Alternative routes of acetyl-CoA synthesis identified by comparative genomic analysis: involvement in the lipid production of oleaginous yeast and fungi

Tayvich Vorapreeda,¹ Chinae Thammarongtham,¹
Supapon Cheevadhanarak^{2,3} and Kobkul Laoteng¹

Correspondence

Kobkul Laoteng
kobkul@biotec.or.th or
laotengk@yahoo.com

¹Biochemical Engineering and Pilot Plant Research and Development Unit, National Center for Genetic Engineering and Biotechnology at King Mongkut's University of Technology Thonburi, Bangkhuntien, Bangkok 10150, Thailand

²School of Bioresources and Technology, King Mongkut's University of Technology Thonburi, Bangkhuntien, Bangkok 10150, Thailand

³Pilot Plant Development and Training Institute, King Mongkut's University of Technology Thonburi, Bangkhuntien, Bangkok 10150, Thailand

For a bio-based economy, microbial lipids offer a potential solution as alternative feedstocks in the oleochemical industry. The existing genome data for the promising strains, oleaginous yeasts and fungi, allowed us to investigate candidate orthologous sequences that participate in their oleaginicity. Comparative genome analysis of the non-oleaginous (*Saccharomyces cerevisiae*, *Candida albicans* and *Ashbya gossypii*) and oleaginous strains (*Yarrowia lipolytica*, *Rhizopus oryzae*, *Aspergillus oryzae* and *Mucor circinelloides*) showed that 209 orthologous protein sequences of the oleaginous microbes were distributed over several processes of the cells. Based on the 41 sequences categorized by metabolism, putative routes potentially involved in the generation of precursors for fatty acid and lipid synthesis, particularly acetyl-CoA, were then identified that were not present in the non-oleaginous strains. We found a set of the orthologous oleaginous proteins that was responsible for the biosynthesis of this key two-carbon metabolite through citrate catabolism, fatty acid β -oxidation, leucine metabolism and lysine degradation. Our findings suggest a relationship between carbohydrate, lipid and amino acid metabolism in the biosynthesis of acetyl-CoA, which contributes to the lipid production of oleaginous microbes.

Received 13 June 2011
Revised 14 September 2011
Accepted 13 October 2011

INTRODUCTION

Lipids are dynamically bioactive molecules that contribute to the regulation of several complex systems of living cells. Besides the medical perspective, the remarkable growth of the lipid field is undoubtedly driven by the demand for feedstock for the oleochemical industry. In addition to the production of n-3 and n-6 polyunsaturated fatty acids that are beneficial to human health, extensive attention is being directed to biodiesel production from micro-organisms to replace non-sustainable petroleum (Liu & Zhao, 2007; Vicente *et al.*, 2010). Certain strains, such as *Rhodospiridium toruloides*, *Lipomyces starkeyi*, *Yarrowia lipolytica* and *Mucor circinelloides*, are known to accumulate substantial amounts of lipids, accounting for more than 20 % of their biomass, and are thus called oleaginous strains (Ageitos *et al.*, 2011; Beopoulos *et al.*, 2009; Meng *et*

al. 2009; Ratledge, 2004). Due to their short cultivation time, their high level of intracellular lipids that are predominantly triacylglycerol (TAG) and their utilization of various substrates, oleaginous yeasts and fungi have become important model systems for alternatives to traditional sources of lipids derived from fossil, animal and plant origins (Beopoulos *et al.*, 2009; Ratledge, 2004). Therefore, an understanding of lipid physiology is required to improve the efficient production of lipids of commercial interest. An integrated approach has been implemented using recent developments for studying the lipid metabolism of these promising micro-organisms. It has been reported that the lipid production of these oleaginous species is enhanced by controlling cultivation or nutritional conditions (Certik *et al.*, 1999; Ruenwai *et al.*, 2010). Based on the biochemistry of the lipid-accumulation process studied by Ratledge & Wynn (2002), the channelling of carbon flux to lipid biosynthesis by nutrient imbalance or by using nitrogen-limited conditions has been reported. The biochemical process is accomplished by the cooperative function of key enzymes, which include acetyl-CoA

Abbreviations: ACL, ATP: citrate lyase; HMM, hidden Markov model.

Four supplementary tables are available with the online version of this paper.

carboxylase, ATP: citrate lyase (ACL) and malic enzymes, which generate the precursor pools required for fatty acid synthesis. Moreover, the formation of lipid droplets, which mainly contain TAG and steryl esters, is a prerequisite for fungal oleaginicacy as well as peroxisomal β -oxidation (Athenstaedt *et al.*, 2006; Beopoulos *et al.*, 2008; Daum *et al.*, 2007). Although these processes seem to be conserved, there is a considerable variation among oleaginous species in accumulated lipid content, the sources of which remain unclear. In addition to physiological manipulation, lipid overproduction in yeast and fungi has been pursued through genetic modification (Kamisaka *et al.*, 2007; Zhang *et al.*, 2007). However, the overexpression of one or a few genes involved in the committing steps of fatty acid biosynthesis does not usually increase the total lipid content of the engineered strains by a substantial amount (Courchesne *et al.*, 2009; Ruenwai *et al.*, 2009).

Currently, the increase in available genome data is accelerating progress toward and understanding of lipid metabolism at the systems level. A topic of particular interest is the genome-wide study of lipid metabolism in the simple eukaryotic model *Saccharomyces cerevisiae* (Dobson *et al.*, 2010; Nookaew *et al.*, 2008), which is an industrial yeast with various commercial applications. Furthermore, a genomic comparison of TAG synthesis between *S. cerevisiae* and *Y. lipolytica* has provided molecular insights into oleaginicacy. It has been shown that the availability of glycerol 3-phosphate is crucial for lipid accumulation, and that acyl-CoA transferases are the limiting enzymes in TAG formation (Beopoulos *et al.*, 2008). The genome-wide studies carried out so far have been directed toward the oleaginous yeasts. Relatively little information from systems-level analysis is available for filamentous fungi with regard to their ability to store a high lipid content. Based on the naturally occurring variation in the lipid contents and fatty acid profiles among species, a comparative analysis of existing sequence data would provide an understanding of the genetic backgrounds of the complex biological systems involved in the lipid metabolism of oleaginous yeast and fungi, and would offer new biotechnological perspectives for developing efficient microbial platforms for the production of valuable lipids. In this work, comparative genome analysis of non-oleaginous and oleaginous micro-organisms was performed to investigate candidate orthologous genes and enzymes related to oleaginicacy. The oleaginous strains studied were selected based on their genome availability and oleaginicacy features, and included *Aspergillus oryzae* (Kaur & Worgan, 1982; Lin *et al.*, 2010; Meng *et al.*, 2009; Santos-Fo *et al.*, 2011), *M. circinelloides* (Vicente *et al.* 2010), *Rhizopus oryzae* (Ratledge, 1986; Suzuki, 1991) and *Y. lipolytica* (Beopoulos *et al.*, 2009), which have been shown to be potential lipid producers (Ageitos *et al.*, 2011; Ratledge, 2004). We also emphasized the metabolism of acetyl-CoA, which is the key two-carbon metabolite in several metabolic processes, particularly in terms of precursor supply for fatty acid synthesis in oleaginous strains.

METHODS

Database retrieval. Protein sequences and gene annotation data from the genome databases on *R. oryzae* RA99-880 (http://www.broadinstitute.org/annotation/genome/rhizopus_oryzae/MultiHome.html) (Ma *et al.*, 2009), *A. oryzae* RIB40 (<http://www.bio.nite.go.jp/dogan/project/view/AO>) (Machida *et al.*, 2005), *Y. lipolytica* CLIB122 (<http://www.genolevures.org/download.html>) (Sherman *et al.*, 2009), *S. cerevisiae* (http://www.genome.jp/kegg-bin/show_organism?org=sce) (Goffeau *et al.*, 1996), *Candida albicans* sc53140, version 21 (http://www.broadinstitute.org/annotation/genome/candida_group/MultiHome.html) (Jones *et al.*, 2004), *Ashbya gossypii* ATCC10895 (ftp://ftp.ncbi.nih.gov/genomes/Fungi/Ermothecium_gossypii_uid10623/) (Dietrich *et al.*, 2004) and *M. circinelloides* CBS277.49, v2.0 (<http://genome.jgi-psf.org/Mucci2/Mucci2.download.ftp.html>) were retrieved.

Identification of orthologous proteins. Orthologous proteins were identified based on the reciprocal best-hit approach, using BLASTP (Altschul *et al.*, 1997) for each pairwise species comparison with a cut-off E-value of 10^{-10} . This E-value, which is a rather high-stringency cut-off, was used to filter out distantly related sequences that might not be genuine orthologous sequences. To investigate the functional relevance of orthologous proteins to metabolic or biological processes, their functional annotations were classified according to the information in the KEGG (Kanehisa *et al.*, 2004), KOG (Tatusov *et al.*, 2003) and GO (Gene Ontology Consortium; Harris *et al.*, 2004) databases. Hidden Markov model (HMM) profiles were also used to confirm the functions of the related orthologous proteins by BLAST search against the Pfam protein family database, with an E-value of 10^{-10} as a lower limit cut-off (Finn *et al.*, 2010). The orthologous proteins were subsequently mapped to relevant metabolic pathways.

Conserved domain identification. The functional domains of the proteins were identified by comparing amino acid sequences between the enzymes of interest and the conserved protein families using Pfam release 24.0 at the Sanger Centre, UK (<http://pfam.sanger.ac.uk/>) and a default E-value cut-off of 1.0 (Finn *et al.*, 2010). The query protein's functional regions were searched against the domains Pfam-A and Pfam-B, which were defined based on expert knowledge, sequence similarity, other protein family databases and the ability of HMM profiles to correctly identify and align the members. The types and locations of the domains identified within a group of the proteins of interest were presented as a graphical summary by the MyDomains – Image Creator, which is available at the ExPASy Proteomics Server (<http://expasy.org/tools/mydomains/>).

Analysis of phylogenetic relationships. Amino acid sequences were aligned using the multi-sequence alignment program available as the MAFFT software (<http://mafft.cbrc.jp/alignment/software/>) (Katoh & Toh, 2008) under the FFT-NS-i algorithm (iterative refinement method). The process was automatically repeated, for a maximum of 1000 iterations, until no improvement in scoring alignment was obtained. The phylogenetic relationship between the groups of interesting enzymes was analysed under the assumption of a Jones–Taylor–Thornton (JTT) substitution model (Jones *et al.*, 1992). These matrices were used to elaborate the phylogenetic tree using the neighbour-joining method (Saitou & Nei, 1987). The phylogenetic tree was then displayed using Archaeopteryx (applet v0.955) (<http://phylosoft.org/archaeopteryx/>).

Prediction of protein subcellular localization. Protein localization was analysed by using the programs TargetP 1.1 (Emanuelsson *et al.*, 2000), MitoProtII – v1.101 (Claros & Vincens, 1996) and PTS1 Predictor (Neuberger *et al.*, 2003) to predict the N-terminal cleavage sites, mitochondrial targeting sequences and peroxisomal target signals, respectively. The statistical values were used according to

the default parameter settings of the programs. The potential cleavage sites and target sites of protein transport were then predicted.

RESULTS

Identification and classification of orthologous genes in oleaginous strains

Based on the ability of yeasts and fungi to produce different amounts of lipids, genome comparisons between oleaginous and non-oleaginous strains were undertaken to unravel cellular metabolic processes that might participate in oleaginic. In this work, the genome sequences of four oleaginous strains, *Y. lipolytica*, *A. oryzae*, *R. oryzae* and *M. circinelloides*, were compared with those of *Ashbya gossypii*, *S. cerevisiae* and *C. albicans*, which are non-oleaginous yeasts. Comparing the seven genomes based on reciprocal best hits with the cut-off E-value of 10^{-10} showed that 209 protein sequences were orthologous among the four oleaginous strains. These proteins were designated 'orthologous oleaginous proteins' (Table 1). Using information from the KEGG database, the orthologous oleaginous sequences were categorized into five classes (Table 1), including genes involved in metabolism (19.6 % of total orthologous oleaginous genes), genetic information processing (18.2 % of total orthologous oleaginous genes), environmental information processing (6.7 % of total orthologous oleaginous genes), cellular processes (9.1 % of total orthologous oleaginous genes) and unclassified functions (46.4 % of total orthologous oleaginous genes) (see Supplementary Table S1). It was noted that nearly a quarter of the orthologous oleaginous proteins (41 sequences) are responsible for metabolic processes or metabolism, as shown in Fig. 1. Of these proteins, we further focused on the data relevant to the metabolism of precursors required for fatty acid synthesis and lipid accumulation.

Comparative study of carbohydrate metabolism in oleaginous and non-oleaginous strains

Partitioning carbons from central carbon metabolism to fatty acid synthesis has been thought to be a potential route for enhancing lipid content. Acetyl-CoA is the essential two-carbon donor molecule for fatty acid synthesis. Considering carbohydrate metabolism based on their genome data, four metabolic routes for the synthesis of acetyl-CoA via pyruvate metabolism were found in the oleaginous and non-oleaginous species. The enzymes involved in the pathways included pyruvate kinase, pyruvate dehydrogenase, malic enzyme, lactate dehydrogenase and acetyl-CoA synthetase, as shown in Fig. 2. The computational identification of the orthologous ACL gene also indicates that oleaginous strains could synthesize acetyl-CoA via ACL in addition to the four general routes, which is consistent with earlier reports that the enzyme plays a crucial role in the direct generation of acetyl-CoA during the lipid accumulation phase (Ratledge, 2002). This

Table 1. Distribution of orthologous oleaginous genes in yeast and fungal genomes

Description	Number of genes
Metabolism	
Carbohydrate metabolism	8
Energy metabolism	2
Lipid metabolism	9
Nucleotide metabolism	1
Amino acid metabolism	14
Glycan biosynthesis and metabolism	3
Metabolism of cofactors and vitamins	3
Metabolism of terpenoids and polyketides	1
Genetic information processing	
Transcription	7
Translation	12
Folding, sorting and degradation	19
Environmental information processing	
Membrane transport	7
Signal transduction	5
Signalling molecules and interaction	2
Cellular processes	
Transport and catabolism	7
Cell motility	2
Cell growth and death	9
Cell communication	1
Unclassified function	
Poorly characterized protein	43
Uncharacterized protein	54

enzyme was not found in the non-oleaginous yeasts studied in this work, indicating the presence of an additional route of acetyl-CoA biosynthesis in the oleaginous strains.

In addition to energy and pyruvate production, glycolysis is responsible for the generation of glycerol 3-phosphate as a building material for the *de novo* synthesis of phospholipids and triglycerides. We identified orthologous oleaginous sequences in the glycolysis pathway, such as that for enolase. In addition, a set of the orthologous oleaginous sequences in carbohydrate metabolism included glycosyl-transferase group I, formyl-CoA transferase, hydroxy-methylglutaryl-CoA lyase, UDP-glucose dehydrogenase and UDP-glucuronate decarboxylase, which had no homologues in the non-oleaginous strains (see Supplementary Table S1).

Acetyl-CoA biosynthesis via lipid metabolism in oleaginous and non-oleaginous strains

In addition to carbohydrate metabolism, acetyl-CoA is also generated via other metabolic processes. Using genome analysis, we found orthologous protein sequences of the oleaginous strains categorized in lipid metabolism, which might contribute to the generation of acetyl-CoA through fatty acid degradation, also called β -oxidation (Fig. 2). In addition to the peroxisomal β -oxidation, fatty acid

Fig. 1. Classification of orthologous oleaginous genes based on their relevant functions and metabolic processes. The bar chart shows the number of genes in each of five major classes, while the pie chart shows the percentage of genes devoted to different aspects of metabolism.

degradation occurs in the mitochondria of oleaginous microbes, as evidenced by the finding of mitochondrion-targeted orthologous proteins involved in β -oxidation. These proteins included acyl-CoA dehydrogenase, 3-hydroxyacyl-CoA dehydrogenase, 2-enoyl-CoA hydratase and acetyl-CoA acetyltransferase, the first two enzymes of which had no homology with any protein sequences of the non-oleaginous strains. As shown in Fig. 2, there was a protein sequence (B10) found only in two closely related genomes belonging to the Zygomycetes (*M. circinelloides* and *R. oryzae*). The result of BLAST analysis showed that this unique Zygomycetes protein (zHAD) showed the highest similarity to 3-hydroxyacyl-CoA dehydrogenase of

bacteria, which catalyses the reduction of 3-hydroxyacyl-CoA to 3-oxoacyl-CoA. The results of Pfam analysis showed that zHAD contained only a functional catalytic domain and had similarity to the peroxisomal multifunctional β -oxidation protein (FOX2) of the yeast *S. cerevisiae*, particularly in the domain of the 3-hydroxyacyl-CoA dehydrogenase (LCHAD) (Fig. 3). However, it did not share a significant homology with the mitochondrial LCHAD of yeasts and fungi, and the mitochondrial targeting sequence was undetectable in the zHAD sequence (see Supplementary Tables S2–S4). The characterization of enzymes involved in the catalysis of 3-hydroxyacyl-CoA is summarized in Table 2. Based on phylogenetic tree

Fig. 2. Schematic diagram of the acetyl-CoA metabolism of non-oleaginous and oleaginous micro-organisms. Sce, Cal, Ago, Yli, Aor, Ror and Mci stand for *S. cerevisiae*, *C. albicans*, *Ashbya gossypii*, *Y. lipolitica*, *A. oryzae*, *R. oryzae* and *M. circinelloides*, respectively. A green background indicates the presence of the metabolic pathways of individual organisms. Upper-case letters followed by numbers indicate the enzyme name abbreviations, of which the names are provided in Table 3.

Fig. 3. Functional domains of the enzymes catalysing the third step of β -oxidation in yeast and fungi. The scale bar indicates the approximate length in amino acids of each protein. (a) FOX2, (b) LCHAD, (c) zHAD.

analysis, these proteins were clustered into three groups, as shown in Fig. 4. Taken together, the zHAD of the Zygomycetes was separately classified into a new group (group III) with unknown localization.

Acetyl-CoA synthesis in oleaginous strains via alternative routes of amino acid metabolism

Acetyl-CoA can be produced via the degradation of branched-chain amino acids (e.g. leucine) in addition to the aforementioned routes. Obviously, some enzymic reactions in leucine degradation are similar to those found in the degradation of fatty acids, including dehydrogenation, carboxylation, hydration and thioester hydrolysis forming acetyl-CoA. By genome analysis, we found the orthologous oleaginous enzymes that might be involved in leucine degradation in the oleaginous strains, which include branched-chain α -keto acid dehydrogenase (BCKDH), isovaleryl-CoA dehydrogenase (IVD), methylcrotonoyl-CoA carboxylase (MCCA) and

hydroxymethylglutaryl-CoA lyase (HMGCL), as shown in Fig. 2 and Table 3. Notably, based on the KEGG database, HMGCL was categorized as an orthologous oleaginous protein in carbohydrate metabolism. These enzymes have been well studied in mammals and are recognized as housekeeping proteins. The deficiency of these enzymes results in deficient catabolism of the carbon skeleton of leucine (Wysocki & Hähnel, 1986; Harris *et al.*, 2005). Unlike non-oleaginous strains, the presence of the additional pathway of leucine degradation in the oleaginous fungi indicates a theoretical route with potential for providing the acetyl-CoA substrate.

With respect to lysine biosynthesis, in oleaginous strains we found the candidate orthologous gene dihydrodipicolinate synthase (DHDPS), which may participate in amino acid synthesis via an alternative route, the α,ϵ -diaminopimelic acid pathway (see Supplementary Table S1). Thus, DHDPS is likely to be responsible for catalysing the first unique step of lysine biosynthesis without consuming the acetyl-CoA precursor; the latter event is usually found in prokaryotes

Table 2. Characteristics of the enzymes that catalyse the reduction of 3-hydroxyacyl-CoA to 3-oxoacyl-CoA in oleaginous and non-oleaginous strains

Characteristic	LCHAD	zHAD	FOX2
Enzyme name	3-Hydroxyacyl-CoA dehydrogenase	3-Hydroxyacyl-CoA dehydrogenase	Peroxisomal multifunctional β -oxidation protein
EC number	1.1.1.35	1.1.1.35	4.2.1.74
Functional type	Monofunctional protein	Monofunctional protein	Multifunctional protein
Organism	<i>Y. lipolytica</i> <i>A. oryzae</i> <i>R. oryzae</i> <i>M. circinelloides</i>	<i>R. oryzae</i> <i>M. circinelloides</i>	<i>Ashbya gossypii</i> <i>S. cerevisiae</i> <i>C. albicans</i> <i>Y. lipolytica</i> <i>A. oryzae</i> <i>R. oryzae</i> <i>M. circinelloides</i>
Amino acid length	271–319	260	891–916
Subcellular localization	Mitochondria	Unknown	Peroxisome

Fig. 4. Evolutionary relationship of the enzymes catalysing the reduction of 3-hydroxyacyl-CoA to 3-oxoacyl-CoA in different micro-organisms. FOX2, LCHAD and zHAD indicate the peroxisomal multifunctional β -oxidation protein, mitochondrial 3-hydroxyacyl-CoA dehydrogenase and Zygomycete 3-hydroxyacyl-CoA dehydrogenase with unknown localization, respectively.

(Dobson *et al.*, 2004) and higher plants (Vauterin *et al.*, 1999). Therefore, it can be postulated that in oleaginous fungi, lysine can be synthesized via two pathways, the α -amino adipic acid and α,ϵ -diaminopimelic acid pathways, whereas the non-oleaginous strains synthesize this essential amino acid only via the α -amino adipate pathway. If the flux of lysine synthesis is directed more toward the α,ϵ -diaminopimelic route, the lipid content may increase as a result of the greater availability of the acetyl-CoA substrate being channelled to the lipid biosynthetic pathway. Moreover, in fungi, lysine is degraded via the reversible reactions of the α -amino adipate pathway. The first half of the pathway begins with the degradation of lysine via saccharopine, 2-amino adipate-6-semialdehyde, 2-amino adipate, glutaryl-dihydrolipoamide and glutaryl-CoA (Xu *et al.*, 2006), after which, the last molecule, glutaryl-CoA, may be modified through a series of acyl-CoA intermediates to produce the end product, acetyl-CoA, which then returns to be a substrate for lysine synthesis. We identified the first half of the pathway in both oleaginous and non-oleaginous strains. Interestingly, the enzymes involved in the latter path of lysine catabolism were specific to the oleaginous strains, including glutaryl-CoA dehydrogenase, enoyl-CoA

hydratase and 3-hydroxyacyl-CoA dehydrogenase (categorized in lipid metabolism), as shown in Fig. 2. Thus, acetyl-CoA generated from lysine metabolism might be mobilized to the cytosol for fatty acid synthesis in oleaginous fungi. Therefore, this set of enzymes might be potential targets for increasing the flux of lipid biosynthesis by increasing the acetyl-CoA supply via amino acid metabolism.

Other orthologous oleaginous genes in the metabolism of other amino acids included 3-deoxy-7-phosphoheptulonate synthase, glutamine synthetase, S-adenosylmethionine (SAM)-dependent 6-methyltransferase and a putative O-methyltransferase that had no homology with the non-oleaginous fungi (see Supplementary Table S1).

DISCUSSION

On the basis of sequence similarity, comparative genome analysis, which is an effective approach for searching possible functional contexts of interest, was conducted in this work to investigate how oleaginous strains accumulate lipids at significant levels. It was hypothesized that this phenotypic characteristic requires a balance of the

Table 3. Comparison of the enzymes involved in acetyl-CoA biosynthesis between oleaginous and non-oleaginous microbes

O, present; —, absent; NA, not assigned; NL, not clearly localized or identified.

Enzyme	Catalytic function	EC number	Oleaginous	Non-oleaginous
Acetyl-CoA generation process				
A1	Pyruvate kinase	2.7.1.40	O	O
A2	Pyruvate dehydrogenase	1.2.4.1	O	O
A3	Malic enzyme	1.1.1.40	O	O
A4	Lactate dehydrogenase	1.1.2.4	O	O
A5	Acetyl-CoA synthetase	6.2.1.1	O	O
A6	ACL	2.3.3.8	O	—
β-Oxidation				
Peroxisomal β -oxidation				
B1	Fatty acyl-coenzyme A oxidase	1.3.3.6	O	O
B2	Peroxisomal multifunctional β -oxidation protein	4.2.1.74	O	O
B3	3-Ketoacyl thiolase	2.3.1.16	O	O
Mitochondrial β -oxidation				
B4	Acyl-CoA dehydrogenase	1.3.99.-	O	—
B5	3-Hydroxyacyl-CoA dehydrogenase	1.1.1.35	O	—
B6	Enoyl-CoA hydratase	4.2.1.17	O	NL
B7	Acetyl-CoA acetyltransferase	2.3.1.9	O	NL
Unclassified β -oxidation				
B8	Fatty acyl-coenzyme A oxidase	1.3.3.6	NA	NA
B9	Enoyl-CoA hydratase	4.2.1.17	NA	NA
B10	3-Hydroxyacyl-CoA dehydrogenase (similar to bacterial protein)*	1.1.1.35	O	—
B11	3-Ketoacyl thiolase	2.3.1.16	NA	NA
Leucine degradation				
Leucine degradation I†				
C1	Branched-chain amino acid aminotransferase	2.6.1.42	O	O
C2	Branched-chain keto acid dehydrogenase E1, β polypeptide	—	O	—
C3	Isovaleryl-CoA dehydrogenase	1.3.99.10	O	—
C4	Methylcrotonoyl-CoA carboxylase	6.4.1.4	O	—
C5	Methylglutaconyl-CoA hydratase	4.2.1.18	—	—
C6	Hydroxymethylglutaryl-CoA lyase	4.1.3.4	O	—
C7	Acetyl-CoA acetoacetyltransferase	2.3.1.9	O	O
Leucine degradation III†				
C1	Branched-chain amino acid aminotransferase	2.6.1.42	O	O
C8	Pyruvate decarboxylase	4.1.1.1	O	O
C9	Alcohol dehydrogenase	1.1.1.1	O	O
Lysine degradation				
Lysine degradation†				
D1	Saccharopine dehydrogenase (L-lysine-forming)	1.5.1.7	O	O
D2	Saccharopine dehydrogenase (L-glutamate-forming)	1.5.1.10	O	O
D3	Amino adipate-semialdehyde dehydrogenase large subunit	1.2.1.31	O	O
D4	2-Amino adipate transaminase	2.6.1.39	NA	NA
D5	2-Oxoglutarate dehydrogenase E1 component	1.2.4.2	O	O
D6	2-Oxoglutarate dehydrogenase E2 component	2.3.1.61	O	O
Glutaryl-CoA degradation†				
D7	Glutaryl-CoA dehydrogenase	1.3.99.7	O	—
D8	Enoyl-CoA hydratase	4.2.1.17	O	—
D9	3-Hydroxyacyl-CoA dehydrogenase	1.1.1.35	O	—
C7	Acetyl-CoA acetoacetyltransferase	2.3.1.9	O	O

*Found in Zygomycetes.

†Based on MetaCyc database.

synthesis, mobilization and degradation of lipids to maintain such high lipid contents in the cells under particular conditions (Beopoulos *et al.*, 2009; Daum *et al.*, 2007). According to the comparative study of the seven genomes, the identified orthologous sequences shared by the oleaginous yeast and fungi, and the core oleaginous genes, were distributed over several processes and aspects of metabolism. Some of them were also found in the non-oleaginous strains, suggesting that not all orthologous oleaginous genes are responsible for oleaginicacy. If the increase of carbon fluxes to fatty acid synthesis depends on substrate availability, as described elsewhere (Oliver *et al.*, 2009), then in particular the additional routes of acetyl-CoA biosynthesis that do not exist in the non-oleaginous microbes might play a role in the extent of lipid storage. Typically, the key metabolite acetyl-CoA is independently synthesized by several metabolic pathways, which are available as reference pathways in the KEGG database. Although there are some reports that *A. oryzae* and *R. oryzae* contain lipids at high levels (Kaur & Worgan, 1982; Meng *et al.*, 2009; Ratledge, 1986; Suzuki, 1991), not many efforts have been made to study the lipid physiology of these fungi, particularly in terms of lipid production. However, from the genome data that are currently available, we identified the missing enzymes that catalyse acetyl-CoA synthesis in the oleaginous strains, which were localized in the cytosol and mitochondria. Apart from ACL, which is a well-known cytosolic enzyme shared among oleaginous micro-organisms (Ratledge, 2002), the orthologous oleaginous enzymes localized in mitochondria might be responsible for generating acetyl-CoA through the degradation of branched-chain amino acids (leucine and lysine) and fatty acids. It is possible that a portion of the acetyl-CoA generated from the mitochondrial pathways is transported into the cytosol for fatty acid synthesis. This transport may be accomplished via specific carnitine-dependent translocation, as reported elsewhere (Strijbis & Distel, 2010). The close relationship between carbohydrate, lipid and amino acid metabolism, particularly in acetyl-CoA synthesis, in the oleaginous microbes, may allow them to tailor the physiology of lipid accumulation to their nutritional regime.

In fatty acid catabolism, diverse β -oxidation pathways have been found in eukaryotes, depending on the characteristics and subcellular locations of the corresponding enzymes. In this work, a complete set of the mitochondrial enzymes required for the four steps of β -oxidation, in which each enzyme contained a monofunctional domain, was observed in the oleaginous strains. Not all of the enzymes exist in the non-oleaginous strains, consistent with the previous report of Shen & Burger (2009). None of the enzymes could be detected in the non-oleaginous yeast *S. cerevisiae*, whereas the genes encoding the enzymes catalysing the third and fourth steps were absent in *C. albicans*. Elsewhere, it has also been found that *Y. lipolytica* possesses an incomplete set of the mitochondrial β -oxidation enzymes acyl-CoA dehydrogenase, 2-enoyl-CoA hydratase and 3-ketoacyl-CoA thiolase

(Beopoulos *et al.*, 2011). However, we computationally identified the missing enzyme for the third step, 3-hydroxyacyl-CoA dehydrogenase, as having a mitochondrial targeting sequence in the oleaginous strains that has not been previously reported. This result indicates that the fatty acid degradation in the oleaginous microbes occurs in the mitochondria, where all β -oxidation enzymes are located, in addition to the peroxisomal system. This mitochondrial β -oxidation seems to be the type I pathway, which is responsible for breaking down short- and medium-chain fatty acids, similar to what has been documented in mammals (Shen & Burger 2009). Furthermore, the Zygomycetes oleaginous fungi had three isoforms with 3-hydroxyacyl-CoA dehydrogenase activity, including fungal-like (FOX2 or MFE), mammalian-like (LCHAD) and bacterial-like (zHAD) enzymes, which were localized in the peroxisomes, mitochondria and an unknown compartment, respectively (Fig. 4). While the sequence of mitochondrial LCHADs of the oleaginous microbes was similar to that of the mammalian enzymes, the zHAD might have evolved from a bacterial ancestor. The coupling of mitochondrial fatty acid oxidation to the citric acid cycle and to oxidative phosphorylation has been reported to be essential for providing energy for the activation of fatty acids (Schulz, 1996). Although the sequence similarity-based prediction provided a good indication of enzyme location, the precise subcellular localization and functional characteristics of the putative zHAD should be further investigated experimentally. The advantage of multiple subcellular localizations of enzymes with the same functional activities for two-carbon shortening in the oleaginous strains has not been well elucidated. This phenomenon might accompany lipid turnover or be involved in the regulation of lipid homeostasis in the high-lipid-accumulating strains. Remarkably, the β -oxidation pathway has been considered as a potential target for enhancing the production of fatty acid-related products (Dulermo & Nicaud, 2011; Picataggio *et al.*, 1992; Waché *et al.*, 2003).

The cellular processes in oleaginous microbes to efficiently metabolize acetyl-CoA through amino acid metabolism appear to be the putative routes of leucine and lysine degradation and of lysine biosynthesis. The alternative route of lysine biosynthesis (α,ϵ -diaminopimelic acid pathway) in the oleaginous microbes, which is acetyl-CoA-independent, may be to overcome insufficient supply of acetyl-CoA for fatty acid synthesis under particular conditions. There is evidence showing a relationship between lipid accumulation and leucine metabolism in *S. cerevisiae* (Kamisaka *et al.*, 2007). It has been found that the expression of the *LEU2* gene or the addition of leucine into the culture medium leads to an increase in the total fatty acid content of the yeast. The additional route of leucine degradation in the oleaginous microbes is probably associated with their physiology of lipid accumulation.

A wide variety of channels for synthesizing the two-carbon substrate were distributed over subcellular compartments

that govern several aspects of the cellular metabolism of oleaginous microbes. In plants, the subcellular organization of acetyl-CoA metabolism for different metabolic purposes has been reported (Oliver *et al.*, 2009). The peroxisomal and mitochondrial acetyl-CoA pools are used for both catabolic and anabolic processes, whereas the pools generated from the cytosol and plastid are responsible for anabolic processes. The unique acetyl-CoA biosynthetic machinery of the oleaginous microbes may affect the extent of lipid accumulation in addition to carbohydrate and energy metabolism. In addition to the significance of the acetyl-CoA supply for fatty acid synthesis, other orthologous oleaginous sequences identified in metabolism, including that for diacylglycerol O-acyltransferase, might be involved in the lipid accumulation process, as previously reported. The regulation of lipid accumulation in oleaginous strains is coupled with acyltransferases to facilitate the incorporation of fatty acids into storage lipids, especially TAG (Courchesne *et al.*, 2009). It has been reported that NADPH pools and energy charges are also essential for fatty acid synthesis and lipid accumulation in oleaginous micro-organisms (Ratledge, 2002). However, the enzymes responsible for NADPH generation in the cells, such as malic enzymes, 6-phosphogluconate dehydrogenase and glucose-6-phosphate dehydrogenase, were not orthologous oleaginous sequences. In fact, not all culture conditions provide the promising phenotype of high lipid content (>20% of dry biomass), even in the same strains. Therefore, the mechanism controlling the enzyme function or the expression of the corresponding genes should not be excluded (Laoteng *et al.*, 2011; Song *et al.*, 2001). Apart from the orthologous enzymes, the existence of multiple isoforms of other enzymes involved in lipid metabolism (e.g. the acyltransferase and lipase families), which have different substrate specificities and subcellular localizations, probably links to the obese phenotype.

In summary, the newly identified orthologous sequences that underlie several aspects of the metabolism of oleaginous microbes shed light on how specific individual genotypes contribute to the different physiological characteristics of lipid production. Comparative genome analysis points to a relationship between carbohydrate, lipid and amino acid metabolism in the biosynthesis of the important intermediate metabolite acetyl-CoA, which may relate to the ability of oleaginous microbes to accumulate high levels of lipids.

ACKNOWLEDGEMENTS

This work was supported by the Higher Education Research Promotion and National Research University Project of Thailand, Office of the Higher Education Commission. The *M. circinelloides* sequence data were produced by the US Department of Energy Joint Genome Institute (<http://www.jgi.doe.gov/>) in collaboration with the user community.

REFERENCES

- Ageitos, J. M., Vallejo, J. A., Veiga-Crespo, P. & Villa, T. G. (2011). Oily yeasts as oleaginous cell factories. *Appl Microbiol Biotechnol* **90**, 1219–1227.
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402.
- Athenstaedt, K., Jolivet, P., Boulard, C., Zivy, M., Negroni, L., Nicaud, J. M. & Chardot, T. (2006). Lipid particle composition of the yeast *Yarrowia lipolytica* depends on the carbon source. *Proteomics* **6**, 1450–1459.
- Beopoulos, A., Mrozova, Z., Thevenieau, F., Le Dall, M. T., Hapala, I., Papanikolaou, S., Chardot, T. & Nicaud, J. M. (2008). Control of lipid accumulation in the yeast *Yarrowia lipolytica*. *Appl Environ Microbiol* **74**, 7779–7789.
- Beopoulos, A., Cescut, J., Haddouche, R., Uribealarea, J. L., Molina-Jouve, C. & Nicaud, J. M. (2009). *Yarrowia lipolytica* as a model for bio-oil production. *Prog Lipid Res* **48**, 375–387.
- Beopoulos, A., Nicaud, J. M. & Gaillardin, C. (2011). An overview of lipid metabolism in yeasts and its impact on biotechnological processes. *Appl Microbiol Biotechnol* **90**, 1193–1206.
- Certik, M., Megova, J. & Horenitzky, R. (1999). Effect of nitrogen sources on the activities of lipogenic enzymes in oleaginous fungus *Cunninghamella echinulata*. *J Gen Appl Microbiol* **45**, 289–293.
- Claros, M. G. & Vincens, P. (1996). Computational method to predict mitochondrially imported proteins and their targeting sequences. *Eur J Biochem* **241**, 779–786.
- Courchesne, N. M. D., Parisien, A., Wang, B. & Lan, C. Q. (2009). Enhancement of lipid production using biochemical, genetic and transcription factor engineering approaches. *J Biotechnol* **141**, 31–41.
- Daum, G., Wagner, A., Czabany, T. & Athenstaedt, K. (2007). Dynamics of neutral lipid storage and mobilization in yeast. *Biochimie* **89**, 243–248.
- Dietrich, F. S., Voegeli, S., Brachat, S., Lerch, A., Gates, K., Steiner, S., Mohr, C., Pöhlmann, R., Luedi, P. & other authors (2004). The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome. *Science* **304**, 304–307.
- Dobson, R. C., Griffin, M. D., Roberts, S. J. & Gerrard, J. A. (2004). Dihydrodipicolinate synthase (DHDDS) from *Escherichia coli* displays partial mixed inhibition with respect to its first substrate, pyruvate. *Biochimie* **86**, 311–315.
- Dobson, P. D., Smallbone, K., Jameson, D., Simeonidis, E., Lanthaler, K., Pir, P., Lu, C., Swainston, N., Dunn, W. B. & other authors (2010). Further developments towards a genome-scale metabolic model of yeast. *BMC Syst Biol* **4**, 145.
- Dulermo, T. & Nicaud, J. M. (2011). Involvement of the G3P shuttle and β -oxidation pathway in the control of TAG synthesis and lipid accumulation in *Yarrowia lipolytica*. *Metab Eng* **13**, 482–491.
- Emanuelsson, O., Nielsen, H., Brunak, S. & von Heijne, G. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol* **300**, 1005–1016.
- Finn, R. D., Mistry, J., Tate, J., Coghill, P., Heger, A., Pollington, J. E., Gavin, O. L., Gunasekaran, P., Ceric, G. & other authors (2010). The Pfam protein families database. *Nucleic Acids Res* **38** (Database issue), D211–D222.
- Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C. & other authors (1996). Life with 6000 genes. *Science* **274**, 546–567.

- Harris, M. A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R., Eilbeck, K., Lewis, S., Marshall, B. & Gene Ontology Consortium (2004). The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res* 32 (Database issue), D258–D261.
- Harris, R. A., Joshi, M., Jeoung, N. H. & Obayashi, M. (2005). Overview of the molecular and biochemical basis of branched-chain amino acid catabolism. *J Nutr* 135 (Suppl.), 1527S–1530S.
- Jones, D. T., Taylor, W. R. & Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* 8, 275–282.
- Jones, T., Federspiel, N. A., Chibana, H., Dungan, J., Kalman, S., Magee, B. B., Newport, G., Thorstenson, Y. R., Agabian, N. & other authors (2004). The diploid genome sequence of *Candida albicans*. *Proc Natl Acad Sci U S A* 101, 7329–7334.
- Kamisaka, Y., Tomita, N., Kimura, K., Kainou, K. & Uemura, H. (2007). *DGA1* (diacylglycerol acyltransferase gene) overexpression and leucine biosynthesis significantly increase lipid accumulation in the $\Delta snf2$ disruptant of *Saccharomyces cerevisiae*. *Biochem J* 408, 61–68.
- Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. & Hattori, M. (2004). The KEGG resource for deciphering the genome. *Nucleic Acids Res* 32 (Database issue), D277–D280.
- Katoh, K. & Toh, H. (2008). Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform* 9, 286–298.
- Kaur, P. & Worgan, J. T. (1982). Lipid production by *Aspergillus oryzae* from starch substrates. *Eur J Appl Microbiol Biotechnol* 16, 126–130.
- Laoteng, K., Čertík, M. & Cheevadhanarak, S. (2011). Mechanisms controlling lipid accumulation and polyunsaturated fatty acid synthesis in oleaginous fungi. *Chem Pap* 65, 97–103.
- Lin, H., Cheng, W., Ding, H. T., Chen, X. J., Zhou, Q. F. & Zhao, Y. H. (2010). Direct microbial conversion of wheat straw into lipid by a cellulolytic fungus of *Aspergillus oryzae* A-4 in solid-state fermentation. *Bioresour Technol* 101, 7556–7562.
- Liu, B. & Zhao, Z. (2007). Biodiesel production by direct methanolysis of oleaginous microbial biomass. *J Chem Technol Biotechnol* 82, 775–780.
- Ma, L. J., Ibrahim, A. S., Skory, C., Grabherr, M. G., Burger, G., Butler, M., Elias, M., Idnurm, A., Lang, B. F. & other authors (2009). Genomic analysis of the basal lineage fungus *Rhizopus oryzae* reveals a whole-genome duplication. *PLoS Genet* 5, e1000549.
- Machida, M., Asai, K., Sano, M., Tanaka, T., Kumagai, T., Terai, G., Kusumoto, K., Arima, T., Akita, O. & other authors (2005). Genome sequencing and analysis of *Aspergillus oryzae*. *Nature* 438, 1157–1161.
- Meng, X., Yang, J., Xu, X., Zhang, L., Nie, Q. & Xian, M. (2009). Biodiesel production from oleaginous microorganisms. *Renew Energy* 34, 1–5.
- Neuberger, G., Maurer-Stroh, S., Eisenhaber, B., Hartig, A. & Eisenhaber, F. (2003). Motif refinement of the peroxisomal targeting signal 1 and evaluation of taxon-specific differences. *J Mol Biol* 328, 567–579.
- Nookaew, I., Jewett, M. C., Meechai, A., Thammarongtham, C., Laoteng, K., Cheevadhanarak, S., Nielsen, J. & Bhumiratana, S. (2008). The genome-scale metabolic model iIN800 of *Saccharomyces cerevisiae* and its validation: a scaffold to query lipid metabolism. *BMC Syst Biol* 2, 71.
- Oliver, D. J., Nikolau, B. J. & Wurtele, E. S. (2009). Acetyl-CoA – life at the metabolic nexus. *Plant Sci* 176, 597–601.
- Picataggio, S., Rohrer, T., Deanda, K., Lanning, D., Reynolds, R., Mielenz, J. & Eirich, L. D. (1992). Metabolic engineering of *Candida tropicalis* for the production of long-chain dicarboxylic acids. *Biotechnology (N Y)* 10, 894–898.
- Ratledge, C. (1986). The potential of microorganisms for oil production – a review of recent publications. In *Proceedings of World Conference on Emerging Technologies in the Fats and Oils Industry*, pp. 318–330. Edited by A. R. Baldwin. Champaign, IL: American Oil Chemists' Society.
- Ratledge, C. (2002). Regulation of lipid accumulation in oleaginous micro-organisms. *Biochem Soc Trans* 30, 1047–1050.
- Ratledge, C. (2004). Fatty acid biosynthesis in microorganisms being used for Single Cell Oil production. *Biochimie* 86, 807–815.
- Ratledge, C. & Wynn, J. P. (2002). The biochemistry and molecular biology of lipid accumulation in oleaginous microorganisms. *Adv Appl Microbiol* 51, 1–51.
- Ruenwai, R., Cheevadhanarak, S. & Laoteng, K. (2009). Overexpression of acetyl-CoA carboxylase gene of *Mucor rouxii* enhanced fatty acid content in *Hansenula polymorpha*. *Mol Biotechnol* 42, 327–332.
- Ruenwai, R., Cheevadhanarak, S., Rachdawong, S., Tanticharoen, M. & Laoteng, K. (2010). Oxygen-induced expression of Δ^6 -, Δ^9 - and Δ^{12} -desaturase genes modulates fatty acid composition and lipid content in *Mucor rouxii*. *Appl Microbiol Biotechnol* 86, 327–334.
- Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4, 406–425.
- Santos-Fo, F., Fill, T. P., Nakamura, J., Monteiro, M. R. & Rodrigues-Fo, E. (2011). Endophytic fungi as a source of biofuel precursors. *J Microbiol Biotechnol* 21, 728–733.
- Schulz, H. (1996). Oxidation of fatty acids. In *Biochemistry of Lipids, Lipoproteins and Membranes*, pp. 75–99. Edited by D. E. Vance & J. E. Vance. Amsterdam: Elsevier Science.
- Shen, Y. Q. & Burger, G. (2009). Plasticity of a key metabolic pathway in fungi. *Funct Integr Genomics* 9, 145–151.
- Sherman, D. J., Martin, T., Nikolski, M., Cayla, C., Souciet, J. L., Durrens, P. & Génolevures Consortium (2009). Génolevures: protein families and synteny among complete hemiascomycetous yeast proteomes and genomes. *Nucleic Acids Res* 37 (Database issue), D550–D554.
- Song, Y., Wynn, J. P., Li, Y., Grantham, D. & Ratledge, C. (2001). A pre-genetic study of the isoforms of malic enzyme associated with lipid accumulation in *Mucor circinelloides*. *Microbiology* 147, 1507–1515.
- Strijbis, K. & Distel, B. (2010). Intracellular acetyl unit transport in fungal carbon metabolism. *Eukaryot Cell* 9, 1809–1815.
- Suzuki, O. (1991). Recent trends of oleochemicals by biotechnology. In *Proceedings of World Conference on Oleochemicals into the 21st Century*, pp. 221–230. Edited by T. H. F. Applewhite. Champaign, IL: American Oil Chemists' Society.
- Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kiryutin, B., Koonin, E. V., Krylov, D. M., Mazumder, R., Mekhedov, S. L. & other authors (2003). The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4, 41.
- Vauterin, M., Frankard, V. & Jacobs, M. (1999). The *Arabidopsis thaliana* *dhdps* gene encoding dihydrodipicolinate synthase, key enzyme of lysine biosynthesis, is expressed in a cell-specific manner. *Plant Mol Biol* 39, 695–708.
- Vicente, G., Bautista, L. F., Gutierrez, F. J., Rodriguez, R., Martinez, V., Rodriguez-Frometa, R. A., Ruiz-Vazquez, R. M., Torres-Martinez, S. & Garre, V. (2010). Direct transformation of fungal biomass from submerged cultures into biodiesel. *Energy Fuels* 24, 3173–3178.
- Waché, Y., Aguedo, M., Nicaud, J. M. & Belin, J. M. (2003). Catabolism of hydroxyacids and biotechnological production of lactones by *Yarrowia lipolytica*. *Appl Microbiol Biotechnol* 61, 393–404.

Wysocki, S. J. & Hähnel, R. (1986). 3-Hydroxy-3-methylglutaryl-coenzyme A lyase deficiency: a review. *J Inherit Metab Dis* **9**, 225–233.

Xu, H., Andi, B., Qian, J., West, A. H. & Cook, P. F. (2006). The α -aminoadipate pathway for lysine biosynthesis in fungi. *Cell Biochem Biophys* **46**, 43–64.

Zhang, Y., Adams, I. P. & Ratledge, C. (2007). Malic enzyme: the controlling activity for lipid production? Overexpression of malic enzyme in *Mucor circinelloides* leads to a 2.5-fold increase in lipid accumulation. *Microbiology* **153**, 2013–2025.

Edited by: J. M. Becker