 *EcWRKY75* 基因全长 cDNA 的获得 利用 Trimmer 截去 EST 序列两端的载体序列,再利用编写的 Perl 脚本提取可能含有 WRKY 氨基酸的 EST 序列。再利用 NCBI 中 Blastx 对可能含有 WRKY 氨基酸的 EST 在非冗余蛋白序列库中进行比对;利用 CAP3 软件对获得的 EST 序列进行拼接;进一步利用 NCBI 中 ORF Finder 对含有 WRKY 保守结构域的 EST 搜索开放阅读框,得到 WRKY 转录因子全长基因。

1.2.2 *EcWRKY75* 基因编码的蛋白质生物信息学分析 利用 ExPASy 服务器中的 ProtParam、Compute pI/Mw 和 ProtScale 软件预测 *EcWRKY75* 蛋白的氨基酸组成、分子量、等电点、疏水性/亲水性、不稳定系数及脂肪系数等理化性质;利用在线软件 PSORT,预测 *EcWRKY75* 蛋白的位置;利用 SOPMA 预测该蛋白二级结构;利用 ExPASy 中 SWISS-MODEL 预测该蛋白的三维结构;利用 ProtFun 预测该基因的功能分类。

2 结果与分析

2.1 基因序列和结构分析

以 WRKY 保守域对应的 DNA 序列为探针,利用自编的 Perl 脚本搜索赤桉 EST 序列数据库,得到 110 条序列,并利用 CAP3 对获得的 EST 序列进行拼接,获得 19 个 contigs 和 35 个 singlets。利用 Blastx 对拼接获得的序列进行比对,对获得的 15 条包含 WRKY 结构域的序列利用 CAP3 进行再次拼接,对获得的 2 个 contig 序列再次进行 Blastx 注释,并进行 ORF 查找,获得 1 条具有完整开放阅读框的序列。

该序列长 690 bp,开放阅读框(44~622 bp)为 579 bp,以 ATG 为起始密码子,TGA 为终止密码子,编码 192 个氨基酸。其推导的氨基酸序列 N 端具有 1 个 WRKYGQK 的保守结构域,C 端有 1 个 C_{X4-5} C_{X22-23} HxH 的结构域(见图 1)。因此认为是 WRKY 基因家族 Group II 的成员。由于其与拟南芥 WRKY 家族成员中的 *AtWRKY75* 高度同源,故命名为 *EcWRKY75*。

2.2 *EcWRKY45* 蛋白理化性质分析

利用 ExPASy 中的 ProtParam 程序预测了 *EcWRKY75* 的氨基酸组成、分子量、等电点、疏水性/亲水性、不稳定系数及脂肪系数等理化性质。由表 1 可知,*EcWRKY75* 蛋白由 20 种氨基酸组成,其中 Ser(S)的含量最高为 10.4%,其次 Gly(G)的含量为 9.4%;Cys(C)和 Trp(W)的含量最低分别 1.0%和 0.5%。

表 1 *EcWRKY75* 的氨基酸组成

Table 1 Amino acid components of *EcWRKY75*

氨基酸 Amino acids	含量/% Content	氨基酸 Amino acids	含量/% Content
Ala(A)	6.8	Leu(L)	4.7
Arg(R)	7.3	Lys(K)	7.3
Asn(N)	2.1	Met(M)	2.6
Asp(D)	4.2	Phe(F)	5.2
Cys(C)	1.0	Pro(P)	6.2
Gln(Q)	6.2	Ser(S)	10.4
Glu(E)	5.2	Thr(T)	4.2
Gly(G)	9.4	Trp(W)	0.5
His(H)	3.6	Tyr(Y)	3.6
Ile(I)	3.6	Val(V)	5.7

```

1      ATTTCCTCTTTCATCCCTCATACGACCAACTGTTGCGACCTCATGAAAGACACCAGAT
                                         M E R H Q M
61     GTTGTTCCTCGCTCTCGGCAGATGTCAGGCCGTCCTCGGCGCAGCCGCGCTGACCTCCTT
      L F P L S A D V R P S P A Q P P L T S F
121    CCGGTTTCGCGTTCGAAGCCGAGGTTCTCGTGTCCAGAGCAGCGCGGCTTCGGCAGTGA
      P F A F E A G G S R V Q S S A G F G S D
181    TCTGATGAGCTCGTCCGATGAGGTTCTGGGATGAGCGGCGCGAAGACGGGATCGC
      L M S S S V H G F L G L S G A E D G I A
241    CCAGAGATCAAAGCCGAGTACGATGCTGCTCAGGGAGCTCGGTATCCGGGGTCTGTA
      Q R S K A E Y D A A H G S S V S G G A E
301    GAAAATTGGGAAGAAGAAGGGGAGAAGAAGTCCGAAGGCCGAGAATCGCTTTCAGAC
      K I G K K K G E K K V R R P R I A F Q T
361    CCGGAGTCAGGTGGACATACTGGACGATGTTATCGATGGAGGAAGTATGGACAGAAAGC
      R S Q V D I L D D G Y R W R K Y G Q K A
421    AGTCAAGAACAACAGATTCCAGAAAGTACTATCGCTGCACGATCCGGGATGCAATGT
      V K N N R F P R S Y Y R C T H P G C N V
481    AAAGAAACAAGTCCAGAGGCTAAGCAAAGACGAGGCGATGGTGGTGACCACTACGAAGG
      K K Q V Q R L S K D E G M V V T T Y E G
541    GACTCACACCCATCCCATCCAAAATCGACTGAGAATTGGAACACATCTTGAGTCAGAT
      T H T H P I Q K S T E N F E H I L S Q M
601    GCAGATCTACCTCCCTTTTCATTGACCTAGAATCTGCTGAATCTTAAACCAATCATC
      Q I Y P P F *
661    ATGCTTAGCTGCAAGCACATTGTAAGAG
    
```

图1 *EcWRKY75* 基因 cDNA 序列及其推导的氨基酸序列

* 为终止密码子。

Fig. 1 cDNA sequence and putative amino acid sequence of *EcWRKY75*

* is termination codon.

由表 2 可知,该蛋白的分子量为 21.395 KD,等电点为 9.72,其不稳定系数为 48.45,大于 40,说明该蛋白不稳定;其平均疏水性为-0.725,表明该蛋白为水溶性蛋白。

表 2 *EcWRKY75* 氨基酸一级结构特征的预测结果

Table 2 Primary structure of amino acid of *EcWRKY75*

EcWRKY75 一级结构特征 Characteristics of primary structure	预测结果 Predicting outcomes
氨基酸数目/个 Amino acids number	192
分子量/KD Mw	21.395
等电点 pI	9.72
负电荷残基 Asp+Glu	18
正电荷残基 Arg+Lys	28
不稳定系数 II	48.45
脂肪系数 AI	55.89
平均疏水性 GRAVY	-0.725

疏水作用是蛋白高级结构构成的一种重要的

作用力,从疏水性方面也可以推断出该蛋白的不稳定性。由图 2 分析可知,在该蛋白的 N 端部分表现出疏水性,从第 60 位的氨基酸以后基本都表现出亲水性,且该蛋白的平均疏水性为-0.725,因此该蛋白整体表现为亲水性。

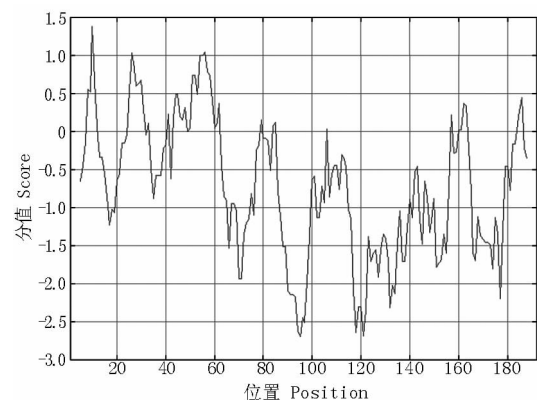


图 2 *EcWRKY75* 氨基酸序列的疏/亲水性分析

Fig. 2 Predicted hydrophobicity/hydrophilicity of putative amino acid sequence of *EcWRKY75*

2.3 *EcWRKY75* 蛋白的高级结构

利用 SOPMA 程序预测了 *EcWRKY75* 蛋白二级结构,结果表明,*EcWRKY75* 蛋白由 α 螺旋、

延伸链、 β 折叠和无规则卷曲四种结构组成,无规则卷曲的比例最高为 50%,其次为 α 螺旋(25.52%), β 折叠的比例最低为 8.33%(见表 3)。利用 ExPASy 中 SWISS-MODEL 程序预测了其蛋白的三级结构(见图 3)。

表 3 *EcWRKY75* 蛋白二级结构预测结果

Table 3 Secondly structure of *EcWRKY75*

二级结构类型 Secondly structure	氨基酸数量 Amino acids number	百分比/% Proportion
α 螺旋 α helix	49	25.52
延伸链 Extending chain	31	16.15
β 折叠 β folding	16	8.33
无规则卷曲 Random curl	96	50.00

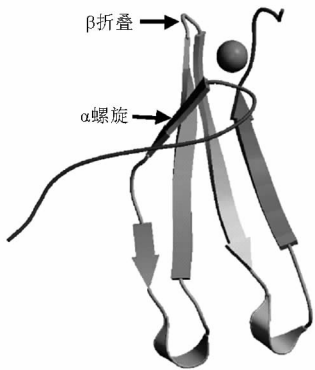


图 3 *EcWRKY75* 蛋白三级结构预测

Fig. 3 Predicted tertiary structure of *EcWRKY75* protein

2.4 *EcWRKY75* 蛋白的亚细胞定位

基因编码的蛋白质在细胞中的分布在一定程度上反映了该基因的功能。研究采用 PSPORT 程序预测 *EcWRKY75* 的可能分布位置,从表 4 可以看出,该蛋白分布微体和核内的可能性最大,分别为 0.462 和 0.300,而线粒体基质空间与内质网膜的可能性则为 0.100 和 0。

表 4 基于 PSPORT 预测的

EcWRKY75 亚细胞定位

Table 4 Subcellular localization of *EcWRKY75* predicted by PSPORT

亚细胞定位 Subcellular location	可能性 Probability
微体 Micro body	0.462
核 Nuclear	0.300
线粒体基质空间 Mitochondrial matrix	0.100
内质网膜 Endoplasmic reticulum membrane	0

2.5 *EcWRKY75* 蛋白的功能分类

研究利用 ProtFun 程序对其进行了功能分类,结果表明,*EcWRKY75* 最重要的功能可能是转录、信号转导和转录调控,其概率分别为 0.150、0.129 和 0.123,这也反映了该基因可能是转录因子(见表 5)。

表 5 *EcWRKY75* 功能分类结果

Table 5 Function prediction of *EcWRKY75*

序号 No.	功能分类 Function	概率 Probability
1	信号转导	0.129
2	受体	0.007
3	激素	0.002
4	结构蛋白	0.005
5	转运体	0.025
6	离子通道	0.009
7	电压门控离子通道	0.005
8	阳离子通道	0.010
9	转录	0.150
10	转录调控	0.123
11	胁迫应答	0.008
12	免疫应答	0.015
13	生长因子	0.018
14	金属离子转运	0.009

3 结论与讨论

随着基因组计划和 EST 计划的深入开展,测序技术不断改进,许多物种的基因组测序工作已经完成,产生了大量的数据信息,基因组研究从此由结构研究转入到功能研究。获得基因全长是功能研究的必要前提。在海量数据的基础上,电子克隆相较于图位克隆、cDNA 全长基因筛选和 RACE 等方法更加快速,高效,成本更低廉,更具针对性^[18]。

相关研究表明,*WRKY75* 基因在植物的生物胁迫和非生物胁迫、生长发育及衰老过程中的重要作用,而 *WRKY75* 基因的克隆和功能研究较少,该研究针对赤桉的 *WRKY75* 进行电子克隆;对基因结构、蛋白质的理化性质、一级结构及高级结构进行了分析;预测 *EcWRKY75* 蛋白分布在细胞核中,这与 *WRKY75* 的亚细胞定位一致^[19];并预测了该基因在细胞核中具有转录、转录调控及信号转导等功能。该研究对 *EcWRKY75* 的结构和功能进行了初步分析,为 *EcWRKY75* 的实验克隆和功能分析提供了参考依据,也为进一步结合桉树高效遗传转化体系^[19] 实现桉树分子遗传改良提供优良基因资源。

参考文献:

- [1] 祁述雄. 中国桉树[M]. 北京: 中国林业出版社, 1989.
- [2] 蓝贺胜. 福建省赤桉种源/家系的试验研究[J]. 广西林业科学, 2006, 35(2): 57-61.
- [3] 吴良栋, 陈强, 余青云, 等. 福建南平引种赤桉和邓恩桉试验[J]. 林业科技开发, 2004, 18(6): 20-22.
- [4] 王懿然. 桉树生物学概论[M]. 北京: 科学出版社, 2010.
- [5] Eulgem T, Rushton P J, Robatzek S, et al. The WRKY superfamily of plant transcription factors[J]. Trends Plant Sciences, 2000, 5(5): 199-206.
- [6] Tripathi P, Rabara R C, Langum T J, et al. The WRKY transcription factor family in Brachypodium distachyon[J]. BMC Genomics, 2012, 13: 270-290.
- [7] 仇玉萍, 荆邵娟, 付坚, 等. 13 个水稻 WRKY 基因的克隆及其表达谱分析[J]. 科学通报, 2004, 49(18): 1860-1869.
- [8] Rushton P J, Somssich I E, Ringler P, et al. WRKY transcription factors[J]. Trends Plant Sciences, 2010, 15(5): 247-258.
- [9] Devaiah B N, Karthikeyan A S, Raghothama K G. WRKY75 transcription factor is a modulator of phosphate acquisition and root development in Arabidopsis[J]. Plant Physiology, 2007, 143(4): 1789-1801.
- [10] Devaiah B N, Raghothama K G. Transcriptional Regulation of Pi Starvation Responses by WRKY75[J]. Plant Signal Behav, 2007, 2(5): 424-425.
- [11] Rishmawi L, Pesch M, Juengst C, et al. Non-Cell-Autonomous regulation of root hair patterning genes by WRKY75 in Arabidopsis [J]. Plant Physiology, 2014, 165(1): 186-195.
- [12] Encinas-Villarejo S, Maldonado A M, Amil-Ruiz F, et al. Evidence for a positive regulatory role of strawberry (Fragaria x ananassa) Fa WRKY1 and Arabidopsis At WRKY75 proteins in resistance[J]. Journal of Experimental Botany, 2009, 60(11): 3043-3065.
- [13] Li Z, Peng J, Wen X, et al. Gene network analysis and functional studies of senescence-associated genes reveal novel regulators of Arabidopsis leaf senescence[J]. Journal of Integrative Plant Biology, 2012, 54(8): 526-539.
- [14] 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 Reports, 2013, 32(10): 1589-1599.
- [15] Ranjan A, Pandey N, Lakhwani D, et al. Comparative transcriptomic analysis of roots of contrasting Gossypium herbaceum genotypes revealing adaptation to drought [J]. BMC Genomics, 2012, 13: 680-702.
- [16] 黄骥, 张红生, 曹雅君, 等. 水稻功能基因的电子克隆策略[J]. 中国水稻科学, 2002, 16(4): 295-298.
- [17] 胡铄, 萧浪涛. 生物信息学在新基因全长 cDNA 电子克隆中的应用[J]. 生物技术通报, 2007(4): 93-96.
- [18] 孔祥瑞. 茶树功能候选基因电子克隆的可行性分析[J]. 分子植物育种, 2014, 12(2): 332-337.
- [19] 王鹏良, 姜福星, 蔡玲, 等. 广林巨尾桉 9 号遗传转化体系的建立[J]. 林业科技开发, 2013, 27(3): 76-80.

In Silico Cloning and Bioinformatic Analysis of *EcWRKY75* in *Eucalyptus camaldulensis*

WANG Peng-liang¹, YANG Li-ping², LI Tie-ying³, WANG Hua-yu², GAO Lei⁴, QIN Wen-xia³, GAN Si-ming^{5,6}

(1. Guangxi Forestry Research Institute, Guangxi Key Laboratory of Superior Timber Trees Resource Cultivation, Nanning, Guangxi 530002; 2. Qinzhou Forestry Research Institute, Qinzhou Plant Biological Engineering Technology Research Center, Qinzhou, Guangxi 535000; 3. Guangxi State-owned Sanmenjiang Forestry Farm, Liuzhou, Guangxi 545006; 4. Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan, Hubei 430074; 5. Research Institute of Tropical Forestry, Chinese Academy of Forestry, Guangzhou, Guangdong 510520; 6. State Key Laboratory of Tree Genetics and Breeding, Beijing 100091)

Abstract: The gene *WRKY75* plays key roles in the development, abiotic stress and biotic stress of plants, it is essential to study. Based on the EST sequences of *Eucalyptus camaldulensis*, *EcWRKY75* was cloned in silico, and its bioinformatic analysis was carried out. The results indicated that *EcWRKY75* with a ORF of 579 bp encoded 192 amino acid with the molecular weight of 21.395 KD and the isoelectric point of 9.72; the encoded protein belonged to the Group II with a conserved domain of WRKYGQK and a domain of C₂H₂, the protein participated in transcription, transcription regulation or signal transduction in the nuclei, which laid a foundation for the identification and functional analysis of *EcWRKY75*.

Key words: *Eucalyptus camaldulensis*; *EcWRKY75*; in silico cloning; bioinformatic analysis