ORIGINAL RESEARCH PAPER

Annotation and analysis of malic enzyme genes encoding for multiple isoforms in the fungus *Mucor circinelloides* CBS 277.49

Wanwipa Vongsangnak · Yingtong Zhang · Wei Chen · Colin Ratledge · Yuanda Song

Abstract Based on the newly-released genomic data of *Mucor circinelloides* CBS 277.49, we have annotated five genes encoding for malic enzyme: all code for proteins that contain conserved domains/motifs for malic acid binding, NAD^+ binding and $NAD(P)^+$ binding. Phylogenetic analysis for malic enzyme genes showed that genes ID 78524 and 11639 share $\sim 80\%$ amino acid identity and are grouped in cluster 1; genes ID 182779, 186772 and 116127 share $\sim 66\%$ amino acid identity are grouped in cluster 2. Genes ID 78524, 11639 and 166127 produce proteins that are localized in the mitochondrion, while the products from genes 182779 and 186772 are localized in the cytosol. Based on the comparative analysis published previously by Song et al. (Microbiology 147:1507–1515, 2001), we propose that malic enzyme genes ID 78524, 166127, 182779, 186772, 11639, respectively, represent protein isoforms I, II, III/IV, V, and VI.

Keywords Fatty acid biosynthesis · Malic enzyme isoforms · Malic enzyme genes · *Mucor circinelloides* · NADPH

W. Vongsangnak
Center for Systems Biology, Soochow University,
Suzhou 215006, People's Republic of China
e-mail: wanwipa@suda.edu.cn

Y. Zhang · W. Chen · C. Ratledge · Y. Song (✉)
State Key Laboratory of Food Science and Technology,
Jiangnan University, Wuxi 214122, Jiangsu,
People's Republic of China
e-mail: ysong@jiangnan.edu.cn

Y. Zhang
e-mail: yingzhang111@hotmail.com

W. Chen
e-mail: weichen@jiangnan.edu.cn

C. Ratledge
e-mail: c.ratledge@hull.ac.uk

C. Ratledge
Department of Biological Sciences, University of Hull,
Hull HU6 7RX, UK

Introduction

Malic enzyme catalyzes the oxidative decarboxylation of malic acid to pyruvate and simultaneously generates NADH/NADPH for cell metabolism. In oleaginous fungi, such as *Mucor circinelloides* and *Mortierella alpina*, both of which have been or continue to be used for the commercial production of microbial oils, malic enzyme is $NADP^+$-specific and provides NADPH for fatty acid biosynthesis (Wynn et al. 1999). Lipid accumulation is closely linked with the activity of malic enzyme; inhibition of malic enzyme activity leads to the complete abolition of lipid accumulation (Wynn et al. 1997) and, importantly, the extent of lipid accumulation in these two fungi is governed by malic enzyme being the rate-controlling activity. No other NADPH-generating enzyme can replace malic enzyme for this role. In a later study with *Mc. circinelloides* CBS 108.16 (Song et al. 2001), as many as at least six

isoforms of malic enzyme were found, most of which were detected under anaerobic/microaerophilic growth conditions (see Fig. 1). The aerobic isoforms, III and IV, were demonstrated to be linked with lipid accumulation. A similar phenomenon has been found in *Mt. alpina* with at least seven isoforms being present (Zhang and Ratledge 2008). Again, only isoforms D/E in *Mt. alpina* (equivalent to III/IV in *Mc. circelloides*) were associated with lipid biosynthesis. The role of malic enzyme isoforms III/IV and D/E has been confirmed in *Mc. circinelloides* (Zhang et al. 2007): over-expression of the genes coding for isoform III and isoform D both led to a 2.5-fold increase in lipid accumulation but, as in the wild-type organism, both isoforms were converted into their less-active counterparts—isoform IV and isoform E resulting once more in cessation of lipid biosynthesis. Based on the identical *N*-terminal amino acid sequences of isoforms III and IV, it was suggested that these two

isoforms may be encoded by a single gene, and that isoform IV is probably formed by post-translational modification (Song et al. 2001), however the exact mechanism of the conversion of isoforms III/D to isoforms IV/E has yet to be determined.

So far, only the genes coding for isoform II (Li et al. 2005) and III/IV (Zhang et al. 2007) in *Mc. circinelloides* CBS 108.16 have been identified so far. Recently, genomic sequences for *Mc. circinelloides* CBS 277.49 have been released at JGI database (http://genome.jgi-psf.org/Mucci2/Mucci2.download.ftp.html). Although from the released genomic data, five potential genes had been predicted, the biological functions of these genes, and the potential links to those isoforms previously found are not clear. Therefore in this study, we have annotated malic enzyme genes by different bioinformatics tools and can now provide potential links between these annotated genes and the various malic enzyme isoforms. We hereby show for the first
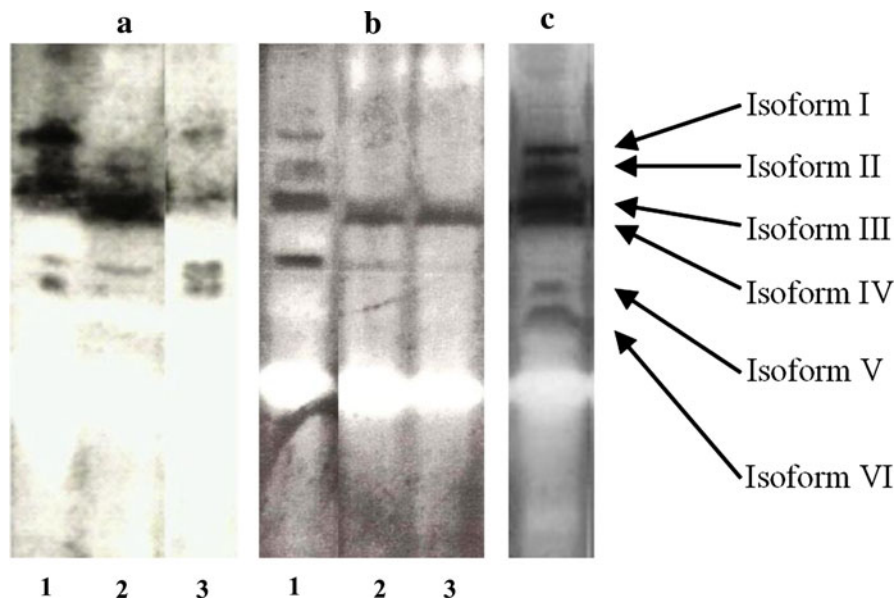

Fig. 1 Multiple isoforms of malic enzyme in *Mc. circinelloides* CBS 108.16. The isoforms were separated by non-denaturing PAGE and visualized by activity staining. *Mc. circinelloides* CBS 108.16 was grown either in 1 l stirred bottles in Kendrick medium (Kendrick and Ratledge 1992) as described previously (Wynn et al. 1997), or in a 5 l fermenter in modified Kendrick medium containing 12 g glucose l^{-1}, 0.5 g ammonium chloride l^{-1} and 0.1 g yeast extract l^{-1}. Continuous culture was carried out in a 5 l chemostat with a working volume of 4 l, aerated at 1.7 l min^{-1}, and stirred at 700 rev min^{-1}. The pH was maintained at 5.5–6.5, and dilution rate was set at 0.04 h^{-1}. Anaerobic growth was achieved by passing pure N_2 through the air inlet. The dissolved oxygen tension (DOT) was measured

using a galvanic oxygen electrode calibrated to 100% of full scale deflection immediately prior to inoculation with the fermenter being flushed with air at 1.7 l min^{-1}. **a** Cell-free extracts were prepared from the cells grown in 1 l stirred bottles at 30°C for 60 h, and under different gaseous environment: (*1*) sealed bottle without aeration (mimicking a microaerophilic condition); (*2*) with aeration; (*3*) CO_2 (99.999% pure) was passed through the medium. **b** Cell-free extracts were prepared from the cells grown in chemostat culture at anaerobic and aerobic conditions: (*1*) DOT = 0% (achieved by passing 99.999% pure N_2 through the medium; (*2*) DOT = 75%; (*3*) DOT = 85%. **c** Cell-free extracts were prepared from mixed cells of a (*1*) and a (*2*), photograph from Song et al. 2001)

time a relationship between the complete genetic background and the six isoforms of malic enzyme in *Mc. circinelloides*.

Methods

Annotation of malic enzyme genes

For annotation of malic enzyme genes, 11,719 protein sequences of *Mc. circinelloides* CBS 277.49 from JGI database (http://genome.jgi-psf.org/Mucci2/Mucci2. download.ftp.html) were explored by comparative sequence alignment analysis against malic enzymes family from Uniprot database (http://www.uniprot. org/) and non-redundant (NR) protein sequences from various organisms deposited in NCBI (www.ncbi.nlm. nih.gov). Function assignment and protein domain analysis were performed using different bioinformatics tools (e.g. BLAST (Altschul et al. 1990), HMMER (Eddy 1998)) and biological databases (i.e. Interproscan (Quevillon et al. 2005), GO (Harris et al. 2004), Pfam (Finn et al. 2008), COGs (Tatusov et al. 2000), TIGRFAMs (Haft et al. 2003), Gene3D (Lees et al. 2010), SUPERFAMILY (Gough and Chothia 2002), PROSITE (Sigrist et al. 2010), and PANTHER (Thomas et al. 2003).

Analysis of phylogenetic relationship of malic enzyme isoforms

Protein sequences of malic enzyme genes were aligned using ClustalW2 multiple alignment tool available at EBI database (http://www.ebi.ac.uk/ Tools/msa/clustalw2/). For aligning protein sequences, the Gonnet protein weight matrix (Gonnet et al. 1992) was set with a gap penalty of 10 for opening and 0.2 for extension to generate the multiple alignments. The ClustalW2-aligned multiple sequences were used to create a phylogenetic tree by the neighbor-joining method (Saitou and Nei 1987) using % identity. Jalview version 2.6.1 (www.jalview.org/) was used for phylogenetic tree visualization.

Prediction of malic enzyme subcellular localization

Localization of individual gene encoding malic enzyme was predicted using the subCELlular LOcalization predictor (CELLO) version 2.5.0 under a two-level Support Vector Machine system (Yu et al. 2006) with features analysis of amino acid composition, *N*-peptide composition, partitioned sequence composition, physico-chemical composition and neighboring sequencing composition.

Calculation of protein theoretical molecular weight and isoelectric point (*p*I)

The theoretical molecular weight and *p*I of malic enzyme isoforms were calculated using "compute *p*I/ Mw tool" from ExPASy proteomics server of Swiss Institute of Bioinformatics (http://expasy.org/tools/).

Results

Analysis of annotated function and protein domain of malic enzyme isoforms

Using the comparative protein sequence analysis and functional assignment, we have identified five putative genes encoding malic enzymes in the genome of *Mc. circinelloides* CBS 277.49. Protein domains and families of these putative five genes ID 182779, 186772, 166127, 78524 and 11639 were further analyzed.

All five genes were conserved in protein domains at different positions within an individual malic enzyme gene. These domains are for malic acid binding, NAD^+ binding, and $NAD(P)^+$ binding. They also share conserved motifs for a malic acid binding site (see Fig. 2). The protein sequences of genes ID 78524 and ID 11639 shared about 80% amino acid identity and have very high sequence homology to known malic enzyme gene of *Mc. circinelloides* CBS 108.16 (*mce*) under assession name of Q875H8_MUCCI (http://www.ncbi.nlm.nih.gov/protein/Q875H8) with amino acid identity of about 94 and 78%, respectively. This is illustrated by phylogenetic tree of malic enzyme genes grouped together in cluster 1 as presented in Fig. 3. The other three malic enzyme genes, ID 182779, 186772 and 166127, indicate protein sequences that share about 66% identity and high sequence homology to known malic enzyme gene of *Mc. circinelloides* CBS 108.16 (*mce2*) under accession name of A6XP72 (Zhang et al. 2007) with amino acid identity of about 90, 71 and 61%, respectively. As expected, these three genes grouped in the same
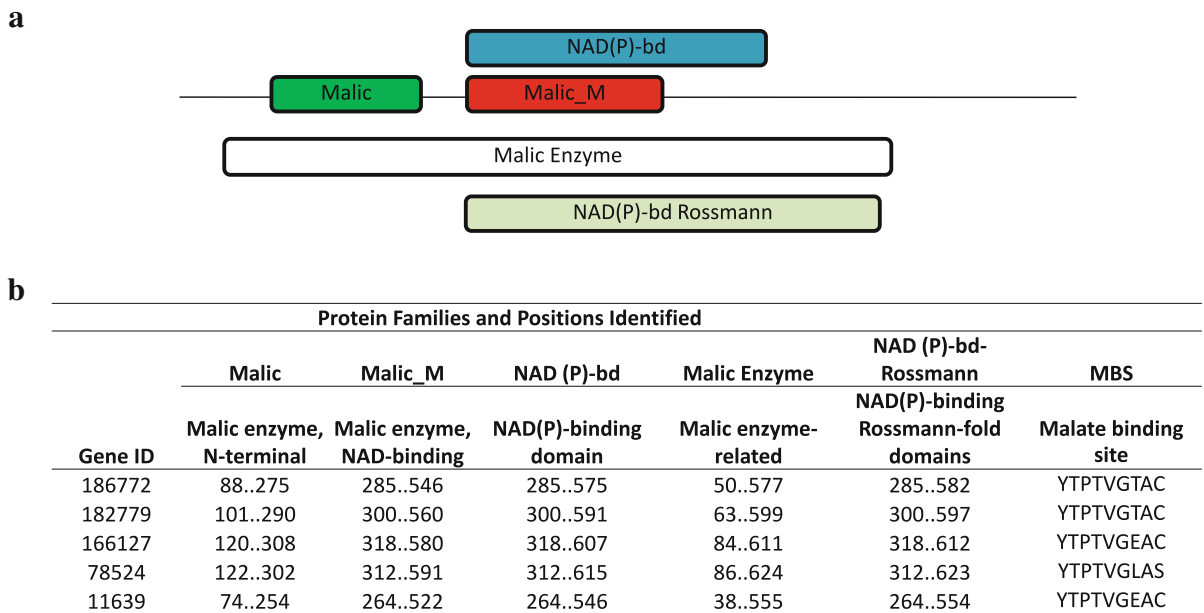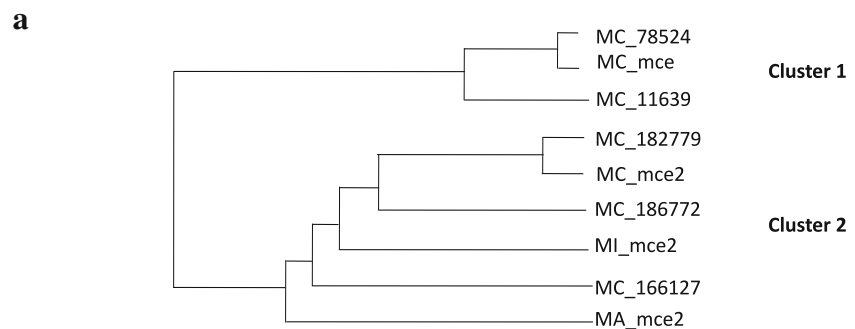
a

b

	Protein Families and Positions Identified					
	Malic	**Malic_M**	**NAD (P)-bd**	**Malic Enzyme**	**NAD (P)-bd-Rossmann**	**MBS**
Gene ID	Malic enzyme, N-terminal	Malic enzyme, NAD-binding	NAD(P)-binding domain	Malic enzyme-related	NAD(P)-binding Rossmann-fold domains	Malate binding site
186772	88..275	285..546	285..575	50..577	285..582	YTPTVGTAC
182779	101..290	300..560	300..591	63..599	300..597	YTPTVGTAC
166127	120..308	318..580	318..607	84..611	318..612	YTPTVGEAC
78524	122..302	312..591	312..615	86..624	312..623	YTPTVGLAS
11639	74..254	264..522	264..546	38..555	264..554	YTPTVGEAC

Fig. 2 Annotation of different genes encoding malic enzyme in *M. circinelloides* CBS 277.49. **a** Overall structure of predicted protein domains of malic enzyme. **b** Protein domain families and their positions identified in different genes encoding malic enzyme

cluster 2 (see Fig. 3). With respect to protein localization, we also can assign the malic enzyme from the different genes to different subcellular localizations. Genes ID 78524, 11639 and 166127 are putatively localized in mitochondria, while 182779 and 186772 are putatively localized in cytosol (see Fig. 3).

Calculation of the molecular weight and *p*I for the malic enzyme isoforms

Using ExPASy proteomics tools (Swiss Institute of Bioinformatics), we calculated the theoretical values for the molecular weights and *p*I values for the five

Fig. 3 Comparative protein sequences analysis of different malic enzyme gene product. **a** Phylogenetic analysis of multiple protein sequences alignments, MC refers to *Mc. circinelloides*, MI refers to *Mt. isabellina*, MA refers to *Mt. alpina*. **b** Assignment of putative gene functions, putative localizations and putative isoforms for malic enzymes

a

b

Gene ID	Putative protein name based homology	Putative protein localization	Putative Isoform based phylogenetic clustering profile
186772	mce2	Cytosol	
182779	mce2	Cytosol	Isoform III and/or Isoform IV
166127	mce2	Mitochondria	
78524	mce	Mitochondria	Isoform II ?
11639	mce	Mitochondria	

isoforms of the identified malic enzyme genes (see Table 1).

Discussion

Mucor circinelloides CBS 108.16 contains at least six isoforms of malic enzyme (Song et al. 2001). Based on their mobility on native PAGE, we designated those protein bands as isoforms I, II, III, IV, V and VI from top to bottom (see Fig. 1c), among which, isoforms I, II and VI were expressed under anaerobic/microaerophilic conditions with VI being expressed only under CO_2 enrichment. Isoforms III and V were expressed under both aerobic and anaerobic conditions but only isoform IV, derived from isoform III following nitrogen limitation from the culture medium, was specific for aerobic conditions (Figs. 1a, b). Genes encoding isoforms II and III/IV were identified (Li et al. 2005; Zhang et al. 2007), but genetic information for other isoforms remained unknown because lack of genomic information of *Mc. circinelloides*.

Now, for the first time, we have identified five malic enzyme genes in *Mc. circinelloides* CBS 277.49. All five gene encoded products contain highly conserved domains for malic acid binding, $NAD(P)^+$ binding and NAD^+ binding (see Fig. 2). The finding of NAD^+-binding domains of these malic enzymes in *Mc. circinelloides* is somehow surprising, as so far no NAD^+-malic enzyme activity has been detected in this fungus. Each of the five genes is highly homologous to known malic enzyme genes in fungi of the same family such as *Mc. circinelloides* CBS 108.16, *Mt. alpina* and *Mt. isabellina,* and is highly homologous to each other within the same cluster. Among the

five genes encoded products, protein ID 182779 shares highest homology with protein ID 186772 with about 69% amino acid identity, and both are localized in cytosol. In addition gene ID 182779 is highly homologous to, and shares 90% of its amino acid identity with, the known malic enzyme gene, *mce2,* in *Mc. circinelloides* CBS 108.16, which encodes isoforms III/IV associated with lipid accumulation. Considering that the *N*-terminal amino acid sequences of all the five genes are different from each other and the fact that isoform III and IV share an identical *N*-terminal amino acid sequence (Song et al. 2001), we propose therefore that gene ID 182779 encodes malic enzyme isoforms III/IV.

Of the three genes identified that encode for mitochondria proteins, genes ID 78524 and 166127 must code for the anaerobic isoforms, I or II, because both protein molecular weight and *p*I values of these two genes are higher than that arising from gene ID 182779 (isoforms III/IV) (see Table 1). In addition, the molecular weight and *p*I of the other mitochondrial gene ID, 11639, are lower than those from gene ID 182779 and, for structurally similar proteins, the molecular weight and *p*I are the key factors for their mobility on native PAGE.

This conclusion is further supported by the mitochondrial localization of the proteins from these two genes and, indeed, a mitochondrial malic enzyme has been reported to be associated with anaerobic growth in *Saccharomyces cerevisiae* (Boles et al. 1998). Between these two genes, the *p*I of the malic enzyme from gene ID 78524 is close to that from gene ID 166127 (0.05 difference) but its molecular weight is significantly larger (see Table 1). Thus, we predict that gene ID 78524 encodes for isoform I and gene ID 166127 encodes for isoform II. The protein sequence of gene ID 78524 shows 94% amino acid identity to known malic enzyme gene, *mce,* in *Mc. circinelloides* CBS 108.16 (Li et al. 2005), which was suggested as coding for isoform II based on its mitochondria localization and its association with anaerobic growth. However, that conclusion is arguable because both I and II are mitochondrial proteins and are expressed only under anaerobic growth conditions and among all the anaerobic isoforms, I, II and VI, only isoform VI is associated with growth under CO_2 enrichment (Song et al. 2001).

The other two genes, ID 186772 and ID 11639, must encode isoforms V or VI as both the molecular

Table 1 Calculation of the molecular weight and *p*I for malic enzyme gene product using ExPASy protein analysis tools, and prediction of the protein isoforms in *Mc. circinelloides* CBS 277.49

Gene ID	Molecular weight (Da)	*p*I	Putative isoforms
78524	70179.82	6.37	I
166127	68549.01	6.43	II
182779	67161.42	5.72	III/IV
186772	64925.65	5.26	V
11639	61110.97	5.53	VI

weight and the pI values of the encoded proteins are smaller than that from gene ID 182779 (isoforms III/IV) (see Table 1). Although the molecular weight of the protein from gene ID 186772 is significantly larger than that from gene ID 11639, the pI of the former is slightly smaller than that of the other, therefore it is debatable whether gene ID 186772 encodes for isoform V and gene ID 11639 encodes for isoform VI, or vice versa. However, among all five genes, gene ID 186772 shares the highest homology with ID 182779 (69% amino acid identity) and the gene product is localized in the cytosol. It is very unlikely, therefore, that this gene would code for anaerobic isoform VI. Besides, gene ID 11639 shares highest homology with ID 78524 coding for an anaerobic malic enzyme that is localized in mitochondria. Therefore, we suggest that gene ID 186772 encodes malic enzyme isoform V and 11639 encode isoform VI. This conclusion agrees with our previous finding and confirmed experiments in this study that isoform V is expressed under both aerobic and anaerobic conditions, while isoform VI is expressed only under CO_2-enriched anaerobic or microaerophilic conditions suggesting that it is a pyruvate-carboxylating enzyme.

Conclusion

Our analysis for the multiple malic enzyme genes in *Mc. circinelloides* CBS 277.49 suggests that genes ID 78524, 166127, 182779, 186772 and 11639 encode malic enzyme isoforms I, II, III/IV, V and VI, respectively. Isoform IV probably arises by post-transcriptional modification of isoform III but all other isoforms are single protein products of individual genes. As it is only isoforms III/IV that are associated with lipid biosynthesis and accumulation, we can now focus attention on the mechanism by which isoform IV is produced knowing that is not the product of a separate gene. This study serves as a basis for malic enzyme isoform characterization of *Mc. circinelloides* and relevant oleaginous fungi, e.g. *Mt. alpina*.

References

Altschul SF, Gish W, Miller W et al (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Boles E, de Jong-Gubbels P, Pronk JT (1998) Identification and characterization of *MAE1*, the *Saccharomyces cerevisiae* structural gene encoding mitochondrial malic enzyme. J Bacteriol 180:2875–2882

Eddy SR (1998) Profile hidden markov models. Bioinformatics 14:755–763

Finn RD, Tate J, Mistry J et al (2008) The pfam protein families database. Nucleic Acids Res 36:D281–D288

Gonnet GH, Cohen MA, Benner SA (1992) Exhaustive matching of the entire protein sequence database. Science 256:1443–1445

Gough J, Chothia C (2002) Superfamily: HMMs representing all proteins of known structure. SCOP sequence searches, alignments and genome assignments. Nucleic Acids Res 30:268–272

Haft DH, Selengut JD, White O (2003) The TIGRFAMs database of protein families. Nucleic Acids Res 31:371–373

Harris MA, Clark J, Ireland A et al (2004) The gene ontology GO database and informatics resource. Nucleic Acids Res 32:D258–D261

Kendrick A, Ratledge C (1992) Lipid formation in the oleaginous mould *Entomophthora exitalis* grown in continuous culture: effects of growth rate, temperature and dissolved oxygen tension on polyunsaturated fatty acids. Appl Microbiol Biotechnol 37:18–22

Lees J, Yeats C, Redfern O et al (2010) Gene3D: merging structure and function for a thousand genomes. Nucleic Acids Res 38:D296–D300

Li Y, Adams IP, Wynn JP et al (2005) Cloning and characterization of gene encoding a malic enzyme involved in anaerobic growth in *Mucor circinelloides*. Mycol Res 109:461–468

Quevillon E, Silventoinen V, Pillai S et al (2005) InterPro Scan: protein domains identifier. Nucleic Acids Res 33:W116–W120

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Sigrist CJ, Cerutti L, de Castro E et al (2010) PROSITE, a protein domain database for functional characterization and annotation. Nucleic Acids Res 38:D161–D166

Song Y, Wynn JP, Li Y et al (2001) A pre-genetic study of the isoforms of malic enzyme associated with lipid accumulation in *Mucor circinelloides*. Microbiology 147:1507–1515

Tatusov RL, Galperin MY, Natale DA et al (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res 28:33–36

Thomas PD, Kejariwal A, Campbell MJ et al (2003) PANTHER: a browsable database of gene products organized

by biological function, using curated protein family and subfamily classification. Nucleic Acids Res 31:334–341

Wynn JP, Hamid AA, Ratledge C (1997) Sesamol as an inhibitor of growth and lipid metabolism in *Mucor circinelloides* via its action on malic enzyme. Lipids 32:605–610

Wynn JP, Kendrick A, Ratledge C (1999) The role of malic enzyme in the regulation of lipid accumulation in filamentous. Microbiology 145:1911–1917

Yu CS, Chen YC, Lu CH et al (2006) Prediction of protein subcellular localization. Proteins 64:643–651

Zhang Y, Ratledge C (2008) Multiple isoforms of malic enzyme in the oleaginous fungus, *Mortierella alpina*. Mycol Res 112:725–730

Zhang Y, Adams IP, 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