Sains Malaysiana 53(9)(2024): 3031-3044
http://doi.org/10.17576/jsm-2024-5309-10

Exploring the Protease Diversity of Psychrophilic Yeast, *Glaciozyma antarctica* through Genome Mining Analysis

(Meneroka Kepelbagaian Protease Yis Psikrofili, *Glaciozyma antarctica* melalui Analisis Perlombongan Genom)

Norfarhan Mohd-Assaad[1], Rohaiza Ahmad Redzuan[2] Mohd Faizal Abu Bakar[3], Nor Muhammad Mahadi[2], Abdul Munir Abdul Murad[2], Rosli Md. Illias[4], Doris Quay Huai Xia[1], Izwan Bharudin[2], Farah Diba Abu Bakar[2] & Shazilah Kamaruddin[2,*]

[1]*Department of Applied Physics, Faculty of Science & Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia*
[2]*Department of Biological Sciences and Biotechnology, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia*
[3]*Malaysia Genome & Vaccine Institute, Jalan Bangi Lama, 43000 Kajang, Selangor, Malaysia*
[4]*Department of Bioprocess Engineering, Faculty of Chemical and Energy Engineering, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia*

Received: 20 February 2024/Accepted: 17 July 2024

ABSTRACT

Proteases are one of the most significant classes of enzymes, holding immense physiological relevance and extensive industrial applications. The genome of *Glaciozyma antarctica* was fully sequenced, showing 7,857 open reading frames that offer an intriguing opportunity to investigate its proteolytic repertoire. This study aims to unveil the protease landscape of *G. antarctica*, a psychrophilic yeast that produces cold-active enzymes that offer remarkable benefits, particularly in the food and pharmaceutical industries. In this work, we performed a comprehensive analysis to identify the diverse families of proteases encoded within the *G. antarctica* genome and compare them with proteases from other mesophilic and thermophilic fungi in the MEROPS database. The sequence similarity searches resulted in the identification of 195 open reading frames predicted to encode for proteases in *G. antarctica* with a high number of intracellular proteases. These findings suggest an evolved system for protein quality control and turnover, essential for cell viability and adaptation to environmental stressors. The MEROPS classification analysis showed an abundance of metalloproteases, constituting 38% of the total protease genes, a proportion surpassing that found in other yeast and fungal genomes studied. This reflects the vital role of metalloproteases in the cold adaptation of microbes in the Antarctic region. This unique profile not only sheds light on the adaptive mechanisms of psychrophilic organisms but also presents a rich reservoir of potential cold-active proteases for various applications. The findings of this study provide a foundation for targeted enzyme discovery and engineering, unlocking new frontiers in industrial biotechnology and extremophile biology.

Keywords: Cold active enzyme; cold adaptation; comparative genomics; peptidase; polar microbiology

ABSTRAK

Protease adalah salah satu kelas enzim penting yang mempunyai peranan fisiologi yang besar dan aplikasi industri yang luas. Genom *G. antarctica* telah dijujuk sepenuhnya dan memaparkan sejumlah 7,857 rangka bacaan terbuka yang membuka peluang menarik untuk kajian himpunan enzim proteolitiknya. Kajian ini mendedahkan landskap protease *Glaciozyma antarctica*, yis psikofilik yang menghasilkan enzim aktif sejuk yang menawarkan manfaat yang luas terutamanya dalam industri makanan dan farmaseutik. Dalam kajian ini, satu analisis komprehensif telah dijalankan untuk mengenal pasti kepelbagaian keluarga protease yang dikodkan dalam genom *G. antarctica* dan melakukan analisis perbandingan dengan protease daripada kulat mesofil dan termofil dalam pangkalan data MEROPS. Analisis carian persamaan jujukan molekul telah mengenal pasti sebanyak 195 bingkai bacaan terbuka yang diramalkan sebagai gen mengekod protease dalam genom *G. antarctica* dengan bilangan gen mengekod protease intrasel adalah yang tertinggi. Penemuan ini mencadangkan satu evolusi dalam sistem kawalan kualiti dan kadar pusing ganti protein yang penting untuk kelangsungan hidup *G. antarctica* dan penyesuaian kepada tekanan alam sekitar. Analisis pengelasan MEROPS *G. antarctica* menunjukkan yis ini mempunyai bilangan gen mengekod metaloprotease yang tinggi iaitu kira-kira 38% daripada gen mengekod protease keseluruhan

di dalam genom. Peratusan ini adalah yang tertinggi jika dibandingkan dengan genom yis dan kulat yang telah dikaji. Ini mencerminkan peranan penting metaloprotease dalam penyesuaian mikrob sejuk di rantau Antartika. Profil protease yang unik ini bukan sahaja memberikan gambaran mekanisme penyesuaian organisme psikofil tetapi juga menyediakan reservoir genom yang kaya dengan protease aktif sejuk yang berpotensi untuk aplikasi yang pelbagai. Hasil kajian ini menyediakan asas untuk penemuan dan kejuruteraan enzim secara bersasar serta penerokaan sempadan ilmu baharu dalam bioteknologi industri dan biologi ekstremofil.

Kata kunci: Adaptasi sejuk; enzim aktif sejuk; genom perbandingan; mikrobiologi kutub; peptidase

INTRODUCTION

The interest in psychrophiles has recently increased due to their ability to produce novel enzymes and secondary metabolites that allow them to survive in harsh and cold conditions (Parvizpour et al. 2021; Yusof, Hashim & Bharudin 2021). Despite the environmental constraints, cold-adapted microorganisms can thrive in extreme cold conditions by evolving molecular mechanisms, such as the regulation of membrane fluidity, maintenance of protein synthesis, production of cold-acclimation proteins, mechanisms for freeze tolerance or avoidance, and alteration of protein functions that drive metabolism and cell cycle (Baeza et al. 2017; Bharudin et al. 2018). In addition, the ability of psychrophiles to produce a diverse array of hydrolytic enzymes allows it to degrade a wide range of polymeric substances, enabling the assimilation of various carbon sources. This metabolic versatility contributes to the recycling and mineralisation of organic matter in its Antarctic environment (Alcaíno, Cifuentes & Baeza 2015). For example, psychrophiles from Antarctic regions can thrive and produce proteases on a wide variety of substrates, indicating their potential for breaking down protein-rich substrates such as night soil (Dube, Singh & Alam 2001).

Protease (EC 3.4) is one of the most important groups of hydrolytic enzymes and represents a wide array of enzymes capable of acting on various proteinaceous substrates that are most abundant in soil of organic nitrogen (Geisseler et al. 2010; Schimel & Bennett 2004). These enzymes are associated with diverse essential functions in all living organisms, such as the synthesis of necessary biomolecules and cell differentiation, cell division, protein turnover and cascade-involved apoptosis, signal transduction, and vast infection regarding organism's life cycle (Gimenes, Silveira & Tambourgi 2019; Gurumallesh et al. 2019; Naveed et al. 2021). These proteolytic enzymes can be found in the cell cytoplasm and tethered to the cell surface. Although microorganisms can regulate protease production and secretion based on their needs for carbon and nitrogen, some proteases are secreted constitutively by microorganisms into the environment to initiate protein degradation (Geisseler & Horwath 2008).

Proteases comprise a substantial set of hydrolytic enzymes that catalyse the peptide bonds between amino acid residues, which results in the hydrolysis of proteins (Białkowska et al. 2016; Tavano et al. 2018). These enzymes are classified as exopeptidases or endopeptidases, depending on the site of action on polypeptide chains. Exopeptidases act on the ends of polypeptide chains while endopeptidases act randomly in the inner regions of polypeptide chains. Based on the catalytic residue found in the active sites, these peptidases can be further divided into six groups: serine, aspartic, cysteine, metallo, glutamic acid, and threonine protease (Gimenes, Silveira & Tambourgi 2019; Song et al. 2023). Besides being essential for life, they are also widely used in food, beverage, leather, detergent, pharmaceutical, and textile industries (Christensen et al. 2022; Martorell et al. 2019; Sarmiento, Peralta & Blamey 2015). Microbial protease has rapidly gained attention for industrial purposes due to their high yield production, lower energy consumption, consistent production, eco-friendliness, non-toxicity, and cost-effectiveness (Abada 2019; Choi, Han & Kim 2015; Liu & Kokare 2017).

An obligate psychrophilic yeast, *Glaciozyma antarctica* PI12 was isolated from sea ice near Casey Research Station, Antarctica, and the draft genome was successfully analysed and published (Firdaus-Raih et al. 2018). It has been reported that this yeast encodes numerous unique proteins with novel structures and functions facilitating the organism's growth and survival in extremely cold conditions such as producing cold-active enzymes (Kamaruddin et al. 2022; Mohamad Nor et al. 2020; Yusof, Hashim & Bharudin 2021). The need for organisms to overcome the detrimental effect of low temperature suggests that a wide variety of cold-active enzymes might be functional for industrial applications. Some useful properties of these enzymes for biotechnological applications include high catalytic efficiency at low temperatures leading to lower concentrations of enzymes being used, avoidance of heating steps in industrial settings and selective inactivation by moderate heating (Feller 2013; Kuddus 2018). Unusual flexibilisation of protein structure in cold conditions permits the enzymes to have reduced activation energy, which results in high catalytic efficiency and active at low optimum temperature. Due to the reduction in energy consumption and cost, the exploitation of cold-active proteases from psychrophile microbes for industrial applications is fast becoming an attractive option (Furhan 2020; Joseph, Kumar & Ramteke 2019; Yang et al. 2023). The advancement in genomics and bioinformatics approaches, including accessibility to the complete genome sequences of several mesophilic and thermophilic yeast and fungal species, provides a prominent

way to carry out a comprehensive analysis to investigate the distribution pattern of proteases in *G. antarctica*. In this paper, we incorporated multiple bioinformatics analyses to sort out proteases that may have potential in industrial and biotechnological applications and may be essential in the ecological role of *G. antarctica* cold-adaptation process.

MATERIALS AND METHODS

PROTEASE IDENTIFICATION AND PROTEASE FAMILY ASSIGNMENT

The gene sequences from *G. antarctica* PI12 genome (BioProject Accession PRJNA202387) were blasted against NCBI non-redundant database, UniProtKB/SwissProt database, and UniProtKB/TrEMBL database (http://www.uniprot. org/) with the identity and E-value scores set at >30% and <1e-04, respectively. Each significant hit was then used as query against the MEROPS database of peptidase sequence collection (Rawlings et al. 2014) database for protease family assignment. We included proteases from thermophilic (*Chaetomium thermophilum*), mesophilic (*Neurospora crassa*, *Arthroderma gypseum*, *Saccharomyces cerevisiae*, and *Cryptococcus neoformans*) and psychrophilic (*Rhodotorula glutinis* and *Geomyces destructans*) fungi, in which the whole genome sequences are complete to compare the distribution of peptidases across different fungal classes. The peptidase sequences of each fungal species were retrieved from the MEROPS database (accessed on 23rd March 2021) for the comparative analysis of peptidase distribution alongside the peptidases in *G. antarctica*.

PHYLOGENETIC RECONSTRUCTION OF SEVERAL FUNGI OF DIFFERENT GROWTH TEMPERATURES

The phylogenetic relationship among all fungal species included in this study was inferred using the 18 rRNA sequence available in the SILVA database (Quast et al. 2013). The sequences were aligned using MAFFT v7.453 (Katoh & Standley 2013) with --auto option and 1,000 iterative refinement cycles. The phylogenetic analysis was performed using the maximum-likelihood approach in MEGA X (Kumar et al. 2018; Stecher, Tamura & Kumar 2020) using a TN93+G model with 1,000 replications. The nucleic acid substitution model was chosen according to the Bayesian information criterion output from the model selection test embedded in MEGA X.

PREDICTION OF EXTRACELLULAR PROTEASES

Extracellular proteases secreted by *G. antarctica* were sorted based on the presence of secretion signal, prediction of subcellular location, and lack of transmembrane domain. Secretory signal peptides were identified using SignalP 4.1 (http://www.cbs.dtu.dk/services/SignalP/) (Petersen et al. 2011), transmembrane domains were identified using Phobius (http://phobius.cgb.ki.se/) (Kall, Krogh & Sonnhammer 2007), and subcellular location of peptidases were predicted using WoLF PSORT (https://wolfpsort.hgc.jp/) (Horton et al. 2007).

RESULTS AND DISCUSSION

Genome mining of *G. antarctica* was performed to identify putative protease sequences using the BLAST tool against UniProt and NCBI databases. Sequence identity at >30% is a well-established practice for homolog similarity in selecting genes due to its balance of evolutionary, structural, functional, and practical considerations (Ding & Dokholyan 2006). Using this threshold in this analysis helps in identifying protease homologs, ensuring functional relevance, and facilitating gene function prediction and annotation. Lower E-values suggest a higher degree of similarity and stronger evidence that the sequences are homologous (Choudhuri 2014; Pearson 2013). The identification of protease genes was based on the presence of a protease family conserved domain in the query sequence. Homolog proteases were then appended to the MEROPS database for peptidase family assignment. A total of 195 open reading frames (ORFs) were predicted as peptidase family genes and classified into six families, which are threonine protease, serine protease, metalloprotease, cysteine protease, aspartic protease, and uncharacterised proteins (Figure 1). Predominant ORFs were predicted enriched with metalloprotease genes (74 genes), and metalloprotease was determined as the largest group of peptidase families in *G. antarctica*, followed by serine protease, cysteine protease, threonine protease, and aspartic protease with 40, 37, 16, and 9 genes, respectively. On the contrary, 20 genes were observed as uncharacterised proteins. These peptidase gene families were further assigned into sub-peptidase families. Within each peptidase family, the number of members per subfamily varied, ranging from 1 to 16 genes (Figure 1). The peptidase families C19, T1, M41, S9, and A1 constitute the highest number of peptidases in *G. antarctica* with 16, 14, 11, 8, and 7 peptidases in each subfamily, respectively. Collectively, these five peptidase subfamilies account for approximately 30% of the total 53 classified peptidase subfamilies identified in the *G. antarctica* genome. This indicates that these families play an outsized role in the peptidase complement of this psychrophilic yeast. The prevalence of these peptidase subfamilies likely reflects their importance in enabling *G. antarctica* to thrive in its cold, Antarctic environment. The diverse array of hydrolytic enzymes produced by these peptidases likely allows the bacterium to efficiently degrade a wide range of polymeric substances and assimilate various carbon sources - a key metabolic adaptation for surviving in its nutrient-

limited habitat (Yin et al. 2024). In contrast, there are some subfamilies containing only a single member, as illustrated in Figure 1. These group may represent enzymes that have evolved to perform very specific functions or adaptations within the host organism. Overall, the abundance and distribution diversity of peptide families in *G. antarctica* may reflect the specialised roles these enzymes play.

The distribution of peptidase families and their corresponding super-families across microbial taxa were then compared across different fungal species inhabiting different temperature ranges. The phylogenetic relationship of *G. antarctica* and eight fungal species (three psychrophilic, four mesophilic, and one thermophilic) was reconstructed and compared (Figure 2). In general, extremophile fungi contain a smaller number of peptidases compared to their mesophilic counterparts. Mesophilic fungi colonise a wide range of areas surrounded by a variety of nutrient availability, hence the requirement for different types of enzymes to utilise each type of nutrient (Sabath et al. 2013). Meanwhile, extremophiles that inhabit low variability habitats require only enzymes that can metabolise nutrients efficiently at specific temperatures and tend to have smaller genomes, hence a smaller metabolic network. Among the four extremophiles, *G. antarctica* harboured the largest peptidase content, with 195 peptidases in total.

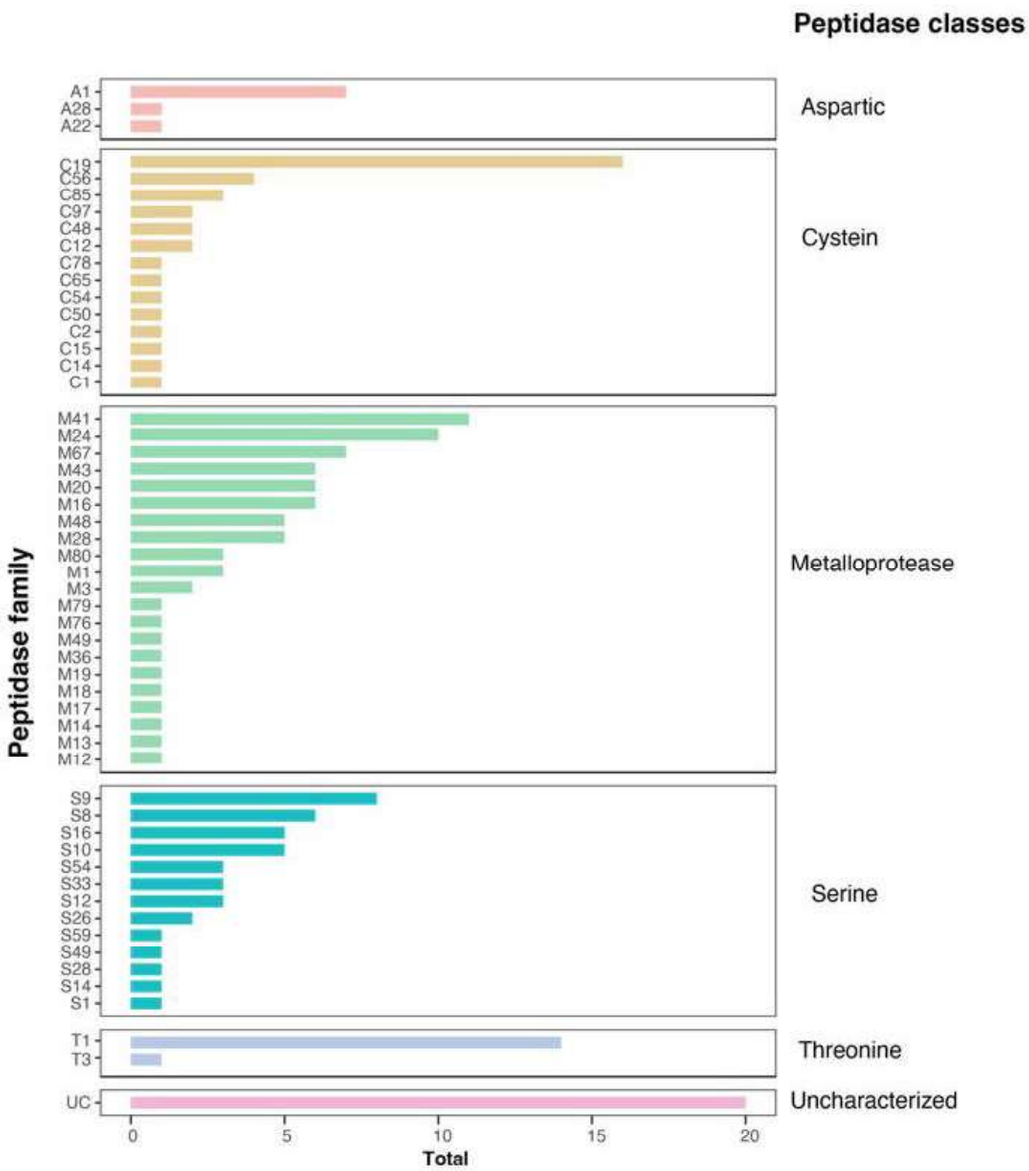

FIGURE 1. Distribution of peptidase families in *G. antarctica*

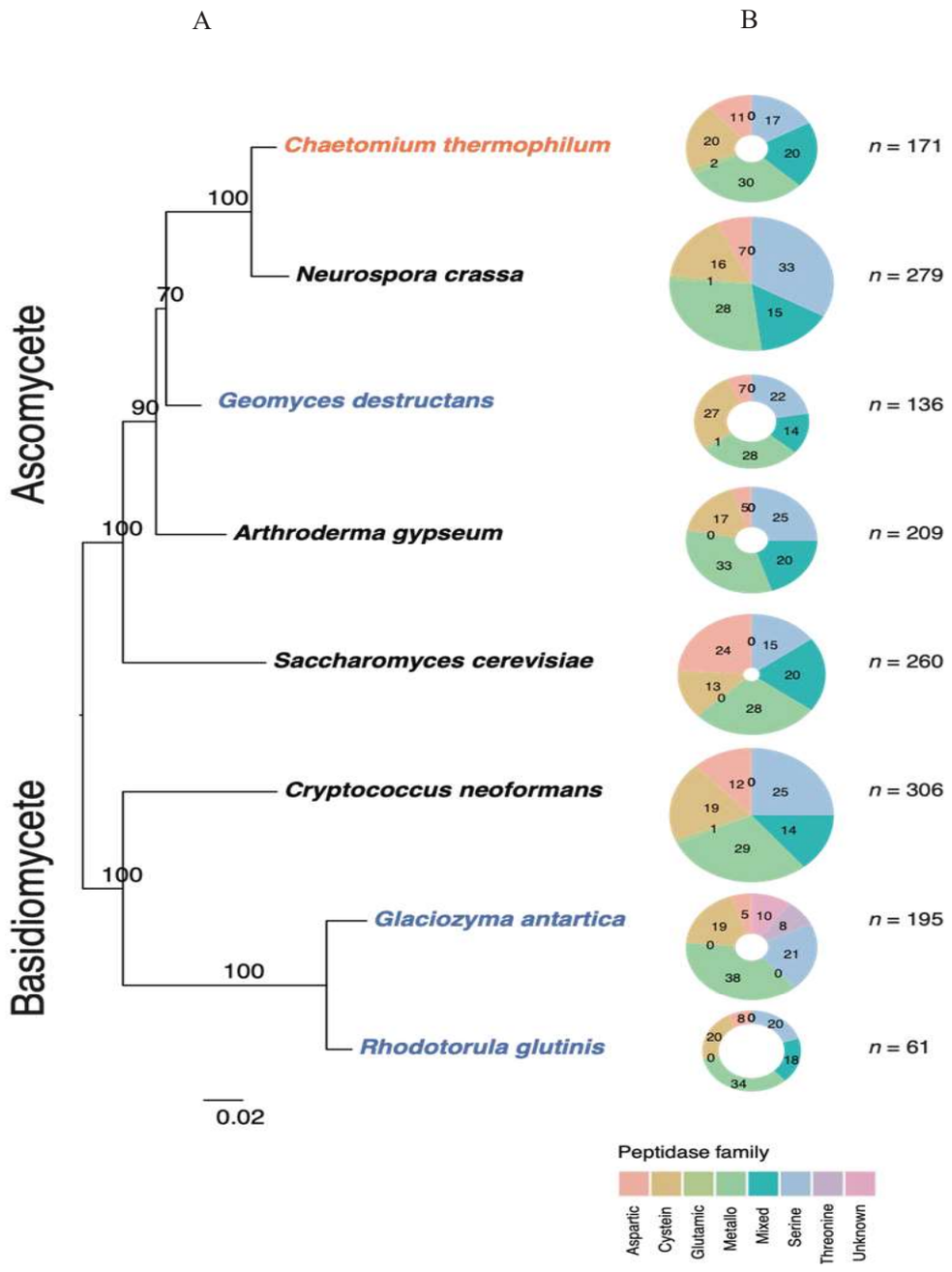

FIGURE 2. Phylogenetic tree of eight fungal species and the peptidase family distribution. (A) Species highlighted in orange, blue, and black indicate thermophiles, psychrophiles, and mesophiles, respectively, and (B) Pie charts show the proportion of different protease families identified in each species with the total n number of the proteases identified in each genome. The size of the pie chart is relative to the total proteases identified in the respective species

Comparative analysis of peptidase families showed that the distribution of peptidases in fungi and yeasts is enriched with metalloproteases proteolytic enzymes, which metalloproteases is known to be the most densely populated catalytic class of protease in many organisms (Ugalde et al. 2010). Psychrophilic yeast *G. antarctica* and *R. glutinis* possess an immense family of metalloprotease with 38% and 34%, respectively, followed by 33% and 30% from mesophilic fungal *A. gypseum* and thermophilic fungal *C. thermophilum*, respectively. Previous studies discovered

that microbes isolated from various environmental areas in Antarctica secrete cold-active metalloprotease (Matsui et al. 2017; Santos et al. 2015; Vazquez, Coria & Mac 2004; Zhou et al. 2013). Furthermore, metalloprotease is among the major extracellular proteases identified in bacteria isolated from the Antarctic coastal sediments (Zhou et al. 2013). This complies with the role of metalloproteases as a regulator in essential biological processes, which include nutrient absorption, protein turnover, cell fusion, cell adhesion and migration, and extracellular matrix remodelling for microbial defense (Wu & Chen 2011). The high redundancy of metalloproteases in psychrophilic organisms could be attributed to their important ecological roles in the recycling of organic substances (Vazquez, Coria & Mac 2004). The abundance of metalloprotease genes presents in the genome of these psychrophilic yeast implies the involvement in the modification and adaptation of biological processes within a cold environment. Some metalloproteases can perform efficient catalysis in extreme conditions (Adekoya & Sylte 2009). This finding suggests its vital role in Antarctic regions, which may explain the high proportion of this peptidase family in *G. antarctica* and *R. glutinis*.

The presence of many predicted proteases in the genome of *G. antarctica*, both intracellular and extracellular, suggests that proteolytic activity is an essential part of its biology and adaptation to the Antarctic environment. The list of extracellular and intracellular proteases can be referred to in Tables 1, 2 and 3. Figure 3 shows that out of 195 proteases identified in the genome of *G. antarctica*, 44 proteins were predicted to be extracellular proteases by at least one of the three bioinformatics software tools used in the analysis (WoLF PSORT, SignalP 4.1, and Phobius). Secretion signal is identified from the presence of a signal peptide (also known as signal sequence, targeting signal, localisation signal, localisation sequence, transit peptide, leader sequence, or leader peptide), which consists of a short peptide (16-30 long). This signal presents at the N-terminus of many newly synthesised proteins that are destined toward the secretory pathway (Kapp et al. 2009; Wu et al. 2020). Besides that, 151 genes were predicted as intracellular proteases. This is likely an important adaptation for survival in the extreme Antarctic environment, where nutrient availability is limited and cellular damage from cold temperatures is a constant challenge.

Intracellular proteases play a vital role in the regulation of protein turnover and quality control within cells (Sommerfield & Darwin 2022). The high number of intracellular proteases identified in *G. antarctica* suggests that this species has evolved a robust system for protein quality control and turnover, which is essential for cell viability and adaptation to environmental stressors. These proteases serve to eliminate proteins that are damaged, misfolded, or no longer needed. By degrading these proteins, they prevent their accumulation, which could otherwise impede the normal functioning of cells. This process helps to maintain cellular integrity and ensure optimal protein quality and functionality (Burgos et al. 2020).

Extracellular proteases, on the other hand, are involved in the degradation of extracellular matrix components and the processing of signalling molecules. They actively degrade various exterior proteins, providing carbon and nitrogen essential for nutrient acquisition for themselves and potentially other neighbouring microorganisms. Additionally, these proteases promote carbon and nitrogen recycling and also energy metabolism in many organisms (Cheng et al. 2021). In the case of *G. antarctica*, these proteases may play a role in the degradation of food particles and other organic matter, allowing *G. antarctica* to access the nutrients needed for survival.

It is also important to note that the proteases from *G. antarctica* that have been identified and classified in this study could potentially be utilised in industrial applications, particularly in the food and pharmaceutical industries, where proteases are commonly used for their ability to break down proteins and peptides. Therefore, further research into the properties and functions of these predicted extracellular proteases could have practical applications beyond the study of Antarctic krill adaptation.

CONCLUSION

The availability of the genome sequence of *G. antarctica* has allowed the analysis of peptidase repertoire in this psychrophile yeast. The *G. antarctica* genome consists of at least 195 putative proteases which were assigned into 7 peptidase families and 53 subfamilies by the MEROPS classification system. Our findings provide an overview of peptidase diversity in *G. antarctica* that provides a foundation and information on the evolution of proteolysis in psychrophiles. In this work, by predicting functional extracellular enzymes from *G. antarctica*, we identified essential cold proteases involved in adaptability processes that could be exploited for applications in biotechnology and industry. The exploration of potential cold-adapted enzymes from *G. antarctica* will help in developing a new source of cold-adapted enzymes, especially cold-adapted proteases that present several advantages and possible use in several industrial processes. Therefore, further research into the properties and functions of these predicted proteases could have practical applications beyond the realm of cold-adapted organisms. By delving deeper into their characteristics, we can unlock their full potential and explore novel avenues for their use in various industrial processes. These advancements in enzyme-based technologies and bioprocessing have the potential to significantly improve efficiency and product development across industries. Hence, it is imperative to continue investigating and harnessing the practical applications of these proteases beyond their role in the adaptation of cold-adapted organisms.

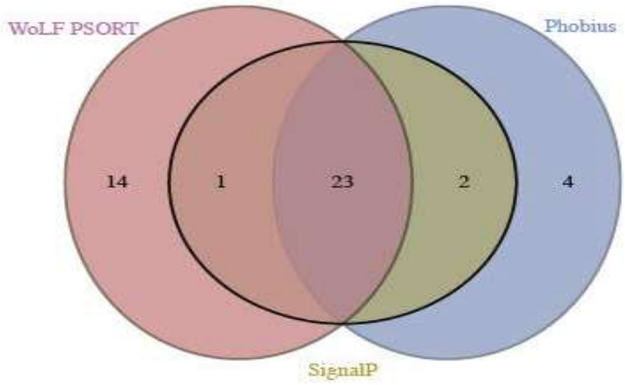

FIGURE 3. Venn diagram of the output from three bioinformatic softwares used to identify secreted proteases in *G. antarctica*. A total of 44 proteins were predicted as extracellular proteases by at least one of the three identifiers (WoLF PSORT, SignalP 4.1, and Phobius) used in this study

TABLE 1. List of proteases predicted as extracellular protein

GA Gene ID	Description
LAN_01_079	Putative peptidase S8 family protein
LAN_01_112	Putative peptidase A1 family protein
LAN_02_200	Putative peptidase M43B family protein
LAN_02_205	Putative peptidase M43B family protein
LAN_02_206	Putative peptidase M43B family protein
LAN_03_399	Putative peptidase S8 family protein
LAN_04_531	Putative peptidase
LAN_06_141	Putative peptidase A1 family protein
LAN_06_252	Putative peptidase S8 family protein
LAN_06_367	Putative peptidase M28E subfamily protein
LAN_08_059	Putative extracellular metalloprotease protein
LAN_08_202	Putative peptidase
LAN_08_241	Putative peptidase S8 family protein
LAN_11_125	Putative metallopeptidase-domain protein
LAN_12_170	Putative peptidase S10 family protein
LAN_12_254	Putative peptidase A1 family protein
LAN_12_491	Putative peptidase S10 family protein
LAN_13_044	Putative peptidase A1 family protein
LAN_13_322	Putative peptidase A1 family protein
LAN_16_048	Putative amino peptidase
LAN_16_282	Putative peptidase S28 family protein
LAN_16_401	Putative extracellular metalloproteinase
LAN_16_715	Putative subtilisin-like serine protease
LAN_03_764	Putative 'GDXG' lipolytic enzyme family protein
LAN_06_355	Putative peptidase M24 family protein
LAN_07_014	Putative peptidase S8 family protein
LAN_08_393	Putative isonitrile hydrataseIsonitrile hydratase

continue to next page

continue from previous page

LAN_08_396	Putative PfpI endopeptidase protein
LAN_09_226	Putative peptidase S9B family protein
LAN_10_037	Putative CAAX prenyl protease
LAN_10_238	Putative peptidase M20A family protein
LAN_11_286	Putative peptidase S9A family protein
LAN_13_316	Putative peptidase M14 family protein
LAN_14_108	Putative peptidase M20A family protein
LAN_16_234	Putative serine carboxypeptidase
LAN_16_413	Putative carboxypeptidase
LAN_16_565	Putative peptidase M19 family protein
LAN_05_095	Putative eIF-3 subunit
LAN_07_145	Putative ubiquitin carboxyl-terminal hydrolase
LAN_12_234	Putative methionine aminopeptidas
LAN_15_219	Putative signal peptidase complex subunit
LAN_14_292	Putative peptidase M12B domain containing protein
LAN_03_704	Putative peptidase M20A family protein
LAN_08_019	Putative nucleophile aminohydrolase

TABLE 2. List of proteases predicted as extracellular protein by three software

GA Gene ID	Description
LAN_01_079	Putative peptidase S8 family protein
LAN_01_112	Putative peptidase A1 family protein
LAN_02_200	Putative peptidase M43B family protein
LAN_02_205	Putative peptidase M43B family protein
LAN_02_206	Putative peptidase M43B family protein
LAN_03_399	Putative peptidase S8 family protein
LAN_04_531	Putative peptidase
LAN_06_141	Putative peptidase A1 family protein
LAN_06_252	Putative peptidase S8 family protein
LAN_06_367	Putative peptidase M28E subfamily protein
LAN_08_059	Putative extracellular metalloprotease protein
LAN_08_202	Putative peptidase
LAN_08_241	Putative peptidase S8 family protein
LAN_11_125	Putative metallopeptidase-domain protein
LAN_12_170	Putative peptidase S10 family protein
LAN_12_254	Putative peptidase A1 family protein
LAN_12_491	Putative peptidase S10 family protein
LAN_13_044	Putative peptidase A1 family protein
LAN_13_322	Putative peptidase A1 family protein
LAN_16_048	Putative amino peptidase
LAN_16_282	Putative peptidase S28 family protein
LAN_16_401	Putative extracellular metalloproteinase
LAN_16_715	Putative subtilisin-like serine protease

TABLE 3. List of proteases predicted as intracellular protein

GA Gene ID	Description
LAN_01_205	Putative peptidase S26B family protein
LAN_01_234	Putative peptidase S33 family protein
LAN_01_245	Putative AAA ATPase
LAN_01_306	Putative peptidase M16 family protein
LAN_01_352	Putative RING-type zinc finger domain-containing protein
LAN_02_035	Putative thioredoxin domain-containing protein
LAN_02_074	Putative 26S proteasome regulatory subunit protein
LAN_02_085	Putative 26S protease subunit protein
LAN_02_155	Putative ubiquitin carboxyl-terminal hydrolase
LAN_02_160	Putative proteasome subunit alpha
LAN_02_208	Putative peptidase M43B family protein
LAN_02_218	Putative FACT complex subunit protein
LAN_02_220	Putative SAC3 family protein
LAN_02_246	Putative RING-type zinc finger domain-containing protein
LAN_03_039	Putative peptidase M24A family protein
LAN_03_073	Putative peptidase M24A family protein
LAN_03_115	Putative peptidase M16 family protein
LAN_03_130	Putative peptidase M16 family protein
LAN_03_182	Putative AAA ATPase
LAN_03_187	Putative C2H2-type zinc finger and OUT domain-containing protein
LAN_03_188	Putative peptidase C19 family protein
LAN_03_192	Putative MCM family protein
LAN_03_241	Putative AAA ATPase
LAN_03_400	Putative peptidase T1B family protein
LAN_03_411	Putative peptidase M48 family protein
LAN_03_630	Putative 20S proteasome, A and B subunits-containing protein
LAN_03_679	Putative peptidase S54 family protein
LAN_03_736	Putative OTU domain-containing protein
LAN_03_740	Putative peptidase A22B family protein
LAN_04_142	Putative DNA damage-inducible protein 1
LAN_04_161	Putative peptidase S54 family protein
LAN_04_178	Putative peptidase C14B family protein
LAN_04_182	Putative MPN (JAB/Mov34) domain-containing protein
LAN_04_229	Putative peptidase C19 family protein
LAN_04_293	Putative peptidase T1B family protein
LAN_04_337	Putative eIF-3 subunit
LAN_04_367	Putative beta-lactamase/transpeptidase-like protein
LAN_04_499	Putative proteasome subunit
LAN_04_500	Putative peptidase T1B family protein
LAN_04_519	Putative peptidase
LAN_05_020	Putative thimet oligopeptidase
LAN_05_022	Putative biquitin carboxyl-terminal hydrolase 5
LAN_05_189	Putative peptidase T1A family protein

continue to next page

continue from previous page

LAN_05_214	Putative midasin family protein
LAN_05_306	Putative peptidase C19 family protein
LAN_05_327	Putative peptidase S9B family protein
LAN_06_019	Putative AAA ATPase
LAN_06_102	Putative peptidase S49 family protein
LAN_06_114	Putative MCM family protein
LAN_06_251	Putative peptidase M67 family protein
LAN_06_332	Putative MCM family protein
LAN_06_347	Putative proteasome subunit S2 family protein
LAN_07_045	Putative activator 1 small subunits family protein
LAN_07_153	Putative O-sialoglycoprotein endopeptidase
LAN_07_168	Putative protein YME1 homolog
LAN_07_181	Putative ubiquitin carboxyl-terminal hydrolase
LAN_08_097	Putative MPN (JAB/Mov34) domain-containing protein
LAN_08_132	Putative peptidase C48 family protein
LAN_08_164	Putative SPCS2 family protein
LAN_08_173	Putative peptidase M28 family protein
LAN_08_211	Putative peptidase S9C family protein
LAN_08_324	Putative alpha/Beta hydrolase protein
LAN_08_376	Putative peptidase M1 family protein
LAN_08_444	Putative peptidase S54 family protein
LAN_09_080	Putative ubiquitin carboxyl-terminal hydrolase
LAN_09_082	Putative ubiquitin carboxyl-terminal hydrolase
LAN_09_091	Putative ATP-dependent Clp protease proteolytic subunit
LAN_09_185	Putative pyroglutamyl-peptidase
LAN_09_237	Putative carboxypeptidase
LAN_10_019	Putative peptidase M49 family protein
LAN_10_180	Putative beta-lactamase
LAN_10_181	Putative beta-lactamase
LAN_10_211	Putative trans-sulfuration enzyme
LAN_10_251	Putative nucleoporin-like protein
LAN_10_260	Putative proteasome component region PCI
LAN_10_286	Putative 26S proteasome regulatory subunit protein
LAN_10_343	Putative peptidase M17 family protein
LAN_10_378	Putative separin
LAN_10_399	Putative peptidase T1A family protein
LAN_10_450	Putative AMSH-like protease
LAN_10_465	Putative proteasome alpha subunit
LAN_11_077	Putative peptidase C19 family protein
LAN_11_098	Putative GPCR, family 2, secretin-like protein
LAN_11_147	Putative proteasome alpha subunit
LAN_11_160	Putative gamma-glutamyltransferase family protein
LAN_11_177	Putative WLM domain-containing protein
LAN_11_218	Putative proteasome subunit S5A family protein
LAN_11_221	Putative peptidase M20A family protein

continue to next page

continue from previous page

LAN_11_309	Putative peptidase M24B family protein
LAN_11_387	Putative peptidase M22 family protein
LAN_11_403	Putative proteasome subunit
LAN_11_452	Putative peptidase M24B family protein
LAN_11_509	Putative peptidase C12 family protein
LAN_12_013	Putative peptidase M24 family protein
LAN_12_065	Putative proteasome component PRE2
LAN_12_092	Putative proteasome subunit S1 family protein
LAN_12_101	Putative lon protease homolog, mitochondrial
LAN_12_173	Putative peptidase M16 family protein
LAN_12_182	Putative peptidase C19 family protein
LAN_12_230	Putative ruvB family protein
LAN_12_286	Putative peptidase M24B family protein
LAN_12_337	Putative peptidase M16 family protein
LAN_12_358	Putative peptidase M20A family protein
LAN_12_361	Putative peptidase S33 family protein
LAN_12_368	Putative peptidase M1 family protein
LAN_12_433	Putative mitochondrial inner membrane protease
LAN_12_483	Putative alanine/arginine aminopeptidase
LAN_12_526	Putative CSN1 family protein
LAN_12_561	Putative Nudix hydrolase
LAN_12_566	Putative peptidase C56 family protein
LAN_13_137	Putative peptidase M28 family protein
LAN_13_200	Putative peptidase M48 family protein
LAN_13_201	Putative peptidase M48 family protein (incomplete)
LAN_13_221	Putative peptidase C56 family protein
LAN_13_276	Putative mitochondrial chaperone bcs1
LAN_13_296	Putative AAA ATPase
LAN_13_390	Putative peptidase M24B family protein
LAN_13_450	Putative peptidase M13 family protein
LAN_13_473	Putative peptidase S33 family protein
LAN_13_474	Putative cysteine protease ATG4
LAN_14_099	Putative peptidase C12 family protein
LAN_14_161	Putative peptidase
LAN_14_243	Putative peptidase C1 family protein
LAN_14_324	Putative alpha/Beta hydrolase protein
LAN_15_097	Putative OTU domain-containing protein
LAN_15_114	Putative OTU domain-containing protein
LAN_15_154	Putative ATP-dependent protease
LAN_15_175	Putative peroxisomal biogenesis factor
LAN_15_267	Putative ubiquitin carboxyl-terminal hydrolase
LAN_15_327	Putative mitochondrial intermediate peptidase
LAN_16_011	Putative peptidase M48A family protein
LAN_16_034	Putative proteasome subunit alpha
LAN_16_209	Putative peptidase M28 family protein

continue to next page

continue from previous page

LAN_16_248	Putative peptidase S10 family protein
LAN_16_256	Putative mitochondrial inner membrane protease subunit
LAN_16_314	Putative peptidase A1 family protein
LAN_16_362	Putative ubiquitin carboxyl-terminal hydrolase
LAN_16_385	Putative proteasome subunit beta
LAN_16_479	Putative peptidase C2 family protein
LAN_16_483	Putative aspartyl aminopeptidase
LAN_16_557	Putative proteasome subunit beta
LAN_16_589	Putative peroxisome biosynthesis protein
LAN_16_797	Putative UPF0326 family protein
LAN_16_809	Putative peptidase M16 family protein
LAN_16_822	Putative mitochondrial respiratory chain complexes assembly protein
LAN_16_884	Putative mitochondrial metalloendopeptidase
LAN_17_048	Putative PCI domain-containing protein
LAN_17_077	Putative ubiquitin carboxyl-terminal hydrolase
LAN_17_080	Putative ubiquitin carboxyl-terminal hydrolase
LAN_17_152	Putative dipeptidyl aminopeptidase
LAN_21_006	Putative peptidase C78 family protein

ACKNOWLEDGEMENTS

This research was supported by the Ministry of Science and Technology, Malaysia through the research grant 02-05-20-SF007.

REFERENCES

Abada, E.A. 2019. Application of microbial enzymes in the dairy industry. In *Enzymes in Food Biotechnology*, edited by Kuddus, M. Massachusetts: Academic Press. pp. 61-72.

Adekoya, A.O. & Sylte, I. 2009. The thermolysin family (M4) of enzymes: Therapeutic and biotechnological potential. *Chemical Biology & Drug Design* 73: 7-16.

Alcaíno, J., Cifuentes, V. & Baeza, M. 2015. Physiological adaptations of yeasts living in cold environments and their potential applications. *World Journal of Microbiology and Biotechnology* 31(10): 1467-1473.

Baeza, M., Alcaíno, J., Cifuentes, V., Turchetti, B. & Buzzini, P. 2017. Cold-active enzymes from cold-adapted yeasts. In *Biotechnology of Yeasts and Filamentous Fungi,* edited by Sibirny, A. Springer, Champ. pp. 297-324.

Bharudin, I., Abu Bakar, M.F., Hashim, N.H.F., Mat Isa, M.N., Alias, H., Firdaus-Raih, M., Md Illias, R., Najimudin, N., Mahadi, N.M., Abu Bakar, F.D. & Abdul Murad, A.M. 2018. Unravelling the adaptation strategies employed by *Glaciozyma antarctica* PI12 on Antarctic sea ice. *Marine Environmental Research* 137: 169-176.

Białkowska, A., Gromek, E., Florczak, T., Krysiak, J., Szulczewska, K. & Turkiewicz, M. 2016. Extremophilic proteases: Developments of their special functions, potential resources and biotechnological applications. In *Grand Challenges in Biology and Biotechnology*, edited by Rampelotto, P.H. Switzerland: Springer International Publishing. pp. 399-444.

Burgos, R., Weber, M., Martinez, S., Lluch-Senar, M. & Serrano, L. 2020. Protein quality control and regulated proteolysis in the genome-reduced organism *Mycoplasma pneumoniae*. *Molecular Systems Biology* 16: e9530.

Cheng, J.H., Wang, Y., Zhang, X.Y., Sun, M.L., Zhang, X., Song, X.Y., Zhang, Y.Z., Zhang, Y. & Chen X.L. 2021. Characterization and diversity analysis of the extracellular proteases of thermophilic *Anoxybacillus caldiproteolyticus* 1A02591 from deep-sea hydrothermal vent sediment. *Frontiers in Microbiology* 12: 643508.

Choi, J.M., Han, S.S. & Kim, H.S. 2015. Industrial applications of enzyme biocatalysis: Current status and future aspect. *Biotechnology Advances* 33: 1443-1454.

Choudhuri, S. 2014. Sequence alignment and similarity searching in genomic databases. In *Bioinformatics for Beginners: Genes, Genomes, Molecular Evolution, Databases and Analytical Tools,* edited by Choudhuri, S. Massachusetts: Academic Press. pp. 133-155.

Christensen, L.F., García-Béjar, B., Bang-Berthelsen, C.H. & Hansen, E.B. 2022. Extracellular microbial proteases with specificity for plant proteins in food fermentation. *International Journal of Food Microbiology* 381: 109889.

Ding, F. & Dokholyan, N.V. 2006 Correction: Emergence of protein fold families through rational design. *PLOS Computational Biology* 2(10): e149.

Dube, S., Singh, L. & Alam, S.I. 2001. Proteolytic anaerobic bacteria from lake sediments of Antarctica. *Enzyme and Microbial Technology* 28(1): 114-121.

Feller, G. 2013. Psychrophilic enzymes: From folding to function and biotechnology. *Scientifica* 2013: 512840.

Firdaus-Raih, M., Hashim, N.H.F., Bharudin, I., Bakar, M.F.A., Huang, K.K., Alias, H., Lee, B.K.B., Isa, M.N.M., Mat-Sharani, S., Sulaiman, S., Tay, L.J., Zolkefli, R., Noor, Y.M., Law, D.S.N., Rahman, S.H.A., Md-Illias, R., Bakar, F.D.A., Najimudin, N., Murad, A.M.A. & Mahadi, N.M. 2018. The *Glaciozyma antarctica* genome reveals an array of systems that provide sustained responses towards temperature variations in a persistently cold habitat. *PLoS ONE* 13(1): e0189947.

Furhan, J. 2020. Adaptation, production, and biotechnological potential of cold-adapted proteases from psychrophiles and psychrotrophs: Recent overview. *Journal of Genetic Engineering and Biotechnology* 18(1): 36.

Geisselar, D. & Horwath, R.W. 2008. Regulation of extracellular protease activity in soil in response to different sources and concentrations of nitrogen and carbon. *Soil Biology & Biochemistry* 40: 3040-3048.

Geisseler, D., Horwath R.W., Joergensen, R.G. & Ludwig, B. 2010. Pathways of nitrogen utilization by soil microorganisms- A review. *Soil Biology & Biochemistry* 40: 2058-2067.

Gimenes, N.C., Silveira, E. & Tambourgi, E.B. 2021. An overview of proteases: Production, downstream processes and industrial applications. *Separation and Purification Reviews* 50(3): 223-243.

Gurumallesh, P., Alagu, K., Ramakrishnan, B. & Muthusamy, S. 2019. A systematic reconsideration on proteases. *International Journal of Biological Macromolecules* 128: 254-267.

Horton, P., Park, K.J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C.J. & Nakai, K. 2007. WoLF PSORT: Protein localization predictor. *Nucleic Acids Research* 35: 585-587.

Joseph, B., Kumar, V. & Ramteke, P.W. 2019. Psychrophilic enzymes: Potential biocatalysts for food processing. In *Enzymes in Food Biotechnology*, edited by Kuddus, M. Amsterdam: Elsevier. pp. 817-825.

Kall, L., Krogh, A. & Sonnhammer, E.L. 2007. Advantages of combined transmembrane topology and signal peptide prediction-the Phobius web server. *Nucleic Acids Research* 35: 429-432.

Kamaruddin, S., Redzuan, R.A., Minor, N., Mohd, W., Wan, K., Tab, M., Jaafar, N.R., Rodzli, N.A., Jonet, M.A., Bharudin, I., Yusof, N.A., Quay, D. & Xia, H. 2022. Biochemical characterisation and structure determination of a novel cold-active proline iminopeptidase from the psychrophilic yeast, *Glaciozyma antarctica* PI12. *Catalysts* 12: 722.

Kapp, K., Schrempf, S., Lemberg, M.K. & Dobberstein, B. 2009. Post-targeting functions of signal peptides. In *Protein Transport into the Endoplasmic Reticulum*, edited by Zimmermann, R. Boca Raton: CRC Press. pp. 1-16.

Katoh, K. & Standley, D.M. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution* 30(4): 772-780.

Kuddus, M. 2018. Cold-active enzymes in food biotechnology: An updated mini review. *Journal of Applied Biology & Biotechnology* 6(3): 58-63.

Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. 2018. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution* 35(6): 1547-1549.

Liu, X. & Kokare, C. 2017. Microbial enzymes of use in industry. In *Biotechnology of Microbial Enzymes Production, Biocatalysis and Industrial Applications*, edited by Brahmachari, G. Amsterdam: Elsevier. pp. 267-298.

Martorell, M.M., Ruberto, L.A.M., de Figueroa, L.I.C. & Mac Cormack, W.P. 2019. Antarctic yeasts as a source of enzymes for biotechnological applications. In *Fungi of Antarctica*, edited by Rosa, L. Switzerland: Springer Cham. pp. 285-304.

Matsui, M., Kawamata, A., Kosugi, M. & Imura, S. 2017. Diversity of proteolytic microbes isolated from Antarctic freshwater lakes and characteristics of their cold-active proteases. *Polar Science* 13: 82-90.

Mohamad Nor, N., Hashim, N.H.F., Quay, D.H.X., Mahadi, N.M., Illias, R.M., Abu Bakar, F.D. & Murad, A.M.A. 2020. Functional and structural analyses of an expansin-like protein from the antarctic yeast *Glaciozyma antarctica* PI12 reveal strategies of nutrient scavenging in the sea ice environment. *International Journal of Biological Macromolecules* 144: 231-241.

Naveed, M., Nadeem, F., Mehmood, T., Bilal, M., Anwar, Z. & Amjad, F. 2021. Protease - A versatile and ecofriendly biocatalyst with multi-industrial applications: An updated review. *Catalysis Letters* 151(2): 307-323.

Parvizpour, S., Hussin, N., Shamsir, M.S. & Razmara, J. 2021. Psychrophilic enzymes: Structural adaptation, pharmaceutical and industrial applications. *Applied Microbiology and Biotechnology* 105(3): 899-907.

Pearson, W.R. 2013. An introduction to sequence similarity ("homology") searching. *Current Protocols in Bioinformatics* doi: 10.1002/0471250953.bi0301s42

Petersen, T.N., Brunak, S., Von Heijne, G. & Nielsen, H. 2011. SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nature Methods* 8(10): 785-786.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J. & Glockner, F.O. 2013. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research* 41: 590-596.

Rawlings, N.D., Waller, M., Barrett, A.J. & Bateman, A. 2014. MEROPS: The database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Research* 42: 503-509.

Sabath, N., Ferrada, E., Barve, A. & Wagner, A. 2013. Growth temperature and genome size in bacteria are negatively correlated, suggesting genomic streamlining during thermal adaptation. *Genome Biology and Evolution* 5(5): 966-977.

Santos, A.F., Pires, F., Jesus, H.E., Santos, A.L.S., Peixoto, R., Rosado, A.S., D'avila-Levy, C.M. & Branquinha, M.H. 2015. Detection of proteases from *Sporosarcina aquimarina* and *Algoriphagus antarcticus* isolated from Antarctic soil. *Anais da Academia Brasileira de Ciencias* 87(1): 109-119.

Sarmiento, F., Peralta, R. & Blamey, J.M. 2015. Cold and hot extremozymes: Industrial relevance and current trends. *Frontiers in Bioengineering and Biotechnology* 3: 148.

Schimel, J.P. & Bennett, J. 2004. Nitrogen mineralization: Challenges of a changing paradigm. *Ecology* 85(3): 591-602.

Sommerfield, A.G. & Darwin, A.J. 2022. Bacterial carboxyl-terminal processing proteases play critical roles in the cell envelope and beyond. *Journal of Bacteriology* 204: e00628-21.

Song, P., Zhang, X., Wang, S., Xu, W., Wang, F., Fu, R. & Wei, F. 2023. Microbial proteases and their applications. *Frontier Microbiology* 14: 1236368.

Stecher, G., Tamura, K. & Kumar, S. 2020. Molecular Evolutionary Genetics Analysis (MEGA) for macOS. *Molecular Biology and Evolution* 37(4): 1237-1239.

Tavano, O.L., Berenguer-Murcia, A., Secundo, F. & Fernandez-Lafuente, R. 2018. Biotechnological applications of proteases in food technology. *Comprehensive Reviews in Food Science and Food Safety* 17(2): 412-436.

Ugalde, A.P., Ordóñez, G.R., Quirós, P.M., Puente, X.S. & Lopez-Otín, C. 2010. Metalloproteases and the Degradome. *Methods in Molecular Biology* 622: 3-29.

Vazquez, S.C., Coria, S.H. & Mac, W.P. 2004. Extracellular proteases from eight psychrotolerant antarctic strains. *Microbiological Research* 159: 157-166.

Wu, J.W. & Chen, X.L. 2011. Extracellular metalloproteases from bacteria. *Applied Microbiology and Biotechnology* 92: 253-262.

Wu, Z., Yang, K.K., Liszka, M.J., Lee, A., Batzilla, A., Wernick, D., Weiner, D.P. & Arnold, F.H. 2020. Signal peptides generated by attention-based neural networks. *ACS Synthetic Biology* 9(8): 2154-2161.

Yang, Z., Huang, Z., Wu, Q., Tang, X. & Huang, Z. 2023. Cold-adapted proteases: An efficient and energy-saving biocatalyst. *International Journal of Molecular Sciences* 24: 8532.

Yin, Q., He, K., Collins, G., De Vrieze, J. & Wu, G. 2024. Microbial strategies driving low concentration substrate degradation for sustainable remediation solutions. *npj Clean Water* 7(1): 52.

Yusof, N.A., Hashim, N.H.F. & Bharudin, I. 2021. Cold adaptation strategies and the potential of psychrophilic enzymes from the antarctic yeast, *Glaciozyma antarctica* PI12. *Journal of Fungi* 7(7): 528.

Zhou, M-Y., Wang, G-L., Li, D., Zhao, D-L., Qin, Q-L., Chen, X-L., Chen, B., Zhou, B-C., Zhang, X.Y. & Zhang, Y.Z. 2013. Diversity of both the cultivable protease-producing bacteria and bacterial extracellular proteases in the Coastal Sediments of King George Island. *Antarctica* 8(11): e79668.

*Corresponding author; email: shazilah@ukm.edu.my