MALAYSIAN JOURNAL OF BIOCHEMISTRY & MOLECULAR BIOLOGY

The Official Publication of The Malaysian Society For Biochemistry & Molecular Biology (MSBMB)
http://mjbmb.org

LipExPS: A VALIDATED WEB-BASED PREDICTION TOOL FOR RECOMBINANT LIPASE EXPRESSION IN *Pichia pastoris*

Fatima Gogo Mayaki [1, 2, 4], Arpah Abu[5], Abu Bakar Salleh[2], Siti Nurbaya Oslan[1, 2, 3*]

[1]*Department of Biochemistry, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, 43400 Serdang Selangor, Malaysia*
[2]*Enzyme and Microbial Technology Research Centre, Universiti Putra Malaysia, 43400 Serdang Selangor, Malaysia*
[3]*Institute of Bioscience, Universiti Putra Malaysia, 43400 Serdang Selangor, Malaysia*
[4]*Department of Biochemistry, Faculty of Natural Sciences, Ibrahim Badamasi Babangida University, P.M.B. 11 Lapai, Niger State, Nigeria*
[5]*Institute of Biological Sciences, Faculty of Sciences, University of Malaya, 50603 Kuala Lumpur, Malaysia*

*Corresponding Author: snurbayaoslan@upm.edu.my

History
Received: 19th December 2020
Accepted: 5th March 2021

Keywords:

Lipase, Prediction System, Promoter, Pichia pastoris, Database

Abstract

The production of recombinant lipase in *Pichia pastoris* expression system has become popular due to its reliability and reproducibility. But, choosing the best *P. pastoris* strain and promoter has been a major issue, leading to time consumption. Hence, there is a need to develop an *in silico* tool that can assist in the prediction of recombinant lipase production before cloning. LipExPS is a web-based resource that uses a newly curated database (LipExDB) and is devised to provide a platform for predicting the best promoter/host combination to express a lipase gene in *P. pastoris*. LipExDB was constructed to store lipase expression data that have been successfully expressed in *P. pastoris*. LipExPS used a BLAST heuristic algorithm to retrieve data from LipExDB to predict the best promoter/host combination when a new lipase is intended to be expressed in *P. pastoris* system. The competency of LipExPS was experimentally validated using a bacterial lipase gene (ARM lipase). Initially, the amino acid sequence of ARM lipase was submitted into LipExPS followed by evaluating the scores given by the system. Then, based on the given scores and suggestions, the ARM lipase gene was cloned into *P. pastoris* strain GS115 using alcohol oxidase (AOX) and glyceraldehyde-3-phosphate (GAP) promoters. From the experiment carried out, ARM lipase was successfully expressed in *Pichia* system using the constructs. Expression of the recombinant ARM lipase in *P. pastoris* strain GS115 confirmed that LipExPS is a reliable web-based tool to predict the expression of recombinant lipases in *P. pastoris*.
URL: LipExPS - http://103.18.1.10/lipex/index.html; LipExDB - http://kbioimage.my/lipex/indexDB.php

INTRODUCTION

Lipases are key enzymes that catalyze the hydrolysis and synthesis of triglycerides [1]. They are ubiquitous in nature and play an important role in the metabolism of lipids [2,3]. Lipases are important in various industrial applications such as in the pharmaceutical and food industries [4]. In order to meet industrial needs, a method known as molecular cloning and expression is employed to enhance the yield [5]. This technique facilitates the process of controlling and regulating the host cell carrying the protein of interest to achieve high levels of expression. *Pichia pastoris* is yeast that has been excessively used to express recombinant proteins. However, the selection of promoter

and host strains is important prior to cloning a gene [6]. Presently, the choice of promoter is made by first screening several promoters (alcohol oxidase – AOX; glyceraldehyde-3-phosphate – GAP; formaldehyde dehydrogenase – FLD) which in turn increases the time required to select the best promoter to use when cloning the lipase of interest [7-9]. Indeed, once the promoter is determined the most suitable *P. pastoris* strain needs to be selected in conjunction to suit the regulatory system. In fact, the negative results or any failures are not reported by the researchers. Thus, make the chances of getting high recombinant protein expression using the selected promoter and host strain become harder. Hence, there is a need for a prediction tool to assist in the choice of promoter and host strain before cloning.

Some available promoter databases includes Eukaryotic Promoter Database [10]; TRED [11]; CircuitDB connects [12]; TRANSFAC database [13]; RegNetwork [14]; JASPER [15]; *Saccharomyces cerevisiae* promoter database (SCPD) [16]; Yeast Promoter Atlas (YPA) [17]; Yeast Search For Transcriptional Regulators and Consensus tracking (YEASTRACT) [18]; YCRD [19]. Out of all the promoter databases presently available, none is available to store the promoters used to express a gene in *Pichia pastoris*. The (Basic Local Alignment Search Tool (BLAST) algorithm has been implemented in a few software such as CUDASW++ 2.0 [20] PSI-Search [21]; CS-BLAST [22]; BLAST [23]; PSI-BLAST (Altschul et al. 1997); ScalaBLAST [25]; SAM-T08 [26]; SWAPHI [27]; SWIPE [28]. However, BLAST algorithm and the database have not been implemented to be used in predicting and recording the protein expressions in *P. pastoris* system.

There are many transcription factor binding site analyses such as promoters of genes associated with cancer [29], single-molecule analysis of transcription factor binding at transcription sites in live cells [30], plant transcription factors analysis [31] and transcription factor binding site prediction [32]. But, these systems are not related to the prediction system for recombinant protein expression in a cloning/expression host.

In order to facilitate the molecular expression of lipase gene in *P. pastoris* system, this study sought to develop a computational method to predict the expression prior to cloning. This tool called LipExPS implements a pipeline for lipase expression that takes advantage of existing lipase expression data. Lipases expressed in *P. pastoris* expression host were curated manually and BLAST heuristic algorithm was implemented to compare the existing sequences of lipase expression data to the new lipase sequence. To validate LipExPS, an amino acid sequence of a lipase was selected and submitted to LipExPS. Then, the predicted promoter and host strain were used in the laboratory experiment.

MATERIALS AND METHODS

System Implementation and Requirement to Access LipExPS

LipExPS was written in the Perl and BLAST heuristic algorithm (BHA) for sequence comparison and alignment. HyperText Markup Language (HTML), Cascading Style Sheets (CSS) and JavaScript were used for the interface design. This system consisted of two modules; LipExPS and LipExDB. Internet access is needed to use the LipExPS webserver. LipExPS webserver can open on Android and other operating systems. Also, any search engine can open LipExPS.

LipExPS Module

The general flow of LipExPS module is illustrated in Fig 1(a) and it can be divided into graphical user interface (GUI), algorithm implementation and the database. The GUI was designed to facilitate the use of the system. The algorithm was implemented for sequence comparison, while the database was used to store and manage the data.

LipExPS Implementation

Fig 1(b) shows the flow chart of LipExPS implementation. The algorithm used in the prediction was BHA. BHA was implemented in Perl scripting language where it identified homologous sequences using a heuristic method. The comparison was done starting with 3 letters at a time. Then, the algorithm was implemented in three steps [33]. The first step was building the hash table, called the query index, the second step was to identify hits and the third step was to find similar words and compute the score achieved.

LipExDB Module

Fig 1(c) shows the workflow of LipExDB development, where data was collected from public databases and journal articles. The search was done using keywords such as “lipase expression”, “*Pichia pastoris*”, “yeast expression system”, “lipase”, “enzyme”, “yeast expression”, “recombinant protein in yeast”, “lipase expressed in yeast”, and “lipase expressed in *P. pastoris*”. Public databases such as Genbank [34] and PubMed [35] were used for the search. Expression data and sequence for lipases were manually curated in a non-redundant manner by not repeating any information and documented in LipExDB.

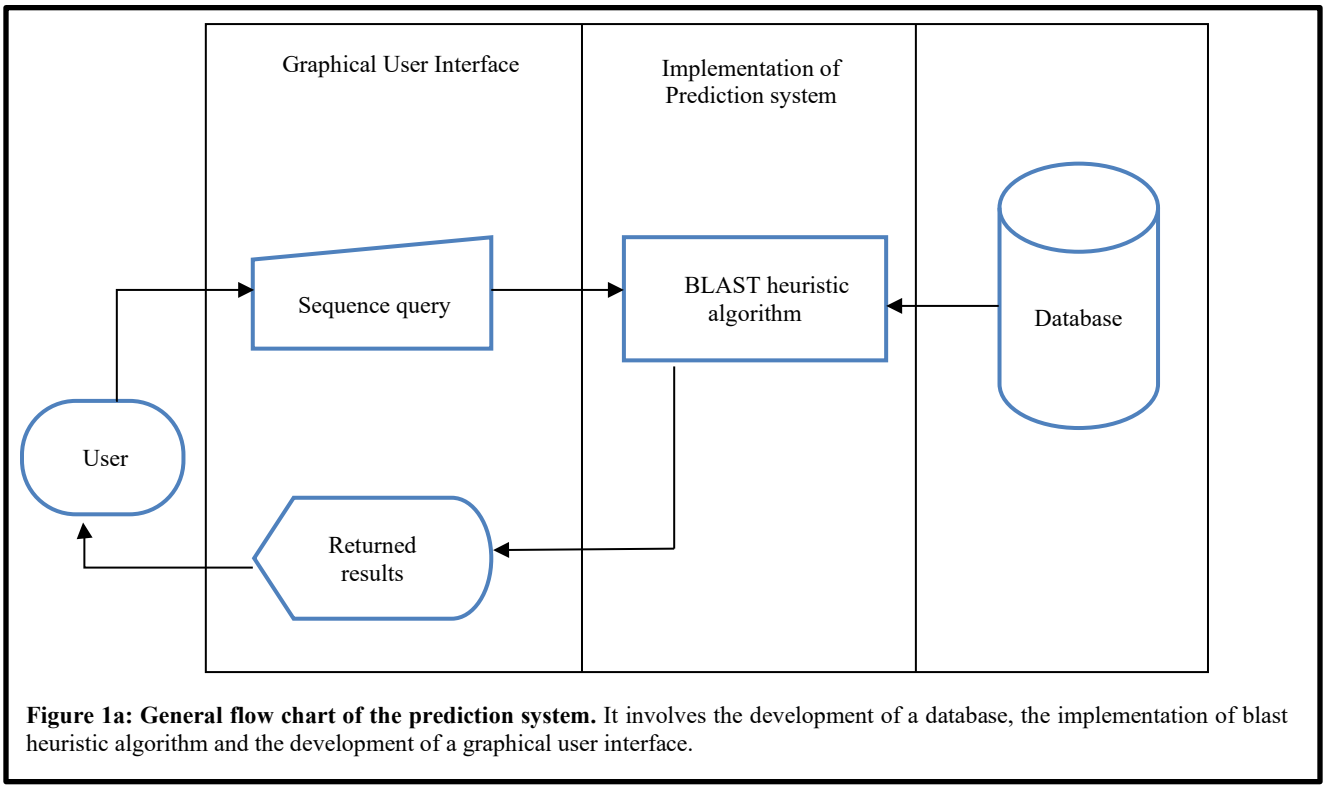


Figure 1a: General flow chart of the prediction system. It involves the development of a database, the implementation of blast heuristic algorithm and the development of a graphical user interface.

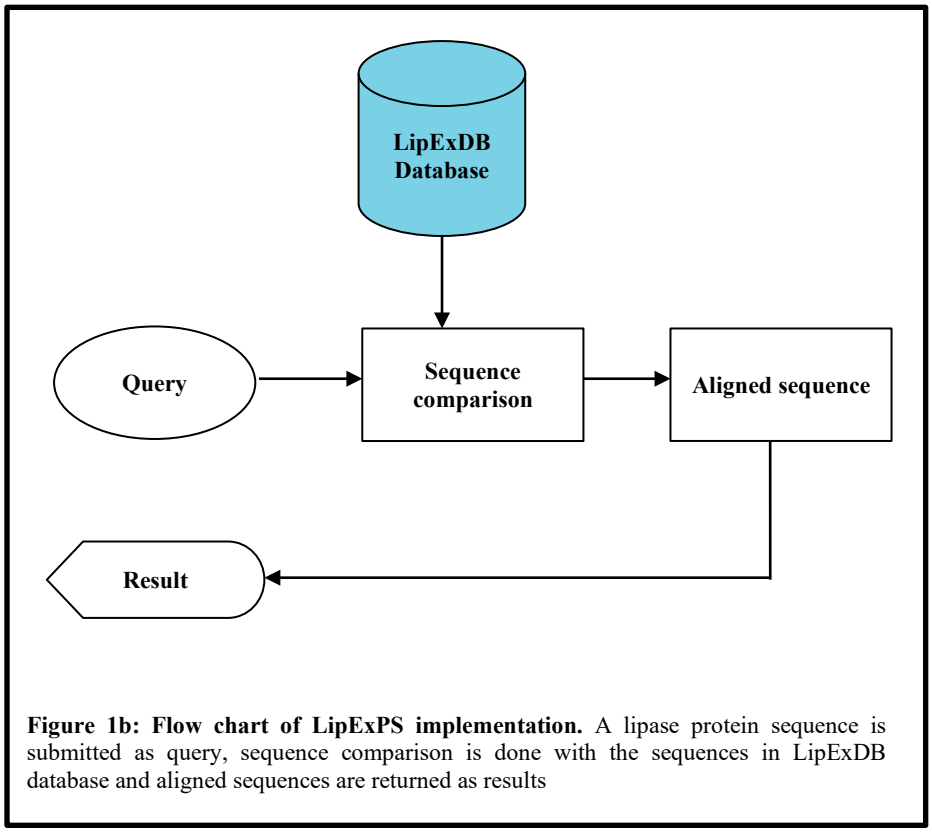


Figure 1b: Flow chart of LipExPS implementation. A lipase protein sequence is submitted as query, sequence comparison is done with the sequences in LipExDB database and aligned sequences are returned as results

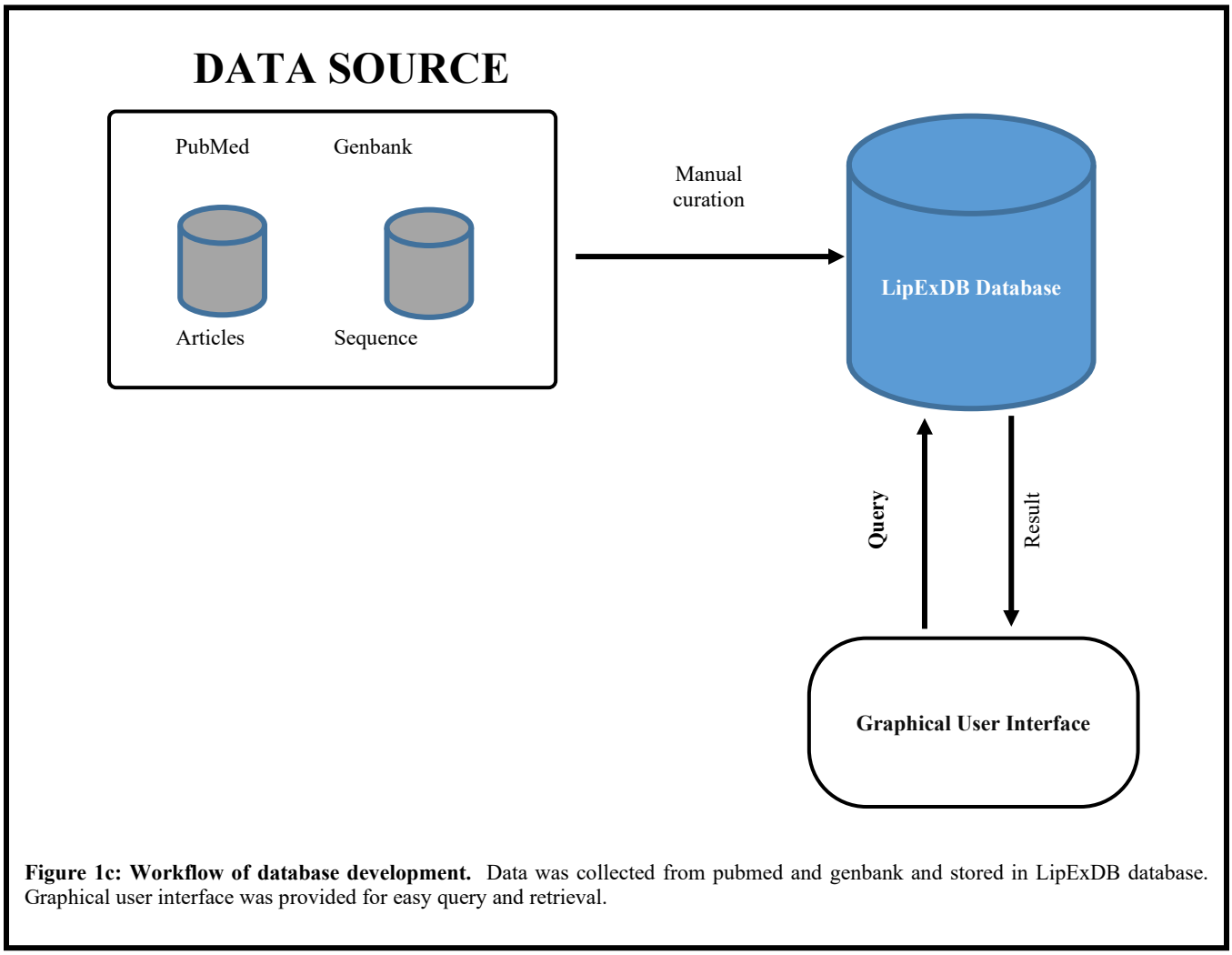


Figure 1c: Workflow of database development. Data was collected from pubmed and genbank and stored in LipExDB database. Graphical user interface was provided for easy query and retrieval.

LipExDB Implementation

The LipExDB database was implemented in MySQL, permitting users to search for lipases according to specific criteria. The web interface of LipExDB database was written in Hypertext Preprocessor (PHP) version 7.0 and JavaScript under an Apache webserver running on a Windows operating system.

Eight intra connected tables were constructed in MySQL. The tables contained information about lipase name gene characteristics, source, citing information and linked to each other using one-to-many relationship as shown in Fig 1(d); (i) The gene table provides information about the list of lipase genes and their symbol in the database, (ii) publication provides information on the cited literatures, (iii) source provides information on where the lipase gene was isolated, (iv) sample provides information on the accession number, sequences, and characteristics of the lipases in the database, (v) strain provides information on the strains of *P. pastoris* used to express the curated lipases in the database, (vi) vector provides information on the expression vectors of *P. pastoris* used to clone the curated lipases in the database, (vii) promoter provides information on the promoters of *P. pastoris* used to clone and express the curated lipases in the database, (viii) spvs provides information on the yield and the level of expression of the lipases curated in the database.

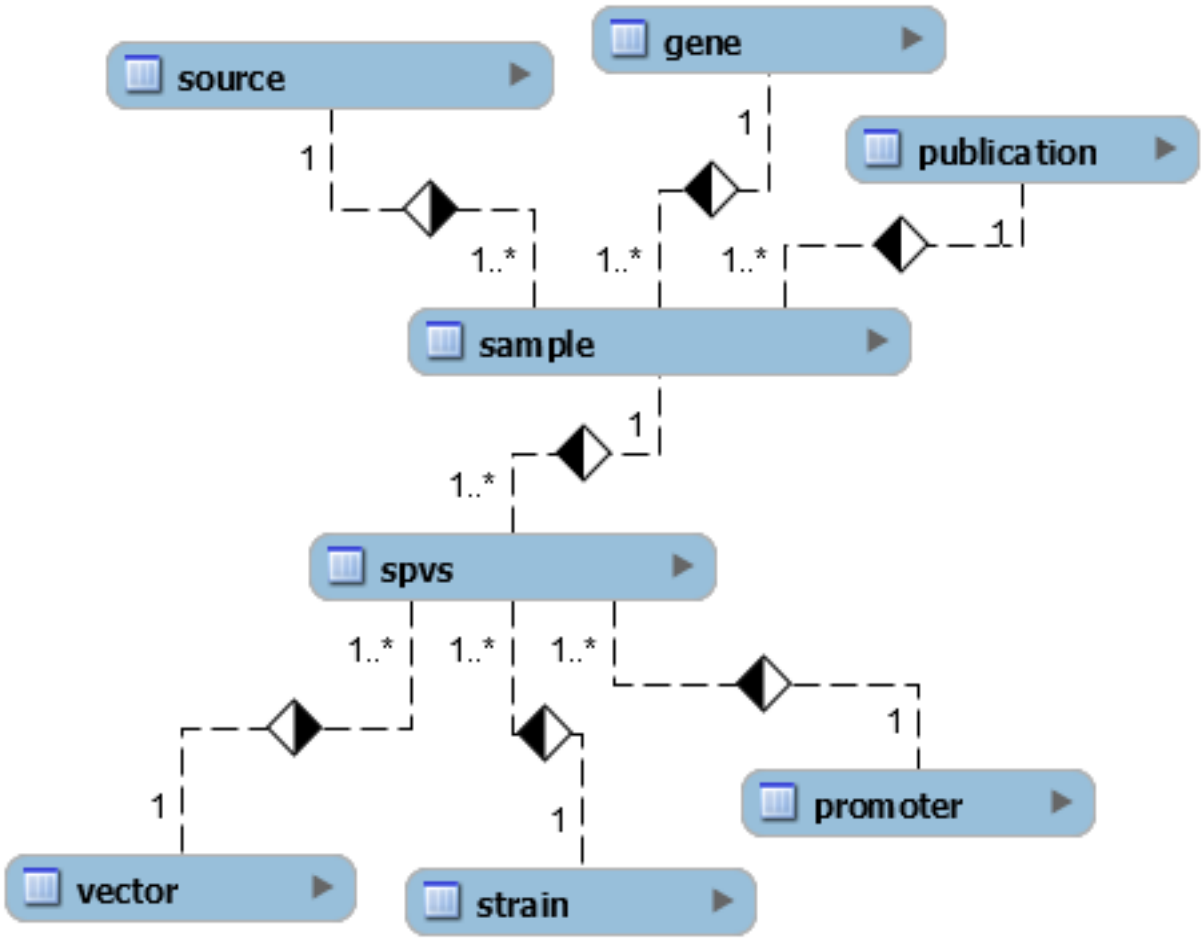


Figure 1d: Database schema. The schema was designed using MYSQL workbench 6.3 CE. The schema consists of eight tables which have one to many relationships with each other

LipExPS Validation

Strains, Plasmids and Media

Pichia pastoris strain GS115, plasmids (pPICZαB and pGAPZαA), and *Escherichia coli* strain TOP10 were purchased from Invitrogen, USA. Luria Bertani medium (LB) [1% (w/v) tryptone, 0.5% (w/v) yeast extract, 1% (w/v) NaCl, pH 7, 1.5% (w/v) agar] supplemented with 25 µg/mL of Zeocin was used to grow and maintain *E. coli* containing plasmid. Yeast Peptone Dextrose medium (YPD) [1% (w/v) yeast extract, 2% (w/v) peptone, 2% (w/v) dextrose] was used for *P. pastoris* growth. YPDS (same composition as YPD with 1M sorbitol) supplemented with 100 µg/mL of Zeocin was used for *P. pastoris* transformants. Yeast Peptone Tryptic Glycerol (YPTG) [1% (w/v) yeast extract, 2% (w/v) peptone, 4×10^{-5}% (w/v) biotin, 0.2% (w/v) tryptic soy broth, 1% (w/v) glycerol] was used for mass cultivation, and YPTM culture medium (same as YPTG except with glycerol replaced with 1% (w/v) methanol) was used for lipase expression [36, 37].

Cloning of ARM Lipase into pPICZαB and pGAPZαA and Transformation into *P. pastoris*

The ARM lipase from Ebrahimpour [38] was used for LipExPS validation. The cloning was done according to method suggested by Sambrook [39]. The ARM lipase gene was cloned into two plasmids (pPICZαB and pGAPZαA). Forward primers flanked with *Eco*RI restriction site (underlined letters) (5'-3') (CTGCTGAATTCTTGCGGCTTCGCGAGCCAA) and (CTACTGAATTCGCGGCTTCGCG AGCCAA) were used for pPICZαB and pGAPZαA, respectively. The reverse primer flanked with *Xba*I restriction site (underlined letters) (5'-3') (GCGCTCTAGATTAGGTTGCAAGCT CGCCAA) was used for both plasmids. Then, both plasmids and insert were digested with *Eco*RI and *Xba*I (New England BioLabs, USA) followed by ligation. The recombinant plasmids were confirmed *via* sequencing followed by BLAST analysis. Then, the plasmids were linearized, and transformed into *P. pastoris* strain GS115 using electroporation method (Genepulser, BioRad) [40].

Expression of Recombinant ARM Lipase in *P. pastoris* Strain GS115

A single colony of the recombinant *P. pastoris* carrying the pPICZαB/ARM and pGAPZαA/ARM were grown in 10 mL of YPD broth at 30 ºC with 250 rpm overnight. Then 1 mL culture of GS115/pPICZαB/ARM was transferred into 50 mL YPTG medium for 24 hours 24 hours followed by replacing YPTG with 50 mL YPTM medium. 1% (v/v) methanol was used to induce the culture every 24 hours interval for 48 hours. Then, the cells were harvested and supernatants were kept for the lipase assay. The growth of transformants constitutively expressing the ARM lipase was carried out in YPD medium (Invitrogen, USA). One (1) mL of GS115/pGAPZαA/ARM overnight culture was inoculated into 100 mL of YPD in 500 mL flask and incubated at 30 ºC with 250 rpm for 48 hours. *Pichia* harboring the empty plasmids (pPICZαB and pGAPZαA) were used as the controls. Next, the culture was spun and the supernatant was collected for the lipase assay. Lipase assay was performed according to the method described by Kwon and Rhee [41] using Tris buffer pH 8.0 at 65 ºC and olive oil was used as a substrate. One unit of lipase activity was defined as the rate of free fatty acid formed in µmole per minute.

RESULTS

Prediction System: LipExPS

LipExPS is a web-based tool to predict recombinant lipase expression in *P. pastoris* (http://103.18.1.10/lipex/index.html). The development of LipExPS was done by implementing BHA for sequence comparison of a query sequence with the available sequences in LipExDB database.

Fig 2(a) and 2(b) show the Graphical User Interface (GUI) of LipExPS. A bacterial lipase from *Geobacillus* sp. strain ARM [38] amino acid sequence (EF042975) was submitted into the query box. The database (Lipase database) and the scoring matrix (BLOSUM62) were selected and then the 'Run' button was clicked to submit the query (Fig 2a). After a few seconds, the results of the sequence comparison were displayed as shown in Fig 2(b). The result page consisted of several parameters which included an auto generated "Query ID, a "Job Title" provided by the user (here we used ARM lipase), the "Query Length" which was the total length of the submitted query sequence and the "Scoring matrix" as chosen on the query page. Other statistical parameters such as "score", "E-value", "identities" and "Target Id" were displayed in the result page. Here, the link was provided to take the user to the query page of LipExDB, as shown in Fig 3(a).

Figure 2a: Query page of LipExPS. The query submitted was lipase protein sequence and the scoring matrix selected was BLOSUM62

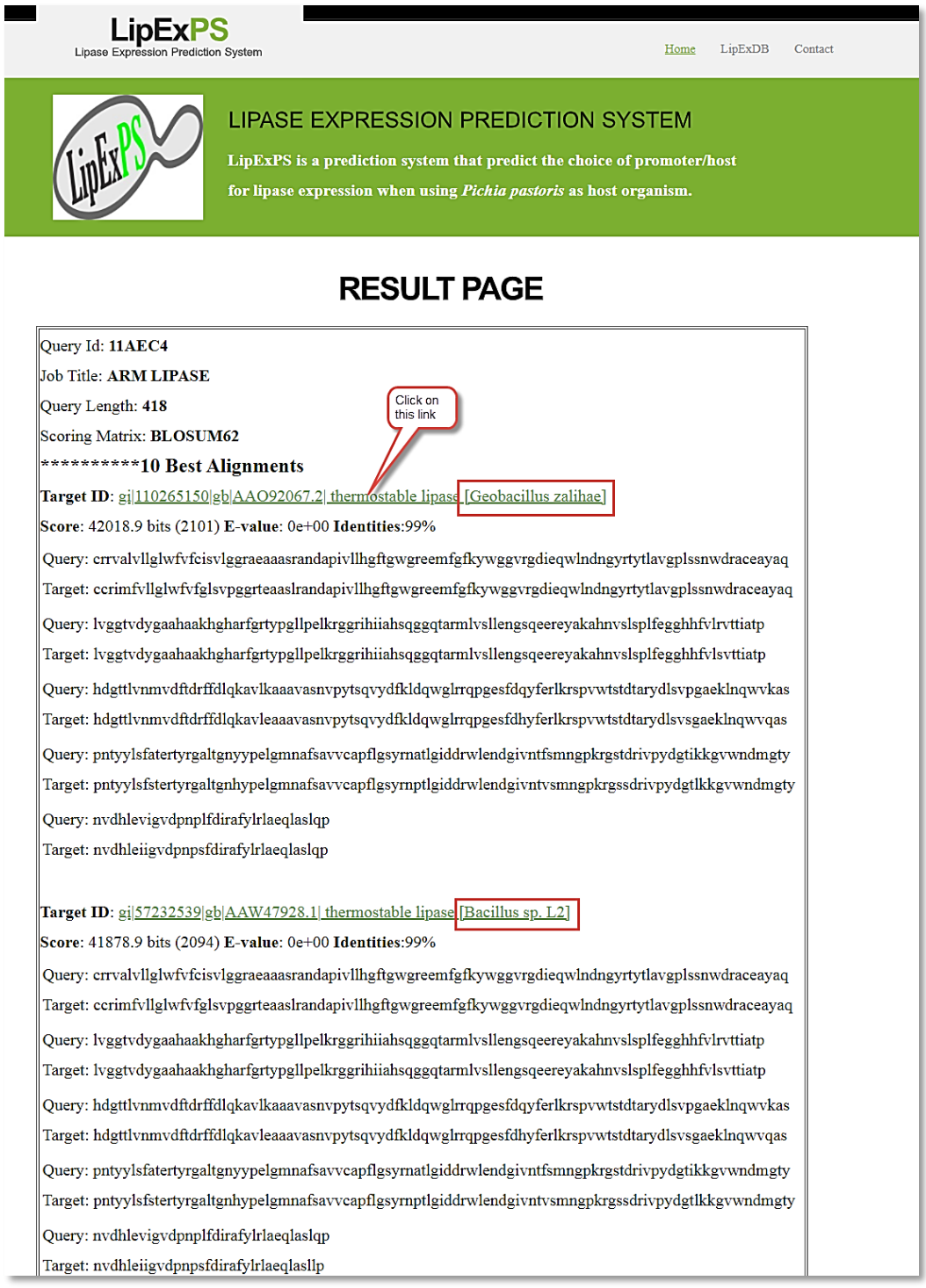
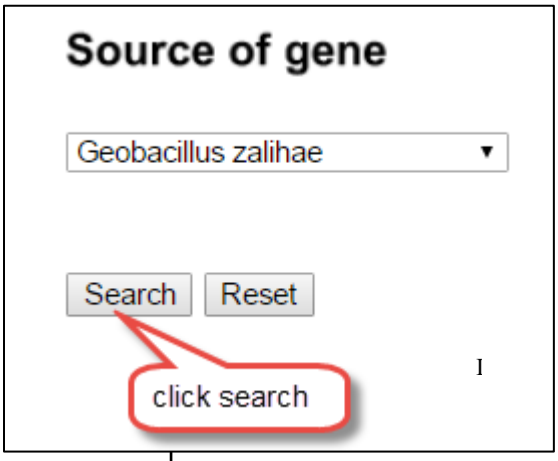


Figure 2b: Result page of LipExPS. This page displays the query id, job title, query length, scoring matrix, target id and the alignments results which include other statistical parameters such as e-value, score and identities.

Fig 3(a) and 3(b) show the query and result pages of LipExDB, respectively. The name of the Target IDs was used as represented by the rectangle box in Fig 2(b) for the first and second hits: *Geobacillus zalihae* and *Bacillus* sp. L2, respectively. These organisms were used to search the information in LipExDB (Fig 3a). Clicking on the search button would return the result page of LipExDB (Fig 3b). The red boxes represented the promoters and strains used to clone and express the lipases similar to ARM lipase.



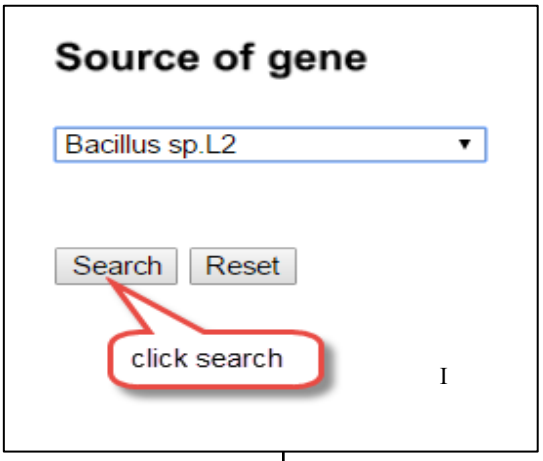


Figure 3a: Query page of LipExDB

Promoter/Strain	AOX	FLD	GAP
GS115	33.33 %	0 %	0%
X-33	33.33 %	0 %	0 %
KM71H	33.33 %	0 %	0 %
SMD1168	0 %	0 %	0 %
SMD1168H	0 %	0 %	0 %
KM71	0 %	0 %	0 %

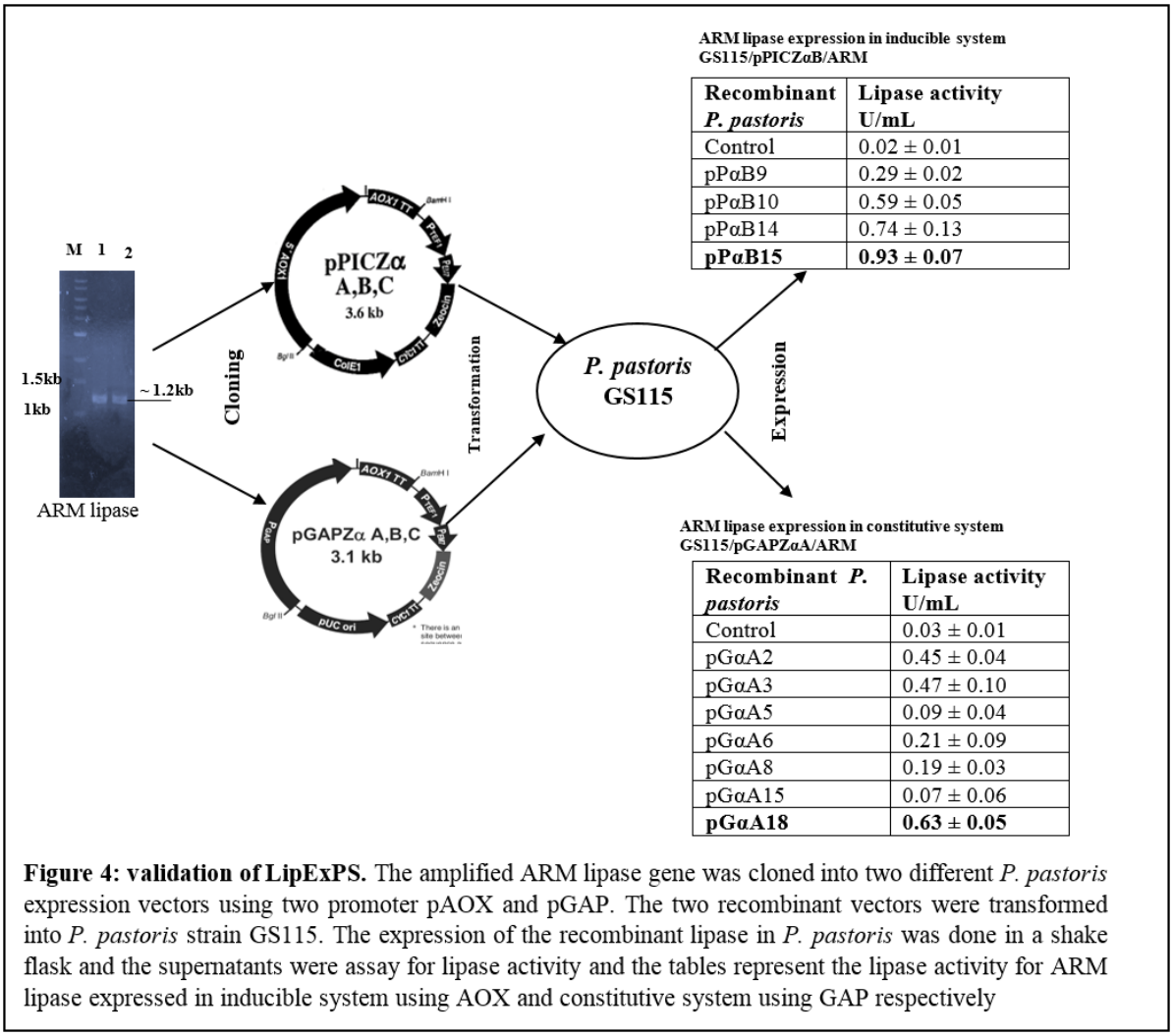
Promoter/Strain	AOX	FLD	GAP
GS115	66.67 %	0 %	33.33%
X-33	0 %	0 %	0 %
KM71H	0 %	0 %	0 %
SMD1168	0 %	0 %	0 %
SMD1168H	0 %	0 %	0 %
KM71	0 %	0 %	0 %

Figure 3b: Result page of LipExDB

Figure 3: Query and result pages of LipExDB after user clicked on the target id link

LipExPS validation

Fig 4 shows the validation process of LipExPS. The amplified ARM lipase was cloned into pPICZαB and pGAPZαA. The plasmids and strain were selected based on the predicted result from LipExPS and LipExDB (Fig 2b and 3b). The tables in Fig 4 show the lipase activity for GS115/pPICZαB/ARM and GS115/pGAPZαA/ARM. pPαB15 and pGαA18 have shown the highest lipase activities in GS115/pPICZαB/ARM and GS115/pGAPZαA/ARM, respectively. It demonstrated that the inducible expression system containing the AOX promoter resulted in higher lipase activity than the constitutive GAP promoter.



ARM lipase expression in inducible system GS115/pPICZαB/ARM

Recombinant *P. pastoris*	Lipase activity U/mL
Control	0.02 ± 0.01
pPαB9	0.29 ± 0.02
pPαB10	0.59 ± 0.05
pPαB14	0.74 ± 0.13
pPαB15	**0.93 ± 0.07**

ARM lipase expression in constitutive system GS115/pGAPZαA/ARM

Recombinant *P. pastoris*	Lipase activity U/mL
Control	0.03 ± 0.01
pGαA2	0.45 ± 0.04
pGαA3	0.47 ± 0.10
pGαA5	0.09 ± 0.04
pGαA6	0.21 ± 0.09
pGαA8	0.19 ± 0.03
pGαA15	0.07 ± 0.06
pGαA18	**0.63 ± 0.05**

Figure 4: validation of LipExPS. The amplified ARM lipase gene was cloned into two different *P. pastoris* expression vectors using two promoter pAOX and pGAP. The two recombinant vectors were transformed into *P. pastoris* strain GS115. The expression of the recombinant lipase in *P. pastoris* was done in a shake flask and the supernatants were assay for lipase activity and the tables represent the lipase activity for ARM lipase expressed in inducible system using AOX and constitutive system using GAP respectively

LipExDB Database

Journal articles (150) from various databases were searched with the respective keywords, but only 42 reports on lipase expression with 20 lipase genes were found. The UniProtKB accession numbers of the curated lipases as reported by published results were AF229435, AF073953, AJ012632, AJ320260, AY260764, AY855077, DQ831123, FJ536288, M93283, M93284, P19515, P32948, P41365, S65092, X64703, X64704, X78032, X95309, and Z30645. The data curated were defined in Table 1 and classified into structured and unstructured data before storing in the database. The structured data was stored in a relational model and the unstructured data was stored using a flat file model. This information was documented in the database. The link for LipExDB is provided in http://kbioimage.my/lipex/indexDB.php. Fig 5 shows the GUI of LipExDB. This page contained a navigation bar with "Home", "Prediction", "Search", "Contact" and side bar with Frequently Asked Questions (FAQs) such as "*What is a gene*?", "*What is a promoter*?", "*What is strain*?", "*What is a vector*?", "*What is Pichia pastoris*?" and "*What is yeast*?".

Table 1: Structured and data definition of LipExDB.

Table	Attributes	Description
Gene	Id Name Symbol	Represents the gene of interest
Promoter	Id Name Symbol	Enables the expression of the gene in the host
Publication	Id Type Title Author Year Journal	Represents the source of the journal
Sample	Id Gene_id Accession_number Sequence Type Source_id Characteristics Publication_id	Links gene, publication and source tables together
Spvs	Id Sample_id Promoter_id Vector_id Strain_id Yield	Links sample-, promoter, vector and strain tables together and it also contains the yield (level of expression)
Source	Id Species_name Genus Authorship Year Family Class Phylum Kingdom Sorder	Represents the organism where the gene was isolated from
Strain	Id Name	Represents the host organism which expressed the lipase
Vector	Id Name	Represents the carrier of the cloned gene into the host strain

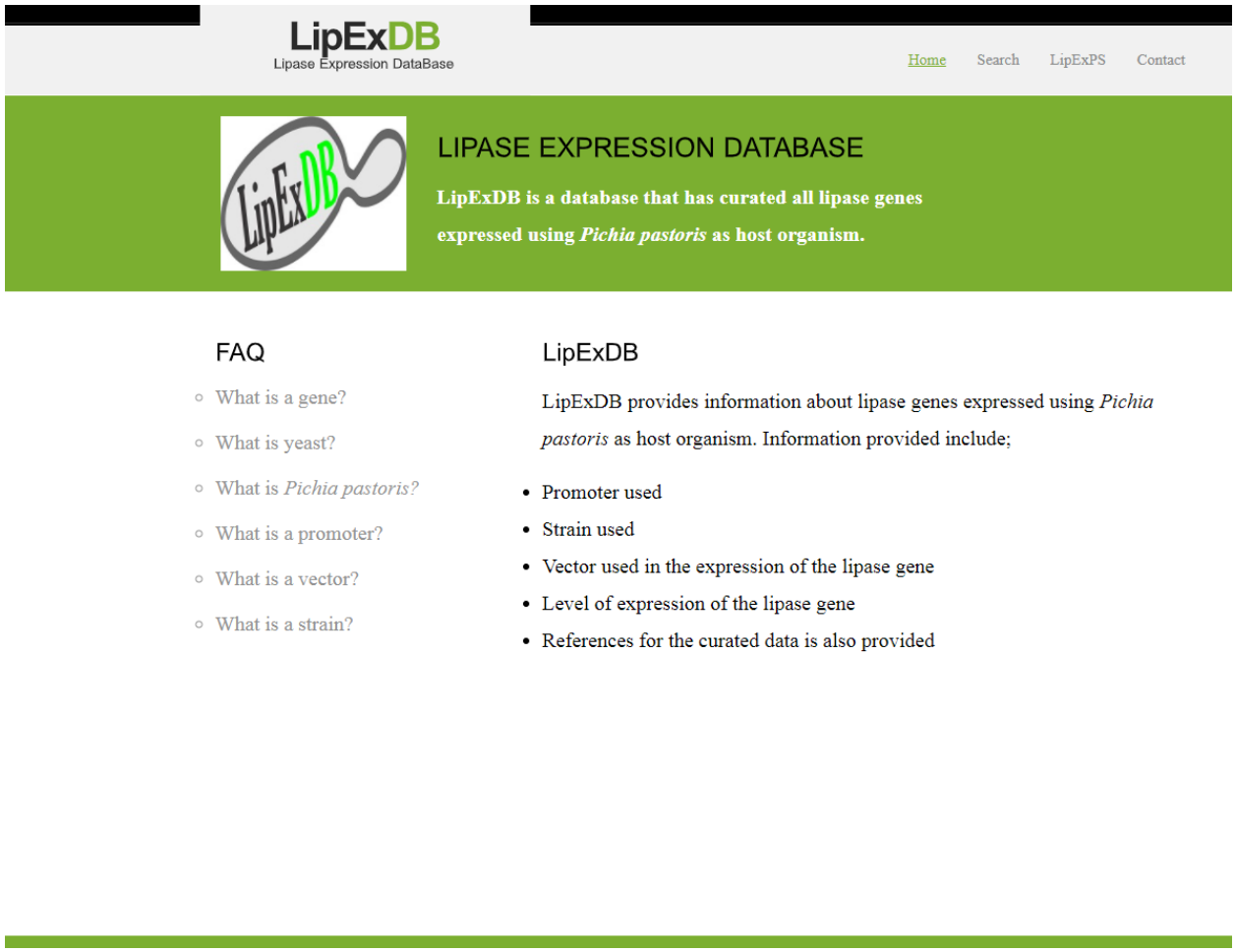


Figure 5: Graphical user interface of LipExDB. This page has four main menus which are home, prediction, search and contact. The prediction menu links the database to the prediction page. The search menu is use to query the database. Contact menu is for enquiry and comments.

DISCUSSION

The LipExPS is a web-based tool to predict recombinant lipase expression in *P. pastoris*. This website performs a search against the curated LipExDB using a BLAST heuristic algorithm to find similar matches to the new submitted lipase query sequence. The result page displayed the query sequence aligned with curated sequences in LipExDB. Other statistical results were also returned in order to justify the validity of the returned alignments. The statistical results such as score, E-value and identity were important to show the strength of the alignment (Fig 2b). The alignment was considered valid and acceptable if the given score was above 1500 with 0 E-value and the identity above 75% [33]. If the returned alignment result met the above requirement, the user can click on the target id of the returned alignment to get more information about the promoter/host to be used to clone and express the query lipase (Fig 3). On the other hand, a query sequence cannot be shorter than 20 amino acids as shorter queries would affect the accuracy of the alignment but in a case where it was shorter than 20 amino acids users should focus more on the identity. LipExPS has a simple interface and user query fields. A user guide on how to use LipExPS was also provided on the website.

The BLAST algorithm used in this study was a common algorithm that has been used in [42], [43]. This algorithm was implemented for binary prediction in previous studies and primarily for searching sequence databases [44]. By contrast, in this study, the search was performed based on curated lipase expression data. A prediction was made based on the similarity of the sequences, providing both a promoter and a host strain to be used in the cloning of a new lipase gene.

ARM lipase was chosen to validate the LipExPS prediction system experimentally because this lipase has

not been expressed in *P. pastoris*. The ARM lipase gene was successfully cloned into two different plasmids (pPICZαB and pGAPZαA). As predicted by LipExPS, ARM lipase was productively expressed in *P. pastoris* strain GS115 using AOX and GAP promoters resulting in the highest level of expression at 0.93U/mL and 0.63U/mL, respectively after 48 hours. This un-optimized result proved that LipExPS can predict the expression of recombinant lipase in *P. pastoris*. Similar trends were observed for T1 lipase [36] and L2 lipase [45] where the AOX promoter gave higher expression levels than the GAP promoter.

LipExDB was developed to store lipase data, perform data management and retrieve data. GUI of LipExDB contained few functions for data retrieving and record management. Users can communicate interactively with LipExDB through the user-friendly GUI provided. Data currently presented in LipExDB were specific to lipases expressed in *P. pastoris.* LipExDB can provide users with information for each lipase expressed in P. pastoris, such as the type of lipase, accession number, optimal pH and temperature, promoter and vector used in the cloning process *P. pastoris* strain used in the expression of the particular lipase. Only 20 lipases were present in this database. Even though many attempts and keywords used, only a few lipases have been successfully expressed in *P. pastoris*. In general, researchers did not report the negative result where their attempts to express the particular lipases in *P. pastoris* failed. LipExDB has provided the users with the opportunity of knowing what gene, promoter, vector, yeast, *P. pastoris* and strain are. It also has three drop-down lists where the users can select to retrieve the information about the reported lipase.

Curating is an ongoing process, so as new lipase genes are expressed and articles are published, LipExDB database will be updated. So far, the study focused on lipase expression in *P. pastoris* data. This study can be further improved to retrieve data from other hydrolytic enzymes or proteins, which will be deposited in the future. This platform and strategy could be extended to document and predict other recombinant proteins in future use.

ACKNOWLEDGMENT

This study was supported by the Universiti Putra Malaysia, GP-IPM - Geran Putra Inisiatif Putra Muda (grant no. 9426600) which was awarded to the fourth author.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest regarding the publication of this manuscript.

REFERENCES

1. Jallouli, R., Ali, M.B. and Charfeddine, M. (2016) Heterologous overexpression and biochemical characterization of the (galactophospho) lipase from *Fusarium solani* in *Pichia pastoris* that is expressed in planta. *Int. J. Biol. Macromol.* **84**,94–100.

2. Jallouli, R., Parsiegla, G. and Carrire, F. (2017) Efficient heterologous expression of *Fusarium solani* lipase, FSL2, in *Pichia pastoris*, functional characterization of the recombinant enzyme and molecular modeling. *Int. J. Biol. Macromol.* **94**, 61–71.

3. Eom, G.T., Lee, S.H. and Song, B.K. (2013) High-level extracellular production and characterization of *Candida antarctica* lipase B in *Pichia pastoris*. *J. Biosci. Bioeng*. **116**,165–70.

4. Ray, A. (2012) Application of Lipase in Industry. *Asian J. Pharm. Technol.* **2**,33–37.

5. Li, S., Jendresen, C.B. and Grnberger, A. (2016) Enhanced protein and biochemical production using CRISPRi-based growth switches. *Metab. Eng*. **38**,274–284.

6. Jin, Z., Han, S.Y., Zhang, L. (2013) Combined utilization of lipase-displaying *Pichia pastoris* whole-cell biocatalysts to improve biodiesel production in co-solvent media. *Bioresour. Technol.* **130**,102–109.

7. Cruz-Ramírez, M.G., Rivera-Ríos, J.M. and Téllez-Jurado, A. (2012) Screening for thermotolerant ligninolytic fungi with laccase, lipase, and protease activity isolated in Mexico. *J. Environ. Manage.* **95**,S256–S259.

8. Zhang, X., Li, X. and Xia, L. (2015) Expression of a thermo-alkaline lipase gene from Talaromyces thermophilus in recombinant *Pichia pastoris*. *Biochem. Eng. J.* **103**,263–269.

9. Robert, J.M., Lattari, F.S. and Machado, A.C. (2017) Production of recombinant lipase B from *Candida antarctica* in *Pichia pastoris* under control of the promoter PGK using crude glycerol from biodiesel production as carbon source. *Biochem. Eng. J.* **118**,123–131.

10. Dreos, R., Ambrosini, G. and Cavin Périer, R. (2014) The Eukaryotic Promoter Database: expansion of EPDnew and new promoter analysis tools. *Nucleic Acids Res*. 1–5.

11. Jiang, C., Xuan, Z. and Zhao, F. (2007) TRED: A transcriptional regulatory element database, new entries and other development. *Nucleic Acids Res.* **35**,140–143.

12. Friard, O., Re, A. and Taverna, D. (2010) CircuitsDB: a database of mixed microRNA/transcription factor feed-forward regulatory circuits in human and mouse. *BMC Bioinformatics*. 11,435.

13. Matys, V. (2006) TRANSFAC(R) and its module TRANSCompel(R): transcriptional gene regulation in eukaryotes. *Nucleic Acids Res*. **34**,D108–D110

14. Liu, B., Liu, F. and Wang, X. (2015) Pse-in-One: A web server for generating various modes of pseudo components of DNA, RNA, and protein sequences. *Nucleic Acids Res.* **43**,W65–W71.

15. Mathelier, A., Zhao, X. and Zhang, A.W. (2014) JASPAR 2014: An extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **42**,142–147

16. Zhu, J. and Zhang, M.Q. (1999) SCPD: a promoter database of the yeast *Saccharomyces cerevisiae*. *Bioinformatics*. **15**,607–611.

17. Chang, D.T.H., Huang, C.Y. and Wu, C.Y. (2011) YPA: An integrated repository of promoter features in *Saccharomyces cerevisiae*. *Nucleic Acids Res*. **39**,647–652

18. Teixeira, M.C., Monteiro, P.T. and Guerreiro, J.F. (2014) The YEASTRACT database: An upgraded information system for the analysis of gene and genomic transcription regulation in *Saccharomyces cerevisiae*. *Nucleic Acids Res*. **42**,161–166.

19. Wu, W.S., Hsieh, Y.C. and Lai, F.J. (2016) YCRD: Yeast combinatorial regulation database. *PLoS One*. **11**,1–12.

20. Liu, Y., Schmidt, B. and Maskell, D.L. (2010) CUDASW++2.0: enhanced Smith-Waterman protein database search on CUDA-enabled GPUs based on SIMT and virtualized SIMD abstractions. *BMC Res. Notes.* **3**,93.

21. Li, W., McWilliam, H. and Goujon, M. (2012) PSI-Search: Iterative HOE-reduced profile SSEARCH searching. *Bioinformatics.* **28**,1650–1651.

22. Alva, V., Nam, S.Z. Söding, J. (2016) The MPI bioinformatics Toolkit as an integrative platform for advanced protein sequence and structure analysis. *Nucleic Acids Res.* **44**,34.

23. Altschul, S.F., Gish, W., Miller, W., et el. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**,403–10.

24. Altschul, S.F., Madden, T.L. and Schäffer, A.A. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402.

25. Oehmen, C.S. and Baxter, D.J. (2013) ScalaBLAST 2.0: Rapid and robust BLAST calculations on multiprocessor systems. *Bioinformatics*. **29**,797–798.

26. Karplus, K. (2009) SAM-T08:HMM-based protein structure prediction. *Nucleic Acids Res.* **37**,492–497.

27. Liu, Y., Tran, T.T. and Lauenroth, F. (2014) SWAPHI-LS: Smith-Waterman Algorithm on Xeon Phi coprocessors for Long DNA Sequences. *2014 IEEE Int. Conf. Clust. Comput. Clust.* 2014,257–265.

28. Rognes, T. (2011) Faster Smith-Waterman database searches with inter-sequence SIMD parallelisation. *BMC Bioinformatics*. **12**,221.

29. Bell, R.J.A., Rube, H.T. and Kreig, A. (2015) The transcription factor GABP selectively binds and activates the mutant TERT promoter in cancer. *Science.* (80) **348**,1036–1039.

30. Witalison, E., Thompson, P. and Hofseth, L. (2015) HHS Public Access. *Curr. Drug Targets.* **16**,700–710.

31. Franco-Zorrilla, J.M., López-Vidriero, I. and Carrasco, J.L. (2014) DNA-binding specificities of plant transcription factors and their potential to define target genes. *Proc. Natl. Acad. Sci.* **111**, 2367–2372.

32. Jayaram, N., Usvyat, D. and Martin, A.C. (2016) Evaluating tools for transcription factor binding site prediction. *BMC Bioinformatics*. 1–12.

33. Dwyer, R.A. (2003) *Genomic perl: From bioinformatics basics to working code*. Vol. 1. Cambridge University Press.

34. GenBank [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology information; [1982]-2016.

35. PubMed [Internet]. Bethesda (MD) National Library of Medicine (US). [1946] - 2017.

36. Oslan, N.S., Salleh, A.B and Rahman, R.N.Z.R.A. (2014) *Pichia pastoris* as a host to overexpress the thermostable T1 lipase from *Geobacillus zalihae*. *GSTF J. Biosci.* 3.

37. Oslan, N.S., Salleh, A.B. and Rahman, R.N.Z.R.A. (2015) "A Newly Isolated Yeast as an Expression Host for Recombinant Lipase." *Cellular and Molecular Biology Letters* **20**(2): 279–93.

38. Ebrahimpour, A., Rahman, R.N.Z.R.A., Basri, M, et al. (2011) High level expression and characterization of a novel thermostable, organic solvent tolerant, 1,3-regioselective lipase from *Geobacillus* sp. strain ARM. *Bioresour. Technol*. **102**,6972–6981.

39. Sambrook, J., Fritsch E.F. and Maniatis T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

40. Lin-Cereghino, J., Wong, W.W., Xiong, S. et al. (2005) Condensed protocol for competent cell preparation and transformation of the methylotrophic yeast *Pichia pastoris*. *Biotechniques*. **38,**44–48.

41. Kwon, D.Y. and Rhee, J.S. (1986) A simple and rapid colorimetric method for determination of free fatty acids for lipase assay. *J. Am. Oil Chem. Soc* 6**3**,89–92.

42. Drozdetskiy, A., Cole, C., Procter, J., et al. (2015) JPred4: A protein secondary structure prediction server. *Nucleic Acids Res*. **43**,W389–W394.

43. Szcześniak, M.W., Kabza, M. and Karolak, J.A. (2017) KTCNlncDB-a first platform to investigate lncRNAs expressed in human keratoconus and non-keratoconus corneas. *Database (Oxford)*. 2017,1–6.

44. Zhang, B.J., Gao, H., Chai, Z. and Yang, G. (2016) Identification of DNA-binding proteins using multi-features fusion and binary firefly optimization algorithm. *BMC Bioinformatics*. vol. 17, 1–12.

45. Sabri, S., Rahman, R.N.Z.R.A., Leow, T.C., et al. (2009) Secretory expression and characterization of a highly Ca^{2+}-activated thermostable L2 lipase. *Protein Expr Purif.* **68**,161–6.