

# 雨生红球藻 $\beta$ -胡萝卜素酮化酶基因的 cDNA 和基因组 DNA 克隆及序列分析

滕长英<sup>1</sup>, 张立<sup>1</sup>, 秦松<sup>2</sup>, 曾呈奎<sup>2</sup>

(1. 烟台师范学院 生命科学学院, 山东 烟台 264025; 2. 中国科学院 海洋研究所, 山东 青岛 266071)

**摘要:** 根据已知的  $\beta$ -胡萝卜素酮化酶( $\beta$ -carotene ketolase, BKT) 的 cDNA 序列设计引物, 利用长距离 PCR 扩增获得了 *bkt* 基因的 cDNA 完整编码片段及基因组 DNA 片段, 并对这些片段进行了序列测定。cDNA 和基因组 DNA 比对结果表明该基因至少包括 5 个内含子, 它们的剪切位点皆符合 GU-AG 规律。在比对结果中同时还发现一个 19bp 的短序列重复, 该重复序列在 *bkt* 的 cDNA 和基因组 DNA 上都发生了多次重复, 推测在 BKT 的表达调控中可能具有一定功能。另外, 在序列比对过程中还发现本实验所获得的 cDNA 和基因组 DNA 序列与前人报道的 cDNA 序列之间都存在多处单碱基差异和 19bp 的短序列差异。

**关键词:** 雨生红球藻(*Haematococcus pluvialis*);  $\beta$ -胡萝卜素酮化酶; cDNA; 基因组 DNA; 克隆; 序列分析

中图分类号: Q31

文献标识码: A

文章编号: 1000-3096(2006)08-0020-08

雨生红球藻(*Haematococcus pluvialis*)是一种单细胞淡水绿藻, 在环境胁迫条件下可以迅速诱导合成虾青素, 其积累量可高达细胞干重的 4%<sup>[1]</sup>, 是其他虾青素合成生物体积累量的一到几个数量级。雨生红球藻是目前首选的天然虾青素合成藻。虾青素是具有最强抗氧化活性的类胡萝卜素, 其抗氧化活性比其它类胡萝卜素高约 10 倍, 比  $\alpha$ -生育酚高 500 倍, 已经被认为是“超级维生素 E”<sup>[2]</sup>。这种超强的抗氧化能力赋予虾青素广泛的医用价值, 如提高人体的免疫能力、抗肿瘤活性等<sup>[3-5]</sup>。除此之外虾青素也是水产养殖中不可或缺的饲料添加剂。红球藻中虾青素是在一系列酶的作用下合成的, 其中  $\beta$ -胡萝卜素酮化酶( $\beta$ -carotene ketolase, BKT) 是虾青素合成的关键酶<sup>[6]</sup>。Steinbrenner 等<sup>[7]</sup> 和 Grunewald 等<sup>[8]</sup> 的实验已表明在诱导虾青素大量合成的环境胁迫下, 虾青素合成酶——八氢番茄红素合成酶、 $\beta$ -胡萝卜素酮化酶和羟化酶的 mRNA 水平在胁迫后数小时内迅速提高, 在虾青素积累达高峰前到达最高水平, 这说明虾青素的合成调控部分发生在合成酶基因的转录水平上, 意味着从分子水平上研究虾青素合成酶的转录调控将有可能解决虾青素的生产问题。国外已经通过 cDNA 文库筛选获得了  $\beta$ -胡萝卜素酮化酶完整的 cDNA 编码序列<sup>[9-10]</sup>, 但它们的基因组 DNA 序列目前国内外都没有报道。为了找出与该基因表达

相关的调控序列, 作者根据已知的 cDNA 序列设计引物, 利用长距离 PCR 扩增 *bkt* 基因 cDNA 的完整编码片段及基因组 DNA 片段, 进行序列测定, 并对其核苷酸一级结构进行了分析。

## 1 材料与 方法

### 1.1 雨生红球藻的来源与培养

实验所用材料由中国科学院海洋研究所刘建国研究员提供, 采用 MCM 培养基<sup>[11]</sup>, 1000lx 的光照强度, 于 20℃ 以 12 h: 12 h 的光/暗周期进行静止培养。1000lx 下培养的细胞经 5000lx 的高光照强度处理 3 d 后用于 RNA 提取。

### 1.2 菌株、克隆载体及工具酶

大肠杆菌菌株 DH5 $\alpha$  为本实验室保存; RNA 提取试剂盒 RNeasy Plant Mini Kit 为 QIAGEN 公司产品; 反转录酶 MMLV RNase H<sup>-</sup> 为美国 Promega 公司产品; T-A 克隆载体 pMD 18-T Vector Kit 和长距离 PCR DNA 聚合酶 LA Taq<sup>TM</sup> 均购自大连宝生物工程

收稿日期: 2004-04-29; 修回日期: 2004-08-20

基金项目: 中国科学院重点创新项目(KZCX3-SW-215)

作者简介: 滕长英(1975), 女, 博士, 讲师, 研究方向: 藻类分子生物学及基因工程, E-mail: lzhangt@sina.com; 秦松, 通讯联系人, E-mail: sqin@ms.qdio.ac.cn

公司; Taq DNA 聚合酶和胶回收试剂盒购自上海生物工程公司; T<sub>4</sub> DNA 连接酶为 TaKaRa 公司产品。

### 1.3 红球藻 RNA 提取

取约 10<sup>6</sup> 个经高光强处理的新鲜细胞, 液氮冷冻研磨后按 RNeasy Plant Mini Kit 的操作说明提取红球藻总 RNA。

### 1.4 引物设计及合成

利用引物设计软件 OMIGA2, 根据 Kajiwara<sup>[9]</sup> 报道的 BKT 的 cDNA 序列在起始密码子处和终止密码子外端分别设计上下游引物:

1P1: 5'-CACTATCACATGCCATCG 3',

3P2: 5'-TG CATCAA ACTCCATCCTA-3'。

根据 BKT 的 cDNA 5' 序列设计另一对引物:

4P1: 5'-TCCGTCCTCTGCCAAATG 3',

4P2: 5'-AGCTGCC TGTGCCTCAA A 3' 扩增起始密码子上游的序列(图 1)。引物由上海生物工程公司合成。



图 1 *bkt* 引物设计示意图

Fig.1 Chart of *bkt* primers design

### 1.5 PCR 扩增获得 *bkt* 的基因组 DNA 片段

每 25 μL 反应体系取 10~50 ng 红球藻 DNA 为模板, 分别以 1P1/3P2 和 4P1/4P2 为引物, 在 DNA 聚合酶 LA Taq<sup>TM</sup> 的作用下按下列 PCR 循环参数进行 PCR 扩增: 94℃ 4 min, 94℃ 40 s, 60℃ 40 s, 72℃ 2 min, 32 个循环, 72℃ 再延伸 10 min。

### 1.6 RT-PCR 扩增获得 *bkt* 的 cDNA 片段

按下列组成调制反转录反应体系: 2 μL 10 × PCR 缓冲液、2 μL 2.5 mmol/L dNTP、1 μL 4 × 10<sup>7</sup> U/L RNase 抑制剂、0.2 μL 200 U/μL MMLV RNase H<sup>-</sup>、1 μL 10 μmol/L 引物 3p2、1 μL 总 RNA (约 0.5 ng)、1 μL 20 mmol/L DTT、12 μL 无 RNase 的 H<sub>2</sub>O, 42℃ 保温 2 h 合成 *bkt* 的 cDNA 第一链。取第一链反应产物 2 μL, 以 1P1 和 3P2 为引物按下列 PCR 循环参数进行 PCR 扩增: 94℃ 4 min, 94℃ 40 s, 60℃ 40 s, 72℃ 1 min, 32 个循环, 72℃ 再延伸 10 min。

### 1.7 T-A 克隆及测序

各 PCR 产物用琼脂糖凝胶电泳回收纯化后, 连

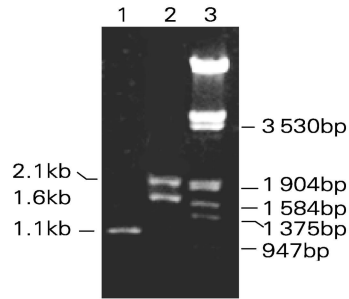
接到 pMD 18-T Vector 载体上, 转化 DH5α 感受态细胞, 用 PCR 法筛选阳性单克隆, 测序由上海康生生物公司完成。

### 1.8 序列分析

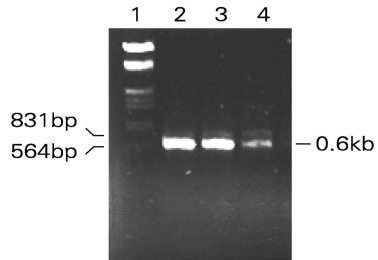
所获得的 *bkt* cDNA 序列、基因组 DNA 序列与 Kajiwara<sup>[9]</sup> 公布的 *bkt* cDNA 序列, 用软件 CLUSTAL 1.8 进行序列比对。

## 2 结果与分析

### 2.1 *bkt* cDNA 及基因组 DNA 片段的扩增



A



B

图 2 *bkt* cDNA 及基因组 DNA 的扩增

Fig. 2 PCR amplifications of *bkt* cDNA and genomic DNA

A. *bkt* D13/C13

1: *bkt* C13; 2: *bkt* D13; 3: λ DNA/EcoRI+ *Hind* III

B. *bkt* D4

1: λ DNA/EcoRI+ *Hind* III; 2: 58℃; 3: 60℃; 4: 62℃

以雨生红球藻的总 RNA 为模板, 1P1/3P2 为特异引物, 经过反转录和 PCR 扩增两个过程后, 产生一条约 1.1 kb 的产物, 命名为 *bkt* C13, 见图 2A。用聚合酶 LA Taq<sup>TM</sup> 进行长距离 PCR 扩增 *bkt* 的基因组 DNA 片段。由图 2A 的电泳结果可见 1P1/3P2 引导合成了两个大小不同的产物, 分别命名为 *bkt* D13 1,

*bkt* D13-2, 它们的分子量分别约为 2.1kb, 1.6kb。引物对 4P1/4P2 分别在 58, 60 和 62℃ 的退火温度下都能引导合成一约 0.6kb 的产物, 命名为 *bkt* D4, 见图 2B。

### 2.2 PCR 产物的克隆与测序

分别切取 cDNA 扩增产物 (*bkt* C13) 和 1P1/3P2 引导合成的基因组 DNA 扩增产物中较大的一条带 (*bkt* D13-1) 以及 4P1/4P2 引导合成的 *bkt* D4, 凝胶电泳回收后按适当比例分别连接到 T-克隆载体 pMD 18-T Vector 中。用 PCR 法筛选阳性克隆, 进

一步通过质粒大小验证后提交试剂公司测序。以上三个片断皆从正反两个方向分别进行测序, 最后进行拼接和检验, 测得的序列拼接结果分别见图 3。所获得的基因组 DNA 序列已提交 GenBank, 登陆号为 AY534876。

### 2.3 序列分析

*bkt* D13 和 *bkt* D4 的序列拼接后, 与本文所获得的 cDNA 序列和 Kajiwara<sup>[9]</sup> 报道的 cDNA 序列用序列比对软件 CLUSTAL 1.8 进行序列比较, 结果见图 3。

2	1	-----							
3	1	TCCGTCCTCT	GCCAAATCTC	GCGTCGGGAC	CTGCCTAAAT	CGAAGAATGC	ACGTCGCATC		
1	1	TCCGTCCTCT	GCCAAATCTC	GCGTCGGGAC	CTGCCTAAAT	CGAAGAATGC	ACGTCGCATC		
Con									
2	61	-----							
3	61	GGCACTAATG	GTCGAGCAGA	AAGGCAGTGA	GGCAGCTGCT	TGCAGCCAG	ACGTCTTGAG		
1	61	GGCACTAATG	GTCGAGCAGA	AAGGCAGTGA	GGCAGCTGCT	TGCAGCCAG	ACGTCTTGAG		
Con									
2	121	-----	--ACAGTATC	ACATGCCATC	CGAGTCGTCA	GACGCAGCTC	GTCCTGCGTT		
3	121	GGCGTGGGCG	ACACAGTATC	ACATGCCATC	CGAGTCGTCA	GACGCAGCTC	GTCCTGCGTT		
1	121	GGCGTGGGCG	ACACAGTATC	ACATGCCATC	CGAGTCGTCA	GACGCAGCTC	GTCCTGCGTT		
Con			*****	*****	*****	*****	*****	*****	*
2	181	GAAGCAGGCC	TATAAACCTC	CAGCATCTGA	CGCCAAGGGC	ATCACTATGG	CGCTGACCAT		
3	181	GAAGCAGGCC	TATAAACCTC	CAGCATCTGA	CGCCAAGGGC	ATCACTATGG	CGCTGACCAT		
1	181	GAAGCAGGCC	TATAAACCTC	CAGCATCTGA	CGCCAAGGGC	ATCACTATGG	CGCTGACCAT		
Con			*****	*****	*****	*****	*****	*****	*****
2	241	CATTGGCACC	TGGACCGCAG	TGTTTTTACA	CGCAATATTG	CAAATCAGGC	TACCGACATC		
3	241	CATTGGCACC	TGGACCGCAG	TGTTTTTACA	CGCAATATTG	CAAATCAGGC	TACCGACATC		
1	241	CATTGGCACC	TGGACCGCAG	TGTTTTTACA	CGCAATATTG	CAAATCAGGC	TACCGACATC		
Con			*****	*****	*****	*****	*****	*****	*****
2	301	CATGGACCAG	CTTCACTGGT	TGCCTGTGTC	CGAAGCCACA	GCCCAGCTGT	TGGGCGGAAG		
3	301	CATGGACCAG	CTTCACTGGT	TGCCTGTGTC	CGAAGCCACA	GCCCAGCTGT	TGGGCGGAAG		
1	301	CATGGACCAG	CTTCACTGGT	TGCCTGTGTC	CGAAGCCACA	GCCCAGCTGT	TGGGCGGAAG		
Con			*****	*****	*****	*****	*****	*****	*****
2	361	CAGCAGCCTA	TGTCACATCG	CGCAGTCTT	CATTGTACTT	GAGTTCTGT	ACACTG----		
3	361	CAGCAGCCTA	TGTCACATCG	CGCAGTCTT	CATTGTACTT	GAGTTCTGT	ACACTG[GT]GG		
1	361	CAGCAGCCTA	TGTCACATCG	CGCAGTCTT	CATTGTACTT	GAGTTCTGT	ACACTG----		
Con			*****	*****	*****	*****	*****	*****	*****
2	421	-----							
3	421	GGCGATCAGA	GTATACTGTC	TACCAGCGTA	TGCATCAGCT	TTGCTGTTAT	GCTCCCATAG		
1	421	-----							
Con									
2	481	-----							
3	481	CTTACATCGT	AAGCGCTTGT	GTCCTTTGA	CATGCGTAAT	CAACTCTAAC	TTTTGTCATG		
1	481	-----							
Con									

2	541	---GTCTATT	CATCACCACG	CATGATGCAA	TGCATGGCAC	CATAGCTTTG	AGGAACAGGC	
3	541	<u>CAG</u> GTCTATT	CATCACCACG	CATGATGCAA	TGCATGGCAC	CATAGCTTTG	AGGAACAGGC	
1	541	---GTCTATT	CATCACCACA	CATGACGCAA	TGCATGGCAC	CATAGCTTTG	AGGCACAGGC	
Con		*****	*****	*****	*****	*****	*****	
2	601	AGCTCAATGA	TCTCCTTGGC	AACATCTGCA	TATCACTGTA	CGCCTGGTTT	GACTACAGCA	
3	601	AGCTCAATGA	TCTCCTTGGC	AACATCTGCA	TATCACTGTA	CGCCTGGTTT	GACTACAGCA	
1	601	AGCTCAATGA	TCTCCTTGGC	AACATCTGCA	TATCACTGTA	CGCCTGGTTT	GACTACAGCA	
Con		*****	*****	*****	*****	*****	*****	
2	661	TGCCGCACCG	CAAG-----	-----	-----	-----	-----	
3	661	TGCCGCACCG	CAAG <u>GT</u> GAGG	AAGCGTCTGA	CCTCCTTCTC	TGCCAGTTCA	ACACATTACA	
1	661	TGCTGCATCG	CAAG-----	-----	-----	-----	-----	
Con		*** ** *	****					
2	721	-----	-----	-----	-----	CACTGGGAGC	ACCACAACCA	
3	721	GTCGCTGAGA	CTTGTGCGCT	GCGCTGCTTG	TCCTTC <u>G</u> CAG	CACTGGGAGC	ACCACAACCA	
1	721	-----	-----	-----	-----	CACTGGGAGC	ACCACAACCA	
Con						*****	*****	
2	781	TACTGGCGAA	GTGGGGAAAG	ACCCTGACTT	CCACAA <u>A</u> GGA	AATCC <u>T</u> GGCC	TTGTCCCCTG	
3	781	TACTGGCGAA	GTGGGGAAAG	ACCCTGACTT	CCACAA <u>A</u> GGA	AATCC <u>T</u> GGCC	TTGTCCCCTG	
1	781	TACTGGCGAA	GTGGGGAAAG	ACCCTGACTT	CCACAA <u>A</u> GGA	AATCC <u>C</u> GGCC	TTGTCCCCTG	
Con		*****	*****	*****	*****	*****	*****	
2	841	GTTGCCAG-	-----	-----	-----	-----	-----	
3	841	GTTGCCAG <u>G</u> -	<u>T</u> AAGGCTCAT	CAGTGGCTCT	GTTGCAGTGC	TGGTGGGCTA	TGCAAGGAAT	
1	841	GTTGCCAG-	-----	-----	-----	-----	-----	
Con		*****						
2	901	-----	-----	-----	-----	-----	-----	
3	901	ACATCTTGCT	GCACACAATC	CCTCCAGAGG	TGTATGGCAT	ATGTGGACAA	AACAAGTGTG	
1	901	-----	-----	-----	-----	-----	-----	
Con								
2	961	-----	-----	-----	CT	TCATGTCCAG	CTACATGTCC	CCATGGCAGT
3	961	GCAAGGTGTG	GTGCTGCCTG	CCGCGC <u>A</u> GCT	TCATGTCCAG	CTACATGTCC	C <u>T</u> GTTGGCAGT	
1	961	-----	-----	-----	CT	TCATGTCCAG	CTACATGTCC	C <u>T</u> GTTGGCAGT
Con					**	*****	*****	
2	1021	TTGCCCGGCT	GGCATGGTGG	GCAGTGGTGA	TGCAAA <u>C</u> GTT	GGGGGCCCC	ATGGC <u>G</u> AATC	
3	1021	TTGCCCGGCT	GGCATGGTGG	GCAGTGGTGA	TGCAAA <u>C</u> GTT	GGGGGCCCC	ATGGC <u>G</u> AATC	
1	1021	TTGCCCGGCT	GGCATGGTGG	GCAGTGGTGA	TGCAAA <u>C</u> GCT	GGGGGC <u>C</u> CC	ATGGC <u>A</u> AATC	
Con		*****	*****	*****	*****	*****	*****	
2	1081	TCCTAGTCTT	CATGGCTGCA	GCCCCAATCT	TGTCAGCATT	CCACCTCTTC	TACTTCG---	
3	1081	TCCTAGTCTT	CATGGCTGCA	GCCCCAATCT	TGTCAGCATT	CCGCCTCTTC	TACTTCG <u>G</u> TG	
1	1081	TCCTAGTCTT	CATGGCTGCA	GCCCCAATCT	TGTCAGCATT	CCGCCTCTTC	TACTTCG---	
Con		*****	*****	*****	*****	**	*****	
2	1141	-----	-----	-----	-----	-----	-----	
3	1141	TGTGCCTCTT	GACCCCAAG	CGTAACCCTG	GACATGCACC	AGTCAGCTTG	CATGGTTTCC	
1	1141	-----	-----	-----	-----	-----	-----	
Con								
2	1201	-----	-----	-----	-----	-----	-----	
3	1201	TAGCTTTGTG	CATGCTGAGC	TCGTGGTGAT	GACTGCGGTG	CTCCGGCCAG	CTTTCATATA	
1	1201	-----	-----	-----	-----	-----	-----	
Con								

2	1261	-----	-----	-----	-----	-----	-----
3	1261	AACTACCGGT	GACATCTGCT	AGTGCCAGGG	CTGTGCACAC	CATGGTGATG	CAGATCATGC
1	1261	-----	-----	-----	-----	-----	-----
Con							
2	1321	-----	-----	-----	-----	-----	-----
3	1321	ACATCATCAG	CACCACAGCG	TCATAGCCTT	GTGTGTACAG	TGTTTCTGCC	CTGCTTGCTG
1	1321	-----	-----	-----	-----	-----	-----
Con							
2	1381	-----	-----	-----	-----	-----	-----
3	1381	ATGCTTGTGA	AGTCCTGACG	CCCCTGACAC	CCCAGTAGCC	CAAATGGAGG	TGGCAACTGG
1	1381	-----	-----	-----	-----	-----	-----
Con							
2	1441	-----	-----	-----	-----	----- GCAC	TTACCTGCCA
3	1441	CCAATAGTGT	GGTGTCTGCG	TCTATGCTCG	TGGTGCTGGT	TTGCAGGCAC	TTACCTGCCA
1	1441	-----	-----	-----	-----	----- GCAC	TTACCTGCCA
Con						****	*****
2	1501	CACAAGCCTG	AGCCAGGCC	TGCAGCAGGC	TCTCAGGTCA	TGTCCTGGTT	CAGGGCCAAG
3	1501	CACAAGCCTG	AGCCAGGCC	TGCAGCAGGC	TCTCAGGTCA	TGTCCTGGTT	CAGGGCCAAG
1	1501	CACAAGCCTG	AGCCAGGCC	TGCAGCAGGC	TCTCAGGTGA	TGCCCTGGTT	CAGGGCCAAG
Con		*****	*****	*****	*****	** * *****	*****
2	1561	ACAAGTGAGG	CATCTGATGT	GATGAGCTTC	CTGACATGCT	ACCACCTTGA	CCTG-----
3	1561	ACAAGTGAGG	CATCTGATGT	GATGAGCTTC	CTGACATGCT	ACCACCTTGA	CCTGCTGGGT
1	1561	ACAAGTGAGG	CATCTGATGT	GATGAGTTC	CTGACATGCT	ACCACCTTGA	CCTG-----
Con		*****	*****	*****	*****	*****	****
2	1621	-----	-----	-----	-----	-----	-----
3	1621	GCTAACCCAC	ACTGACAAAG	CTGCAAGCCA	CCTCTGCTGG	CTGCGTGTAG	TGCATCAGCA
1	1621	-----	-----	-----	-----	-----	-----
Con							
2	1681	-----	-----	-----	-----	-----	-----
3	1681	AGCATGGTGC	ATAGACCCAA	CAAGTGCCT	GGTAAGTGTG	TCCCAGTCCC	AACCTGGCCT
1	1681	-----	-----	-----	-----	-----	-----
Con							
2	1741	-----	-----	-----	-----	-----	-----
3	1741	GGGATGGCAT	GCTGCATGTC	ACTTGACCAG	CTTTTGCTG	TGCCCTGGCT	GTCCTGTCGC
1	1741	-----	-----	-----	-----	-----	-----
Con							
2	1801	-----	CACTGGGAGC	ACCACAGGTG	GCCCTTTGCC	CCCTGGTGGC	AGCTGCCCA
3	1801	TGCCTTGCAG	CACTGGGAGC	ACCACAGGTG	GCCTTTGCC	CCCTGGTGGC	AGCTGCCCA
1	1801	-----	CACTGGGAGC	ACCACAGGTG	GCCCTTTGCC	CCCTGGTGGC	AGCTGCCCA
Con			*****	*****	* *****	*****	*****
2	1861	CTGCCGCCG	CTGCTGGGC	GTGGCCTGGT	GCCTGCCTTG	GCATGACCTG	GTCCTCCGC
3	1861	CTGCCGCCG	CTGCTGGGC	GTGGCCTGGT	GCCTGCCTTG	GCATGACCTG	GTCCTCCGC
1	1861	CTGCCGCCG	CTGTCCGGGC	GTGGCCTGGT	GCCTGCCTTG	GCATGACCTG	GTCCTCCGC
Con		*****	*****	*****	*****	*****	*****
2	1921	TGGTGACCCG	GCGTCTGCAC	AAGAGTGTCA	TGCTACAGGG	TGCTGGGGCC	AGTGGCAGCG
3	1921	TGGTGACCCG	GCGTCTGCAC	AAGAGTGTCA	TGCTACAGGG	TGCTGGGGCC	AGTGGCAGCG
1	1921	TGGTGACCCA	GCGTCTGCAC	AAGAGTGTCA	TGCTACAGGG	TGCTCGGGCC	AGTGGCAGCG
Con		*****	*****	*****	*****	*****	*****
2	1981	CAGTGCATCT	T----- GTA	TGGGGCTACC	GCTGTGCCAC	TGAGCACTGG	GCGTGCCACT

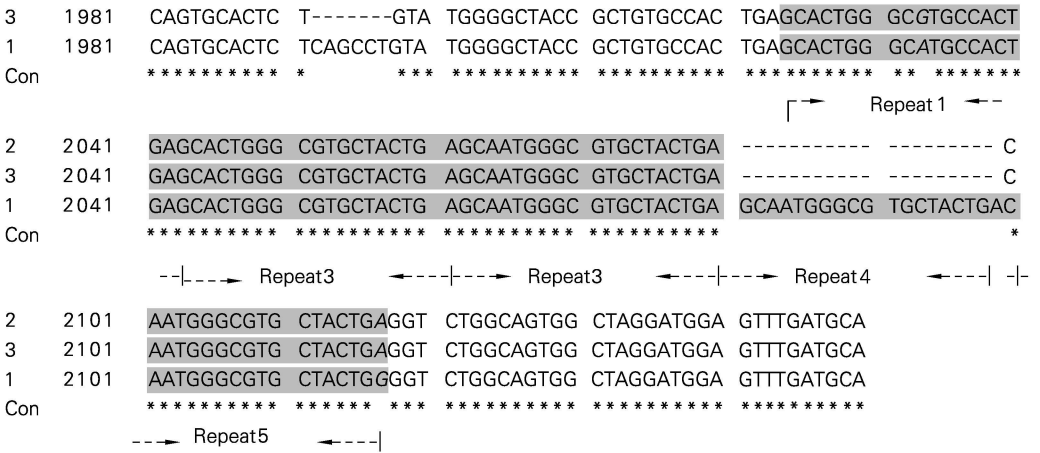


图3 *bkt* 的 DNA 序列与 cDNA 序列的比较

Fig. 3 The alignment of *bkt* DNA sequence with cDNA sequence

2 为 C13 的序列; 3 为 D13 和 D4 的拼接序列; 1 为 Kajiwara 报道的 cDNA 序列; Con 表示共有序列; 阴影和斜体表示差异序列; 粗体 ATG 和 TGA 分别为起始和终止密码子

2 indicates the sequence of C13; 3 indicates the joint sequence of D13 and D4; 1 indicates the sequence of cDNA reported by Kajiwara; Con indicates the consensus sequence; The shady and italic bases indicate the different nucleotides in between; The bold ATG and TGA indicate initial and stop codon

序列比较结果表明 *bkt* 基因为断裂基因, 图 3 显示了该基因的内含子和外显子序列, 在该基因序列中共有 5 个内含子和 6 个外显子。表 1 列出了 *bkt* 的所有内含子和外显子的大小、剪切位点序列以及外显子在 β 胡萝卜素酮化酶多肽编码链上的位置。与动物相比, 该基因的内含子都比较小, 其中最大的一个为 349bp, 最小的一个只有 85bp。由表 1 的剪切位点序列可以看出 *bkt* 基因的所有内含子-外显子交界处都存在 GT-AG 序列, 这证明雨生红球藻断裂基因的内含子与其它物种的内含子剪切规律相符。在 *bkt* 中, 不仅 GT-AG 是内含子两端的共有序列, 而且在 5' 剪切位点的外显子 3' 末端还存在一富含 G 的延伸(见表 1)。

一般在位于 3' 剪切位点上游 20 和 50 个核苷酸之间的位置上, 内含子还具有一称作分支点的重要位点, 在某些种属如酵母、哺乳动物等存在一段共有或相似序列。但在 *bkt* 基因的内含子中没有发现具有共同或相似序列的分支点。

图 3 中同时用阴影和斜体标出的碱基为 cDNA 和 DNA 序列间存有差异的核苷酸, 在所测得的 *bkt* 基因组 DNA 序列(D4-D13)中共存在 33 个与 C13 序

列或 Kajiwara 报道的 cDNA 序列不同的碱基, 其中有 28 个碱基与 Kajiwara 报道的 cDNA 序列不同, 但却与 C13 的序列一致, 这说明该 28 个碱基不是作者在 PCR 过程或测序中造成的错误信息。这种差异有三种可能解释: (1) Kajiwara 等人所用的材料与作者的材料存在少量遗传差异, 尽管属于同一物种; (2) Kajiwara 等人与作者所用的材料相同, 但是在雨生红球藻基因组中该基因存在一个以上的拷贝, 这两个结果来自不同的拷贝; (3) 该基因存在等位基因的差异, 这两个不同结果来自存在差异的等位基因。

除了上述的单碱基差异外, 作者获得的 cDNA 序列和 DNA 序列与 Kajiwara 报道的 cDNA 序列还存在两处较大的差异, 即分别在图 3 的比对结果中 1992~1997 和 2081~2099 位置处缺少两段核苷酸序列, 或者说后者比作者的序列多出两段寡核苷酸。为了分析在 2081~2099 位置处的 19bp 序列 GCAATGGGCGTGCTACTGA 的来源, 作者在 *bkt* 的 cDNA 和 DNA 序列中查找这段序列。结果发现这段序列在 Kajiwara 报道<sup>[9]</sup> 的 cDNA 中紧接着该序列的上下游共有 5 个非常相似的拷贝(图中用阴影标示出), 这是一种短序列重复。这种由 5~50 个短的重复单位串联组成的序列, 在分子生物学上称为小卫星(minisatellite)或数目可变的串联重复(variable number tandem repeat, VNTR) 序列。这两个重复序

列的差异发生在终止密码子之后,也即 3' 非翻译区 (3' UTR),因而肯定不会影响 BKT 酶的活性。但是

不能排除这种序列重复对 BKT 酶的表达水平存在影响。

表 1 *bkt* 中内含子和外显子的组织

Tab. 1 The organizations of exons and introns in *bkt*

外显子			内含子		
位置	大小(bp)	5' 剪切位点	大小(bp)	分类	3' 剪切位点
1~ 274	274	CACTGgtgggcega	127	1	atgcagGTCTA
275~ 405	131	GCAAGgtgaggaa	85	0	cgcagCACTG
406~ 494	89	GCCAGgtaaggc	139	2	cgcagCTTCA
495~ 643	149	CTTCGgtgtgtgc	349	1	ttgcagGCACT
644~ 771	128	ACCTGgtgggtgc	270	0	ttgcagCACTG
772~ 860	89				

表中位置指外显子在编码序列中所处的位置,其中 1 指在翻译起始密码子 ATG 中 A 的位置,860 为终止密码子的位置。5' 剪切位点和 3' 剪切位点中大写字母为外显子部分,小写字母标示内含子的序列。

### 3 讨论

随着对更多内含子的鉴定和功能分析,人们逐渐认识到内含子在基因的表达中担负一定的调控功能,已经有多个具有调控功能的内含子被鉴定。如 *Serrate* 基因编码果蝇 (*Drosophila*) 的一个跨膜蛋白,控制果蝇的翅膀和 haltere 的发育,它的表达在果蝇的胚胎和幼虫发育阶段受到复杂的时空调控。功能鉴定实验已证明该基因的第三个启动子中含 *Serrate* 基因后期转录的增强子成分<sup>[12]</sup>。人类酸性麦芽糖酶基因的其中一个内含子是一个具有组织特异性的沉默子<sup>[13]</sup>。

内含子在基因表达中具有调控作用的另一个证据是 cDNA 在真核细胞中的表达效率常常比包括内含子的基因组基因低一些。例如,早期的实验证明重组的 SV-40 cDNAs 只能产生很少的 mRNA 甚至没有 mRNA 的转录产生,但是当在其中插入一个内含子后,这种重组 cDNA 就可以大量表达了<sup>[14]</sup>。至今已经证明在包括哺乳动物、昆虫、植物等多种生物体中,多个遗传转化的内源基因的表达都需要一个或多个内含子的存在<sup>[15-17]</sup>。内含子对基因表达的影响程度差别很大,从没有明显作用到提高表达量达 400 倍<sup>[15-17]</sup>。

各种环境胁迫对虾青素代谢的诱导作用,暗示虾青素合成酶的表达在红球藻中受到复杂的调控作用,很难排除 *bkt* 基因的内含子在其中行使一定的功能。Marquardt 等在红藻 (*Galdieria sulphuraria*) 中已经发现了编码捕光色素结合蛋白的 *lhc4* 基因能够产生两种成熟的 mRNA,有证据提示这种可变剪切与光捕获蛋白的表达调控有关。这说明藻类基因

的内含子同样具有表达调控功能的可能性。下一步作者将通过重组和转化实验验证雨生红球藻 *bkt* 基因的各个内含子对该基因的表达水平是否存在影响。

通过对  $\beta$ -胡萝卜素酮化酶基因的 cDNA 序列与基因组序列的比较,在该基因的 3' 非翻译区发现了多个长度为 19bp 的串联重复序列。Lers 等<sup>[18]</sup> 从盐藻 (*Dunaliella bardawil*) 中分离到一个  $\beta$ -胡萝卜素合成相关蛋白的基因 (*cbr*) 的起始密码子上游,也发现了三个 21bp 长的串联重复,它的部分序列与 LDL 受体蛋白的固醇调节因子的结合位点序列相似。通过在 GenBank 中查找,没有在其他核酸序列中发现与这两个序列相似的序列,是否这一重复序列在  $\beta$ -胡萝卜素酮化酶基因的表达调控中起到重要的作用,需要进一步的实验验证。

#### 参考文献:

- [1] Boussiba S. Carotenogenesis in the green alga *Haematococcus pluvialis*: cellular physiology and stress response[J]. *Physiol Plant*, 2000, 108: 111-117.
- [2] Miki W. Biological functions and activities of animal carotenoids[J]. *Pure Appl Chem*, 1991, 63: 141-146.
- [3] Kobayashi M, Sakamoto Y. Singlet oxygen quenching ability of astaxanthin esters from the green alga *Haematococcus pluvialis*[J]. *Biotechnol Lett*, 1999, 21: 265-269.
- [4] Jyonouchi H, Sun S, Cross M. Astaxanthin, a carotenoid without vitamin A activity, augments antibody response in cultures including T-helper cell clones and suboptimal doses of antigen[J]. *J Nutr*, 1995, 125 (10): 2483-2492.
- [5] Tanaka T, Morishita Y, Suzai M, et al. Chemopre-

- vention of mouse urinary bladder by the naturally occurring carotenoid astaxanthin [J]. **Carcinogenesis**, 1994, **15**(1): 15-19.
- [6] Grunewald K, Hirschberg J, Hagen C. Ketocarotenoid biosynthesis outside of plastids in the unicellular green alga *Haematococcus pluvialis*[J]. **J Biol Chem**, 2001, **276**(8): 6 023-6 029.
- [7] Steinbrenner J, Linden H. Regulation of two carotenoid biosynthesis genes coding for phytoene synthase and carotenoid hydroxylase during stress-induced astaxanthin formation in the alga *Haematococcus pluvialis* [J]. **Plant Physiol**, 2001, **125**: 810-817.
- [8] Grunewald K, Eckert M, Hirschberg J, *et al.* Phytoene desaturase is localized exclusively in the chloroplast and up-regulated at the mRNA level during accumulation of secondary carotenoids in *Haematococcus pluvialis* (Volvocales, Chlorophyceae) [J]. **Plant Physiol**, 2000, **122**: 1 261-1 268.
- [9] Kajiwara S, Kakizono T, Saito T, *et al.* Isolation and functional identification of a novel cDNA for astaxanthin biosynthesis from *Haematococcus pluvialis*, and astaxanthin synthesis in *Escherichia coli* [J]. **Plant Molecular Biology**, 1995, **29**: 343-352.
- [10] Lotan T, Hirschberg J. Cloning and expression in *Escherichia coli* of the ketocarotenoid canthaxanthin in *Haematococcus pluvialis* [J]. **FEBS Lett**, 1995, **364**: 125-128.
- [11] Borowitzka M A, Huisman J M, Osborn A. Culture of the astaxanthin producing green alga *Haematococcus pluvialis* [J]. **J Appl Phycol**, 1991, **3**: 295-304.
- [12] Bachmann A, Knust E. Dissection of cis regulatory elements of the *Drosophila* gene *Serrate* [J]. **Dev Genes Evol**, 1998, **208**: 346-351.
- [13] Yan B, Raben N. Identification and characterization of a tissue-specific silencer element in the first intron of the human acid maltase gene [J]. **Hum Genet**, 2001, **109**(2): 186-90.
- [14] Gruss P, Lai C J, Dhar R, *et al.* Splicing as a requirement for biogenesis of functional 16S mRNA of simian virus 40 [J]. **Proc Natl Acad Sci**, 1979, **76**: 4 317-4 321.
- [15] Buchman A R, Berg P. Comparison of intron-dependent and intron-independent gene expression [J]. **Mol Cell Biol**, 1988, **8**: 4 395-4 405.
- [16] Duncker B P, Davies P L, Walker V K. Intron boost transgene expression in *Drosophila melanogaster* [J]. **Mol Gen Genet**, 1997, **254**: 291-296.
- [17] Bourdon V, Harvey A, Lonsdale D M. Introns and their positions affect the translational activity of mRNA in plant cells [J]. **EMBO Rep**, 2001, **2**: 394-398.
- [18] Lers A, Levy H, Zamir A.  $C\sigma$  regulation of a gene homologous to early light-induced genes in higher plants and  $\beta$ -carotene biosynthesis in the alga *Dunaliella bardawil* [J]. **J Biol Chem**, 1991, **266**(21): 13 698-13 705.

## Cloning and sequence analysis of $\beta$ carotenoid ketolase cDNA and genomic DNA in *Haematococcus pluvialis*

TENG Chang-ying<sup>1</sup>, ZHANG Li<sup>1</sup>, QIN Song<sup>2</sup>, Chengkui Tseng<sup>2</sup>

(1. Yantai Normal College, Yantai 264025, China; 2. Institute of Oceanology, the Chinese Academy of Sciences, Qingdao 266071, China)

Received: Apr., 29, 2004

**Key words:** *Haematococcus pluvialis*;  $\beta$ -carotenoid ketolase; cDNA; genomic DNA; clone; sequence analysis

**Abstract:** This paper had designed primers according to the cDNA sequence reported by Kajiwara, cloned the cDNA and genomic DNA fragments by LD-PCR, and obtained their entire sequences by sequencing. The alignment in between cDNA, genomic DNA and cDNA reported by Kajiwara shows that this gene includes five introns at least and all the exon-intron junctions conform to GU-AG splicing rule. Through alignment a 19bp long short repeat unit was found, which exists in cDNA and DNA of *bkt* for several times. They may have some functions in BKT expression regulation. The diversities from single nucleotides and 19bp short sequence between cDNA sequence reported by Kajiwara and DNA sequences obtained in this paper suggested they have different origins.

(本文编辑:张培新)