Malays. Appl. Biol. (2015) 44(1): 167–172

GENOME MINING FOR GLYCOSIDE HYDROLASES FROM THE PSYCHROPHILIC YEAST *Glaciozyma antarctica* PI12

NOORAISYAH, M.N.[1], SITI NUR HASANAH, M.Y.[1], MAHADI, N.M.[2], ABU BAKAR, F.D.[1] and MURAD, A.M.A.[1*]

[1]*School of Biosciences and Biotechnology, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor Malaysia*
[2]*Malaysia Genome Institute, Jalan Bangi, 43000 Kajang, Selangor, Malaysia*
[*]*Email: munir@ukm.edu.my*

ABSTRACT

Glycoside hydrolases are enzymes that hydrolyse glycosidic bonds in carbohydrate chains to produce simple molecules. In fungi, glycoside hydrolases are important enzymes that hydrolyse complex carbohydrates into simple sugars that can subsequently be consumed for energy metabolism. *Glaciozyma antarctica* is a psychrophilic yeast isolated from sea ice in Antarctica. The *G. antarctica* genome has been completely sequenced, and a total of 7,857 genes have been predicted. The objective of the present study was to determine different classes of glycoside hydrolases from the *G. antarctica* genome and predict the localisation of these enzymes. Using genome mining, a total of 97 *G. antarctica* genes were predicted to encode glycoside hydrolases. The majority of the enzymes, including endoglucanases, xylanases, and chitinases, were identified from GH family 5 (12 genes), followed by GH family 45 (11 genes), GH family 10 (11 genes) and GH family 18 (9 genes). The secreted glycoside hydrolase enzymes were primarily endoglucanases from GH family 45, and these enzymes degrade celluloses in the cell walls of plants and algae. Extracellular glycoside hydrolases have been implicated as important in nutrient scavenging and organic decomposition in Antarctic sea ice.

Key words: Glycoside hydrolases, psychrophiles, *Glaciozyma antarctica*, genome mining

INTRODUCTION

Glaciozyma antarctica PI12 is a psychrophilic yeast isolated from Antarctic sea ice (Hashim *et al*., 2013). Psychrophiles have evolved a complex range of adaptation strategies to survive in cold environments. Several mechanisms have been suggested for cold adaptation, such as changes in membrane fluidity, the secretion of exopolysaccharides and antifreeze proteins and the production of enzymes that optimally function at subfreezing temperatures (Kawahara, 2002). The Malaysian Genome Institute has fully sequenced the *G. antarctica* genome, comprising 7,857 predicted genes (http://www.genomemalaysia.gov.my/glaciozyma_antarctica/). Glycoside hydrolases are enzymes that hydrolyse or cleave the glycosidic bonds between carbohydrate-carbohydrate or carbohydrate-non carbohydrate moieties, such as proteins and lipids. Based on the carbohydrate active enzymes (CAZy) database (http://www.cazy.org/), 133 glycoside hydrolase families have been identified to date. The families are grouped based on amino acid sequence similarities. In fungi, glycoside hydrolases are required to facilitate infection, gain nutrition and degrade organic matter in the surrounding environment (Zhao *et al*., 2013). Based on a comparative analysis of the glycoside hydrolases in fungi, these enzymes are extracellularly secreted. The characterisation of these secreted proteins would facilitate the prediction of the function of this enzyme in terms of the hydrolysis of polysaccharide materials. A detailed study of these enzymes will increase the current understanding of how glycoside hydrolases facilitate *G. antarctica* survival in the extreme Antarctic sea ice habitat and provide information on the nutrition patterns of this yeast. In addition, glycoside hydrolases are a group of enzymes with industrial importance, and cold active glycoside hydrolases have great potential in food, pharmaceutical and detergent industries.

* To whom correspondence should be addressed.

MATERIALS AND METHODS

All known glycoside hydrolase protein sequences were downloaded from NCBI non-redundant database and the UniprotKB/Swissprot and UniprotKB/TrEMBL databases (http://www.uniprot.org/). These protein sequences were blasted against the *G. antarctica* genome database available at the Malaysian Genome Institute (http://27.126.156.144/glaciozyma_antarctica/). All matching sequences based on a set parameter were filtered into a list of potential glycoside hydrolase enzymes for *G. antarctica*. The set parameters were 1) sequence identity equal or higher than 30% and 2) an e-value equal or lesser than $1x10^{-5}$. Once the list of potential enzymes was obtained, a functional analysis of the sequences was performed to identify potential glycoside hydrolase domains using the InterProScan software available at EMBL/EBI (http://www.ebi.ac.uk/Tools/pfa/iprscan/) (Quevillon *et al*., 2005). Alternatively, the putative function for each sequence was also confirmed through a blast analysis against the Protein Data Bank (PDB database) (http://www.rcsb.org/pdb/home/home.do) available at NCBI Blast (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The enzymes were classified into different glycoside hydrolase families according to the CAZY database (http://www.cazy.org) (Lombard *et al*., 2014). The signal peptide prediction was performed using three software programs: SignalP 4.1 (http://www.cbs.dtu.dk/services/SignalP/) (Petersen *et al*., 2011), Signal 3L (http://www.csbio.sjtu.edu.cn/bioinf/Signal-3L/) and Phobius (http://phobius.sbc.su.se/).

RESULTS

Among the 7857 genes in the *G. antarctica* genome, a total of 97 genes (1.24%) were predicted to encode glycoside hydrolases (Table 1). The prediction was based on the presence of a glycoside hydrolase domain or a glycoside hydrolase binding subgroup in the sequence. Subsequently, these proteins were grouped into a respective glycoside hydrolase family, according to a CAZy database (Fig. 1). The number of hydrolase-encoding genes seems low in this fungus, as a comparative analysis of fungi genomes revealed a large number of CAZy enzymes, ranging from 171-285 genes per genome (Damasio *et al*., 2014). For instance, *Trichoderma reesei* contains 200 glycoside hydrolase enzymes, while *Aspergilus niger* contains 250 glycoside hydrolase enzymes. The low number of glycoside hydrolases in *G. antarctica* might suggest that in psychrophiles, these enzymes might contain different coding sequences for glycoside hydrolase domains. A recent comparative analysis of glycoside hydrolases in fungi revealed the distribution of common GH families in fungi. The most prevalent families were GH5, GH13, GH31, and GH61 (Zhao *et al*., 2013). However, in the *G. antarctica* genome, the distribution of the most prevalent GH family is GH5 (12 genes), followed by GH45 (11 genes), GH10 (11 genes) and GH18 (9 genes). These families belong to mannanases, endoglucanases, cellulases, and chitinases, respectively. Notably, two genes fall into unclassified families due to non-similarity with any sequence in the database, although these sequences contain glycoside hydrolase domains. These two domains are named GH-type carbohydrate-binding subgroup domain (LAN_11_346) and galactose binding domain (LAN_14_284).

We also screened the genes listed in Table 1 for the presence of signal peptides in each of the sequences. A total of 46% of the glycoside hydrolases in the list had a signal peptide, suggesting that these enzymes could be secreted (Fig. 2). The majority of the secreted enzymes were endoglucanases, followed by xylanases (Table 2). Other secreted enzymes identified were mannanases, α-amylases, lichenases, chitinases and mannosidases.

DISCUSSION

The extreme cold environment in which *G. antarctica* inhabits requires special modifications, including scavenging for scarce nutrients in the environment, for the growth and survival of these microorganisms. Glycoside hydrolase is an enzyme that degrades carbohydrate linkages to generate carbon sources. The release of various extracellular enzymes in the cold environment might reflect the metabolic adaptations of psychrophilic microorganisms for significant functions, such as the decomposition of organic materials and nutrient cycling. This idea is supported by the presence of organic carbon and nitrogen sources originating from melting glacier ice (Brizzio *et al*., 2007), the abundance of organic matter from the death and lysis of the sea-ice organisms, and the secretion of organic polymers by algae and bacteria (Yu *et al*., 2009), suggesting that psychrophilic yeasts, such as *G. antarctica* induce the secretion of extracellular enzymes into the environment. The majority of glycoside hydrolases present in *G. antarctica* include GH 5, GH 45, GH10, and GH 18 families. For GH families 5 and 45, most of the enzymes comprise endo-β-glucanases, which hydrolyse -1,3 or -1,4 linkages in beta glucans. GH 45 is also the largest group of signal peptide-containing enzymes, suggesting that the members of this group could be extracellularly secreted from *G. antarctica*. The substrates for endo-β-glucanases are beta glucans, namely cellulose, laminarin and lichenin. Cellulose

Table 1. List of proteins with glycoside hydrolase domains identified in the *G. antarctica* genome

	G. antarctica gene identification number*	Annotation	E-value	GH classification
1	LAN_01_115	Glucosidase I (*Pseudozyma antarctica*)	0	GH63
2	LAN_01_126	Barwin-related endoglucanase (*Hordeum vulgare*)	8e-18	GH45
3	LAN_01_290	Barwin-related endoglucanase (*Laccaria bicolor*)	5e-36	GH45
4	LAN_01_370	Endoglucanase-like protein (*Cryptosporangium arvum*)	7e-05	GH45
5	LAN_02_055	Endo-glycoceramidase (*Rhodococcus sp.)*	8e-05	GH5
6	LAN_02_121	Exo-beta-1,3-glucanase (*Candida albicans*)	2e-22	GH5
7	LAN_02_122	1,3-beta-glucosidase (*C. albicans*)	8e-07	GH5
8	LAN_02_123	1,3-beta-glucosidase (*C. albicans*)	1e-07	GH5
9	LAN_03_017	Lichenase (*Clostridium thermocellum*)	7e-06	GH26
10	LAN_03_046	Endo-1,3(4)-beta-glucanase, GH 16 (*Bacillus licheniformis*)	5e-12	GH16
11	LAN_03_240	Glycosyl hydrolase 5 (cellulase A) (*Triticum aestivum*)	0	GH5
12	LAN_03_255	Transglycosylase SLT domain protein (*Neosartorya fischeri)*	1e-25	GH23
13	LAN_03_593	Alpha-glucosidase (*Ruminococcis obeum*)	3e-76	GH31
14	LAN_03_646	Mannosyl oligosaccharide glucosidase (*Saccharomyces cerevisiae*)	0	GH63
15	LAN_03_748	Endo-β-1,3-glucanase (*Thermotoga petrophila*)	4e-08	GH16
16	LAN_03_749	Endo-β-1,3-glucanase (*Rhodosporidium toruloides*)	9e-10	GH16
17	LAN_04_028	Alpha-galactosidase (*Ajellomyces capsulatus*)	3e-35	GH23
18	LAN_04_169	Barwin-related endoglucanase (*Dichomitus squalens*)	.61	GH45
19	LAN_04_221	Chitin deacetylase (*Aspergillus nidulans*)	5e-13	GH10
20	LAN_04_222	Chitin deacetylase (*A. nidulans*)	2e-17	GH10
21	LAN_04_430	Beta-galactosidase (*Escherichia coli*)	1e-52	GH72
22	LAN_05_070	Beta-hexosaminidase (*Streotomyces coelicolor*)	1e-93	GH20
23	LAN_05_143	Beta-D-glucan exohydrolase (*Aspergillus aculeatus*)	6e-26	GH3
24	LAN_05_160	Endo-1,3(4)-beta-glucanase (*A. nidulans*)	4e-08	GH16
25	LAN_05_489	Chitinase (*Bacillus circulans*)	2e-17	GH18
25	LAN_05_492	Polysaccharide deacetylase (*R. toruloides*)	6e-08	GH16
27	LAN_05_524	Barwin-related endoglucanase (*Stereum hirsutum*)	1e-09	GH45
28	LAN_05_528	Barwin-related endoglucanase (*Gloeophyllum trabeum*)	3e-05	GH45
29	LAN_06_001	Glycosidase (*Auricularia delicata*)	3e-06	GH88
30	LAN_06_080	Polysaccharide deacetylase (*Roseomonas cervicalis*)	1e-33	GH10
31	LAN_06_221	Polysaccharide deacetylase (*Verticillium dahliae*)	1e-09	GH10
32	LAN_06_222	Polysaccharide deacetylase (*Moniliophthora roreri*)	4e-36	GH10
33	LAN_07_007	B-mannosidase (*R. toruloides*)	1e-39	GH2
34	LAN_07_048	Barwin-like endoglucanase (*Neurospora tetrasperma)*	5e-06	GH45
35	LAN_07_165	Endo-beta-mannanase (*R. toruloides*)	8e-05	GH5
36	LAN_07_220	Exo-beta-1,3-glucanase (*Rhizoctonia solani*)	2e-08	GH55
37	LAN_07_233	Exo-beta1,3-glucanase (*Laccaria bicolor*)	8e-10	GH55
38	LAN_08_020	Endo-1,4-beta-D-mannanase (*R. toruloides*)	6e-52	GH5
39	LAN_08_021	Endo-1,4-beta-D-mannanase (*R. toruloides*)	1e-45	GH5
40	LAN_08_171	Alpha-1,2 mannosidase (*Rhodotorula glutinis*)	7e-42	GH47
41	LAN_08_180	Exo-oligoxylanase (*Macrophomina phaseolina*)	4e-65	GH10
42	LAN_08_216	Alpha-galactosidase (*Trametes versicolor*)	8e-55	GH31
43	LAN_08_305	Mannosidase (*Fomitiporia mediterranea*)	2e-90	GH92
44	LAN_09_198	Glucan beta-glucosidase (*Pseudozyma aphidis*)	2e-108	GH17
45	LAN_09_234	Mannan endo-1,6-alpha-mannosidase (*Verticillium dahliae*)	4e-26	GH76
46	LAN_09_291	Alpha-1,2 mannosidase (*R. toruloides*)	5e-68	GH47
47	LAN_10_095	Alpha-glucosidase (*R. solani*)	4e-21	GH31
48	LAN_10_097	Glucoamylase (*R. toruloides*)	3e-60	GH15
49	LAN_10_172	Chitinase (*R. toruloides*)	2e-43	GH18
50	LAN_10_180	Beta-lactamase (*R. toruloides*)	1e-06	GH3
51	LAN_10_181	Beta-lactamase (*R. toruloides*)	5e-14	GH3
52	LAN_10_189	Glycosidase (*Melampsora larici-populina*)	5e-05	GH16
53	LAN_10_210	Barwin-like endoglucanase (*R. toruloides*)	9e-07	GH45
54	LAN_10_275	Beta-lactamase (*R. toruloides*)	1e-120	GH3
55	LAN_11_034	Mannan endo-1,6-alpha-mannosidase (*Schizosaccharomyces octosporus*)	5e-27	GH76
56	LAN_11_156	Endo-β-1,3-glucanase (*M. larici-populina*)	2e-06	GH16
57	LAN_11_306	Cytosolic endo-beta-N-acetylglucosaminidase (*Morus notabilis*)	4e-49	GH85
58	LAN_11_346	Galactose mutarotase (*Lactococcus lactis*)	2e-05	GH-type carbohydrate binding subgroup domain
59	LAN_11_456	Chitinase (*M. larici-populina*)	5e-14	GH18
60	LAN_11_482	Oligo-1,6-glucosidase (*R. toruloides*)	1e-105	GH13
61	LAN_11_485	Endo-1,4-beta mannanase (*R. toruloides*)	2e-55	GH5
62	LAN_11_506	Endo-beta-mannanase (*R. toruloides)*	7e-48	GH5
63	LAN_12_102	Trehalase (*R. toruloides)*	2e-06	GH37

Table 1 continue...

Table 1 continued...

64	LAN_12_113	Chitinase (*R. toruloides)*	5e-47	GH18
65	LAN_12_116	Galactose mutarotase (*R. toruloides)*	6e-05	GH16
66	LAN_12_147	Endo-beta-D-1,4-mannanase (*R. toruloides)*	5e-80	GH5
67	LAN_12_265	Alpha-glucosidase (*R. toruloies)*	2e-06	GH13
68	LAN_12_427	Alpha-1,2-mannosidase ((*R. toruloides)*	1e-150	GH47
69	LAN_12_521	Alpha-mannosidase (*R. toruloides)*	1e-128	GH38
70	LAN_13_036	Alpha-amylase (*Cryptococcus neoformans*)	0	GH13
71	LAN_14_046	Chitinase (*Puccinia triticina*)	7e-54	GH18
72	LAN_14_077	Glucoamylase (*M. larici-populina)*	2e-54	GH15
73	LAN_14_140	Expansin (*R. toruloides*)	5e-06	GH45
74	LAN_14_164	Cellulase (*R. toruloides)*	9e-11	GH5
75	LAN_14_165	Endo-beta-1,3-glucanase (*R. toruloides)*	9e-05	GH81
76	LAN_14_244	Beta-1,3-xylanase (*R. toruloides)*	1e-35	GH26
77	LAN_14_270	Chitin deacetylase (*R. toruloides)*	6e-09	GH10
78	LAN_14_284	Allantoicase (*Glarea lozoyensis*)	5e-84	Galactose-binding domain like
79	LAN_14_333	Galactose mutarotase (*R. solani*)	3e-37	GH31
80	LAN_15_276	Chitin deacetylase (*R. toruloides)*		GH18
81	LAN_15_282	Chitin deacetylase (*Puccinia graminis)*	6e-11	GH10
82	LAN_15_284	Chitin deacetylase (*P. graminis*)	2e-13	GH10
83	LAN_15_297	Chitin deacetylase (*P. graminis)*	1e-06	GH10
84	LAN_16_072	Polyphosphoinositide phosphatase (*R. solani*)	0	GH31
85	LAN_16_363	Beta-1,3-xylanase (*Dacryopinax sp.*)	8e-76	GH26
86	LAN_16_504	Hexose-6-phosphate mutarotase (*R. toruloides)*	3e-32	GH10
87	LAN_16_507	Glucan synthase (*Metarhizium anisopliae)*	3e-30	GH23
88	LAN_16_574	N-acetylhexosaminidase (*Postia placenta*)	8e-132	GH20
89	LAN_16_632	Expansin (*R. toruloides)*	1e-41	GH45
90	LAN_16_634	Expansin (*R. toruloides)*	8e-34	GH45
91	LAN_16_648	Alpha-1,6-mannanase *(R. toruloides)*	6e-30	GH76
92	LAN_16_668	Alpha-amylase (*M. larici-populina*)	0	GH13
93	LAN_16_703	Alpha-amylase (*Microbotryum violaceum*)	0	GH13
94	LAN_16_718	Endo-beta-1,3-glucanase (*M. larici-populina*)	8e-06	GH16
95	LAN_16_736	Cold-induced thioredoxin domain-containing protein (*C. neoformans*)	0	GH45
96	LAN_16_754	Exo-alpha-1,6-mannosidase (*Elizabethkingia anophelis*)	4e-21	GH126
97	LAN_16_887	Chitinase (*Coniophora puteana*)	5e-58	GH18

*The identification number assigned to each gene in the *G. antarctica* genome database (http://27.126.156.144/glaciozyma_antarctica/)

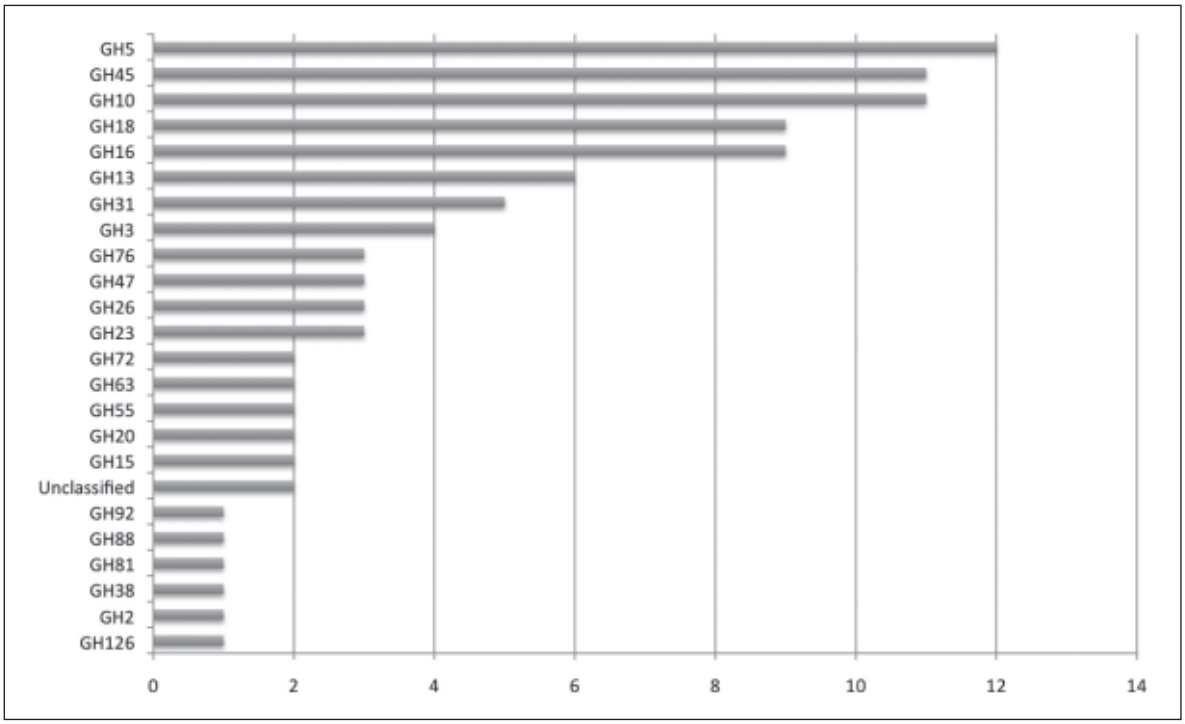

Fig. 1. Distribution of glycoside hydrolase families in *G. antarctica.*

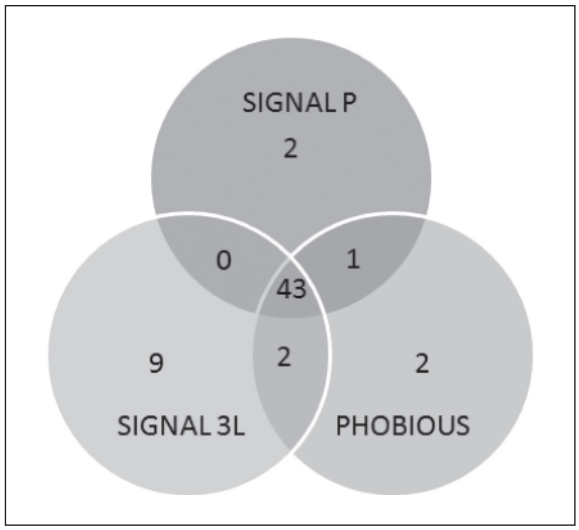

Fig. 2. Venn diagram of the three bioinformatics software programmes used to predict the signal peptide sequence in the glycoside hydrolases of *G. antarctica*. A total of 46 proteins were predicted to contain a signal peptide in at least two software programmes.

Table 2. Summary of *G. antarctica* glycoside hydrolases that contain signal peptides

Family	Predicted activities	Number of proteins with signal peptides
GH45	Endoglucanase	9
GH10	Xylanase	7
GH16	Endo-1,3(4)-β-Glucanase	5
GH5	Endo-1,4-β-D-Mannanase	4
GH76	α-1,6-mannanase	3
GH31	α-glucosidase	3
GH26	Lichenase	3
GH18	Chitinase	3
GH55	Exo-β-1,3-glucanase	2
GH88	β-glucuronyl hydrolase	1
GH72	β-1,3-glucanosyltransglycosylase	1
GH63	α-glucosidase	1
GH47	α-mannosidase	1
GH23	Lysozyme	1
GH20	β-hexosaminidase	1
GH13	α-amylase	1

is an extremely important component of the primary cell wall of plants, while laminarin is a storage glucan in brown algae and lichenan is a complex glucan identified in certain lichen species. The secreted enzymes might facilitate nutrient scavenging by *G. antarctica*. Studies have previously shown that fungi isolated from the interior structural woods at Ross Island, Antarctica secrete endoglucanases into the environment (Duncan *et al*., 2006). These enzymes are active at low temperatures, as indicated through the cellulose degradation observed at 4ºC and 15ºC.

Thus, in the present study, we classified 97 putative genes from *G. antarctica* into 23 different glycoside hydrolase families from the 133 CAZy GH families available. These groups primarily comprise extracellular enzymes, including endoglucanases, xylanases and chitinases, which are important for nutrient utilisation for the survival of these microorganisms. The optimal activity of glycoside hydrolases at low temperatures facilitates the study of cold active enzyme mechanisms and the potential application of these enzymes in biotechnology. Indeed, extracellular cold enzymes have useful functions in food, detergent and biofuel industries. Therefore, it is important to elucidate the biochemical properties of these annotated glycoside hydrolases to understand the unique properties of these enzymes.

ACKNOWLEDGEMENTS

The authors would like to thank Malaysian Genome Institute (MGI) for providing access to the *Glaciozyma antarctica* Genome Database. This study was financially supported through the Ministry of Science and Technology, Malaysia (MOSTI), grant no. 02-05-20-SF0007 and 10-05-16-MB002.

REFERENCES

Brizzio, S., Turchetti, B., de Garcia, V., Libkind, D., Buzzini, P. & van Broock, M. 2007. Extracellular enzymatic activities of Basidiomycetous yeasts isolated from glacial and subglacial waters of northwest Patagonia (Argentina). *Canadian Journal of Microbiology*, **53(4)**: 519-525.

Damasio, A.R.L., Rubio, M.V., Oliveira, L.C., Segato, F., Dias, B.A., Citadini, A.P., Paixao, D.A. & Sequina, F.M. 2014. Understanding the function of conserved variations in the catalytic loop of fungal glycoside hydrolase family 12. *Biotechnology and Bioengineering,* **111(8)**: 1494-1505.

Duncan, S.M., Farrell, R.L., Thwaites, J.M., Held, B.W., Arenz, B.E., Jurgens, J.A. & Blanchette, R.A. 2006. Endoglucanase-producing fungi isolated from Cape Evans historic expedition hut on Ross Island, Antarctica. *Environmental Microbiology*, **8(7)**: 1212-1219.

Hashim, N.H.F., Bharudin, I., Law, D.S.N., Higa, S., Bakar, F.D.A., Nathan, S., Rabu, A., Kawahara, H., Illias, R.M., Najimudin, N., Mahadi, N.M. & Murad, A.M.A. 2013. Characterization of Afp1, an antifreeze protein from the psychrophilic yeast *Glaciozyma antarctica* PI12. *Extremophiles*, **17**: 63-73.

Kawahara, H. 2002. The structure and functions of ice crystal-controlling proteins from bacteria. *Journal of Bioscience and Bioengineering*, **94(6)**: 492-496.

Lombard, V., Ramulu, H.G., Drula, E., Coutinho, P.M. & Henrissat, B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research*, **42**: D490-495.

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.

Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R. & Lopez, R. 2005. InterProScan: Protein domains identifier. *Nucleic Acids Research*, **33(SUPPL. 2)**: W116-W120.

Yu, Y., Li, H., Zeng, Y. & Chen, B. 2009. Extracellular enzymes of cold-adapted bacteria from Arctic sea ice, Canada Basin. *Polar Biology*, **32(10)**: 1539-1547.

Zhao, Z., Liu, H., Wang, C. & Xu, J.R. 2013. Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. *BMC Genomics*, **14**: 274.