Characterization and overexpression of esterases-encoding Lip900 and Lip3954 through metagenomic sequencing of paddy soil

Yu Pei Chen\(^{1,2,3,4}\), Pei Chi Lee\(^4\), Hong Tan Wu\(^{1,2,3}\), Guey Horng Wang\(^{1,2,3}\), Chai Fang Tsai\(^5\) and Chiu Chung Young\(^{5,6}\)

\(^1\)Department of Medical Technology, Xiamen Medical College, Xiamen, China. 
\(^2\)Application Technique Engineering Center of Natural Cosmeceuticals College of Fujian Province, Xiamen Medical College, Xiamen, China 
\(^3\)Xiamen Key Laboratory of Traditional Chinese Medicine Bioengineering, Xiamen Medical College, Xiamen, China 
\(^4\)Department of Biological Science and Technology, China University of Science and Technology, Taipei, Taiwan 
\(^5\)Department of Soil and Environmental Sciences, National Chung Hsing University, Taichung, Taiwan 
\(^6\)Innovation and Development Center of Sustainable Agriculture, National Chung Hsing University, Taichung, Taiwan

E-mail: ccyoung@mail.nchu.edu.tw

Abstract. A lot of industrial genes can be explored from the metagenomic sequences. In this study, metagenome sequencing of paddy soil was carried out, and several putative open reading frames (ORFs) involved in the lipolytic activity can be identified. Lipolytic enzymes are widely used in different industrial applications, such as biodiesel production, bioremediation, and waste treatment. To verify the lipolytic enzymes of assembly ORFs, two putative genes encoding esterase, namely Lip900 and Lip3594, which shared 47.6% and 43.7% identities with the uncharacterized esterase proteins, were synthesized and constructed with pET-30a for Escherichia coli overexpression. Lip900 and Lip3594 belonging to VI and XII families were successfully obtained and characterized. The results of recombinant proteins indicated that Lip900 and Lip3594 preferred to hydrolyze short-length p-nitrophenyl (p-NP) esters such as p-NP butyrate (C4). The optimal temperature and pH for both Lip900 and Lip3594 were 30°C and pH 7, respectively. Nevertheless, Lip3594 had a higher relative activity than Lip900 when the temperature was over 40°C. The effect of various reagents on Lip900 and Lip3594 activities was determined. The inhibition of Lip900 and Lip3594 was observed in the presence of MgSO\(_4\), MnSO\(_4\), NiSO\(_4\), and sodium dodecyl sulfate (SDS). However, the addition of ethylenediaminetetraacetic acid (EDTA) can improve the lipolytic activity, indicating these esterases without metal ions as the cofactor. Moreover, Lip900 and Lip3594 were resistant to methanol, ethanol, and butanol. Over 81.6% of the relative activity of Lip900 can be attained when these organic solvents of alcohol were added to 10%. These results revealed that Lip900 and Lip3594 have potential applications in various industries.

1. Introduction
Lipolytic enzymes can be applied in various industrial areas including biodiesel production, bioremediation, and waste treatment. They were grouped into eight families according to their amino acid sequences [1]. Furthermore, the new families were successively found to reach fourteen families, which family I was recognized as the true lipase and family II to XIV was esterase [2]. In the lipase engineering database, lipolytic enzymes were classified into three classes, GGGX, GX, and Y, on the basis of the oxyanion hole, which stabilizes the catalytic reaction [3]. The development of high-yield production has attracted considerable attention. Therefore, the heterologous expression of an esterase from various organisms in E. coli, Bacillus subtilis, Trichoderma reesei, Pichia pastoris, and Saccharomyces cerevisiae has been thoroughly investigated, and esterases have been successfully produced [4].

Several studies reported that metagenomics library construction is used to screen the functional genes [5-6]. However, the technique of DNA library construction and screening is time-consuming. Recently, metagenomic sequencing has been a powerful alternative and an already tool for cost reduction [7]. The novel genes can be explored from the metagenomic sequencing which shows a valuable source. In this study, the esterases were screened through the metagenomics sequencing of paddy soil. The novel esterases, Lip900, and Lip3954 which shared only 47.6% and 43.7% identities with the unpublished sequences, were successfully synthesized and overexpressed in E. coli for characterization. The recombinant esterases were purified, their structures were predicted, and the activities were studied.

2. Materials and methods

2.1. Metagenomic sequencing of the paddy soil

Soil DNA obtained from the Taiwan Agricultural Research Institute, Council of Agriculture, Executive Yuan (24°01′N, 120°41′E) was conducted through next-generation sequencing (NGS) [8]. The metagenomic sequencing and genome de novo analysis were performed by the Sequencing-Tech company (Taipei, Taiwan). The Kyoto Encyclopedia of Genes and Genomes (KEGG) was utilized for the analysis of metabolic pathways [9], and ORFs were annotated by MetaGeneMark and BLASTP of National Center for Biotechnology Information (NCBI). The phylogenetic tree was carried out by the neighbor-joining method in MEGA-X software with 1000 bootstrap replicates [10].

2.2. Overexpression plasmid construction of esterase

Two putative ORFs encoding esterase were designated as Lip900 and Lip3954. Lip900 and Lip3954 with EcoRV adaptors at 5′ and 3′ sequences were synthesized and provided by GeneDireX Inc. (Las Vegas, NV, USA). The pET-30a was treated with EcoRV restriction enzyme, and the linear pET-30a was ligated with the synthesized DNA. Two overexpression plasmids for E. coli were obtained, namely pLip900 and pLip3954. The two plasmids transferred into E. coli C43(DE3) which was induced by adding IPTG to 0.1 mM and incubated at 25°C overnight. The cells were disrupted with lysozyme, and the recombinant proteins were purified by a Ni-NTA purification system (Invitrogen, Carlsbad, CA, USA). Protein content was determined by the standard curve of BSA protein with different concentration on the SDS-PAGE.

2.3. Enzymatic activity assay of esterase protein

Lipolytic activity was detected through the detection of p-nitrophenyl (p-NP) esters as substrates by using a spectrophotometer at OD_{405}. The mixture in the enzymatic assay contained the purified esterase and various p-NP esters: p-NP butyrate (C4), p-NP valerate (C5), p-NP octanoate (C8), p-NP decanoate (C10), p-NP dodecanoate (C12), p-NP myristate (C14), and p-NP palmitate (C16). The enzymatic reaction included 0.25 mM p-NP esters (C4–C10), 2.5% ethanol, 50 mM Tris-HCl (pH 7), and the recombinant protein with a final volume of 800 μL. To detect the lipolytic activity with long-chain p-NP esters (C12-C16), 0.2 mM p-NP esters, 2.5% ethanol, 0.1% gum arabic, 0.1% deoxycholate, 50 mM Tris-HCl (pH 7), and recombinant protein were used. The enzymatic mixture
was incubated at 30°C and pH 7 for 10 min. The optimal temperature and pH of enzymatic activity were measured between 20°C–70°C and pH 4–10 using p-NP butyrate as the substrate. The 50 mM sodium acetate/acetic acid buffer, 100 mM sodium phosphate, and glycine/NaOH buffers were used for the assays at pH 4.0–5.0, pH 6.0–8.0, and pH 9.0–10. The reaction was interrupted with 95°C incubation for 15 min. To analyze the effect of different reagents on esterases, various chemical compounds were used containing 10 mM MgSO₄, MnSO₄, NiSO₄, EDTA, SDS and organic solvents of alcohol such as methanol, ethanol, and butanol.

2.4. Nucleotide sequence accession number
The accession numbers of Lip900 and Lip3954 genes have been submitted to the NCBI. The accession numbers are MK641590 and MK641591.

3. Results and discussion

3.1 Esterases identification in metagenomic sequencing
High-throughput sequencing provides a large amount of gene information which can be utilized in the enzymatic engineering. Several metagenomes have been conducted in various fields, such as soil, animal gut, and marine to establish the microbial structure [11-13]. In this study, metagenomic sequencing of paddy soil was carried out with 55.68 Gb sequences by the Illumina Genome Analyze, and the metagenomic project had been deposited at DDBJ/ENA/GenBank under accession numbers BioProject PRJNA498754 and BioSample SAMN10334111 [14]. To explore the lipolytic enzyme in the metagenome, the KEGG pathway of lipid metabolism was analyzed (Fig. 1). The genes involved in fatty acid metabolism were predominant, followed by glycerolipid metabolism and fatty acid biosynthesis. Further observation of glycerolipid metabolism could reveal genes related to the esterase. The result indicated that the triacylglycerol acylhydrolases (lipases) (EC 3.1.1.3) could be found.

![Figure 1. Gene classification of paddy soil metagenome by using KEGG analysis. The red box indicates the sequencing reads which can be identified in the metagenome of paddy soil.](image)

In this study, the database search and ORF prediction further provide the information of putative esterase genes. Of the hypothetical esterases, two putative ORFs (L900 and L3594) encoding esterase with 481 and 308 amino acids with the low identical match were identified (Table 1). These two ORFs were designated as Lip900 and Lip3594; they had 47.6% and 43.7% identities with uncharacterized lipases from *Bradyrhizobium* sp. and uncultured bacterium, respectively.
Table 1. Lipase search from the data mining of putative open reading frame with over 250 amino acids in length based on the NCBI BLAST result.

| No.   | Closed sequence                        | Amino acid length | Identity (%) | Coverage (%) |
|-------|----------------------------------------|-------------------|--------------|--------------|
| L529  | lipase [Pseudomonas sp. GM30]          | 276               | 83.7         | 100          |
| L900  | putative Lipase [Bradyrhizobium sp. STM 3843] | 481               | 47.6         | 93           |
| L1145 | esterase/lipase [Leptolyngbya sp. PCC 7375] | 288               | 67.9         | 99           |
| L1620 | lipase [Tsukamurella paurometabola]    | 280               | 75.4         | 100          |
| L1699 | lysophospholipase [Pyrimononas methylaliphagota] | 300               | 63.5         | 66           |
| L2145 | phospholipase C [Candidatus Koribacter versatilis] | 253               | 72.4         | 100          |
| L2830 | phospholipase [Methylcystis sp. SB2]   | 281               | 71.7         | 98           |
| L3207 | phospholipase C [Candidatus Koribacter versatilis] | 284               | 52.5         | 99           |
| L3594 | esterase/lipase [uncultured bacterium] | 308               | 43.7         | 87           |
| L3766 | phospholipase [Candidatus Solibacter usitatus] | 277               | 65.8         | 100          |

3.2 Characterization of Lip900 and Lip3594

Figure 2. Phylogeny construction of esterase sequences, and prediction of Lip900 and Lip3594 protein structures by using SWISS-MODEL [15].
The lipolytic enzymes can be divided into 14 families by Rao et al. [2]. A phylogenetic tree was constructed by the deduced amino acids for the assessment of esterase relationship among the Lip900, Lip3594 and other esterases from various organisms (Fig. 2). The result suggested that the two genes were classified into two families, VI and XII. Furthermore, three-dimensional structures of Lip900 and Lip3594 were predicted based on the SWISS-MODEL [15]. A typical α/β-hydrolase fold of Lip900 and Lip3594 with a central β-sheet barrel surrounded by α-helices can be found [16]. In addition, the nucleophilic serine, acidic aspartate or glutamate, and histidine residues were conserved in the catalytic triad of esterase [16-17]. According to the alignment of deduced amino acids, the results revealed that the predicted catalytic residues were observed at Ser314, Asp337, and His410 for Lip900 and Ser145, Asp195, and His262 for Lip3594 (Fig. 3). The conserved G-X-S-X-G motif also can be perceived in both genes [1].

![Figure 3](image3.png)

**Figure 3.** The alignment of deduced amino acids of esterases. The filled triangles indicated the assumed active site of serine, aspartate, and histidine.

### 3.3 Esterase activity assay

To confirm the property of Lip900 and Lip3594, *E. coli* was used as a heterologous host for esterase expression. Lip900 and Lip3594 were synthesized and introduced into the expression vector of pET-30a including a His-tag fusion of N-terminus. The recombinant proteins were purified, and various p-NP esters were utilized for demonstrating the esterase activity. The result indicated that Lip900 and Lip3594 tended to be the hydrolysis of short-length p-NP esters (C4)(Fig. 4). This was in agreement with other esterases from *Thalassomonas agarivorans*, and *Aeromonas* sp. [18]. The temperature and pH profiles of the recombinant Lip900 and Lip3594 were further determined. The optimal temperature of Lip900 and Lip3594 was 30°C. At 20°C, both relative activities were dropped below 30%. However, the residual activity of Lip3594 remained 72% at 50°C and was higher than that of Lip900 with 21%. The optimal pH of Lip900 and Lip3594 was pH 7. When pH was below 5, and above 9, a rare activity of Lip900 and Lip3594 was measured.

![Figure 4](image4.png)

**Figure 4.** The optimal enzymatic activity of (a) p-NP esters, (b) temperature and (c) pH by Lip900 and Lip3594.

### 3.4 Effect of various reagents on Lip900 and Lip3594

The effect of different chemical reagents and metal ions on the recombinant esterase activities was shown in Fig. 5. Lip900 was obviously inhibited by MgSO₄, MnSO₄, NiSO₄, and SDS; besides, the remarkable inhibition of Lip3594 was observed by addition of NiSO₄ and SDS. By contrast, EDTA can improve a considerable stimulation of esterase activity. This implied that metal ions were not required as the enzymatic cofactor. The similar result was reported for the esterases from
**Psychrobacter cryohalolentis** and **Yarrowia lipolytica** [19-20]. High organic solvents usually result in lost enzymatic activity [21]. Rare effect of organic solvents on Lip900 was perceived as methanol, ethanol, and butanol below 5%. Additionally, no enzymatic activity loss on Lip3594 was found as ethanol below 10%. When these organic solvents were gradually increased, the enzymatic activity of Lip900 and Lip3594 was slowly decreased. These results suggested that Lip900 and Lip3594 had a high tolerance to organic solvent of methanol and ethanol, respectively. These enzymes were more stable than that of esterases from **Bacillus methylotrophicus** and **Yarrowia lipolytica** [20, 22].

**Figure 5.** (a) Effect of various reagents on Lip900 and Lip3594. Different concentration of methanol, ethanol, and butanol was added into (b) Lip900 and (c) Lip3594 reaction with pH 7 at 30°C.

**4. Conclusions**

Esterases were the new trend for industrial application and account for 21% of the worldwide industrial enzyme market. They are the third largest commercialized enzymes after proteases and carbohydrate enzymes [23]. In this study, two genes, Lip900, and Lip3594, involved in lipolytic activity were identified through the paddy soil of metagenomic sequencing. According to the classification of phylogeny, Lip900 and Lip3594 were belonging to family VI and XII. These enzymes can be successfully overexpressed in **E. coli**, and the enzymatic profile can be characterized in detail. Lip900 and Lip3594 had the ability for hydrolyzing short-length p-nitrophenyl esters. In addition, they were both sensitive to the metal ion of nickel. Addition of EDTA into the recombinant esterases can enhance the enzymatic activity. On the other hand, Lip900 and Lip3594 were obviously resistant to methanol and ethanol, respectively. This property was useful for biodiesel application. Taking together, the metagenomic sequencing-based strategy had the potential to be applied in the various fields for mining the novel genes.

**Acknowledgments**

This work was supported by the Ministry of Science and Technology, Taiwan (Grant 102-2313-B-157-003-MY3), the Educational and Scientific Research Program for Young Scholar Sponsored by Educational Department of Fujian Province (JAT170715), Fujian, China, and the Innovation and Development Center of Sustainable Agriculture from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan.

**References**

[1] Arpigny J L and Jaeger K E 1999 *Biochem J* **343** Pt 1 pp 177-183
[2] Rao L, Xue Y, Zheng Y, Lu J R and Ma Y 2013 *PLoS One* **8** e60645
[3] Widmann M, Juhl P B and Pleiss J 2010 *BMC Genomics* **11** p 123
[4] Valero F 2012 *Methods Mol Biol* **861** pp 161-178
[5] Lu M, Dukunde A and Daniel R 2019 *Appl Microbiol Biotechnol* 10.1007/s00253-019-09695-1
[6] Istvan P, Souza A A, Garay A V, Dos Santos D F K, de Oliveira G M, Santana R H, Lopes F A C, de Freitas S M, Barbosa J and Kruger R H 2018 *Biotechnol Lett* **40** pp 1395-1406
[7] Bilal T, Malik B and Hakeem K R 2018 *J Microbiol Methods* **155** pp 65-69
[8] Chen Y P, Liaw L L, Kuo J T, Wu H T, Wang G H, Chen X Q, Tsai C F and Young C C 2019 *J
Biosci Bioeng 128 pp 274-282

[9] Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita K F, Itoh M, Kawashima S, Katayama T, Araki M and Hirakawa M 2006 Nucleic Acids Res 34 D354-357

[10] Kumar S, Stecher G and Tamura K 2016 Mol Biol Evol 33 pp 1870-1874

[11] Gastauer M, Vera M P O, de Souza K P, Pires E S, Alves R, Caldeira C F, Ramos S J and Oliveira G 2019 Sci Data 6 190008

[12] Qin J, Li R, Raes J, Arumugam M, Burgdorf K S, Manichanh C, Nielsen T, Pons N, Levenez F and Yamada T, et al. 2010 Nature 464 pp 59-65

[13] Muangchinda C, Rungsihiranrut A, Prombutara P, Soonglerdsongpha S and Pinyakong O 2018 J Hazard Mater 357 pp 119-127

[14] Chen Y P, Liaw L L, Kuo J T, Wu H T, Wang G H, Chen X Q, Tsai C F and Young C C 2019 J. Biosci. Bioeng. 128 pp 274-282

[15] Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Gallo Cassarino T, Bertoni M and Bordoli L, et al. 2014 Nucleic Acids Res 42 W252-258

[16] Casas-Godoy L, Duquesne S, Bordes F, Sandoval G and Marty A 2012 Methods Mol Biol 861 pp 3-30

[17] Sanishvili R, Yakunin A F, Laskowski R A, Skarina T, Evdokimova E, Doherty-Kirby A, Lajoie G A, Thornton J M, Arrowsmith C H and Savchenko A, et al. 2003 J Biol Chem 278 pp 26039-26045

[18] Chen Y P, Liang S S, Hwang I E, Chien M C, Wang G H and Liaw L L 2018 MATEC Web of Conferences 169 p 01018

[19] Novototskaya-Vlasova K A, Petrovskaya I E, Rivkina E M, Dolgikh D A and Kirpichnikov M P 2013 Biochemistry (Mosc) 78 pp 385-394

[20] Li T, Zhang W, Hao J, Sun M and Lin S X 2019 Biotechnol Rep (Amst) 21 e00295

[21] Gupta M N, Batra R, Tyagi R and Sharma A 1997 Biotechnol. Prog. 13 pp 284-288

[22] Sharma P, Sharma N, Pathania S and Handa S 2017 J Genet Eng Biotechnol 15 pp 369-377

[23] World enzymes 2014 Industry study with forecasts for 2017 & 2022 (Cleveland: The Freedonia Group)