Genome mining of fungal lipid-degrading enzymes for industrial applications

Tayvich Vorapreeda,¹ Chinea Thammarongtham,¹ Supapon Cheevadhanarak²,³ and Kobkul Laoteng⁴

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

¹Biochemical Engineering and Pilot Plant Research and Development Laboratory, National Center for Genetic Engineering and Biotechnology (BIOTEC) 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
⁴Bioprocess Technology Laboratory, Bioresources Technology Unit, National Center for Genetic Engineering and Biotechnology (BIOTEC), National Science and Technology Development Agency (NSTDA), Thailand Science Park, Pathum Thani 12120, Thailand

Lipases are interesting enzymes, which contribute important roles in maintaining lipid homeostasis and cellular metabolisms. Using available genome data, seven lipase families of oleaginous and non-oleaginous yeast and fungi were categorized based on the similarity of their amino acid sequences and conserved structural domains. Of them, triacylglycerol lipase (patatin-domain-containing protein) and steryl ester hydrolase (abhydro_lipase-domain-containing protein) families were ubiquitous enzymes found in all species studied. The two essential lipases rendered signature characteristics of integral membrane proteins that might be targeted to lipid monolayer particles. At least one of the extracellular lipase families existed in each species of yeast and fungi. We found that the diversity of lipase families and the number of genes in individual families of oleaginous strains were greater than those identified in non-oleaginous species, which might play a role in nutrient acquisition from surrounding hydrophobic substrates and attribute to their obese phenotype. The gene/enzyme catalogue and relevant informative data of the lipases provided by this study are not only valuable toolboxes for investigation of the biological role of these lipases, but also convey potential in various industrial applications.

INTRODUCTION

Lipids are major components of cell and organelle membranes. Moreover, they are also stored as metabolic fuels in particular forms, such as lipid particles and adipose tissue. Dissimilar to proteins, lipid molecules do not contain discrete structural domains that enable their sorting to target compartments. Indeed, the transport machinery and other elements, particularly lipid-degrading enzymes, must exist in the cells for their dynamic features. Lipases are biocatalysts, and are water-soluble enzymes responsible for the hydrolysis of the ester bonds of water-insoluble or hydrophobic substrates. Eukaryotic cells are able to produce several lipases with discrimination in their localizations (extracellular, membrane-bound and intracellular activities) and properties. Enzymes catalysing the hydrolysis of triacylglycerol and steryl ester are known as triacylglycerol lipase and steryl ester lipase, respectively (Müllner & Daum, 2004; Kohlwein, 2010). Phospholipases are also important in diverse biological processes, such as membrane homeostasis, nutrient acquisition and generation of signal molecules (Köhler et al., 2006). Extracellular lipases are recognized as digestive enzymes for hydrolysis of hydrophobic substrates into small molecules (i.e. fatty acids), which can be transported into the cells as an energy or nutrient source (Fickers et al., 2005b; Grillitsch & Daum, 2011). Indeed, lipases play key roles not only in lipid degradation, but are also indirectly involved in lipid uptake and trafficking. Furthermore, the lipolytic enzymes contribute to cell homeostasis through lipid turnover and signal transduction processes (Müllner & Daum, 2004; Fickers et al., 2005a; Athenstaedt & Daum, 2006; Kohlwein, 2010; Kohlwein et al., 2013).

Abbreviation: pI, isoelectric point.

Three supplementary tables and one supplementary figure are available with the online Supplementary Material.

Received 28 January 2015
Accepted 18 June 2015
From an industrial viewpoint, microbial lipases have gained enormous attention due to their biotechnological potential for catalysing a number of useful reactions, including esterification, transesterification, chemoselective, regioselective and stereoselective reactions, and other functions (Mustranta et al., 1993; Anderson et al., 1998; Schmid & Verger, 1998; Xu et al., 2002; Hasan et al., 2006; Taniguchi et al., 2006; Vakhru & Kour, 2006; Fickers et al., 2011). Therefore, they are commonly used in processes in biotechnological industries, including food (De Maria et al., 2007; Rajan et al., 2008; Keng et al., 2009), cosmetics and personal care products, oleochemicals (Xia et al., 1996; Lai et al., 2005; Saisubramanian et al., 2006; Goswami et al., 2013) and biomedical industries as well as bioremediation (Lanciotti et al., 2005; Cammarota & Freire, 2006) and biosensor applications (Albertsson & Varma, 2003; Kartel et al., 2007; Sandoval et al. 2010). Nowadays, yeast and fungi are very attractive lipase-producing systems for industrial applications due to the versatility of their properties and ease of mass production (Fickers et al., 2011; Singh & Mukhopadhyay, 2012). Major species of yeast and fungi of commercial importance for lipase production are recognized, such as Candida sp., Penicillium sp., Rhizopus sp., Aspergillus sp., Mucor sp. and Yarrowia lipolytica (Schmid & Verger, 1998; Treichel et al., 2010; Singh & Mukhopadhyay, 2012). In addition to microbial species, the production yield and titre of lipases rely on media composition, and bioprocess and operation variables (Marek & Bednarski, 1996; Fickers et al., 2004). There is considerable interest in industrial production of lipases with desired properties (i.e. stability and co-factor independence) and substrate specificities for wider applications. Isolation and screening of microorganisms with such potentials is one strategy used to meet the target. With the advances in omics technologies and available genome databases, a more rational approach could be adopted not only for studying the genome-wide function of lipases across organisms, but also for accelerating the production development of commercially important lipases. In this study, we investigated the diversity of lipases in 11 yeast and fungal strains using bioinformatics tools, which is an empowering approach to identify enzyme targets with commercial interest by means of genome mining. Six genomes of oleaginous species, Yarrowia lipolytica, Rhodosporidium toruloides, Aspergillus oryzae, Mortierella alpina, Mucor circinelloides and Rhizopus oryzae, were also included. These oleaginous microorganisms are industrially important due to their lipid phenotypes, accumulating substantial amounts of lipids (more than 20% of biomass) (Ageitos et al., 2011; Meng et al., 2009; Ratledge, 2004). Some of them are promising strains for production of essential polyunsaturated fatty acids (Certi & Shimizu, 1999; Zhu et al., 2002; Sakuradani, 2010; Laoteng et al., 2011). A gene/enzyme catalogue and relevant informative data of the lipases provided by this study are also valuable toolboxes for further investigating the biological roles of the lipases and for searching for promising enzymes with specific purpose.

**METHODS**

Combinatorial approaches based on amino acid sequence similarity, conserved structural domains and data mining of associated biochemical function through available databases (Arpigny & Jaeger, 1999) were used to identify and categorize lipid-degrading enzymes in oleaginous and non-oleaginous yeast and fungi.

**Gene retrieval and identification of putative lipases.** An initial set of lipogenic gene sequences was derived from public databases, including the National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov/), DOE Joint Genome Institute (JGI; http://www.jgi.doe.gov/), and Fungal Genome Initiative (FGI; http://www.broadinstitute.org/), and was used as a query protein set (see Table S1, available in the online Supplementary Material). Together, the relevant biochemical information supporting the gene data, particularly in lipid catabolism, was provided in protein annotation. Subsequently, each protein in the query dataset was used to identify the homologous proteins in yeast and fungi by subjecting it to BLASTP search (Altschul et al., 1997) with cut-off E-value inclusion using a threshold set to $10^{-10}$ against targeted genomes. This E-value was used to filter out distantly related sequences that might be not genuine homologous sequences. For analysing the lipase protein sequences in classified families, BLAST was performed using an E-value of $10^{-20}$. Among the 11 target genomes of yeast and fungi, eight of them were partially or completely annotated and provided by the JGI or JGI-IMG ER databases (Markowitz et al., 2009), these included Saccharomyces cerevisiae s328c, Schizosaccharomyces pombe 972h, Ashbya gossypii ATCC 10895, Pichia pastoris GS 115, Hansenula polymorpha NCYC 495 leu1.1 v2.0, Yarrowia lipolytica CLIB122, Aspergillus oryzae RIB40 and Mucor circinelloides CBS2774.9 v2.0. The genomes of Mortierella alpina ATCC 32222 and Rhodosporidium toruloides MTCC 457 were obtained from NCBI, and the Rhizopus oryzae RA99-880 genome was obtained from the BROAD Institute (http://www.broadinstitute.org/scientific-community/data). To confirm the relevant functions of homologous proteins in lipid catabolism, their sequences were classified according to the information in the RefSeq (Pruitt et al., 2007) and Pfam protein family databases (Punta et al., 2012).

**Identification of conserved domains and motifs.** The homologous proteins were classified into particular lipase families based on their conserved domains. 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/) (Punta et al., 2012). The functional regions of query proteins were searched against the domains Pfam-A and Pfam-B using a default E-value cut-off of 1.0, which was defined based on expert knowledge, sequence similarity, other protein family databases, and the ability of HMM profiles to correctly identify and align the members.

Conservative sequence motifs of individual lipase families were extracted from an alignment using the multi-sequence alignment program MAFFT (http://mafft.cbrc.jp/alignment/software/) (Katoh & Standley, 2013) under the FFT-NS-i algorithm (iterative refinement method). The process was repeated automatically for a maximum of 1000 iterations until no further improvement in alignment scoring was achieved.

**Subcellular localization analysis.** Protein localization was analyzed using the support vector machine system implemented in the program CELLO version 2.5 (http://cello.life.nctu.edu.tw/) (Yu et al., 2006). The standard statistical values were used for CELLO algorithm training through multiple feature vectors based on multiple n-peptide composition, which provides the accuracy of predictive results.

**Other phenotypic characterization.** Each predicted lipase protein was submitted to the Compute pi/MW tool at the ExPASy server.
(http://web.expasy.org/compute_pi/) (Gasteiger et al., 2003) for computing protein isoelectric point (pI) and molecular mass properties on the basis of the deduced amino acid sequences.

Philius, a dynamic Bayesian networks based approach available at the Yeast Resource Center (http://www.yeastrc.org/philius/), which is a probabilistic framework of Philius deriving informative confidence for differentiating between transmembrane and globular proteins (Reynolds et al., 2008), was used to analyse transmembrane topology and signal peptides of individual lipase proteins. According to the transmembrane topology analysis, if the proteins have no significance in transmembrane helix features, the proteins are considered as globular water-soluble proteins.

RESULTS

Cataloguing of lipase enzymes in yeast and fungi

A set of yeast and fungal genome sequences was employed to generate a catalogue of lipase enzymes. Based on homology sequence analysis with the cut-off E-value of 10^{−10}, the result showed that there was a differentiation in terms of number of genes and families of lipases among the 11 genomes. Although the E-value of BLAST for identifying putative lipase sequences was set at 10^{−16}, it was found that the members in individual classification groups shared strong sequence similarity to each other within the same group at a higher stringency of E-value (10^{−20}). Using a combination of their amino acid sequence similarity and informative publications relevant to known protein families with specified biochemical function, the lipases were categorized into seven families according to their conserved structural domains and functionality (Table 1). The last family, putative hormone-sensitive like lipase, had not been reported previously in fungi, which might be due to lack of experimental studies. The distribution of coding genes in individual lipase families is presented in Table 2 (a list of lipase genes is shown in Table S2). Among the genomes studied, only two species, Y. lipolytica and Muc. circinelloides, contained all lipase families. Considering gene number, the largest number of lipases (50 coding genes) was found in Asp. oryzae, which was distributed into six families. Of them, the carboxylesterase family (EC 3.1.1.1) was the largest group (52 % of total lipase genes in Asp. oryzae), which is involved in hydrolysis of compounds containing ester bonds (Converti et al., 2002; Molinari et al., 2000). Conversely, the smallest set of lipases was detected in Rho. toruloides.

Noticeably, two families of lipases, which contained the abhydro_lipase domain (PF04083) and patatin domain (PF01734), were found in all 11 yeast and fungal genomes. The steryl ester hydrolase enzymes having an abhydro_lipase domain are involved in steryl ester degradation (Köffel et al., 2005; Müllner et al., 2005; Köffel & Schneider, 2006), whereas the patatin-domain-containing lipases (triacylglycerol lipases) play essential roles in breakdown of triglyceride (Kienesberger et al., 2009; Rajakumari & Daum, 2010; Yazawa et al., 2012). As a result, these two major groups of lipases seem to be ubiquitous enzymes, which might be essential for yeast and fungal cells to either utilize or mobilize neutral lipids.

Based on physicochemical properties, the PGAP1-domain- and abhydro_lipase_3-domain-containing proteins, which were predicted as soluble intracellular lipases, were predicted as mitochondrial and cytosolic enzymes, respectively. Our results indicated that these two lipases were not ubiquitous enzymes. The acylglycerol lipase (PGAP1) family was not found in Sch. pombe or Asp. oryzae genomes, whereas the enzyme family containing the abhydro_lipase_3 domain was absent in Sac. cerevisiae, Ash. gossypii and Rho. toruloides. Possibly, the enzymic function of the soluble intracellular lipases may be compensated by the membrane-bound lipases, such as triacylglycerol lipase (patatin) and steryl ester hydrolase (abhydro_lipase) family proteins.

Extracellular lipases were classified into three families, which contained lipase-class3, PLA2_B and COesterase domains that are specific for triacylglycerols, phospholipids and carboxyl esters, respectively. Our results showed that each family of extracellular lipases was not distributed among all yeast and fungi studied in contrast to the ubiquitous enzymes identified, containing abhydro_lipase and patatin domains. However, at least one family of extracellular lipases was found in each individual genome, suggesting that the acquisition of lipid metabolic capability is essential for oleaginous and non-oleaginous yeast and fungal strains. For example, only one family, lipase-class3-domain-containing protein, was found in Ash. gossypii, in which it has been proposed to be involved in nutrition acquisition (Stahmann et al., 1994). Notably, only three oleaginous strains, including Y. lipolytica, Asp. oryzae and Muc. circinelloides, contained all extracellular lipase families. Mostly, the number of extracellular lipase genes of oleaginous species was larger than that of non-oleaginous species, which might serve to facilitate the digestion of hydrophobic compounds to more simple forms that can be more easily adsorbed, transported and accumulated into the cells.

Ubiquitous lipases: steryl ester hydrolase family (abhydro_lipase-domain-containing protein)

The results of sequence similarity and evolutionarily conserved units of proteins showed that the steryl ester hydrolase family in yeast and fungi was also classified into two groups, TGL1-like and YEH1/YEH2-like proteins. The TGL1-like proteins contained two alpha/beta-hydrolase-associated lipase regions, including abhydro_lipase (PF04083) and abhydro_lipase_1 (PF00561) domains, whereas YEH1/YEH2-like proteins contained only an abhydro_lipase domain, according to the pfam database (Fig. 1 and Table S3). Based on genome sequence analysis, 24 steryl ester hydrolytic enzymes were found among 11 species studied. Of them, the group of TGL1-like proteins (15 sequences; 62.5 % of the total abhydro_lipase sequences) was slightly larger than the YEH1/YEH2-like protein group except in Rho. toruloides (Table 1, Fig. 1). Obviously,
<table>
<thead>
<tr>
<th>Family</th>
<th>Description</th>
<th>Consensus domain</th>
<th>(G/A)xSxG motif</th>
<th>Predicted EC</th>
<th>Localization</th>
<th>Members</th>
<th>Length (aa)</th>
<th>Mol. mass (KDa)</th>
<th>pI range</th>
</tr>
</thead>
<tbody>
<tr>
<td>F-I</td>
<td>Steryl ester hydrolase</td>
<td>PF04083 Abhydro_lipase</td>
<td>Most (21/24)</td>
<td>3.1.1.13</td>
<td>Plasma membrane (96 %) Mitochondria (4 %)</td>
<td>24</td>
<td>371–787</td>
<td>43–87</td>
<td>5.39–8.91</td>
</tr>
<tr>
<td>F-II</td>
<td>Triacylglycerol lipase</td>
<td>PF01734 Patatin PF11815 DUF3336</td>
<td>Most (21/28)</td>
<td>3.1.1.3</td>
<td>Plasma membrane (79 %) Nuclear (14 %) Mitochondria (7 %)</td>
<td>28</td>
<td>375–910</td>
<td>42–102</td>
<td>5.65–9.62</td>
</tr>
<tr>
<td>F-III</td>
<td>Acylglycerol lipase</td>
<td>PF07819 PGAP1</td>
<td>All</td>
<td>3.1.1.23</td>
<td>Mitochondria (64 %) Cytoplasm (18 %) Plasma membrane (9 %) Extracellular (9 %)</td>
<td>11</td>
<td>232–612</td>
<td>26–66</td>
<td>5.57–9.00</td>
</tr>
<tr>
<td>F-IV</td>
<td>Extracellular TGL lipase</td>
<td>PF01764 Lipase class3</td>
<td>All</td>
<td>3.1.1.3</td>
<td>Extracellular (82 %) Plasma membrane (14 %) Cytoplasm (2 %) Mitochondria (2 %) Extracellular (81 %) Cytoplasm (19 %)</td>
<td>43</td>
<td>233–555</td>
<td>26–61</td>
<td>5.08–9.03</td>
</tr>
<tr>
<td>F-V</td>
<td>Phospholipase</td>
<td>PF01735 PLA2_B</td>
<td>All</td>
<td>3.1.1.5</td>
<td>Extracellular (81 %) Cytoplasm (19 %)</td>
<td>21</td>
<td>399–850</td>
<td>44–94</td>
<td>4.10–6.83</td>
</tr>
<tr>
<td>F-VI</td>
<td>Carboxylesterase</td>
<td>PF00135 COesterase</td>
<td>All</td>
<td>3.1.1.1</td>
<td>Extracellular (58 %) Cytoplasm (32 %) Plasma membrane (5 %) Others (5 %)</td>
<td>40</td>
<td>140–736</td>
<td>15–80</td>
<td>4.63–8.25</td>
</tr>
<tr>
<td>F-VII</td>
<td>Putative hormone-sensitive-like lipase</td>
<td>PF07859 Abhydrolase_3</td>
<td>All</td>
<td>3.1.1.-</td>
<td>Nuclear (8 %)</td>
<td>Mitochondrial (24 %) Plasma membrane (22 %)</td>
<td>37</td>
<td>265–899</td>
<td>29–96</td>
</tr>
</tbody>
</table>
Table 2. Distribution of lipid-degrading enzymes in yeast and fungal genomes based on lipase gene families

<table>
<thead>
<tr>
<th>Species</th>
<th>Lipase family</th>
<th>F-I</th>
<th>F-II</th>
<th>F-III</th>
<th>F-IV</th>
<th>F-V</th>
<th>F-VI</th>
<th>F-VII</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sac. cerevisiae</td>
<td></td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Sch. pombe</td>
<td></td>
<td>3</td>
<td>3</td>
<td>–</td>
<td>–</td>
<td>5</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>H. polymorpha</td>
<td></td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>–</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Pic. pastoris</td>
<td></td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>–</td>
<td>–</td>
<td>1</td>
</tr>
<tr>
<td>Ash. gossypii</td>
<td></td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Y. lipolytica</td>
<td></td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>16</td>
<td>4</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Rho. toruloides</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>–</td>
<td>1</td>
<td>2</td>
<td>–</td>
</tr>
<tr>
<td>Asp. oryzae</td>
<td></td>
<td>2</td>
<td>3</td>
<td>–</td>
<td>7</td>
<td>4</td>
<td>26</td>
<td>8</td>
</tr>
<tr>
<td>Mor. alpina</td>
<td></td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>–</td>
<td>–</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>Muc. circinelloides</td>
<td></td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td>Rhi. oryzae</td>
<td></td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>6</td>
<td>1</td>
<td>–</td>
<td>9</td>
</tr>
</tbody>
</table>

the YEHI/YEHE2-like protein group was absent in the Sch. pombe, Mor. alpina and Muc. circinelloides genomes. We also found the putative key consensus sequences, GxSxG or serine-centre motif and GxxR motif, shared among steryl ester hydrolytic enzymes of the yeast and fungi (Fig. 2). The GxSxG motif has been proposed to be the residues involved in interfacial binding of enzyme to substrate, which is located near the centre of the protein (Upton et al., 1995; Wong & Schotz, 2002). It has been reported that another consensus motif (GxxR) may play a role in substrate recognition by contacting the sugar or methyl moieties that another consensus motif (GxxR) may play a role in substrate recognition by contacting the sugar or methyl moieties that another consensus motif (GxxR) may play a role in substrate recognition by contacting the sugar or methyl moieties that another consensus motif (GxxR) may play a role in substrate recognition by contacting the sugar or methyl moieties that another consensus motif (GxxR) may play a role in substrate recognition by contacting the sugar or methyl moieties that another consensus motif (GxxR) may play a role in substrate recognition by contacting the sugar or methyl moieties that another consensus motif (GxxR) may play a role in substrate recognition by contacting the sugar or methyl moieties that another consensus motif (GxxR) may play a role in substrate recognition by contacting the sugar or methyl moieties that another consensus motif (GxxR) may play a role in substrate recognition by contacting the sugar or methyl moieties that another consensus motif (GxxR) may play a role in substrate recognition by contacting the sugar or methyl moieties that another consensus motif (GxxR) may play a role in substrate recognition by contacting the sugar or methyl moieties that another consensus motif (GxxR) may play a role in substrate recognition by contacting the sugar or methyl moieties that another consensus motif (GxxR) may play a role in substrate recognition by contacting the sugar or methyl moieties.

The transmembrane topology analysis showed that most of the sequences of the steryl ester hydrolase (abhydro_lipase) family contained transmembrane domains at the N terminus, indicating that they render a characteristic of membrane proteins. However, the subcellular localization predicted by CELLO showed that these steryl ester hydrolytic enzymes were located at the plasma membrane with the exception of Sac. cerevisiae YLC020C, which was predicted as a mitochondrial protein. There was no difference in the enzyme topology and localization between the TGL1-like and YEHI/YEHE2-like proteins. Taken together, it can be concluded that the steryl ester hydrolytic enzymes in these yeast and fungi were membrane-associated proteins. It has been proposed that a monolayer of lipid particles is a preferred target for the membrane-anchoring region of steryl ester hydrolase, rather than a common bilayer membrane (Köffel et al., 2005).

Ubiquitous lipases: triacylglycerol lipase family (patatin-domain-containing protein)

For another ubiquitous lipase family, the triacylglycerol lipases of the yeast and fungi contained two consensus domains, including an unknown consensus domain (DUF3336, PF11815) and a patatin or patatin-like phospholipase (PF01734) domain, which were classified according to the Pfam database. The protein sequences identified, containing a patatin domain, which had high similarity to the known triacylglycerol lipases, were thus classified as triacylglycerol lipase (patatin) family, in which 28 sequences were distributed among 11 genomes of yeast and fungi. We found that these proteins contained four consensus regions, including GTK, GxSxG and two unidentified motifs, by multiple sequence alignment as shown in Fig. 3. The GTK motif, which is an uncharacterized motif, was located near the N-terminus of triacylglycerol lipases. For the catalytic motif (GxSxG), the serine residue of several triacylglycerol lipases was substituted with other amino acids, such as alanine (A), glycine (G) and lysine (K). It has been reported that the amino acid substitution found in Sch. pombe SPCC1450.16c (Ptl1p) does not affect triacylglycerol lipase activity (Yazawa et al., 2012). As such, the central serine residue in the GxSxG motif might not be required for enzyme activities and protein topology. Besides, the analysis of consensus regions revealed that many triacylglycerol lipases contained unknown motifs (GxTxFxEAY and LLNxtTAP) as shown in Fig. 3. These consensus residues of unknown motifs at specific positions were under strong selective pressure and, therefore, might be a functionally relevant signature of triacylglycerol lipases.

The pl of triacylglycerol lipases in non-oleaginous strains was in the range of 7.19–9.53, whereas those of oleaginous strains exhibited a wider range (pI=5.65–9.62) (Fig. S1) due to finding of homologous sequences of acidic triacylglycerol lipases (pI<7). The localization analysis using CELLO showed that most of the putative proteins, containing a patatin-like phospholipase domain, were located in the plasma membrane. This finding coincides with experimental studies indicating that they are required for mobilization of triacylglycerol from plasma membranes to store as lipid particles (Athenstaedt et al., 1999; Athenstaedt & Daum, 2005). In addition, the results of protein topology analysis indicated that most of the triacylglycerol lipases in yeast and fungi, particularly patatin-domain-containing
Fig. 1. Sequence motif and domain architecture in lipase families of yeasts and fungi. SP and TM represent signal peptide and transmembrane regions, respectively.

F-I. Steryl ester hydrolase (abhydro_lipase family)

- YALI0E32035g: TM G×S×G, TGL1-like
- YALI0E00528g: TM G×S×G, YEH1/YEH2-like

F-II. Triacylglycerol lipase (patatin family)

- YALI0F10010g: TM G×S×G, PF11815 PF01734

F-III. Acylglycerol lipase (PGAP1 family)

- YALI0E31515g: G×S×G, PF07819

F-IV. Extracellular triacylglycerol lipase (lipase class3 family)

- YALI0A10439g: SP G×S×G, PF01764

F-V. Phospholipase (PLA2_B family)

- YALI0E16060g: SP G×S×G, PF01735

F-VI. Carboxylesterase (COesterase family)

- YALI0B08030g: G×S×G, Intracellular protein
- Rhodo_protein4903: SP G×S×G, Extracellular protein

F-VII. Putative hormone-sensitive-like lipase (abhydrolase_3 family)

- YALI0E17655g: G×S×G, PF07859
lipases, contained an N-terminal membrane-spanning region that is a common feature of this protein group.

Other lipase families with different substrate utilization
Apart from the ubiquitous lipases, we found sequences coding for several lipase families in yeast and fungi which had a differentiation in compartmentalization and specificity on lipid substrates, including acylglycerol lipase, extracellular triacylglycerol lipase, phospholipase, carboxylesterase and putative hormone-sensitive-like lipase. All of them contained the serine-centre motif (GxSxG).

**Acylglycerol lipase family (PGAP1-domain-containing protein)** Using the \textit{Sac. cerevisiae} Tgl2 protein sequence coding for PGAP1-domain-containing acylglycerol lipase as a query, 11 homologous proteins were found in the yeast and fungal genomes studied except for those of \textit{Sch. pombe} and \textit{Asp. oryzae}. This result suggests

**Fig. 2.** Multiple alignments of amino acid sequences of steryl ester hydrolases. Alignment was made using MAFFT software. Arrowheads indicate the amino acid residues constituting the consensus motifs. Number in brackets indicate number of amino acid residues. Black and grey letters represent identical and similar residues, respectively.

http://mic.sgmjournals.org 1619
that the TGL2-like protein might not be essential in yeast and fungi, which is in agreement with previous studies describing the growth ability of the deletion strain of the TGL2 gene on minimal medium (Van Heusden et al., 1998; Ham et al., 2010). However, it has been found that the TGL2 protein is required for the viability of cells treated with antimitotic drugs (Ham et al., 2010). In addition to the serine-centre motif, the acylglycerol lipase family (PGAP1-domain-containing proteins) shared an unidentified motif (SLLTxxTPH) (Fig. 4). The localization analysis by CELLO showed four major groups of the putative lipases, including mitochondrial (64%), cytoplasmic (18%), plasma membrane (9%) and extracellular (9%) proteins. The above result is consistent with the experimental observation that lipolytic activity of this family occurs in the mitochondria (Ham et al., 2010).

### Extracellular triacylglycerol lipase family (lipase-class3-domain-containing protein)
As shown in Table 1, the sequence analysis showed that there were 43 homologous genes coding for putative extracellular triacylglycerol lipases (lipase class3) in the yeast and fungal genomes, particularly in oleaginous strains (34 proteins; 79% of lipase class3). However, these genes did not exist in *S. cerevisiae*.

---

**Fig. 3.** Multiple alignments of amino acid sequences of triacylglycerol lipases. Alignment was made using MAFFT software. Arrows indicate the amino acid residues constituting the consensus motifs. Numbers in brackets indicate number of amino acid residues. Black and grey letters represent identical and similar residues, respectively.

**Fig. 4.** Multiple alignments of amino acid sequences of acylglycerol lipase. Alignment was made using MAFFT software. Arrows indicate the amino acid residues constituting the consensus motifs. Black and grey letters represent identical and similar positions, respectively.
pombe, Rho. toruloides or Mor. alpina. As for other lipases, the multiple sequence alignment of the lipase class3 of yeast and fungi showed the serine-centre motif (GxSxG), which was located near the centre of all aligned sequences, which is a signature characteristic of lipases (Wong & Schotz, 2002; Fickers et al., 2005a). In addition, these lipase-class3-domain-containing proteins were presumably located to the extracellular or plasma membrane with the exception of Sac. cerevisiae YJR107W, which was predicted as a cytoplasmic protein. From the results of the silico analysis, about 20 amino acid residues of putative signal sequences were detected at the N-termini of extracellular triacylglycerol lipases (lipase class3 family). Previous studies have reported that the putative signal sequence was not cleaved and might serve as an anchor domain targeting to the cell wall (Fickers et al., 2005a; Liu et al., 2010).

Phospholipase family (PLA2_B-domain-containing protein) According to homology sequence analysis, we found that 21 protein sequences coding for phospholipase enzymes were distributed among eight genomes of yeast and fungi, in which several species had more than one phospholipase enzyme (Tables 2 and S3). Homologous proteins did not exist in H. polymorpha, Ash. gossypii and Mor. alpina. The PLA2_B domain (PF01735), which is a consensus unit, was found in all of these homologue and paralogue sequences. The compartmentalization analysis by CELLO indicated that the phospholipase family proteins were mostly extracellular proteins (17 protein sequences; 81 % of total phospholipases), and a minor portion belonging to oleaginous strains was located in the cytoplasm (4 protein sequences; 19 % of total phospholipases) (Table S3). Similar to other lipase families, the phospholipase enzymes also had a consensus pentapeptide (GxSxG). BLASTP analysis revealed the highest sequence similarity of extracellular phospholipases to phospholipase B (PLB) or lyso phospholipase; approximately 20 amino acid residues as the putative signal peptide cleavage site were detected at their N-termin of the extracellular phospholipase (PLB group). It has been postulated that PLB might be secreted across the plasma membrane (Farn et al., 2001; Shen et al., 2004). Obviously, the cytoplasmic phospholipases shared homology with phospholipase A (PLA) involved in the removal of the sn-1 acyl group from phospholipids to produce 2-acyllysophospholipids (lysophosphatid acids) (Watanabe et al., 1999; Richmond & Smith, 2011), which act as precursors in phosphatidic acid biosynthesis. Recent studies have reported that the rate of lipolysis and the level of cellular phosphatidic acid can influence the size of lipid droplets (Fei et al., 2011; Ayciriex et al., 2012).

Carboxylesterase family (COesterase-domain-containing protein) The COesterase domain (PF00135) of 40 amino acid residues, which belongs to the esterase family containing alpha/beta hydrolase folds, was found in seven yeast and fungal genomes. These sequences encoded enzymes in the esterase (COesterase) family, which are commonly designated carboxylesterase or lipase of type B similar to triacylglycerol lipases. Some of them contained more than one coding sequence of carboxylesterase, including Y. lipolytica (4 isoforms), Rho. toruloides (2 isoforms) and Mor. alpina (5 isoforms). Interestingly, there were 26 homologous proteins coding for carboxylesterase in the Asp. oryzae genome (Table 2). The majority of the predicted proteins in the carboxylesterase family were extracellular proteins (23 protein sequences; 57.5 % of total COesterase) and cytoplasmic proteins (13 protein sequences; 32.5 % of total COesterase). These results are consistent with previous studies demonstrating that both extracellular and intracellular activities of carboxylesterase enzymes have been detected in several micro-organisms (Donaghy & McKay, 1992; Molinari et al., 2000; Dröge et al., 2005; Zorn et al., 2005). The N-terminal signal peptide, which is anchored to the cell wall, was detected in the extracellular carboxylesterases similar to other extracellular lipase proteins, such as extracellular triacylglycerol lipase (lipase class3) and phospholipase (PLA2_B) families. Nevertheless, the secretion process of this family has not been clarified. The finding of cytoplasmic carboxylesterases in yeast and fugal genomes is in agreement with the experimental study of Y. lipolytica and other fungi, documenting that they are categorized as cell-bound lipases (Fickers et al., 2005b). Interestingly, the extracellular carboxylesterase enzymes were present in four oleaginous strains, including Rho. toruloides, Asp. oryzae, Mor. alpina and Muc. circinelloides (Fig. 5 and Table S3), which might utilize surrounding carbon sources in the form of carboxyl ester and then store the derived fatty acids in the cell.

Putative hormone-sensitive-like lipase family (Abhydrolase_3-domain-containing protein) Based on homology search, we found protein sequences corresponding to the Pfam entry abhydrolase_3 domain (alpha/beta hydrolase fold; PF07859) in several yeast and fungal genomes studied, excluding those of Sac. cerevisiae, Ash. gossypii and Rho. toruloides. Although the alpha/beta hydrolase fold family is one of the largest groups of structurally related enzymes containing eight strands connected by helices, which have diverse catalytic functions (Holmqquist, 2000), there is little information on this lipase family. As a result of in silico analysis, we postulate that this lipase family consists of intracellular enzymes, which was predicted to localize in the cytoplasm (17 proteins; 46%), mitochondria (9 proteins; 24%), plasma membrane (8 proteins, 22%) and nucleus (3 proteins; 8%). In addition to the signature characteristic of lipases (GxSxG motif), the C-terminal amino acid sequences of this family in yeast and fungi showed similarities with the hormone-sensitive lipases of mammals, which are capable of hydrolysing triacylglycerols, diacylglycerols, monoacylglycerols and cholesterol esters (Kraemer & Shen, 2002; Holm 2003). It has been reported that these mammalian enzymes play a role in the regulation of lipid storage and energy homeostasis (Kraemer & Shen, 2002; Holm 2003). Thus, the yeast and fungal enzymes containing the abhydrolase_3 domain are called ‘hormone-sensitive-like lipases’, which might be involved in lipid homeostasis.
DISCUSSION

A wide diversity of lipase families has developed in microorganisms throughout evolution for survival in their surrounding natural environments. The discrete structural motif of amino acid sequences, particularly the serine-centre motif (GxSxG), which possesses the characteristic catalytic triad of lipases, could facilitate identification of homologous enzymes. We classified lipases by in silico analysis of conserved motifs and domains, subcellular

Fig. 5. A functional schematic diagram of the lipid-degrading enzymes of yeast and fungi. Ago, Ash. gossypii; Sce, Sac. cerevisiae; Spo, Sch. pombe; Hpo, H. polymorpha; Ppa, Pic. pastoris; Yli, Y. lipolytica; Rto, Rho. toruloides; Aor, Asp. oryzae; Mal, Mor. alpine; Mci, Muc. circinelloides; Ror, R. oryzae. SE, steryl ester; TAG, triacylglycerol; PL, phospholipid; CE, carboxyl ester. Green bars indicate the presence of lipase enzymes of individual genomes. Upper-case letters F-I to F-VI indicate the abbreviation of lipase families.

Lipid biosynthesis & lipid accumulation
localization and enzyme topology, thereby inferring a divergence of seven families among 11 genomes of yeast and fungi. The existence of two ubiquitous enzymes, steryl ester hydrolase and triacylglycerol lipase, in all genomes studied indicates that they might have arisen from common ancestral genes and diverged into distinct evolutionary events. Moreover, these two lipase activities cannot be compensated by other enzyme activities (Athenstaedt & Daum, 2003; Köffel et al., 2005; Müllner et al., 2005). In addition, the two families might contribute essential roles in lipid mobilization or trafficking for maintaining lipid homeostasis in the cells. Nevertheless, it has been reported that their functions are non-essential under common growth conditions, which might be explained by the functional redundancy of lipase gene products (Athenstaedt & Daum, 2005; Landrositz et al., 2005; Daum et al., 2007). However, these ubiquitous lipases may turn to be important for supporting cell viability in the absence of carbon sources (van Zutphen et al., 2014).

In yeast, deletion of genes encoding the lipid-droplet-mobilized proteins, which included triacylglycerol lipase (patatin-domain-containing protein) and steryl ester hydrolase (abhydro_lipase-domain-containing protein), showed alterations in lipid droplet composition, size and number (Wagner et al., 2009; Koolwein, 2010; Koolwein et al. 2013). According to enzyme topology in this study, the two elementary lipases behave as integral membrane proteins, which might be involved in the formation of lipid monolayers in addition to plasma membrane as suggested in previous reports (Athenstaedt et al., 1999; Athenstaedt & Daum, 2005). Unfortunately, no tool for localizing proteins embedded in lipid particles is available. Nevertheless, it has been reported that these ubiquitous lipase families are not involved in the lipid accumulation process in WT strains of oleaginous microbes. The transcript levels of genes associated with degradation of triacylglycerol (patatin-domain-containing lipase) and steryl ester (abhydro_lipase-domain-containing lipases) were not significantly changed during the nitrogen limitation phase or lipid accumulation phase (Morin et al., 2011; Rismani-Yazdi et al., 2012). We also found that a number of genes encoding triacylglycerol lipase and steryl ester hydrolase were rather similar between oleaginous and non-oleaginous strains (Table 2). The abhydro_lipase domain-containing proteins in yeast and fungi required for steryl ester hydrolysis also have a role in maintaining free sterol homeostasis in cells (Kölbel et al., 2005; Müllner et al., 2005; Köffel & Schneiter, 2006). The steryl ester hydrolase family was paralogous to the mammalian acid lipase family, which includes the lysosomal acid lipase, the gastric triacylglyceride lipase and cholesteryl ester hydrolase. Very interestingly, the discovery of acidic (pI<7) extracellular triacylglycerol lipases (or lipase-class3-domain-containing protein) only in oleaginous yeast and fungi suggests their potential for application in food and feed industries for hydrolysis and/or modification of triacylglycerols under acidic conditions to improve nutritional properties (Mhetras et al., 2009).

The extracellular lipase proteins, which facilitate nutrient adsorption from the surrounding medium, have important roles in yeast and fungal growth, and implication for industrial interest. Our results showed that at least one family of extracellular lipases was found in each species of yeast and fungi. For example, Ash. gossypii contained only extracellular triacylglycerol lipase family (lipase-class3-domain-containing protein), whereas the carboxylesterase family (COesterase-domain-containing proteins) was found in Mor. alpina. Interestingly, particular oleaginous yeast and fungi (excluding Rho. toruloides and Mor. alpina) had several extracellular lipase proteins belonging to different types of lipase families when compared with non-oleaginous species. The functional complexity of lipase enzymes of individual genomes is shown in Fig. 5. Notably no gene encoding extracellular carboxylesterase was found in non-oleaginous species. Furthermore, the extracellular triacylglycerol lipase was dominant in oleaginous species that might deduce their ability of triacylglycerol utilization in extracellular media more than those of non-oleaginous species. The highest number of lipase-class3-domain-containing genes was observed in Y. lipolytica (Tables 2 and S3).

The first genetically characterized gene coding for extracellular triacylglycerol lipase was reported in yeast, the LIP2 gene of Y. lipolytica (Pignéde et al., 2000). The complete genome sequence of Y. lipolytica possesses 16 lipase-coding genes that are homologous with the known extracellular triacylglycerol lipase gene (LIP2) (Fickers et al., 2011) and only three isoforms (Lip2p, Lip7p and Lip8p) have been characterized so far (Pignéde et al., 2000; Fickers et al., 2005a). It has been documented that the extracellular triacylglycerol lipase (lipase class3) is associated with the cell wall before being released into the culture medium by mild alkali (Fickers et al., 2004; Liu et al., 2010). At the transcriptomic level of Y. lipolytica, the overexpression of an extracellular lipase-coding gene (LIP15) was detected during lipid accumulation (Morin et al., 2011). The role of extracellular triacylglycerol lipase may be associated with the lipid channelling process for supplying an additional pool of fatty acids for triacylglycerol synthesis and lipid accumulation (Morin et al., 2011). Thus, the extracellular triacylglycerol lipase family identified in this work might be a potential target to improve the lipid accumulation process in yeast and fungi for production of biodiesel and other lipid-based products. However, the physiological function of each lipase family should be further clarified through experimental study in addition to direct sequence comparison. It can be summarized that the set of genes encoding lipases of oleaginous yeast and fungi is larger in terms of gene families and number of genes when compared with that of non-oleaginous species. Thus, oleaginous species would be promising sources for production of diversified lipases with considerable industrial applications. For example, extracellular triacylglycerol lipases are biocatalysts for food processing, drug synthesis and wastewater treatment (López del Castillo-Lozano et al., 2007; Wu et al., 2009; Fickers et al., 2011;
In conclusion, the divergence of lipolytic enzymes in yeast and fungi on the basis of genome analysis expands perspectives not only for better understanding dynamic processes of lipid metabolism and related metabolic pathways, but also for searching for potential enzymes of particular interest in individual industrial applications.

ACKNOWLEDGEMENTS

This work was supported by the Higher Education Research Promotion and National Research University Project of Thailand, Office of the Higher Education Commission.

REFERENCES


Edited by: D. Mattanovich