WHOLE GENOME SEQUENCING, COMPARATIVE GENOMICS AND VIRULENCE FACTORS ANALYSES OF *Meyerozyma guilliermondii* STRAIN SO EXPRESSION SYSTEM

By

ROBIATUL AZILAH BINTI ZAINUDIN

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

December 2021

FBSB 2021 41

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

WHOLE GENOME SEQUENCING, COMPARATIVE GENOMICS AND VIRULENCE FACTORS ANALYSES OF *Meyerozyma guilliermondii* STRAIN SO EXPRESSION SYSTEM

By

ROBIATUL AZILAH BINTI ZAINUDIN

December 2021

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

Meyerozyma guilliermondii strain SO isolated from spoiled orange has been developed as a free-inducer expression system and attains a positive impact in industrial recombinant proteins production. The comprehension on genomic features is necessitated to cater its competency to perform as an expression host. Furthermore, it may enhance the yield of production at lower cost in the absence of an inducer. Therefore, the complete genome data of *M. guilliermondii* strain SO representing the host system from the perspective of genome arrangement, polymorphic variants, the composition of genes and the association in metabolic pathway is prerequisite in genomic comparative and toxicity analyses. Thus, the genome data were generated from Illumina Hiseq 4000 sequencing platform and assembled into 51 scaffolds successfully accumulated into 10.63 Mbp. These enclosed 5,335 CDS genes and 5,349 protein sequences with 43.72% GC content. About 99.29% of it were annotated to public databases. These data were employed to conduct a comparison of *M. guilliermondii* intraspecies strains which comprises of SO, ATCC 6260, YLG18 and RP-YS-11. The study discovered 99.18% genes similarity among these strains and subsequently embarking high accuracy analysis. Besides, the evaluation of established yeast expression systems, *Komagataella pastoris* and *Saccharomyces cerevisiae* with our in-house strain SO and the reference strain of *M. guilliermondii* were carried out comparatively to identify the consensus domain or subdomain that putatively responsible to perform as an expression host. A non-expression yeast species, *Candida albicans* was included in the investigation to structure normalization. This interspecies study revealed 666 homologous genes with 55 consensus regions of genome identified exclusively in *M. guilliermondii* and both expression hosts. Hence, the connectivity enzymes that played pivotal roles during carbon metabolism particularly on the utilization of methane was accessed. The study recognised an absence of alcohol oxidase (AOX) enzyme in strain SO which contributed to the factor of methanol-independency. This eventually highlighted the strength of *M. guilliermondii* strain SO to perform as a forthcoming free-inducer alternative host for recombinant protein expression. Additionally, the selected

potential virulence factors in *M. guilliermondii* strain SO were determined from system-level insights. The algorithm of Hidden Markov Model detected in silico indication of proteases (SAP), phospholipases (PLC and PLD) and hemolysin (MAM3) motifs in the genome which possessed 85% similarity to *C. albicans*, a pathogenic yeast that caused candidiasis and triggering safety concerns. Hence, the investigation of apportioning virulence factors in strain SO to predict SAP, PLC, PLD and MAM3 were executed and identified the resemblance of *C. albicans* with the expect value $2.4e^{-107}$, $9.5e^{-200}$, $0.0e^{+00}$ and $1.2e^{-258}$, respectively. Accordingly, these significant genes possibly play roles in pathogenicity. The topology of phylogenetic analysis constructed strain SO and *C. albicans* branches from the same node and clustered together as a clade to signify molecular relatedness and congeneric among these species. Nevertheless, *in vitro* analysis in quantifying the level of expression need to be investigated from the assay to quantify the enzymatic activity which may and may not activate strain SO as an opportunistic pathogenic yeast, subsequently, certifying the toxicity status of *M. guilliermondii* strain SO.

Abstrak tesis yang dikemukakan kepada Senat of Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Sarjana Sains

PENJUJUKAN GENOM, PERBANDINGAN GENOMIK DAN ANALISIS FAKTOR KE ATAS SISTEM PENGEKSPRESIAN YIS *Meyerozyma guilliermondii* STRAIN SO

Oleh

ROBIATUL AZILAH ZAINUDIN

Disember 2021

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

Meyerozyma guilliermondii strain SO yang diperolehi melalui isolasi ke atas buah oren yang rosak telah berjaya menghasilkan sistem pengekspresian tanpa induksi dan memberi kesan positif ke atas industri penghasilan protein rekombinan. Kefahaman mengenai ciri-ciri genomik adalah perlu bagi mempertingkatkan kebolehcapaiannya sebagai hos. Ini sekaligus meningkatkan penghasilan produksi pada kos yang rendah tanpa kehadiran induksi. Oleh itu, genom data *M. guilliermondii* strain SO yang lengkap menunjukkan sistem hos dari perspektif penyusunan genom, polimorfik varian, komposisi gen dan hubung-kait dalam rangkaian metabolik diperlukan sebagai asas dalam analisis perbandingan genomik dan toksisiti. Seterusnya, melalui platfom Illumina Hiseq 4000, data genomik telah dijujuk 51 scaffold berjaya menjana 10.63 Mbp data. Ia merangkumi 5,349 jujukan protein dan 5,335 gen diterjemah dilokasi pengekodan dengan 43.72% kandungan GC. Sekitar 99.29% daripadanya berpadanan dengan pengkalan data awam. Data ini kemudiannya digunakan bagi menjalankan perbandingan sesama spesies *M. guilliermondii* strain SO, ATCC 6260, YLG18 dan RP-YS-11. Kajian mendapati terdapat 99.18% persamaan gen di antara strain-strain *M. guilliermondii* dan ini seterusnya membuktikan ketepatan aras tinggi analisis yang dijalankan. Selain itu, semakan ke atas sistem ekspresi yis *Komagataella pastoris* dan *Saccharomyces cerevisiae* terhadap strain SO kajian kami dan strain rujukan *M. guilliermondii* dijalankan secara perbandingan bagi mengenalpasti kesamaan domain dan subdomain yang dianggarkan berperanan sebagai hos pengekspresian. Spesies hos bukan yis pengekspresi, *Candida albicans* turut dikaji bagi membentuk normalisasi. Kajian antara spesies mengenal pasti 666 gen homolog bersamaan 55 kawasan turutan penganjuran genom secara khususnya dalam *M. guilliermondii* dan kedua-dua hos pengekpresi. Kemudian, hubung kait enzim yang berperanan penting semasa metabolik karbon, khasnya dalam penggunaan metana diperhati. Kajian mendapati ketiadaan enzim alkohol oksida (AOX) di dalam strain SO menyumbang kepada faktor bebas-metanol. Hal ini menunjukkan kekuatan *M. guilliermondii* strain SO sebagai hos alternatif untuk penghasilan protein rekombinan tanpa induksi. Sebagai tambahan, beberapa

kebarangkalian faktor virulensi yang terpilih dikenal pasti dalam *M. guilliermondii* strain SO. Algoritma model Markov tersembunyi telah mengesan secara in silico kehadiran jujukan motif enzim protease (SAP), phospholipase (PLC dan PLD) dan hemolisin (MAM3) di dalam genom, di mana membawa 85% persamaan dengan *C. albicans*, yis patogenik yang menyebabkan candidiasis dan kebimbangan dari aspek keselamatan. Oleh itu, kajian kehadiran faktor virulensi dalam strain SO seperti penjangkaan SAP, PLC, PLD dan MAM3 dilaksanakan dan mengenal pasti kemiripan *C. albicans* dengan parameter nilai jangkaan masing-masing $2.4e^{-107}$, $9.5e^{-200}$, $0.0e^{+00}$ dan $1.2e^{-258}$. Berdasarkannya juga, gen yang signifikan ini berkemungkinan berperanan dalam patogenisiti. Topologi analisis filogenetik menunjukkan cabang konstruksi strain SO dan *C. albicans* berasal dari nodus dan kelompok yang sama bagi membuktikan kaitan molekular dan taksonomi kedua spesies ini. Walau bagaimanapun, kajian lanjutan secara *in vitro* perlu bagi mengukur aras pengekspresian aktiviti enzim melalui asai dan berkemampuan mengaktifkan peluang strain SO sebagai yis patogenik, seterusnya mengesahkan status toksisiti *M. guilliermondii* strain SO.

ACKNOWLEDGEMENTS

All praises to the Almighty Allah, the most Gracious and Merciful, Who is omnipotent and all giving, for affording me the strength and determination to complete this study. I would like to express my gratitude to my supervisor Assoc. Prof. Dr. Siti Nurbaya Oslan for her guidance, continued support and encouragement throughout this work. I am particularly grateful for my co-supervisors Prof. Dato Dr. Abu Bakar Salleh, Assoc. Prof. Dr. Suriana Sabri and Dr. Arpah Abu for providing me the needed support, good comments and invaluable suggestions.

Thank you to the most important figures who always pray, support and encourage me both mentally and physically, my parents, Zalila Hairuddin and Zainudin Abdul Aziz. My special appreciation and gratitude towards the love of my life, Muhammad Ar-Rasyiid Ahmad Dahlan and my children; Muhammad Ariiq Rizqullah, Muhammad Aqiil Rizqullah and Muhammad Al Haqq Rizqullah for their endless love, encouragement, prayer, financial and motivational support throughout my Master journey.

My special thanks to my labmates from Enzyme Technology Laboratory, IBS and EMTech members, who helped me a lot in completing this project. I also want to thank Graduate Research Fellowship for sponsoring my Master Degree and IPS Grant (GP-IPS/2016/9513300).

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
Associate Professor
Faculty of Biotechnology and Biomolecular Sciences
Universiti Putra Malaysia
(Chairman)

Suriana Sabri, PhD
Associate Professor
Faculty of Biotechnology and Biomolecular Sciences
Universiti Putra Malaysia
(Member)

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

Arpah Abu, PhD
Senior Lecturer
Faculty of Science
Institute of Biological Sciences
Universiti Malaya
(Member)

_____
ZALILAH MOHD SHARIFF, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 09 February 2023

Declaration by the Graduate Student

I hereby confirm that:

- this thesis was 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 Univeristi 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, prosters, reports, lecture notes, learning modules or any other material 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 Universiti Putra Malaysia (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: Robiatul Azilah Binti Zainudin

Declaration by the Members of Supervisory Committee

This is to confirm that:

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

Signature: _____
Name of Chairman
of Supervisory
Commitee: _____

Signature: _____
Name of Member
of Supervisory
Commitee: _____

Signature: _____
Name of Member
of Supervisory
Commitee: _____

Signature: _____
Name of Member
of Supervisory
Commitee: _____

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABREVIATIONS

ABI	Applied Biosystems
AOX	Alcohol oxidase
ATCC	American type culture collection
ATPase	Adenosine triphosphate enzyme
BLAST	The basic local alignment search tool
BLASTN	Search nucleotide databases using a nucleotide query
BLASTP	Compares a protein query to a protein database
BLASTX	Search protein databases using a translated nucleotide query
Bp	Base pair
BUSCO	Benchmarking universal single-copy orthologs
BWA	Burrows-Wheeler aligner
CAGR	Compound annual growth rate
CDS	Coding region sequence
CTG clade	Reassigned leu cug codons to serine
Chr	Chromosome
EC	Enzyme commission
E-value	Expected value
FASTQ	Text-based format; (sequence and quality scores)
FLD	Formaldehyde dehydrogenase
GAP	Glyceraldehydes-3-phosphate dehydrogenase
Gbp	Gigabase pair
GC content	Guanine-cytosine content
gDNA	Genomic DNA
GO	Gene ontology

GRAS	Generally recognized as safe
HGAP	Hierarchical genome-assembly process
HMM	Hidden Markov model
HOXD	Homeobox protein
IgG	Immunoglobulin G
ITS	Internal transcribed spacer
Kbp	Kilobase pair
KEGG	Kyoto encyclopedia of genes and genomes
KOALA	KEGG orthology and links annotation
L50	Smallest contigs;length covers 50% genome size
LCBs	Locally collinear blocks
MAUVE	Multiple alignment of conserved genomic sequence with rearrangements
Mbp	Mega base pair
mRNA	Messenger ribonucleic acid
MSA	Multiple sequence alignment
MUMs	Maximal unique matching subsequences
MUSCLE	Multiple sequence comparison by log- expectation
Mut^{+}	Methanol utilization plus
Mut^{s}	Methanol utilization slow
Mut^{-}	Methanol utilization minus
N50	Shortest contig length to cover 50% genome
NGS	Next-generation sequencing
NJ	Neighbor Joining
pPICZαB	*Pichia pastoris* protein secreting expression vectors
nm	Nanometer

P_{AOX1}	AOX1 promoter
Pfam	The protein families database
pH	Potential of hydrogen
PHMMER	Profile hidden Markov model
PLC	Phospholipase C
PLD	Phospholipase D
RM	Ringgit Malaysia
Rpm	Revolutions per minute
rRNA	Ribosomal ribonucleic acid
S	Svedberg unit; sedimentation rate
SAP	Aspartic protease
SMRT	Single molecule real-time
TAE	Tris base, acetic acid and EDTA buffer
TB	Terabyte
tRNA	Transfer ribonucleic acid
U/g	Unit per gram
U/ml	Units per millilitre
UniProtKB	UniProt Knowledgebase
USD	United States Dollar
V	Volts
v/v	Volume/volume percentage
WGS	Whole genome sequencing
w/v	Weight/volume percentage
YPD	Yeast Extract-Peptone-Dextrose media
YPTM	Yeast Extract-Peptone-Tryptic-Methanol media

CHAPTER 1

INTRODUCTION

1.1 Background

The advancement on the production of recombinant proteins offers significant potential for therapeutic and industrial enzymes. The conventional strategies are merged with molecular technology to yield higher protein at lower cost. Yeasts are unicellular eukaryotic microbial that were discovered to provide capability to growth robustly on simple media, capable to accommodate genetic modifications and incorporate post-translational modifications. Pertaining to the advantages of yeast cellular machinery, the production of functional protein in a large amount *via* recombinant DNA approach to regulate heterologous gene mechanism is achievable (Nielsen, 2014). Several commercial products manipulated from heterologous protein secretion are available in the market, for example, insulin, vaccine against hepatitis B, detergents and paper pulp (Porro *et al.,* 2005).

A locally isolated ascomycetous species from spoiled orange identified as *Meyerozyma guilliermondii* strain SO (GenBank JN084128) (Oslan *et al.,* 2012) has been developed as a prospective system for heterologous protein expression providing an alternative to the intensively used species, *Komagataella pastoris* (formerly known as *Pichia pastoris*) (Oslan *et al.,* 2015). In fact, this novel strain is capable to express heterologous recombinant enzyme such as lipase (Oslan *et al.,* 2015), α-amylase (Mohamad *et al.,* 2020), protease and diamino oxidase (Mahyon, 2017). Moreover, the competency of this yeast to compatibly host an expression vector mediated by alcohol oxidase (AOX) and formaldehyde dehydrogenase (FLD) promoters successfully proved the commencement of mRNA transcription independently without being initiated by any inducer such as methanol or methylamine (Mohamad *et al.,* 2020). The outstanding achievement on demonstrating the ability to perform as a yeast expression system obliquely could reduce the production cost, minimize methanol toxicity effects and would innovate the technology of enzyme research.

Significantly, *M. guilliermondii* shares a common approach to *K. pastoris* in regulating the expression of recombinant protein. The compatibility of this strain using pPICZαB vector likewise in *K. pastoris* features AOX1 promoter (P_{AOX1}) which is responsible to initiate the metabolism process in peroxisome of methylotrophic yeast, thus, represented as a control element for heterologous gene expression (Chiruvolu *et al.,* 1997). The AOX1 gene particularly utilizes methanol as a carbon source to control the transcription of foreign protein *via* repression / derepression mechanism and undergoes an oxidation process to compose formaldehyde and hydrogen peroxide as a byproduct (Cregg *et al.,* 1989). Furthermore, BLAST algorithm identifies the promoter in *M. guilliermondii* strain SO is 100 percent identical to the AOX1 promoter in *K. pastoris* expression system (Oslan *et al.,* 2015). Besides, prior study remarkably discovered that strain SO required shorter cultivation time to produce heterologous protein as compared to *K. pastoris*,

therefore, worthwhile to be established as the next commercial expression system.

The complete genomic data of *M. guilliermondii* strain SO is necessary in order to construct a model of expression host. To date, the available genomic data of *M. guilliermondii* in public databases reported are from 6 strains, where ATCC 6260 is recognised as representative genome. Moreover, each strain may reveal nucleotide polymorphism and demonstrated heterogeneity. The urgency of having its own whole genome sequencing (WGS) data is crucial for further modification, hence, leading the objective of this study. The establishment of WGS pipeline is embedded according to the Illumina next-generation sequencing technology platform and performed bioinformatics analysis, from assembly, annotation and finalization through interspecies and intraspecies comparative analysis.

Eventually, the inception of this novel strain as prospective yeast expression system deemed an emergence study regarding its toxicological concern to determine 'Generally Recognized as Safe' (GRAS) status. So, the identification of virulence factors is decisive to implicate adverse effects cause by the yeast.

1.2 Problem statement

Inadequate information on genomic data of *M. guilliermondii* strain SO is pivotal to comprehend and manipulate the competency of the host as an expression system. Furthermore, prior studies have reported the species possesses similarity to *Candida albicans*, the opportunistic human pathogenic yeast. Yet, the candida-like virulence proteins of *M. guilliermondii* strain SO have not been identified/analysed.

1.3 Objectives

A comprehensive understanding of yeast expression system performed by *M. guilliermondii* was achieved in this study through objectives as followed;

i Acquiring, assembling and annotating the full genome sequence of *M. guilliermondii* strain SO.

ii Comparing the full genome of *M. guilliermondii* strain SO intraspecies and interspecies of yeast expression system.

iii Predict the potential virulence factors in silico from *M. guilliermondii* strain SO proteome.

REFERENCES

Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Retrieved from http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed December 2020.

Anthony, C. (1982). The Biochemistry of Methylotrophs. *Academic Press* (10): 269-295

Altschul, S.F., Gish, W., Miller, W. *et al.* (1990). Basic local alignment search tool. *Journal of Molecular Biology* 215:403-410.

Baghban, R., Farajnia, S., Rajabibazl, M. *et al.* (2019). Yeast expression systems: overview and recent advances. *Molecular Biotechnology* 61(5): 365–384.

Bairoch, A., Apweiler, R., Wu, C. *et al.* (2005). The Universal Protein Resource (UniProt). *Nucleic acids research*, *33* (Database issue): D154–D159.

Bairwa, G., Hee Jung, W. & Kronstad J.W. (2017). Iron acquisition in fungal pathogens of humans. *Metallomics* 9(3):215–227.

Balamurugan, V., Reddy, G.R. & Suryanarayana, V.V.S. (2007). *Pichia pastoris*: A notable heterologous expression system for the production of foreign proteins – Vaccines. *Indian Journal of Biotechnology* 6: 175–186.

Bandana, K., Jashandeep, K. & Jagdeep, K. (2018). Phospholipases in bacterial virulence and pathogenesis. *Advances in Biotechnology & Microbiology* 10(5): 106–113.

Bennett, D. E., McCreary, C. E. & Coleman, D. C. (1998). Genetic characterization of a phospholipase C gene from *Candida albicans*: presence of homologous sequences in Candida species other than Candida albicans. *Microbiology* 144: 55–72.

Besemer, J., Lomsadze, A., & Borodovsky, M. (2001). GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Research* 29(12): 2607–2618.

Boetzer, M, Henkel, C.V., Jansen, H.J. *et al.* (2011). Scaffolding pre-assembled contigs using SSPACE. Bioinformatics 27(4): 578-579.

Busby, S. & Ebright, R.H. (1994). Promoter structure, promoter recognition, and transcription activation in prokaryotes. *Cell* 79(5): 743–746.

Butler, G., Rasmussen M.D., Lin, M.F. *et al.*, (2009). Evolution of pathogenicity and sexual reproduction in eight *Candida* genomes. *Nature* 459(7247):657–662.

Calderone, R.A & Fonzi W.A. (2001). Virulence factors of *Candida albicans*. *Trends Microbiol* 9(7): 327-335.

Çelik, E. and Çalık, P. (2012). Production of recombinant proteins by yeast cells.

Biotechnology Advances 30: 1108–1118.

Chauvel, M., Nesseir, A., Cabral, V. *et al.,* (2012). A versatile overexpression strategy in the pathogenic yeast Candida albicans: identification of regulators of morphogenesis and fitness. *PLoS One* 7(9): 45912.

Chaves, A.L.S., Trilles, L., Alves, G.M. *et al.,* (2020). A case-series of bloodstream infections caused by the *Meyerozyma guilliermondii* species complex at a reference center of oncology in Brazil. *Medical Mycology* 59 (3):235-243.

Cherry, J.M., Hong, E.L., Amundsen, C, *et al.,* (2012) *Saccharomyces* Genome Database: the genomics resource of budding yeast. *Nucleic Acids Research* 40(D1):700–705.

Chiruvolu, V., Cregg, J. M., & Meagher, M. M. (1997). Recombinant protein production in an alcohol oxidase-defective strain of Pichia pastoris in fedbatch fermentations. *Enzyme and Microbial Technology*, 21(4), 277–283.

Conesa, A., & Götz, S. (2008). Blast2GO: A comprehensive suite for functional analysis in plant genomics. *International journal of plant genomics*: 619832.

Cregg, J.M., Barringer, K.J. Hessler, A.Y. *et al.,* (1985). *Pichia pastoris* as a host system for transformations. *Molecular and Cellular Biology* 5(12): 3376–3385.

Cregg, J., Madden, K., Barringer, K. *et al.,* (1989). Functional characterization of the two alcohol oxidase genes from the yeast Pichia pastoris. *Molecular and cellular biology* 9(3): 1316–1323.

Daly, R. & Hearn M.T.W. (2005). Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool in protein engineering and production. *Journal of Molecular Recognition* 18(2): 119–138.

Darling, A.C.E., Mau, B., Blattner, F.R. *et al.,* (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Research* 14(7): 1394–1403.

Department of Statistics Malaysia. (2020). Current Population Estimates, Malaysia, 2020. Retrieved from https://www.dosm.gov.my/v1/index.php?r=column/cthemeByCat&cat=155&bul_id=OVByWjg5YkQ3MWFZRTN5bDJiaEVhZz09&menu_id=L0pheU43NWJwRWVSZklWdzQ4TlhUUT09. Accessed December, 2020.

Dolan, J.W., Bell, A.C., Hube, B. *et al.,* (2004). *Candida albicans* PLD1 activity is required for full virulence. *Medical Mycology* 42: 439–447.

Dong, Y., Hu, J., Fan, L. *et al.,* (2017). RNA-Seq-based transcriptomic and metabolomic analysis reveal stress responses and programmed cell death induced by acetic acid in *Saccharomyces cerevisiae*. *Scientific Reports* 7: 42659.

Dujon, B. (2010). Yeast evolutionary genomics. *Nature Reviews Genetics* 11: 512–524.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32(5): 1792–1797.

Engel, S.R., Dietrich, F.S., Fisk, D.G. *et al.*, (2014). The Reference Genome Sequence of *Saccharomyces cerevisiae*: Then and Now. *G3 (Bethesda)* 4: 389–398.

Entian, K.D, Schuster, T., Hegemann J.H. *et al.*, (1999) Functional analysis of 150 deletion mutants in *Saccharomyces cerevisiae* by a systematic approach. *Molecular Genetics and Genomics* 262(4-5):683-702.

Fam, J.P. (2018). Metabolic profiling of *Meyerozyma guilliermondii* strain SO and a recombinant strain SO2 expressing lipase. Masters thesis, Universiti Putra Malaysia.

Fakruddin, M., Hossain, M.N. & Ahmed, M.M. (2017). Antimicrobial and antioxidant activities of *Saccharomyces cerevisiae* IFST062013, a potential probiotic. *BMC Complement Altern Med* 17:64.

Finn, R.D., Coggill, P., Eberhardt, R.Y., *et al.* (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic acids research* 44(D1): D279–D285.

Fitzpatrick, D.A., Logue, M.E., Stajich, J.E. *et al.*, (2006). A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis. *BMC Evolutionary Biology* 6: 99.

Geiss, G.K., Bumgarner, R. E., Birditt, B. *et al.*, (2008). Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nature Biotechnology*, 26(3): 317–325.

Ghannoum, M.A. (2000). Potential role of phospholipases in virulence and fungal pathogenesis. *Clinical Microbiology Reviews* 13(1): 122–143.

Ghareib, M., Youssef, K.A & Khalil, A.A. (1988). Ethanol tolerance of *Saccharomyces cerevisiae* and its relationship to lipid content and composition. *Folia Microbiol (Praha)* 33(6): 447–452.

Girmenia, C., Pizzarelli, G., Cristini, F. *et al.*, (2006). *Candida guilliermondii* fungemia in patients with hematologic malignancies. *Journal of Clinical Microbiology* 44(7): 2458–2464.

Goffeau, A., Barrell B.G., Bussey, H. *et al.* (1996). Life with 6000 genes. *Science* 274(5287): 546, 563–567.

Hak-Min K., Sungwon J., Oksung C. *et al.* (2021). Comparative analysis of 7 short-read sequencing platforms using the Korean Reference Genome: MGI and Illumina sequencing benchmark for whole-genome sequencing, *GigaScience* 10(3).

Hannon, G.J. (2010) FASTX-Toolkit. Retrieved from http://hannonlab.cshl.edu/fastx_toolkit. Accessed December 2020.

Hawksworth, D.L. & Lucking, R. (2017). Fungal diversity revisited:2.2 to 3.8 million species. *Microbiology Spectrum* 5(4).

Herrmann, G.F., Krezdorn, C., Malissard, M. *et al.,* (1995). Large-scale production of a soluble human β-1,4-Galactosyltransferase ising a *Saccharomyces cerevisiae* expression system. *Protein Expression and Purification* 6(1): 72-78.

de Hoog, G., Ahmed, S., Danesi, P. *et al.,* (2018). Distribution of pathogens and outbreak fungi in the fungal kingdom. *Emerging and epizootic fungal infections in animals*, *Springer Dordrecht*: 3-16.

Hyde, K.D., Abdullah, M.S.A., Andersen, B. *et al.,* (2018). The world's ten most feared fungi. *Fungal Diversity* 93: 161–194.

Ireland, R.J., & Joy, K.W. (1983). Purification and properties of an asparagine aminotransferase from Pisum sativum leaves. *Archives of Biochemistry and Biophysics*, 223(1): 291–296.

Jabra-Rizk, M.A., Falkler, W.A. & Meiller, T.F. (2004). Fungal biofilms and drug resistance. *Emerging Infectious Diseases* 10(1): 14–19.

Jackson, B. E., Wilhelmus, K. R., & Hube, B. (2007). The Role of Secreted Aspartyl Proteinases in *Candida albicans* Keratitis. *Investigative Opthalmology & Visual Science* 48(8): 3559.

Jenkins, G.M. & Frohman, M.A. (2005). Phospholipase D: a lipid centric review. *Cellular and Molecular Life Sciences* 62(19-20): 2305–2316.

Jeon, S.A., Park, J. L., Park, S. J. *et al.* 2021). Comparison between MGI and Illumina sequencing platforms for whole genome sequencing. *Genes & genomics* 43(7): 713–724.

Kanehisa, M. and Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res*. 28: 27-30.

Kunze, D., Melzer, I., Bennett, D. *et al.,* (2005). Functional analysis of the phospholipase C gene CaPLC1 and two unusual phospholipase C genes, CaPLC2 and CaPLC3, of *Candida albicans*. *Microbiology* 151: 3381–3394.

Kurtzman, C.P., (2009). Biotechnological strains of *Komagataella (Pichia) pastoris* are *Komagataella phaffii* as determined from multigene sequence analysis. *Journal of Industrial Microbiology Biotechnology*. 36(11):1435-1438.

Kurtzman, C.P. (2011). *Meyerozyma* Kurtzman & M. Suzuki (2010): The Yeasts, A Taxonomic Study. *The Yeast* (5): 621–624.

Lagesen, K, Hallin, P.F., Roedland, E. *et al.* (2007). RNammer: consistent annotation of rRNA genes in genomic sequences. *Nucleic Acids Research.*

Li, S.F., Wang, J., Dong, R. *et al.,* (2020). Chromosome-level genome assembly,

annotation and evolutionary analysis of the ornamental plant Asparagus setaceus. *Horticulture Research* 7: 48.

Lo, H. J., Kohler, J.R., DiDomenico B. *et al.,* (1997). Nonfilamentous *C. albicans* mutants are avirulent. *Cell* 90: 939–949.

Locey, K.J. & Lennon J.T. (2016). Scaling laws predict global microbial diversity. *Proceedings of the National Academy of Sciences* 113(21): 5970–5975.

Love, K.R., Shah, K.A., Whittaker, C.A. *et al.,* (2016). Comparative genomics and transcriptomics of *Pichia pastoris. BMC Genomics* 17: 550.

Lowe, T.M. & Eddy, S.R. (1997). tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research.* 25: 955-964.

Mahyon, N.I. (2017). Structural investigation of alcohol oxidase from *Meyerozyma guilliermondii* and the use of its promoter for recombinant protein expression. Masters thesis, Universiti Putra Malaysia.

Mamo, J. & Assefa, F. (2018). The Role of Microbial Aspartic Protease Enzyme in Food and Beverage Industries. *Journal of Food Quality* :1–15.

Mardis, E.R. (2008). Next-generation DNA sequencing methods. *Annual Review of Genomics and Human Genetics* 9: 387-402.

Miller, J. R., Koren, S., & Sutton, G. (2010). Assembly algorithms for next-generation sequencing data. *Genomics,* 95(6): 315–327.

Mohamad, N.S.N., Leow C.T., Oslan S.N. *et al.,* (2018). Molecular expression of a recombinant thermostable bacterial amylase from *Geobacillus stearothermophilus* SR74 using methanol-free *Meyerozyma guilliermondii* strain SO yeast system. *BioResources* 15 (2): 3161–3172.

Mohan, V., & Ballal M. (2008). Proteinase and phospholipase activity as virulence factors in Candida species isolated from blood. *Revista Iberoamericana de Micología* 25: 208–210.

Mukherjee, S., Mukherjee, N., Saini, P. *et al.,* (2014). Molecular evidence on the occurrence of co-infection with *Pichia guilliermondii* and *Wuchereria bancrofti* in two filarial endemic districts of India. *Infectious Disease of Poverty*, *3*(13): 1–10.

Nadalin, F., Vezzi, F., & Policriti, A. (2012). GapFiller: a *de novo* assembly approach to fill the gap within paired reads. *BMC Bioinformatics*, *13 Suppl 14*(Suppl 14), S8.

Naglik, J.R., Challacombe, S.J. & Hube, B. (2003). *Candida albicans* secreted aspartyl proteinases in virulence and pathogenesis. *Microbiology and Molecular Biology Reviews* 67(3): 400–428.

Nakamura, Y., Kanemaru, K., Shoji, K. *et al.,* (2020). Phosphatidylinositol-specific phospholipase C enhances epidermal penetration by *Staphylococcus aureus*.

Scientific Reports 10: 17845.

Navarro-Arias, M.J, Hernández-Chávez M.J, Garcia-Carnero L.C *et al.*, (2019). Differential recognition of *Candida tropicalis, Candida guilliermondii, Candida krusei,* and *Candida auris* by human innate immune cells. *Infect Drug Resist* 12: 783-794.

Nielsen, K. (2014). Protein expression-yeast. *Laboratory Methods in Enzymology: Protein Part A* 536: 133-147.

Ohama, T., Suzuku, T., Mori, M. *et al.,* (1993). Non-universal decoding of the leucine codon CUG in several Candida species. *Nucleic acids research* 21(17): 4039–4045.

Oslan S.N., Salleh A.B., Raja A.R.R.N.Z *et al.,* (2012). Locally isolated yeasts from Malaysia: identification, phylogenetic study and characterization. *Acta Biochimica Polonica* 59: 225–229.

Oslan S.N., Salleh A.B., Raja A.R.R.N.Z *et al.,* (2015). A newly isolated yeast as an expression host for recombinant lipase. *Cellular and Molecular Biology Letters* 20: 279–293.

Pais, F.S.M., Ruy, P.C., Oliveira, G. *et al.,* (2014). Assessing the efficiency of multiple sequence alignment programs. *Algorithms for Molecular Biology* 9(1): 4.

Papon, N., Savini, V., Lanoue, A. *et al.,* (2013). *Candida guilliermondii*: biotechnological applications, perspectives for biological control, emerging clinical importance and recent advances in genetics. *Current Genetics* 59: 73-90.

Pfaller, M.A., Diekema D. J., Gibbs D. L. *et al.,* (2010). Results from the ARTEMIS DISK global antifungal surveillance study, 1997 to 2007: a 10.5-year analysis of susceptibilities of *Candida* species to fluconazole and voriconazole determined by CLSI standardized disk diffusion. *J. Clin. Microbiol* 48:1366–1377.

Ponting, C. P., & Kerr, I. D. (1996). A novel family of phospholipase D homologues that includes phospholipid synthases and putative endonucleases: identification of duplicated repeats and potential active site residues. *Protein science: a publication of the Protein Society*, *5*(5): 914–922.

Porro, D., Sauer, M., Paola B. *et al.,* (2005). Recombinant protein production in yeasts. *Molecular Biotechnology* 31: 245–259.

Potter, S.C., Luciani, A., Eddy, S.R. *et al.,* (2018). HMMER web server: 2018 update. *Nucleic Acids Research* 46: 200–204.

Quinlan, A.R. & Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)*, *26*(6), 841–842.

Radzi S.N.F. (2020). Toxicity studies of *Meyerozyma guilliermondii* strain SO using zebrafish as a model, Final Year Project Thesis, Universiti Putra Malaysia.

Romanos, M.A., Scorer, C.A. & Clare, J.J. (1992). Foreign gene expression in yeast: A Review. *Yeast* 8: 423–488.

Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4(4): 406–425.

dos Santos, A.L.S & de Arau´jo, R.M.S (2005). Candida guilliermondii isolated from HIV-infected human secretesa 50 kDa serine proteinase that cleaves a broad spectrum of proteinaceous substrates. *FEMS Immunology and Medical Microbiology* 43: 13-20.

Santos, M.A.S., Gomes, A.C., Santos, M.C. *et al.*, (2011). The genetic code of the fungal CTG clade. *Comptes Rendus Biologies* 334(8-9): 607–611.

Santos, S.C., Teixeira, M.C., Cabrito, T.R. *et al.*, (2012). Yeast toxicogenomics: genome-wide responses to chemical stress with imp,act in environment health pharmacology & biotechnology. *Frontier in Genetics* 3: 63.

Schreiber, B., Lyman, C. A., Gurevich, J. *et al.*, (1985). Proteolytic activity of *Candida albicans* and other yeasts. *Diagnostic Microbiology and Infectious Disease*, 3(1), 1–5.

Sewalt, V., Shanahan, D., Gregg, L. *et al.*, (2016). The Generally Recognized as Safe (GRAS) process for industrial microbial enzymes. *Industrial Biotechnology* 12(5): 295-302.

Sibirny, A.A. & Boretsky, Y.R. (2009). *Pichia guilliermondii*. Yeast biotechnology: diversity and applications. *BMC Chemistry* :113–134.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., *et al.* (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics (Oxford, England)* 31(19): 3210–3212.

Singh, D.K., Tóth, R. & Gácser, A. (2020). Mechanisms of pathogenic *Candida* species to evade the host complement attack. *Frontiers in Cellular and Infection Microbiology* 10: 94.

Smith, M.R, Liu Y.L, Matthews, N.T, *et al*., (1994). Phospholipase C-gamma 1 can induce DNA synthesis by a mechanism independent of its lipase activity. *Proceedings of the National Academy of Sciences of the United States of America* 91(14): 6554–6558.

Srivastava, V.K., Suneetha, K.J. & Kaur, R. (2014). A systematic analysis reveals an essential role for high-affinity iron uptake system, haemolysin and CFEM domain-containing protein in iron homoeostasis and virulence in *Candida glabrata*. *Biochemical Journal* 463: 103–114.

Takagi, S., Tsutsumi, N., Terui, Y. *et al.* (2019). Engineering the expression system for *Komagataella phaffii* (*Pichia pastoris*): an attempt to develop a methanol-free expression system. *FEMS Yeast Research* 19(6).

Ter-Hovhannisyan, V., Lomsadze, A., Chernoff, Y.O. *et al.* (2008). Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Research* 18: 1979–1990.

Tsai, M.H., Hsu, J.F., Yang, L.Y. *et al.* (2018). Candidemia due to uncommon *Candida* species in children: new threat and impacts on outcomes. *Scientific Reports* 8:15239.

Tseng, T.Y., Chen, T.C., Ho, C.M. *et al.* (2017). Clinical features, antifungal susceptibility and outcome of *Candida guilliermondii* fungemia: An experience in a tertiary hospital in mid-Taiwan. *Journal of Microbiology, Immunology and Infection* 51: 552–558.

Velayuthan, R.D., Samudi, C., Singh, H.K.L. *et al.* (2018). Estimation of the burden of serious human fungal infections in Malaysia. *Journal of Fungi* 4(38).

Vieira, G.A.M., Souza, C.T., Silva, C.L. *et al.* (2018). Comparison of yeasts as hosts for recombinant protein production. *Microorganisms* 6(2): 38.

Weinhandl, K., Winkler, M., Glieder, A. *et al.* (2014). Carbon source dependent promoters in yeasts. *Microb Cell Fact* 13: 5.

Wheeler, D. & Bhagwat, M. (2007). BLAST QuickStart: Example-Driven Web-Based BLAST Tutorial. Comparative Genomics: Volumes 1 and 2. *Humana Press*.

Wheeler, T.J., Clements, J. & Finn, R.D. (2014). Skylign: a tool for creating informative, interactive logos representing sequence alignments and profile hidden Markov models. *BMC Bioinformatics* 15: 7.

Wilfinger, W.W. (1997). Effect of pH and ionic strength on the spectrophotometric assessment of nucleic acid purity. *BioTechniques* 22: 474–481.

World Health Organization. (2019). Global spending on health: A world in transition. Retrieved from https://www.who.int/health_financing/documents/health-expenditure-report-2019.pdf?ua=1. Accessed December, 2020.

Wu, Z., Liu, Y., Feng, X. *et al.* (2014). Candidemia: incidence rates, type of species, and risk factors at a tertiary care academic hospital in China. *International Journal of Infectious Diseases* 22: 4-8.

Yan, W., Zhang, S., Wu, M. *et al.* (2019). The draft genome sequence of *Meyerozyma guilliermondii* strain YLG18, a yeast capable of producing and tolerating high concentration of 2-phenylethanol. *Genome Reports* 9: 441.

Yang, M., Jensen, L.T., Gardner, A.J. *et al.* (2005). Manganese toxicity and *Saccharomyces cerevisiae* Mam3p, a member of the ACDP (ancient conserved domain protein) family. *Biochemical Journal* 386: 479–487.

Zerbino, D. R., & Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome research* 18(5): 821–829.

Zhao, C. & Wang, Z. (2018). GOGO: An improved algorithm to measure the semantic similarity between gene ontology terms. *Sci Rep* 8 (15107).