# In Silico Identification and Classification of *Glaciozyma antarctica* Small Secreted Protein in Adaptation to Cold

**Nur Nisrina Che Zainudin[1], Shuhaila Mat-Sharani[2*], Norfarhan Mohd-Assaad[1*], Nur Izyan Shahanim Yusoff[1], Farrah Affifi Mohamad[1] and Babalola Abdulhafeez[2]**

[1] Department of Applied Physics, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM-Bangi, Selangor
[2] Biomedicine Programme, School of Health Sciences, Universiti Sains Malaysia,
16150 Kubang Kerian, Kelantan, Malaysia

***Abstract-*** Small Open Reading Frames (sORFs) are short sequence that were previously overlooked within genomes but have gained increasing recognition for their diverse functional roles. This study investigates the landscape of sORF proteins within the secretomes of extremophilic microorganisms, focusing on cold adaptation. Through computational analysis of 17 genomes, encompassing psychrophilic, mesophilic, and thermophilic organisms, 3,058 sORFs were identified, of which 173 were classified as secretome proteins. Psychrophilic organisms exhibited the highest proportion of secretome proteins. Functional annotation revealed diverse molecular functions, biological processes, and cellular components associated with cold adaptation, including mitochondrial assembly and transmembrane transport. However, explicit annotations related to stress response pathways or cold adaptation were lacking, presenting opportunities for future investigations. This research enhances understanding of molecular mechanisms underlying cold adaptation and underscores the importance of sORF proteins in extremophilic organisms. Future works may involve exploring stress response pathways, structural bioinformatics, and machine learning algorithms to advance knowledge of microbial adaptation to extreme conditions.

***Index Terms-*** extremophiles, mesophilic, psychrophilic, secretome, sORFs prediction, thermophilic

## I. Introduction

Antarctica the Earth's southernmost continent, is one of the most extreme places, environments, characterized by powerful winds, extremely low air humidity, significant levels of solar radiation-particularly ultraviolet (UV) radiation-limited precipitation, and freezing temperatures [1]. Organisms inhabiting this region have evolved remarkable adaptations strategies to survive and thrive these extreme conditions. *Glaciozyma antarctica* is a psychrophilic yeast isolated from Antarctica that can survive at cold temperatures below 0 °C. Psychrophilic microorganisms have adapted physiologically to the severe cold of the arctic, deep sea, and alpine areas through such as antifreeze proteins synthesis, enzyme kinetics regulation, and membrane fluidity maintenance [2].

The genome sequences of numerous psychrophilic microorganisms have been published, including the archaeon *Methanococcoides burtonii* [3], and bacteria such as *Colwellia psychrerythraea* 34H [4], *Psychromonas ingrahamii* [5], *Pseudoalteromonas haloplanktis* TAC125 [6] and *Psychrobacter arcticus* 273-4 [7]. Most psychrophilic microorganisms shared similar adaptation mechanisms that enable survival at low temperature environments. These mechanisms include the synthesis of unsaturated fatty acids to enhance membrane flexibility, the production of multiple transporter proteins, and the expression of chaperones and cold-shock proteins [8].

One of the critical aspects of an organism's adaptation to extreme environmental conditions is its secretome, which comprises the entire set of proteins secreted by the organism into its extracellular space. Secreted These proteins play important roles in various cellular processes, including nutrient acquisition, defense against environmental stress, and interactions with the surrounding environment [9]. In the case of *G. antarctica*, the secretome is likely highly specialized to function at low temperatures and withstand the unique environmental stresses of Antarctic ecosystem.

The identification and classification of the *G. antarctica* secretome using in silico methods can provide a comprehensive overview of the proteins involved in cold adaptation. In silico analysis utilizes of computational techniques to predict and analyze protein sequences, structures, and functions. These methods aid in identifying potential secreted proteins and provide insights into their functions and potential roles in cold adaptation. Studying the genome of a psychrophilic eukaryotic microbe can also reveal cold adaptation strategies that are independent of the significant insulative barrier mechanisms available to vertebrate eukaryotes [2].

## II. Methods

Protein sequences from 17 extremophilic microorganisms, including *Glaciozyma antarctica*, were obtained from the GlacIER portal [10] and NCBI RefSeq [11]. These include 9 psychrophiles, 5 mesophiles, and 3 thermophiles spanning fungi, bacteria, and archaea.

sORFs were identified using a Python script developed in-house, applying a cut-off of fewer than 80 amino acids, following thresholds proposed by [12]. Validation was performed by comparing the predicted sORFs against total number of known proteins.

To predict secretory sORFs, signal peptides were identified using SignalP v5.0 [13] and Phobius [14]. Transmembrane domains were filtered using DeepTMHMM [15], and mitochondrial-targeting sequences were excluded using TargetP v2 [13]. Further filtering was done with ScanProsite [16] for ER retention signals and NetGPI [17] for GPI-anchored proteins. The remaining sequences were considered refined secretome candidates.

Functional annotation of the predicted secretory sORFs was carried out using Blast2GO [18], incorporating BLASTP against the NCBI nr database, InterPro domain analysis, and Gene Ontology (GO) mapping to determine biological processes, molecular functions, and cellular components.

## III. Discussion

The data retrieval process resulted in the acquisition of 17 genomes, comprising 9 psychrophilic, 5 mesophilic and 3 thermophilic organisms. These genomes were taxonomically classified into 9 bacterial, 6 fungal, and 2 archaeal genomes. The selection criteria, as detailed in the methodology, were based on two primary factors: i) relevance to temperature adaptation, where their scientific name is mentioned in the paper related to temperature adaptation. ii) the availability of corresponding proteomic data in the NCBI database. This targeted selection enhances the robustness of the analysis, enabling a more comprehensive investigation of temperature adaptation mechanisms among different organisms.

To provide additional context for the selected genomes in this study, their optimal growth temperatures (OGT) were examined. OGT is a key determinant of an organism's environmental adaptation, influencing its biological processes and metabolic activities. Psychrophilic organisms characterized by their OGT typically below 20°C, are adapted to cold environments while mesophilic organisms exhibit OGTs within moderate range of 20°C to 40°C. Thermophilic organisms display OGTs above 40°C [19]. The complete list of selected genomes is provided in Supplementary Table S1. This approach establishes transparency in the data collection process and ensures that the chosen genomes are well-suited for addressing the research objectives related to temperature adaptation.

Following the collection of proteomic data, a script was executed to identify Small Open Reading Frames (sORFs) within the dataset, applying a cut-off value of 80 amino acids. This threshold was chosen based on [12], who reported that short coding sequences have a median length of 79 codons and are preferentially found in functionally monocistronic transcript. These transcripts exhibit mRNA characteristics, albeit shorter and structurally simpler than canonical protein-coding mRNAs. Open reading frames (ORFs) are stretches of DNA bounded by a stop codon, constituting a crucial genomic characteristic utilized for the identification of protein-coding genes. Traditionally, open reading frames (ORFs) with fewer than 100 codons, referred to as sORFs, were excluded from gene annotations [20]. However, recent advances in next-generation sequencing and proteomics have led to the dentification of numerous sORFs-encoded proteins, also known as micropeptides. These proteins are sometimes referred as polycistronic peptides when translated from a polycistronic mRNA or as short open reading frame (sORF)-encoded polypeptides (SEPs) [21]. While the molecular structure and functional mechanisms of many recently identified sORFs remain poorly understood, well-characterized small proteins provide valuable insights into the diverse molecular roles of sORFs in system biology [22].

The results presented in Supplementary Table S2 provide a comprehensive overview of sORFs across various genome. As shown in the table above, the proportion of sORFs compared to the total number of protein-coding genes is relatively small, ranging from approximately 1 to 11%. Among the total of 78,684 proteins analyzed across all organisms included in this study, only 3,058 were classified as sORFs. Even for Pseudomonas destructans, which has the highest number of proteins (9,405), only 162 (1.72%) were identified as sORFs.

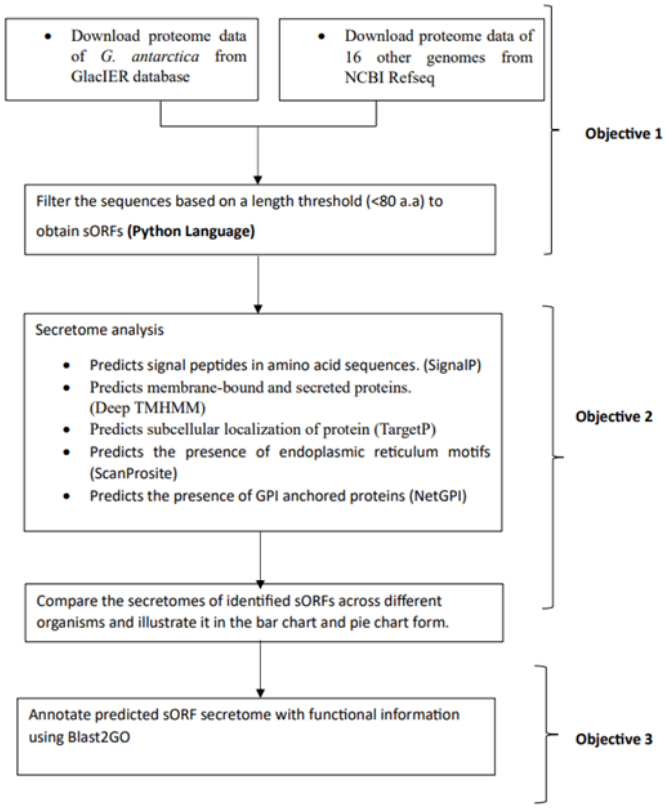


Figure 1 Overall workflow

According to published literature, at least 299 genes in Saccharomyces cerevisiae is likely to encode sORFs amino acids sequences of fewer than 100 residues. The results, as depicted in Figure 1, demonstrate considerable variation in the percentage of sORFs across different genomes, ranging from as low as 1.09% in Cryptococcus neoformans to as high as 11.76% in *Geobacillus stearothermophilus*. Additionally, this finding reveals notable discrepancies in the average length of sORFs among different genomes, highlighting substantial variability in the size of these sORFs. For instance, *C. neoformans* exhibits a relatively low number of sORFs, whereas *Methanococcoides burtonii* showcases significantly higher proportion. This variation underscores the diverse genetic architectures across species and suggests potential differences in the functional roles or evolutionary pressures shaping the type and length distribution of sORFs within genomes.

The validation of the result sORFs is validated by developing a script to compare the total number of sORFs to the total number of proteins encoded by each genome. This comparison enables an assessment of the relative abundance of sORFs within the genomic context and identification of any potential discrepancies. Such validation enhances confidence in the biological relevance of the identified sORFs.

The secretome was predicted in a similar manner to the guidelines of the previously described pipeline in 'Secretome Analysis for a New Strain of the Blackleg Fungus Plenodomus lingam Reveals Candidate Proteins for Effectors and Virulence Factors' [23-24] in the article titled 'In Silico Characterization of the Secretome of the Fungal Pathogen Thielaviopsis punctulate, the Causal Agent of Date Palm Black Scorch Disease'.

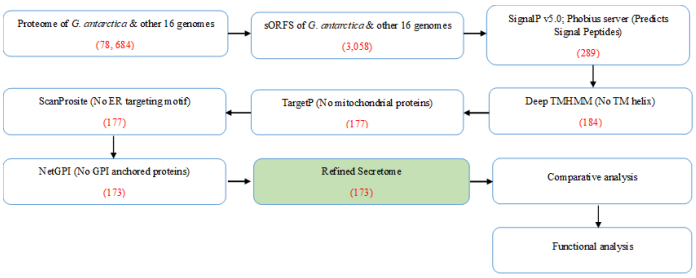


Figure 2 Result of sORF secretome prediction for all 17 genomes

The methodology used to predict the secretome of 17 genomes are illustrated in Figure 2 with the number of sORFs in each step. Using a combination of SignalP and Phobius server, of the 3058 total sORFs, 289 sORFs were predicted to have a signal peptide at their N-terminal region. Among these 289 sORFs, 105 transmembrane proteins were excluded using DeepTMHMM and 7 mitochondrial-target proteins identified by TargetP were discarded. Then, the remaining 184 sORFs were scanned for an endoplasmic reticulum (ER)-targeting signal to exclude the proteins that remain in the endoplasmic reticulum by the ScanProsite webserver. Three sORFs were then identified by NetGPI to harbor glycophosphatidylinositol (GPI) motifs, resulting finally in a list of 173 refined secreted sORFs. Thus, the bioinformatic pipeline predicted a total of 173 (5.65%) of the entire sORFs of 17 genomes as secretome and these were used for further analysis. Overall, the number of refined secretome in all 17 genomes ranged from 1 to 22. The result in Supplementary Table S2 shown the number of refined secretome for each genome and their percentage from the total of sORFs and proteins. Interestingly, only 1 sORFs secretome predicted for *H. lacusprofundi* from the total of 265 sORFs using the pipeline in Figure 2.

The comparative analysis also was conducted between the temperature groups which are psychrophiles, mesophiles and thermophiles. This was shown in Figure 3 in the form of horizontal bar chart. From the total of 173 sORFs secretome, 112 (65%) of them were psychrophiles, 37 (21%) were mesophiles and 24 (14%) were thermophiles. Further exploration of these temperature-specific sORFs may elucidate their functional roles and shed light on the adaptive strategies employed by microorganisms inhabiting diverse environmental niches.

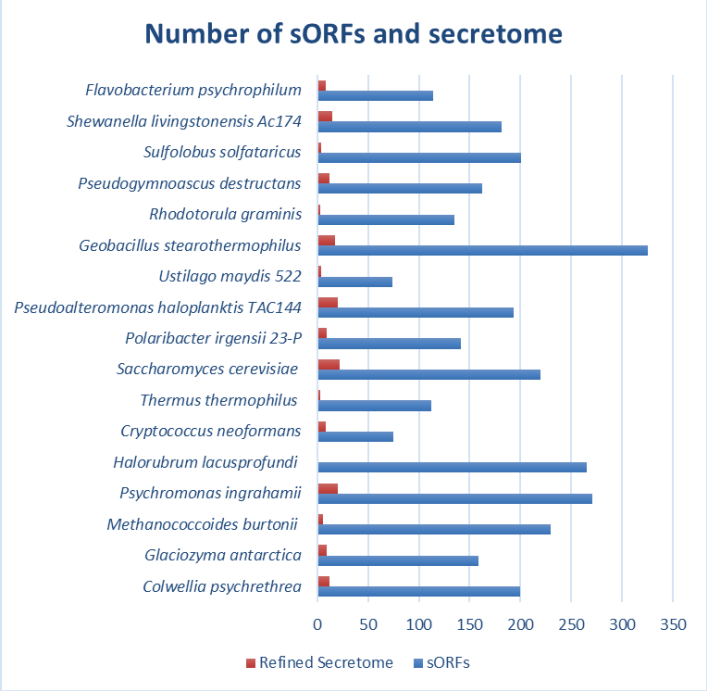


Figure 3 Number of sORFs and secretome for across all genomes

The functional analysis of predicted secretomes represents a critical aspect of understanding the molecular mechanisms underlying cellular communication, intercellular signaling, and more. In this project, bioinformatics tools and techniques were utilized to predict and characterize the sORFs secretomes of organisms, aiming to uncover the functional roles and biological significance of secreted proteins.

Firstly, all the 173 secretome of sORFs predicted using the secretome pipeline were loaded in the Blast2GO interface. Then, all of them is going through sequence similarity search using

BLASTP. There are 12 sequences that do not have any match of sequence similarity in the known database. Then, InterPro is run using default parameters to find any conserved domains and protein signature. This integration allows users to identify functional domains and motifs associated with their sequences, enhancing the understanding of protein structure and function. By leveraging InterPro annotations, Blast2GO users can explore the functional diversity and evolutionary relationships of proteins within their datasets, facilitating comprehensive functional annotation and analysis workflows.

Next, after running InterPro analysis in Blast2GO, one crucial step is mapping the InterPro annotations to Gene Ontology (GO) terms. This process involves associating the protein domains, families, and motifs identified by InterPro with standardized terms from the GO database, which categorizes gene products based on their biological processes, molecular functions, and cellular components. The mapping of InterPro annotations to GO terms is automated, leveraging the rich annotation resources available in both databases. Blast2GO utilizes algorithms to match InterPro entries to corresponding GO terms based on predefined mappings and associations.

The total of 173 sORFs secretome were blast searched against the nr database and classified according to the three major Gene Ontology (GO) classes of molecular function, biological process and cellular component. Out of 173 sORFs secretome, only 59 could be classified according to GO classes. This resulted in 48 sORFs being classified into biological process, 68 sORFs classified to be involved in molecular function and 75 sORFs being classified into cellular components (Figure 4). The number of genes resulting from the Gene Ontology classification are higher than the total number of predicted sORFs because one gene can be associated with multiple classes.

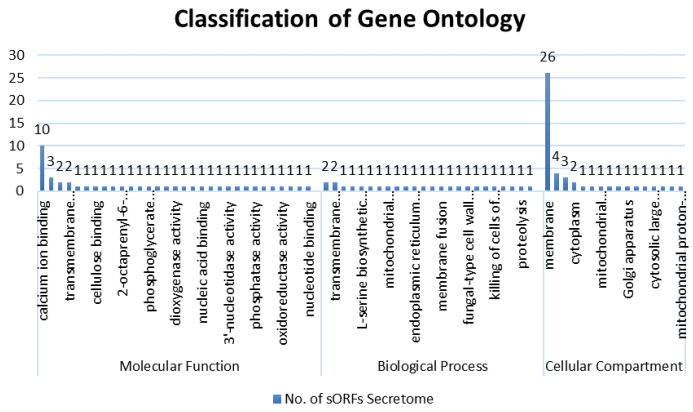


Figure 4 The distribution of functional analysis of 173 sORFs secretomes

In the cellular component classification, there were 75 predicted sORFs classified into mitochondrion (1), membrane (27), cytoplasm (2), cellular component (13), proteasome complex (1), cytochrome complex assembly (1), plasma membrane (4), oxidoreductase activity (1), regulation of DNA-templated transcription (1), DNA binding (3), DNA-binding transcription factor activity (1), cytosol (1), endoplasmic reticulum membrane (1), extracellular region (4), mitochondrial intermembrane space (1), cytosolic large ribosomal subunit (1), fungal-type vacuole (5), cell periphery (1) side of membrane (1), fungal-type cell wall (1), mitochondrial proton-transporting ATP synthase complex, coupling factor F(o) (1), endoplasmic reticulum (1), Golgi apparatus (1), and cellular anatomical entity (1). Under the molecular function classification the predicted sORFs were assigned to function associated to 2-octaprenyl-6-methoxyphenol hydroxylase activity (1), oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, NAD(P)H as one donor, and incorporation of one atom of oxygen (1), FAD binding (1), transposase activity (1), transmembrane transporter activity (4), hydrolase activity (1), hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds (1), NADPH-hemoprotein reductase activity (1), FMN binding (1), flavin adenine dinucleotide binding (1), NADP binding (1), calcium ion binding (10), phosphatase activity (1), phosphoglycerate dehydrogenase activity (1), cellulose binding (1), NAD binding (1), copper ion binding (1), copper chaperone activity (1), nucleic acid binding (1), peptidase activity (1), structural constituent of ribosome (1), metal ion binding (2), structural constituent of cell wall (1), SNAP receptor activity (1), long-chain fatty acid-CoA ligase activity(1), nucleotide binding, 3'-nucleotidase activity (1), 2',3'-cyclic-nucleotide 2'-phosphodiesterase activity (1), enoyl-[acyl-carrier-protein] reductase (NADH) activity (1), nitronate monooxygenase activity (1), and dioxygenase activity (1). For biological process GO classification, they were classified into mitochondrial respiratory chain complex I assembly (1), ubiquinone biosynthetic process (1), DNA transposition (1), transmembrane transport (4), response to toxic substance (1), carbohydrate metabolic process (2), L-serine biosynthetic process (1), mitochondrial cytochrome c oxidase assembly (1), proteolysis (1), translation (1), fungal-type cell wall organization (1), proton motive force-driven ATP synthesis (1), protein secretion (1), intracellular transport (1), endoplasmic reticulum to Golgi vesicle-mediated transport (1), killing of cells of another organism (1), defense response to bacterium (1), long-chain fatty acid metabolic process (1), and nucleotide catabolic process (1).

The cellular component, molecular function, and biological process classifications of predicted sORFs provide valuable insights into the potential roles of these peptides in stress response and cold adaptation across psychrophilic, mesophilic, and thermophilic organisms. Notably, the diverse localization of sORFs within cellular compartments such as the membrane, cytoplasm, and organelles like mitochondria and the endoplasmic reticulum suggests their involvement in temperature-dependent stress responses and adaptation mechanisms. Molecular function assignments, including oxidoreductase activity and transmembrane transporter activity, underscore the importance of redox regulation and membrane transport in coping with temperature fluctuations. Biological processes associated with sORFs, such as response to toxic substances and defense

responses to bacteria, suggest potential strategies employed by organisms to withstand environmental stresses at different temperature ranges. Additionally, the presence of sORFs associated with protein secretion and intracellular transport hints at the dynamic interplay between cellular compartments in orchestrating stress adaptation processes. These findings offer a comprehensive framework for exploring the molecular mechanisms underlying temperature-dependent stress responses and adaptation strategies in psychrophilic, mesophilic, and thermophilic organisms.

## IV. Conclusion

In conclusion, this study has provided valuable insights into the landscape of small Open Reading Frame (sORF) proteins within the secretomes of various extremophilic microorganisms, especially towards cold adaptation. Through computational analysis, from a total of 78, 684 proteins, a total of 3,058 sORFs were identified across 17 genomes, with notable variation in their abundance and length distribution. There is a total of 173 sORFs secretome predicted. The comparative analysis of sORF secretomes across different temperature groups revealed intriguing patterns, with psychrophilic organisms exhibiting the highest proportion of refined secretome proteins. Furthermore, functional annotation of the predicted secretome proteins offers a glimpse into the potential roles of these proteins in cold adaptation and other cellular processes. The sORFs secretome analysis revealed a diverse array of molecular functions, biological process and cellular components associated with cold adaptation, including mitochondrial assembly, transmembrane transport, and enzymatic activities. However, despite their presence, the explicit annotations related to stress response pathways or cold adaptation were apparent. Nevertheless, this presents an opportunity for future investigations. Overall, this research contributes to our understanding of the molecular mechanisms underlying cold adaptation and highlights the importance of sORF proteins in extremophilic organisms.

Further analysis could involve exploring stress response pathways through pathway enrichment analysis or comparative proteomics with known stress-responsive proteins. Structural bioinformatics approaches could elucidate the structural features and potential interactions of these proteins, while machine learning algorithms could aid in predicting sequence motifs associated with cold adaptation. These efforts will advance our understanding of microbial adaptation to extreme conditions.

## Acknowledgment

This research was supported by the Ministry of Higher Education Malaysia (FRGS grant number FRGS/1/2022/ICT01/USM/03/2). The research was performed at the School of Health Sciences, Universiti Sains Malaysia (USM).

## References

[1] N. A. Yusof, N. H. F. Hashim, and I. Bharudin, "Cold adaptation strategies and the potential of psychrophilic enzymes from the Antarctic yeast, *Glaciozyma antarctica* PI12," *J. Fungi*, vol. 7, no. 7, p. 528, 2021, doi: 10.3390/jof7070528.

[2] M. Firdaus-Raih et al., "The *Glaciozyma antarctica* genome reveals an array of systems that provide sustained responses towards temperature variations in a persistently cold habitat," PLoS ONE, vol. 13, no. 1, e0189947, 2018. [Online]. Available: https://doi.org/10.1371/journal.pone.0189947.

[3] M. A. Allen *et al.*, "The genome sequence of the psychrophilic archaeon, *Methanococcoides burtonii*: the role of genome evolution in cold adaptation," *ISME J.*, vol. 3, no. 9, pp. 1012–1035, 2009. [Online]. Available: https://doi.org/10.1038/ismej.2009.45.

[4] B. A. Methé *et al.*, "The psychrophilic lifestyle as revealed by the genome sequence of *Colwellia psychrerythraea* 34H through genomic and proteomic analyses," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 31, pp. 10913–10918, 2005. [Online]. Available: https://doi.org/10.1073/pnas.0504766102.

[5] M. Riley *et al.*, "Genomics of an extreme psychrophile, *Psychromonas ingrahamii*," *BMC Genomics*, vol. 9, p. 210, 2008. [Online]. Available: https://doi.org/10.1186/1471-2164-9-210.

[6] C. Médigue *et al.*, "Coping with cold: the genome of the versatile marine Antarctica bacterium *Pseudoalteromonas haloplanktis* TAC125," *Genome Res.*, vol. 15, no. 10, pp. 1325–1335, 2005. [Online]. Available: https://doi.org/10.1101/gr.4126905.

[7] H. L. Ayala-del-Río *et al.*, "The genome sequence of *Psychrobacter arcticus* 273-4, a psychroactive Siberian permafrost bacterium, reveals mechanisms for adaptation to low temperature growth," *Appl. Environ. Microbiol.*, vol. 76, no. 7, pp. 2304–2312, 2010. [Online]. Available: https://doi.org/10.1128/AEM.02101-09.

[8] I. Bharudin *et al.*, "Unravelling the adaptation strategies employed by *Glaciozyma antarctica* PI12 on Antarctic sea ice," *Mar. Environ. Res.*, vol. 137, pp. 169–176, 2018. [Online]. Available: https://doi.org/10.1016/j.marenvres.2018.03.007.

[9] S. Ranganathan and G. Garg, "Secretome: clues into pathogen infection and clinical applications," *Genome Med.*, vol. 1, no. 11, p. 113, 2009. [Online]. Available: https://doi.org/10.1186/gm113.

[10] GlacIER, "GlacIER: *Glaciozyma antarctica* Integrated Exploration Resource," 2025. [Online]. Available: https://nibm.my/glacier/. [Accessed: Mar. 9, 2025].

[11] N. A. O'Leary *et al.*, "Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation," *Nucleic Acids Res.*, vol. 44, no. D1, pp. D733–D745, 2016. [Online]. Available: https://doi.org/10.1093/nar/gkv1189.

[12] J. P. Couso and P. Patraquim, "Classification and function of small open reading frames," *Nat. Rev. Mol. Cell Biol.*, vol. 18, pp. 575–589, 2017. [Online]. Available: https://doi.org/10.1038/nrm.2017.58.

[13] J. J. Almagro Armenteros *et al.*, "SignalP 5.0 improves signal peptide predictions using deep neural networks," *Nat. Biotechnol.*, vol. 37, no. 4, pp. 420–423, 2019. [Online]. Available: https://doi.org/10.1038/s41587-019-0036-z.

[14] L. Käll, A. Krogh, and E. L. Sonnhammer, "Advantages of combined transmembrane topology and signal peptide prediction—The Phobius web server," *Nucleic Acids Res.*, vol. 35, Web Server issue, pp. W429–W432, 2007. [Online]. Available: https://doi.org/10.1093/nar/gkm256.

[15] J. Hallgren *et al.*, "DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks," *Cold Spring Harb. Lab.*, 2022. [Online]. Available: https://doi.org/10.1101/2022.04.08.487609.

[16] E. de Castro *et al.*, "ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins," *Nucleic Acids Res.*, vol. 34, Web Server issue, pp. W362–W365, 2006. [Online]. Available: https://doi.org/10.1093/nar/gkl124.

[17] M. H. Gíslason, H. Nielsen, J. J. Almagro Armenteros, and A. R. Johansen, "Prediction of GPI-anchored proteins with pointer neural networks," *Curr. Res. Biotechnol.*, vol. 3, pp. 6–13, 2021. [Online]. Available: https://doi.org/10.1016/j.crbiot.2021.01.001.

[18] A. Conesa and S. Götz, "Blast2GO: a comprehensive suite for functional analysis in plant genomics," *Int. J. Plant Genomics*, vol. 2008, p. 619832, 2008. [Online]. Available: https://doi.org/10.1155/2008/619832.

[19] V. A. Lyubetsky, O. A. Zverkov, L. I. Rubanov, and A. V. Seliverstov, "Optimal growth temperature and intergenic distances in bacteria, archaea, and plastids of rhodophytic branch," BioMed Res. Int., vol. 2020, p. 3465380, 2020. [Online]. Available: https://doi.org/10.1155/2020/3465380.

[20] N. Jain, F. Richter, I. Adzhubei, A. J. Sharp, and B. D. Gelb, "Small open reading frames: a comparative genetics approach to validation," *BMC Genomics*, vol. 24, no. 1, p. 226, 2023. [Online]. Available: https://doi.org/10.1186/s12864-023-09311-7.

[21] J. Crappé, W. Van Criekinge, and G. Menschaert, "Little things make big things happen: a summary of micropeptide encoding genes," *EuPA Open Proteomics*, vol. 3, pp. 128–137, 2014. [Online]. Available: https://doi.org/10.1016/j.euprot.2014.02.006.

[22] D. Schlesinger and S. J. Elsässer, "Revisiting sORFs: overcoming challenges to identify and characterize functional microproteins," *FEBS J.*, vol. 289, no. 1, pp. 53–74, 2022. [Online]. Available: https://doi.org/10.1111/febs.15769.

[23] N. A. Bouqellah, N. A. Elkady, and P. F. Farag, "Secretome analysis for a new strain of the blackleg fungus Plenodomus lingam reveals candidate proteins for effectors and virulence factors," J. Fungi, vol. 9, no. 7, p. 740, 2023. [Online]. Available: https://doi.org/10.3390/jof9070740.

[24] B. V. Chellappan et al., "In silico characterization of the secretome of the fungal pathogen Thielaviopsis punctulata, the causal agent of date palm black scorch disease," J. Fungi, vol. 9, no. 3, p. 303, 2023. [Online]. Available: https://doi.org/10.3390/jof9030303.

AUTHORS

**First Author** – Nur Nisrina Che Zainudin, BSc (Hons) Bioinformatics, Universiti Kebangsaan Malaysia
**Second Author** – Shuhaila Mat-Sharani, PhD, Universiti Sains Malaysia,
**Third Author** – Norfarhan Mohd-Assaad, PhD, Universiti Kebangsaan Malaysia
**Fourth Author** – Nur Izyan Shahanim Yusoff, BSc (Hons) Bioinformatics, Universiti Kebangsaan Malaysia
**Fifth Author** – Farrah Affifi Mohamad, BSc (Hons) Bioinformatics, Universiti Kebangsaan Malaysia,.
**Sixth Author** – Babalola Abdulhafeez, MSc, Universiti Sains Malaysia,

**Correspondence Author** – Shuhaila Mat-Sharani, and Norfarhan Mohd-Assaad