African Journal of Biotechnology Vol. 9(13), pp. 1864-1870, 29 March, 2010
Available online at http://www.academicjournals.org/AJB
ISSN 1684–5315 © 2010 Academic Journals

Full Length Research Paper

In silico cloning and bioinformatic analysis of PEPCK gene in *Fusarium oxysporum*

Li He[1,2], Zhou Guoying[1,2]*, Zhang Huai-yun[1], Li Lin[2] and Liu Jun-ang[1,2]

[1]Central South University of Forestry and Technology, Biotechnology Core Facilities, Changsha, 410004, China.
[2]Central South Universities of Forestry and Technology, College of Forestry, Changsha, 410004, China.

Accepted 26 February, 2010

Phosphoenolpyruvate carboxykinase (PEPCK), a critical gluconeogenic enzyme, catalyzes the first committed step in the diversion of tricarboxylic acid cycle intermediates toward gluconeogenesis. According to the relative conservation of homologous gene, a bioinformatics strategy was applied to clone *Fusarium oxysporum* phosphoenolpyruvate carboxykinase gene (*PEPCK*) by blasting search of EST database with homologous gene cDNA of *Neurospora crassa* and identified. Some characters of the PEPCK that were analyzed and predicted by the tools of bioinformatics in the following aspects include the composition of amino acid sequences, physical and chemical properties, O-glycosylation site, hydrophobicity or hydrophilicity, secondary and tertiary structure of the protein and function. These results showed that the full-length of PEPCK was 1771 bp and it contained a complete ORF (1575 bp), encoded 524 amino acids, which is much conserved in ascomycetes. The calculated molecular weight of PEPCK was 58358.2 Da, theoretical pI of 6.84. It has 20 α-helices, 37 sheets, and 12 glycosylation sites. It was a hydrophilic and stable protein with active site, ATP -binding site, metal-binding site and substrate-binding site.

Key words: *Fusarium oxysporum*, PEPCK, *in silico* cloning, EST.

INTRODUCTION

Teaoil camellia (*Camellia oleifera*) is a special woody plant for edible oil production which is rich in vitamins in Southern China. Fusarium wilt caused by *Fusarium oxysporum* is a common disease in *C. oleifera*, and always results in significant economic losses (Zhou et al., 2007). Phosphoenolpyruvate carboxykinase (PEPCK), a critical gluconeogenic enzyme, catalyzes the first committed step in the diversion of tricarboxylic acid cycle intermediates toward gluconeogenesis. It catalyzes the reversible decarboxylation and phosphorylation of oxaloacetate to yield phosphoenolpyruvate and carbon dioxide, using a nucleotide molecule (ATP) for the phosphoryl transfer, and has a strict requirement for divalent metal ions for activity (Reymond at al., 1992; Leegood and Ap Ree, 1978).

The expressed sequence tags (EST) database of *F. oxysporum* was screened by information-probe of cDNA sequence of PEPCK of *Neurospora crassa*. We utilized the method of *in silico* cloning to successfully achieve the cDNA of PEPCK gene from *F. oxysporum*. The sequence of PEPCK of *F. oxysporum* was analyzed by bioinformatics methods, including open reading frame (ORF) analysis, BLAST, protein structure, physical and chemical properties prediction and phylogenetic analysis.

MATERIALS AND METHODS

Bioinformatics tools

http://www.ncbi.nlm.nih.gov/
http://mobyle.pasteur.fr/cgi-bin/portal.py?form=cap3
http://www.ncbi.nlm.nih.gov/gorf/gorf.html
http://blast.ncbi.nlm.nih.gov/Blast.cgi
http://frodo.wi.mit.edu/primer3
http://www.expasy.org/tools/protparam.html
http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html

*Corresponding author. E-mail: csuftlihe@163.com.

Abbreviations: PEPCK, Phosphoenolpyruvate carboxykinase; **EST,** expressed sequence tags; **dbEST,** expressed sequence tags database; **NCBI,** national center for biotechnology information.

http://www.expasy.org/cgi-bin/protscale.pl
http://www.cbs.dtu.dk/services/YinOYang/
http://swissmodel.expasy.org/
http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi
http:// www. umass. edu/microbio/ rasmol
RasMol and Cn3D software

In silico cloning of *PEPCK* of *F. oxysporum*

The sequence of *PEPCK* of *N. crassa* (Accession Number: XM_955062) in NCBI acting as the information probe, was sent into the EST database of *F. oxysporum* to do the BLAST searching (Expect = 10), and then the generated contigs (http://mobyle. pasteur.fr/cgi-bin/portal.py?form = cap3) with high identity to the probe and coming from the same organism, *F. oxysporum*, was collected and assembled into a longer novel EST sequences as second probe. The above step was not repeated until the newly generated probe can not be elongated. The sequence of newly generated probe (contig) was analyzed by bioinformatics methods, including open reading frame (ORF) analysis, BLAST, protein structure, physical and chemical properties prediction and phylogenetic analysis.

Experimental verification

RT-PCR was used to experimentally verify validity of assembled contig. Total RNA of *F. oxysporum* was extracted using the Fungal RNA kit (Omega, USA) and cDNA synthesized using the Reverse-IT 1st strand synthesis kit (TaKaRa). Primers were designed using the Primer3 program. PCR was performed with primers Pk-F (5'-CAAGACTTCCCTGCACCCTA-3') and Pk-R (5'-GTGAAGTTCT CGCGGAAGAG-3'). Amplification was carried out at 94℃, 2 min; 94℃, 1 min, 56℃, 1 min, 72℃, 1 min for 36 cycles; then 72℃, 10 min. PCR products were separated on 1.2% agarose gel. A candidate DNA band was recovered and cloned into pMD18-T vector (TaKaRa). Then the product was confirmed by sequencing from both orientations.

RESULTS

ORF prediction and amino acids sequence deduction of *F. oxysporum* PEPCK

By searching the NCBI database, a 1771 bp contig was obtained by assembling these EST sequences of *F. oxysporum* by CAP3 program. A pair primer was designed and a 1610 bp RT-PCR product were amplified and sequenced. Identity comparing between the assembled contig and the 1610 bp RT-PCR product showed that the 100% of the two sequence was the same which indicated that the *in silico* cloning results was credible. Searching the cDNA sequence for potential coding regions by ORF finder (NCBI), an entire open reading frame (ORF) of 524 amino acids was detected with a potential start codon at the153rd site and a stop codon at the 1727th site (Figure 1).

Similarity analysis and phylogenetic tree construction of PEPCK

The sequence analysis and similarity comparison with the PEPCK of *F. oxysporum*, *Gibberella sea*, *N. crassa*, *Aspergillus niger*, *Magnaporthe grisea*, *Podospora anserine*, and *Aspergillus flavus* were done by the softwares of BLASTN and BLASTP in NCBI (www. ncbi. nlm. nih. gov). The phylogenic tree was constructed basing on the deduced amino acids by the program of Distance Tree of Results. The results (Figure 2) showed that there were high identities in nucleotide acid level, no less than 80% identity in nucleotide acid level and no less than 85% identity (90% similarity) in amino acid level. The phylogenic tree suggested that *F. oxysporum* PEPCK (Small triangle pointing: Unnamed protein product) was more closely related to the PEPCK of ascomycetes (in the rectangle) than the other species PEPCK genes. Basidiomycetes, cellular slime molds take second place; the least genetically related was ciliates. Evolution relationships between *F. oxysporum* PEPCK and other species PEPCK homologs revealed in the phylogenetic tree were in agreement with the concept of traditional taxonomy.

Physical and chemical properties of PEPCK

The physical and chemical properties of PEPCK of *F. oxysporum* were analyzed utilizing online service (http:// www.expasy.org/tools/protparam.html). The results showed that molecular weight: 58358.2 Da, theoretical pI: 6.84, formula: $C_{2603}H_{4032}N_{704}O_{777}S_{23}$, total number of atoms: 8139, extinction coefficient: 77405(280nm). The estimated half-life was: 30 h (mammalian reticulocytes, *in vitro*). The instability index (II) was computed to be 39.06; this classifies the protein as stable. Aliphatic index: 74.10. Grand average of hydropathicity (GRAVY): - 0.359. The PEPCK was composed of 20 kinds of Amino Acid. And Ala, Gly, Leu, Thr are the most abundant components but low content of Gln and Trp.

Secondary structure of PEPCK

The secondary structure of PEPCK of *F. oxysporum* was predicted utilizing online service (http://npsa-pbil.ibcp.fr/ cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma. html). The results showed that PEPCK of *F. oxysporum* consisted of 28.63% alpha helix, 18.89% extended strand, 44.08% random coil and 8.40% beta turn.

Hydrophilicity prediction of PEPCK

Hydrophilicity of *F. oxysporum* PEPCK was predicted utilizing Program of ProtScale (Kyte and Doolittle). The results showed that the site of 136 (Glu) is the most hydrophilic (Score: -2.444) and 225 (Phe) the most hydrophobic (Score: 1.678) (Figure 3). We come to the conclusion that the PEPCK of *F. oxysporum* is hydrophilic protein.

GCACGAGGCACAAATCAATGCTTCAGGAACCGGTTTATCAAGATGCTTCCCAACAACGTTAACAAGACTTCCCTGC

ACCCTACCGGTGTTACTCCTCACCAGGAGCACACCGAGCTGGAACAGGAGCTCCACGACAAGGGCTCACATCGATT

at gacccggt gt cgccat t t at t t ccccaaaccct t cat t cgct gccct caacaaggaat t gccct ggt t ct acaagaccgggaaccgct

M T R C R H L F P Q T L H S L P S T R N C P G S T R P G T A

t at caacct t ccagcggt gccct gact gcct act ct ggt gccaagact ggt cggt cccct ct t gacaagcgaat t gt cgaggaggct t ct

Y Q P S S G A L T A Y S G A K T G R S P L D K R I V E E A S

t ccaaggacaacat ct ggt ggggacct gt caacaagcccat gact cct gaggt ct ggaagat caaccgagaacgt gct gt cgat t acct c

S K D N I W W G P V N K P M T P E V W K I N R E R A V D Y L

aacacacgaagccgt at ct at gt cat cgat ggt t t cgccggct gggacgagaagt accgcat ccgagt ccgagt t at ct gcgcccgt gcc

N T R S R I Y V I D G F A G W D E K Y R I R V R V I C A R A

t accat gct ct ct t cat gcgaaacat gct t at ccgacct t cacgagaggagct caaggact t ccaccccgact acaccat ct acaat gcc

Y H A L F M R N M L I R P S R E E L K D F H P D Y T I Y N A

ggcaagt t ccccgccaacagat acact gagggt at gacct ct ggt acct ccgt t gccat caact t cgagcagaaggagat ggt cat t ct c

G K F P A N R Y T E G M T S G T S V A I N F E Q K E M V I L

ggt accgagt acgccggt gagat gaagaagggt gt ct t cact gt cct ct t ct acgagat gcct at caagcacaacgt cct cact ct ccac

G T E Y A G E M K K G V F T V L F Y E M P I K H N V L T L H

t cct ct gccaacgagggcaagaacggcgat gt t acact ct t ct t cggt ct ct ct ggaact ggcaagaccact ct ct ccgccgaccccaac

S S A N E G K N G D V T L F F G L S G T G K T T L S A D P N

cgagct ct cat t ggt gacgacgagcact gct ggt ccgacaacggt gt ct t caacat cgagggaggt t gct acgct aagaccat t ggcct g

R A L I G D D E H C W S D N G V F N I E G G C Y A K T I G L

t cggccgagaaggagcccgat at ct t t ggcgct at ccgat acggt t ct gt cct cgagaacgt cgt ct t cgaccct ct cacccgt gaggt c

S A E K E P D I F G A I R Y G S V L E N V V F D P L T R E V

gact acgacgat gccaccct cact gagaacacccgat gt gcct accccat cgagt acat t t ccaacgccaagat t cct t gcct gt ct ccc

D Y D D A T L T E N T R C A Y P I E Y I S N A K I P C L S P

aact cccccct ccaacat cat cct cct cacat gcgacgcccgcggt gt t ct gcccccct at ct ccaagct cgaccgcaaccagaccat gt t c

N S P S N I I L L T C D A R G V L P P I S K L D R N Q T M F

cact t cat ct ct ggt t acacct ccaagat ggccggt act gaggacggcgt caccgagccccaggct acct t ct ccagct gct t cgcccag

H F I S G Y T S K M A G T E D G V T E P Q A T F S S C F A Q

ccat t cct t gct ct gcaccccat gaagt acgccaagat gct t gccgacaagat cgagact cacaaggccaacgcct ggct cct caacacc

P F L A L H P M K Y A K M L A D K I E T H K A N A W L L N T

ggt t gggt t ggt gccggt t t cgct cagggt ggcaagcgat gccct ct caagt acact cgt gccat t ct cgat gccat ccact ct ggcgaa

G W V G A G F A Q G G K R C P L K Y T R A I L D A I H S G E

ct t gccaat gt cgagt at gagaact at gaggt t t t caacct ccaggt t cccaagacct gccccaacgt t ccct cagagct t ct caacccc

L A N V E Y E N Y E V F N L Q V P K T C P N V P S E L L N P

t ccaaggct t ggaccgct ggcgaggacagct t caagaccgaggt t gt caagct gggcaagct ct t ccgcgagaact t caccaagt acgag

S K A W T A G E D S F K T E V V K L G K L F R E N F T K Y E

agcgaggct accgaggat gt cgt caaggct ggt cccgt t gt ct aaGAAGAGGAAGAATGAGAGTAAGAAGGGAACAAAAGGGAAAGCGA

S E A T E D V V K A G P V V *

Figure 1. Nucleotide sequence and deduced amino acid sequence of *PEPCK* of *Fusarium oxysporum*. The asterisk indicates the stop codon.

Predicted O-glycosylation site of *PEPCK*

The O-glycosylation site of PEPCK was predicted utilizing online service (http://www.cbs.dtu.dk/services/YinOYang/). The results showed that there were 12 glycosylation sites of *PEPCK* of *F. oxysporum* at 2(T), 18(T), 24(S), 25(T), 34(S), 35(S), 42(S), 163(T), 267(T), 378(T), 385(S) and 469(T). The site of 35(S) with the highest possibility, 42(S) and 469(T) took second place, the others relatively less possibility (Figure 4).

Advanced Structure of PEPCK

The structure prediction from the primary structure to advanced structure is an important task in the protein research field. In order to study relationship between the advanced structure and functions, the advanced structures of the putative amino acid sequences from PEPCK gene known whole coding region was predicted by software of predicting protein in internet (http://swissmodel.expasy.org/). This project involved the use of

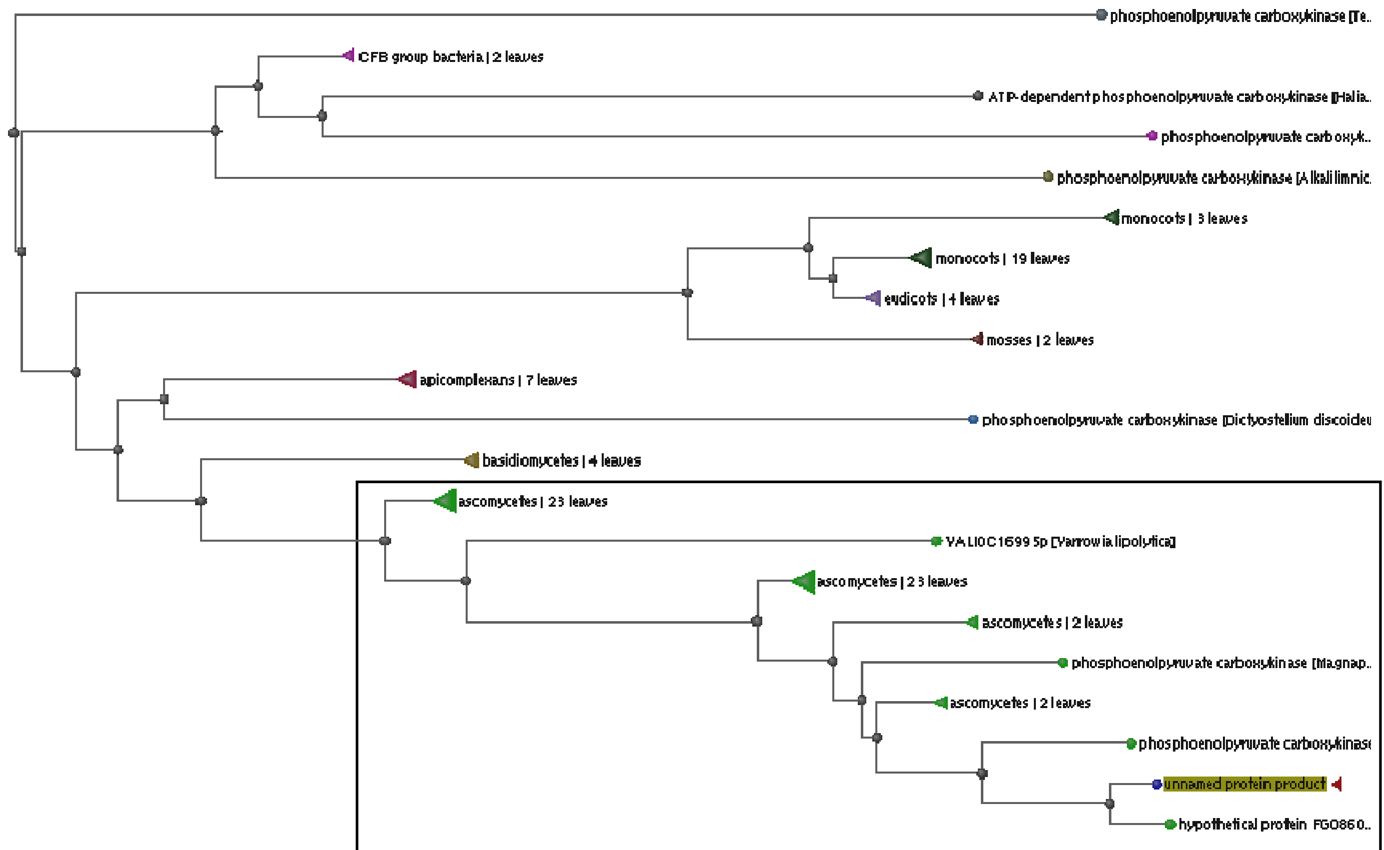

Figure 2. The position of *F. oxysporum* PEPCK in phylogenetic tree.

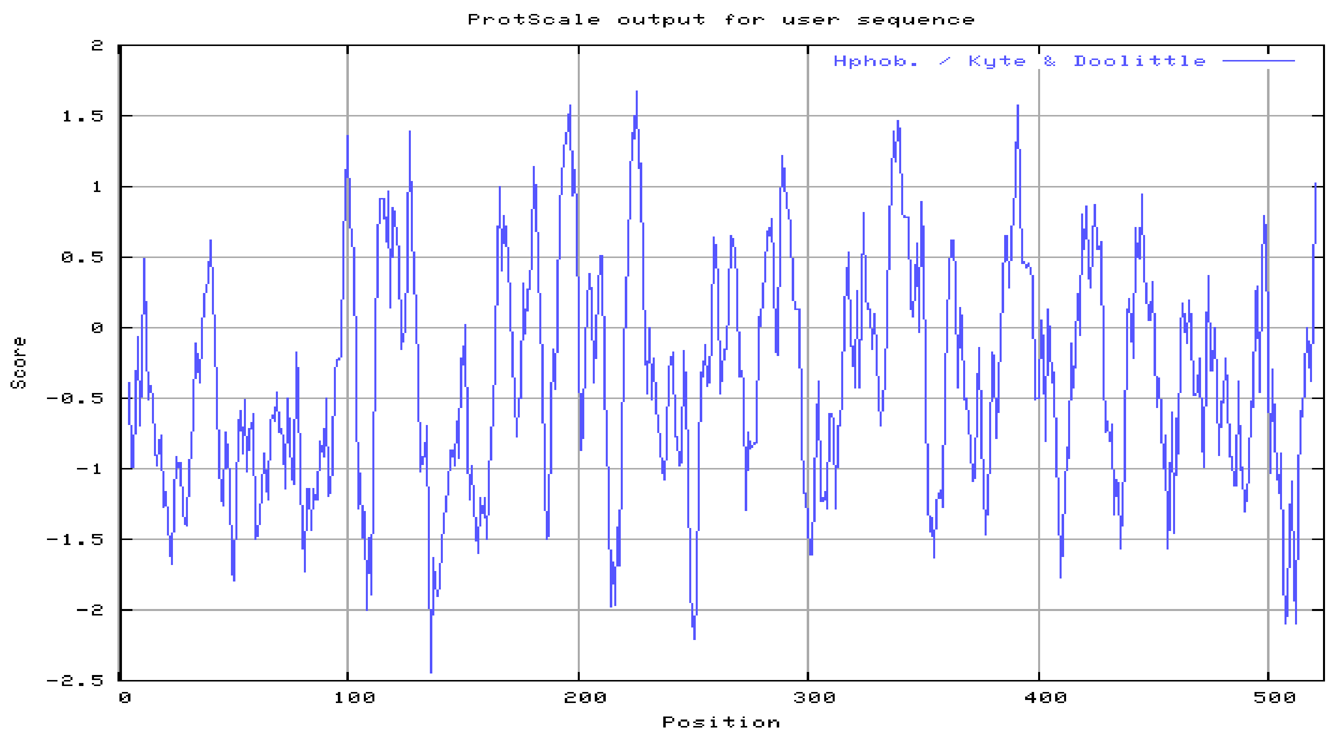

Figure 3. Hydrophilicity profile of *F. oxysporum* PEPCK.

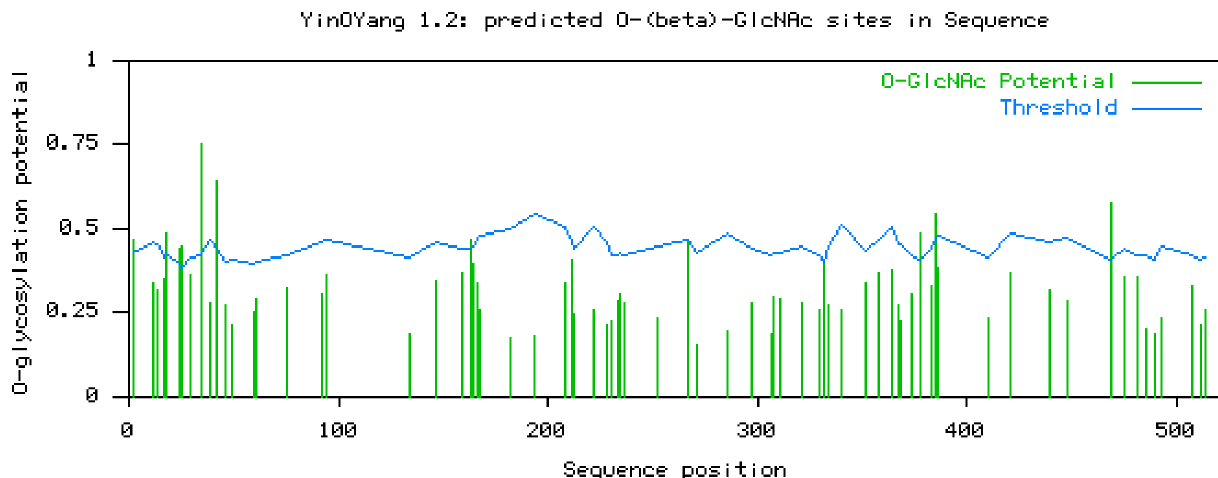

Figure 4. Predicted O-glycosylation site of PEPCK.

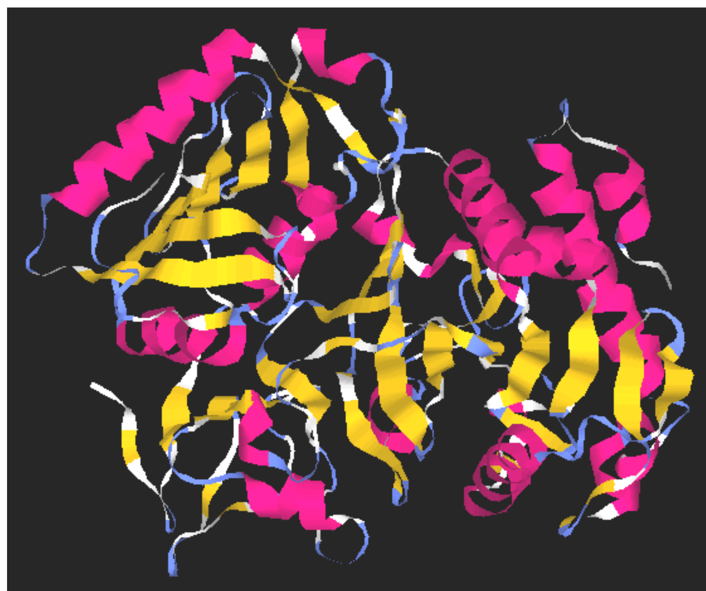

Figure 5. Three-dimensional structure of PEPCK.

the molecular visualization program RasMol (Sayle and Milner-White, 1995) and computer simulated amino acids of the PEPCK. The three-dimensional structure of PEPCK showed that there were 20 α-helices, 37 sheets and some irregular coiled peptides (Figure 5).

Prediction and analysis functional domain of PEPCK

The functional domain of *F. oxysporum* PEPCK was predicted and analyzed base on http:// www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi. Three-dimensional structure of PEPCK was constructed and viewed with Cn3D 4.1 software. The results (Figure 6) showed that there were some active sites (Figure 7A), ATP binding site (Figure 7B), metal-binding site (Figure 7C), and a substrate-binding site (Figure 7D). PEPCK is separated into two phylogenetic groups based on their nucleotide substrate specificity; the PEPCK of *F. oxysporum* belongs to the ATP-dependent groups.

DISCUSSION

In silico cloning, a kind novel method developed in recent years for functional gene identification by using genome and EST database, compared to the traditional methods, such as the molecular hybridization, the construction of

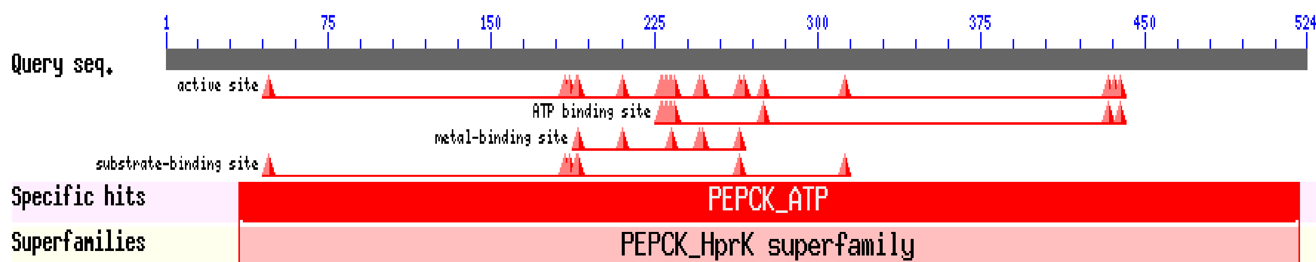

Figure 6. Functionaldomain of *F. Oxysporum* PEPCK.

Figure 7. 3D-structure of functional domain of *F. oxysporum* PEPCK. A: active site; B: ATP binding site; C: metal-binding site; D: substrate-binding site.

cosmid genomic library, and the screening of EST library, has lots of advantages compared to laboratory cloning. The advantages of *in silico* cloning include low cost, high efficiency, easy operation, etc (Zhang at al., 2005; Feng at al., 2004). With the increment of EST data and accomplishment of fungi genome sequencing, it would become possible and feasible to isolate and identify the functional genes from fungi by *in silico* cloning.

The successful examples in which the technique of *in silico* cloning was applied strongly support the fact that it is absolutely a robust and feasible tool for gene cloning and presents some advantages, compared to the

traditional methods (Chen at al., 2001; Huminiecki and Bicknell, 2000; Keeling and Palmer, 2001; Lescure at al., 1999). In this study, the full gene cDNA of *F. oxysporum* PEPCK was obtained primarily by searching and splicing the EST sequences by *in silico* cloning and the structural and functional were analyzed and predicted using bio-informatics methods successfully. The results revealed that it is a convenient technique for cloning novel gene by searching EST database with homologous gene of model living things. To our knowledge, it was the first report successfully applying the technique of *in silico* cloning in PEPCK of *F. oxysporum*. This research achievement will provide theory and reference for relative research. Further studies should be carried out on *F. oxysporum* PEPCK.

ACKNOWLEDGEMENTS

The research of this dissertation is Supported by Youth Scientific Research Foundation of Central South University of Forestry and Technology of China (No.2009001A); the Special Funds for Public Welfare Industry of Forestry of China（No.200904006）; The National Eleventh Five-Year Program on Science and Technology in China (No. 2009BADB1B0204); we are grateful to the free biological soft-wares (tools) and the dbEST in NCBI.

REFERENCES

Chen Y, Zhao YH, Wu R (2001). *In silico* cloning of mouse Muc5b gene and upregulation of its expression in mouse asthma model. Am. J. Respir. Crit. Care Med. 164: 1059-1066.

Feng YJ, Zhang HM, Jiang MG, Lan XW (2004). *In silico* cloning of full length cDNA of *Cryphonectria parasitica* ubiquitin conjugated enzyme gene(*CpUBC*). Chinese J. Bioinformatics, 2: 5-9.

Huminiecki L, Bicknell R (2000). *In silico* cloning of novel endothelial specific genes. Genome Res. 10: 1796-1806.

Keeling PJ, Palmer JD (2001). Lateral transfer at the gene and subgenic levels in the evolution of eukaryotic enolase. Proc. Natl. Acad. Sci . USA, 98: 10745-10750.

Leegood RC, Ap Ree T (1978). Phosphoenolpyruvate carboxykinase and gluconeogenesis in cotyledons of *Cucurbita pepo.* Biochim. Biophys. Acta. 524: 207-218.

Lescure A, Daniel G, Philippe C, Alain K (1999). Novel selenoproteins identified *in silico* and in vivo by using a conserved RNA structural motif. J. Biol. Chem. 274: 38147-38154.

Reymond P, Geourjon C, Roux B, Durand R, Fevre M (1992). Sequence of the phosphoenolpyruvate carboxykinase encodeing cDNA from the rumen anaerobic fungus *Neocallimastix frontalis*: comparison of the amino acid sequence with animals and yeast. Gene, 110: 57-63.

Sayle RA, Milner-White EJ (1995). RASMOL: biomolecular graphics for all. Trends Biochem. Sci. 20: 374.

Zhou GY, Song GT, Li H (2007). Present Situation and Countermeasures to Control *Camellia oliefera* Pest and disease. Journal of Central South University of Forestry Technology 6: 179-182.

Zhang HM, Jiang MG, Feng YJ (2005). *In silico* cloning of *MgEno-1* cDNA from *Magnaporthe grisea*. China J. Bioinformatics, 4: 57-61.