UNIVERSITI PUTRA MALAYSIA

DEVELOPMENT AND VALIDATION OF PROMOTER PREDICTION SYSTEM FOR RECOMBINANT LIPASE EXPRESSION IN Pichia pastoris

MAYAKI FATIMA GOGO

FBSB 2017 13

DEVELOPMENT AND VALIDATION OF PROMOTER PREDICTION SYSTEM FOR RECOMBINANT LIPASE EXPRESSION IN *Pichia pastoris*

By

MAYAKI FATIMA GOGO

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirements for the Degree of Master of Science

April 2017

© COPYRIGHT UPM

COPYRIGHT

All material contained within the thesis, including without limitation text, logos, icons, photographs and all other artwork, is copyright material of Universiti Putra Malaysia unless otherwise stated. Use may be made of any material contained within the thesis for non-commercial purposes from the copyright holder. Commercial use of material may only be made with the express, prior, written permission of Universiti Putra Malaysia.

Copyright © Universiti Putra Malaysia

© COPYRIGHT UPM

DEDICATION

TO MY LATE MOTHER

© COPYRIGHT UPM

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the degree of Master of Science

DEVELOPMENT AND VALIDATION OF PROMOTER PREDICTION SYSTEM FOR RECOMBINANT LIPASE EXPRESSION IN *Pichia pastoris*

By

MAYAKI FATIMA GOGO

April 2017

Chairman : Siti Nurbaya Oslan, PhD
Faculty : Biotechnology and Biomolecular Sciences

One of the most reliable methods to improve the heterologous protein production is by conducting molecular cloning and expression of a targeted gene in an organism. This process requires extensive assessment on different promoters and hosts resulting in wastage of time, energy and resources. BLAST heuristic algorithm can predict lipase expression in *Pichia pastoris.* This study aims to develop a promoter prediction system for recombinant lipase expression in commonly used yeast expression system, *P. pastoris* using BLAST heuristic algorithm. Published experimental results of lipase expressed in *P. pastoris* were curated manually from publicly available databases and used for lipase database development. Two databases were developed, to store, organize and manage the curated data using MySQL and protein sequences in FASTA format using flat file. As for prediction, a BLAST heuristic method was implemented for sequence alignment and similarity comparison using Perl scripting language. The prediction system consists of two modules; Lipase Expression Database (LipExDB) and Lipase Expression Prediction System (LipExPS). LipExDB is a module for storing curated data and providing users with interface to query and retrieve information; and LipExPS is a module for users to upload the query sequence and retrieve the results of sequence alignment and other statistical parameters such as percentage of similarity and e-value. A user interface was provided for LipExPS. ARM lipase was used to validate the LipExPS. The predicted result of ARM lipase was used to clone into pGAPZαA and pPICZαB vectors followed by transformation into *P. pastoris*. The expression of ARM lipase was monitored in *P. pastoris* strain GS115. Based on the experimental validation, ARM lipase was successfully expressed in *P. pastoris* strain GS115 using GAP and AOX promoters as predicted by LipExPS. LipExPS was able to address the problem of promoter/host choice in lipase expression when using *P. pastoris* as a host by predicting the promoters that can be used to express the gene submitted to LipExPS. LipExDB would provide researchers with more information on lipase expression in *P. pastoris*. The development of this prediction system would facilitate the expression of lipase gene in *P. pastoris* thus saving resources, time and energy.

© COPYRIGHT UPM

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia dalam memenuhi keperluan untuk Ijazah Sarjana Sains

PEMBANGUNAN DAN PENGESAHAN SISTEM RAMALAN PROMOTER UNTUK PENGEKSPRESAN LIPASE REKOMBINAN DI *Pichia pastoris*

Oleh

MAYAKI FATIMA GOGO

April 2017

Pengerusi : Siti Nurbaya Oslan, PhD
Fakulti : Bioteknologi dan Sains Biomolekul

Salah satu kaedah yang boleh digunakan untuk meningkatkan pengeluaran protein heterologus adalah dengan menjalankan pengklonan dan ekspresi molekular gen yang dikehendaki di dalam organisma. Proses ini memerlukan pemilihan menyeluruh promoter dan hos yang berbeza, di mana akan mengakibatkan pembaziran masa, tenaga dan sumber. Algoritma heuristik BLAST boleh meramalkan ekspresi lipase dalam *Pichia pastoris*. Kajian ini bertujuan untuk membangunkan sistem ramalan promoter untuk mengekspreskan lipase rekombinan dalam sistem ekspresi yis *P. pastoris* menggunakan algoritma heuristik BLAST. Hasil eksperimen yang telah diterbitkan dalam pangkalan data umum, dikumpulkan secara manual dan digunakan untuk pembangunan pangkalan data lipase. Dua pangkalan data telah dibangunkan untuk menyimpan, menyusun dan menguruskan data yang dikumpul menggunakan MySQL dan jujukan protein dalam format FASTA menggunakan fail rata. Bagi pembangunan sistem ramalan, kaedah heuristik BLAST yang telah dilaksanakan untuk penjajaran jujukan protein dan perbandingan persamaan menggunakan bahasa pengatucaraan Perl. Sistem ramalan ini terdiri daripada dua modul; pangkalan data ekspresi lipase (LipExDB) dan sistem ramalan ekspresi lipase (LipExPS). LipExDB adalah modul untuk menyimpan data yang terkumpul disamping menyediakan antara muka pengguna untuk pertanyaan dan mendapatkan maklumat; manakala LipExPS adalah modul untuk pengguna memuatnaik jujukan pertanyaan dan mendapatkan keputusan penjajaran jujukan berserta parameter statistik yang lain seperti peratus persamaan dan e-nilai. Antara muka pengguna telah disediakan bagi kegunaan LipExPS. ARM lipase telah dipilih untuk mengesahkan sistem LipExPS. Hasil ramalan telah digunakan untuk proses mengklonkan ARM lipase ke dalam pGAPZαA dan pPICZαB vektor, diikuti dengan transformasi ke dalam *P. pastoris*. Ekspresi ARM lipase dipantau dalam *P. pastoris* GS115. Berdasarkan pengesahan eksperimen itu, ARM lipase berjaya diekspreskan di dalam *P. pastoris* GS115 menggunakan promoter GAP dan AOX seperti yang diramalkan oleh LipExPS. LipExPS berjaya menangani masalah pilihan promoter/hos dalam ekspresi lipase di dalam *P. pastoris*. Sistem ramalan promoter (LipExPS) ini boleh digunakan dalam pengekspresan gen di dalam

© COPYRIGHT UPM

P. pastoris. LipExDB juga akan memberikan maklumat lebih lanjut mengenai ekspresi lipase dalam *P. pastoris*. Pembangunan sistem ramalan ini akan memudahkan pengekspresan lipase dalam *P. pastoris* sekaligus menjimatkan sumber, masa dan tenaga.

© COPYRIGHT UPM

ACKNOWLEDGEMENTS

Alhamdulillah, all glory and thanks be to Almighty Allah for his mercy and blessings all through this thesis. A lot of people have contributed to the completion of this thesis.

First, I would like to thank my supervisor Dr. Siti Nurbaya Oslan for her encouragement, advice, guidance and support all through my research work. I appreciate the effort of my co-supervisors: Prof. Dato' Dr. Abu Bakar Salleh and Dr. Arpah Abu for their support, time, and patience towards the completion of this study. Also my appreciation goes to Enzyme and Technology (EMTech) Research group for their support all through my study. Thanks to my lab mates IBS and DG4 (UM).

To my late mother (late Haj Fatima N Mayaki) who saw the start of this study but not the end. Thanks for your support, love, and prayers to see me succeed and may Allah grant her Al-janatual Firdausi (Amin).

To my father Alh S. N Mayaki, my siblings (Jummai, Isah, Mohammed, Amina and Abubakar), relatives and friends thanks for your support, love and care.

Finally, I will like to appricate Tertiary Education Trust Fund (TETFund) and Ibrahim Badamasi Babangida University Lapai for their financial assistance.

© COPYRIGHT UPM

I certify that a Thesis Examination Committee has met on 4 April 2017 to conduct the final examination of Mayaki Fatima Gogo on her thesis entitled "Development and Validation of Promoter Prediction System for Recombinant Lipase Expression in *Pichia pastoris*" in accordance with the Universities and University Colleges Act 1971 and the Constitution of the Universiti Putra Malaysia [P.U.(A) 106] 15 March 1998. The Committee recommends that the student be awarded the Master of Science.

Members of the Thesis Examination Committee were as follows:

Umi Kalsom binti Md Shah, PhD
Associate Professor
Faculty of Biotechnology and Biomolecular Sciences
Universiti Putra Malaysia
(Chairman)

Noor Azmi Shaharuddin, PhD
Senior Lecturer
Faculty of Biotechnology and Biomolecular Sciences
Universiti Putra Malaysia
(Internal Examiner)

Roswanira binti Ab. Wahab, PhD
Senior Lecturer
Universiti Teknologi Malaysia
Malaysia
(External Examiner)

NOR AINI AB. SHUKOR, PhD
Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 6 July 2017

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Master of Science. The members of the Supervisory Committee were as follows:

Siti Nurbaya Oslan, PhD
Senior Lecturer
Faculty of Biotechnology and Biomolecular Sciences
Universiti Putra Malaysia
(Chairman)

Abu Bakar Salleh, PhD
Professor
Faculty of Biotechnology and Biomolecular Sciences
Universiti Putra Malaysia
(Member)

Arpah Abu, PhD
Senior Lecturer
Faculty of Science
University of Malaya
(Member)

_____
ROBIAH BINTI YUNUS, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

© COPYRIGHT UPM

Declaration by graduate student

I hereby confirm that:

- this thesis is my original work;
- quotations, illustrations and citations have been duly referenced;
- this thesis has not been submitted previously or concurrently for any other
- degree at any other institutions;
- intellectual property from the thesis and copyright of thesis are fully-owned by Universiti Putra Malaysia, as according to the Universiti Putra Malaysia (Research) Rules 2012;
- written permission must be obtained from supervisor and the office of Deputy Vice-Chancellor (Research and Innovation) before thesis is published (in the form of written, printed or in electronic form) including books, journals, modules, proceedings, popular writings, seminar papers, manuscripts, posters, reports, lecture notes, learning modules or any other materials as stated in the Universiti Putra Malaysia (Research) Rules 2012;
- there is no plagiarism or data falsification/fabrication in the thesis, and scholarly integrity is upheld as according to the Universiti PutraMalaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) and the Universiti Putra Malaysia (Research) Rules 2012. The thesis has undergone plagiarism detection software.

Signature: _____ Date: _____

Name and Matric No: Mayaki Fatima Gogo , GS41883

© COPYRIGHT UPM

Declaration by Members of Supervisory Committee

This is to confirm that:

- the research conducted and the writing of this thesis was under our supervision;
- supervision responsibilities as stated in the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) were adhered to

Signature: _____
Name of Chairman of Supervisory Committee: Dr. Siti Nurbaya Oslan

Signature: _____
Name of Member of Supervisory Committee: Professor Dr. Abu Bakar Salleh

Signature: _____
Name of Member of Supervisory Committee: Dr. Arpah Abu

© COPYRIGHT UPM

TABLE OF CONTENTS

		Page
ABSTRACT		i
ABSTRAK		ii
ACKNOWLEDGEMENTS		iv
APPROVAL		vi
DECLERATION		vii
LIST OF TABLES		xi
LIST OF FIGURES		xii
LIST OF ABBREVIATIONS		xiv
CHAPTER		
1	**INTRODUCTION**	1
2	**LITERATURE REVIEW**	3
2.1	Computational prediction system	3
2.1.1	Types of prediction system	4
2.1.2	Framework of a prediction system	4
2.2	Prediction algorithms	5
2.2.1	Needleman-Wunsch (NW) algorithm	6
2.2.2	Smith-Waterman (SW) algorithm	6
2.2.3	BLAST heuristic algorithm	6
2.3	Graphical User Interface (GUI)	9
2.4	Database	10
2.4.1	Biological Database	10
2.4.2	Data modelling	12
2.4.3	Database Management System (DBMS)	13
2.5	Yeast expression system	14
2.6	*Pichia pastoris*	14
2.7	Promoter in *Pichia pastoris*	15
2.8	Lipase	16
3	**MATERIALS AND METHODS**	17
3.1	General workflow of LipExPS	17
3.2	Development environment settings	17
3.2.1	Hardware	17
3.2.2	Software	18
3.3	Lipase Expression Database (LipExDB) development	19
3.3.1	Data collection	20
3.3.2	Designing database schema	20
3.3.3	Information retrieval	21
3.3.4	Graphical User interface (GUI) design	22
3.4	Prediction System	25
3.4.1	Algorithm	25
3.5	Validation of the LipExPS	26

© COPYRIGHT UPM

3.5.1 Prediction of ARM lipase by LipExPS 26
3.5.2 Cloning of ARM lipase gene in *P. pastoris* expression system 26
3.5.3 Sequencing of recombinant ARM lipase gene 29
3.5.4 Transformation of recombinant plasmids into *P. pastoris* 30
3.5.5 Direct PCR analysis of the Pichia transformants 30
3.5.6 Screening of lipase activity 30
3.5.7 Protein expression in *P. pastoris* strain GS115 31

4 RESULTS AND DISCUSSION 32
4.1 Functional system requirement 32
4.2 Non-functional system requirement 32
4.3 LipExPS architecture 32
4.4 LipExDB schema 32
4.4.1 Data annotation 35
4.5 The Database – LipExDB 37
4.6 Lipase Expression Database (LipExDB) 37
4.6.1 Query data 38
4.6.2 Retrieve data 38
4.7 Lipase Expression Prediction System (LipExPS) 42
4.7.1 Query page of LipExPS 42
4.7.2 Result page of LipExPS 42
4.8 Validation of the LipExPS 46
4.8.1 Cloning of ARM lipase gene in *P. pastoris* expression system 46
4.8.2 Transformation into *E. coli* strain TOP10 48
4.9 Transformation and selection of recombinant *P. pastoris* 52
4.10 Screening of lipase activity 53
4.11 Expression of ARM lipase transformants in *P. pastoris* 53
4.11.1 Expression under inducible alcohol oxidase (*AOX1*) promoter 53
4.11.2 Expression under constitutive glyceraldehyde- 3-phosphate dehydrogenase (*GAP*) promoter 55
4.12 Evaluation of LipExPS 56

5 CONCLUSION AND RECOMMENDATION FOR FUTURE RESEARCH 57
5.1 Conclusion 57
5.2 Recommendation for future research 57

REFERENCES 59
APPENDICES 74
BIODATA OF STUDENT 114
PUBLICATION 115

© COPYRIGHT UPM

LIST OF TABLES

Table		Page
1	Existing prediction systems in molecular biology	3
2	Boolean operators	5
3	Existing software for sequence similarity search	8
4	Existing Biological Databases Systems	11
5	Server- and Client-side hardware tools	17
6	Software's used in the development process	18
7	Sequencing primers	29
8	Structured vocabularies of LipExDB	35
9	Examples of naming examples	36
10	ARM lipase expression in inducible system	55
11	ARM lipase expression in constitutive system	56

© COPYRIGHT UPM

LIST OF FIGURES

Figure		Page
1	Prediction system framework	4
2	General workflow of LipExPS	18
3	Framework of LipExDB development	19
4	LipExDB Entity Relational (ER) diagram	21
5	LipExDB information retrieval framework	23
6	LipExPS information retrieval framework	24
7	LipExPS flow chart	26
8	LipExPS architecture	33
9	LipExDB schema	34
10	Lipase database (LipExDB) MySQL	37
11	Home page of LipExDB	38
12	Contact page of LipExDB	39
13	Query page of LipExDB	40
14	Result page of LipExDB	41
15	Rooted phylogenetic tree of LipExDB sequences	43
16	Query page of LipExPS	44
17	Result page of LipExPS	45
18	Submission of ARM lipase protein sequence	47
19	Prediction result page of ARM lipase query	48
20	ARM lipase prediction result after clicking of the target id	49
21	Gel electrophoresis of PCR product for the gene encoding ARM lipase	50
22	Gel electrophoresis of recombinant pGAPZαA/ARM	51

© COPYRIGHT UPM

23	Gel electrophoresis of recombinant pGAPZαA/ARM	51
24	The transformation plate of GS115/pGAPZαA/ARM	52
25	The transformation plate of GS115/pPICZαB/ARM	53
26	Screening of lipase activity in GS115/pPICZαB/ARM	54
27	Screening of lipase activity in GS115/pGAPZαA/ARM	55

© COPYRIGHT UPM

LIST OF ABBREVIATIONS

BLAST	Basic Local Alignment Search Tool
BLOSUM	BLOck SUbstitution Matrix
bp	base pair
CSS	Cascading Style Sheets
CGI	Common Gateway Interface
CMD	Command Prompt
da	Daltons
db	Database
DBMS	Database Management System
dH_2O	distilled water
DNA	Deoxyribonucleic Acid
ETDA	Ethylene di-amine tetrachloro acetate
g/L	gram per liter
g	gram
GNU	GNU's Not Unix
GUI	Graphical User Interface
h	hour/s
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
IP	Internet Protocol
IR	Information Retrieval
JDBC	Java DataBase Connectivity

© COPYRIGHT UPM

kb	kilo base
KDa	Kilo Dalton
L	Liter
LAN	Local Area Network
LB	Luria Bertani
M	Molar
mg	milligram
mL	milliliter
µg	microgram
µl	microliter
min	minutes
mM	milimolar
nm	nanometer
OD	Optical Density
OD_{600}	Optical Density at 600nm
ODBC	Open DataBase Connectivity
ORF	Open Reading frame
OSS	Open Source Software
PCs	Personal Computers
PCR	Polymerase Chain Reaction
PERL	Practical Extraction and Reporting Language
PHP	Hypertext Preprocessor
RDBs	Relational Databases

© COPYRIGHT UPM

RE	Restriction Enzyme
RNA	Ribonucleic Acid
rpm	rotation per minute
RT	Room Temperature
sp	species
SQL	Structured Query Language
t	Incubation time (h)
TCA	Tricholoacetic Acid
U	Unit
UI	User Interface
UV	Ultraviolet
U/ml	Unit per milliliter
U/mg	Unit per milligram
URL	Uniform Resource Locator
V	Volt
v/v	Volume per volume
WAN	Wide Area Network
w/v	Weight per volume
WWW	World Wide Web
µF	microfarad
YPD	Yeast extract, Peptone, and Dextrose
YPDS	Yeast extract, Peptone, Dextrose and Sorbitol
YPTG	Yeast extract, Peptone, Tryptic soy broth, biotin and Glycerol
YPTM	Yeast extract, Peptone, Tryptic soy broth, biotin and Methanol

© COPYRIGHT UPM

CHAPTER 1

INTRODUCTION

In computational biology, predictions are used in various aspects such as in predicting protein-protein interactions, protein structure and transcription bind site, gene expression and drug discovery and design (Hartner *et al.,* 2008; Barbosa *et al.,* 2015). In order for a prediction system to exist an algorithm is needed. Algorithms are set of statistical formulas used to solve a problem (Parikh & Boyd, 2013). There are different categories of algorithms such as simple recursive algorithms, backtracking algorithms, divide-and-conquer algorithms, dynamic programming algorithms, greedy algorithms and branch-and-bound algorithms (Seddon *et al.,* 2013; Fernandez *et al.,* 2017).

Widely used in computational biology and bioinformatics is dynamic programming algorithms (Badr *et al.,* 2014). These algorithms are used for optimization problems because it can remember past results and use it to find new results (Xia, 2013; Jou *et al.,* 2016; Reeb *et al.,* 2016). These algorithms are used for protein folding, RNA structure prediction and sequence alignment (Bandyopadhyay *et al.,* 2006). Examples of sequence alignment dynamic programming algorithms are Basic Local Alignment Search Tool (BLAST) (Altschul *et al.,* 1990) and FAST All (FASTA) (Lipman & Pearson, 1985). Both BLAST and FASTA are used for nucleotides and proteins sequence analysis (Neumann *et al.,* 2014). Algorithms are implemented on databases to make predictions (Wang & Bryant, 2014).

Databases are organized collection of information which are arranged for easy access, management and updating (Medvedeva *et al.,* 2015). They are classified based on the information content. The functions of a database include storing data, managing data and indexing data (Alonso & Korth, 1993; Kroger, 2001). The functions of a database management system (DBMS) are data definition, updating and retrieving information (Altman, 2004; Murovec *et al.,* 2015). Most databases curate data such as protein data (Finn *et al.,* 2016), weather data (Field *et al.,* 2015) and pollution data (Falchi *et al.,* 2016). Examples of protein data include enzymes and hormones.

Enzymes play an important role in many industries which makes their production also crucial. Enzymes are important in detergent, fermentation, textiles, paper and dairy industries (Ray, 2012). There are many enzymes available that suit industrial use such as lipases, amylases and proteases. The hydrolysis of fats and oil is carried out by the enzyme lipase (triacylglycerol ester hydrolases, EC 3.1.1.3) with subsequent release of monoacylglycerols, diacylglycerols, free fatty acids, and glycerol (Villeneuve *et al.,* 2000). These enzymes are ubiquitous in nature. The hydrolysis of triglycerides takes place at the interface between the insoluble substrate and water (Quyen *et al.,* 2003). These enzymes have also been used broadly in hydrolysis of a variety of esters and enantio- selective synthesis (Nars *et al.,* 2014). Antioxidants, biodiesel production, cosmetic, syntheses of antitumor, biosensor construction, waste water treatment are applications of this enzyme. Lipases are largely produced from microbes

© COPYRIGHT UPM

and specifically bacterial play an important role in the production of lipases (Hasan *et al.*, 2009; Guillén *et al.*, 2011).

For industrial enzyme production, genes are cloned and expressed in various hosts such as yeast, bacteria and mammalian cells in order to increase their productivity (Cereghino & Cregg, 2000). In this study, the focus is on the development of a promoter prediction system for lipase expression in *Pichia pastoris*. *Pichia (Komagataella) pastoris*, a methylotrophic yeast that use methanol as sole carbon and energy source. Due to the ability of this yeast to utilize methanol, carry out post-translational modifications, higher yield and easy to manipulate it was evaluated as a potential source of single-cell protein (Cos *et al.*, 2006; Lin-Cereghino *et al.*, 2013).

At the moment, only few methods are available for the enhancement of higher expression of proteins in *P. pastoris* (Nocon *et al.*, 2014). Most of the methods involve the optimization of the expression vectors and none is available to choose a promoter/host prior to cloning (Puxbaum *et al.*, 2015; Zhou *et al.*, 2015). Prior to cloning, the choice of promoter/host is done by trial and error method. BLAST heuristic algorithm can predict promoter/host prior to cloning of lipase in *P. pastoris*. Hence, this study focussed on the development of a promoter prediction system for recombinant lipase expression in *P. pastoris*.

The main research objective is:

To develop and validate a promoter prediction system for recombinant lipase expression in *P. pastoris* using BLAST heuristic algorithm.

In order to achieve this objective, the following sub-objectives were conducted:

1. To develop a lipase expression database from *P. pastoris.*
2. To implement BLAST heuristic algorithm in Perl.
3. To validate the prediction system experimentally.

© COPYRIGHT UPM

REFERENCES

Abeel, T., Saeys, Y., Rouzé, P., & Van de Peer, Y. (2008). ProSOM: core promoter prediction based on unsupervised clustering of DNA physical profiles. *Bioinformatics (Oxford, England)*, *24*(13), i24-31.

Acland, A., Agarwala, R., Barrett, T., Beck, J., Benson, D. A., Bollin, C., Zbicz, K. (2014). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, *42*(D1), D7–D17.

Adames, N. R., Wilson, M. L., Fang, G., Lux, M. W., Glick, B. S., & Peccoud, J. (2015). GenoLIB: A database of biological parts derived from a library of common plasmid features. *Nucleic Acids Research*, *43*(10), 4823–4832.

Ahmad, M., Hirz, M., Pichler, H., & Schwab, H. (2014). Protein expression in *Pichia pastoris*: recent achievements and perspectives for heterologous protein production. *Applied Microbiology and Biotechnology*, *98*(12), 5301–17.

Alkan, F., & Erten, C. (2014). BEAMS: Backbone extraction and merge strategy for the global many-to-many alignment of multiple PPI networks. *Bioinformatics*, *30*(4), 531–539.

Alonso, R., & Korth, H. F. (1993). Database system issues in nomadic computing. *ACM SIGMOD Record*, *22*(2), 388–392.

Aloulou, A., Grandval, P., De Caro, J., De Caro, A., & Carrière, F. (2006). Constitutive expression of human pancreatic lipase-related protein 1 in *Pichia pastoris*. *Protein Expression and Purification*, *47*(2), 415–421.

Altman, R. B. (2004). Building successful biological databases. *Briefings in Bioinformatics*, *5*(1), 4–5.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–10.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research*, *25*(17), 3389–3402.

Alva, V., Nam, S.-Z., Söding, J., & Lupas, A. N. (2016). The MPI bioinformatics Toolkit as an integrative platform for advanced protein sequence and structure analysis. *Nucleic Acids Research*, *44*(April 2016), gkw348.

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics*, *31*(2), 166–169.

Andreatta, M., & Nielsen, M. (2015). Gapped sequence alignment using artificial neural networks: Application to the MHC class i system. *Bioinformatics*, *32*(4), 511–517.

© COPYRIGHT UPM

Attrill, H., Falls, K., Goodman, J. L., Millburn, G. H., Antonazzo, G., Rey, A. J., Marygold, S. J. (2016). Flybase: Establishing a gene group resource for *Drosophila melanogaster*. *Nucleic Acids Research*, *44*(D1), D786–D792.

Badr El Din Ahmed, A., & Sayed Elaraby, I. (2014). Data Mining: A prediction for Student's Performance Using Classification Method. *World Journal of Computer Application and Technology*, *2*(2), 43–47.

Bandyopadhyay, S., Mitra, R., Cameron, M., Williams, H. E., Cannane, A., Eddy, S. R., … Barton, G. J. (2006). A Parallel Pairwise Local Sequence Alignment Algorithm. *BMC Bioinformatics*, *7*(1), 139–146.

Bâra, A., & Andreescu, A. (2014). Data Model for SIPAMER Prototype, *V*(4), 39–48.

Barbosa, L. C. B., Garrido, S. S., & Marchetto, R. (2015). BtoxDB: A comprehensive database of protein structural data on toxin–antitoxin systems. *Computers in Biology and Medicine*, *58*, 146–153.

Baxevanis, A. D. (2001). The Molecular Biology Database Collection: an updated compilation of biological database resources. *Nucleic Acids Research*, *29*(1), 1–10.

Blin, K., Wolf, T., Chevrette, M. G., Lu, X., Schwalen, C. J., Kautsar, S. A., Medema, M. H. (2017). antiSMASH 4.0—improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Research*, *1854*(1), 1019–1037.

Boldyreff, C., Lavery, J., Nutter, D., & Rank, S. (2003). Open-Source Development Processes and Tools, (1), 15–18.

Botha, I., Bâra, A., Oprea, S., & Jordão, T. C. (2012). Integrating XML Technology with Object-Relational Databases into Decision Support Systems 1, *III*(1), 11–20.

Bou Ali, M., Ben Ali, Y., Aissa, I., & Gargouri, Y. (2014). Eukaryotic Expression System *Pichia pastoris* Affects the Lipase Catalytic Properties: A Monolayer Study. *PloS One*, *9*(8), e104221.

Bou Ali, M., Ben Ali, Y., Karray, A., Fendri, A., & Gargouri, Y. (2011). Purification and characterization of the first recombinant bird pancreatic lipase expressed in *Pichia pastoris*: the turkey. *Lipids in Health and Disease*, *10*(1), 24.

Brunel, L., Neugnot, V., Landucci, L., Boze, H., Moulin, G., Bigey, F., & Dubreucq, E. (2004). High-level expression of *Candida parapsilosis* lipase/acyltransferase in *Pichia pastoris*. *Journal of Biotechnology*, *111*(1), 41–50.

Buneman, P., Buneman, P., Cheney, J., Cheney, J., Tan, W., Tan, W., Vansummeren, S. (2008). Curated Databases. *Science*, 1–12.

© COPYRIGHT UPM

Cameron, M., Williams, H. E., & Cannane, A. (2004). Improved gapped alignment in BLAST. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *1*(3), 116–129.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Riddell, A. (2016). Journal of Statistical Software Stan : A Probabilistic Programming Language. *Journal of Statistical Software*, *23*(5).

Catoni, E., Brocca, S., & Schmid, R. D. (1997). Overexpression of lipase A and B of *Geotrichum candidum* in *Pichia pastoris* : High-level production and some properties of functional expressed lipase B. *Biotechnology*, *11*(9), 689–695.

Celik, E., & Calık, P. (2012). Production of recombinant proteins by yeast cells. *Biotechnology Advances*, *30*(5), 1108–18.

Cereghino, G. P. L., Cereghino, J. L., Ilgen, C., & Cregg, J. M. (2002). Production of recombinant proteins in fermenter cultures of the yeast *Pichia pastoris*, 329–332.

Cereghino, J. L., & Cregg, J. M. (2000). Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*, *24*.

Cheng, X., Hofmann, H., & Cook, D. (2011). MissingDataGUI : A Graphical User Interface for Exploring Missing Values in Data, 1–13.

Chowdhury, G. (2010). Basic concepts of information retrieval systems. *Introduction to Modern Information Retrieval*, 1–13.

Cook, M. J., O'Brien, T. J., Berkovic, S. F., Murphy, M., Morokoff, A., Fabinyi, G., Himes, D. (2013). Prediction of seizure likelihood with a long-term, implanted seizure advisory system in patients with drug-resistant epilepsy: A first-in-man study. *The Lancet Neurology*, *12*(6), 563–571.

Cos, O., Ramón, R., Montesinos, J. L., & Valero, F. (2006). Operational strategies, monitoring and control of heterologous protein production in the methylotrophic yeast *Pichia pastoris* under different promoters: a review. *Microbial Cell Factories*, *5*, 17.

Cos, O., Resina, D., Ferrer, P., Montesinos, J. L., & Valero, F. (2005). Heterologous production of *Rhizopus oryzae* lipase in *Pichia pastoris* using the alcohol oxidase and formaldehyde dehydrogenase promoters in batch and fed-batch cultures. *Biochemical Engineering Journal*, *26*(2–3), 86–94.

Cos, O., Serrano, A., Montesinos, J. L., Ferrer, P., Cregg, J. M., & Valero, F. (2005). Combined effect of the methanol utilization (Mut) phenotype and gene dosage on recombinant protein production in *Pichia pastoris* fed-batch cultures. *Journal of Biotechnology*, *117*, 321–335.

Cronin, C., & Nikolov, N. S. (2014). Visualisation of GitHub ' s Public Data, (October 2014), 2–7.

© COPYRIGHT UPM

Daily, J. (2016). Parasail: SIMD C library for global, semi-global, and local pairwise sequence alignments. *BMC Bioinformatics*, *17*(1), 81.

Dam, J. K. Van. (2016). Identifying source code programming languages through natural language processing.

De Schutter, K., Lin, Y. C., Tiels, P., Van Hecke, A., Glinka, S., Weber-Lehmann, J., Callewaert, N. (2009). Supplementary: Genome sequence of the recombinant protein production host *Pichia pastoris*. *Nature Biotechnology*, *27*(6), 561–566.

Derbyshire, M. K., Gonzales, N. R., Lu, S., He, J., Marchler, G. H., Wang, Z., & Marchler-Bauer, A. (2015). Improving the consistency of domain annotation within the Conserved Domain Database. *Database : The Journal of Biological Databases and Curation*, *2015*(7), 1–8.

Dong, D. D. (2014). An efficient algorithm for global alignment of protein-protein interaction networks, (2008), 1–9.

Dwyer, A. (2003). Genomic Perl: From Bioinformatics Basics to Working Code: Cambridge University Press.

Ebrahimpour, A., Rahman, R. N. Z. R. A., Basri, M., & Salleh, A. B. (2011). High level expression and characterization of a novel thermostable, organic solvent tolerant, 1,3-regioselective lipase from *Geobacillus* sp. strain ARM. *Bioresource Technology*, *102*(13), 6972–6981.

Eom, G. T., Lee, S. H., Song, B. K., Chung, K. W., Kim, Y. W., & Song, J. K. (2013). High-level extracellular production and characterization of *Candida antarctica* lipase B in *Pichia pastoris*. *Journal of Bioscience and Bioengineering*, *116*(2), 165–70.

Falchi, F., Cinzano, P., Duriscoe, D., Kyba, C. C. M., Elvidge, C. D., Baugh, K., Furgoni, R. (2016). The new world atlas of artificial night sky brightness. *Science Advances*, *2*(6), e1600377–e1600377.

Fan, C., Xiao, F., & Wang, S. (2014). Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques. *Applied Energy*, *127*, 1–10.

Feng, G., An, B., Yang, F., Wang, H., & Zhang, L. (2017). Relevance popularity : A term event model based feature selection scheme for text classification, 1–15.

Fernandez, A., Salvador-Carulla, L., Choi, I., Calvo, R., Harvey, S. B., & Glozier, N. (2017). Development and validation of a prediction algorithm for the onset of common mental disorders in a working population. *Australian & New Zealand Journal of Psychiatry*, 486741770450.

© COPYRIGHT UPM

Fernández, L., Pérez-Victoria, I., Zafra, A., Benítez, P. L., Morales, J. C., Velasco, J., & Adrio, J. L. (2006). High-level expression and characterization of *Galactomyces geotrichum* (BT107) lipase I in *Pichia pastoris*. *Protein Expression and Purification*, *49*(2), 256–264.

Ferrer, P., Alarcón, M., Ramón, R., Dolors Benaiges, M., & Valero, F. (2009). Recombinant *Candida rugosa* LIP2 expression in *Pichia pastoris* under the control of the AOX1 promoter. *Biochemical Engineering Journal*, *46*(3), 271–277.

Fickers, P. (2014). *Pichia pastoris* : a workhorse for recombinant protein production, *2*(3), 354–363.

Field, R. D., Spessa, A. C., Aziz, N. A., Camia, A., Cantin, A., Carr, R., Wang, X. (2015). Development of a Global Fire Weather Database. *Natural Hazards and Earth System Sciences*, *15*(6), 1407–1423.

Finn, J. T., Brownscombe, J. W., Haak, C. R., Cooke, S. J., Cormier, R., Gagne, T., & Danylchuk, A. J. (2010). Applying network methods to acoustic telemetry data: Modeling the movements of tropical marine fishes. *Ecological Modelling*, *293*, 139–149.

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., Bateman, A. (2016). The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Research*, *44*(D1), D279–D285.

Frakes, W. B., William B., Baeza-Yates, R. (1992). *Information retrieval : data structures & algorithms*. Prentice Hall.

Gadhoumi, K., Lina, J. M., Mormann, F., & Gotman, J. (2016). Seizure prediction for therapeutic devices: A review. *Journal of Neuroscience Methods*, *260*(29), 270–282.

Galperin, M. Y. (2006). The Molecular Biology Database Collection: 2006 update. *Nucleic Acids Research*, *34*(Database issue), D3–D5.

Goevert, K., Cloutier, R., Roth, M., & Lindemann, U. (2016). Concept of System Architecture Database Analysis, 410–414.

Granata, I., Sangiovanni, M., Maiorano, F., Miele, M., Guarracino, M. R., Pabinger, S., Campbell, J. (2016). Var2GO: a web-based tool for gene variants selection. *BMC Bioinformatics*, *17*(S12), 135–140.

Guillén, M., Benaiges, M. D., & Valero, F. (2011). Comparison of the biochemical properties of a recombinant lipase extract from *Rhizopus oryzae* expressed in *Pichia pastoris* with a native extract. *Biochemical Engineering Journal*, *54*(2), 117–123.

© COPYRIGHT UPM

Hartner, F. S., Ruth, C., Langenegger, D., Johnson, S. N., Hyka, P., Lin-Cereghino, G. P., Glieder, A. (2008). Promoter library designed for fine-tuned gene expression in *Pichia pastoris*. *Nucleic Acids Research*, *36*(12), e76.

Hasan, F., Shah, A. A., & Hameed, A. (2009). Methods for detection and characterization of lipases: A comprehensive review. *Biotechnology Advances*, *27*(6), 782–98.

Hashemikhabir, S., Neelamraju, Y., & Janga, S. C. (2015). Database of RNA binding protein expression and disease dynamics (READ DB). *Database*, *2015*, bav072.

Holmquist, M., Tessier, D. C., & Cygler, M. (1997). High-level production of recombinant *Geotrichum candidum* lipases in yeast *Pichia pastoris*. *Protein Expression and Purification*, *11*(1), 35–40.

Huang, J., Yang, Z., Guan, F., Zhang, S., Cui, D., Guan, G., & Li, Y. (2013). A novel mono- and diacylglycerol lipase highly expressed in *Pichia pastoris* and its application for food emulsifier preparation. *Process Biochemistry*, *48*(12), 1899–1904.

Jahic, M., Gustavsson, M., Jansen, A. K., Martinelle, M., & Enfors, S. O. (2003). Analysis and control of proteolysis of a fusion protein in *Pichia pastoris* fed-batch processes. *Journal of Biotechnology*, *102*(1), 45–53.

Jin, Z., Han, S. Y., Zhang, L., Zheng, S. P., Wang, Y., & Lin, Y. (2013). Combined utilization of lipase-displaying *Pichia pastoris* whole-cell biocatalysts to improve biodiesel production in co-solvent media. *Bioresource Technology*, *130*, 102–109.

Jou, J. D., Jain, S., Georgiev, I. S., & Donald, B. R. (2016). BWM*: A Novel, Provable, Ensemble-based Dynamic Programming Algorithm for Sparse Approximations of Computational Protein Design. *Journal of Computational Biology*, *23*(6), 413–424.

Karplus, K. (2009). SAM-T08, HMM-based protein structure prediction. *Nucleic Acids Research*, *37*(SUPPL. 2), 492–497.

Kaur, K., & Rani, R. (2013). Modeling and querying data in NoSQL databases. *Proceedings - 2013 IEEE International Conference on Big Data, Big Data 2013*, 1–7.

Kouker, G., & Jaeger, K. E. (1987). Specific and sensitive plate assay for bacterial lipases. *Applied and Environmental Microbiology*, *53*(1), 211–213.

Krainer, F. W., Dietzsch, C., Hajek, T., Herwig, C., Spadiut, O., & Glieder, A. (2012). Recombinant protein expression in *Pichia pastoris* strains with an engineered methanol utilization pathway. *Microbial Cell Factories*, *11*(1), 22.

© COPYRIGHT UPM

Kramer, C., Dahl, G., Tyrchan, C., & Ulander, J. (2016). A comprehensive company database analysis of biological assay variability. *Drug Discovery Today*, *21*(8), 1213–1221.

Kroger, P. (2001). Molecular Biology Data : Database Overview , Modelling Issues , and Perspectives. *Database*.

Kwon, D. Y., & Rhee, J. S. (1986). A simple and rapid colorimetric method for determination of free fatty acids for lipase assay. *Journal of the American Oil Chemists' Society*, *63*(1), 89–92.

Lambert, C., Léonard, N., De Bolle, X., & Depiereux, E. (2002). ESyPred3D: Prediction of proteins 3D structures. *Bioinformatics (Oxford, England)*, *18*(9), 1250–1256.

Larsen, M. W., Bornscheuer, U. T., & Hult, K. (2008). Expression of *Candida antarctica* lipase B in *Pichia pastoris* and various *Escherichia coli* systems. *Protein Expression and Purification*, *62*(1), 90–97.

Lee, G. C., Lee, L. C., Sava, V., & Shaw, J. F. (2002). Multiple mutagenesis of non-universal serine codons of the *Candida rugosa* LIP2 gene and biochemical characterization of purified recombinant LIP2 lipase overexpressed in *Pichia pastoris*. *The Biochemical Journal*, *366*(2), 603–611.

Li, W., McWilliam, H., Goujon, M., Cowley, A., Lopez, R., & Pearson, W. R. (2012). PSI-Search: Iterative HOE-reduced profile SSEARCH searching. *Bioinformatics*, *28*(12), 1650–1651.

Licari, D., Baiardi, A., Biczysko, M., Egidi, F., Latouche, C., & Barone, V. (2015). Implementation of a graphical user interface for the virtual multifrequency spectrometer: The VMS-draw tool. *Journal of Computational Chemistry*, *36*(5), 321–334.

Lin-Cereghino, G. P., Stark, C. M., Kim, D., Chang, J., Shaheen, N., Poerwanto, H., Lin-Cereghino, J. (2013). The effect of α-mating factor secretion signal mutations on recombinant protein expression in *Pichia pastoris*. *Gene*, *519*(2), 311–7.

Lipman, D. J., & Pearson, W. R. (1985). Rapid and sensitive protein similarity searches. *Science*, *227*(4693), 1435–1441.

Liu, C., Wang, X. S., Nayak, K., Huang, Y., & Shi, E. (2015). ObliVM: A programming framework for secure computation. *Proceedings - IEEE Symposium on Security and Privacy*, *2015–July*, 359–376.

Liu, Y., & Schmidt, B. (2014). SWAPHI: Smith-waterman protein database search on Xeon Phi coprocessors. *Proceedings of the International Conference on Application-Specific Systems, Architectures and Processors*, 184–185.

© COPYRIGHT UPM

Liu, Y., Schmidt, B., & 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 Research Notes*, *3*, 93.

Liu, Y., Tran, T. T., Lauenroth, F., & Schmidt, B. (2014). SWAPHI-LS: Smith-Waterman Algorithm on Xeon Phi coprocessors for Long DNA Sequences. *2014 IEEE International Conference on Cluster Computing, CLUSTER 2014*, 257–265.

Liu, Y., Wirawan, A., & Schmidt, B. (2013). CUDASW++ 3.0: accelerating Smith-Waterman protein database search by coupling CPU and GPU SIMD instructions. *BMC Bioinformatics*, *14*, 117.

Lyakhov, A. O., Oganov, A. R., Stokes, H. T., & Zhu, Q. (2013). New developments in evolutionary structure prediction algorithm USPEX. *Computer Physics Communications*, *184*(4), 1172–1182.

Mackenzie, C. O., & Grigoryan, G. (2017). Protein structural motifs in prediction and design. *Current Opinion in Structural Biology*, *44*, 161–167.

Manning, C. D., & Raghavan, P. (2008). Introduction to Information Retrieval. *Americas*, *32*, 10013–2473.

Margreitter, C., Petrov, D., & Zagrovic, B. (2013). Vienna-PTM web server: a toolkit for MD simulations of protein post-translational modifications. *Nucleic Acids Research*, *41*(Web Server issue), 422–426.

Mattanovich, D., Graf, A., Stadlmann, J., Dragosits, M., Redl, A., Maurer, M., Gasser, B. (2009). Genome, secretome and glucose transport highlight unique features of the protein production host *Pichia pastoris*. *Microbial Cell Factories*, *8*, 29.

Matthes, M., Preusse, M., Zhang, J., Schechter, J., Mayer, D., Lentes, B., Trümbach, D. (2014). Mouse IDGenes: a reference database for genetic interactions in the developing mouse brain. *Database : The Journal of Biological Databases and Curation*, *2014*, 1–16.

Mcgowan, J., Sampson, M., Salzwedel, D. M., Cogo, E., Foerster, V., & Lefebvre, C. (2016). GUIDELINE STATEMENT PRESS Peer Review of Electronic Search Strategies : 2015 Guideline Statement. *Journal of Clinical Epidemiology*, *75*, 40–46.

Medvedeva, Y. A., Lennartsson, A., Ehsani, R., Kulakovskiy, I. V., Vorontsov, I. E., Panahandeh, P., Drabløs, F. (2015). EpiFactors: a comprehensive database of human epigenetic factors and complexes. *Database*, *2015*, bav067.

Minning, S., Schmidt-Dannert, C., & Schmid, R. D. (1998). Functional expression of *Rhizopus oryzae* lipase in *Pichia pastoris*: High-level production and some properties. *Journal of Biotechnology*, *66*(2–3), 147–156.

© COPYRIGHT UPM

Minning, S., Serrano, A., Ferrer, P., Solá, C., Schmid, R. D., & Valero, F. (2001). Optimization of the high-level production of *Rhizopus oryzae* lipase in *Pichia pastoris*. *Journal of Biotechnology*, *86*(1), 59–70.

Murovec, B., Kolbl, S., & Stres, B. (2015). Methane Yield Database: Online infrastructure and bioresource for methane yield data and related metadata. *Bioresource Technology*, *189*, 217–223.

Muth, T., Weilnböck, L., Rapp, E., Huber, C. G., Martens, L., Vaudel, M., & Barsnes, H. (2014). DeNovoGUI: An open source graphical user interface for de novo sequencing of tandem mass spectra. *Journal of Proteome Research*, *13*(2), 1143–1146.

Naccache, S. N., Federman, S., Veeraraghavan, N., Zaharia, M., Lee, D., Samayoa, E., Chiu, C. Y. (2014). A cloud-compatible bioinformatics pipeline for ultrarapid pathogen identification from next-generation sequencing of clinical samples. *Genome Research*, *24*(7), 1180–1192.

Nars, G., Saurel, O., Bordes, F., Saves, I., Remaud-Siméon, M., André, I., Marty, A. (2014). Production of stable isotope labelled lipase Lip2 from *Yarrowia lipolytica* for NMR: investigation of several expression systems. *Protein Expression and Purification*, *101*, 14–20.

Needleman, S. B., & Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, *48*(3), 443–453.

Neumann, R. S., Kumar, S., & Shalchian-Tabrizi, K. (2014). BLAST output visualization in the new sequencing era. *Briefings in Bioinformatics*, *15*(4), 484–503.

Neyshabur, B., Khadem, A., Hashemifar, S., & Arab, S. S. (2013). NETAL: A new graph-based method for global alignment of protein-protein interaction networks. *Bioinformatics*, *29*(13), 1654–1662.

Nguyen, T. H., & Shirai, K. (2015). Topic Modeling based Sentiment Analysis on Social Media for Stock Market Prediction. *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 1354–1364.

Nocon, J., Steiger, M. G., Pfeffer, M., Sohn, S. B., Kim, T. Y., Maurer, M., Mattanovich, D. (2014). Model based engineering of *Pichia pastoris* central metabolism enhances recombinant protein production. *Metabolic Engineering*, *24*, 129–138.

Noé, L., & Kucherov, G. (2004). Improved hit criteria for DNA local alignment. *BMC Bioinformatics*, *5*, 149.

© COPYRIGHT UPM

Nuzzo, A., Carapezza, G., Di Bella, S., Pulvirenti, A., Isacchi, A., Bosotti, R., Freedman, T. (2016). KAOS: a new automated computational method for the identification of overexpressed genes. *BMC Bioinformatics*, *17*(S12), 5–14.

Oehmen, C. S., & Baxter, D. J. (2013). ScalaBLAST 2.0: Rapid and robust BLAST calculations on multiprocessor systems. *Bioinformatics*, *29*(6), 797–798.

Oslan, N. S., Salleh, A. B., Rahman, R. N. Z. R. A., Leow, T. C., Basri, M. (2014). *Pichia pastoris* as a host to overexpress the thermostable T1 lipase from *GeoBacillus zalihae*. *GSTF Journal of BioSciences*, *3*(1).

Ozorhon, B., Karatas, C. G., & Demirkesen, S. (2014). A Web-based Database System for Managing Construction Project Knowledge. *Procedia - Social and Behavioral Sciences*, *119*(1995), 377–386.

Parikh, N., & Boyd, S. (2013). Proximal Algorithms. *Foundations and Trends in Optimization*, *1*(3), 123–231.

Patil, T. R. (2013). Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification. *International Journal Of Computer Science And Applications, ISSN: 0974-1011*, *6*(2), 256–261.

Pierce, B. G., Wiehe, K., Hwang, H., Kim, B. H., Vreven, T., & Weng, Z. (2014). ZDOCK server: Interactive docking prediction of protein-protein complexes and symmetric multimers. *Bioinformatics*, *30*(12), 1771–1773.

Piñero, J., Bravo, À., Queralt-Rosinach, N., Gutiérrez-Sacristán, A., Deu-Pons, J., Centeno, E., Furlong, L. I. (2016). DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Research*, *45*, gkw943.

Pivert, O., Slama, O., Smits, G., & Thion, V. (2016). SUGAR: A graph database fuzzy querying system. *Proceedings - International Conference on Research Challenges in Information Science*, *2016–Augus*, 0–1.

Porubská, J., Giertlová, A., Morochovičová, M., Kováčiková, E., & Porubský, O. (2014). The Slovak national food composition database: New management system DaRiS. *Journal of Food Composition and Analysis*, *34*(1), 26–38.

Potvin, G., Ahmad, A., & Zhang, Z. (2012). Bioprocess engineering aspects of heterologous protein production in *Pichia pastoris*: A review. *Biochemical Engineering Journal*, *64*, 91–105.

Prytuliak, R., Volkmer, M., Meier, M., & Habermann, B. H. (2017). HH-MOTiF: de novo detection of short linear motifs in proteins by Hidden Markov Model comparisons. *Nucleic Acids Research*, *40*(11), D242–D251.

Puxbaum, V., Mattanovich, D., & Gasser, B. (2015). Quo vadis? The challenges of recombinant protein folding and secretion in *Pichia pastoris*. *Applied Microbiology and Biotechnology*, *99*(7), 2925–2938.

© COPYRIGHT UPM

Quyen, D. T., Schmidt-Dannert, C., & Schmid, R. D. (2003). High-level expression of a lipase from *Bacillus* thermocatenulatus BTL2 in *Pichia pastoris* and some properties of the recombinant lipase. *Protein Expression and Purification*, *28*(1), 102–110.

Ravikumar, G., & Khaparde, S. A. (2015). CIM oriented graph database for network topology processing and applications integration. *Proceedings of the Universities Power Engineering Conference*, *2015–Novem*(Cim).

Ray, A. (2012). Application of Lipase in Industry. *Asian Journal of Pharmaceutical Technology*, *2*(2), 33–37.

Reeb, J., Hecht, M., Mahlich, Y., Bromberg, Y., & Rost, B. (2016). Predicted Molecular Effects of Sequence Variants Link to System Level of Disease. *PLoS Computational Biology*, *12*(8), 1–14.

Resina, D., Maurer, M., Cos, O., Arnau, C., Carnicer, M., Marx, H., Ferrer, P. (2009). Engineering of bottlenecks in *Rhizopus oryzae* lipase production in *Pichia pastoris* using the nitrogen source-regulated FLD1 promoter. *New Biotechnology*, *25*(6), 396–403.

Resina, D., Serrano, A., Valero, F., & Ferrer, P. (2004). Expression of a *Rhizopus oryzae* lipase in *Pichia pastoris* under control of the nitrogen source-regulated formaldehyde dehydrogenase promoter. *Journal of Biotechnology*, *109*(1–2), 103–113.

Rognes, T. (2011). Faster Smith-Waterman database searches with inter-sequence SIMD parallelisation. *BMC Bioinformatics*, *12*(1), 221.

Rotticci-Mulder, J. C., Gustavsson, M., Holmquist, M., Hult, K., & Martinelle, M. (2001). Expression in *Pichia pastoris* of *Candida antarctica* lipase B and lipase B fused to a cellulose-binding domain. *Protein Expression and Purification*, *21*(3), 386–392.

Sabri, S., Rahman, R. N. Z. R. A., Leow, T. C., Basri, M., & Salleh, A. B. (2009). Secretory expression and characterization of a highly Ca^{2+}-activated thermostable L2 lipase. *Protein Expression and Purification*, *68*(2), 161–6.

Sadowski, I., Breitkreutz, B. J., Stark, C., Su, T. C., Dahabieh, M., Raithatha, S., Tyers, M. (2013). The PhosphoGRID Saccharomyces cerevisiae protein phosphorylation site database: Version 2.0 update. *Database*, *2013*, 1–10.

Schaible, J., Gottron, T., & Scherp, A. (2014). Survey on common strategies of vocabulary reuse in linked open data modeling. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8465 LNCS*, 457–472.

Schmidt, T., Bergner, A., & Schwede, T. (2014). Modelling three-dimensional protein structures for applications in drug design. *Drug Discovery Today*, *19*(7), 890–897.

© COPYRIGHT UPM

Sebban-Kreuzer, C., Deprez-Beauclair, P., Berton, A., & Crenon, I. (2006). High-level expression of nonglycosylated human pancreatic lipase-related protein 2 in *Pichia pastoris*. *Protein Expression and Purification*, *49*(2), 284–91.

Seddon, J. M., Reynolds, R., Yu, Y., & Rosner, B. (2013). Validation of a prediction algorithm for progression to advanced macular degeneration subtypes. *JAMA Ophthalmology*, *131*(4), 448–55.

Sextos, A. G., & Balafas, G. K. (2011). Using the New Sap2000 Open Application Programming Interface To Develop an Interactive Front-End for the Modal Pushover Analysis of Bridges. *Compdyn 2011*, *2011*(May), 25–28.

Shahidan, N. H., Rahman, R. N. Z. R. A., Leow, T. C., Rosfarizan M., Basri, M., Salleh, A. B. (2011). The effect of carbon sources on the expression level of thermostable L2 lipase in *Pichia pastoris*. *African Journal of Biotechnology*, *10*(62), 13528–13535.

Shen, S., Sulter, G., Jeffries, T. W., & Cregg, J. M. (1998). A strong nitrogen source-regulated promoter for controlled expression of foreign genes in the yeast *Pichia pastoris*. *Gene*, *216*(1), 93–102.

Söding, J., Biegert, A., & Lupas, A. N. (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Research*, *33*(SUPPL. 2), 244–248.

Sridhar, V. B., Tian, P., Dale, A. M., Devor, A., & Saisan, P. A. (2014). Neurovascular Network Explorer 1.0: a database of 2-photon single-vessel diameter measurements with MATLAB(®) graphical user interface. *Frontiers in Neuroinformatics*, *8*(May), 56.

Stadlmayr, G., Mecklenbräuker, A., Rothmüller, M., Maurer, M., Sauer, M., Mattanovich, D., & Gasser, B. (2010). Identification and characterisation of novel *Pichia pastoris* promoters for heterologous protein production. *Journal of Biotechnology*, *150*(4), 519–29.

Surribas, A., Stahn, R., Montesinos, J. L., Enfors, S. O., Valero, F., & Jahic, M. (2007). Production of a *Rhizopus oryzae* lipase from *Pichia pastoris* using alternative operational strategies. *Journal of Biotechnology*, *130*(3), 291–299.

Suzuki, H., & Kasahara, M. (2017). Acceleration of Nucleotide Semi-Global Alignment with Adaptive Banded Dynamic Programming. *bioRxiv*, (2015), 1–8.

Szklarczyk, D., Franceschini, A., Kuhn, M., Simonovic, M., Roth, A., Minguez, P., Von Mering, C. (2011). The STRING database in 2011: Functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Research*, *39*(SUPPL. 1), 561–568.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Von Mering, C. (2015). STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Research*, *43*(D1), D447–D452.

© COPYRIGHT UPM

Tang, S. J., Shaw, J. F., Sun, K. H., Sun, G. H., Chang, T. Y., Lin, C. K., Lee, G. C. (2001). Recombinant expression and characterization of the *Candida rugosa* lip4 lipase in *Pichia pastoris*: comparison of glycosylation, activity, and stability. *Archives of Biochemistry and Biophysics*, *387*(1), 93–98.

Thomas, U., Hirzinger, G., Rumpe, B., Schulze, C., & Wortmann, A. (2013). A new skill based robot programming language using UML/P Statecharts. *Proceedings - IEEE International Conference on Robotics and Automation*, 461–466.

Toups, M. A. (2016). A study of three paradigms for storing geospatial data: distributed-cloud model, relational database, and indexed flat file, 44.

Tryka, K. A., Hao, L., Sturcke, A., Jin, Y., Wang, Z. Y., Ziyabari, L., Feolo, M. (2014). NCBI ' s Database of Genotypes and Phenotypes : dbGaP, *42*(December 2013), 975–979.

Vaidehi, V., Monica, S., S, M. S. S., Deepika, M., & Sangeetha, S. (2008). A Prediction System Based on Fuzzy Logic. *Proceedings of the World Congresson Engineering and Computer Science*.

Vijayakumar, K. R., & Gowda, L. R. (2013). Rice (*Oryza sativa*) lipase: Molecular cloning, functional expression and substrate specificity. *Protein Expression and Purification*, *88*(1), 67–79.

Villeneuve, P., Muderhwa, J. M., Graille, J., & Haas, M. J. (2000). Customizing lipases for biocatalysis: a survey of chemical, physical and molecular biological approaches. *Journal of Molecular Catalysis B: Enzymatic*, *9*(4–6), 113–148.

Vogl, T., & Glieder, A. (2013). Regulation of *Pichia pastoris* promoters and its consequences for protein production. *New Biotechnology*, *30*(4), 385–404.

Wang, J., Duncan, D., Shi, Z., & Zhang, B. (2013). WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. *Nucleic Acids Research*, *41*(Web Server issue), 77–83.

Wang, L., Wu, Y., Deng, Y., Kim, B., Pierce, L., Krilov, G., Abel, R. (2015). Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *Journal of the American Chemical Society*, *137*(7), 2695–2703.

Wang, X., Sun, Y., Shen, X., Ke, F., Zhao, H., Liu, Y., Yan, Y. (2012). Intracellular expression of Vitreoscilla hemoglobin improves production of *Yarrowia lipolytica* lipase LIP2 in a recombinant *Pichia pastoris*. *Enzyme and Microbial Technology*, *50*(1), 22–28.

Wang, Y., & Bryant, S. H. (2014). NCBI PubChem BioAssay Database, (Md), 1–11.

Wang, Y., Srivastava, K. C., Shen, G. J., & Wang, H. Y. (1995). Thermostable alkaline lipase from a newly isolated thermophilic *Bacillus*, strain A30-1 (ATCC 53841). *Journal of Fermentation and Bioengineering*, *79*(5), 433–438.

© COPYRIGHT UPM

Winkelmann, R., Harrington, J., & Jänsch, K. (2017). EMU-SDMS: Advanced Speech Database Management and Analysis in R. *Computer Speech & Language*, *0*, 1–19.

Wu, D., Yu, X. W., Wang, T. C., Wang, R., & Xu, Y. (2011). High yield Rhizopus chinenisis prolipase production in *Pichia pastoris*: Impact of methanol concentration. *Biotechnology and Bioprocess Engineering*, *16*(2), 305–311.

Wu, S., & Zhang, Y. (2008). MUSTER: Improving protein sequence profile-profile alignments by using multiple sources of structure information. *Proteins: Structure, Function and Genetics*, *72*(2), 547–556.

Xia, X. (2013). DAMBE5: A comprehensive software package for data analysis in molecular biology and evolution. *Molecular Biology and Evolution*, *30*(7).

Xu, B., & Yang, Z. (2013). PamlX: A graphical user interface for PAML. *Molecular Biology and Evolution*, *30*(12), 2723–2724.

Xu, H., Ang, Y. S., Sevilla, A., Lemischka, I. R., & Ma'ayan, A. (2014). Construction and validation of a regulatory network for pluripotency and self-renewal of mouse embryonic stem cells. *PLoS Computational Biology*, *10*(8), e1003777.

Yamada, K., & Tomii, K. (2014). Revisiting amino acid substitution matrices for identifying distantly related proteins. *Bioinformatics*, *30*(3), 317–325.

Yamada, R., Kimoto, Y., & Ogino, H. (2016). Combinatorial library strategy for strong overexpression of the lipase from *GeoBacillus thermocatenulatus* on the cell surface of yeast *Pichia pastoris*. *Biochemical Engineering Journal* (Vol. 113).

Yang, J., & Liu, L. (2010). Codon optimization through a two-step gene synthesis leads to a high-level expression of *Aspergillus niger* lip2 gene in *Pichia pastoris*. *Journal of Molecular Catalysis B: Enzymatic*, *63*(3–4), 164–169.

Yang, S., Kuang, Y., Li, H., Liu, Y., Hui, X., Li, P., Wu, D. (2013). Enhanced production of recombinant secretory proteins in *Pichia pastoris* by optimizing Kex2 P1' site. *PloS One*, *8*(9), e75347.

Yang, Y., Faraggi, E., Zhao, H., & Zhou, Y. (2011). Improving protein fold recognition and template-based modeling by employing probabilistic-based matching between predicted one-dimensional structural properties of query and corresponding native properties of templates. *Bioinformatics*, *27*(15), 2076–2082.

Yonekura-Sakakibara, K., Fukushima, A., & Saito, K. (2013). Transcriptome data modeling for targeted plant metabolic engineering. *Current Opinion in Biotechnology*, *24*(2), 285–290.

© COPYRIGHT UPM

Yu, C. S., Cheng, C. W., Su, W. C., Chang, K. C., Huang, S. W., Hwang, J. K., Lu, C. H. (2014). CELLO2GO: A Web Server for Protein subCELlular LOcalization Prediction with Functional Gene Ontology Annotation. *PloS One*, *9*(6), e99368.

Yu, M., Lange, S., Richter, S., Tan, T., & Schmid, R. D. (2007). High-level expression of extracellular lipase Lip2 from *Yarrowia lipolytica* in *Pichia pastoris* and its purification and characterization. *Protein Expression and Purification*, *53*(2), 255–63.

Yu, X. W., Wang, L. Le, & Xu, Y. (2009). *Rhizopus chinensis* lipase: Gene cloning, expression in *Pichia pastoris* and properties. *Journal of Molecular Catalysis B: Enzymatic*, *57*(1–4), 304–311.

Zahid, S. K., Hasan, L., Khan, A. A., & Ullah, S. (2015). A novel structure of the Smith-Waterman Algorithm for efficient sequence alignment. *2015 3rd International Conference on Digital Information, Networking, and Wireless Communications, DINWC 2015*, 6–9.

Zêzere, J. L., Pereira, S., Tavares, A. O., Bateira, C., Trigo, R. M., Quaresma, I., Verde, J. (2014). DISASTER: A GIS database on hydro-geomorphologic disasters in Portugal. *Natural Hazards*, *72*(2), 503–532.

Zhao, X., Xie, W., Lin, Y., Lin, X., Zheng, S., & Han, S. (2013). Combined strategies for improving the heterologous expression of an alkaline lipase from *Acinetobacter radioresistens* CMC-1 in *Pichia pastoris*. *Process Biochemistry*, *48*(9), 1317–1323.

Zhou, W. J., Yang, J. K., Mao, L., & Miao, L. H. (2015). Codon optimization, promoter and expression system selection that achieved high-level production of *Yarrowia lipolytica* lipase in *Pichia pastoris*. *Enzyme and Microbial Technology*, *71*, 66–72.

© COPYRIGHT UPM