Pertanika J. Sci. & Technol. 30 (1): 777 - 797 (2022)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Determination of Putative Vacuolar Proteases, PEP4 and PRB1 in a Novel Yeast Expression Host *Meyerozyma guilliermondii* Strain SO using Bioinformatics Tools

Okojie Eseoghene Lorrine[1,2], Raja Noor Zaliha Raja Abd. Rahman[1,3,4], Joo Shun Tan[5], Raja Farhana Raja Khairuddin[6], Abu Bakar Salleh[1] and Siti Nurbaya Oslan[1,2,3]*

[1]Enzyme and Microbial Technology Research Centre, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia
[2]Department of Biochemistry, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia
[3]Laboratory of Vaccine and Biomolecules (VacBio 5), Institute of Bioscience, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia
[4]Department of Microbiology, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia
[5]School of Industrial Technology, Universiti Sains Malaysia, 11800 USM, Pulau Pinang, Malaysia
[6]Faculty of Science and Mathematics, Universiti Pendidikan Sultan Idris, 35900 UPSI, Tanjong Malim, Perak, Malaysia

ABSTRACT

Meyerozyma guilliermondii strain SO, a newly isolated yeast species from spoilt orange, has been used as a host to express the recombinant proteins using methylotrophic yeast promoters. However, as a novel yeast expression system, the vacuolar proteases of this yeast have not been determined, which may have contributed to the low level of heterologous protein secretions. Thus, this study aimed to determine intra- and extracellular proteolytic activity and identify the putative vacuolar proteases using bioinformatics techniques. A clear zone was observed from the nutrient agar skimmed milk screening plate. Proteolytic activity of 117.30 U/ml and 75 U/ml were obtained after 72 h of cultivation for both extracellular and intracellular proteins, respectively. Next, the Hidden Markov model (HMM) was used to detect the presence of the vacuolar proteases (PEP4 and PRB1) from the strain SO proteome. Aspartyl protease (PEP4) with 97.55% identity to *Meyerozyma* sp. JA9 and a serine protease (PRB1) with 70.91% identity to

ARTICLE INFO

Article history:
Received: 26 April 2021
Accepted: 05 July 2021
Published: 04 January 2022

DOI: https://doi.org/10.47836/pjst.30.1.42

E-mail addresses:
gs50664@student.upm.edu.my (Okojie Eseoghene Lorrine)
rnzaliha@upm.edu.my (Raja Noor Zaliha Raja Abd. Rahman)
jooshun@usm.my (Joo Shun Tan)
rfrk@fsmt.upsi.edu.my (Raja Farhana Raja Khairuddin)
abubakar@upm.edu.my (Abu Bakar Salleh)
snurbayaoslan@upm.edu.my (Siti Nurbaya Oslan)
* Corresponding author

ISSN: 0128-7680
e-ISSN: 2231-8526

Candida albicans were revealed. The homology with other yeast vacuolar proteases was confirmed *via* evolutionary analysis. PROSPER tool prediction of cleavage sites postulated that PEP4 and PRB1 might have caused proteolysis of heterologous proteins in strain SO. In conclusion, two putative vacuolar proteases (PEP4 and PRB1) were successfully identified in strain SO. Further characterization can be done to understand their specific properties, and their effects on heterologous protein expression can be conducted *via* genome editing.

Keywords: Hidden Markov model, phylogenetic tree, secretion, vacuolar proteases, yeast expression system

INTRODUCTION

Proteases are proteolytic enzymes that can break long chain-like protein molecules into shorter fragments, peptides, and eventually into their component's amino acids (Barrett et al., 1998). Yeast proteases are secreted into the extracellular space, and the majority of which are abound in the vacuolar, cell wall, plasma membrane, and Golgi apparatus (Delic et al., 2013; Feyder et al., 2015). A major limitation for efficient secretion of heterologous proteins is the undesired travel to the vacuoles where the proteins are degraded and recycled (Forgac, 2000; Li & Kane, 2009). Besides, descriptions of proteolytic degradation of recombinant proteins have been attributed to the release of vacuolar proteases to the exterior as a result of lysis of the cells as opposed to product mistargeting (Kobayashi et al., 2000; Sinha et al., 2005) and could mostly be averted by disruption of the two major vacuolar proteases PEP4 (aspartyl) and PRB1(serine) or yapsin-type proteases (Gleeson et al., 1998; Idiris et al., 2010; Silva et al., 2011; Wu et al., 2013). Some studies have reported the reduced proteolysis of secreted heterologous proteins using yapsin (vacuolar protease) mutant strains of *Pichia pastoris* and *Saccharomyces cerevisiae* (Copley et al., 1998; Kerry-Williams et al., 1998; Bourbonnais et al., 2000; Egel-Mitani et al., 2000; Yao et al., 2009; Werten & de Wolf, 2005).

In bioinformatics, hidden Markov models are feasibly the most common statistical models (Haussler et al., 1993), whose applications are increasingly popular in the statistical analysis of biological sequences with complex correlations. Some examples of such applications include gene prediction (Krogh et al., 2001), protein modeling (Krogh et al., 1994), copy number reconstruction (Wang et al., 2007), sequence alignment (Hughey & Krogh, 1996), functional segmentation of the genome (Ernst & Kellis, 2012) and identification of ancestral DNA segments (Falush et al., 2003; Tang et al., 2006; Li & Durbin, 2011). In addition, in some studies, it has been used to identify genomic regions that contain regulatory information, such as the *cis*-regulatory modules (CRMs) and transcription factor binding sites (TFBSs) (Crowley et al., 1997; Frith et al., 2001; Rajewsky et al., 2002; Bailey & Nobel, 2003; Sinha et al., 2003).

In addition, several tools are utilized for the recognition of sorting signals as well as the prediction of subcellular localization of proteins from their amino acid sequences

(Imai & Nakai, 2020). For predicting signal peptides and their cleavage sites, many prediction methods, such as SPEPlip (Fariselli et al., 2003), SignalP 4.0 (Petersen et al., 2011), Phobius (Krogh et al., 2007), and DeepSig (Savojardo et al., 2018), have been developed. Moreover, SignalP has been further elevated as a deep neural network-based method that combines conditional random field classification and optimized transfer learning (SignalP-5.0; Armenteros et al., 2019). Other online tools that are used to predict possible protease cleavage for recombinant proteins include PROSPER (Song et al., 2012), MEROPS peptidase database (Rawlings et al., 2018), and ExPASy PeptideCutter tool (Gasteiger et al., 2005).

In recent times, yeasts have served as an alternative expression system in recombinant gene technology than the insect-baculovirus, bacterial and mammalian hosts (Gellissen et al., 2005). Unlike the prokaryotic expression system, yeasts are easily modified genetically and present a simple large-scale fermentation profile (Mattanovich et al., 2012). Additionally, the secret large amount of glycosylated proteins (Gellissen et al., 2005; Boer et al., 2007; Johanna et al., 2017). Recombinant protein production in yeasts can be targeted for food and non-food industries (Hensing et al., 1995). Furthermore, yeast cells act as an advantageous host for several productions of recombinant proteins in that they are considered as GRAS (Generally Recognized as Safe) by the American Food and Drug Administration (FDA) (Kim et al., 2014; Martinez et al., 2012; Nevoigt, 2008; Ahmad et al., 2014; Gellissen et al., 2005; Madzak et al., 2004; Van Ooyen et al., 2006).

A novel yeast expression system *M. guilliermondii* strain SO, which was isolated from spoiled orange (Oslan et al., 2012), had been reported to be an expression host for thermostable T1 lipase from *Geobacillus zalihae* under the regulation of alcohol oxidase promoter (pAOX1) with 14 (U/ml) in comparison to the commercial expression system, *P. pastoris*, with 81 (U/ml) (Oslan et al., 2014; Oslan et al., 2015). Interestingly, the T1 lipase could also be expressed without methanol induction (Abu et al., 2017). Nevertheless, strain SO had been used to express thermostable α-amylase using formaldehyde dehydrogenase promoter (pFLD) with a yield of 26 (U/mL) after 24 h of cultivation without induction (Nasir et al., 2020). Furthermore, W200R protease and diamine oxidase had been expressed in strain SO (Mahyon, 2017). Several optimization strategies had been carried out, yet a low expression level of these enzymes was recorded. This limitation has deterred the strain SO from being used widely as the expression host. This occurrence may be due to the proteolytic activity of the native vacuolar protease(s) of strain SO regarding the secretory pathway of heterologous proteins in yeast expression systems. Thus, it is imperative to identify the vacuolar protease(s) gene to enable future genomic disruption that might give a better yield of secreted heterologous proteins in strain SO as an expression host.

Recently, the genome/proteome of strain SO has been deposited in the Genbank (BioProject: PRJNA547962). With the availability of various bioinformatics tools, this

study aimed to determine the proteolytic level of strain SO and identify the putative vacuolar protease(s) of *M. guilliermondii* strain SO.

MATERIALS AND METHODS

Strain

M. guilliermondii strain SO was used during this study obtained from the previous study (Oslan et al., 2012) preserved in 80% glycerol stock solution.

Inoculum and Production Media Composition

Strain was grown for biomass production on YPD medium containing 1% (w/v) yeast extract, 2% (w/v) peptone, and 2% (v/v) dextrose; for expressed protease in the culture supernatant and pellet fractions, YPT medium containing 1% (w/v) yeast extract, 2% (w/v) peptone and 0.2% (w/v) tryptic soya broth and 4×10^{-5}% (w/v) biotin were used (Oslan et al., 2015).

Preparation of Inoculum

Wild type strain SO was streaked on YPD plate (1% (w/v) yeast extract, 2% (w/v) peptone, 2% (w/v) Bacteriological agar, and 2% (w/v) dextrose). A 10 mL YPD broth was inoculated using a single colony and incubated at 30°C overnight with an agitation of 250 rpm.

Preparation of Skim Milk agar for Qualitative Assay

The enzyme activity was also screened qualitatively on a skim milk agar plate prepared by dissolving 2.5 g of skim milk into 50 ml distilled water and autoclaved at 115°C for 5 min. Furthermore, 2 g of nutrient agar was dissolved in 50 ml of distilled water and autoclaved at 121°C for 15 min. Finally, the solutions were mixed in sterile conditions and poured into sterile petri-dishes (Rahman et al., 1994).

Determination of Protease Activity

For sample analysis, a prepared inoculum was incubated overnight at 30°C at 250 rpm. Subsequently, 1% (v/v) of the inoculum was transferred into a 100 mL YPD medium in a conical flask for biomass production. First, the culture media were grown for 24 h in an incubator shaker at 30°C with a shaking speed of 250 rpm. Then, the cells were centrifuged at 3000 × *g* at room temperature for 10 min and resuspended in 100 ml of YPT. Next, the cell was incubated at 30°C in a shaker at 250 rpm for three days. 5 ml yeast culture from the flask was harvested at 4000 x *g* for 10 min at 4°C every 24 h until three days. 5 ml of assay buffer (0.1 M Tris-HCl-2 mM $CaCl_2$ at pH 7.0) was used to resuspend the pellet obtained from the initial harvested culture and sonicated for 6 min with pause interval of

10 sec. Next, the sample was centrifuged at 10 000 × *g* for 7 min at 4°C. Then, a protease assay for extracellular and intracellular activity was undertaken.

Qualitative Protease Activity

For screenings, filtered crude extracellular and intracellular enzymes using cellulose acetate membrane filter with a pore size of 0.22 µm were obtained in sterile micro-centrifuge tubes. Furthermore, wells were bored on the skimmed milk agar plate using a sterile wire loop. Subsequently, 100 µl of enzymes was pipetted into the holes and incubated overnight at 30°C (Rahman et al., 1994).

Quantitative Protease Assay

The proteolytic activities were determined based on the method described by Rahman et al. (1994) using azocasein as the protein substrate with slight modification. The substrate mixture was freshly prepared by mixing 0.5% (w/v) azocasein in 0.1 M Tris-HCl-2 mM $CaCl_2$ at pH 7.0. A reaction mixture (100 µl of enzyme solution-supernatant and 1 ml of substrate mixture) was shaken in a horizontal water bath shaker for 30 min at 30°C. A 1.1 ml of 10% (w/v) TCA was used, and the sample was incubated at 25°C for 30 min to terminate the reaction. Subsequently, 1ml of the sample was transferred into a 1.5 ml tube and were later centrifuged at 10 000 × *g* for 7 min. Subsequently, 700 µl supernatant was transferred into a 1.5 ml tube and mixed with 700 µl of 1 M NaOH, and the absorbance of the mixture was measured at 450 nm where one unit (U) of protease activity is defined as the rate of the reaction that gives an increase of 0.001 absorbance unit at 450 nm per minute under the assay condition.

Statistical Analysis

The experimental data were further analyzed statistically using a t-test for paired two samples for means of expression level. SPSS version 2.0 was used to compute and evaluate the experimental results. $P<0.05$ was considered to indicate a statistically significant difference (Kirkwood & Sterne, 2003).

Bioinformatics Tools (Hardware and Software Used)

The hypothetical protease search was conducted on Windows 7, and the comparative genomic analysis was performed on iMac with MacOs Sierra with 10.12.1 version as the operating system.

National Center for Biotechnology Information (NCBI). The reference vacuolar proteases model sequence used for this study were *P. pastoris* aspartyl protease (PEP4) and

serine protease (PRB1) sequences retrieved from NCBI. The sequences were downloaded from http://www.ncbi.nlm.nih.gov/.

Hidden Markov Model. This installation was used to investigate the sequence obtained from the database for sequence homologies. It was downloaded from http://hmmer.org/download.html and installed in iMac. Xcode was also installed as a dependency for HMMER.

Molecular Evolutionary Genetic Analysis 7(MEGA7). It was downloaded and installed in Windows 7 from http://www.megasoftware.net/. Reported to be suited for analyzing protein and DNA sequence data from different living sources (Kumar et al., 2016). The retrieved proteases sequences from different yeast species from the database, NCBI, were aligned using the MUSCLE tool of MEGA 7.

Identification of the Vacuolar Protease(s) Sequence in *M. guilliermondii* Strain SO

Bioinformatics analysis of *M. guilliermondii* strain SO genome/proteome was performed using the nucleotide sequences deposited at the NCBI database (https://www.ncbi.nlm.nih.gov) BioProject: PRJNA547962. In addition, the *P. pastoris* sequence for PEP4 (accession number: CCA39046) and PRB1 (accession number: CCA36690) obtained from NCBI (Küberl et al., 2011; Valli et al., 2016) were taken as reference vacuolar protease model sequences because they had been well characterized and have been found similar to the aspartyl and serine proteases sequence of *S. cerevisiae* (Van Den Hazel et al., 1996). The position-specific iterated blast PSI-BLAST algorithm was selected using a cut-off E-value=0.01 using the Basic Local Alignment Search Tool (BLAST) with the reference proteins (refseq_protein) to search the database. From the third iteration values of the blast results, selection of closely related protease sequences of other yeast species was carried out based on percentage identity and alignment coverage ranging from 49%–99%. A total of 27 sequences of serine proteases and 24 sequences of aspartyl proteases were selected. Therefore, multiple sequence alignment (MSA) of the selected sequences for each model was created using MEGA7 software by MUSCLE tool, and the aligned files were saved in a FASTA format as PEP4_proteinaseA.fas and PRB1_proteinaseB.fas, respectively.

Identification of Homologous Vacuolar Proteases Sequence using Hidden Markov Model Software

The whole proteome of *M. guilliermondii* strain SO was obtained from GenBank, BioProject: PRJNA547962. Vacuolar aspartyl and serine protease models were assigned as PEP_ProteinaseA.hmm and PRB1_proteinaseB.hmm, respectively, to build the models. Primarily, the models were searched against *M. guilliermondii* strain, SO proteome as the

directory contained the targeted proteome of strain SO. The models were built using a set command with an E-value of 0.01. Subsequently, the search results (hits) were observed using ATOM software.

Analysis of the Hits Output

The hits output results for both vacuolar protease groups were further analyzed to identify the protein names and species homology. Hence, pairwise sequence alignment was carried out to determine the percentage similarities of the output of the hits with the reference model of *P. pastoris* vacuolar aspartyl and serine proteases, respectively. First, the hits output sequences were searched against the GenBank database NCBI (National Center for Biotechnology Information, National Library of Medicine) to identify similarities using the BLASTp search program. Next, multiple sequence alignment was carried out with the reference sequence used to build the HMM models using MEGA7 software. Subsequently, evolutionary analysis was inferred by constructing a phylogenetic tree using the maximum-likelihood method (Jones et al., 1992).

Furthermore, the molecular weight and theoretical pI of strain SO vacuolar aspartyl and serine proteases were predicted using ExPASy software tools (http://web.expasy.org/protparam/). The presence of signal peptide was inferred by using SignalP [http://www.cbs.dtu.gk/services/SignalP/]. An *in-silico* prediction of the cleavage sites on thermostable T1 lipase integrated with the identified vacuolar proteases model was also inferred using PROSPER, an online software (http://lightning.med.monash.edu.au/PROSPER/).

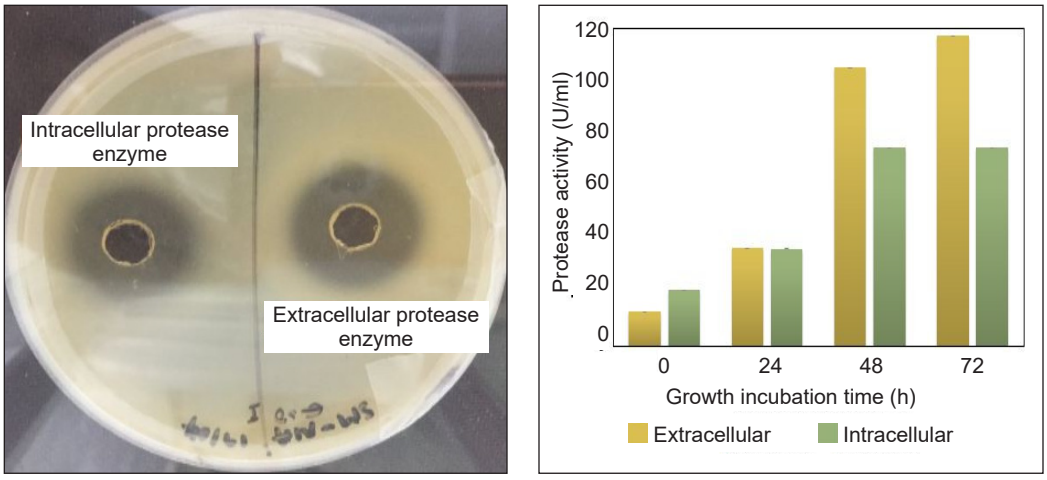
RESULTS

Determination of Native Protease(s) Activities of *M. guilliermondii* Strain SO

The quantitative and qualitative activity of strain SO native protease(s) were determined using azocasein and skimmed milk as the protein substrate. Qualitatively, the proteolysis zones could be directly observed by the presence of the clear halos around the wells for both extracellular and intracellular enzymes of M. *guilliermondii* strain SO native protease after 24 h of incubation time at 30^{o}C (Figure 1a).

P. pastoris strain X-33 was used as a positive control. The yeasts were grown in a YPD medium providing the yeast cells with a carbon source for biomass production. After generating mass, the cells were resuspended into a YPT medium (Clare et al., 1991). The cell culture was incubated at 30^{o}C for three days, and protease activity was determined along with the growth incubation time. The fermentation time course in shake flask for protease production by *M. guilliermondii* strain SO using 100µl (2mg) of the crude enzymes showed that the maximum protease activity (117.30 U/ml and 75 U/ml) were obtained after 72 h of cultivation for both extracellular and intracellular samples (Figure 1b). This

study showed that the native protease of *M. guilliermondii* strain SO exhibited a substrate specificity for azocasein and skim milk (Rahman et al., 1994). Suryawanshi and Pandya (2017) reported on the efficacy of screening of proteolytic activity using skimmed milk agar plates for screening and identification of alkaline proteases producing fungi.



(a) (b)

Figure 1. Total protease activity of *M. guilliermondii* strain SO: (a) Qualitative assay showing the clear zones of proteolysis due to the hydrolysis of the peptide bonds of the substrate by the enzymes; and (b) Quantitative native protease activity with standard deviation calculated.

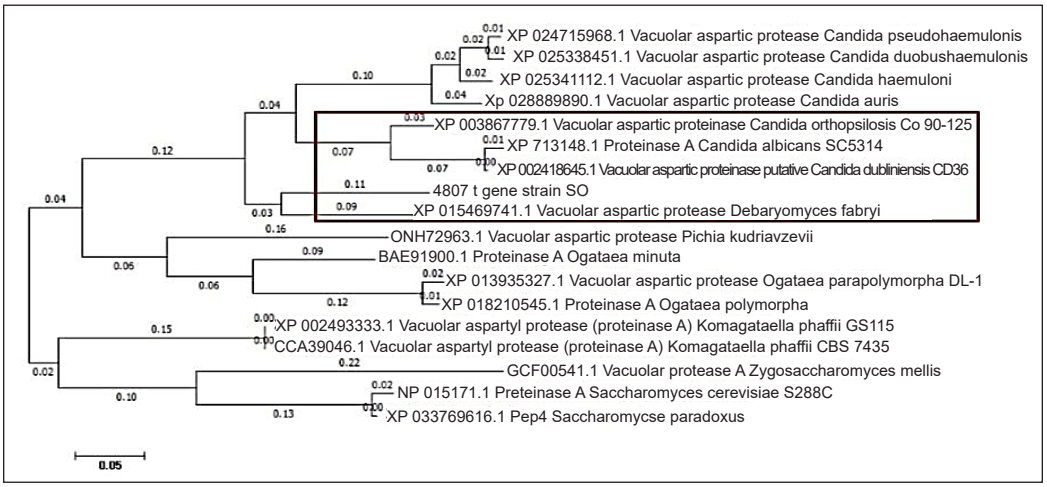
Identification of the Vacuolar Protease(s) Sequence in *M. guilliermondii* Strain SO Using HMM Strategy

HMM strategy was used to determine and identify the protease(s) sequence in *M. guilliermondii* strain SO proteome deposited in the GenBank with accession number: PRJNA547962. A total of 14 hits were observed from the aspartyl model built, and three hits were observed from the serine model built to detect protease sequence from *M. guilliermondii* proteome. The first hit observed from the results for each group was significantly matched to the multiple sequences aligned file because the E-value was chosen within the cut-off value of 0.01.

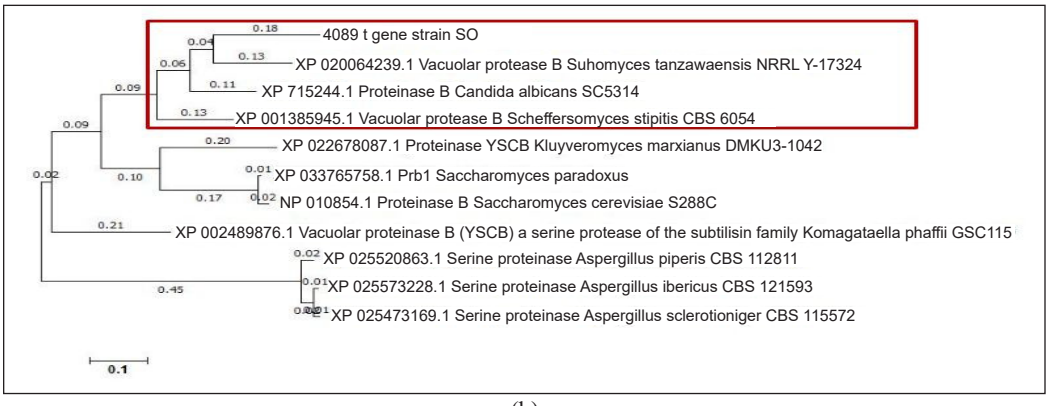
Analysis of the Hidden Markov Model Hits Output

The amino acid sequence of both vacuolar aspartyl and serine proteases were aligned with corresponding amino acids of *P. pastoris* vacuolar aspartic and serine proteases, respectively (Appendix - Tables S2a & S2b). Based on the pairwise alignment tools, the output of the first hit of serine protease and aspartyl protease of strain SO showed that it had 770 and 408 amino acid residues, respectively, and were 40.90% and 68.3% identical to *P. pastoris* PRB1 serine protease and PEP4 aspartyl protease respectively.

The protein BLAST results revealed that the serine protease sequence of strain SO had 70.91% identity with *C. albicans* (accession number: XP7152441). While aspartyl protease sequence of strain SO had 97.55% identity with *Meyerozyma* sp. JA9 (accession number: RLV83957.1). Furthermore, the protein sequences of the two groups of proteases were used to construct a phylogenetic tree with other yeasts vacuolar serine and aspartyl proteases selected from the reference protein sequence database. A query coverage of 80% above, e-value of 1e-15, and types of yeast family, which include Saccharomyceteceae, Debaryomyceteceae, Metschnokowiaceae, Aspergilleceae, Phaffomycetaceae, and Pichiaceae, were used. The phylogenetic trees drawn to scale using the maximum likelihood method (Jones, 1991) (Figures 2a & 2b) showed the nearness of the respective identified vacuolar protease of strain SO (aspartyl 4807t (PEP4) and serine 4089t (PRB1)) to Debaryomyceteceae yeast family.



(a)

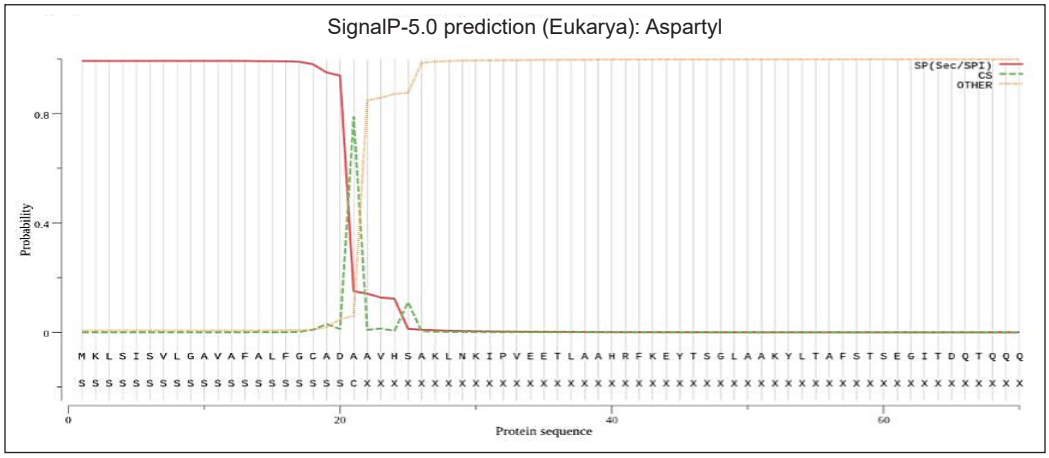


(b)

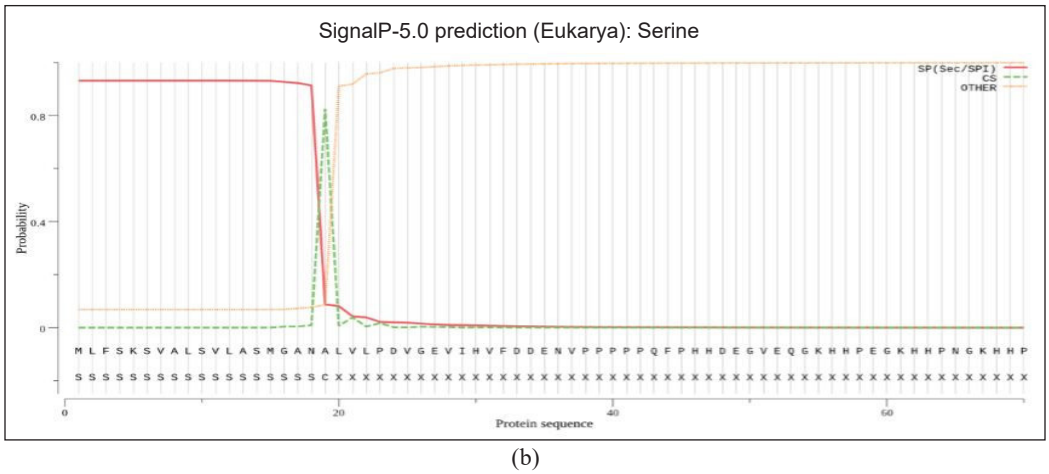
Figure 2. Phylogenetic tree construction: (a) Dendrogram showing the closeness of *M. guilliermondii* strain SO aspartyl protease (4807t gene) with the Debryomyceteceae family; and (b) Dendrogram showing the closeness of *M. guilliermondii* strain SO serine protease (4089t gene) with the Debryomyceteceae family.

Determination of Signal Peptide of Vacuolar Proteases of Strain SO

The Signal peptide (Sec/SPI) was predicted with the cleavage site for aspartyl protease of strain SO between positions 21 and 22 of the amino acid sequence ADA-AV (Alanine, Aspartic acid, Alanine–Alanine, Valine) with a probability of 0.7896. Furthermore, it showed that the protein was secreted extracellularly with an approximate molecular weight and theoretical pI of 44.2kDa and 4.57, respectively (Gasteiger et al., 2005) (Figure 3a). Likewise, for serine protease, the signal peptide was predicted with the cleavage site between positions 19 and 20 of the amino acid sequence ANA-LV (Alanine, Asparagine, Alanine–Leucine, Valine) with a probability of 0.8258. This result revealed that the protein was secreted extracellularly with approximate molecular weight and theoretical pI of 83.3kDa and 5.50, respectively (Gasteiger et al., 2005) (Figure 3b).



(a)



(b)

Figure 3. Assessment of the signal-peptides (SPs) of *M. guilliermondii* strain SO vacuolar proteases using SignalP software: (a) The signal peptide of aspartyl protease of *M. guilliermondii* strain SO predicted graph; and (b) Signal peptide of serine protease of *M. guilliermondii* strain SO predicted graph

More so, with PROSPER online protease cleave prediction tool, three and five cleavage sites respectively on recombinant protein T1 lipase from *G. zalihae* (Leow et al., 2004) as a substrate used for the integration (Appendix - Tables S3a & S3b) was predicted using model(s) enzymes. The prediction infers the possibility of degradation of the recombinant protein (T1 lipase) into peptides that would invariably reduce the expressed protein.

DISCUSSION

The native protease(s) activity over time was examined in the wild-type *M. guilliermondii* strain SO. Protease activity was observed to increase as the incubation time increased.

In other studies, the current strategies implemented to considerably reduce the effects of the native protease(s) in yeasts cells for recombinant protein secretion is by identifying the genes of the vacuolar proteases based on the secretory pathway of yeasts expression system and after that developing protease-deficient mutants (Copley et al., 1998; Gonzalez-Lopez et al., 2002; Jonson & Rehfeld 2004; Kang et al., 1998; Chung & Park, 1998). Presently, the proteolytic activities detected in both the extracted crude intra- and extracellular enzymes of *M. guilliermondii* strain SO were the total protease activity. The specific aspartyl and serine protease activities could only be detected once these two proteases are purified and characterized accordingly. Hence in this study, we report on the approach used to identify the two potential vascular proteases using genome/proteomic data of *M. guilliermondii* strain SO (BioProject: PRJNA547962).

Investigation and evaluation of the significant difference between the two variables obtained from the extracellular and intracellular protease activities of strain SO over time were reported. Statistically, according to Kirkwood and Sterne (2003), the null hypothesis is accepted because there is no significant difference between extracellular protease activity and intracellular protease activity as the alpha value of 0.05 is less than the t-value of 0.27 (Appendix - Table S1). HMM has been reviewed by various studies as an efficient and popular statistical tool, which utilizes mathematical principles to analyze data (Singh et al., 2012; Yoon, 2009; Gales & Young, 2007). It is an analytical approach with hidden states (Gomez-lopera et al., 2017). The vacuolar proteases hits from the HMM search yielded results contributing to the construction of vacuolar aspartyl and serine null mutant strain(s) in our future study. Several studies have reported on the identification of the vacuolar proteases (PEP4 and PRB1), which are mainly the bottlenecks of secretory heterologous proteins in a yeast expression system (Gleeson et al., 1998; Van den Hazel et al., 1996), which this study had identified in the novel yeast expression host *M. guilliermondii* strain SO proteome. The phylogenetic trees construction further revealed that the respective identified vacuolar protease of strain SO (aspartyl 4807t and serine 4089t) showed more nearness to the Debaryomyceteceae yeast family.

Signal peptides are short peptides located at the N-terminal of secretory proteins. In several industrial and scientific fields, disease diagnosis, and laboratory techniques, signal peptides have attracted great interest, particularly in producing recombinant protein, which determines efficient translocation from heterologous species (Zamani et al., 2015; Negahdaripour et al., 2017; Mousavi et al., 2017). The presence of the signal peptide in the native vacuolar protease(s) of strain SO has increased the secretion of these proteins in the medium. This finding might have explained why the proteolytic activity was higher in crude extracellular proteins environment as compared to intracellular. Furthermore, other studies have reported that the signal peptides have increased the secretion of the heterologous proteins at significant levels (Low et al., 2013; Mergulhao et al., 2005; Ohmuro-Matsuyama & Yamaji, 2017).

The signal-peptide was assessed by calculating the D-scores of Signal Peptides (SPs) using SignalP 5.0. The software predicts likely cleavage sites in ORF and signals peptides sequences. The web-based software uses an HMM algorithm and a neural network to compute a D-score to the input polypeptide sequence, using an average of the mean and the maximum score, which infers the possibility of a cleavage site.

Furthermore, PROSPER, a tool used for predicting protease substrate cleavage sites (Song et al., 2012), also predicted the possible cleavage sites of the enzyme model(s) for thermostable T1 lipase. Based on the analysis conducted in this study, *M. guilliermondii* strain SO vacuolar proteases have been identified as aspartyl protease and serine protease showing homology with other yeasts vacuolar aspartyl and serine proteases, respectively as inferred by the comparative sequence study, bioinformatics analysis, and evolutionary analysis conducted by the construction of the phylogenetic tree.

CONCLUSION

This study successfully determined and identified two vacuolar proteases, PEP4 (4807t) and PRB1 (4089t) of *M. guilliermondiii* strain SO employing activity assays and bioinformatics analysis approaches. The observed relationship with other yeasts' vacuolar proteases was indicated as the respective identified putative vacuolar proteases of strain SO were clustered among other yeasts' PEP4 and PRB1 proteases, respectively. PROSPER tool detected both PEP4 and PRB1 cleavage sites of the T1 lipase sequence. Signal peptides were also present in these proteases, thus confirming their secretory features. However, for the explicit understanding of their specific properties, the further characterization needs to be conducted in future research.

The identified vacuolar proteases PEP4 and PRB1 can be targeted for gene knockout to develop protease-deficient strains of *M. guilliermondii* strain SO and optimization of the cultivation parameters that may contribute to the numbers of protease mutant strains used to enhance secretion of heterologous proteins in yeast expression systems.

ACKNOWLEDGEMENTS

The authors wish to thank the Ministry of Education, Malaysia, for the Fundamental Research Grant (Code: FRGS/1/2019/STG05/UPM/02/1) and the Ministry of Education, Nigeria, allowing OEL to develop their career further.

REFERENCES

Abu, M. L., Nooh, H. M., Oslan, S. N., & Salleh, A. B. (2017). Optimization of physical conditions for the production of thermostable T1 lipase in *Pichia guilliermondii* strain SO using response surface methodology. *BMC Biotechnology, 17*, Article 78. https://doi.org/10.1186/s12896-017-0397-7

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*, 5301-5317. https://doi.org/10.1007/s00253-014-5732-5

Armenteros, J. J. A., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., & Brunak, S. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology, 37*, 420-423. https://doi.org/10.1038/s41587-019-0036-z

Bailey, T. L., & Noble, W. S. (2003). Searching for statistically significant regulatory modules. *Bioinformatics, 19*(suppl.2), 1116-1125. https://doi.org/10.1093/bioinformatics/btg1054

Barrett, A. J., Rawlings, N. D., & Woessner, J. F. (1998). *Handbook of proteolytic enzymes*. Academic Press.

Boer, E., Steinborn, G., Kunze, G., & Gellissen, G. (2007). Yeast expression platforms. *Applied Microbiology and Biotechnology, 77*, 513-523. https://doi.org/10.1007/s00253-007-1209-0

Bourbonnais, Y., Larouche, C., & Tremblay, G. M. (2000). Production of full-length human pre-elafin, an elastase specific inhibitor, from yeast requires the absence of a functional yapsin 1(Yps1p) endoprotease. *Protein Expression and Purification, 20*, 485-491. https://doi.org/10.1006/prep.2000.1338

Chung, B. H., & Park, K. S. (1998). Simple approach to reducing proteolysis during the secretory production of human parathyroid hormone in *Saccharomyces cerevisiae. Biotechnology and Bioengineering, 57*, 245-249. https://doi.org/10.1002/(SICI)1097-0290(19980120)57:2

Clare, J. J., Rayment, F. B., Ballantine, S. P., Sreekrishna, K., & Romanos, M. A. (1991). High-level expression of tetanus toxin fragment C in *Pichia pastoris* strains containing multiple tandem intergrations of the gene. *Biotechnology Journal, 9*, 455-60. https://doi.org/10.1038/nbt0591-455

Copley, K. S., Alm, S. M., Schooley, D. A., & Courchesne, W. E. (1998). Expression, processing and secretion of a proteolytically sensitive insect diuretic hormone by *Saccharomyces cerevisiae* requires the use of a yeast strain lacking genes encoding the Yap3 and Mkc7 endoproteases found in the secretory pathway. *Biochemical Journal, 330*, 1333-1340. https://doi.org/10.1042/bj3301333

Crowley, E. M., Roeder, K., & Bina, M. (1997). A statistical model for locating regulatory regions in genomic DNA. *Journal of Molecular Biology, 268*(1), 8-14. https://doi.org/10.1006/jmbi.1997.0965

Delic, M., Valli, M., Graf, A. B., Pfeffer, M., Mattanovich, D., & Gasser, B. (2013). The secretory pathway: Exploring yeast diversity. *FEMS Microbiology Reviews, 37*(6), 872-914. https://doi.org/10.1111/1574-6976.12020

Egel-Mitani, M., Andersen, A. S., Diers, I. I., Hach, M., Thim, L., Hastrup, S., & Vad, K. (2000). Yield improvement of heterologous peptides expressed in yps1-disrupted *Saccharomyces cerevisiae* strains. *Enzyme and Microbial Technology, 26*, 671-677. https://doi.org/10.1016/s0141-0229 (00)00158-7

Ernst, J., & Kellis, M. (2012). ChromHMM: Automating chromatin-state discovery and characterization. *Nature Methods, 9*, 215-216. https://doi.org/10.1038/nmeth.1906.

Falush, D., Matthew, S., & Jonathan, K. P. (2003). Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics, 164*, 1567-1587.

Fariselli, P., Finocchiaro, G., & Casadio, R. (2003). SPEPlip: The detection of signal peptide and lipoprotein cleavage sites. *Journal of Bioinformatics, 19*, 2498-2499. https://doi.org/10.1093/bioinformatics/btg360

Feyder, S., De Craene, J. O., Bär, S., Bertazzi, D. L., & Friant, S. (2015). Membrane trafficking in the yeast *Saccharomyces cerevisiae* model. *International Journal of Molecular Sciences, 16*(1), 1509-1525. https://doi.org/10.3390/ijms16011509

Forgac, M. (2000). Structure, mechanism and regulation of the clathrincoated vesicle and yeast vacuolar H(+)-ATPase. *Journal of Experimental Biology, 203*, 71-80.

Frith, M. C., Hansen, U., & Weng, Z. (2001). Detection of *cis*-element clusters in higher eukaryotic DNA. *Journal of Bioinformatics, 17*, 878-889.

Gales, M., & Young, S. (2007). The application of hidden Markov models in speech recognition. *Foundation and Trends in Signal Processing, 1*(3), 195-304. https://doi.org/10.1561/2000000004

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch A. (2005). Protein identification and analysis tools on the ExPASy server. In J. M. Walker (Ed.), *The proteomics protocols handbook* (pp 571-607). Humana Press.

Gellissen, G., Kunze, G., Gaillardin, C., Cregg, J. M., Berardi, E., Veenhuis, M., & Van Der Klei, I. (2005). New yeast expression platforms based on methylotrophic *Hansenula polymorpha* and *Pichia pastoris* and on dimorphic *Arxula adeninivorans* and *Yarrowia lipolytica*–a comparison. *FEMS Yeast Research, 5*(11), 1079-1096. http://doi.org/10.1016/j.femsyr.2005.06.004

Gleeson M. A. G., White C. E., Meininger D. P., & Komives E. A. (1998). Generation of protease-deficient strains and their use in heterologous protein expression. In D. R. Higgins & J. M. Cregg (Eds.), *Pichia protocols* (pp. 81-94). Humana Press. https://doi.org/10.1385/0-89603-421-6:81

Gomez-Lopera, J. F., Matinez-Aroza, J., Roman-Roldan, R., Roman-Galvez, R., & Blanco-Navarro, D. (2017). The evaluation problem in discrete semi-hidden Markov models. *Mathematics and Computers in Simulation, 137*, 350-365. https://doi.org/10.1016/j.matcom.2016.12.002

Gonzalez-Lopez, C. I., Szabo, R., Blanchin-Roland, S., & Gaillardin, C. (2002). Genetic control of extracellular protease synthesisin the yeast *Yarrowia lipolytica. Genetics, 160*, 417-427.

Haussler, D., Krogh, A., Mian, I. S., & Sjolander, K. (1993). Protein modeling using hidden Markov models: Analysis of globins. *In 1993 Proceedings of the Twenty-sixth Hawaii International Conference on System Sciences (Vol. 1, pp. 792-802). IEEE Publishing.* https://doi.org/*10.1109/HICSS.1993.270611*

Hensing, M. C., Rouwenhorst, M., Heijnen, R. J., Van, J. J., Dijken, J. P., & Pronk, J. T. (1995). Physiological and technological aspects of large-scale heterologous-protein production with yeasts. *Antonie Van Leeuwenhoek, 67*(3), 261-279. https://doi.org/10.1007/BF00873690

Hughey, R., & Krogh, A. (1996). Hidden Markov models for sequence analysis: Extension and analysis of the basic method. *Bioinformatics, 12*(2), 95-107. https://doi.org/10.1093/bioinformatics/12.2.95

Idiris, A., Tohda, H., Kumagai, H., & Takegawa, K. (2010). Engineering of protein secretion in yeast: Strategies and impact on protein production. *Applied Microbiology and Biotechnology, 86*, 403-417. https://doi.org/10.1007/s00253-010-2447-0

Imai, K., & Nakai, K. (2020) Tools for the recognition of sorting signals and the prediction of subcellular localization of proteins from their amino acid sequences. *Frontiers in Genetics, 11*, Article 607812. https://doi.org/10.3389/fgene.2020.607812

Johanna, M., Smeekens, H. X., & Wu, R. (2017). Global analysis of secreted protein and glycoproteins in *Saccharomyces cerevisiae*. *Journal of Proteome Research, 16*(2), 1039-1049. https://doi.org/10.1021/acs.jproteome.6b00953

Jones, D. T., Taylor, W. R., & Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Computer Applications in the Biosciences, 8*, 275-282. https://doi.org/10.1093/bioinformatics/8.3.275

Jones, E. W. (1991). Tackling the proteases problem in *Saccharomyces cerevisiae*. *Methods in Enzymology, 194*, 428-453. https://doi.org/10.1016/0076-6879(91)94034-a

Jonson, L., & Rehfeld, J. F. (2004). Enhanced peptide secretion by gene disruption of CYM1, a novel protease in *Saccharomyces cerevisiae. European Journal of Biochemistry, 271*(23-24), 4788-4797. https://doi.org/10.1111/j.1432-1033.2004.04443.x

Kang, H. A., Kim, S. J., Choi, E. S., Rhee, S. K., & Chung, B. H. (1998). Efficient production of intact human parathyroid hormone in a *Saccharomyces cerevisiae* mutant deficient in yeast aspartic protease 3 (YAP3). *Applied Microbiology and Biotechnology, 50*,187-192. https://doi.org/10.1007/s002530051275

Kerry-Williams, S. M., Gilbert, S. C., Evans, L. R., & Balance, D. J. (1998). Disruption of the *Saccharomyces cerevisiae* YAP3 gene reduces the proteolytic degradation of secreted recombinant human albumin. *Yeast, 14*, 161-169. https://doi.org/10.1002/(SICI)1097-0061(19980130)14:2

Kim, H., Yoo, S. J., & Kang, H. A. (2014). Yeast synthetic biology for the production of recombinant therapeutic proteins. *FEMS Yeast Research, 15*, 1-16. https://doi.org/10.1111/1567-1364.12195

Kirkwood, B., & Sterne, J. (2003). *Essential medical statistics* (2nd Ed.). Wiley-Blackwell.

Kobayashi, K., Kuwae, S., Ohya, T., & Ohda, T. (2000). High-level expression of recombinant human serum albumin from the methylotrophic yeast *Pichia pastoris* with minimal protease production and activation. *Journal of Bioscience and Bioengineering, 89*, 55-61. https://doi.org/10.1016/s1389-1723(00)88050-0

Krogh, A., Brown, M., Mian, I. S., Sjölander, K., & Haussler, D. (1994). Hidden Markov models in computational biology: Applications to protein modeling. *Journal of Molecular Biology, 235*(5), 1501-1531. https://doi.org/10.1006/jmbi.1994.1104

Krogh, A., Larsson, B., von Heijne, G., & Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *Journal of Molecular Biology, 305*(3), 567-580. https://doi.org/10.1006/jmbi.2000.4315

Krogh, A., Sonnhammer, E. L. L., & Ka, L. (2007). Advantages of combined transmembrane topology and signal peptide prediction - The Phobius web server. *Nucleic Acids Research, 35*, W429-W432. https://doi.org/10.1093/nar/gkm256

Küberl, A., Schneider, J., Thallinger, G. G., Anderl, I., Wibberg, D., Hajek, T., Jaenicke, S., Brinkrolf, K., Goesmann, A., Szczepanowski, R., Pühler, A., Schwab, H., Glieder, A., & Pichler, H. (2011). High-quality genome sequence of *Pichia pastoris* CBS7435. *Journal of Biotechnology, 154*(4), 312-320. https://doi.org/10.1016/j.jbiotec.2011.04.014

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution, 33*(7) 1870-1874. https://doi.org/10.1093/molbev/msw054

Leow, T. C., Rahman, R. N. Z. R. A., Basri, M., & Salleh, A. B. (2004). High level expression of thermostable lipase from *Geobacillus specie* strain T1. *Bioscience, Biotechnology, and Biochemistry, 68*(1), 96-103. https://doi.org/10.1271/bbb.68.96

Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature, 475*(7357), 493-496. https://doi.org/10.1038/nature10231

Li, S. C., & Kane, P. M. (2009). The yeast lysosome-like vacuole: Endpoint and crossroads. *Biochimica et Biophysica Acta, 1793*, 650-663. https://doi.org/10.1016/j.bbamcr.2008.08.003

Low, K. O., Mahadi, N. M., & IIias, R. M. (2013). Optimization of signal peptide for recombinant protein secretion in bacterial hosts. *Applied Microbiology and Biotechnology, 97*(9), 3811-3826. https://doi.org/10.1007/s00253-013-4831-z

Madzak, C., Gaillardin, C., & Beckerich, J. M. (2004). Heterologous protein expression and secretion in the non-conventional yeast *Yarrowia lipolytica*: A review. *Journal of Biotechnology, 109*, 63-81. https://doi.org/10.1016/j.jbiotec.2003.10.027

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

Martínez, J. L., Liu, L., Petranovic, D., & Nielsen, J. (2012). Pharmaceutical protein production by yeast: Towards production of human blood proteins by microbial fermentation. *Current Opinion in Biotechnology, 23*, 965-971. https://doi.org/10.1016/j.copbio.2012.03.011

Mattanovich, D., Branduardi, P., Dato, L., Gasser, B., Sauer, M., & Porro, D. (2012). Recombinant protein production in yeasts. *Methods in Molecular Biology, 824*, 329-358. https://doi.org/10.1007/978-1-61779-433-9_17

Mergulhao, F., Summers, D. K., & Monteri, G. A. (2005). Recombinant protein secretion in *Escherichia coli. Biotechnology Advances, 23*(3), 177-202. https://doi.org/10.1016/j.biotechadv.2004.11.003

Mousavi, P., Mostafavi, Z., & Morowvat, M. H. (2017). *In silico* analysis of several signal peptides for the excretory production of reteplase in *Escherichia coli. Current Proteomics, 14*(4), 326-335. https://doi.org/10.2174/1570164614666170809144446

Nasir, N. M., Leow, C., Oslan, S., Salleh, A., & Oslan, S. (2020). Molecular expression of a recombinant thermostable bacterial amylase from *Geobacillus stearothermophilus* SR74 using methanol-free *Meyerozyma guilliermondii* strain SO yeast system. *BioResource, 15*(2), 3161-3172

Negahdaripour, M., Nezafat, N., & Hajighahramani, N. (2017). *In silico* study of different signal peptides for secretory production of interleukin- 11 in *Escherichia coli. Current Protein and Peptide Science, 14*(2), 112-121. https://doi.org/10.2174/1570164614666170106110848

Nevoigt, E. (2008). Progress in metabolic engineering of *Saccharomyces cerevisiae*. *Microbiology and Molecular Biology Reviews, 72*, 379-412. https:/doi.org/10.1128/MMBR.00025-07

Ohmuro-Matsuyama, Y., & Yamaji, H. (2017). Modification of a signal sequence for antibody secretion from insect cells. *Journal of Cytotechnology, 70*(3), 891-898. https://doi.org/10.1007/s10616-017-0109-0

Oslan, S. N., Salleh, A. B., Rahman, R. A., Basri, M., & Chor, A. L. (2012). Locally isolated yeasts from Malaysia: Identification, phylogenetic study and characterization. *Acta Biochimica Polonica, 59* (2), 225-229.

Oslan, S. N., Salleh, A. B., Rahman, R. A., Zaliha, R. N., Leow, T. C., Sukamat, H., & Basri, M. (2015). A newly isolated yeast as an expression host for recombinant lipase. *Cellular and Molecular Biology Letters, 20*(2), 279-293. https://doi.org/10.1515/cmble-2015-0015

Oslan, S. N., Salleh, A. B., Rahman, R. A., Zaliha, R. N., Leow, A. T. C., & Basri, M. (2014). *Pichia pastoris* as a host to overexpress the thermostable T1 lipase from *Geobacillus zalihae*. *Journal of Biosciences, 3*(1), 7-17. https://doi.org/10.5176/2251-3140_3.1.45

Petersen, T. N., Brunak, S., von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nature Methods, 8*, 785-786. https://doi.org/10.1038/nmeth.1701

Rahman, R. N. Z. A., Razak, C. N., Ampon, K., Basri, M., Yunus, W. M. Z. W., & Salleh, A. B. (1994). Purification and characterization of a heat stable alkaline protease from *Bacillus stearothermophilus* F1. *Applied Microbiology and Biotechnology, 40*, 822-827. https://doi.org/10.1007/BF00173982

Rajewsky, N., Vergassola, M., Gaul, U., & Siggia, E. D. (2002). Computational detection of genomic *cis*-regulatory modules applied to body patterning in the early Drosophila embryo. *BMC Bioinformatics, 3*, Article 30. https://doi.org/10.1186/1471-2105-3-30

Rawlings, N. D., Barrett, A. J., Thomas, P. D., Huang, X., Bateman, A., & Finn, R. D. (2018). The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Research, 46*, D624-D632. https://doi.org/10.1093/nar/gkx1134

Savojardo, C., Martelli, P. L., Fariselli, P., & Casadio, R. (2018). DeepSig: Deep learning improves signal peptide detection in proteins. *Journal of Bioinformatics, 34*, 1690-1696. https://doi.org/10.1093/bioinformatics/btx818

Silva, C. I., Teles, H., Moers, A. P., & Eggink, G. (2011). Secreted production of collagen-inspired gel-forming polymers with high thermal stability in *Pichia pastoris. Biotechnology and Bioengineering, 108***,** 2517-525. https://doi.org/10.1002/bit.23228

Singh, B., Kapur, N., & Kaur, P. (2012). Speech recognition with Hidden Markov Model: A review. *International Journal of Advanced Research in Computer Science and Software Engineering, 2*(3), 401-403.

Sinha, J., Plantz, B. A., Inan, M., & Meagher, M. M. (2005). Causes of proteolytic degradation of secreted recombinant proteins produced in methylotrophic yeast *Pichia pastoris*: Case study with recombinant ovine interferon-tau. *Biotechnology and Bioengineering, 89*,102-112. https://doi.org/10.1002/bit.20318

Sinha, S., van Nimwegen, E., & Sigga, E. D. (2003). A probabilistic method to detect regulatory modules. *Journal of Bioinformatics, 19*(Suppl. 1), i292-i301. https://doi.org/10.1093/bioinformatics/btg1040

Song, J., Tan, H., Perry, A. J., Akutsu, T., Webb, G. I., & Whisstock, J. C. (2012). PROSPER: An integrated feature-based tool for predicting protease substrate cleavage sites. *PLoS ONE, 7*(11), Article e50300. https://doi.org/10.1371/journal.pone.0050300

Suryawanshi, H. K., & Pandya, N. D. (2017). Screening, identification of alkaline proteases producing fungi from soil of different habitats of Amalner Tahsil [Maharashtra] and their applications. *International Journal of Applied Sciences and Biotechnology, 5*(3), 397-402. https://doi.org/10.3126/ijasbt.v5i3.18304

Tang, H., Coram, M., Wang, P., Zhu, X., & Risch, N. (2006). Reconstructing genetic ancestry blocks in admixed individuals. *American Journal of Human Genetics, 79*, 1-12. https://doi.org/10.1086/504302

Valli, M., Tatto, N. E., Peymann, A., Gruber, C., Landes, N., Ekker, H., Thallinger, G. G., Mattanovich, D., Gasser, B., & Graf, A. B. (2016). Curation of the genome annotation of *Pichia pastoris* (*Komagataella phaffii)* CBS7435 from gene level to protein function. *FEMS Yeast Research, 16*(6), 1-12. https://doi.org/10.1093/femsyr/fow051

Van Den Hazel, H. B., Morten, C. K. B., & Jakob, R. W. (1996). Biosynthesis and function of yeast vacuolar proteases: Review. *Yeast, 12*, 1-16. https://doi.org/10.1002/(SICI)1097-0061(199601)12:1<1

Van Ooyen, A. J., Dekker, P., Huang, M., Olsthoorn, M. M., Jacobs, D. I., Colussi, P. A., & Taron, C. H. (2006). Heterologous protein production in the yeast *Kluyveromyces lactis*. *FEMS Yeast Research, 6*, 381-392. https://doi.org /10.1111/j.1567-1364.2006.00049.x

Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S. F. A., & Bucan, M. (2007). PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Research, 17*(11), 1665-1674. https://doi.org/10.1101/gr.6861907

Werten, M. W., & de Wolf, F. A. (2005). Reduced proteolysis of secreted gelatin and Yps1-mediated alpha-factor leader processing in a *Pichia pastoris* kex2 disruptant. *Applied and Environmental Microbiology, 71*, 2310-2317. https://doi.org/10.1128/AEM.71.5.2310-2317.2005

Wu, M., Shen, Q. M., Yang, Y., & Zhang, S. (2013). Disruption of YPS1 and PEP4 genes reduces proteolytic degradation of secreted HAS/PTH in *Pichia pastoris GS115. Journal of Industrial Microbiology and Biotechnology, 40*, 589-599. https:doi.org/10.1007/s10295-013-1264-8

Yao, X. Q., Zhao, H. L., Xue, C., Zhang, W., Xiong, X. H., Wang, Z. W., Li, X. Y., & Liu, Z. M. (2009). Degradation of HSA-AX15 (R13K) when expressed in *Pichia pastoris* can be reduced via the

disruption of YPS1 gene in this yeast. *Journal of Biotechnology, 139*, 131-136. https://doi.org/10.1016/j.jbiotec.2008.09.006

Yoon, B. (2009). Hidden Markov Models and their applications in biological sequence analysis. *Current Genomics, 10*(6), 402-415. https://doi.org/10.2174/138920209789177575

Zamani, M., Nezafat, N., & Negahdaripour, M. (2015). *In silico* evaluation of different signal peptides for the secretory production of human growth hormone in *E. coli. International Journal of Peptide Research and Therapeutics, 21*(3), 261-268. https://doi.org/10.1007/s10989-015-9454-z

APPENDIX

Table S1
Summary of statistical analysis of native protease activity assay of M. guilliermondii *strain SO*

	Extracellular	Intracellular
Mean	68.16666667	51.91666667
Variance	2604.481481	751.3611111
Observations	4	4
Pearson Correlation	0.994528638	
Hypothesized Mean Difference	0	
df	3	
t Stat	1.357279359	
P(T<=t) one-tail	0.133888423	
t Critical one-tail	2.353363435	
P(T<=t) two-tail	0.267776845	
t Critical two-tail	3.182446305	

Statistical analysis of the two variables obtained from the extracellular and intracellular protease activities of *M. guilliermondii* strain SO over time, shows that the t-value is greater than alpha value of 0.05

Table S2a
Pairwise sequence alignment of aspartyl protease hits from the HMM strategy with reference P. pastoris *aspartyl protease (PEP4)*

Hit(s)	Amino acid count	% identity	% similarity	Gaps	Score
4807	408	68.3%	79.8%	3.4	1510.0
4057	434	18.3%	31.7%	37.4	230
4064	388	20.8%	36.1%	28.3	246
1971	402	23.8%	38.4%	28.2	255
999	393	22.7%	35.8%	32.4	217
4978	414	25.2%	36.7%	26.5	296
4979	384	26.4%	38.6%	28.7	298
4980	446	24.3%	39.9%	19.1	244.5
4707	390	22.9%	41.6%	26.0	307
341	399	19.7%	29.8%	38.3	219
1972	504	23.8%	39.6%	18.1	280
70	549	22.3%	36.6%	28.1	264
57	334	23.7%	40.3%	23.9	259
2	344	19.6%	32.0%	37.8	177.5

Pairwise sequence alignment. **(a)** Showing aspartyl protease sequence hit output. Each hit (14) outputs were pairwise aligned with reference sequence using the EMBL-EBI search and sequence analysis tools. The first hits output from the HMM analysis with the lowest E-value shows the closest similarity with the reference model and has the highest score value. **(b)** Showing serine protease sequence hit output. Each hit (3) outputs were pairwise aligned with reference sequence using the EMBL-EBI search and sequence analysis tools. The first hits output from the HMM analysis with the lowest E-value shows the closest similarity with the reference model and has the highest score value.

Table S2b

Pairwise sequence alignment of serine protease hits from the HMM strategy with reference P. pastoris *serine protease (PRB1)*

Hit(s)	Amino acid count	% identity	% similarity	Gaps	Score
4089	770	40.9%	53.5%	30.7	1574.5
5205	1048	20.3%	27.5%	59.0	987
3337	381	26.3%	37.7%	35.1	542

Table S3a

Predicted potential cleavage sites (red) on thermostable T1 lipase using PROSPER aspartyl protease model

Position	Segment	N-fragment	C-fragment	Score
172	TLVN▮MVDF	21.40 kDa	26.16 kDa	1.14
156	HFVL▮SVTT	19.68 kDa	27.89 kDa	1.11
124	MLVS▮LLEN	15.74 kDa	31.82 kDa	1.06

Predicted cleavage sites using PROSPER online tool. **(a)** Showing three (3) predicted potential cleavage sites on thermostable T1 lipase using PROSPER aspartyl protease model. **(b)** Showing five (5) predicted potential cleavage sites on thermostable T1 lipase using PROSPER serine protease model.

Table S3b

Shows five (5) predicted potential cleavage sites (red) on thermostable T1 lipase using PROSPER serine protease model

Position	Segment	N-fragment	C-fragment	Score
360	DHLE▮IIGV	44.33 kDa	3.24 kDa	1.27
127	SLLE▮NGSQ	16.10 kDa	31.47kDa	1.02
149	PLFE▮GGHH	18.70 kDa	28.86 kDa	1.02
189	AVLE▮AAAV	23.46 kDa	24.11 kDa	1.00
315	RWLE▮NDGI	38.70 kDa	8.86 kDa	0.98