

Construction of a cDNA Library and Cloning of an Arabinosidase cDNA from *Armillariella tabescens*^{*}

YAO Dong-sheng¹ HUANG Hui¹ ZHAO Long² XIE Chun-fang¹ LIU Da-ling^{1**}

(1 Laboratory of Biotechnology, Life Science and Technology College, Jinan University Guangzhou 510632, China)

(2 Department of Biological Sciences and Biotechnology, Tsinghua University Beijing 100084, China)

Abstract The expression cDNA library of *A. tabescens* was constructed by SMART technique, which use λ TriplEx2 as a vector. The titer and the percentage of the constructed library were about 1.0×10^6 pfu/ml and 98.3% respectively, and the titer and the capacity of the amplified library were about 3.1×10^8 pfu/ml and 4.2×10^{10} . The library was used to provide expressed sequence tags (ESTs). 147 Expressed Sequence Tags (ESTs) were gained from 176 clones, which were selected randomly and sequenced at the 5' end. The sequences were submitted to the EMBL database. Blasting the sequences in the GenBank, 43 of them were found that they have significant similarity with data in GenBank. EST AJ620046 has significant similarity with the arabinosidase of *Bacteroides thetaiotaomicron*. Using SMART-RACE a full-length cDNA of AJ620046 was successfully obtained. In order to initially characterize the biochemical properties of AJ620046, the ORF of AJ620046 named AF was cloned and expressed in *Pichia Pastoris* yeast. Recombinant pHIL-S₁-AF constructed by inserting AF into pHIL-S1 was transformed into *Pichia Pastoris* GS115. Preliminary experiments indicated that AJ620046 was expressed as a 32 kDa protein in recombinant yeast.

Key Words *Armillariella tabescens* Bioinformatics Arabinosidase Cloning

Armillariella tabescens is a fungus of *Typholomataceae*. It is parasitism on the body or root of tree and can cause many trees' decayed diseases^[1]. According to the literature, *A. tabescens* is a non-toxic, edible fungus with curative properties in the treatment of cholecystitis, hepatitis, appendicitis and otitis media, and some of the fungus can decompose mycotoxins^[2-4].

The expression cDNA library of *A. tabescens* was constructed by SMART technique that uses λ TriplEx2 as vector^[5-6]. The library was used to provide Expressed Sequence Tags (ESTs). 147 ESTs were gained from the cDNA library of *A. tabescens*. Analyzing these

sequences by bioinformatics^[7-9], EST AJ620046 has significant similarity with arabinosidase of *Bacteroides thetaiotaomicron*; the ORF of AJ620046 was cloned and expressed in *Pichia Pastoris* yeast.

1 Materials and Methods

1.1 Materials

A. tabescens (EJLY9812) and *E. coli* DH5 α was kept by our lab; *E. coli* XL1-Blue, *E. coli* BM25.8 and λ TriplEx2 vector were offered by the Clontech kits; pMD18-T was purchased from TaKaRa. cDNA library construction kits was purchased from Clontech; Oligotex mRNA extract kits was purchased from QIAGEN; TRIzol Reagent was purchased from Gibco GRL; lamada capsid Gigapack III Gold packaging extract was purchased from Stratagene cloning systems; *Bam*HI, *Hind*III, *Eco*RI, *Sal*I and *Xho*I was purchased from TaKaRa; SMART

Received February 21, 2004 Accepted May 10, 2005

* This work was supported by Grants from National "863" High-Tech Project (No. 2002AA213011), National Natural Science Foundation (No. 30270043) and Guangdong Key Task Project (No. C30109)

** Corresponding author, E-mail: tldl@jnu.edu.cn

RACE cDNA Amplification Kit was purchased from CLONTECH; *Pichia Pastoris* GS115, GS115 Albumin, GS115 β -gal and pHIL-S1 vector were purchased from Invitrogen; 4-Nitrophenyl- α -l-arabinofuranoside (pNPAF) was purchased from Sigma; PCR primer was synthesized by AUCT Company in Beijing.

1.2 Methods

1.2.1 Construction of *A. tabescens* cDNA library Total RNA was isolated from tissues of *A. Tabescence* using TRIZOL reagent. mRNA was enriched by oligo-(dT) cellulose chromatography. The cDNA library was prepared using the SMARTTM cDNA Library Construction Kit.

1.2.2 Random cDNA clones sequencing Following the SMARTTM cDNA Library Construction Kit, the cDNA library was transformed into competent *E. coli* XL1-Blue; white colonies were picked and translated into *E. coli*. BM25.8. DNA sequencing was completed by BIOASIA Company in Shanghai. Those cDNA inserts greater than 200bp were used in bioinformatics analysis.

1.2.3 Sequencing of full-length cDNA of AJ620046

The complete 5' sequence of AJ620046 was obtained by the 5' RACE method as described by SMARTTM Technology Overview. (5' primer: 5'-GAGGAAGAGGTGGGGATTGTGGCGAC-3' and PCR conditions: 94°C for 1 min, 5 cycles of 94°C for 30 s and 72°C for 4 min, 5 cycles of 94°C for 30 s and 70°C for 4 min, and 25 cycles of 94°C for 30 s and 68°C for 4 min, 72°C for 10 min). The PCR product was sequenced and spliced with the EST AJ620046, the full-length of AJ620046 was obtained.

In order to obtain the ORF of AJ620046 (named AF), end to end PCR was performed with the forward primer (5'-GCACTCGAGAAATGATTACGATGGG-AATCCTC-3', containing a *Xho*I restriction site) and the reversed primer (5'-CATGGATCCTACTTACTGCTTT-AGCACCGTG-3', containing a *Bam*H restriction site) by touchdown PCR (PCR conditions: 95°C for 1 min, 5 cycles of 95°C for 30 s and 72°C for 4 min, 5 cycles of 95°C for 30 s and 70°C for 4 min, 25 cycles of 95°C for 30 s and 68°C for 4 min, 72°C for 10 min).

1.2.4 Expression in yeast and determination of product AF was inserted into *Pichia* expression vector pHIL-

S₁, the recombinant pHIL-S₁-AF digested with *Sal*I was transformed into *P. pastoris* strain GS115 by spheroplasting method. Transformants GS115/pHIL-S₁-AF and negative control GS115/pHIL-S₁ were cultivated following the *Pichia Pastoris* expression Kit. The supernatant of GS115/pHIL-S₁-AF and GS115/pHIL-S₁ after filtration was concentrated 10 fold and kept at -80°C. SDS-PAGE was performed to monitor protein production.

The activity of arabinosidase was determined using a UV-visible spectrum as described by Pablo et al.^[10]. The reaction group was made up of 10 μ l 20mmol/L pNPAF, 180 μ l 50mmol/L citrate buffer (pH6.2) and 10 μ l supernatant (contain enzyme). The reaction was performed at 30°C for 20 min, and then stopped by addition of 0.75mol/L NaCO₃. The reaction group replaced the supernatant with boiled supernatant (at 100°C for 20min) was as negative control group 1, then the reaction stopped by NaCO₃ when the reaction begin was as negative control group 2. The enzymatic relative activity was defined as the percentage of p-nitrophenol released from pNPA.

1.2.5 Bioinformatic analysis Nucleotide and amino acid homology searches were conducted by using BLASTn and BLASTx programmes on the NCBI network service (<http://www.ncbi.nlm.nih.gov/>). We define a match if BLAST detects a hit with an E-value $\leq 1e-10$ and/or an overlapping region ≥ 80 base pair (bp) with $\geq 50\%$ identity as significant match^[11]. The ORF program Finder (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>). The protein Superfamily prediction analysis by Superfamily1.61 (<http://supfam.mrc-lmb.cam.ac.uk/SUPERFAMILY/hmm.html>). Structure and function domain search using programmes of Simple Modular Architecture Research Tool (<http://coot.embl-heidelberg.de/SMART/>).

2 Results

2.1 Synthesis of *A. tabescens* cDNA

The total RNA extraction was analyzed on 1.0% agarose gel (Fig. 1), OD260/OD280 = 1.958. The majority of the cDNA produced from LD-PCR was

analyzed on 1.0% agarose gel and a homogeneous smear of PCR product ranging from 0 to 2kb in size was observed, collecting the cDNA of 7 to 10 fragments to construct the library (Fig.2).

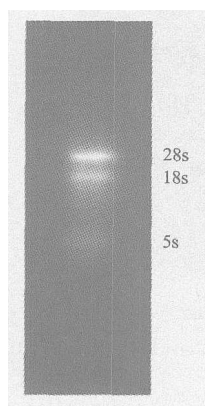


Fig.1 Electrophoresis of total RNA from *A. tabescens*

2.2 Quality assessment of *Armillariella tabescens* cDNA libraries

The titer and the percentage of the constructed library were about 1.0×10^6 pfu/ml and 98.3%; meanwhile, the titer and the capacity of the amplified library were about 4.2×10^{10} pfu/ml and 4×10^{10} .



Fig.2 Electrophoresis of cDNA fractionated

M: Marker; Lane 1 ~ 16: cDNA treated with *Sfi*I

2.3 Analysis of EST

ESTs more than 200 bp was used for bioinformatic analysis. The 147 ESTs of *A. tabescens* has been submitted to the EMBL database, Accession Numbers were AJ609126, AJ609394-AJ609417, AJ616844-AJ616846, AJ617271-AJ617280, AJ619997-AJ620105. 43 of the, *A. tabescens*, ESTs showed homology to previously described genes from other species (Table 1). 104 of them were found that they have no significant similarity with data in GenBank or EMBL.

Table 1 The results of the BLASTX analysis of ESTs from *A. tabescens* cDNA library

Submitted clones	Homology protein	Submitted clones	Homology protein
AJ609407	glyceraldehyde-3-phosphate dehydrogenase (CAD29456)	AJ620046	Arabinosidase (Q8A1K6)
AJ609411	NADH-ubiquinone oxidoreductase (Q00673)	AJ620047	aspartate-tRNA ligase (NP_588352)
AJ616846	alcohol dehydrogenase repressor protein (T41296)	AJ620051	ribosomal protein 10 (CAD70957)
AJ617272	Small heat shock protein (AAM78595)	AJ620052	small heat shock protein (AAM78595)
AJ617277	Dehydrogenase (NP_252789)	AJ620055	Transferase (NP_302039)
AJ617278	Small heat shock protein (AAM78595)	AJ620058	SIS ₁ protein (CAA72798)
AJ619997	carboxylic acid transport protein (AAO33825)	AJ620064	heat shock protein 90 (AAP44977)
AJ619998	PMPK (EAA16141)	AJ620066	GTP-binding protein (NP_593350)
AJ620001	Hydrophobin (CAA74987)	AJ620069	apolipoprotein A-I binding protein (NP_659146)
AJ620004	40s ribosomal proteins10 (NP_595605)	AJ620074	Small heat shock protein (AAM78595)
AJ620007	NADP-dependent malic enzyme (AAN63880)	AJ620075	Delta 1-pyrroline-5-carboxylate reductase (NP_594706)
AJ620010	c cytochrome P450(BAA33717)	AJ620085	alcohol dehydrogenase (AAM26268)
AJ620015	GTP-binding protein CG2522-PA (NP_524887)	AJ620089	guanosine-diphosphatase (CAD27295)
AJ620024	phenylalanine ammonium lyase (CAA09013)	AJ620091	transcription factor iib 70 kd subunit (NP_596265)
AJ620026	Cell division cycle 5-like (BAB62527)	AJ620092	Small heat shock protein (AAM78595)
AJ620027	40s ribosomal protein s11-2 (NP_594003)	AJ620094	GTP-binding protein (NP_594403)
AJ620028	LPS biosynthesis protein (NP_266373)	AJ620095	Beta-actin (BAA77815)
AJ620031	Sm-like protein(AAO50813)	AJ620098	PHD finger protein (NP_587791)
AJ620032	tripeptidyl peptidase A(BAC56232)	AJ620100	Down stream factor of ras (NP_593671)
AJ620035	integral membrane protein (NP_594041)	AJ620044	Lysophospholipase (NP_594165)
AJ620039	2-oxoglutarate dehydrogenase (NP_273993)	AJ620045	NADPH oxidase isoform 2 (AAQ74977)
AJ620040	pyruvate decarboxylase (AAP75899)		

2.4 Full-length cDNA of AJ620046

5' RACE PCR was generated a single prominent band with an approximate size of 400bp (Fig.3). The PCR product was transformed into *E. coli* DH5 α and digested with *Xho*I and *Bam*HI, a 400 bp band was observed by gel electrophoresis (Fig.4). Splicing the sequence of 5' RACE and the primal sequence (EST AJ620046), 1016-bp full-length cDNA was obtained for AJ620046 of *A. tabescens*. Using end-to-end PCR, a 924 bp open reading frame (ORF) of AJ620046 was obtained (Fig. 5), which encodes for a 308 amino acid protein, named AF. AF was inserted into *Pichia* expression vector pHIL-S₁ and tested by restriction analysis (Fig. 6).

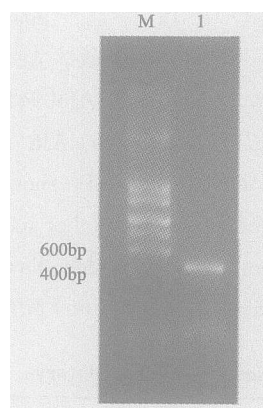


Fig.3 The result of 5' race products

M: Marker 1kb ladder; 1: 5'PCR products

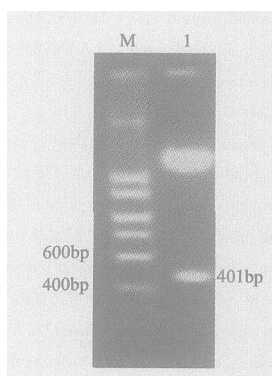


Fig.4 Restriction analysis of 5' race

1: PCR products/ *Eco*RI + *Hind*III(DH5 α); M: Marker

2.5 Expression in *P. pastoris* GS115 and determination of product

Use SDS/PAGE to analyze secreted protein in the media. Compared with the negative control, the

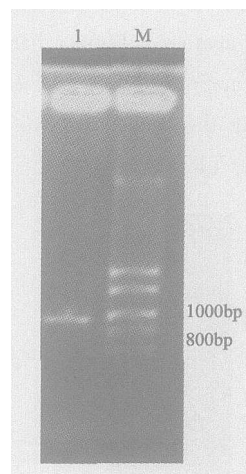


Fig.5 The result of end to end PCR

1: End to end PCR products; M: Marker

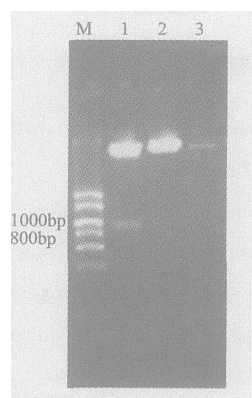


Fig.6 Restriction analysis of plasmids pPAF

1: Plasmid pHIL-S₁ containing the AF inserts; M: Marker

supernatant of Transformant GS115/pHIL-S₁-AF had a 32 kDa protein band (Fig. 7). The enzymatic relative activity determination showed that the percentage of released p-nitrophenol from pNPA of reaction group was 4%, the negative control group 1 was 0% and the negative control group 2 was 1.5%. These results direct that the supernatant contained arabinosidase.

2.6 Bioinformatic analysis of AJ620046

The full-length sequence of AJ620046 was submitted to the EMBL database to update. The complete sequence of AJ620046 is 1016bp and the putative open reading frame (ORF) is extended from a start codon at nucleotide 7 to a stop codon at nucleotide 930. The deduced translation product is a 308 amino acid residual protein (named AF) with a predicted molecular mass of 32916.04 Daltons; the isoelectric point (pI) is 4.636. At the 5' end of the coding region,

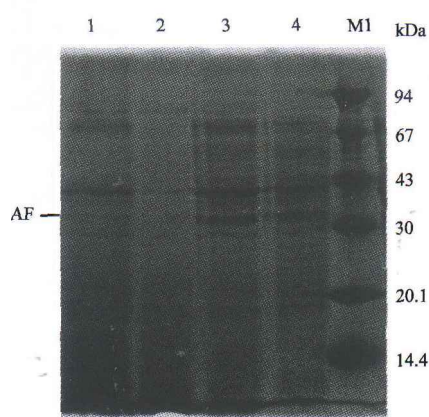


Fig.7 SDS-PAGE analysis of supernatant of recombinant GS115/pHIL-S₁-AF

1: Induced supernatant of negative control transformant GS115/pHILS₁ (concentrated 10 fold); 2: Induced supernatant of negative control transformant GS115/pHILS₁. 3: Induced supernatant of recombinant GS115/pHIL-S₁-AF (concentrated 10 fold); 4: Induced supernatant of recombinant GS115/pHIL-S₁-AF; M1: Protein molecular-mass marker

AF displays a strong hydrophobic region and a signal peptide sequence. 1 ~ 30 amino acid residue (aa) of AF was a typical hydrophobic domain. 5 ~ 27 aa of AF was a transmembrane domain, the mature peptide chain began at 22 aa, which is likely to be a shearing site, and at 5 ~ 27 aa is a signal peptide. The PredictProtein of CUBIC results show that AF has high homology with data of the arabinosidase of *Bacteroides thetaiotaomicron*. Superfamily analysis shows that AF has two superfamilies, one is alpha-L-arabinanase (synonym of arabinosidase) at 78 ~ 219 aa, the other is DFP domain at 25 ~ 304 aa (Fig. 8). The analytic result of the structure and function domain shows that AF has a typical SCOP domain at 25 ~ 304 aa (Fig. 9).

3 Discussion

To our knowledge, this is the first description of a cDNA library from *A. tabescence*. The construction and

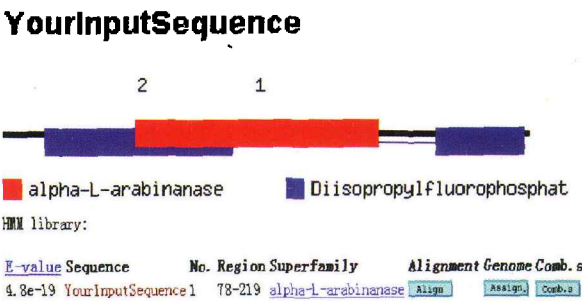


Fig.8 The result of AF superfamily

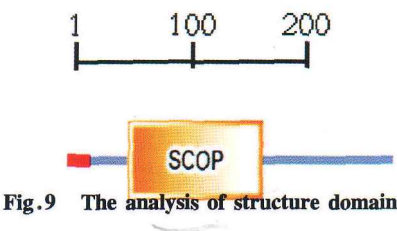


Fig.9 The analysis of structure domain

assessment of the cDNA library is a valuable resource for more detailed EST investigation and identification of full-length genes of interest. To ensure synthesize integrated cDNA, the SMART™ cDNA library construction kits were used to construct a relatively high percentage full-length cDNA library. The expression cDNA library of *A. tabescens* was constructed using λTriplEx2 as vector to enhance the possibility of low quantity genes. 147 ESTs were gained from the cDNA library of *A. tabescens*. Making Blastx search in

GenBank, AJ617272, AJ617278, AJ620038, AJ620052, AJ620074, AJ620092, AJ620064 all displayed more than 50% identity to heat shock protein, AJ617274, AJ620004, AJ620027 displayed more than 50% identity to 40s ribosomal protein, which shows that the two kind of protein are in high quantity in the cDNA library of *A. tabescens*. Based on our subjective criteria for *A. tabescens* gene identification of a minimum aa sequence identity of 25% over a contiguous series of 80 aa^[11], 43 of the *A. tabescens* ESTs showed homology to previously described genes from other species, showed in table 1. Analyzing the 5' region of AJ620046 full-length cDNA, a conservative sequence Kozak was detected at the second ATG: ACGATGG. The conservative sequence Kozak (G/ANNATGG) is the best promote sequence, available to the effective expression of

gene^[12]. The downstream of it has the termination codon TAA and 36bp Poly (A)⁺. These results suggest that the cDNA sequence of AJ620046 is a full-length sequence^[13, 14]. AJ620046 sequence shows significant similarity to the superfamily of alpha-L-arabinanase, and contains a typical SCOP domain. The AF was cloned and expressed in *Pichia Pastoris* yeast; the induced supernatant of recombinant GS115/pHIL-S₁-AF can decompose pNPA and release p-nitrophenol. Preliminary experiments indicated that AJ620046 was a novel Hemicellulose gene, which has the arabinosidase activity. All data from the present study implied that AJ620046 is a new member of arabinosidase.

Searching the new hemicelluloses from fungi, cloning and reforming the relative genes by genetic engineering to obtain mass production of high activity hemicelluloses fitted for the industry application is very important. This work provided important basic research for further exploitation of *A. tabescence*.

References

- [1] Mao X L. The Macrofungi in China. Henan Science & Technology Press. 2000.187
- [2] 陈伟强.假密环菌杂多糖 Y-HW 对急性肝损伤小鼠血浆中 TNF- α 和 IFN- γ 水平的影响. 广东药学院学报, 2000, 16 (1): 44 ~ 45
Chen W Q. Academic journal of GuangDong college of pharmacy. 2000, 16 (1): 44 ~ 45
- [3] Liu D L, Yao D S, Liang R, et al. Detoxification of aflatoxin B1 by enzymes isolated from *Armillariella tabescens*. Food Chem Toxicol, 1998, 36:563 ~ 574
- [4] Yao D S, Liang R, Liu D L, et al. Screening of the fungus whose multienzyme system has catalytic detoxification activity towards aflatoxin B1 (Part I). Ann N Y Acad Sci, 1998, 864:579 ~ 585
- [5] Seth D, Gorrell M D, McGuinness P H, et al. SMART amplification maintains representation of relative gene expression: quantitative validation by real time PCR and application to studies of alcoholic liver disease in primates. Biochem Biophys Methods, 2003, 55:53 ~ 66
- [6] Reid E A, Cao Z, Wang Y L, et al. Molecular cloning and identification of a putative PKC α cDNA from *Limulus polyphemus* brain. Life Sciences, 2003, 72:961 ~ 976
- [7] Yu L, Frith M C, Suzuki Y, et al. Characterization of genomic organization of the adenosine A2A receptor gene by molecular and bioinformatics analyses. Brain Research, 2004, 1000:156 ~ 173
- [8] Bettinotti M P, Olsen A, Stroneck D. The use of bioinformatics to identify the genomic structure of the gene that encodes neutrophil antigen NB1, CD177. Clinical Immunology, 2002, 102(2):138 ~ 144
- [9] Hsu S Y T. Bioinformatics in reproductive biology-functional annotation based on comparative sequence analysis. Journal of Reproductive Immunology, 2004, 63(1):75 ~ 83
- [10] Pablo D I, Alessandra P, Jeannette S, et al. An a-L-arabinofuranosidase from *Penicillium purpurogenum*: production, purification and properties. Journal of Biotechnology, 2000, 76: 253 ~ 258
- [11] 张成岗, 贺福初. 生物信息学方法与实践. 北京: 科学出版社, 2002. 63
Zhang CH G, He F CH. Bioinformatic methods and practice. Beijing: Science Press, 2002.63
- [12] Huang T, Sahud M A. Association of C807T, P1^A, and 5 C/T Kozak genotypes with density of glycoprotein receptors on platelet surface. Thrombosis Reserch, 2003, 112(3):147 ~ 150
- [13] Christian Z, Peter S J, Jia B W, et al. Control of μ opioid receptor expression by modification of cDNA 5'-and 3'-noncoding regions. Molecular Brain Research, 2000, 79(1 ~ 2):159 ~ 162
- [14] Jun Z, Richard E C, Daniele R, et al. Cloning of the murine kritl cDNA reveals novel mammalian 5' coding exons. Genomics, 2000, 70(3):392 ~ 395

假密环菌 cDNA 文库的构建及其阿拉伯糖苷酶基因的克隆

姚冬生¹ 黄辉¹ 赵龙² 谢春芳¹ 刘大岭¹

(1 暨南大学生命科学与技术学院 生物技术实验室 广州 510632 2 清华大学生物科学与技术系 北京 100084)

摘要 用 SMART 技术构建了载体为 λ TriplEx2 的假密环菌的表达型 cDNA 文库。文库滴度为 1.0×10^6 pfu/ml, 重组率约为 98.3%, 扩增后滴度为 3.1×10^8 pfu/ml, 容量约为 4.2×10^{10} 。从文库中随机挑选 176 个克隆进行 5'端测序, 得到 147 条表达序列标签(EST), 并将测序结果提交到 EMBL 数据库。随机测序结果表明: 插入片段长度均在 200 ~ 800 之间。测序结果经过生物信息学分析, 发现有 43 条序列与已知序列有明显同源性, 其中序列 AJ620046 与多形拟杆菌的阿拉伯糖苷酶序列有很高的序列一致性。用 SMART-RACE 技术成功获得了 AJ620046 的全长 cDNA, 克隆了 AJ620046 的开放阅读框 AF, 并成功构建了重组质粒 pHIL-S₁-AF, 在毕赤酵母菌中进行了初步表达。

关键词 假密环菌 生物信息学 阿拉伯糖苷酶 克隆