

巴氏杜氏藻 ζ -胡萝卜素脱氢酶基因的克隆与分析

肖岚, 劳永民, 姜建国

(华南理工大学生物科学与工程学院, 广东广州 510006)

摘要: ζ -胡萝卜素脱氢酶 (*ZDS*) 基因是胡萝卜素生物合成途径中的一个关键酶之一。根据实验室已获得的巴氏杜氏藻 *ZDS* 基因的 cDNA 序列, 设计引物, 通过分段 PCR 的方法, 获得 *ZDS* 基因的编码区序列。然后根据获得编码区序列, 利用染色体步移的方法获得其两端侧翼序列: 启动子和终止子, 并利用生物信息学工具对获得序列进行分析。实验获得的完整 *ZDS* 基因全长 11896 bp, 其中编码区序列 (从 "ATG" 到 "TAA") 长度为 6435, 编码区上游序列 4091 bp (包括 33 bp 的 5'UTR 序列), 编码区下游序列 1370 bp (包括 411 bp 的 3'UTR 序列)。将编码序列与 cDNA (开放阅读框) 进行比对, 发现整个编码区序列含有 12 个外显子和 11 个内含子, 其中内含子长度达 4686 bp, 约为外显子长度的 2.68 倍, 内含子均以 GT 开始, AG 收尾, 属于最常见的内含子类型。通过生物信息学分析, 发现 *ZDS* 启动子中具有多种转录因子结合位点, 包括与光调控有关的 GAGA box, ASF1; 与植物黄化反应有光的 ACGT box 等。

关键词: 巴氏藻; ζ -胡萝卜素脱氢酶; 编码区; 启动子; 生物信息学

文章编号: 1673-9078(2012)7-764-767

Clone and Analysis ζ - Carotene Desaturase of *Dunaliella bardawil*

XIAO Lan, LAO Yong-min, JIANG Jian-guo

(School of Bioscience and Bioengineering, South China University of Technology, Guangzhou 510006, China)

Abstract: ζ -Carotene desaturase (*ZDS*) is one of the key enzymes in the carotenoid biosynthetic pathway. According to the full-length sequence of the *ZDS* cDNA in *D. bardawil* that our laboratory has obtained before, the coding region sequence of *ZDS* were isolated by the approach of PCR and then flanking sequence of the coding region were amplified by the method of genome walking, including promoter and terminator of *ZDS*, followed by analysis of these sequences by bioinformatics tools. The whole sequence of *ZDS* was 11896 bp. It contained 6435 bp coding region (from ATG to TAA), 4091 bp upstream of the coding region, and 1370 downstream of the coding region. Blastn the coding region and ORF of *ZDS* showed that there were 12 exons and 11 introns in coding region. The total length of the introns was 4086 bp, about 2.68 fold of the total length of exons (1749bp). Each intron started with GT and end with AG, belonging to the most common style of exon. By analyzing the promoter of *ZDS* gene, we found that there are many transcription factor binding sites, including "GAGA box", ASF1, related with plant light regulation; "ACGT box", required for etiolation-induced expression of early responsive to dehydration.

Key words: *Dunaliella bardawil*; ζ -carotene desaturase; coding region; promoter; bioinformatics

ζ -胡萝卜素脱氢酶 (*ZDS*) 是植物类胡萝卜素代谢途径中的一个关键酶。目前, 一些植物以及微藻的 *ZDS* 基因的 cDNA 序列被克隆并研究, 重点探索其在类胡萝卜素代谢途径中的作用。Rodrigo MJ 等^[1]研究果实 (甜橙) 成熟过程中果皮与叶子中的 *PSY*、*PDS*、*ZDS* 基因对类胡萝卜素积累的作用, 发现在果实成熟时, *PDS* 和 *ZDS* 基因表达量的增加会促进类胡萝卜素的积累。Jing Ji 等^[2]转基因烟草中过表达黄龙胆的 *ZDS* 基因, 使得烟叶中 β -胡萝卜素的含量提高了 91%, 烟花中的含量提升了 49%。

本课题组一直以来都以杜氏藻 (包括 *D. salina* 和 *D. bardawil*) 为研究对象, 重点研究类胡萝卜素代谢

途径的各种酶基因的克隆以及对其代谢调控机理的研究。本实验中, 以巴氏杜氏藻 (*D. bardawil*) 为研究对象, 利用实验室已经克隆得到的 *D. bardawil* 的 *ZDS* 基因 cDNA 序列, 克隆并研究 *ZDS* 基因: 利用 *ZDS* 的 cDNA 序列克隆基因的编码区序列; 利用染色体步移技术扩增已 *ZDS* 编码区序列的侧翼调控序列; 利用生物信息学工具分析 *ZDS* 基因, 包括 *ZDS* 基因编码区序列的组成、启动子转录起始位点的预测及其转录因子结合位点的分析等。

1 料与方法

1.1 材料

1.1.1 藻种、菌种与载体

巴氏杜氏藻 (*Dunaliella bardawil*) 购于中国科学院水生生物研究所 (武汉), 由实验室保存, 盐藻培养

收稿日期: 2012-05-21

基金项目: 国家自然科学基金项目 (31171631)

通讯作者: 姜建国

参照 Ben-Amotz^[3]和周世水^[4]的方法; 宿主菌为大肠杆菌 *E. coli* GT-116, 由实验室保存; 克隆载体为 PCR 2.1 vector 购自 Invitrogen 公司。

1.1.2 主要试剂

基因组提取试剂盒 Universal Genomic DNA Extraction Kit Ver.3.0, 基因组步移试剂盒 Genome Walking Kit, 连接试剂盒 DNA Ligation Kit Ver.2.0, 以及各种 PCR Enzymes 和 DNA Markers 购自 TaKaRa 公司; 胶回收试剂盒 E.Z.N.A.TM Gel Extraction Kit, 核酸纯化试剂盒 E.Z.N.A. Cycle-PureKit、质粒提取试剂盒 E.Z.N.A. TM Plasmid Mini Kit 购自 Omega 公司; TA 克隆载体 PCR 2.1 Vector 购自 Invitrogen 公司

1.2 ZDS 编码区序列的获得

表 1 获得编码区序列的引物设计

Table 1 Primer design for getting the coding sequence

引物名	引物序列
CR-1S	5'-"CACTCGCACCAGAATAAGG"-3'
CR-1A	5'-"CAGCCAGAGGAAACACAC"-3'
CR-2S	5'-"ATGCGTGTGCGCCATTGTG"-3'
CR-2A	5'-"GTAGCACCCGAAGAAGAC"-3'
CR-3S	5'-"TGTCAGTCTCAACAAAGG"-3'
CR-3A	5'-"GATGGTCAACATTCAAGCG"-3'
CR-4S	5'-"GGCGAGTCTAAGCAATCC"-3'
CR-4A	5'-"CGAAGCAGGAGACGAATG"-3'
CR-5S	5'-"AGAGCAGCAAAACAGTATCC"-3'
CR-5A	5'-"GTAGTCTTGCTTGGTGTAG"-3'
CR-6S	5'-"ATCAACCTTTTTGCCCTGTGTGC"-3'
CR-6A	5'-"GGTATCTCGTCACCAATAGAAACAGC"-3'

由于 ZDS 基因编码区序列由内含子和外显子组成, 引物设计应尽量不要跨越内含子。NCBI 搜索与杜氏盐藻亲缘性较近的物种, 利用 clustalX 软件比对它们的 ZDS 氨基酸序列, 我们可以得到一些相对保守的区域。根据这些保守的氨基酸序列找出对应的 cDNA 序列, 并设计引物, 可提高实验的成功率, 设计引物参见表 1。实验中采用分段 PCR 的方法(图 1), 即先得到编码区的一段序列, 然后根据该序列和 cDNA 序列设计引物向两端扩增, 并利用两段序列的重叠区域进行巢式 PCR 反应, 排除非特异性条带, 分段获得 ZDS 基因的编码区序列。最后将获得序列进行拼接从而得到完整的 ZDS 基因编码区序列。

1.3 ZDS 启动子与终止子的获得

实验采用使用 Takara 公司的染色体步移试剂盒 Genome Walking Kit, 运用特异引物 PCR 的步移方法, 扩增编码区序列两端的未知序列: 启动子和终止子。根据已经获得的 DbZDS 基因编码区的序列, 分别设

计两组同向且退火温度较高的特异性引物 (SP 引物, 参见表 2), 利用三次巢式 PCR 反应从两端克隆 ZDS 基因的启动子和终止子。取 1st、2nd、3rd PCR 反应液各 5 μL, 使用 1% 的琼脂糖凝胶进行电泳。切胶回收 SP3 清晰的电泳条带, 以 SP3 Primer 为引物对 PCR 产物进行 DNA 测序。

表 2 获取启动子与终止子的步移引物

Table 2 Primer walking for getting the promoters and terminators

引物名称	引物序列
启动子	P-SP1 5'-"GCAGGTAGAGCAAGCGGTTCG"-3'
	P-SP2 5'-"GAAGACGAGGTGGGGAACGA"-3'
	P-SP3 5'-"CCGGCAGACTCTTCCCTTAG"-3'
终止子	T-SP1 5'-"CCAATGACCTACCAAGTCGCCT"-3'
	T-SP2 5'-"CGCATCGCCTGAAGACTAGAAG"-3'
	T-SP3 5'-"TGTGATGCTACGGAGCCTTGAG"-3'

1.4 生物信息学分析

DbZDS 基因 cDNA 开放阅读框分析与核酸翻译: EditSeq; 保守结构域分析: NCBI_Blastn, NCBI_CDD (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>); 序列相似性搜索与编码区序列分析: NCBI_Blast (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>); 多重序列比对与系统发育树构建: Clustal X 1.8 与 MEGA 4.0; 启动子转录因子结合位点分析: PlantPAN (<http://plantpan.mbc.nctu.edu.tw/>)

2 结果

2.1 ZDS 编码区序列

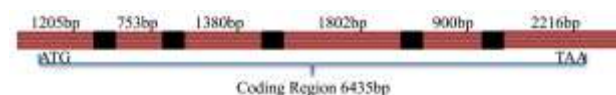


图 1 ZDS 基因编码区序列的获得

Fig.1 Obtainment of ZDS gene coding region

通过 6 次 PCR 反应得到 6 条核苷酸片段, 参见图 1, 图中黑色部分为两片段的重叠区域。其中第一个核酸片段 (1205 bp) 出现单个引物克隆出一段序列的情况, 即在测序结果中出现前端序列与末端的引物序列完全互补的情况。为了验证这个结论, 利用单个引物 CR-1A 进行 PCR 反应, 得到了相同的条带和测序结果。将该 1205 bp 序列与 cDNA 比对, 发现该序列超出 ZDS 基因上游起始密码子 ATG 390bp (包含 33 bp 的 5'-UTR), 其与 cDNA 序列的重叠部分包含两个内含子和外显子。第六个片段扩增到终止密码子下游 411 bp, 且为 3'-UTR 序列。将得到的 6 段序列拼接, 得到完整的 ZDS 基因编码区序列, 长度为 6435 bp。

2.2 DbZDS 启动子与终止子

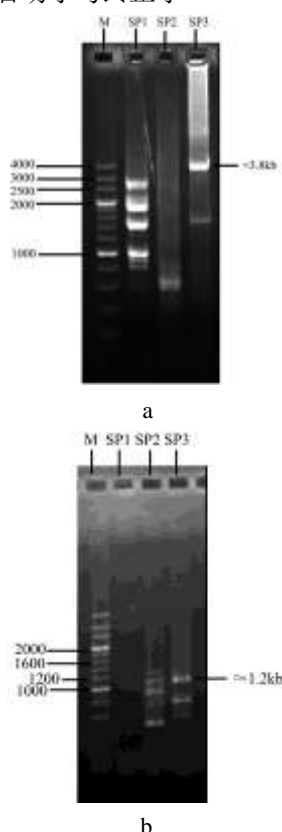


图 2 ZDS 启动子与终止子的获得

Fig.2 Obtainment of promoter and terminator of ZDS gene

注：a：启动子基因组步移；b：终止子基因组步移。

电泳结果参见图 2，图 2(a)图为启动子三条特异引物 PCR 结果，SP3 引物 PCR 得到大小分别为 3.8 kb 和 1.5 kb 左右的核酸片段。测序结果反馈后，发现 1.5 kb 和 3.8 kb 片段的序列与编码区序列存在 37 bp 的重叠区域，而 1.5 kb 片段是 3.8 kb 片段的部分；图 2(b) 图为终止子三条特异引物的 PCR 结果，SP3 引物 PCR 得到大小为 1.1 Kb、780 bp、500 bp 左右的核酸片段，胶回收 1.1 Kb 和 780 bp 条带并测序，1.1 kb 条带与已知序列存在 121 bp 的重叠区域。

2.3 生物信息学分析

2.3.1 ZDS 完整基因序列分析



图 3 ZDS 全基因结构

Fig.3 The structure of full-length genomic DNA of ZDS

将获得的 ZDS 基因的编码区序列与启动子和终止子序列拼接，获得完整的 *D. bardawil* 的 ZDS 基因序列，序列全长 11896 bp，其中编码区序列长度为 6435 bp，编码区序列上游 4091 bp (ATG 上游含 33 bp 的

5'-UTR)，编码区下游序列长度为 1370 bp (TAA 下游，含 411 bp 的 3'-UTR 序列)。将其与 ZDS 基因的 cDNA 序列进行比对，发现编码区含有 12 个外显子和 11 个内含子；外显子长度为 1749 bp，内含子长度达 4686 bp，约为外显子长度的 2.68 倍；内含子均以 GT 开始，AG 收尾，属于最常见的内含子类型。

表 3 ZDS 基因启动子转录因子结合位点分析结果

Table 3 Transcription factor binding site in the ZDS gene

promoter				
转录因子	特征序列	转录因子描述	物种	参考文献
GATA box-127	GATA	光调控，组织特异性表达	拟南芥	[5]
CDC5	NGCTCAGCGCN	转录激活因子	拟南芥	[6]
Dof2	NNNWAAAGCNN	Dof 蛋白结合位点，调控转录	玉米	[7]
ACGT	ACGT	干旱诱导黄化表达	拟南芥	[8]
W box	TTGAC	被水杨酸诱导产生的 WRKY DNA 结合蛋白特异性识别，	拟南芥	[9]
T Box	ACTTTG	“T box”突变会导致光诱导基因转录水平下降	拟南芥	[10]

2.3.2 DbZDS 保守结构域分析

利用 NCBI 对 ZDS 蛋白的保守结构域进行分析，ZDS 中具有一个含 Rossmann 折叠的 NAD(P)结合结构域 (92-158 氨基酸)；1,5-二磷酸核酮糖合成酶结构域，1,5-二磷酸核酮糖是在光合作用中卡尔文循环里起重要作用的一种五碳糖；八氢番茄红素脱氢酶结构域；与氨基酸转运与代谢有关的单胺氧化酶、甘氨酸/D 型氨基酸氧化酶结构域等。大部分结构域位于 ZDS 蛋白的 50 至 160 个氨基酸之间。

2.3.3 ZDS 启动子生物信息学分析

选定 5 个物种 (Arabidopsis、Bean、Potato、Tobacco 及 other)，利用 PlantPAN 在线服务器检索 ZDS 基因启动子转录因子结合位点。发现 ZDS 基因的启动子序列中存在多种调控元件，包括与光调控、组织特异性表达有关的 GATA 盒与 ASF1 结合位点；干旱诱导植物黄化反应有关的“ACTG”盒；调控转录的 Dof 蛋白结合位点 Dof 2 等，表 3 中列出了部分可能存在的转录因子结合位点。

3 结论

3.1 巴氏杜氏藻 (*D. bardawil*) 和杜氏盐藻 (*D. salina*)

是目前商业化开发最成功的两种杜氏藻,这得益于它们在胁迫条件下,对 β -胡萝卜素的富集能力,最高可达细胞干重的14%。近些年来关于杜氏藻基因的研究越来越广泛,而对于杜氏藻类胡萝卜素代谢途径各种酶的基因的研究相对较少。

3.2 实验重点研究 *D. bardawil* 藻类胡萝卜素生物合成途径中的一个关键酶—ZDS的基因。实验中利用分段PCR,巢式PCR以及序列拼接的方法获得ZDS基因完整的编码区序列,对于一些内含子多且长的基因来说,分段PCR的方法可以避免实验的盲目性,提高实验的成功率。染色体步移技术可以有效获取与已知序列相邻的未知序列,主要应用于启动子与终止子的克隆、步移获得新物种中基因的非保守区域、鉴定T-DNA或转座子的插入位点、染色体测序工作中的空隙填补等方面^[1]。本实验利用特异引物染色体步移的方法扩增得到ZDS基因的启动子和终止子,并对获得的完整ZDS基因进行生物信息学分析,包括ZDS基因编码区序列结构,ZDS蛋白的保守结构域,ZDS基因启动子的转录因子结合位点分析。

参考文献

- [1] Rodrigo MJ, Marcos JF, Zacarías L. Biochemical and molecular analysis of carotenoid biosynthesis in flavedo of orange (*Citrus sinensis* L.) during fruit development and maturation [J]. J. Agric. Food Chem., 2004, 52: 6724-6731
- [2] Ji J, Wang G, Wang JH, et al. Functional analysis of multiple carotenogenic genes from *Lycium barbarum* and *Gentiana lutea* L. for their effects on β -carotene production in transgenic tobacco [J]. Biotechnol Lett, 2009, 31: 305-312
- [3] Ben-Amotz A, Avron M. The biotechnology of cultivating the halotolerant alga *Dunaliella* [J]. Trends Biotechnol, 2002, 18(6): 257-263
- [4] 周世水,姜建国,林炜铁,等.盐藻生长及其 β -胡萝卜素累积的作用因子和最适条件探讨[J].食品与发酵工业,2002,28(11): 1-3
- [5] Lam E, Chua NH. ASF-2: A factor that binds to the cauliflower mosaic virus 35S promoter and a conserved GATA motif in cab promoters [J]. Plant Cell, 1989, 1: 1147-1156
- [6] Hirayama T, Shinozaki K. A cdc5+ homolog of a higher plant, *Arabidopsis thaliana* [J]. Proceedings of the National Academy of Sciences, 1996, 93: 13371-13376
- [7] Yanagisawa S, Schmidt RJ. Diversity and similarity among recognition sequences of Dof transcription factors [J]. The Plant Journal, 1999, 17: 209-214
- [8] Simpson SD, Nakashima K, Narusaka Y, et al. Two different novel cis-acting elements of *erd1*, a *clpA* homologous *Arabidopsis* gene function in induction by dehydration stress and dark-induced senescence [J]. The Plant Journal, 2003, 33: 259-270
- [9] Yu D, Chen C, Chen Z. Evidence for an important role of WRKY DNA binding proteins in the regulation of NPR1 gene expression [J]. Plant Cell, 2001, 13: 1527-1540
- [10] Chan CS, Guo L, Shih MC. Promoter analysis of the nuclear gene encoding the chloroplast glyceraldehyde-3-phosphate dehydrogenase B subunit of *Arabidopsis thaliana* [J]. Plant Molecular Biology, 2001, 46: 131-141
- [11] 刘博,苏乔,汤敏谦,等.应用于染色体步移的PCR扩增技术的研究进展[J].遗传,2006,28(5):587-595