

# 竹子捕光叶绿素 a/ b 结合蛋白基因全长的 克隆和序列分析\*

高志民<sup>1</sup> 李雪平<sup>1</sup> 彭镇华<sup>2</sup> 岳永德<sup>1</sup>

(1. 国际竹藤网络中心 国家林业局竹藤科学与技术重点开放实验室 北京 100102; 2. 中国林业科学研究院林业研究所 北京 100091)

**摘要:** 采用 RT-PCR 技术与 RACE 技术,从绿竹中克隆到捕光叶绿素 a/ b 结合蛋白基因全长 cDNA,命名为 *cab-BO1* (GenBank 登记号:EF061137)。该基因长 1 102 bp,从第 64 ~ 861 bp 含有 1 个开放阅读框,编码 265 个氨基酸。*cab-BO1* 编码的氨基酸中第 64 ~ 232 位包括典型的捕光叶绿素 a/ b 结合蛋白功能域,第 9 ~ 11 位为蛋白激酶 C 的磷酸化位点,第 27 ~ 52 位为泛素结构功能域信号,第 55 ~ 58 位为酪氨酸激酶 (CK2) 磷酸化位点。序列相似性分析结果表明:*cab-BO1* DNA 序列和编码的氨基酸序列与玉米、小果野蕉、小麦、水稻等的 *cab* 基因及其氨基酸序列的相似性分别在 85 % 和 89 % 以上,显示了 *cab* 家族基因在进化过程中的保守性。

**关键词:** 绿竹; 叶绿素 a/ b 结合蛋白基因 (*cab*); 基因克隆; 序列分析

中图分类号: S718.46; Q943.2 文献标识码: A 文章编号: 1001-7488(2007)03-0034-05

## Cloning and Characterization of a Full-Length Gene Encoding the Light Harvesting Chlorophyll a/ b Binding Protein in Bamboo

Gao Zhimin<sup>1</sup> Li Xueping<sup>1</sup> Peng Zhenhua<sup>2</sup> Yue Yongde<sup>1</sup>

(1. Key Laboratory of Bamboo and Rattan Science and Technology of State Forestry Administration International Center for Bamboo and Rattan Beijing 100102; 2. Research Institute of Forestry, CAF Beijing 100091)

**Abstract:** A full-length cDNA of *cab* gene was cloned from the first strand of bamboo (*Bambusa oldhamii*) cDNA through RT-PCR and RACE methods, named as *cab-BO1* (*cab* gene from *B. oldhamii*). The length of *cab-BO1* (GenBank accession number: EF061137) is 1 102 bp, which contains an open reading frame encoding 265 amino acids from 64<sup>th</sup> to 861<sup>th</sup> position. The bioinformatics analysis indicated that the protein encoded by *cab-BO1* has a chlorophyll a/ b binding domain (64<sup>th</sup> ~ 232<sup>th</sup> position), a protein kinase C phosphorylation site (9<sup>th</sup> ~ 11<sup>th</sup> position), a ubiquitin domain signature (27<sup>th</sup> ~ 52<sup>th</sup> position), and a casein kinase II phosphorylation site (55<sup>th</sup> ~ 58<sup>th</sup> position). The DNA sequence and encoding amino acid sequence of *cab-BO1* showed high similarity with the *cab* genes of *Triticum aestivum*, *Musa acuminata*, *Zea mays* and *Oryza sativa*, more than 85 % and 89 % respectively. The result indicates a clue that cDNA of *cab* family genes are conserved.

**Key words:** bamboo (*Bambusa oldhamii*); light harvesting chlorophyll a/ b binding protein (*cab*); gene cloning; characterization

由于光合作用对人类生存和发展的重要性,关于植物光合作用的研究一直是生物学的热点。光合特性反映植物对外界物质的同化能力,竹子具有生长快、产量高等特点,这意味着竹子具有较强的同化能力。通过对毛竹 (*Phyllostachys edulis*) 叶片净光合速率与光照强度、气温多元相关分析表明,日光合量与平均气温呈极显著线性正相关,呼吸速率与气温呈极显著相关,根据回归方程模型估算毛竹的年净光合量、年呼吸量分别为  $1.81 \times 10^4$ 、 $2.00 \times 10^4$  mg CO<sub>2</sub> · dm<sup>-2</sup> · a<sup>-1</sup> (黄承才等,2000)。对 9 个毛竹种源新竹光合性状的研究结果表明,所有种源净光合速率在 1 年中有 2 个峰值,第 1 次都在 5 月,但第 2 次峰值随南北不同种源带有所差异,各种源带净光合速率随纬度的升高而呈现增大的趋势,各种源光补偿点的变幅在 570 ~ 913 lx,其中中带种源最低,能更有效地利用光能 (陈存及等,2001)。对雷竹 (*Phyllostachys praecox* f. *prevernalis*)、绿竹 (*Bambusa oldhamii*) 的研究也得到了类似的结果,不同种源的净光合速率日变化呈双峰曲线,有明显的“午休”现象 (郑炳松等,2001; 黄勇,2003)。竹子作为高效生物质能源植物之一,其光合作用有着独特的生理规

收稿日期:2006-11-10。

基金项目:国家林业局 948 项目 (2005-4-38; 2004-4-58; 2004-4-60)。

\* 彭镇华为通讯作者。

律,真正搞清竹子的高效光合机制仅从形态、生理角度进行研究有一定的局限性,如何更好地利用和调控竹子的光合效率,提高目标产量,需要在现有基础上开展竹子光合作用的分子机制研究。

捕光色素结合蛋白基因属于光合系统基因(Raghvendra, 1998),其编码的蛋白与色素所形成的色素蛋白复合体是一类捕获光能并把能量迅速传至反应中心引起光化学反应的色素蛋白复合体,在类囊体膜中除了进行光能的吸收和传递之外,在维持类囊体膜的结构,调节激发能在光系统 和光系统 之间的分配,光保护以及对各种环境的适应等过程中都起着重要的作用(孙钦秒等,2000)。绿竹是我国原产的优良速生笋材两用丛生竹种之一。本研究以绿竹为材料进行光合捕光蛋白基因的克隆,以期从分子水平来揭示竹子高效光合作用的机制奠定基础。

## 1 材料与方法

### 1.1 总 RNA 提取与 cDNA 合成

采用 Invitrogen 公司的 Trizol reagent 提取绿竹(源于浙江省温州)幼嫩新鲜叶片的 RNA(Gao *et al.*, 2006),用 Promega 公司的反转录试剂盒合成 cDNA。采用 Clontech 公司的 SMART<sup>TM</sup> RACE 试剂盒合成 3 cDNA。

### 1.2 基因的克隆与测序

根据禾本科植物捕光叶绿素 a/b 结合蛋白基因家族中的同源基因序列设计 PCR 引物,由北京三博远志科技有限公司、上海生工生物工程技术有限公司合成。

引物 1: 5'-AGCATCTCGGTAATGGACCTGAC-3'; 引物 2: 5'-CCTGGACGAA GAATCCGAACA-3';

引物 3: 5'-GCATCGCCCTCGGTAGAAC-3'; 引物 4: 5'-GTCGTCCGTGGCTACTACAAG-3';

引物 5: 5'-CGTGCTCTACCTCGGTCCTCTCC-3'; 引物 6: 5'-TGACCCCGAGACCTTCGCTAAGAACC-3'。

以绿竹 cDNA 为模板,进行梯度 PCR。引物 1 和引物 2 反应条件为:94 预变性 5 min; 94 1 min, 55 ~ 65 1 min, 72 1.5 min, 39 个循环; 72 延伸 10 min。引物 1 和引物 4 反应条件为:94 预变性 5 min; 94 1 min, 55 ~ 65 1 min, 72 2 min, 39 个循环; 72 延伸 10 min。引物 2 和引物 3 反应条件为:94 预变性 5 min; 94 1 min, 55 ~ 60 1 min, 72 2 min, 39 个循环; 72 延伸 10 min。引物 3 和引物 4 反应条件为:94 预变性 5 min; 94 1 min, 55 ~ 60 1 min, 72 1.5 min, 39 个循环; 72 延伸 10 min。引物 5 和引物 6 的 3-RACE 反应条件 1 为:94 30 s, 72 3 min, 5 个循环; 94 30 s, 70 30 s, 72 3 min, 5 个循环; 94 30 s, 68 30 s, 72 3 min, 25 个循环。引物 5 和引物 6 的 3-RACE 反应条件 2 为:94 30 s, 68 30 s, 72 3 min, 39 个循环。

PCR 产物电泳分析后用上海申能博彩生物科技有限公司试剂盒回收,按照 Promega 的 pGEM-T easy 载体快速连接试剂盒操作流程,将回收的 DNA 片段连接到载体上,转化大肠杆菌(*Esherichia coli*)DH5 菌株,经蓝白斑筛选,提取阳性克隆质粒并酶切图谱分析后,再将单克隆送北京三博远志科技有限公司测序。

### 1.3 序列的生物信息学分析

利用 DNASTAR、SMART 和 Clustl W 等生物软件分析测定 cDNA 序列及其编码蛋白质的结构特点,并与国际核酸和蛋白质数据库联网进行 blast 比较分析。

## 2 结果与分析

### 2.1 基因保守区片段的克隆与分析

用 Promega 公司的反转录试剂盒合成 cDNA 作模板,引物 1 和引物 2、引物 1 和引物 4、引物 3 和引物 2、引物 3 和引物 4 的 PCR 产物,在 1% 的琼脂糖凝胶中电泳,EB 染色结果如图 1 所示,引物 1 和引物 2 的产物在 500 bp 左右有一条亮带,引物 1 和引物 4 的产物、引物 3 和引物 2 的产物在 700 bp 左右都有一条亮带,与预测的基因片段基本相符,初步确定为目的基因片段。而引物 3 和引物 4 没有扩增产物(没有显示)。

用回收试剂盒回收 PCR 产物,并与 pGEM-T easy 载体连接,转化后得到的阳性克隆送公司测序。结果表明:引物 1 和引物 2 PCR 产物的插入片段为 553 bp,引物 1 和引物 4 PCR 产物的插入片段为 718 bp,引物 3 和引物 2 PCR 产物的插入片段为 754 bp,与推断相符。应用 DNASTAR 软件和 Blast 在线软件,将测序的序列与已经报道的基因进行同源性比较,发现引物 1 和引物 4 PCR 产物和引物 3 和引物 2 PCR 产物的插入片段与 *cab* 基因家族有着较高的同源性,而且引物 3 和引物 2 PCR 产物的插入片段包含了基因的 5 末端。

### 2.2 3-RACE 扩增与分析

为了获得基因的全长,根据测序克隆插入片段的序列分别设计了 3-RACE 引物 5 和引物 6,与 SMART<sup>TM</sup> RACE 试剂盒通用引物 (UPM) 配对进行扩增,结果如图 2 所示。引物 5 和 UPM 的 PCR 产物在 800 bp 左右有一条亮带,引物 6 和 UPM 的 PCR 产物在 500 bp、700 bp 左右各有一条亮带,将 PCR 产物回收并与 pGEM-T easy 载体连接后转化,在得到的大量克隆中,选取 2 个阳性克隆进行测序,插入片段大小分别为 809 bp、767 bp,都含有终止密码子和 Poly(A) 结构,相似性为 99%,证明得到了含有基因的 3 末端的片段。

### 2.3 序列分析与 cDNA 全长的克隆

通过对测序基因片段序列的拼接获得了一个 cDNA 全长为 1 102 bp 的基因 (GenBank 登记号: EF061137)。对 EF061137 序列进行分析,结果显示该基因序列从第 64 bp 开始至 861 bp 含有 1 个开放阅读框 (open reading frame, ORF) 和 1 个终止密码子,编码 265 个氨基酸 (图 3)。EF061137 序列第 64 bp 开始的第 1 个起始密码子 AUG 上游第 3 个核苷酸为鸟嘌呤 (G),紧跟其后的核苷酸也为鸟嘌呤 (G),由此可见 AUG 起始密码子前后符合 Kozak 规则 GNAUGG (Kozak, 1999)。此外,EF061137 在 5 端有 63 bp 的非编码区,在 3 端含有 213 bp 的非编码区 (untranslated region, UTR) 和 Poly(A) 28 bp。在转录终止位点附近还有 1 个典型的加尾信号 G AATAA (在图 3 中用下划线标出),上述特征表明,该序列是一个完整的捕光叶绿素 a/b 结合蛋白基因,将其命名为 *cab-BO1* (*cab* gene from *B. oldhamii*)。

经过 SMART 软件和 <http://www.expasy.org/prosite/> 里面 MOTIFSCAN 选相分析, *cab-BO1* 编码的氨基酸中第 64 ~ 232 位包括典型的捕光叶绿素 a/b 结合蛋白功能域 (chlorophyll a/b binding domain), 第 9 ~ 11 位为蛋白激酶 C 的磷酸化位点 (protein kinase C phosphorylation site), 第 27 ~ 52 位为泛素结构功能域信号 (ubiquitin

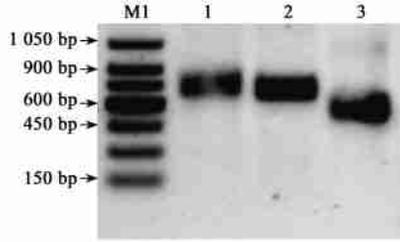


图 1 RT-PCR 产物电泳检测

Fig. 1 RT-PCR products checked by electrophoresis

- 1: 引物 1 和 4 产物 Primer 1 & 4 product;
- 2: 引物 2 和 3 产物 Primer 2 & 3 product;
- 3: 引物 3 和 4 产物 Primer 3 & 4 product;
- M1: 150 bp 分子量标记 150 bp ladder.

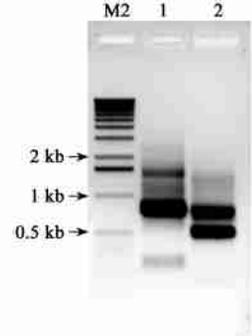


图 2 3-RACE PCR 产物电泳检测

Fig. 2 3-RACE products checked by electrophoresis

- 1: 引物 5 和 UPM 的产物 Product of primer 5 & UPM;
- 2: 引物 6 和 UPM 的产物 Product of primer 6 & UPM;
- M2: 1 kb 分子量标记 1 kb ladder.

```

1  TGCATCCGCTCGTTAGAACACACAGGAGTAGCTAGTACCTCCGGTCGACTTAGGAGC
61  GCAATGGCGACCACCACCATGGCCCTATCCTCCCGGCCCTCGCCGCAAGGCTGTCAAG
    M A T T T M A L S S P A F A G K A V K
121 AACGTGCCGTCGTCGGGTCTCTTCGGCGAGGCCCGCATGACCATGCCGAAGACCGGCGCC
    N V P S S G L F G E A R M T M R K T G A
181 AAGCCCAAGGCGGCCCTCAGGTAGCCCGTGGTACGGCCAGACCCGGTCTCTACCTCGGT
    K P K A A S G S P W Y G P D R V L Y L G
241 CCGCTCTCCGGCGAGCCCCGAGCTACCTGACCCGGGAGTTCGCCGGGACTACGGGTGG
    P L S G E P P S Y L T G E F A G D Y G W
301 GACACCCCGGGCTCTCGGCTGACCCGAGACCTTCGCTAAGAACCAGAGCTGGAGGTG
    D T A G L S A D P E T F A K N R E L E V
361 ATCCACTCCCGCTGGGCCATGCTCGGCGCGCTTGGTGCCTCTCCCGAGCTGCTCGCC
    I H S R W A M L G A L G C V F P E L L A
421 CGCAATGGCGTCAAGTTCGGCGAGCCGTGTGGTTCAAGCCCGCTCCAGATCTTCAGC
    R N G V K F G E A V W F K A G S Q I F S
481 GAGGGCGCCCTCGACTACCTCGCAACCCGAGCCTCATCCACGCGAGAGCATCTCGCC
    E G G L D Y L G N P S L I H A Q S I L A
541 ATCTGGGCTGCCAGGTCGTGCTCATGGGCGCGTCGAGGGGTACCGCATCCCGCGGGC
    I W A C Q V V L M G A V E G Y R I A G G
601 CCGCTCGGCGAGGTTGTCGACCCGCTTACCCGCGCGGACGTTTCGACCCGCTTGGCCTG
    P L G E V V D P L Y P G G S F D P L G L
661 GCCGACGACCCCGAGGCGTTCGCGGAGCTCAAGTGAAGGAATCAAGAACGGCCGCGCTC
    A D D P E A F A E L K V K E L K N G R L
721 GCCATGTTCTCCATGTTTCGGGTTTTTCGTGCAGGCCATCGTACCCGCAAGGGCCACTC
    A M F S M F G F F V Q A I V T G K G P L
781 GAGAACCTCGCCGACCACCTTGGCCGACCCGTCACAAACAACCGCTGGGCTAGTCCCC
    E N L A D H L A D P V N N N A W A Y V P
841 AACTTCGTGCCCGCAAGTGGAGGCGCCCGGAGACTGCGAGCTGTAGGTGTTGAGAGC
    N F V P G K
901 CATGTTAAAGATTAAGATGAATCTCTATAGTTTAAAGTCCAATTCGATTTACTT
961 GTAATCTCTAATCTTCAGAGTTTGTTCCTGTAATCCCTCCCGTACACGTACAGATAA
1021 GCTACAAGGAATGTATTGTCGTTGACTCTATTTCAGATCATCTTTTCTACCGAAAAAA
1081 AAAAAAAAAAAAAAAAAAAAAA
  
```

图 3 *cab-BO1* 基因 (EF061137) 的序列和开放阅读框

Fig. 3 Sequence and open reading frame of *cab-BO1* (EF061137)

domain signature),第 55~58 位为酪氨酸激酶(CK2)磷酸化位点(casein kinase II phosphorylation site)。此外,通过 DNASTAR 软件预测,*cab-B01* 编码的蛋白质等电点和分子量分别为 5.248 和 28 121.25 Da。

应用 Mapdraw 软件对获得基因序列分析发现,对引物 3 和引物 2 的 PCR 产物的测序片段以及引物 5 和 UPM 的 PCR 产物的测序片段都有 *Sac* 单酶切位点,恰好位于两序列的重叠部分。为了获得含有完整序列的基因克隆,首先将引物 3 和引物 2 的 PCR 产物的测序的克隆用 *EcoR* 和 *Sac* 双酶切后,将约 440 bp 的片段与 pBII01 载体连接,获得克隆 A;再将引物 5 和 UPM 的 PCR 产物的测序克隆用 *Pst* 和 *Sac* 双酶切后,将约 690 bp 的片段与 pUC19 载体连接,获得克隆 B;然后用 *Hind* 和 *Sac* 双酶切克隆 A 去除 GUS 片段后作为载体,将克隆 B 用 *Hind* 和 *Sac* 双酶切得到的约 690 bp 作为插入片段,连接后获得含有完整序列(1 102 bp)的基因克隆。



图 4 *cab-B01* 基因 (EF061137) 编码的氨基酸与不同的 *cab* 家族基因编码的氨基酸序列多序列比较

Fig. 4 Multiple sequence alignments of amino acids encoded by *cab-B01* (EF061137) blast with amino acids encoded by the different *cab* family genes

2.4 *cab-B01* 与 *cab* 家族基因的比较分析

通过 blast 软件分析,在 NCBI 核酸数据库中,与 *cab-B01* 有同源性的 *cab* 家族基因或光合系统中的类似基因共有 66 条,这些基因来自玉米 (*Zea mays*)、小果野蕉 (*Musa acuminata*)、小麦 (*Triticum aestivum*)、水稻 (*Oryza sativa*)、大麦 (*Hordeum vulgare*)、松属 (*Pinus*) 植物、美花烟草 (*Nicotiana glauca*)、拟南芥 (*Arabidopsis*)

thaliana)等十几个不同的物种。Blast 比较结果中的 Score 分值为 228~965,其中 *cab-BO1* 与 Y00379 (来源于玉米品种 Golden Cross Bantam) 序列相似性最高,Score 分值为 965,一致性(identity)为 90%(728/804)。在 NCBI 蛋白质数据库中,共搜索到与 *cab-BO1* 编码氨基酸有一定相似性的序列 500 条,Score 分值在 450 以上的就有 31 条,其中相似性最高是玉米(登记号为 P27497),为 92%(248/267)。图 4 是 *cab-BO1* 编码的蛋白质与同源性靠前的 5 个不同物种蛋白质序列的 Clustal W 比较结果,相同氨基酸序列用 \* 标出。

### 3 讨论

在核酸及其编码蛋白的 blast 分析中发现,*cab-BO1* 都与玉米的相似性最高,而来源于玉米的基因(X55892)编码的蛋白(P27497)为光系统的捕光叶绿素 a/b 结合蛋白(Viret *et al.*, 1990),因此推测在绿竹中 *cab-BO1* 编码的蛋白与色素所形成的色素蛋白复合体,作为光受体具有捕获光能并在光系统和系统传递激发能的功能。第 64~232 位包括典型的捕光叶绿素 a/b 结合蛋白功能域,其功能可能是与色素分子结合;第 9~11 位为蛋白激酶 C 的磷酸化位点,第 27~52 位为泛素结构功能域信号,第 55~58 位为酪氨酸激酶(CK2)磷酸化位点,这些位点可能与该基因转录后的调节有关(Allen *et al.*, 1997)。

*cab* 家族基因在进化过程中,其 cDNA 序列是相当保守的(向太和等,2005)。与禾本科中已经报道的 *cab* 家族基因聚类分析发现,绿竹与玉米的亲缘关系最近,接下来是水稻、小麦、大麦和两色蜀黍(*Sorghum bicolor*),这为利用同源基因的进化来推断禾本科各属间的进化关系(Clark *et al.*, 1995)提供了依据。

竹子光合作用的特征是异常的生长速度与生物量的增长,竹子四季常青的特点,使其成为重要的将太阳能转变为生物能量的潜在能源植物(Embaye,2001)。本研究首次从竹子中克隆了捕光叶绿素 a/b 结合蛋白基因,为进一步研究竹子高效光合作用的分子机制奠定了基础,同时 *cab-BO1* 基因的克隆为研究 *cab* 家族基因在竹子中的功能创造了条件。然而,竹子基因功能的研究将是一个漫长而艰难的探索过程,下一步的研究工作需要建立完善的绿竹遗传转化体系,通过基因的过量表达和抑制或敲除来研究基因的功能。

### 参 考 文 献

- 陈存及,邱尔发,梁一池,等. 2001. 毛竹不同种源光合特性研究. 林业科学, 37(6):15-19
- 黄承才,葛滢,常杰. 2000. 中亚热带东部毛竹叶片光合及呼吸的研究. 浙江林业科技, 20(5):14-16,46
- 黄勇. 2003. 绿竹种源的光合特性研究. 福建林业科技, 30(3):50-53,72
- 孙钦秒,冷静,李良璧,等. 2000. 高等植物光系统捕光色素蛋白复合体结构与功能研究的新进展. 植物学通报, 17(4):289-301
- 向太和,王利琳,庞基良. 2005. 水稻(*Oryza sativa* L.)捕光叶绿素 a/b 结合蛋白基因全长 cDNA 的克隆和特性分析. 作物学报, 31(9):1227-1232
- 郑炳松,金爱武,程晓建,等. 2001. 雷竹光合特性的研究. 福建林学院学报, 21(4):359-362
- Allen J F, Nilsson A. 1997. Redox signaling and the structural basis of regulation of photosynthesis by protein phosphorylation. *Physiol Plant*, 100:863-868
- Clark L G, Zhang W, Wendel J. 1995. A phylogeny of the grass family (Poaceae) based on ndhF sequence data. *Syst Bot*, 20:436-460
- Embaye K. 2001. The potential of bamboo as an interceptor and converter of solar energy into essential goods and services: focus on Ethiopia. *International Journal of Sustainable Development and World Ecology*, 8(4):346-355
- Gao Zhimin, Li Xueping, Li Lubin, *et al.* 2006. An effective method for total RNA isolation from bamboo. *Chinese Forestry Science and Technology*, 5(3):52-54
- Kozak M. 1999. Initiation of translation in prokaryotes and eukaryotes. *Gene*, 234(2):187-208
- Raghvendra A S. 1998. *Photosynthesis: a comprehensive treatise*. Cambridge:Cambridge Univ Press, 72-86
- Viret J F, Schantz M L, Schantz R. 1990. Nucleotide sequence of a maize cDNA coding for a light-harvesting chlorophyll a/b binding protein of photosystem. *Journal Nucleic Acids Res*, 18:71-79

(责任编辑 徐红)