Cloning and characterization of *F3PYC* gene encoding pyruvate carboxylase in *Aspergillus flavus* strain (F3)

Sadia Qayyum^{1,3} · Ibrar Khan³ · Zulfiqar Ahmad Bhatti⁴ · Changsheng Peng^{1,2}

Received: 31 January 2017 / Accepted: 6 April 2017 / Published online: 14 July 2017
© Springer-Verlag GmbH Germany 2017

Abstract Pyruvate carboxylase is a major enzyme for biosynthesis of organic acids like; citric acid, fumeric acid, and L-malic acid. These organic acids play very important role for biological remediation of heavy metals. In this study, gene walking method was used to clone and characterize pyruvate carboxylase gene (*F3PYC*) from heavy metal resistant indigenous fungal isolate *Aspergillus flavus* (F3). 3579 bp of an open reading frame which encodes 1193 amino acid protein (isoelectric point: 6.10) with a calculated molecular weight of 131.2008 kDa was characterized. Deduced protein showed 90–95% similarity to those deduced from *PYC* gene from different fungal strains including; *Aspergillus parasiticus*, *Neosartorya fischeri*, *Aspergillus fumigatus*, *Aspergillus clavatus*, and *Aspergillus niger*. Protein generated from the *PYC* gene was a homotetramer (α_4) and having four potential N-linked glycosylation sites and had no signal peptide. Amongst most possible N-glycosylation sites were –N–S–S–I– at 36

amino acid, –N–G–T–V– at 237 amino acid, N–G–S–S– at 517 amino acid, and N–T–S–R– at 1111 amino acid, with several functions have been proposed for the carbohydrate moiety such as thermal stability, pH, and temperature optima for activity and stabilization of the three-dimensional structure. Hence, cloning of *F3PYC* gene from *A. flavus* has important biotechnological applications.

Keywords *Aspergillus flavus* (F3) · Pyruvate carboxylase · Gene cloning · Genome walking

Introduction

Pyruvate carboxylase is ecologically, medically, and industrially important enzyme and is widely distributed in a great variety of organisms such as fungi, bacteria, and plants as well as higher organisms (Dunn et al. 1997; Moller 2001; Pfefferle et al. 2003). Many filamentous fungi uses this enzyme for growth and, in some cases, for acid production (e.g., citric, malic, and fumaric acids) (Kenealy et al. 1986). *PYC* is therefore considered as crucial enzyme for intermediary metabolism, controlling fuel partitioning toward gluconeogenesis, lipogenesis, and insulin secretion (Mukhopadhyay et al. 2001; Branson et al. 2002).

Oxaloacetate is a key intermediate in tricarboxylic cycle (TCA) and is formed in reaction between pyruvate and CO₂ through an ATP-dependent reaction by catalysis of *PYC* enzyme (EC 6.4.1.1). TCA plays amphibolic role in metabolism i.e., as an oxidative device for energy production and also as building blocks for formation of important molecules like porphyrins and many amino acids (Chi et al. 2014). While growing in minimal medium in case of yeasts and fungi, two methods for TCA to work efficiently are the reaction under catalysis of *PYC* and those

Electronic supplementary material The online version of this article (doi:10.1007/s13205-017-0806-6) contains supplementary material, which is available to authorized users.

✉ Changsheng Peng
pcs005@ouc.edu.cn

¹ The Key Lab of Marine Environmental Science and Ecology, Ministry of Education, Ocean University of China, Qingdao 266100, China

² State Key Laboratory of Environmental Criteria and Risk Assessment, Chinese Research Academy of Environmental Sciences, Beijing 100012, China

³ Department of Microbiology, Hazara University, Manshera 21300, Pakistan

⁴ Environmental Sciences Department, COMSATS Institute of Information Technology, Abbottabad 22060, Pakistan

under isocitrate lyase and malate synthase, which form the glyoxylate bypass.

PYC enzyme shares a general mechanism for acid accumulation (e.g., citric, fumaric, malic, oxalic, and itaconic acids) by different filamentous fungi (Kubicek 1988). In this mechanism, all filamentous fungi share two main characteristics: a central role of a cytoplasmic *PYC* and a minor involvement of mitochondrial reactions of the TCA cycle in acid synthesis and accumulation (Kubicek 1988). *PYC* perform very important function in the accumulation of high concentrations of L-malic acid and fumaric acid by *Aspergillus flavus* and *Rhizopus oryzae* (formerly *R. nigricans* and, later, *R. arrhizus*), respectively (Mukhopadhyay et al. 2001). Acid accumulation is dependent on the operation of part of the reductive branch of the TCA cycle (i.e., pyruvate to oxaloacetate to malate to fumarate) (Peleg et al. 1989). Contrary to most eucaryotic organisms, which contain a mitochondrial pyruvate carboxylase (Osmani and Scrutton 1985; Purohit and Ratledge 1988) this enzyme activity is localized exclusively in the cytosolic fraction in several filamentous fungi, such as *A. nidulans* (Osmani and Scrutton 1983), *R. oryzae* (Osmani and Scrutton 1985), and *A. terreus*. However, currently it was observed that in two strains of *Aspergillus niger* which apparently accumulate citric acid, the *PYC* activity was located solely in the mitochondria (Purohit and Ratledge 1988), which is a characteristic of higher eucaryotic cells.

In yeast, a genetic study showed that expression of *Escherichia coli* phosphoenolpyruvate carboxylase can harmonize the phenotypic effects of *PYC* deficiency (Flores et al. 2000). The anaplerotic role of *PYC* is considered important for normal growth in *Rhizobium etli* (Dunn et al. 1996). Two principal pathways for the replenishment of oxaloacetate levels are through carboxylation of pyruvate by PC and from glyoxylate cycle in *Saccharomyces cerevisiae* (Barnett and Kornberg 1960). However, so far, no study on the cloning and characterization of *PYC* gene from *A. flavus* has been reported yet.

Purpose of this study is cloning and characterization of pyruvate carboxylase gene (*F3PYC*) from indigenous fungal strain *A. flavus* (F3) through genome walking.

Materials and methods

Growth media

Fungal isolation, colony, and cell morphology for fungal identification were done by using Malt extract agar (MEA), Potato dextrose agar (PDA), Czapek yeast extract agar (CYA), and Sabouraud dextrose agar (SDA) media, supplemented with 0.05% chloramphenicol (Nonaka et al. 2011; Khan et al. 2014).

Isolation of the pyruvate carboxylase (*PYC*) gene

Genomic DNA extraction

The total genomic DNA of the fungal isolate F3 was isolated and purified by using the methods as described by Al-Samarrai and Schmid (2000).

F3PYC primer designing

The sequence of the *PYC* genes of *Aspergillus* spp. was selected to design primer of their conserved sequences in order to identify the *F3PYC* gene in *A. flavus* (F3) strain (Fig. 1). The genomic DNA was used as the template to amplify 5' and 3' regions of the *F3PYC* gene in (F3) strain with the primes; *F3PYC*-s: 5'-ATYATYAAGGCYGCYTTTGGYGGTGGT-3' and *F3PYC*-a: 5'-ATCTTCATGGCVCTCAGSACVGC-3'.

Polymerase chain reaction (PCR)

The polymerase chain reaction (PCR) for amplification and sequencing of *PYC* from the fungal strain *A. flavus* (F3) was performed by using the *F3PYC*-s and *F3PYC*-a primers. The reaction system (50 µl) was composed of dNTP 0.8 mM, 10 × buffer 2.5 µl, *F3PYC*-s and *F3PYC*-a, H₂O 37 µl, 0.5 mM, MgCl₂ 1.5 mM, Taq DNA polymerase 1.25 U, and template DNA 1.0 µl. The conditions for the PCR amplification were: at 94 °C for 10 min initial denaturation then at 94 °C for 1 min denaturation occurs, annealing temperature at 47 °C for 30 min then extension at 72 °C for 3 min and final extension at 72 °C for 10 min occurs. PCR was run for 32 cycles and PCR cyclor was Gene Amp PCR System 2400 made by Perkin-Elmer.

Gel electrophoresis

Agarose gel DNA electrophoresis was performed by the modified method of Saghai-Marroof et al. (1984). For getting 2% solution, 2 g of agarose (BioRad agarose, Qiagen) was prepared in 98 ml 1× TAE (Tris/Acetate/EDTA) buffer and heated to boiling point in a water bath. Before adding 3 ml ethidium bromide (Et Br) (10 mg l⁻¹ in water to a final concentration of 0.5 µg ml⁻¹) the solution was allowed to cool to 60 °C and thoroughly mixed. In glass plates, the gel was poured. Polymerase chain reaction products were slowly loaded (5 µl) into the wells. 5 V cm⁻¹ voltage was applied to the gel for the electrophoresis run and gDNA and to view PCR product using the Vacutec Gel documentation system and product size confirmed by comparison to D2000 Marker.

Fig. 1 Conserved domains of pyruvate carboxylase gene (*PYC*) from different fungi by Using DNAMAN Program (Lynnon 6BioSoft, Vandreuil, Canada). *PYC* genes sequence was selected to design primer of conserved sequences to identify F3PYC gene in *A. flavus*

Recombination and transformation

Ligation of pGEM-T easy vector into the recovered PCR products was done and transformation into competent cells of *E. coli* JM109 was performed. The transformant were selected on plates with ampicillin. In order to confirm that the PCR products had been ligated into the vector, the colony PCR was used for amplification of cloned *PYC* gene in the *E. coli* strain, respectively. The reaction system and the conditions for PCR amplification were same as described above. The positive transformant *E. coli* containing the recombinant plasmid fragment inserted on the vector were sequenced by Shanghai Sangon Company (http://www.life-biotech.com/company/about_us.html).

Sequencing analyses

Blast on NCBI by using Protein BLASTX analysis of the above obtained sequences was done (<http://blast.ncbi.nlm.nih.gov/Blastx.cgi>). Sequencing was further analyzed for oligonucleotide and deduced amino acid sequence by the DNAMAN Program (Lynnon 6BioSoft, Vandreuil, Canada). The sequence was submitted to GenBank with the accession number KT345940.

Cloning of the full-length *PYC* gene using genome walking

The genomic walking was performed by using a TAKARA Genome Walking Kit (D316) provided with four unique degenerate kit primers with lower annealing temperature i.e., AP1–4. The total six specific primers; Ffsp3/Ffsp2/Ffsp1 and Frsp3/Frsp2/Frsp1 (Table 1), were designed according to the cloned partial sequence *PYC* gene (KT345940) encoding the *PYC* in *A. flavus* (F3). Each set of the specific primers was pointing outwardly from a known sequence end and their designing was based on the Kit manufacturer's basic principles, such as: (1) the distance between each of the two

Table 1 Specific primers used in this work and their sequences

Primer name	Sequences
Frsp3	5'-CAATCATGTAGGCTCATCGGCCTT-3'
Frsp2	5'-AAGAAACCGTAGCCCGGTGGATCA-3'
Frsp1	5'-GACACCTTGTCACCCAAGCTGTCAA-3'
Ffsp1	5'-TGGCTAAGATCAAGAGCCAGATCCG-3'
Ffsp2	5'-TCGTCCAGAAGTTCGGTGATCTCTC-3'
Ffsp3	5'-AAGTTGTTGGCCATCGGCCCTCTTT-3'

primers should be in between 60 and 100 bp, and (2) the length of each primer should be 22–26 nt, GC content should be 45–55% and Tm value should be 60–70 °C. The one of the kit based degenerate primers (AP) was used in combination with the specific primers to obtain the flanking unknown sequences by three nested PCR reactions as described by Khan et al. (2016).

Gel analysis and DNA sequencing

Through agarose gel electrophoresis the PCR products were checked, separated and then sequenced as mentioned above. NCBI ORF finder program was used to analyze the PCR products after sequencing, and aligned with the known sequences of genes encoding the *PYC* from different fungi by NCBI BLASTn, 3579 bp of the *PYC* gene in *A. flavus* (F3) was obtained and the *PYC* gene was named F3PYC (accession number KT345940).

Results and discussion

Isolation of the gene encoding pyruvate carboxylase from *Aspergillus flavus*

The total genomic DNA of the fungal isolate *A. flavus* (F3) was isolated and *PYC* gene was amplified with the help of

Fig. 2 Gel resolving pattern of PCR amplified *PYC* sequences on 2% Agarose gel. **a** Lane *M* 1 kb DNA ladder, Lane *F3* genomic DNA. **b** Lane *M* 1 kb DNA ladder and lane *PYC*, shows PCR products of fungal with about 3000 bp fragment size

degenerate primers based on *PYC* gene sequences submitted in Data Bank (NCBI/GeneBank) according to the method mentioned in “Materials and methods”. A single PCR product of the expected fragment size (about 3000 bp) was achieved (Fig. 2).

Recombination and transformation

Recombination and transformation of PCR amplified expected *PYC* gene was done by using pGEM-T easy vector (https://worldwide.promega.com/products/pcr/pcr-cloning/pgem_t-vector-systems/?activeTab=1) and *E. coli* JM109 competent cells (Code #D9052, TaKaRa) (Fig. 3a).

Furthermore, the recombinant *E. coli* JM109 competent cells were additionally confirmed (further/2nd method of recombinant confirmation) and selected through colony PCR with help of vector-based M-13 primers (Fig. 3b). Furthermore, positive recombinants were sequenced by Shanghai Sangon Company.

Analysis of cloned DNA sequence

The NCBI BLASTn analysis was done in contradiction of entire available NCBI database for the identification of DNA fragment which was being inserted in pGEM-T easy Vector. The putative conserved domain was found as blasted NCBI result showed with the top hits from the search result (Fig. 4, Supplementary Material). Specifically, our submitted query sequence (3000 bp) was 90–99% identical to a *PYC* gene from different fungal strains including: *A. parasiticus* SU-1 |KJK66878.1|, *Neosartoryafischeri* NRRL181 |XP_001267251.1|, *A. fumigatus* Af293 |XP_752054.1|, *A. clavatus* NRRL1|XP_001271664.1|, *A. niger*, CBS 513.88 |XP_001401577.1|, *Aspergillus kawachii* IFO4308 |GAA82689.1|, *Aspergillus ochraceoroseus* |KKK14644.1|,

Fig. 3 **a** The map and sequence reference points of pGEM-T easy Vector used for *PYC* cloning (https://worldwide.promega.com/products/pcr/pcr-cloning/pgem_t-vector-systems/?activeTab=2). **b** The

Gel resolving pattern of the colony PCR. *M* D2000 DNA ladders, lane 1, 2, and 3 were the PCR products with about 750 bp. *M* 1 Kb DNA ladders, lane *a*, *b*, *c*, *d* are PCR product with about bp size

Fig. 4 An excerpt of BLAST result, showing top hits with significant homology to a *PYC* gene from different fungal strains including; *Aspergillus parasiticus* SU-1 [KJK66878.1] (Supplementary material)

and *Aspergillus ruber* CBS135680 [EYE94902], signifying the inserted or cloned DNA fragment in this specific sequence examination after sequence blast was the result of a specific amplification of *PYC* gene from genomic DNA of fungal isolate F3.

Cloning of full-length *PYC* gene encoding partial pyruvate carboxylase in *A. flavus* (F3)

According to the sequence of partial pyruvate carboxylase gene (accession number KT345940) the primers were designed and used for amplification of the full-length *PYC* gene from genomic DNA of *A. flavus* (F3) by genomic walking method as described in “Materials and methods”. Finally, the sequence of the genomic walking PCR products were aligned with the known sequences of the genes

encoding pyruvate carboxylase from different fungi by NCBI BLASTn and BLASTx. In *A. flavus* (F3), the full-length gene which was encoding *PYC* was obtained and it was named *F3PYC* gene (accession number KT345940).

Analysis of the protein sequence deduced from *F3PYC* gene

BLASTn program was used for the alignment and comparison of the protein sequence deduced from the *F3PYC* gene with sequences in the protein databases, which gave the following results. The deduced protein from pyruvate carboxylase gene showed 94 and 89% similarity with *PYC* of *Aspergillus parasiticus* and *Aspergillus terreus*, respectively. The protein deduced from the *PYC* gene had no signal peptide and had the biotin carboxylase (BC) domain,

ATG GCG GCT CCG TTT CGT CAG CCT GAG GAG GCG GTC GAT GAC ACC GAG TTC ATC GAT GAC - 60
 M A A P F R Q P E E A V D D T E F I D D - 20
 CAC CAT GAA CAC CTC CGT GAT ACC GTG CAC CAT CGG TTG CGC GCC AAT TCC TCC ATT ATG -120
 H H E H L R D T V H H R L R A **N S S I** M - 40
 N-glycosylation site
 CAC TTC CAG AAG ATC CTC GTC GCC AAC CGT GGT GAG ATC CCC ATT CGT ATC TTC AGA ACG -180
 H F Q K I L V A N R G E I P I R I F R T - 60
 GCC CAC GAG CTG TCC TTG CAG ACG GTT GCT ATC TAC TCT CAT GAG GAT CGA CTG TCA ATG -240
 A H E L S L Q T V A I Y S H E D R L S M -80
 CAC CGT CAA AAG GCC GAT GAG GCC TAC ATG ATT GGC CAC CGC GGT CAG TAC ACC CCT GTC -300
 H R Q K A D E A Y M I G H R G Q Y T P V -100
 GGT GCG TAC CTG GCG GGC GAT GAG ATC ATC AAG ATC GCC CTG GAG CAC GGT GTC CAG CTG-360
 G A Y L A G D E I I K I A L E H G V Q L -120
 ATC CAC CCG GGC TAC GGT TTC TTG TCC GAG AAC GCC GAC TTC GCC CGC AAG GTT GAG AAC -420
 I H P G Y G F L S E N A D F A R K V E N -140
 GCC GGC ATT GTC TTT GTG GGA CCC ACT CCC GAT ACC ATT GAC AGC TTG GGT GAC AAG GTG -480
 A G I V F V G P T P D T I D S L G D K V -160
 TCG GCC CGT CGG CTG GCC ATT AAG TGC GAG GTC CCT GTC GTT CCG GGT ACG GAG GGC CCC-540
 S A R R L A I K C E V P V P G T E G P -180
 GTC GAG CGC TAT GAG GAG GTC AAG GCG TTC ACA GAC ACC TAT GGC TTC CCC ATC ATC ATC -600
 V E R Y E E V K A F T D T Y G F P I I I -200
 AAG GCT GCC TTT GGC GGT GGT GGC CGT GGT ATG CGT GTG GTC CGT GAC CAG GCC GAG CTG-660
 K A A F G G G R G M R V Y R D Q A E L -220
 CGT GAC TCG TTC GAG CGA GCC ACC TCT GAG GCC CGC TCC GCC TTC GGC AAT GGT ACC GTC -720
 R D S F E R A T S E A R S A F **S N G T V** -240
 N-glycosylation site
 TTC GTC GAG CGC TTC CTC GAC AAA CCC AAG CAC ATT GAA GTC CAG CTT CTG GGT GAC AGC -780
 F V E R F L D K P K H I E V Q L L G D S -260
 CAC GGC AAC GTT GTC CAT CTG TTT GAG CGT GAC TGC TCC GTG CAG CGT CGT CAC CAG AAG -840
 H G N V V H L F E R D C S V Q R R H Q K -280
 GTC GTT GAG GTT GCT CCG GCT AAG GAC CTG CCA GCC GAT GTC CGG GAC CGC ATC CTG GCC-900
 V V E V A P A K D L P A D V R D R I L A -300
 GAT GCT GTG AAG CTG GCC AAG TCC GTC AAC TAC CGT AAC GCC GGT ACA GCT GAG TTC CTG -960
 D A V K L A K S V N Y R N A G T A E F L -320
 GTG GAC CAG CAG AAC CGC CAC TAC TTC ATT GAA ATC AAT CCT CGT ATC CAA GTC GAG CAC -1020
 V D Q Q N R H Y F I E I N P R I Q V E H -340
 ACC ATC ACC GAA GAG ATT ACT GGT ATC GAT ATC GTG GCT GCA CAG ATC CAG ATT GCT GCT -1080
 T I T E E I T G I D I V A A Q I Q I A A -360
 GGT GCA AGC CTC GAG CAA CTG GGC CTG ACT CAG GAC CGC ATC TCC GCC CGC GGA TTT GCC-1140
 G A S L E Q L G L T Q D R I S A R G F A -380
 ATT CAA TGT CGT ATC ACC ACG GAA GAT CCC GCC AAG GGG TTC TCT CCG GAT ACT GGT AAG-1200
 I Q C R I T T E D P A K G F S P D T G K -400
 ATT GAG GTT TAT CGT TCC GCT GGT GGT AAC GGT GTC CGT CTG GAT GGT GGT AAC GGT TTC-1260
 I E V Y R S A G G N G V R L D G G N G F -420
 GCT GGT GCT ATC ATC ACC CCT CAC TAC GAC TCC ATG CTG GTC AAG TGT ACC TGC CGT GGT -1320
 A G A I I T P H Y D S M L V K C T C R G -440
 TCG ACC TAT GAA ATC GCT CGT CGC AAG GTT GTG CGT GCC TTG GTC GAG TTC CGT ATT CGT -1380
 S T Y E I A R R K V V R A L V E F R I R -460
 GGT GTG AAG ACC AAC ATT CCC TTC CTG ACT TCG CTT CTG AGC CAC CCG ACC TTC GTC GAT -1440
 G V K T N I P F L T S L L S H P T F V D -480

Fig. 5 The nucleotide sequence of the *A. flavus* F3 pyruvate carboxylase gene and its bordering regions and amino acid sequence deduced from the F3PYC gene. Translation initiation codon ATG was

positioned at nt +1. Translation terminator codon TAA was located at nt +3579. The amino acid sequences for N-linked glycosylation sites were boxed and in gray (Supplementary material)

GGA AAC TGC TGG ACC ACT TTC ATC GAC GAC ACC CCT GAA TTG TTC TCT CTT GTC GGC AGT -1500
 G N C W T I F I D D T P E L F S L V G S -500
 CAG AAC CGT GCC CAG AAG CTG CTC GCA TAC CTC GGC GAT GTA GCT GTC AAC GGT AGT AGC-1560
 Q N R A Q K L L A Y L G D V A V N G S S -520
 N-glycosylation site
 ATC AAG GGC CAA ATT GGC GAG CCC AAG CTC AAG GGT GAT GTC ATC AAG CCG AAG CTT TTC-1620
 I K G Q I G E P K L K G D V I K P K L F -540
 GAT GCC GAG GGC AAG CCG CTT GAC GTT TCC GCC CCC TGC ACC AAG GGT TGG AAG CAG ATT-1700
 D A E G K P L D V S A P C T K G W K Q I -560
 CTG GAC CGG GAG GGC CCG GCT GCC TTT GCG AAG GCC GTG CGT GCC AAC AAG GGT TGC TTG-1760
 L D R E G P A A F A K A V R A N K G C L -580
 ATC ATG GAT ACT ACC TGG CGT GAC GCC CAC CAG TCT TTG CTG GCC ACC CGT GTG CGT ACC-1820
 I M D T T W R D A H Q S L L A T R V R T -600
 ATC GAC TTG TTG AAC ATC GCC CAT GAG ACC AGC TAC GCC TAC TCC AAT GCG TAC AGT TTG -1880
 I D L L N I A H E T S Y A Y S N A Y S L -620
 GAA TGC TGG GGT GGT GCT ACC TTC GAT GTG GCC ATG CGT TTC CTC TAT GAG GAC CCC TGG-1920
 E C W G G A T F D V A M R F L Y E D P W -640
 GAC CGC CTG CGC AAG ATG CGT AAG GCT GTT CCT AAC ATC CCA TTC CAG ATG TTG CTC CGT -1980
 D R L R K M R K A V P N I P F Q M L L R -660
 GGT GCC AAC GGT GTC GCC TAC TCT TCC CTC CCA GAC AAC GCC ATC TAC CAC TTC TGT AAG -2040
 G A N G V A Y S S L P D N A I Y H F C K -680
 CAG GCT AAG AAG TGC GGT GTC GAC ATT TTC CGT GTT TTC GAC GCC CTC AAC GAT GTC GAT-2100
 Q A K K C G V D I F R V F D A L N D V D -700
 CAG CTC GAG GTC GGT ATC AAG GCT GTT CAT GCT GCC GAG GGT GTT GTC GAG GCC ACC ATG-2160
 Q L E V G I K A V H A A E G V V E A T M -720
 TGC TAC AGC GGT GAC ATG CTG AAC CCC CAC AAG AAG TAC AAC CTG GAG TAC TAC ATG GCC-2220
 C Y S G D M L N P H K K Y N L E Y Y M A -740
 TTG GTG GAT AAG ATT GTA GCC ATG AAG CCT CAC ATC CTT GGT ATC AAG GAT ATG GCC GGT-2280
 L V D K I V A M K P H I L G I K D M A G -760
 GTG CTG AAG CCC CAG GCC GCT CGC CTG TTG GTG GGC TCC ATC CGT CAG CGC TAC CCT GAC-2340
 V L K P Q A A R L L V G S I R Q R Y P D -780
 CTT CCC ATC CAC GTC CAC ACC CAC GAC TCC GCT GGT ACT GGT GTA GCT TCC ATG ATT GCC -2400
 L P I H V H T H D S A G T G V A S M I A -800
 TGT GCC CAG GCG GGT GCC GAC GCC GTG GAC GCC GCG ACC GAC AGC ATG TCC GGT ATG ACC-2460
 C A Q A G A D A V D A A T D S M S G M T -820
 TCC CAG CCT AGC ATT GGT GCC ATT CTG GCC TCT CTT GAG GGC ACT GAG CAA GAC CCC GGT-2520
 S Q P S I G A I L A S L E G T E Q D P G -840
 CTC AAC CTC GCC CAC GTG CGC GCT ATT GAT AGC TAC TGG GCA CAG CTG CGC TTG CTC TAC -2580
 L N L A H V R A I D S Y W A Q L R L L Y -860
 TCT CCT TTC GAG GCG GGT CTC ACT GGC CCC CAG CCT GAG GTC TAC GAG CAC GAT CTC CCT-2640
 S P F E A G L T G P D P E V Y E H E I P -880
 GGT GGT CAG TTG ACC AAG CTT ATC TTC CAG GCC AGT CAG CTC GGC TTG GGC CAG CAG TGG-2700
 G G Q L T N L I F Q A S Q L G L G Q Q W -900
 GCC GAA ACC AAG AAG GCC TAT GAG GCG GCT AAT GAT TTA CTC GGC GAC ATT GTA AAG GTC-2760
 A E T K K A Y E A A N D L L G D I V K V -920
 ACT CCC ACC TCC AAG GTG GTC GGT GAC TTG GCT CAG TTC ATG GTC TCG AAC AAA CTG ACT -2820
 T P T S K V V G D L A Q F M V S N K L T -940
 CCA GAG GAT GTT GTT GAG CGT GCT GGT GAG CTG GAC TTC CCT GGT TCT GTG CTC GAA TTC-2880
 P E D V V E R A G E L D F P G S V L E F -960
 CTC GAA GGT CTC ATG GGA CAG CCC TTC GGT GGA TTC CCC GAG CCA TTG CGC TCC CGC GCC-2940

Fig. 5 continued

the carboxyl transferase (CT) domain, and biotin carboxyl carrier protein (BCCP) domain (Fig. 5, Supplementary Material). The results in Fig. 5 indicated that the calculated

molecular mass of the protein deduced from the *F3PYC* gene was 131.2008 kDa, isoelectric point was 6.10 and the protein contained 1193 amino acids. The calculated

```

L E G L M G Q P F G G F P E P L R S R A -980
CTG CGC GAT CGC CGC AAG CTC GAG AAG CGT CCA GGT CTC TAC CTC GAG CCT TTG GAT TTG-3000
L R D R R K L E K R P G L Y L E P L D L -1000
GCT AAG ATC AAG AGC CAG ATC CGT GAG AAG TTC GGT GCT GCT ACT GAG TAT GAC GTG GCC-3060
A K I K S Q I R E K F G A A T E Y D V A-1020
AGC TAT GCC ATG TAT CCC AAG GTC TTC GAG GAC TAC AAG AAG TTC GTC CAG AAG TTC GGT-3120
S Y A M Y P K V F E D Y K K F V Q K F G -1040
GAT CTC TCC GTC TTG CCC ACA CGG TAC TTC TTG GCC AAG CCT GAG ATT GGC GAG GAG TTC-3180
D L S V L P T R Y F L A K P E I G E E F -1060
CAC GTT GAG CTG GAG AAG GGT AAG GTG CTC ATC CTG AAG TTG TTG GCC ATC GGC CCT CTT-3240
H V E L E K G K V L I L K L L A I G P L -1080
TCA GAG CAG ACT GGT CAG CGT GAG GTC TTC TAC GAA GTC AAC GGT GAG GTG CGC CAG GTC-3300
S E Q T G Q R E V F Y E V N G E V R Q V-1100
GCT GTT GAT GAC AAC AAG GCT TCC GTG GAC AAC ACT TCA CGC CCT AAG GCC GAT GTG GGT-3360
A V D D N K A S V D N T S R P K A D V G -1120
N-glycosylation site
GAC AGC AGC CAG GTC GGT GCT CCT ATG AGC GGT GTG GTT GTT GAA ATC CGT GTC CAC GAT-3420
D S S Q V G A P M S G V V V E I R V H D -1140
GGT CTG GAG GTT AAG AAG GGT GAC CCA CTT GCC GTC CTG AGT GCC ATG AAG ATG GAA ATG-3480
G L E V K K G D P L A V L S A M K M -1160
GTT ATC TCT GCT CCT CAC AGT GGA AAG GTC TCC AGC TTG CTG GTC AAG GAG GGC GAT TCT-3540
V I S A P H S G K V S S L L V K E G D S -1180
GTG GAT GGC CAG GAT CTC GTC TGC AAG ATC GTC AAA GCG TAA
V D G Q D L V C K I V K A *

```

Fig. 5 continued

molecular mass of *P. rubens* PYC is 127.531 kDa, and the number of amino acid residues of their mature proteins of PYC from *P. rubens*, is 1177 (Fu et al. 2015). The N-glycosylation sites of the protein was also analyzed at <http://cbs.dtu.dk/services/NetNGlyc> and the results obtained showed that there were four potential N-linked glycosylation sites of the protein, among them, –N–S–S–I– at 36 amino acid, –N–G–T–V– at 237 amino acid, N–G–S–S– at 517 amino acid, and N–T–S–R– at 1111 amino acid, were the most possible N-glycosylation sites (Fig. 5, Supplementary Material). In contrast, it has been reported that PYC from *P. rubens* have three potential glycosylation sites (Martin et al. 2007), respectively.

PYC determines the yield in a commercial amino acid production process (Pfefferle et al. 2003) and CO₂ fixation rates in *Rhizobium* species (Dunn et al. 1997). Some studies on *Rhizopus* sp. and an *Aspergillus* sp. were performed for acid production (Haarasilta and Taskinen 1977; Pfefferle et al. 2003). It was found that ORF of the cloned gene had 3579 bp encoding 1193 amino acids and was not disrupted by any intron. The size of ORFs encoding PYC from *Penicillium rubens* is 1191 bp (Fu et al. 2015). It mean that size of the *F3PYC* gene cloned in this study and the amino acids deduced from the gene were almost the same as those from other fungi. It was confirmed that most of the genes encoding PYC from fungi have no intron (Fu et al. 2015). Results in our study showed similarity with these studies.

Furthermore, sequence similarity analysis of the nucleotide sequence of pyruvate carboxylase (*F3PYC*)

from *A. flavus* (F3) strain showed ~0.5% difference compared with PYC of *A. parasiticus*, *N. fischeri*, *A. fumigatus*, *A. clavatus*, *A. niger*, *A. kawachii*, *A. ochraceo-roseus*, and *A. ruber*. Moreover, the predicted amino acid sequence holds a conserved domain of the family *Aspergillus* (Fig. 6, Supplementary Material). The protein 3D structure analysis deduced from the *F3PYC* gene at <http://swissmodel.expasy.org/>, showed ribbon model representation of the pyruvate carboxylase from *A. flavus* F3, with homotetramer (α_4) structure (Fig. 7).

No signal peptide was seen in the *F3PYC* deduced from the cloned gene, was a homotetramer (α_4), and containing biotin carboxylase (BC) domain, the carboxyl transferase (CT) domain, the biotin carboxyl carrier protein (BCCP) domain, and the central allosteric domain. In fact, in all eukaryotes and most prokaryotes, PYC is a homotetramer (α_4), with each monomer containing four functional domains: BC domain, CT domain, BCCP domain, and the central allosteric domain (Lietzan and Maurice 2013; Fu et al. 2015; Khan et al. 2017). In the α_4 PYC, the N-terminus of the polypeptide chain having the BC domain. Biotin by using bicarbonate as a substrate in a reaction is carboxylated in the BC domain which is involved in MgATP cleavage to form carboxybiotin, MgADP, and Pi and that requires free Mg²⁺. The CT domain holds a tightly bound transition metal ion and is the central domain in which the carboxyl group from carboxybiotin is transferred to pyruvate to form oxaloacetate. The BCCP domain, located at the C-terminus, contains the lysine residue to which the biotin is attached (Jitrapakdee et al. 2008).

Fig. 6 Multiple alignment of *A. flavus* (F3) pyruvate carboxylase (F3PYC) amino acid sequences with some PYC from other fungi by using DNAMAN Program (Lynnon 6BioSoft, Vandreuil, Canada) (Supplementary material)

Fig. 6 continued

Fig. 7 The ribbon representation of the secondary structure elements of the F3PYC from *A. flavus* F3. This ribbon representation is showing a homotetramer (α₄) (the purple, green, yellow, and blue) at <http://swissmodel.expasy.org/>

Conclusion

PYC enzyme takes part in the accumulation of TCA cycle intermediates in fungi. In study above, full-length F3PYC gene was isolated and characterized (accession number KT345940) from *A. flavus* (F3) strain with ~5.0% amino acid sequence difference compared with the PYC genes of other fungi, with an open reading frame (ORF) containing 3579 bp encoding a protein of 1193 amino acids. The protein deduced from the PYC gene had no signal peptide, was a homotetramer (α₄) and four potential N-linked glycosylation sites, among them, –N–S–S–I– at 36 amino acid, –N–G–T–V– at 237 amino acid, –N–G–S–S– at 517 amino acid, and –N–T–S–R– at 1111 amino acid, were the most possible N-glycosylation sites. If fungus requires CO₂ for its metabolism besides its own reserves, the ten times up-

regulation of the pyruvate carboxylase is needed to incorporate CO₂ for metabolism of fungus to generate more acids to enhance the process of bioremediation because pyruvate carboxylase enzyme incorporates this additional carbon source into oxaloacetate more effectively. This oxaloacetate in turn would be needed for fungal growth activation and other processes like gluconeogenesis and amino acid biosynthesis, etc. (Jitrapakdee et al. 2008). Therefore, our all previous findings show that F3PYC gene isolated from *A. flavus* (F3), which has high potential in the production of organic acid which may be very useful in remediation applications.

Acknowledgements This work was supported by the State Key Laboratory of Environmental Criteria and Risk Assessment (No. SKLECRA2013FP12) and the Public science and technology research funds projects of ocean, China (201105020).

Compliance with ethical standards

Conflict of interest Dear Editor, this is to certify that the reported work in this manuscript entitled “Cloning and characterization of F3PYC gene encoding Pyruvate carboxylase in *Aspergillus flavus* (F3) strain” submitted for publication is an original one and has not been submitted for publication elsewhere. We further certify that paper citation to the previously reported work have been given and no data/table/figure has been quoted verbatim from other publications without giving due acknowledgment and without the permission of the authors. The consent of all the authors of this paper has been obtained for submitting the paper to the “3Biotech Journal”.

References

- Al-Samarrai TH, Schmid J (2000) A simple method for extraction of fungal genomic DNA. *Lett Appl Microbiol* 30:53–56. doi:10.1046/j.1472-765x.2000.00664.x
- Barnett JA, Kornberg HLJ (1960) The utilization by yeasts of acids of the tricarboxylic acid cycle. *J Gen Microbiol* 23:65
- Branson JP, Nežić M, Wallace JC, Attwood PV (2002) Kinetic characterization of yeast pyruvate carboxylase isozyme PYC 1. *Biochemistry* 41:4459–4466
- Chi Z, Wang ZP, Wang GY, Khan I, Chi ZM (2014) Microbial biosynthesis and secretion of L-malic acid and its applications. *Crit Rev Biotechnol*. doi:10.3109/07388551.2014.924474
- Dunn MF, Vargas G, Peralta H, Mora Y, Mora R (1996) Fermentative and aerobic metabolism in *Rhizobium etli*. *J Bacteriol* 178:5960–5970
- Dunn MF, Araiza G, Cevallos MA, Mora J (1997) Regulation of pyruvate carboxylase in *Rhizobium etli*. *FEMS Microbiol Lett* 157:301–306
- Flores C-L, Rodríguez C, Petit T, Gancedo C (2000) Carbohydrate and energy-yielding metabolism in non-conventional yeasts. *FEMS Microbiol Rev* 24:507–529
- Fu J, Hou J, Chen L, Wang M, Shen Y, Zhang Z, Bao X (2015) The yeast BDF1 regulates endocytosis via LSP1 under salt stress. *Curr Microbiol*. 70(5):671–678
- Haarasilta S, Taskinen L (1977) Location of three key enzymes of gluconeogenesis in baker's yeast. *Arch Microbiol* 113:159–161
- Jitrapakdee S, St Maurice M, Rayment I, Cleland WW, Wallace JC, Attwood PV (2008) Structure, mechanism and regulation of pyruvate carboxylase. *Biochem J*. 413:369–387
- Kenealy WE, Zaady JC, du Preez B, Goldberg I (1986) Biochemical aspects of fumaric acid accumulation by *Rhizopus arrhizus*. *Appl Environ Microbiol* 52:128–133
- Khan I, Nazir K, Wang ZP, Liu GL, Chi Z (2014) Calcium malate overproduction by *Penicillium viticola* 152 using the medium containing corn steep liquor. *Appl Microbiol Biotechnol* 98(4):1539–1546
- Khan I, Qayyum S, Ahmed S, Niaz Z, Fatima N, Chi ZM (2016) Molecular cloning and sequence analysis of a PVGOX gene encoding glucose oxidase in *Penicillium viticola* F1 strain and its expression quantitation. *Gene* 592(2):291–302
- Khan I, Qayyum S, Ahmed S, Maqbool F, Tauseef I, Haleem KS, Chi ZM (2017) Cloning and characterization of pyruvate carboxylase gene responsible for calcium malate overproduction in *Penicillium viticola* 152 and its expression analysis. *Gene* 605:81–91
- Kubicek CP (1988) The role of the citric acid cycle in fungal organic acid fermentations. *Biochem Soc Symp* 54:113–126
- Lietzan AD, Maurice MS (2013) Insights into the carboxyltransferase reaction of pyruvate carboxylase from the structures of bound product and intermediate analogs. *Biochem Biophys Res Commun* 441:377–382
- Martin R, Hellwig D, Schaub Y, Bauer J, Walther A, Wendland J (2007) Functional analysis of *Candida albicans* genes whose *Saccharomyces cerevisiae* homologues are involved in endocytosis. *Yeast* 24(6):511–522
- Moller DE (2001) New drug targets for type 2 diabetes and the metabolic syndrome. *Nature* 414:821–827
- Mukhopadhyay B, Concar EM, Wolfe RS (2001) A GTP-dependent vertebrate-type phosphoenolpyruvate carboxykinase from *Mycobacterium smegmatis*. *J Biol Chem* 276:16137–16145
- Nonaka K, Masuma R, Iwatsuki M, Shiomu K, Otaguro K, Omura S (2011) *Penicillium viticola*, a new species isolated from a grape in Japan. *Mycoscience* 52:338–343
- Osmani SA, Scrutton MC (1983) The sub-cellular localisation of pyruvate carboxylase and of some other enzymes in *Aspergillus nidulans*. *Eur J Biochem* 133(3):551–560
- Osmani SA, Scrutton MC (1985) The sub-cellular localisation and regulatory properties of pyruvate carboxylase from *Rhizopus arrhizus*. *Eur J Biochem* 147:119–128
- Peleg YA, Barak MC, Scrutton M, Goldberg I (1989) Malic acid accumulation by *Aspergillus flavus*. III. ¹³C NMR and isoenzyme analyses. *Appl Microbiol Biotechnol* 30:176–183
- Pfefferle W, Mockel B, Bathe B, Marx A (2003) Biotechnological manufacture of lysine. *Adv Biochem Eng Biotechnol* 79:59–112
- Purohit HJ, Ratledge C (1988) Mitochondrial location of pyruvate carboxylase in *Aspergillus niger*. *FEMS Microbiol Lett* 55:129–132
- Saghai-Marouf MA, Soliman KM, Jorgensen RA, Allard RW (1984) Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. *Proc Natl Acad Sci* 81:8014–8019