

## Full Paper

# Comparative transcriptome analysis revealed the improved $\beta$ -carotene production in *Sporidiobolus pararoseus* yellow mutant MuY9

(Received May 8, 2018; Accepted July 6, 2018; J-STAGE Advance publication date: December 13, 2018)

Chunji Li,<sup>1</sup> Bingxue Li,<sup>1,\*</sup> Ning Zhang,<sup>2</sup> Qifan Wang,<sup>1</sup> Wenjing Wang,<sup>1</sup> and Hongtao Zou<sup>1,\*</sup>

<sup>1</sup> College of Land and Environment, Shenyang Agricultural University, Shenyang 110866, People's Republic of China

<sup>2</sup> College of Bioscience and Biotechnology, Shenyang Agricultural University, Shenyang 110866, People's Republic of China

Carotenoids are ubiquitous constituents of living organisms. These structurally diverse pigments have received considerable attention due to their biotechnological applications and potential beneficial effects on human health. In this study, we characterized an over-producing  $\beta$ -carotene mutant of *Sporidiobolus pararoseus*, obtained by ultraviolet mutagenesis, named MuY9. We compared the transcriptome between the wild-type and MuY9. A total of 348 differential expressed genes (DEGs) were found, and only one DEG *crtYB* is involved in carotenoid biosynthesis. The overproduction of  $\beta$ -carotene in MuY9 should be attributed to the up-regulation of *crtYB*. Functional identification of *crtYB* was performed using heterologous complementation in *Escherichia coli*. Our findings indicate that the enzymatic conversions of geranylgeranyl pyrophosphate to phytoene, as well as lycopene to  $\beta$ -carotene, are catalyzed by this *CrtYB*. Furthermore, our insights into the *crtYB* gene should facilitate a more detailed understanding of the carotenogenic pathway in *S. pararoseus*, and advance the development of the genetic engineering for the bio-production of carotenoids.

**Key Words:**  $\beta$ -carotene; carotenoids; *crtYB*; *Sporidiobolus pararoseus*; transcriptome

## Introduction

Carotenoids are one of the most diverse classes of natural occurring fat-soluble pigments (Maiani et al., 2009). They are synthesized by all photosynthesizing species, but also by various fungi and non-photosynthetic bacteria

(Cazzonelli, 2011). Carotenoids, which are among the most effective antioxidants, have received much attention for their potential antimicrobial properties and as anti-cancer agents (Fiedor and Burda, 2014). Due to their bright color and unique biological properties, carotenoids are of value in the pharmaceutical, chemical and food industries (Frengova and Beshkova, 2009).  $\beta$ -carotene ( $\beta$ , $\beta$ -carotene), as a salient member of carotenoids, are natural colorants that occur in fruits and vegetables (Grune et al., 2010).  $\beta$ -carotene consists of 40 carbon atoms in a core structure of 9 fully conjugated double bonds constructed with two  $\beta$ -ionone rings (Cerón-Carrasco et al., 2009). In addition to being an antioxidant, the major function of  $\beta$ -carotene in human nutrition is as an important precursor of vitamin A (Nzamwita et al., 2017). Intervention studies have shown  $\beta$ -carotene represents moderate UV protective effects in the human skin (Biesalski and Obermueller-Jevic, 2001). Currently, approximately 90% of commercialized  $\beta$ -carotene is produced using chemosynthesis (Ajikumar et al., 2008). However, the microbial bioprocess for large-scale  $\beta$ -carotene production has resulted in a strong interest, due to the increasingly restricted rules and regulations applied to obtaining chemicals (Bu et al., 2017).

A number of red yeast species produce a mixture of carotenoids, namely,  $\beta$ -carotene, torulene, torularhodin and  $\gamma$ -carotene, resulting in pigmented colonies (Mannazzu et al., 2015). Among red yeasts, the species *Sporidiobolus* are being currently exploited for the biotechnological production of  $\beta$ -carotene using conventional fed-batch fermentation (Chaiyaso and Manowattana, 2018). Optimization of fermentation conditions, such as changing the nitrogen source and controlling the temperature, also efficiently promote  $\beta$ -carotene production (Han et al., 2012, 2016). However, the low efficiency of  $\beta$ -carotene production by this strategy limits its industrial application.

\*Corresponding authors: Bingxue Li and Hongtao Zou, College of Land and Environment, Shenyang Agricultural University, Shenyang 110866, People's Republic of China.

Tel: +86-24-88487155 Fax: +86-24-88487155 E-mail:libingxue1027@163.com (B. Li); zouhongtao2001@163.com (H. Zou)

None of the authors of this manuscript has any financial or personal relationship with other people or organizations that could inappropriately influence their work.

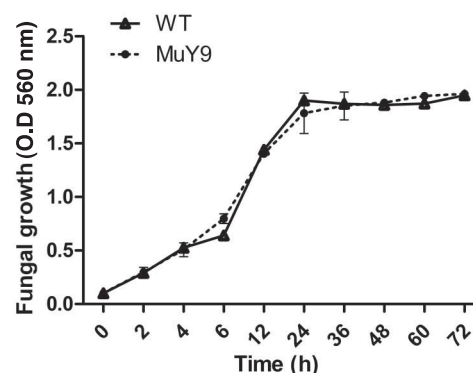
To increase  $\beta$ -carotene yield, a lot of genetic engineering strategies have been developed, such as overexpression of the rate-limiting carotenogenic enzyme in engineered strain (Heider et al., 2014). Moreover, the biosynthesis of carotenoid in red yeasts has been extensively investigated. High carotenoid yields, biosafety, and convenient fermentative operation, indicate *S. pararoeseus* as potential sources for industrial pigment production (Kot et al., 2018). However, despite the considerable number of reports on the bio-production of the carotenoid of *S. pararoeseus* (Li et al., 2017), little is known about the molecular mechanisms underlying  $\beta$ -carotene biosynthesis (Mannazzu et al., 2015). Owing to the unavailability of whole genome information, transcriptome sequencing has become an alternative to identify the precise nature of the genes involved in  $\beta$ -carotene biosynthesis in *S. pararoeseus*. Lycopene cyclase and phytoene synthase are two key enzymes involved in the biosynthetic pathway of  $\beta$ -carotene, encoded by a single gene in red yeasts (Guo et al., 2014). These bifunctional fusion genes commonly contain two domains: (1) the R domain, located near the 5' end, which encodes lycopene cyclase activity; (2) the A domain, which is downstream of domain R and encodes phytoene synthase (Sanz et al., 2011).

Here, we characterized the transcriptome of the *S. pararoeseus* yellow mutant MuY9 and its wild-type, using the Illumina single-end sequencing technology. The transcriptome analysis contains *de novo* assembly, gene annotation, gene classification and enrichment. Comparison of the transcriptome data of MuY9 and its wild-type revealed that significant differences in the expression of genes involved in carotenoid biosynthesis. In addition, we found that the up-regulation of *crtYB* gene is expected to be the main reason for  $\beta$ -carotene overproduction of yellow mutant. Furthermore, this study is the first to clone and functionally identify the *S. pararoeseus crtYB* gene by heterogenous complementary in *E. coli*. Together, our findings suggest that the *S. pararoeseus crtYB* represents an alternative gene source for the reconstruction of the carotenogenic pathway.

## Materials and Methods

**Strains material.** *S. pararoeseus* (WT, CCGMCC 2.5280) ( $\beta$ -carotene yield: 538.0  $\mu\text{g/g}$ ,  $\text{DW}$ ) was separated from strawberry fruit in Shenyang city of Liaoning province, China. A yellow mutant MuY9 ( $\beta$ -carotene yield: 653.2  $\mu\text{g/g}$ ,  $\text{DW}$ ) was obtained through UV-induced mutation. The growing rates of WT and MuY9 of *S. pararoeseus* are shown in Fig. 1. WT and MuY9 were cultured in 250 mL Erlenmeyer baffle flasks containing 50 mL of the YPD medium (10 g/L yeast extract, 20 g/L peptone and 20 g/L glucose) at 28°C on a rotary shaker at 180 rpm. *E. coli* strains DH5 $\alpha$  and BL21 (DE3) were obtained from Takara Bio (Dalian, China) and grown in LB medium at 37°C.

**Analysis of carotenoids.** Following freeze-drying of the cells, carotenoids were resuspended in DMSO-acetone (1:3, v/v) and extracted at 65°C for 30 min. The carotenoids-containing upper hydrophobic phase was collected, dried under a stream of  $\text{N}_2$ , and then dissolved in acetone. Thin-layer chromatography (TLC) analysis of



**Fig. 1.** The growth rate of *S. pararoeseus* WT and MuY9 on YPD medium.

The solid line represents the WT and the dotted line represents MuY9. Yeast growth kinetics was evaluated by optical density at 560 nm after incubation for 0, 2, 4, 6, 12, 24, 36, 48, 60 and 72 h at 28°C at a rotary of 180 rpm. Error bars represent standard deviations of triplicate experiments. WT and Y9 exhibit similar growth trends from 0 h to 72 h under the same culture conditions.

carotenoids was performed under a cyclohexane-ethyl acetate (95:5, v/v) solvent system for 20 min. Subsequently, high-performance liquid chromatography (HPLC) analysis of carotenoids was performed on an Agilent 1100 system equipped with a photodiode array detector (Agilent, USA). Pigments were separated on a reverse-phase column (C18, 2.7  $\mu\text{m}$ , 150  $\times$  4.6 mm, Dikma Diamonsil (2), USA). Isocratic elution system: acetonitrile-methanol-isopropanol (85:10:5, v/v/v) as the mobile phase at a column temperature of 32°C and a flow rate of 1.0 ml/min. Lycopene and  $\beta$ -carotene standard substances were purchased from Dalian Meilun Biotech Co., Ltd (Dalian, China).

**RNA extraction and library preparation.** *S. pararoeseus* cells were harvested by centrifugation at 6,000  $\times g$  for 10 min at 4°C. The total RNA of each sample was extracted using Trizol reagent kit (Invitrogen, USA). The RNA concentration was measured using a Qubit RNA Assay Kit in Qubit 2.0 Fluorometer (Life Technologies, USA). RNA integrity was assessed using the RNA Nano 6000 Assay Kit of the Agilent Bioanalyzer 2100 system (Agilent Technologies, USA). RNA degradation and contamination was monitored on 1% agarose gels. RNA purity was checked using the NanoPhotometer spectrophotometer (IMPLEN, USA). A total amount of 3  $\mu\text{g}$  was used for Illumina mRNA sequencing. Sequencing libraries were generated using NEBNext Ultra RNA Library Prep Kit for Illumina (NEB, USA) following the manufacturer's recommendations, and index codes were added to attribute sequences to each sample. Briefly, mRNA was purified from total RNA using poly-T oligo-attached magnetic beads. Fragmentation was carried out using divalent cations under an elevated temperature in a NEBNext First Strand Synthesis Reaction Buffer (5X). First strand cDNA was synthesized using a random hexamer primer and M-MuLV Reverse Transcriptase (RNase H-). Second strand cDNA synthesis was subsequently performed using DNA polymerase I and RNase H. Remaining overhangs were converted into blunt

**Table 1.** Oligonucleotide primers used in this study.

| Primer name     | Sequence(5'-3')        |
|-----------------|------------------------|
| 26S-F           | AGCGGAGGAAAAGAACTAACA  |
| 26S-R           | TGCGGTCTATCACGGAAAAC   |
| <i>crtYB</i> -F | GAATCAACCCGATACGAAACA  |
| <i>crtYB</i> -R | TAATCTCAAGCGACCATCAAAC |

All primers were synthesized by Shanghai Sangon Ltd. (China).

ends via exonuclease/polymerase activities. After adenylation of 3' ends of DNA fragments, NEB Next Adaptor with a hairpin loop structure were ligated to prepare for hybridization. In order to select cDNA fragments of preferentially 150~200 bp in length, the library fragments were purified with an AMPure XP system (Beckman Coulter, USA). Then 3  $\mu$ L of USER Enzyme (NEB, USA) was used with size-selected, adaptor-ligated cDNA at 37°C for 15 min followed by 5 min at 95°C before PCR. Then PCR was performed with Phusion High-Fidelity DNA polymerase, Universal PCR primers and Index (X) Primer. Finally, PCR products were purified and library quality was assessed on the Agilent Bioanalyzer 2100 system (Bi et al., 2018).

**De novo transcriptome assembly.** Raw reads generated by Illumina Hiseq 2000 were processed to obtain clean reads by removing adaptor sequences, reads containing poly-N, and low quality reads through in-house perl scripts. Q20, Q30, GC-content and sequence duplication level of the clean data were calculated. Clean reads were randomly clipped into short fragments (K-mers) by applying the Trinity program. The K-mers with a certain length of overlap were combined to form longer fragments and contigs. The overlap between these contigs was utilized to build graph components. The clean reads were then mapped back onto the assembled transcriptome using the method of the De Bruijn graph. Finally, the unique assembled transcriptome were further subjected to the process of sequences-splicing redundancy removal using TIGR gene indices clustering tools to acquire non-redundant transcripts called unigenes.

**Functional annotation.** cDNA sequences of unigenes were aligned to the Nr database (NCBI non-redundant protein sequences) via blastx (v 2.2.28,  $E$ -value  $\leq 1 \times 10^{-5}$ ). Information of the top 10 hits of each unigene in Nr was extracted through an in-house script. Blasting against an NT database (NCBI non-redundant nucleotide sequence) was performed using blastn. Annotation in the Pfam (Protein family) database and Swiss-Prot (a manually annotated and reviewed protein sequence database) were performed using hmmscan. Annotation in KOG/COG (Clusters of Orthologous Group of protein) was performed via blastx. Functional GO annotation was performed using the blast2GO program (v3.0). Metabolic pathway analysis was carried out through mapping unigenes to the KEGG database via KEGG Automatic Annotation Server (KAAS). Unique KO IDs encoded by *Sepia* unigenes were extracted and submitted to iPath v2, with the edge width standing for the frequency of each KO IDs (edge width =  $\log_2$  (Frequency+1)  $\times$  5). When a unigene failed to align to any of

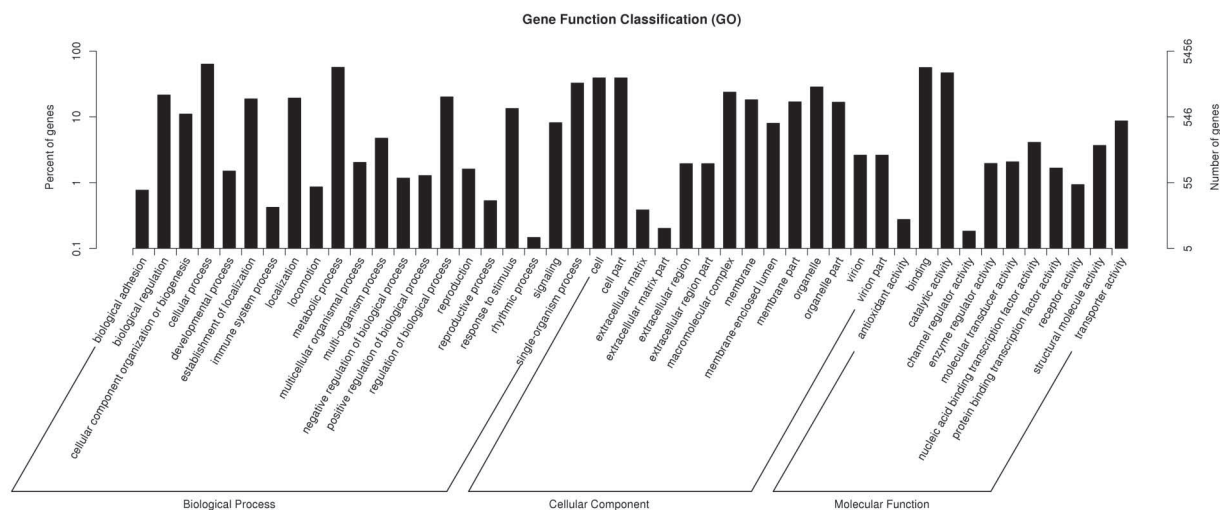
the above databases, the ESTScan software (3.0.3) was introduced to predict the coding regions fragment (ORF) and determine the direction of the sequence (Iseli et al., 1999).

**Differential gene expression analysis.** Differential expression analysis of two groups was performed using the DESeq R package (1.10.1). DESeq provides statistical routines for determining differential expression in digital gene expression data using a model based on the negative binomial distribution. The resulting  $P$  values were adjusted using the Benjamini and Hochberg's approach for controlling the false discovery rate. Genes with an adjusted  $P$ -value  $< 0.05$  found by DESeq were assigned as differentially expressed.

**Cloning and identification of *crtYB*.** For the PCR amplification of *crtYB* with restriction enzymes sites for *Bam*HI and *Xho*I, forward primer *crtYB*-F (CGGGATCCATGGGTTTCGATTACT) and reverse primer *crtYB*-R (CCTCGAGCTACAGGGCCCTCCAG-GCA) were designed on the basis of the complete *crtYB* gene sequence of *S. pararoseus* resulting from DGEs. The carotenogenic gene *crtYB* was amplified from the cDNA of *S. pararoseus* by using KOD DNA polymerase (Toyobo, Japan) under the following conditions: one cycle at 94°C for 5 min, 30 cycles at 98°C for 10 s, 54°C for 30 s, and 68°C for 2 min, and a final extension cycle at 68°C for 7 min. The PCR products were then linked into the cloning vector pTA2 (Toyobo, Japan) and then digested with *Bam*HI and *Xho*I. The expression vector pET30a was digested with the same restriction enzymes. Products were ligated with T4 DNA ligase (Takara, Japan) and then transformed into the *E. coli* DH5 $\alpha$  competent cell, and the positive expression vector was named as pET30a-*crtYB*.

The plasmid pAC-LYC carrying three bacterial (*Erwinia herbicola*) carotenogenic genes, *crtE*, *crtB* and *crtI*, was constructed according to a previously described method (Cunningham et al., 1994). The plasmid pAC-LYCA (*crtB*) carrying two bacterial (*E. herbicola*) carotenogenic genes *crtE* and *crtI*, was obtained by deletion of the adjacent *Nco*I-*Nco*I fragment and then compensated with a fragment of a complete chloramphenicol-resistance gene (GenBank: V00622.1) based on the plasmid pAC-LYC. The previous two expression systems were used to measure the activity of the putative lycopene cyclase/phytoene synthase. For functional analysis of the lycopene cyclase/phytoene synthase, pET30a-*crtYB* was transformed into the GGPP accumulating strain of *E. coli* BL21 (DE3) carrying the reconstructed plasmid pAC-LYCA (*crtB*); plasmid pET30a-*crtYB* was transformed into the lycopene accumulating strain of *E. coli* BL21 (DE3) carrying the plasmid pAC-LYC. The expression vector pET30a lacking *crtYB* gene was used for the control strains. The recombinant *E. coli* BL21 (DE3) strains were cultured in LB medium to OD600 0.5 and induced with 1 mM IPTG. Cultures were grown for 24 h in the dark, and cells were collected by centrifugation at  $6,000 \times g$  (rav 8 cm) for 10 min.

**Real-time PCR analysis.** To confirm our differential expression results of RNA-Seq, we conducted quantitative reverse transcription PCR assays for the selected DEGs



**Fig. 2.** Assembled unigenes were functionally classified by Gene Ontology categorization.

The unigenes related to three main categories: biological process, cellular component, and molecular function.

in the same RNA samples used for RNA-Seq. The total RNA was extracted using the Yeast RNAiso Kit (Takara, Japan), and the RNA was digested with Recombinant DNase I (RNase-free) (Takara, Japan). The digested RNA was used for cDNA synthesis using the M-MLV RTase cDNA Synthesis Kit (Takara, Japan). Quantitative reverse transcription PCR was performed in a StepOne Plus System (Applied Biosystems, USA) using the SYBR Premix Ex Taq II Kit (Takara, Japan) according to the manufacturer's protocol. Primers used are listed in Table 1. A 20  $\mu$ L PCR system was prepared that contained 2  $\mu$ L cDNA templates, 0.8  $\mu$ L of each primer (0.4  $\mu$ M), 10  $\mu$ L SYBR Premix Ex Taq II, 0.4  $\mu$ L ROX Reference Dye (50 $\times$ ) and 6  $\mu$ L RNase-free water. As a negative control, the template DNA was replaced by PCR-grade water. The specificity of the amplified PCR product was identified by performing melting curve analysis using a StepOne Plus system (Applied Biosystems, USA). The 26S rDNA was selected as an internal reference. For analysis of the quantitative results, the relative expression of each gene was calculated by the comparative crossing point (CP) method and presented as  $2^{-\Delta\Delta C_t}$ .

## Results and Discussion

### Illumina sequencing and de novo assembly

Using Illumina Hiseq 2000 platform produces a total of 48,234,372 single-ends raw reads (SE100) for three biological replicates of WT and MuY9, which generated 11.5 Gb of data. The nucleotide analysis displays a GC and Q30 content of 49.63% and 95.79%, respectively. After a trimming process to remove adaptor, primer sequences, poly-A tails and low quality reads, we obtained 47,534,688 cleaned and high-qualified reads. The clean data generated from this study have been submitted to the NCBI Sequence Read Archive (SRA), with the SRA accession of WT1, WT2 and WT3: SRP131948; SRA accession of MuY91: SRR6975874; SRA accession of MuY92: SRR6975879; SRA accession of MuY93: SRR6975878. All clean reads were *de novo* assembled and produced a

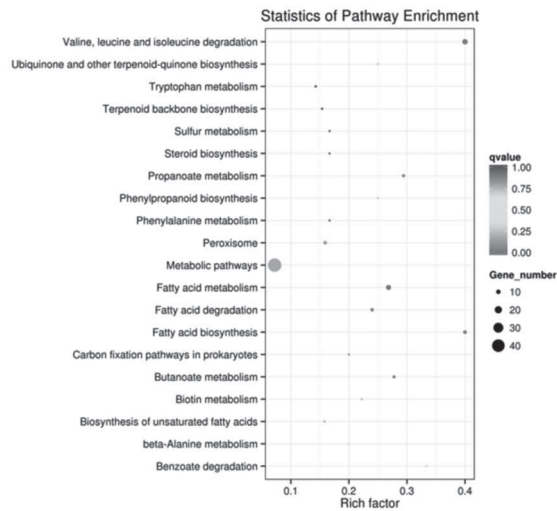
set of 13,586 transcripts with an N50 size of 2371 bp and an average size of 1654 bp using Trinity. The transcriptome library generated 9533 unigenes with an N50 size of 2266 bp and an average size of 1538 bp. Among the transcripts, a total of 2655 (19.54%), 2529 (18.61%), 4403 (32.4%), and 3999 (29.45%) transcripts had size ranging in 200–500 bp, 501–1 kbp, 1001–2 kbp, and >2 kbp, respectively. Among the unigenes, a total of 2232 (23.41%), 1814 (19.02%), 2959 (31.03%) and 2528 (26.54%), with size in the ranges of 200–500 bp, 501–1 kbp, 1001–2 kbp, and >2 kbp, respectively.

### Functional annotation of assembled unigenes

The assembly-derived unigenes were annotated based on the following databases: Nr (NCBI non-redundant protein sequences), Nt (NCBI non-redundant nucleotide sequences), Pfam (Protein family), KOG/COG (Clusters of Orthologous Groups of proteins), Swiss-Prot (a manually annotated and reviewed protein sequence database), KO (KEGG Ortholog database), GO (Gene Ontology). As indicated, a total of 9533 unigenes, 6103(64.02%), 2406(25.24%), 4578(48.02%), 5305(55.65%) and 3778 unigenes showed annotation hits to homologous sequences in NR, NT, KO, SwissProt, PFAM, GO, KOG and SwissProt databases, respectively. A total of 6826 (71.6%) assembled unigenes has been annotated in at least one protein and nucleotide database, and 441 unigenes (4.62%) showed annotation hits to all databases. Details of the distribution of the annotated unigenes in the GO database (classified as biological process, cellular component and molecular function) are shown in Fig. 2.

### DGEs between WT and MuY9

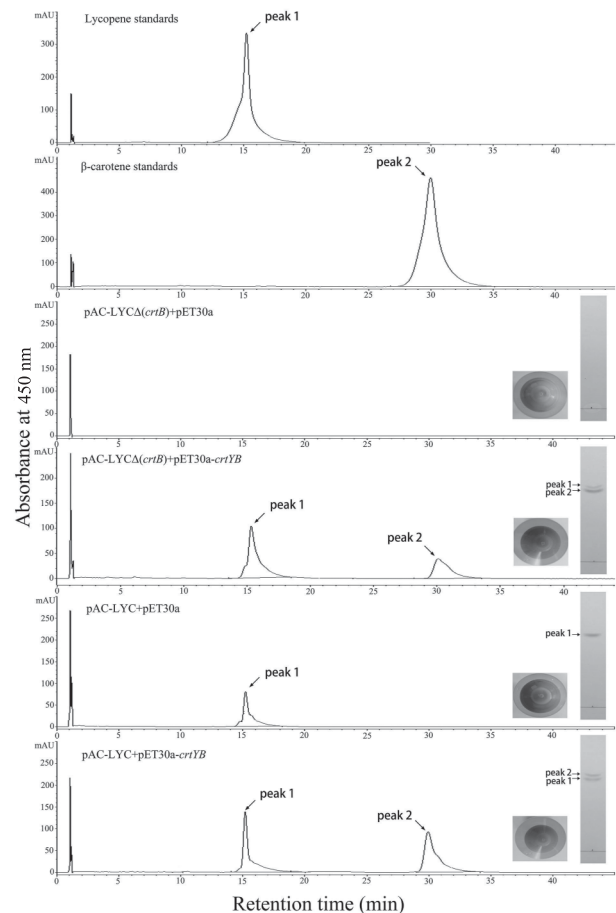
In order to investigate the DEGs between WT and MuY9 samples, the adjusted *p*-value method was employed. The expression level of unigenes with a threshold of  $\text{padj} < 0.05$  is considered as significantly differential expressed (Yang et al., 2016). Moreover, the up-regulated and down-regulated genes were filtered by combining with the  $\log_2$  fold change  $\geq 1$  to avoid false positives. Based on these



**Fig. 3.** The most top 20 enriched KEGG metabolism pathway categories of the differentially expressed genes within the WT and MuY9 datasets.

values, we determined 348 DEGs between WT and MuY9 libraries (Supplementary Table S1). To perform the functional annotation of the 348 DEGs, we subjected the DEGs to GO analysis, including three main categories: biological process, cellular component, and molecular function. Eight significantly-enriched GO terms associated with oxidoreductase activity molecular function (GO: 0016491), oxidation-reduction process biological process (GO: 0055114), monocarboxylic acid metabolic process biological process (GO: 0032787), fatty acid metabolic process biological process (GO: 0006631), monocarboxylic acid biosynthetic process biological process (GO: 0072330), fatty acid biosynthetic process biological process (GO: 0006633), and fatty acid synthase complex cellular component (GO: 0005835), fatty acid synthase activity molecular function (GO: 0004312). It is noteworthy that the significantly-enriched GO terms are mainly concentrated in the biological process category. Furthermore, the KEGG analysis revealed that those 348 DEGs were enriched to 67 canonical reference pathways. The most top 20 enriched key pathways are presented in Fig. 3. The dominant three pathways contain: metabolic pathway (43 DGEs), valine, leucine and isoleucine degradation (12 DGEs), and fatty acid metabolism (11 DGEs). Furthermore, DGE comp1515 (GenBank: MH487469) involved in terpenoid backbone biosynthesis (ko. 00900) were identified. DGE comp1515 encoding acetyl-CoA C-acetyltransferase was up-regulated in MuY9.

In this study, except for those significantly-enriched KEGG pathways mentioned above, we also noticed other important pathways associated with carotenoids biosynthesis (ko. 00906). Fortunately, we found that the gene *comp5921* is predicted to encode phytoene synthase and lycopene cyclase (*crtYB*), which was up-regulated by 1.78-fold in MuY9. The biosynthetic pathway for carotenoid production of red yeast has been documented by Simpson et al. and Mata-Gomez et al. (Mata-Gomez et al., 2014; Simpson et al., 1964). Condensation of two molecules of GGPP leading to phytoene and cyclization of lycopene



**Fig. 4.** The HPLC and TLC analysis of the co-expression product of pAC-LYC/pAC-LYC $\Delta$  (*crtB*) and pET30a-*crtYB*/pET30a in recombinant *E. coli* BL21 (DE3).

transformed to  $\beta$ -carotene via  $\gamma$ -carotene are encoded by a single *crtYB*. We assumed that the overproduction of  $\beta$ -carotene in MuY9 should be attributed to the up-regulation of *crtYB*. This speculation inspired us to verify the function of the putative *crtYB*.

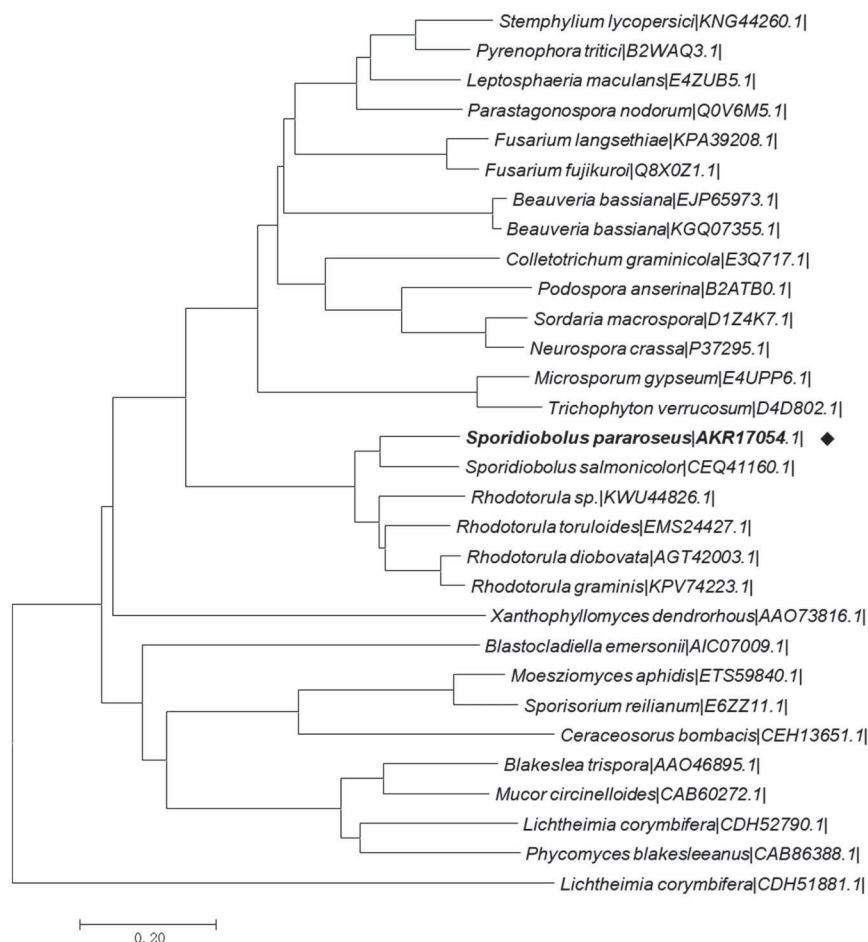
#### Validation of gene expression

To confirm the transcriptome sequencing data, we tested the mRNA expression of *crtYB* by qPCR. These results indicated that the *crtYB* expression levels detected by transcriptome sequencing analysis were mostly consistent with the qPCR (qPCR: up-regulated 2.0-fold). This comparison confirmed the reliability of the transcriptome sequencing results.

#### Cloning of *crtYB* and construction of the plasmid pAC-LYC $\Delta$ (*crtB*)

Based on the *de novo* transcriptome, we obtained a full-length genomic DNA and cDNA clone of the *crtYB* gene by PCR. Comparison of the two sequences suggested that the genomic DNA contains eight introns, with all splicing sites following the GT-AG rule. The cDNA sequences and deduced amino acids of the *crtYB* gene reported in this paper are available in the GenBank databases under the accession number: KR108013.

In order to identify the activity of the putative lycopene cyclase/phytoene synthase of *S. pararoseus*, we con-



**Fig. 5.** Phylogenetic tree constructed with available CrtYB sequences using the neighbor-joining method.

MEGA 7.0 software was used to carry out all evolutionary analysis.

constructed the plasmid pAC-LYCA (*crtB*) according to Guo (Guo et al., 2014). To test whether the plasmid pAC-LYCA (*crtB*) was functional, we cloned the *crtB* gene of *Erwinia herbicola* (Cunningham et al., 1994). And then, we inserted the *crtB* into the expression vector pET30a using T4 DNA ligase, which was named pET30a-*crtB*-ori. *E. coli* BL21 (DE3) cells carrying pAC-LYCA (*crtB*) were transformed with pET30a-*crtB*-ori plasmid, with transformed colonies displaying a color change from colorless to red. Control strains carrying the insert-free plasmid pET30a and pAC-LYCA (*crtB*) showed no change in color. HPLC analysis showed that the accumulating pigments (lycopene) of the recombinant strain carrying pAC-LYCA (*crtB*) transformed with pET30a-*crtB*-ori plasmid were identical with the strain carrying the plasmid pAC-LYC. Together, these results indicated that the plasmid pAC-LYCA (*crtB*) was fully functional in *E. coli*.

#### Functional analysis of putative *crtYB*

In order to identify the function of the putative *crtYB*, we constructed plasmid pET30a-*crtYB*. Functional analysis of the pET30a-*crtYB* expression product was performed via complementation experiments in *E. coli* BL21 (DE3) carrying the plasmid pAC-LYCA (*crtB*) or pAC-LYC. As shown in Fig. 4, the color of thalli of strains carrying pAC-

LYCA (*crtB*) transformed with pET30a-*crtYB* plasmid turned from colorless to yellow, while control transformants carrying the plasmid pET30a lacking the *crtYB* insertion remained colorless; the color of thalli of strains carrying pAC-LYC transformed with pET30a-*crtYB* plasmid turned from red to yellow, while control transformants carrying the plasmid pET30a lacking the *crtYB* gene insertion remained red. To illuminate the nature of the lycopene cyclase/phytoene synthase of *S. pararoseus*, we performed TLC and HPLC analysis. As shown in Fig. 4, the expression product of the *S. pararoseus* CrtYB appeared to be essential for the conversion from GGPP to phytoene as well as lycopene to  $\beta$ -carotene. Therefore, we reasoned that the lycopene cyclase and phytoene synthase activities are encoded by a single *crtYB* gene from *S. pararoseus*. Moreover, our previous study revealed that salt stress increase the yield of carotenoids of *S. pararoseus* (Li et al., 2017). The clone and characterization of *crtYB* described here should provide much-needed insights into the molecular basis underlying the changes of carotenoids production.

#### Comparative sequence analysis and phylogenetic tree construction

The comparative sequence analysis was performed via

NCBI Conserved Domain Search (Marchler-Bauer et al., 2015). The N-terminus and C-terminus of the bifunctional *S. pararoseus* CrtYB are characterized as lycopene cyclase and phytoene synthase domain, respectively. Moreover, there are a substrate binding pocket, a substrate-Mg<sup>2+</sup> binding site and two aspartate-rich regions in the downstream of lycopene cyclase domain of *S. pararoseus*. The lycopene cyclase domain is generally repeated twice within the same polypeptide, as is observed in fungi, archaea, and sphingobacteria (Sanz et al., 2011). However, in *Myxococcus* genome, this domain appears as a single polypeptide, tandemly repeated and usually in a genomic context consistent with a role in carotenoid biosynthesis (Iniesta et al., 2008). As shown in Fig. 5, we constructed the phylogenetic tree by the neighbor-joining method. Phylogenetic profiling analysis was performed to compare the *S. pararoseus* CrtYB to other sequenced fungi. Our analysis showed that the amino acid sequence of the *S. pararoseus* CrtYB exhibits significant homologies with that the *Sporidiobolus salmonicolor* (GeneBank: CEQ41160.1), *Rhodotorula toruloides* (GeneBank: EMS24427.1), *Rhodotorula graminis* (GeneBank: KPV74223.1), and *Rhodospiridium diobovatum* (GeneBank: AGT42003.1).

The biochemistry and evolution of lycopene cyclases and phytoene synthases have been investigated in detail for various microorganisms and higher plants (Maresca et al., 2007). The bifunctional lycopene cyclase/phytoene synthase evolves from the fusion of the ancestral bacterial *crtYc/crtYd*-type and phytoene synthase gene by recombinant processes (Iniesta et al., 2008; Krubasik and Sandmann, 2000). An example of how a carotenogenic fusion gene originated by chromosomal rearrangement was reported earlier for *Rubrivivax gelatinosus* (Ouchane et al., 1997). Furthermore, the amino acid sequence of lycopene cyclases and phytoene synthases from plants and cyanobacteria were shown to be closely related (Sandmann, 2002). Together, these observations indicate that the evolutions of lycopene cyclase and phytoene synthase are from archaea and several groups of bacteria via cyanobacteria and green algae to plants.

## Conclusions

Here, both  $\beta$ -carotene and the comparative transcriptome analysis of *S. pararoseus* and its yellow mutant MuY9 were carried out. Among the DGEs, the increased expression level of gene *crtYB* (1.78-fold) was mainly responsible for the improved  $\beta$ -carotene in MuY9. Moreover, we cloned and functionally identified the bifunctional *crtYB* using heterologous complementation in *E. coli*. Together, the availability of the transcriptome data from *S. pararoseus* made it possible to elucidate the carotenoid biosynthesis pathway in this yeast.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 31271818), the National Key Research and Development Program of China (Project No. 2017YFD200807-3) and the Natural Science Foundation of Liaoning Province of China (Grant No. 2015020763) and the Shenyang Science and Technology Innovation in Agriculture Special Fund (Grant No. F17-150-3-00).

## Conflict of Interest

All authors declare that they have no conflict of interest.

## Supplementary Materials

Supplementary table is available in our J-STAGE site (<http://www.jstage.jst.go.jp/browse/jgam>).

## References

- Ajikumar, P. K., Tyo, K., Carlsen, S., Mucha, O., Phon, T. H. et al. (2008) Terpenoids: opportunities for biosynthesis of natural product drugs using engineered microorganisms. *Mol. Pharmaceut.*, **5**, 167–190.
- Bi, H., Guo, M., Wang, J., Qu, Y., Du, W. et al. (2018) Transcriptome analysis reveals anthocyanin acts as a protectant in *Begonia semperflorens* under low temperature. *Acta Physiol. Plant.*, **40**, 10.
- Biesalski, H. K. and Obermueller-Jevic, U. C. (2001) UV light, beta-carotene and human skin—beneficial and potentially harmful effects. *Arch. Biochem. Biophys.*, **389**, 1–6.
- Bu, X., Sun, L., Shang, F., and Yan, G. (2017) Comparative metabolomics profiling of engineered *Saccharomyces cerevisiae* lead to a strategy that improving  $\beta$ -carotene production by acetate supplementation. *PLoS One*, **12**, e188385.
- Cazzonelli, C. I. (2011) Goldacre Review: Carotenoids in nature: insights from plants and beyond. *Funct. Plant Biol.*, **38**, 833–847.
- Cerón-Carrasco, J. P., Bastida, A., Zúñiga, J., Requena, A., and Miguel, B. (2009) Density functional theory study of the stability and vibrational spectra of the  $\beta$ -carotene isomers. *J. Phys. Chem. A*, **113**, 9899–9907.
- Chaiyaso, T. and Manowattana, A. (2018) Enhancement of carotenoids and lipids production by oleaginous red yeast *Sporidiobolus pararoseus* KM281507. *Prep. Biochem. Biotech.*, **48**, 13–23.
- Cunningham, F. J., Sun, Z., Chamovitz, D., Hirschberg, J., and Gantt, E. (1994) Molecular structure and enzymatic function of lycopene cyclase from the cyanobacterium *Synechococcus* sp strain PCC7942. *Plant Cell*, **6**, 1107–1121.
- Fiedor, J. and Burda, K. (2014) Potential role of carotenoids as antioxidants in human health and disease. *Nutrients*, **6**, 466–488.
- Frengova, G. I. and Beshkova, D. M. (2009) Carotenoids from *Rhodotorula* and *Phaffia*: yeasts of biotechnological importance. *J. Ind. Microbiol. Biot.*, **36**, 163–180.
- Grune, T., Lietz, G., Palou, A., Ross, A. C., Stahl, W. et al. (2010)  $\beta$ -Carotene is an important vitamin A source for humans. *J. Nutr.*, **140**, 2268S–2285S.
- Guo, W., Tang, H., and Zhang, L. (2014) Lycopene cyclase and phytoene synthase activities in the marine yeast *Rhodospiridium diobovatum* are encoded by a single gene *crtYB*. *J. Basic Microb.*, **54**, 1053–1061.
- Han, M., He, Q., and Zhang, W. (2012) Carotenoids production in different culture conditions by *Sporidiobolus pararoseus*. *Prep. Biochem. Biotech.*, **42**, 293–303.
- Han, M., Xu, Z., Du, C., Qian, H., and Zhang, W. (2016) Effects of nitrogen on the lipid and carotenoid accumulation of oleaginous yeast *Sporidiobolus pararoseus*. *Bioproc. Biosyst. Eng.*, **39**, 1425–1433.
- Heider, S. A. E., Peters-Wendisch, P., Netzer, R., Stafnes, M., Brautaset, T. et al. (2014) Production and glucosylation of C50 and C40 carotenoids by metabolically engineered *Corynebacterium glutamicum*. *Appl. Microbiol. Biot.*, **98**, 1223–1235.
- Iniesta, A. A., Cervantes, M., and Murillo, F. J. (2008) Conversion of the lycopene monocyclus of *Myxococcus xanthus* into a bicyclus. *Appl. Microbiol. Biot.*, **79**, 793–802.
- Iseli, C., Jongeneel, C. V., and Bucher, P. (1999) ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, 138–148.
- Kot, A. M., Błażej, S., Gientka, I., Kieliszek, M., and Bryś, J. (2018) Torulene and torularhodin: “new” fungal carotenoids for industry? *Microb. Cell Fact.*, **17**, 49.
- Krubasik, P. and Sandmann, G. (2000) Molecular evolution of lycopene cyclases involved in the formation of carotenoids with ionone end

- groups. *Biochem. Soc. Transact.*, **28**, 806–810.
- Li, C., Zhang, N., Li, B., Xu, Q., Song, J. et al. (2017) Increased torulene accumulation in red yeast *Sporidiobolus pararoseus* NGR as stress response to high salt conditions. *Food Chem.*, **237**, 1041–1047.
- Maiani, G., Periago Castón, M. J., Catasta, G., Toti, E., Cambrodón, I. G. et al. (2009) Carotenoids: Actual knowledge on food sources, intakes, stability and bioavailability and their protective role in humans. *Mol. Nutr. Food Res.*, **53**, S194–S218.
- Mannazzu, I., Landolfo, S., Da Silva, T. L., and Buzzini, P. (2015) Red yeasts and carotenoid production: outlining a future for non-conventional yeasts of biotechnological interest. *World J. Microb. Biot.*, **31**, 1665–1673.
- Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F. et al. (2015) CDD: NCBI's conserved domain database. *Nucleic Acids Res.*, **43**, D222–D226.
- Maresca, J. A., Graham, J. E., Wu, M., Eisen, J. A., and Bryant, D. A. (2007) Identification of a fourth family of lycopene cyclases in photosynthetic bacteria. *PNAS*, **104**, 11784–11789.
- Mata-Gomez, L. C., Montanez, J. C., Mendez-Zavala, A., and Aguilar, C. N. (2014) Biotechnological production of carotenoids by yeasts: an overview. *Microb. Cell Fact.*, **13**, 12.
- Nzamwita, M., Duodu, K. G., and Minnaar, A. (2017) Stability of  $\beta$ -carotene during baking of orange-fleshed sweet potato-wheat composite bread and estimated contribution to vitamin A requirements. *Food Chem.*, **228**, 85–90.
- Ouchane, S., Picaud, M., Vernotte, C., and Astier, C. (1997) Photooxidative stress stimulates illegitimate recombination and mutability in carotenoid-less mutants of *Rubrivivax gelatinosus*. *EMBO J.*, **16**, 4777–4787.
- Sandmann, G. (2002) Molecular evolution of carotenoid biosynthesis from bacteria to plants. *Physiol. Plantarum*, **116**, 431–440.
- Sanz, C., Velayos, A., Alvarez, M. I., Benito, E. P., and Eslava, A. P. (2011) Functional analysis of the *Phycomyces carRA* gene encoding the enzymes phytoene synthase and lycopene cyclase. *PLoS One*, **6**, e23102.
- Simpson, K. L., Nakayama, T. O. M., and Chichester, C. O. (1964) Biosynthesis of yeast carotenoids. *J. Bacteriol.*, **88**, 1688–1694.
- Yang, F., Cao, H., Xiao, Q., Guo, X., Zhuang, Y. et al. (2016) Transcriptome analysis and gene identification in the pulmonary artery of broilers with ascites syndrome. *PLoS One*, **11**, e156045.