African Journal of Agricultural Research Vol. 4 (4), pp. 402-408, April, 2009
Available online at http://www.academicjournals.org/AJAR
ISSN 1991-637X © 2009 Academic Journals

Full Length Research Paper

Cloning and analysis of glyceraldehyde-3-phosphate dehydrogenase gene from *Cordyceps militaris*

Zhenhua Gong[1], Ying Su[1], Lei Huang[1], Juan Lin[2], Kexuan Tang[1] and Xuanwei Zhou[1]*

[1]Plant Biotechnology Research Center, School of Agriculture and Biology, Fudan-SJTU-Nottingham Plant Biotechnology R and D Center, Shanghai Jiao Tong University, Shanghai 200240, People's Republic of China.
[2]State Key Laboratory of Genetic Engineering, School of Life Sciences, Fudan-SJTU-Nottingham Plant Biotechnology R and D Center, Fudan University, Shanghai 200433, People's Republic of China.

Accepted 25 February, 2009

A gene encoding a glyceraldehyde-3-phosphate dehydrogenase (GPD) gene was isolated from *Cordyceps militaris* using degenerate PCR and Thermal Asymmetric Interlaced PCR (TAIL-PCR) technology. Analysis of 4493 bp segments (*Cmgpd*) revealed the cloned gene contains a 2515 bp 5' upstream region, a 1296 bp coding region and a 682 bp 3' downstream region. The coding region contains a 279 bp intron. After cutting the intron, the open reading frame (ORF) with 1017 bp encodes a polypeptide of 338 amino acid residues. The deduced amino acid sequence indicates a proprotein with a molecular weight of 36.18 kDa. There are one TATA box and two possible CAAT boxes lying in the 5' upstream region. The deduced amino acid sequence of *C. militaris* GPD shared different homology (ranging from 77-94%) with *gpd* genes from yeast and filamentous fungi species, such as *Beauveria bassiana*, *Gibberella zeae*, *Myrothecium gramineum*. The cloning of the gene not only provides a basis for the further investigation of its structure, expression and regulation mechanism, but also the upstream promoter of *Cmgpd* has the potential use for directing high and constitutive expression of homologous and heterologous genes.

Key words: Glyceraldehyde-3-phosphate dehydrogenase, *Cordyceps militaris*, TAIL-PCR, promoter.

INTRODUCTION

Glyceraldehyde-3-phosphate dehydrogenase (GPD, E.C. 1.2.1.12) is one of the key enzymes in the Embden Meyerhof Parnas or glocolysis pathway. It catalyzes phosphorylation of glyceraldehyde-3-phosphate to pro-duce 1, 3-diphosphoglycerate. GPD is a tetrameric enzyme composed of four identical subunits. It is an essential enzyme used to maintain life activities through contributing in this way to the formation of ATP and providing additional energy to the cell by reducing NADH to NAD+ and H+ upon its action. Moreover, GPD protein also has many other important functions, such as abiotic stress tolerance (Liu and Yang, 2005). Because of its critical function in every living cell, its expression in *Saccharomyces cerevisiae*, *Aspergillus nidulans* and other eukaryotic organisms is very high representing up to 5% soluble cellular proteins (Punt et al., 1990; Piechaczyk et al., 1984) and its mRNA in yeast also accounts for 2-5% total mRNA (Holland et al., 1978). The abundance of GPD protein or mRNA suggests that the *gpd* gene is regulated by a constitutively and highly active promoter. It has been proved by the fact that the expression of heterologous genes in *S. cerevisiae* (Bitter and Egan, 1984), *Pichia pastoris* (Döring et al., 1998), *Lentinula edodes* (Hirano et al., 2000), *Mucor circinelloides* (Wolff and Arnau, 2002) and *Flammulina velutipes* (Kuo et al., 2004) are successfully directed by their native *gpd* gene promoters.

The medicinal mushroom *Cordyceps militaris* belongs to vegetable wasps and plant worms and is used as a tonic food and herbal medicine. It is a species of family *Cordycipitaceae* in order *Hypocreales* of class *Sordariomycetes*. *C. militaris* is a model fungus in genus *Cordyceps* and can be used to replace *Cordyceps sinensis* (Zhou et al., 2009). Cordycepin (3'-deoxyadenosine) and Cordyceps polysaccharides are pharmacologically active constituents in both *C. militaris* and *C. sinensis*. They play an important role in counteracting tumors, preventing kidney and liver diseases, soothing the lung, staunching

*Corresponding author. E-mail: xuanweizhou@sjtu.edu.cn. Tel: +86-21-34205778. Fax: +86-21-65642425.

Table 1. Primers used in the process of TAIL-PCR.

Feature	Primer name	Sequences direction (5'-3')
Arbitrary degenerate (AD) primers	AD1	NTCGASTWTSGWGTT
	AD2	NGTCGASWGANAWGAA
	AD3	TGWGNAGWANCASAGA
	AD4	AGWGNAGWANCAWAGG
	AD5	STTGNTASTNCTNTGC
	AD6	WGTGNAGWANCANAGA
Nested specific primers used in the first upstream TAIL-PCR	SP1-L1	AACACCGTGGGAGGAGTCATA
	SP2-L1	TGACGAGTGTTCGCCTTTGG
	SP3-L1	AAACTCACGGCGTACTTGACC
Nested specific primers used in the second upstream TAIL-PCR	SP1-L2	CGTGGTAGATGGTCAAGAAGCAA
	SP2-L2	TGTGGGACGGATGGAAAACT
	SP3-L2	GAGGGGTGGGCAAATGGTTT
Nested specific primers used in the third upstream TAIL-PCR	SP1-L3	AAACTTTGACCACCCGCTATG
	SP2-L3	CACGGCGGAACTACCAGAATA
	SP3-L3	CACGACAAGGCTCCCATCCA
Nested specific primers used in the first downstream TAIL-PCR	SP1-R1	TGCTCAAGTATGACTCCTCCCA
	SP2-R1	CTCCCACGGTGTTTTCAAGG
	SP3-R1	TCAACGGCAAGAAGATTCGC
Nested specific primers used in the second downstream TAIL-PCR	SP1-R2	CCACTCCTACACTGCTACCCAG
	SP2-R2	ATGTCCATGCGTGTCCCTACC
	SP3-R2	TCAAGGAGGCTGCTGAGGGC

bleeding and dispersing phlegm (Zhou et al., 2009),but the molecular mechanism of their pharmacological activeties is not fully understood. For better analysis, the synthesis pathway of cordycepin and Cordyceps polysaccharides and genetically modifying the *Cordyceps* species to improve the production of these active components, stable and highly effective transformation system is required in modern fungal research. Here, we reported the cloning and sequencing the *C. militaris gpd* gene (*Cmgpd*). Coding region and 5'- and 3'-flanking sequence of *Cmgpd* were systematically analyzed by bioinformatics method. This study established a basis for clarifying the structure of *C. militaris* GPD protein at the molecular level, and could provide a potential promoter element to a highly effective transformation vector in future.

MATERIALS AND METHODS

Microorganisms and culture conditions

A strain of *C. militaris* was provided by Heilongjiang Xinyisheng Pharmaceutical Co., Ltd. *Escherichia coli* DH5α was kept in our laboratory. *C. militaris* was incubated in Potato/Dextrose liquid medium 5 d at 27 and 128 rpm for DNA and total RNA isolation, and stored in PDA slants at 4 . *E. coli* DH5α was cultured in Luria-Bertani (LB) medium. Liquid *E. coli* DH5α cultures were incubated at 37 and 200 rpm.

Genomic DNA isolation, degenerate PCR and TAIL-PCR

Genomic DNA was prepared from *C. militaris* mycelium (7 d postinoculation) using a CTAB method as previously described (Zhou et al., al, 2008). Based on the amino acid sequences of the conserved regions of fungi GPD protein in order *Hypocreales* published in NCBI, two degenerate primers, gpd-F (5'-ATCAACGGNTTCGGNCGN ATTGG-3') and gpd-R (5'-CATNACGTACATGGGNGCATC-3') were designed. Degenerate PCR was done using these primers and *C. militaris* genomic DNA as template. After that, a core fragment of *C. militaris gpd* gene was obtained. PCR amplification was carried out with an initial denaturation of 94 for 5 min, a 30 cycle of 94 for 30 s/45 for 30 s/72 for 1 min, and a final elongation of 72 for 10 min.

Flanking sequences of this core fragment of *C. militaris gpd* gene were amplified by using Thermal Asymmetric interlaced PCR (TAIL-PCR) technique. All nested specific primers designed based on the obtained nucleotide sequences, and arbitrary degenerate (AD) primers selected from previous literatures (Liu and Whittier, 1995; Liu et al., 1995; Tsugeki et al., 1996) are listed in Table 1. Thermal conditions for TAIL-PCR were modified and a tertiary round TAIL-PCR was added to decrease the amount of contaminating nonspecific products (Table 2). TAIL-PCR reaction mixture in every round TAIL-PCR was described previously (Liu and Whittier, 1995).

RNA isolation and RT-PCR

Total RNA of *C. militaris* was extracted from mycelium (5 d postinoculation) with RNAprep pure Plant Kit (TIANGEN). Total RNA of *C. militaris* was used for first strand cDNA synthesis with M-MLV RTase cDNA Synthesis Kit (TaKaRa). The synthesized cDNA was used directly as template for PCR using degenerate primers (gpd-F /gpd-R) and thermal condition that was the same as above where genomic DNA was used as template.

Cloning and sequencing of PCR products

For analysis of PCR products, each reaction was subjected to 1% agarose gel electrophoresis in TAE buffer. Target bands were cut from agarose gel and purified by EZ-10 Spin Column DNA Gel Extraction Kit (TIANGEN). Target products were ligated into the pMD18-T vector (TaKaRa). Then, plasmid DNA was transformed into *E. coli* DH5α by heat shock method. Colony PCR was used to validate positive clones which were subsequently sequenced by Invitrogen (Shanghai). Seq-

Table 2. TAIL-PCR programs.

Reaction	Cycle no.	Reaction thermal and time condition		
Primary round TAIL-PCR	1	94°C 3min		
	5	94°C 30s	62°C 1min	72°C 2.5min
	2	94°C 30s	25°C 3min (0.4°C /s ramp)	72°C 2.5min (0.3°C /s ramp)
	15	94°C 10s	68°C 1min	72°C 2.5min
		94°C 10s	68°C 1min	72°C 2.5min
		94°C 10s	44°C 1min	72°C 2.5min
	1	72°C 5min	4°C hold	
Secondary round TAIL-PCR	1	94°C 3min		
	5	94°C 10s	64°C 1min	72°C 2.5min
	15	94°C 10s	64°C 1min	72°C 2.5min
		94°C 10s	64°C 1min	72°C 2.5min
		94°C 10s	44°C 1min	72°C 2.5min
	1	72°C 5min	4°C hold	
Tertiary round TAIL-PCR	1	94°C 3min		
	20	94°C 10s	60°C 1min	72°C 2.5min
		94°C 10s	60°C 1min	72°C 2.5min
		94°C 10s	44°C 1min	72°C 2.5min
	1	72°C 5min	4°C hold	

Sequence comparisons to database were performed using BLAST program (National Center for Biotechnology Information).

RESULTS

Cmgpd cloning

Using degenerate primers gpd-F/gpd-R designed based on the conserved regions of GPD proteins in order Hypocreales, PCR was done with *C. militaris* genomic DNA as template, and then a fragment of approximately 650 bp was obtained and contributed as shown in Figure 1. Through sequencing and subsequent blasting against the NCBI database, deduce amino acid sequence of this 657 bp fragment which is interrupted by a 279 bp predicted intron shows a high amino acid identity with *Beauveria bassiana* GPD protein (AAT80324) and *Hypocrea jecorina* GPD protein (ABK33667). It suggests that this 657 bp fragment is a part of coding region in *C. militaris* homolog of *gpd* gene.

Based on this core fragment of *C. militaris gpd* gene, three times upstream and two times downstream TAIL-PCR was used to amplify the whole genomic sequence of *C. militaris gpd* gene. Figure 2 shows an example of TAIL-PCR technique. High amounts of type II nonspecific products primed by the AD primer alone were visible in the primary TAIL-PCR while disappeared in the seconddary TAIL-PCR (Figure 2 indicated by 250 bp band in Lane 16). Some type 3 nonspecific products primed by the nested specific primer alone could appear
in the se-condary TAIL-PCR, but could not be observed
in the tertiary TAIL-PCR (Figure 2 indicated by 2200 bp band in Lane 17). Only type 1 specific target products primed by nested specific primer and AD primer together, which was corroborated through a decrease in product size accord-ing to the position of the nested specific primers, still ap-peared in the tertiary TAIL-PCR. Because the lengths of multiple bands observed in the tertiary TAIL-PCR pro-ducts are ascribed to the annealing position of AD pri-mers in *Cmgpd*, we just should select the longest product in the tertiary round TAIL-PCR for sequencing (Figure 2 indicated by 1470 bp band in lane 12). Flanking sequen-ces of 571 bp, 803 bp and 1407 bp at 5' end were res-pectively obtained after three times of TAIL-PCR. And flanking sequences of 702 bp and 991 bp at 3' end were
respectively obtained after two times of TAIL-PCR. After assembly of all flanking sequences obtained from TAIL-PCR and the core fragment obtained from degenerate PCR, a genomic fragment of 4493 bp in size was obtained. BLAST showed that this fragment contains a full genomic *gpd* gene which is designated *Cmgpd* (*Cordyceps militaris* glyceraldehyde-3-phosphate dehydrogenase, GenBank Accession No. FJ374269).

Characterization of *C. militaris gpd* gene

Through sequence analysis, it was deduced that *Cmgpd* contains a 1296 bp coding region interrupted by a 279 bp intron. (Figure 3) The deduced GPD protein of 338 amino acids has a calculated size of 36.18 kDa and an estimated *IP* value of 6.52. It has been reported that there

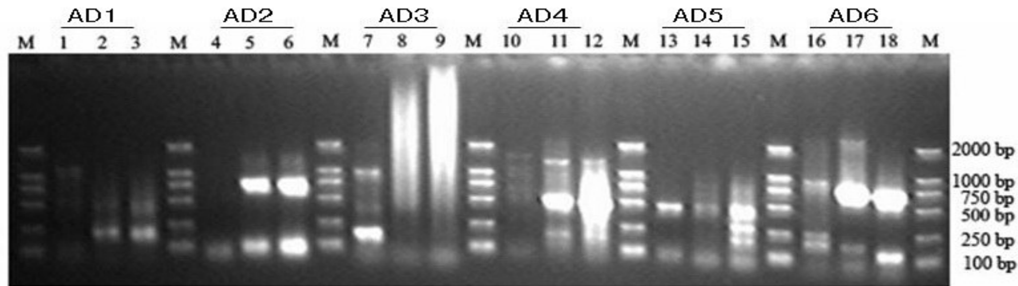

Figure 1. The third TAIL-PCR for 5'-flanking sequence of Cmgpd.

M: DNA Marker (DL 2000 marker); Lanes 1, 4, 7, 10, 13, 16: the primary round TAIL-PCR with specific primer SP1-L3 and one AD primer (AD1-AD6, respectively); Lanes 2, 5, 8, 11, 14, 17: the secondary round TAIL-PCR with specific primer SP2-L3 and one AD primer (AD1-AD6, respectively); Lanes 3, 6, 9, 12, 15, 18: the tertiary round TAIL-PCR with specific primer SP3-L3 and one AD primer (AD1-AD6, respectively).

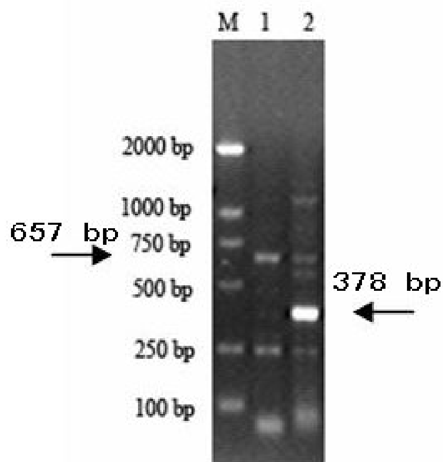

Figure 2. PCR and RT-PCR products using same degenerate primers.

M: DNA Marker (DL 2000 marker); 1: PCR products using *C. militaris* genomic DNA as template and degenerate primers gpd-F/gpd-R; 2: RT-PCR products using *C. militaris* cDNA as template and degenerate primers gpd-F/gpd-R.

there are 9, 9, 6, 5 and 1 introns in *Basidiomycetes*, *Agaricus bisporus gpd1* and *2, Phanerochaete chrysosporium gpd*, *Schizophyllum commune gpd*, and *Ustilago maydis gpd* respectively; in *Ascomycetes*, *A. nidulans*, *Curvularia lunata* and *Cryphonectria parasitica gpd genes* contained 6, 4 and 2 introns, respectively. The predicted intron in *Cmgpd* is just at the position where only one intron is conserved both in basidiomycetes and ascomycetes (Harmsen et al., 1992). The predicted *C. militaris* GPD protein shows high similarities with predicted GPD amino acid sequences of *B. bassiana* (AAT80324), *Gibberella zeae* (XP386433), *Myrothecium gramineum* (ABQ42571) and *A. nidulans* (AAA33307) (94, 88, 86 and 77%, respectively).

It is well known that all GPD enzymes contain a consensus substrate binding region: [ASV]-S-C-[NT]-T-{S}-x-[LIM]. The cysteine in this region is essential for the enzymatic activity since it functions as the binding site in the catalytic region. This pattern is also in the *C. militaris* GPD protein as ASCTTNCL. Another conserved catalytic histidine is found at position 179; conserved lysine and arginine phosphate-binding residues occur at position 194 and 234, respectively (Van and Yoder, 1992) (Figure 3).

Sequence analysis of *Cmgpd* 5' and 3'-flanking sequences

A fragment with 2515 bp upstream of the *Cmgpd* ATG start codon was recognized as the original *Cmgpd* promoter region. Analysis of transcription start sites using Neural Network Promoter Prediction (http://promotor.biosino.org/) found a core promoter sequence: agccacccaggactacaaaacc atttgcccacccctcctcTtcgccgtgc (score=0.15). The Box letter T shows the site of transcription start. Through PlantCARE analysis, one TATA box (TACAAAA) was found in the 5'-flanking region, while two deduced CAAT boxes (CCAAT) found were 1483 and 1732 bp upstream of the *Cmgpd* ATG start codon, respectively. Some typical conserved elements such as the gpd box (897 bp upstream of the ATG) and C+T-rich regions (313 bp upstream of the ATG) were found in *Cmgpd* while compared with promoter regions of *gpd* genes in *Aspergillus* and *Neurospora*. (Figure 3) It suggested that the *Cmgpd* promoter region contains some cis-acting transcript elements. The gpd box was an essential element in regulating the transcription of *gpdA* in *A. nidulans*, herein, when deletion of gpd box resulted in a 50% decrease of activity. Moreover, C+T-rich regions were possibly involved in the accurate starting of transcription (Punt et al., 1990). There also exist some regions of direct-or inverted-repeat sequence in the *Cmgpd* promoter, which are the same as those found in *A. nidulans* PgpdA. The intron/exon borders in *C. militaris gpd* gene (GTAAGT::CAG) and the splicing signal (GCTAAC, 20 bp

```
1     agtgtagtat catagcccaa ggacgtagag ttgaggctct tcagatcaat ggagacgccc ttgtcgatgt tattgaaccc aggattggga ttatggcctc
101   cagatttgac tgagaactgg cagtaattgg ttgtgagaag ctcgacaatg acagacatgt ctcgggacga tgagggcgtc actacacatc cgggatgctg
201   catggaagtg gtggccgcgg atctcggagc tcgggttcac ggcgagcacg ccgaggaaac aacagaactc ttgaagatga ggagaatcga gcagggaagg
301   gcccaagcca ttccactcaa cacacgagct tgtgccaact agagcggaac aggatccttg ccacgactgc attctagaat gccgaattgg agccctgggg
401   tagttagtgg gaatcgagac cgttgtctac gccccgggtt cttcccccag gaaagcggct cgcttgctgg tccaatcttc tgccgagcag ggccggggcc
                  A                                                       a
501   agacggcttt gggcggctgc ttgctctggg ccggacaagg aaggaggaca agccaaagca acgcacccgt tctgctccgg ccctggggcg gaagaagtgg
601   aacaagacgg gcagcgcagc cctgcggtta tcgtgatgcg cccgaacttt gccgcgtcta agtgggacag agcgtgccat tatatagtaa caggaaacca
701   aaggacgaaa atgcagcacc tggttccgat gcttgtgtaa agagcaaaat catagaggtt tgcgcacgac gccgtgctgg cctccccaca cctgcgatgg
801   ttatgcaatt acaagagatt tgatccctct ccccctcttc taccggaatt tgctatgagt gatttttcca ttggcagcgc gctgttttca ctgacctggt
                 A
901   tcttatcggc ggctgcgcag cggagtgcac tgtactggcc tagcgctgtt gcagatacag gatgaacagg gacgacgggg gtacttgagg agtactgagc
1001  ttggtgtgca acaaaggcag gcaaccgggt cgaatcggag ggcagatgct ggtacctcag aggttcccca caccggaccg gcctcagaaa gcgatttaat

1101  tttagcattt tcgcagggcg tcaaacagtg aagggcaaag agaagaaaaa aagtggcgat atacgtaggt aattatttca agacccgctc ccaaaaaaga
                      a
1201  cgcagtgcag tgccaacagg gtgcgttgac gctgccgcat gctgaggctg gtgaatatga ctagccggag acattgctat tcgggccttg tctgttatcg
1301  attcggtgcc caggcgcata ttacggcgtg agtgtgtaat gggaatggag tcagccggcg ccagatcgga ggacgctgaa tgaggaccaa gagtgtactt
1401  ttcccggcgc ggccggacta agacggtttc actaatacgc catggaatca tggatgggag ccttgtcgtg tcatcttggt ggccaatttg tctattctgg
                b                                                c                     CAAT-Box
1501  tagttccgcc gtggccaact tgacggttca ctggttttgt gtcttgatga ctgacaaaga gggccagaga tgagacgacg agctgccggt tgcaccatca
         d                      B
1601  gccaaccgtt cgcagaaatc atcagttgcg ctacagtcaa catagcgggt ggtcaaagtt tgcgtgggca cgggttgccg cacgagttgg aacgtcagta
1701  aagtggggga ggggaagcaa ccagaacgag gccaattgaa tggggatgaag gcgtctggga gggaagcacg agcgtgaatg aatggacggt tgcgtgactc
                                       CAAT-Box                   e
1801  tggtctcgcg atttcgatcg agtcaaggaa cccgtatggt tgcctcttgt cttggagagc tcgtcgaggc ccaagctgcg agtatatggt tggtgatgcg
               C              D                                    E           B
1901  ctgggcattt gcccctctct gcccctccat ggacggacct gcagctcttg ggctggcaag gtgttgcgct acgtcgagat gcggttaggt aagctggcat
                     d                                             c           e
2001  aagctggcat aaggttggct gttcttgaga aatcaggcca gcagtcttgc acctctctgg ctacaggctc tccagcggtc agccaccgcc actccagcgc
                                                           E              b
2101  ctgtttttca ctgacgatgg ccttgttttt ttgttgctct ggcagctggc aagctgggtc ccggcagctg ccacggccaa gccacccagg actacaaaac
         C                ↓                                D                                            TATA-Box
2201  catttgccca cccctcctcT tcgccgtgcg tctcttcttt ttctcccctc cattacacct cattcttcag ttcaaaggaa caaaagtgag ttttccatcc
              transcription start site
2301  gtcccacacc tggctccatc atctcgagct tcattgcttc ttgaccatct accacgctct accctccaac agcttcgctc acccaagagc tgcctgcttt
2401  tcccctcctc tccacgaccc cgccatccac cacttaacct aaagttaccc ctcgctaacc gccaaacgtc ccaacaggtt tttcttctct cacctcactt
2501  ttctaatcaa gaacaATGGC TCCCGTCAAG GTTGGCATCA ACGGCTTCGG CCGCATTGGC CGCATTGTCT TCCGCAACGC CGTCGAGCAC AACGACATCG
1                    M   A   P   V   K   V   G   I   N   G   F   G   R   I   G   R   I   V   F   R   N   A   V   E   H   N   D   I   D
2601  ACGTTGTTGC CGTCAACGAC CCCTTTGTTG AGGTCAAGTA CGCCgtgagt ttcaccaccc gcaccgaata aagctatacc cttgcctgca ccccgacaca
30      V   V   A   V   N   D   P   F   V   E   V   K   Y   A
2701  gccttgctgt tatggctgtg gctgtggctg tggctgtgac gtggtaggat tgtcaatttt ggcaacccgc tcaaacagct atcaggccac gctgccgctg
2801  cttgtgacca ggttatggtt tggttggtaa agcgaaccaa acctcataat cgctctacta gtgccaaagg cgaacactcg tcaacctcat ttgccaataa
2901  gctgctaacc attcatcgcg cagGCCTACA TGCTCAAGTA TGACTCCTCC CACGGTGTTT TCAAGGGCGA GATTGCCATT GACGGCAACG ATCTCGTCGT
44                              A   Y   M   L   K   Y   D   S   S   H   G   V   F   K   G   E   I   A   I   D   G   N   D   L   V   V
3001  CAACGGCAAG AAGATTCGCT TCTACGGCGA GCGCGACCCC GCCGCCATTC CCTGGAAGGA GACCGCCGCC GAGTACGTTG TCGAGTCCAC TGGTGTCTTC
70     N   G   K   K   I   R   F   Y   G   E   R   D   P   A   A   I   P   W   K   E   T   A   A   E   Y   V   V   E   S   T   G   V   F
3101  ACCACCATCG ACAAGGCCAA GGCTCACTTG CAGGGTGGTG CCAAGAAGGT CATCATCTCG GCCCCCTCTG CCGACGCCCC TATGTACGTG ATGGGTGTCA
103   T   T   I   D   K   A   K   A   H   L   Q   G   G   A   K   K   V   I   I   S   A   P   S   A   D   A   P   M   Y   V   M   G   V   N
3201  ACGAGAAGGC TTATGACGGC TCCGCCGACA TCATCTCCAA CGCCTCTTGC ACCACCAACT GCCTGGCTCC CCTCGCCAAG GTTGTCAACG ACAAGTTTGG
137     E   K   A   Y   D   G   S   A   D   I   I   S   N  [A   S   C   T   T   N   C   L]  A   P   L   A   K   V   V   N   D   K   F   G
3301  CATTGTCGAG GGTCTCATGA CCACCATCCA CTCCTACACT GCTACCCAGA AGACTGTCGA TGGCCCCTCT GCCAAGGATT GGCGCGGTGG CCGTGGTGCC
170    I   V   E   G   L   M   T   T   I  [H]  S   Y   T   A   T   Q   K   T   V   D   G   P   S   A  [K]  D   W   R   G   G   R   G   A
3401  GCTCAGAACA TCATCCCCTC CAGCACTGGT GCCGCCAAGG CTGTCGGCAA GGTCATTCCT GAGCTCAACG GCAAGCTTAC TGGCATGTCC ATGCGTGTCC
203   A   Q   N   I   I   P   S   S   T   G   A   A   K   A   V   G   K   V   I   P   E   L   N   G   K   L   T   G   M   S   M  [R]  V   P
3501  CTACCGCCAA CGTTTCCGTT GTCGACCTGA CTGTTCGTCT TGAGAAGGCT GCCAGCTACG ACGCCATCAA GGCCGCCATC AAGGAGGCTG CTGAGGGCCC
237     T   A   N   V   S   V   V   D   L   T   V   R   L   E   K   A   A   S   Y   D   A   I   K   A   A   I   K   E   A   A   E   G   P
3601  CCTCAAGGGT ATTCTCGCTT ACACCGAGGA CGAGCTCGTC TCCTCCGATC TCAACGGCAA CACAAACTCT TCGATTTTCG ATGCTAAGGC CGGTATCTCT
270    L   K   G   I   L   A   Y   T   E   D   E   L   V   S   S   D   L   N   G   N   T   N   S   S   I   F   D   A   K   A   G   I   S
3701  CTCAATGACA ACTTTGTCAA GCTGGTTTCG TGGTACGACA ACGAGTGGGG TTACTCCCGC CGTGTCTTGG ACCTCATCTC CTTCGTTGCC AAGGTCGACG
303   L   N   D   N   F   V   K   L   V   S   W   Y   D   N   E   W   G   Y   S   R   R   V   L   D   L   I   S   F   V   A   K   V   D   A
3801  CTTCCAAATA Ggatgaggac tcctcagcta agccaaggct ataatgtggc ccacgttggc tagctgaggc ggggatgggt aatagagt[aa aaata]tgtta
337     S   K
3901  atg[aagaaat ata]cca[aaaa at]tacgttcc cgattgtatc gtctgctacc ttttcgtcgc ccgcgtggca tcaaacataa tgcaaattcc gaaacaaagg
4001  caattgtcgg tagaaaatgg gagggacgaa caagtgttac atacaacagt aaaatcttaa caacactggc gcca[tttctt] gatatggaac aacaaagggt
4101  ggtacatgaa agacttgtcc cctcatcact ttgcatcatc tgtctcgttg atatttgtca gccacattcc tacaaaaaac gaccaaaacg aaaagacaga
4201  ttcctatcct [tttcat]cttc ccttcctatc ctttgtatcc ttgtcagtt[t ttttc]gcctt ctcctcacaa gtac[atgtat] gccaaaccg[a aataa]ccaga
4301  ttttacatg[a ataaaaaa]aa acgccaaatg tatcatagtt gaatgcagaa aaccacttct ctcatggaaa tgtgttgccc attctgggta taaaaaaaaa
4401  agcacagaga tttcgactct ttcccgtgtt cccatccaaa caaacgctca tcattgttaa atgtgccccg ccgtcaccta ctccgttcct ttt
```

Figure 3. Sequence analysis of glyceraldehydes-3-phosphate dehydrogenase gene (*CmGpd*) and its promoter. Nucleotides are represented in lowercase letters (noncoding) or by capital letters (coding), and amino acid sequences are indicated by bold capital letters. The gpd box is indicated by italic bold lowercase letters and the C+T-rich region is represented in italic shaded lowercase letters. Introns are indicated by italic underlined lowercase letters. Inverted-repeat sequences are indicated by pairs of identical italic lowercase letters above underlined lowercase letters (*a-e*). Direct-repeat sequences are indicated by pairs of identical capital letters above underlined lowercase letters (A-E). Consensus amino acid sequences are represented by bold capital letters in boxes. Sequences important for mRNA 3'-end processing are represented by lowercase letters in boxes

upstream of the 3' splice junction) share virtually homology with corresponding sequences of introns from *B. bassiana Bbgpd* (Liao et al., 2008), *Cochliobolus heterostrophus gpd1* (Van and Yoder, 1992) and other filamentous fungi (Ballance, 1986). Based on the above features, a putative intron found was 167 bp upstream of the *Cmgpd* ATG start codon, the same as the introns found in the 5'-untranslated regions of *A. nidulans gpdA* and *B. bassinana Bbgpd*. However, other consensus elements such as a psk box, a qut box and a qa box in the *A. nidulans gpdA* promoter (Punt et al., 1990) were not found in the *Cmgpd* promoter (Figure 3).

Inspection of the *Cmgpd* 3'-flanking sequence revealed no consensus polyadenylation signals (TATATA) downstream of the stop codon TAG. Graber (1999) classified signal elements in mRNA 3'-end processing into four clusters based on their positions (relative to the 3'-end cleavage site) and putative functions (corresponding to mRNA 3'-end processing in yeast) (Graber et al., 1999). They were called type 1 (efficiency), type 2 (positioning), type 3 (pre-cleavage) and type 4 (downstream), and potential cleavage sites were inside the position of the type 3 clusters. In *Cmgpd* 3'-flanking region, type 2 (AAAATA, AAGAAA, AATATA, AAAAAT), type 3 (TTTCTT), type 4 (TTTCAT, TTTTTC), type 2 (ATGTAT), and type 3 (AAAAAA, AAATAA, AATAAA) of clusters is located successively, which suggests that two potential 3'-end cleavage sites could exist corresponding to two type 3 clusters (Figure 3).

Analysis of intron in *C. militaris gpd* gene

To analyze the *C. militaris gpd* gene expression and confirm the presence of the only intron in *Cmgpd*, RT-PCR was done using the synthesized cDNA of *C. militaris* as template and the degenerate primers gpd-F/gpd-R. This primer pair borders a 657 bp region in the genomic DNA, encompassing the predicted intron. Thus, the corresponding mRNA fragment should be 378 bp long and a fragment with the expected size was detected (Figure 1). After sequencing, comparison of the sequence of this fragment with the relevant region in *Cmgpd* confirmed the predicted position and length of the intron. Moreover, these results suggested that *C. militaris gpd* gene encodes a functional protein and is constitutively expressed in *C. militaris*.

DISCUSSION

Thermal asymmetric interlaced PCR (TAIL-PCR) was first reported by Liu and Whittier (1995) and they used this technique to isolate and sequence the insert end segments from P1 and YAC clones (Liu and Whittier, 1995). TAIL-PCR has been proven to be an efficient and sensitive method to isolate DNA segments flanking known sequences and isolate genes by positional cloning. Compared to other methods, like early inverse PCR and recently reported DW-ACP^{TM} (DNA walking-annealing control primer ^{TM}) method, TAIL-PCR has a number of advantages that facilitate and expedite the procedure of retrieving sequences flanking unknown sequence. For example, no laborious DNA manipulations, such as re-striction cutting or adaptor ligation, are required prior to TAIL-PCR. Since its invention, TAIL-PCR has been wide-ly used (Liu and Huang, 1998; Terauchi and Kahl, 2000; Michiels et al., 2003; Imaizumi et al., 2005). In our study, tertiary round TAIL-PCR was added to increase the abundance of specific products and decrease the nonspecific produces, and therefore no nonspecific products were found after sequencing.

Transformation is an essential part in modern fungal research and is of great importance in genetic modification of diverse fungal species used in technology (Ruiz-Diez, 2002). In the flanking sequence of functional gene in eukaryote, there exist many important transcription and expression regulation factors. It is necessary to detect the unknown flanking sequences which are very important in the gene research of transcription regulation mechanism. Especially in the research of transgenic organisms, the flanking sequence of the insertion site of target gene plays an important role in the normal gene transcription and expression. Generally, strong promoters are required to construct the compact expression cassettes or vectors for transformation. Because little is known about the critical parts of fungal promoters till today, a strong homologous promoter is usually searched for when developing a new transformation system. The glyceraldehyde-3-phosphate dehydrogenase (GPD) promoter is always taken as an ideal candidate in fungi. Although the construction of expression vectors using heterologous *gpd* gene promoter is very common, the expression directed by heterologous *gpd* gene promoters was unsuccessful or low efficient in some species (Casselton and De La Fuente Herce, 1989; Mooibroek et al., 1990). Scientists hope that through utilizing internal promoter, gene in the vector would be highly expressed in the receptor.

Using TAIL-PCR technique and degenerate PCR, we cloned a 4493 fragment (*Cmgpd*) including *C. militaris gpd* gene. The deduced *C. militaris* GPD protein shows a 94% homology with *B. bassiana* predicted GPD protein. *C. militaris* and *B. bassiana* are both entomopathogenic fungi which share many conserved elements in the 5'-flanking region of their *gpd* genes, such as gpd boxs, C+T-rich regions, some regions of direct- or inverted-re-peat sequences and introns in the 5'-untranslated region, with *gpd* gene promoters in *Aspergillus* fungi. It means that a similar regulatory mechanism of *gpd* gene might be shared between entomopathogenic fungi (Liao et al., 2008) and *Aspergillus* fungi. In *C. militaris gpd* gene, 40 out of 61 possible sense codons are used and a pyri-midine is chosen in 95.26% of the cases when a choice between a purine and a pyrimidine is allowed in the third position of codons. Highly expressed genes generally show a more marked codon bias than genes expressed at low levels. This codon usage bias of *C. militaris* GPD protein is similar to that found for highly expressed genes in filamenttous fungi (Punt et al., 1988; Kinnaird and Fincham,

1983), but is clearly different from that in highly expressed genes of *S.cerevisiae* (Bennetzen and Hall, 1982).

In contrast to *A. nidulans gpdA* promoter, *B. bassiana* PBbgpd can drive target genes with a higher expression level and relatively shorter sequence in *B. bassiana* transgenic research (Liao et al., 2008). Thus, *C. militaris* potential gpd promoter could be a strong and constitutively expressed promoter and can be used in the construction of transformation vectors in *C. militaris*. In future, with the advent of high yield and high quality transformed *C. militaris*, *C. sinensis*, which is now confronted with resource starvation and population crisis, would be substituted by "*C. militaris*".

ACKNOWLEDGMENTS

The authors wish to thank Director Xin Wang and Guojun Wang (Heilongjiang Xinyisheng Pharmaceutical Co., Ltd., China) for their help providing the fungal strains. This research is financially supported by the Shanghai Sci-ence and Technology Committee (Project Number: 07DZ-19724), Shanghai Leading Academic Discipline Project (Project Number: B209), and Heilongjiang Xinyisheng Pharmaceutical Co., Ltd., People's Republic of China.

REFERENCES

Ballance DJ (1986). Sequence important for gene expression in filamentous fungi. Yeast 2(4): 229-236.

Bennetzen JL, Hall BD (1982). Codon selection in yeast. J. Biol. Chem. 257 (6): 3026-3031.

Bitter GA, Egan KM (1984). Expression of heterologous genes in *Saccharomyces cerevisiae* from vectors utilizing the glyceraldehyde-3-phosphate dehydrogenase gene promoter. Gene 32 (3): 263-274.

Casselton LA, De La Fuente Herce A (1989). Heterologous gene expression in the basidiomycete fungus *Coprinus cinereus*. Curr. Genet. 16: 35-40.

Döring F, Klapper M, Theis S, Daniel H (1998). Use of the glyceraldehyde-3-phosphate dehydrogenase promoter for production of functional mammalian membrane transport proteins in the yeast *Pichia pastoris*. Biochem. Biophys. Res. Commun. 250 (2): 531-535.

Graber JH, Cantor CR, Mohr SC, Smith TF (1999). Genomic detection of new yeast pre-mRNA 3'-end-processing signals. Nucleic Acids Res. 27 (3): 888-894.

Harmsen MC, Schuren FHJ, Moukha SM, Vanzuilen CM, Punt PJ, Wessels JGH (1992). Sequence analysis of the glyceraldehyde-3-phosphate dehydrogenase genes from the basidomycetes *Schizophyllum commune*, *Phanerochaete chrysosporium* and *Agaricus bisporus*. Curr. Genet. 22 (6): 447-454.

Hirano T, Sato T, Yaegashi K, Enei H (2000). Efficient transformation of the edible basidiomycete *Lentinus edodes* with a vector using a glyceraldehyde-3-phosphate dehydrogenase promoter to hygromycin B resistance. Mol. Gen. Genet. 263 (6): 1047-1052.

Holland MJ, Holland JP (1978). Isolation and identification of yeast messenger ribonucleic acids coding for enolase, glyceraldehyde-3-phosphate dehydrogenase, and phosphoglycerate kinase. Biochemistry 17 (23): 4900-4907.

Imaizumi R, Sato S, Kameya N, Nakamura I, Nakamura Y, Tabata S, Ayabe S, Aoki T (2005). Activation tagging approach in a model legume, Lotus japonicus. J. Plant Res. 118 (6): 391-399.

Kinnaird JH, Fincham JRS (1983). The complete nucleotide sequence of the *Neurospora crassa am* (NADP-specific glutamate dehydrogenase) gene. Gene 26 (2): 253-260.

Kuo CY, Chou SY, Huang CT (2004). Cloning of glyceraldehyde-3-phosphate dehydrogenase gene and use of the gpd promoter for transformation in *Flammulina velutipes*. Appl. Microbiol. Biotechnol. 65 (5): 593-599.

Liao XG, Fang WG, Zhang YJ, Fan YH, Wu XW, Zhou Q, Pei Y (2008). Characterization of a highly active promoter, PBbgpd, in *Beauveria bassiana*. Curr. Microbiol. 57 (2): 121-126.

Liu YG, Whittier RF (1995). Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. Genomics 25 (3): 674-681.

Liu YG, Mitsukawa N, Oosumi T, Whittier RF (1995). Efficient isolation and mapping of *Arabidopsis thaliana* T-DNA insert junctions by thermal asymmetric interlaced PCR. Plant J. 8 (3): 457-463.

Liu YG, Huang N (1998). Efficient amplification of insert end sequences from bacterial artificial chromosome clones by thermal asymmetric interlaced PCR. Plant Mol. Biol. Rep. 16: 175-181.

Liu ZH, Yang Q (2005). Cloning and characterization of glyceraldehyde-3-phosphate dehydrogenase gene from *Chaeto-mium globosum*. Acta Microbiol. Sin. 45 (6): 885-889.

Michiels A, Tucker M, Ende WVD, Laere AV (2003). Chromosomal walking of flanking regions from short known sequences in GC-Rich plant genomic DNA. Plant Mol. Biol. Rep. 21: 295-302.

Mooibroek H, Kuipers AG, Sietsma JH, Punt PJ, Wessels JG (1990). Introduction of hygromycin B resistance into *Schizophyllum commune*: preferential methylation of donor DNA. Mol. Gen. Genet. 222 (1): 41-48.

Piechaczyk M, Blanchard JM, Marty L, Dani C, Panabieres F, El Sabouty S, Fort P, Jeanteur P (1984). Post-transcriptional regulation of glyceraldehyde-3-phosphate-dehydrogenase gene expression in rat tissues. Nucleic Acids Res. 12 (18): 6951-6963.

Punt PJ, Dingemanse MA, Jacobs-Meijsing BJ, Pouwels PH, van den Hondel CA (1988). Isolation and characterization of the glyceraldehyde-3-phosphate dehydrogenase gene of *Aspergillus nidulans*. Gene 69 (1): 49-57.

Punt PJ, Dingemanse MA, Kuyvenhoven A, Soede RD, Pouwels PH, van den Hondel CA (1990). Functional elements in the promoter region of the *Aspergillus nidulans* gpdA gene encoding glyceraldehyde-3-phosphate dehydrogenase. Gene 93 (1): 101- 109 .

Ruiz-Díez B (2002). Strategies for the transformation of filamentous fungi. J. Appl. Microbiol. 92 (2): 189-195.

Terauchi R, Kahl G (2000). Rapid isolation of promoter sequences by TAIL- PCR: the 5'-flanking regions of Pal and Pgi genes from yams (Dioscorea). Mol. Gen. Genet. 263 (3): 554-560.

Tsugeki R, Kochieva EZ, Fedoroff NV (1996). A transposon insertion in the Arabidopsis SSR16 gene causes an embryo-defective lethal mutation. Plant J. 10 (3): 479-489.

Van Wert SL, Yoder OC (1992). Structure of the *Cochliobolus heterostrophus* glyceraldehyde-3-phosphate dehydrogenase gene. Curr. Genet. 22 (1): 29-35.

Wolff AM, Arnau J (2002). Cloning of glyceraldehyde-3-phosphate dehydrogenase-encoding genes in *Mucor circinelloides* (Syn. *racemosus*) and use of the gpd1 promoter for recombinant protein production. Fungal Genet. Biol. 35 (1): 21-29.

Zhou LH (2008). A method for preparation of genome DNA from filamentous fungi with abundant polysaccharose. Hubei Agric. Sci. 47 (4): 379-381.

Zhou XW, Gong ZH, Su Y, Lin J, Tang KX (2009). Cordyceps fungi: natural products, pharmacological functions and developmental products. J. Pharm. Pharmacol. 61 (3):279-291