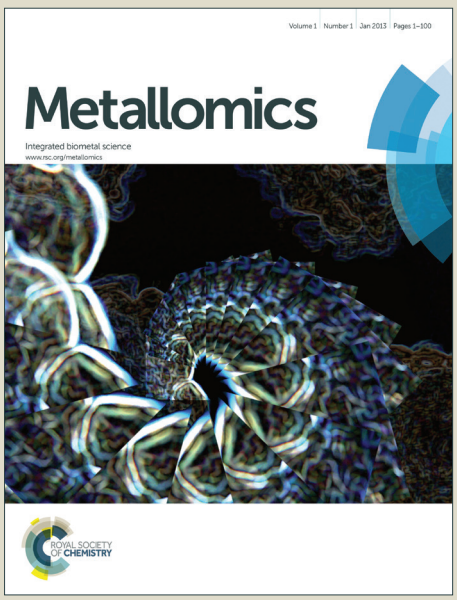
Metallomics

Accepted Manuscript

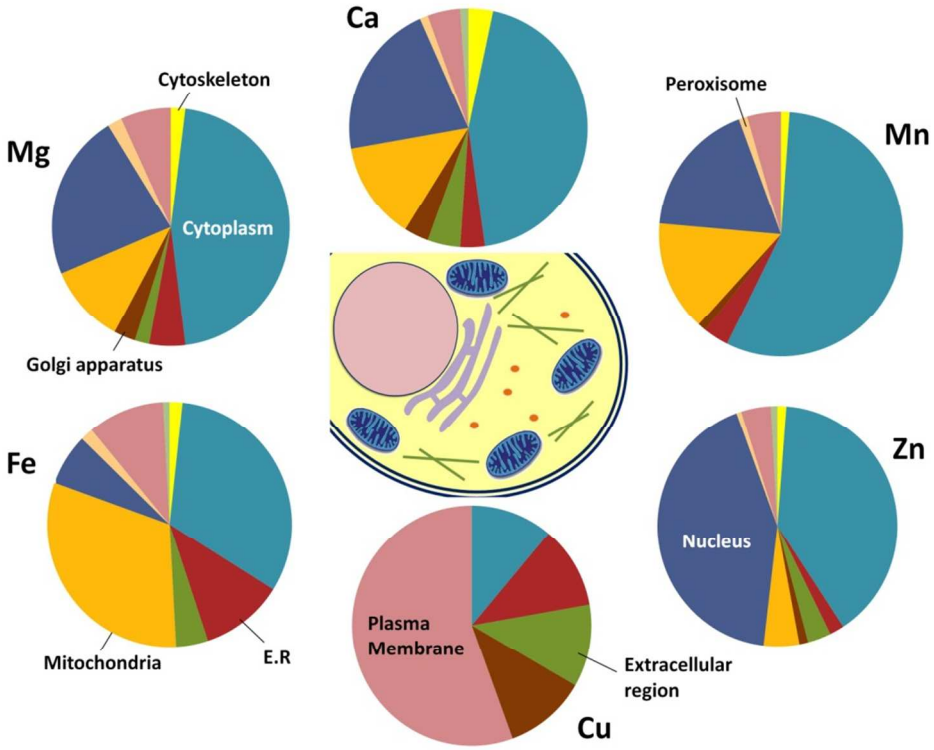


ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/metallomics

Table of contents entry

Psychrophilic metallome of *Glaciozyma antarctica* PI12 predicted by bioinformatic approaches.



Bioinformatics survey of the metal usage by psychrophilic yeast *Glaciozyma antarctica* PI12

Pik Mun Foong[1,2], Roghayeh Abedi Karjiban[1,2], Yahaya M. Normi[1], Abu Bakar Salleh[1], Mohd Basyaruddin Abdul Rahman[1, 2,3*]

[1]Enzyme and Microbial Technology Research Center (EMTech), Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor Darul Ehsan, Malaysia.

[2]Department of Chemistry, Faculty of Science, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor Darul Ehsan, Malaysia.

[3]Structural and Synthetic Biology Research Centre, Malaysia Genome Institute, Jalan Bangi, 43000 Kajang, Selangor Darul Ehsan, Malaysia.

[*] **Corresponding author:**

Mohd Basyaruddin Abdul Rahman

E-mail: basya@upm.edu.my

Phone: +603-8946 6798

Abstract

Metal ions are one of the essential elements which are extensively involved in many cellular activities. With rapid advancements in genome sequencing techniques, bioinformatics approaches have provided a promising way to extract functional information of a protein directly from its primary structure. Recent findings have suggested that the metal content of an organism can be predicted from its complete genome sequences. Characterizing the biological metal usage of cold-adapted organism may help to outline a comprehensive understanding of the metal-partnerships between the psychrophile and its adjacent environment. The focus of this study is targeted towards the analysis of the metal composition of a psychrophilic yeast *Glaciozyma antarctica* PI12 isolated from sea ice of Antarctica. Since the cellular metal content of an organism is usually reflected in the expressed metal-binding proteins, the putative metal-binding sequences from *G. antarctica* PI12 were identified in respect to their sequence homologies, domain compositions, protein families and cellular distribution. Most of the analyses revealed that the proteome was enriched with zinc, and the content of metal decreased in the order of Zn > Fe > Mg > Mn, Ca > Cu. Upon comparison, it was found that the metal compositions among yeasts were almost identical. These observations suggested that *G. antarctica* PI12 could have inherited a conserved trend of metal usage similar to modern eukaryotes, despite its geographically isolated habitat.

Keywords

metallome, metalloproteins, bioinformatics, psychrophilic yeast, zinc ion

Introduction

Metal ions are essential to almost all living organisms. They usually participate as cofactors in many biological processes, regulate cellular activities or provide structural supports. Lacking of these metals can cause proteins to malfunction. Indeed, previous studies have reported that metal-bounded proteins are actually widespread in all organisms at varying composition, and is entwined with their respective habitats and metabolic preferences[1,2]. Therefore, any significant perturbation within the cell and its adjacent environment can influence biological metal composition. However, most of the high-throughput experimental approaches used to identify the complete set of metalloproteins encoded by an organism are still under development and are usually laborious and resource-demanding[3,4]. Since rapid advancement in genome sequencing techniques have generated mountains of sequence data each day, predictive tools that enable scientists to sieve through an organism's genetic blueprint and subsequently analyze in detail the sequence(s) of interest are invaluable. Emergence of bioinformatics approaches has provided a prominent way to identify the putative metal-binding proteins based on the presence of specific metal-binding sites or protein domains in the amino acid sequences, and is readily to solve the question of how many and which proteins may require metal ions to function properly[5,6].

Psychrophiles are defined as organisms which are capable to thrive at very low temperatures in freezing-habitats[7]. Cold temperatures could adversely impact nearly all levels of cell architectures and retard the growth of cell. A complex range of adaptations is hence adopted by psychrophiles in order to withstand harsh and cold environment. Despite renowned as the major spoilage agent in refrigerated and frozen foodstuff, psychrophilic organisms are of great interest to scientists due to their potential biotechnology values[8,9]. They are a precious source of "cold-active enzymes" with high specific activity at low and moderate

temperatures, and are inactivated easily by a moderate increase in temperature[10]. Comprehensive understanding of the characteristics of a psychrophile can thus be beneficially applied for tailoring variants of cold-active enzymes which are adaptable at desired temperatures.

The particular *Glaciozyma antarctica* strain PI12 (previously known as *Leucosporidium antarcticum*) investigated in this study was isolated from Antarctic sea ice near Casey Research Station in temperatures ranging -20 – 15 °C. Antarctica is a unique region with an area of 14 million square kilometers that are mostly covered by ice and snow. Living under such extreme and harsh environment, adaptive strategies are anticipated to be used by this organism to alleviate these stresses. In this study, the relative metal usage in the psychrophilic yeast *G. antarctica* PI12 was examined in order to depict the speciation of metal acquisition for organism surviving under freezing temperatures. We incorporated multiple generic bioinformatics tools to identify the presence of homologous metal-binding proteins, metal-binding domains and protein families from the protein sequences of *G. antarctica* PI12. In addition, several yeast and bacterial counterparts with completely sequenced genomes were also subjected to investigation to compare the overall abundance of the metal-binding proteins in their proteomes. Through such comparison, we attempted to resolve the unique physiological metal biosignature that may possibly be inherited by the psychrophiles.

Procedures

Preparation of BLAST queries

The overall strategy of this project is summarized in Figure 1. In order to trace (putative) metal-binding sites, ensembles of query for BLAST[11] were retrieved from UniProtKB/SWISS-PROT[12] protein knowledgebase [release February 2013]. During the search, metal-related keywords were used as input to identify all proteins that bind a metal ion ('*metalloproteins*'). A total of 17 ensembles of metal-specific queries were obtained: the common (ubiquitous to almost all living organisms) metal-types of "Sodium-binding", "Potassium-binding", "Magnesium-binding", "Calcium-binding", "Manganese-binding", "Iron-binding", "Cobalt-binding", "Nickel-binding", "Copper-binding", "Zinc-binding"; the occasional (rarely been found in biological systems but may crucial for particular organisms) metal-types that of "Lithium-binding", "Vanadium-binding", Molybdenum-binding", "Tungsten-binding"; and the heavy metal-types of "Cadmium-binding", "Mercury-binding" and "Lead-binding". The search results were further clustered using UniRef100[13] database to reduce redundancy. Meanwhile, the translated gene products of psychrophilic yeast *G. antarctica* PI12 (7857 sequences) were obtained from Malaysia Genome Institute and were converted into BLAST-compatible database. BLASTp was initiated with default substitution matrix BLOSUM-62 and cut-off criteria for at least 30% sequence identity and E-value less than 0.001. This was done in order to obtain reliable sequences with significant similarity to other well-characterized and/or annotated metalloproteins in the database.

Identification of functional protein domains through Pfam

The protein domain composition in the protein sequences of *G. antarctica* PI12 were annotated with domain terms defined in Pfam[14] database (release 25.0). Subsequently, lists

of metal-related families and domains were collected by querying the Pfam database with related keywords and clustered into the following groups: "Sodium-binding", "Potassium-binding", "Magnesium-binding", "Calcium-binding", "Manganese-binding", "Iron-binding", "Cobalt-binding", "Nickel-binding", "Copper-binding" and "Zinc-binding". The lists of queries were employed to scan against the domain compositions in *G. antarctica* PI12 obtained as aforementioned. With the assumption that an individual metal-binding sequence contains at least one metal-related protein domain, the resulting matches were manually inspected and the redundant entries were discarded. These sequences were later assembled based on the type of metal bounded.

Classification of protein family using machine learning program

A web-based support vector machine program SVMProt[15] was employed to classify the protein sequences into their respective families according to their sequence features. 10 metal-related properties that included in the classifier were focused on the study: "Sodium-binding", "Potassium-binding", "Magnesium-binding", "Calcium-binding", "Manganese-binding", "Iron-binding", "Cobalt-binding", "Nickel-binding", "Copper-binding" and "Zinc-binding". All the results obtained were tabulated in spreadsheet for further analysis. VENNY[16] and VENNTURE[17] were used to generate the 3-set proportional Venn diagrams and Edwards-Venn diagrams.

Subcellular location prediction of the metal-binding proteins

The distribution of the metal-binding proteins in the cellular compartments of *G. antarctica* PI12 was investigated. In order to derive a consensus prediction, three predictive programs were applied in this study, which were LocTree3[18], WoLF-Psort[19], and WegoLoc[20]. The

putative functions of the metal-binding proteins were associated with their respective locations using web-server GOanna[21].

Comparative analyses of metal content with other yeast counterparts

Three warm-adapted yeast counterparts were selected for comparing the metal content with the psychrophilic *G. antarctica* PI12 since the reference from other cold-adapted yeast is currently unavailable in the sequence databases. These included the mesophile *Saccharomyces cerevisiae* S288c (Taxon identifier: 559292), eukaryotic pathogen *Candida albicans* SC5314 (Taxon identifier: 237561) and the thermo-tolerant *Pichia angusta* ATCC 26012 (Taxon identifier: 871575). Meanwhile, the psychrophilic *Psychrobacter arcticus* 273-4 (Taxon identifier: 259536) and alkaliphilic *Bacillus lehensis* G1 (Taxon identifier: 300825) from bacteria isolates were also included for comparison. The complete protein coding genes for most of the organisms were downloaded from the UniProt and NCBI[22] databases, except for *B. lehensis* G1 which was obtained from Malaysia Genome Institute. The metal compositions in their proteomes were probed using BLASTp with similar query ensembles and parameters stated in previous section.

Results and Discussion

Homologous metal-binding sequences

Approximately a quarter (26%) of the total proteome for *G. antarctica* PI12 are predicted to be metal-bounded by BLAST (Figure 2), with majority of these proteins using essential metals as cofactors. Interestingly, the presence of heavy metal-binding proteins (e.g.: cadmium, mercury and lead) is also recorded (44 sequences) during BLAST analysis. These biologically toxic heavy metals are often speculated for their ability to dislodge the native

metal ions and distort the preferable coordination geometry, causing the proteins to lose their functions[23]. The presence of heavy metal-binding proteins in *G. antarctica* PI12 could possibly serve as an indicator for the existence of heavy metals in their adjacent environment, and may have been assimilated into the cell during metal acquisition. However, due to the reasons that the coverage of heavy metal-binding domains for Pfam domain database is relatively limited, and the coverage of SVMProt have excluded protein families bounded with heavy metals, the detection of heavy metal-binding proteins here is therefore not conclusive to rely upon a single predictive method, but is expected to be clarified as more functional and structural information of the metalloproteins are accumulated in the libraries.

Figure 3 shows the number of metal-binding sequences for *G. antarctica* PI12 recorded by BLAST, Pfam domain composition (Pfam-KW) and SVMProt. Pairwise alignment between two proteins are often useful to locate regions of similarity that usually convey structural, functional or even evolutionary information[24] . Seeking for significant sequence similarity or identity (which occur due to chance) of an unknown protein to a protein deposited in databases with experimentally characterized known function enables the annotation to be transferred, which is regarded as homology-based transfer[25,26]. Presumably, when pairwise sequence identity between the two proteins achieved more than 30%, they are inclined to be similar in structure and function[25]. As in the case for the metal-binding proteins, BLAST has predicted that the *G. antarctica* PI12 proteome is predominantly enriched with zinc-binding proteins, with 1317 significant homologous sequences identified during the search. This is followed by magnesium- and calcium-binding proteins with 643 and 402 sequences respectively.

In fact, protein tends to diverge from its functions more rapidly in nature, and consequently homologous proteins could have evolved with diverse functions. Therefore, inspecting the functional unit of a protein, or also referred to as protein domain, thus offer better accuracy for functional inference. Protein domains are defined as compact regions present in protein structures, with particular conserved amino acid residues folded into a relatively similar conformation[27] which eventually shape the analogous metal-binding site. We scanned for protein domain composition of *G. antarctica* PI12 using the lists of metal-related keywords from Pfam database, and a total of 3925 significant matches was obtained. With the assumption that all of these sequences contain at least one metal-related protein domain, we observed that the manganese-binding proteins (2556 proteins) are more prevalent than zinc proteins (2274 proteins). The contradiction might possibly be due to higher availability of annotated entries associated to manganese-binding domains in Pfam database which subsequently resulted to higher hits obtained during the screening.

To enhance the coverage of prediction, the statistical learning program SVMProt was employed in this study. This web-based program is useful to identify unprecedented metal-binding proteins, especially involving proteins which do not have annotated homologues that may have been overlooked due to lack of significant sequence similarity[28]. In this study, SVMProt described the proteome as tremendously metal-enriched, in which 5458 sequences have been identified by the program as zinc-binding proteins. This is followed in order by magnesium (2466 sequences), calcium (2020 sequences) and iron proteins (1939 sequences).

Figure 4(A) summarizes the results of the mentioned approaches and their overlaps. By overlapping the results obtained, it is noticed that 1750 proteins are identified in parallel by all three methods as putatively metal-bounded, while 3908 proteins are detected by at least

two approaches. There are only 15 proteins which were detected by BLAST; and 182 proteins independently identified by Pfam-KW. Meanwhile, nearly half of the metal-binding sequences predicted by SVMProt (3781 proteins) are detected by either of the other two approaches, and another half of the metal-binding proteins are particularly recorded by SVMProt (3304 proteins). By calculating the number of proteins identified as metal-binding by two out of three approaches, it is discerned that the ensemble of proteins retrieved by BLAST and Pfam-KW (127 sequences) along with BLAST and SVMProt (165 sequences) are relatively close in size. In contrast, the ensemble of proteins detected by Pfam-KW and SVMProt is distinctively more (1866 sequences), which possibly suggest that the protein functional family training sets for SVMProt has covered most of the annotated metal-binding domains available in Pfam database. However, the usage of SVMProt should be mindful of its rather weak discrimination power (with only 62.5% of homologous proteins are recovered during the training)[15] and limited range of metal-binding protein family that are currently supplied for the training sets.

To summarize the results obtained, the putative metals quota for the psychrophilic yeast *G. antarctica* PI12 in combination of all approaches has showed a hierarchy of abundance that descended in the order of Zn > Fe > Mg > Mn, Ca > Cu [Figure 4(B)]. The low number of copper-binding proteins may correlate to the metal's tremendous efficiency in redox activity that under certain circumstances, can appear as a critical challenge to biological systems [29]. The excessive redox activity can cause the generation of highly reactive oxygen species (ROS) in particular hydroxyl radicals through Fenton reaction which can damage the cell. Therefore, the intracellular concentration of copper is tightly control to prevent its toxic potential[27, 30].

Distribution of the metal-binding proteins in cell

The cellular distribution of the putative metal-binding proteins of *G. antarctica* PI12 consensually identified by all the approaches is depicted in Figure 5. Zinc-binding proteins are found in almost all of the cellular compartments, with the largest fraction (392 sequences) being in the cell nucleus. The extensive usages of zinc in the nucleus could be interpreted for the intense requirement of zinc proteins to regulate gene expression and to maintain genome integrity[29,31]. The crucial role of magnesium in stabilizing the nucleic acid structures can also be observed from the presence of 22% magnesium-binding proteins (23 sequences) also located in the nucleus, which accounted as the second most abundant metal after zinc. Meanwhile, the relatively high level of calcium in nucleus (19 sequences) is postulated to be integrated into the typical EF-hand hand motif of calmodulin which is often involved as a signal transducer in cell-cycle events and proliferation[32].

A comparable fraction of zinc- (44 sequences) and iron-binding proteins (37 sequences) has been detected in the mitochondria compartment. Zinc proteins in the mitochondria have been predominantly found to be zinc-containing superoxide dismutase which control the generation of reactive oxygen species, and/or recruiting zinc ions as cofactors in a series of metalloenzymes[33,34]. Indeed, the occurrence of iron proteins in mitochondria is consistent to their prominent role in facilitating electron transfer process via transition between different oxidation states[29, 35]. 13 manganese proteins are found localized in the mitochondria, with most of them are found to catalyze the dismutation of superoxide and prevent the cell from oxidative damage[36]. The presence of magnesium proteins (11 sequences) in this compartment is found in relevance to the enzymes that are involved in energy metabolism such as glycolysis, which are mostly magnesium-dependent[37]; whereas the considerable amount of

calcium proteins (12 sequences) noticed here supports the hypothesis that the mitochondria could be one of the intracellular store for calcium[38].

During the screening, we also noticed an elevated usage of iron proteins (13 sequences) after zinc in endoplasmic reticulum (ER). This in complying with the role of ER as the protein factory in cell that ensures the correctness of protein folding, and thus has an unique oxidizing-folding environment loaded with various protein chaperones and enzymes[39]. Alteration of the redox state of ER could affect protein folding and disulfide bond formation, and hence widespread of metal-dependent superoxide dismutases are anticipated in this compartment.

A substantial proportion of metal-binding proteins have been observed in the cellular cytoplasmic pool, which comprised of over one-fourth of all metal-binding protein identified in each respective metal-type (except copper). A myriad of zinc proteins are found in this compartment (364 sequences), followed by manganese (50 sequences), magnesium (47 sequences), calcium (40 sequences) and iron proteins (38 sequences) in relatively similar fractions. The prevalence of these metal-binding proteins in this compartment may correlated to the shuttling of metals to enter and/or exit the cell compartments in equilibrium[40]. Likewise, the abundance of metal-binding proteins, especially zinc proteins (37 sequences) that are present in the plasma membrane may correspond to various metal transporters and sensors situated in the membrane that play key roles in maintaining the optimal intracellular metal pools of the cell[41].

It is also discerned that some of the metal-binding proteins are localized in the extracellular region, which can be defined as the external space proximal to the outermost membrane of a

cell. This region is often reported to be the reservoir of metal-containing extracellular degrading- and antioxidant-enzymes which are encapsulated in the form of vesicles[42]. The findings have affirmed that the distribution of a metal-binding protein in the cellular compartments has conveyed its functional information.

Comparative metal composition with other counterparts

Several studies have claimed that the compartmentalization of the metal-binding proteins in the cell could be possibly restrained by the bioavailability of the respective metals in the adjacent environment[30,43]. Considerable perturbations in the environment would potentially induce the proteins to evolve with different scaffolds and/or recruit different metal ions in order to confront the stress[44], and this speciation of the elemental composition in an organism can eventually influence its evolution[45,46]. Emphasizing only on prokaryotes, Zerkle and co-worker[46] has attempted to estimate prokaryotic metallomes . They observed that prokaryotes have generally followed a hierarchy of metal concentration as follows: Fe > Zn > Mn > Mo (Molybdenum), Co (Cobalt), Cu > Ni (Nickel) > W (Tungsten), V (Vanadium). A similar estimation has also been performed recently by Cameron and colleagues[1] on the metallomes of hyperthermophiles. Both teams have reported an elevated usage of nickel and cobalt in methanogens and hyperthermophiles, suggesting that these trace elements may be crucial to these extremophiles. However, the reference for metal usage in psychrophiles remained missing. Hence, comparative analyses on the metal usage by other counterparts with different metabolic preference were conducted in this present study. As there is still a lack of other completely-sequenced psychrophilic yeast genomes in the public databank to permit a comparison of metal acquisition between cold-adapted yeasts, a psycrophillic bacterial strain, *Psychrobacter arcticus* isolated from the Siberian Permafrost has been employed here for comparative studies across different taxa. The alkaliphilic bacterium *Bacillus lehensis* was

also incorporated in the analyses as this bacterium has extensive applications in industries, and yet its metal content is poorly described.

The variations of metal composition predicted by BLAST for each metal-type in all four studied yeast proteomes are almost identical, with slight differences noticed in the contents of magnesium and iron (Figure 6). The composition of the magnesium-binding proteins ranged from 18% in *G. antarctica* PI12 and *C. albicans* SC5314 to 21% in *S. cerevisiae* S288c, while the iron contents varied between 9-12%. The presence of cobalt-, nickel-, and copper-binding proteins in the yeasts are rather scarce, with an average of 1-2%. All the proteomes are noticeably zinc-enriched, with zinc compositions constituted 34-38% of the respective metallomes. Andreini and co-authors [6] have justified that prokaryotes generally adopt zinc for catalytic activity, while eukaryotes have additional zinc usages in regulating gene expression, cell compartmentalization, and cell differentiation, which imply the significant requirement of zinc proteins in higher organisms.

On the contrary, a discernible variation in the metal composition is noticed in both the bacterial extremophiles when compared with the metal content of *G. antarctica* PI12 (Figure 7). Magnesium-binding proteins in *P. arcticus* 273-4 proteome are particularly prevalent, which constituted about a quarter (24%) of its metallome. *B. lehensis* G1 on the other hand, has a similar magnesium content with *G. antarctica* PI12 (18%). However, the calcium contents for the two bacterial strains (both 5%) are apparently reduced as compared to *G. antarctica* PI12 (12%). We postulate that the higher acquisition of calcium in yeast may correspond to the role of calcium in regulating cellular signaling which control a vast array of eukaryotic cellular responds[47,48]. In comparison, the fraction of iron-binding proteins in *G. antarctica* PI12 are considerably lower compared to the bacteria. The findings seem to be

consistent with results reported by previous studies [27,35]. Apart from this, it is noticed that elevation in thermophily is paralleled with the increase of metal usage when the metal composition of thermophile *P. angusta* ATCC 26012 is compared with the psychrophiles *G. antarctica* PI12 and P. *arcticus* 273-4 in relative to their proteome sizes (Figure 8). This could be interpreted as an importance of metals especially zinc in providing structural rigidity to the proteins as the temperature increased[6]. Meanwhile, we observed that the cold-adapted bacterium *P. arcticus* 273-4 has in general recruited more metals than *G. antarctica* PI12. The molecular determinant which influences the bulk metals usage of the psychrophilic bacterium in comparison to the yeast regrettably remains a puzzle.

Conclusion

In the present work, we have conducted a survey on the metal preference of a psychrophilic yeast *G. antarctica* PI12 using the integration of multiple sequence-based bioinformatics approaches. The results showed a quarter of its proteome was found to be metal-enriched and was prevalent with zinc. Metal utilization for *G. antarctica* PI12 followed the trend as Zn > Fe > Mg > Mn, Ca > Cu. Predicting the compartmentalization of the metal-binding proteins in the cell has indeed ascertained the wide-ranging participation of metals in cellular activities. However, the psychrophilic yeast has seemed to adopt a trend of metal usage which resembles other yeasts from warm-counterparts, despite its geographically isolated habitat. It can therefore be inferred that the bioavailability of the trace metal surrounded the organism could be the key factor which influence its metallome. In conclusion, the preference of metal acquisition by the psychrophilic yeast *G. antarctica* PI12 can be outlined with the aid of various bioinformatic resources, and the present work provides an alternative to explore the role and function of metal ions in the cellular system from such an integrated view.

Acknowledgments

This work was supported by the Ministry of Science, Technology and Innovation (MOSTI) Malaysia under the Genomics and Molecular Biology Initiative of the Malaysia Genome Institute (GMBI 5487735) and a sponsored Graduate Research Fellowship (GRF) for Pik Mun Foong from Universiti Putra Malaysia.

Conflict of interest

The authors declare that they have no competing interests.

References

1. V. Cameron, C. H. House, S. L. Brantley, A First Analysis of Metallome Biosignatures of Hyperthermophilic Archaea. *Archaea* 2012, *2012*.
2. C. Andreini, I. Bertini, A. Rosato, Metalloproteomes: A Bioinformatic Approach. *Accounts of Chemical Research* 2009, *42*. 1471-1479.
3. A. Passerini, C. Andreini, S. Menchetti, A. Rosato, P. Frasconi, Predicting zinc binding at the proteome level. *BMC Bioinformatics* 2007, *8*. 39.
4. C. Andreini, L. Banci, I. Bertini, A. Rosato, Counting the Zinc-Proteins Encoded in the Human Genome. *Journal of Proteome Research* 2006, *5*. 196-201.
5. X.-F. Wan, D. Xu, Computational methods for remote homolog identification. *Current Protein and Peptide Science* 2005, *6*. 527-546.
6. C. Andreini, L. Banci, I. Bertini, A. Rosato, Zinc through the Three Domains of Life. *Journal of Proteome Research* 2006, *5*. 3173-3178.
7. G. Feller, Psychrophilic enzymes: from folding to function and biotechnology. *Scientifica* 2013, *2013*.
8. C. H. Robinson, Cold adaptation in Arctic and Antarctic fungi. *New phytologist* 2001, *151*. 341-353.
9. R. Margesin, F. Schinner, Properties of cold-adapted microorganisms and their potential role in biotechnology. *Journal of Biotechnology* 1994, *33*. 1-14.
10. M. Turkiewicz, M. Pazgier, H. Kalinowska, S. Bielecki, A cold-adapted extracellular serine proteinase of the yeast *Leucosporidium antarcticum*. *Extremophiles* 2003, *7*. 435-442.
11. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *Journal of Molecular Biology* 1990, *215*. 403-10.
12. A. Bairoch, R. Apweiler, C. H. Wu, W. C. Barker, B. Boeckmann, S. Ferro, E. Gasteiger, H. Huang, R. Lopez, M. Magrane, The universal protein resource (UniProt). *Nucleic Acids Research* 2005, *33*. D154-D159.
13. B. E. Suzek, H. Huang, P. McGarvey, R. Mazumder, C. H. Wu, UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics* 2007, *23*. 1282-1288.
14. A. Bateman, L. Coin, R. Durbin, R. D. Finn, V. Hollich, S. Griffiths-Jones, A. Khanna, M. Marshall, S. Moxon, E. L. Sonnhammer, The Pfam protein families database. *Nucleic Acids Research* 2004, *32*. D138-D141.

15. C. Z. Cai, L. Y. Han, Z. L. Ji, X. Chen, Y. Z. Chen, SVM-Prot: Web-based support vector machine software for functional classification of a protein from its primary sequence. *Nucleic Acids Research* 2003, *31*. 3692-7.
16. J. Oliveros, VENNY:An interactive tool for comparing lists with Venn Diagrams. *BioinfoGP, CNB-CSIC* 2007.
17. B. Martin, W. Chadwick, T. Yi, S.-S. Park, D. Lu, B. Ni, S. Gadkaree, K. Farhang, K. G. Becker, S. Maudsley, VENNTURE-a novel Venn diagram investigational tool for multiple pharmacological dataset analysis. *PloS ONE* 2012, *7*. e36911.
18. R. Nair, B. Rost, Mimicking cellular sorting improves prediction of subcellular localization. *Journal of Molecular Biology* 2005, *348*. 85-100.
19. P. Horton, K.-J. Park, T. Obayashi, N. Fujita, H. Harada, C. J. Adams-Collier, K. Nakai, WoLF PSORT: protein localization predictor. *Nucleic Acids Research* 2007, *35*. W585-W587.
20. S.-M. Chi, D. Nam, WegoLoc: accurate prediction of protein subcellular localization using weighted Gene Ontology terms. *Bioinformatics* 2012, *28*. 1028-1030.
21. F. M. McCarthy, S. M. Bridges, N. Wang, G. B. Magee, W. P. Williams, D. S. Luthe, S. C. Burgess, AgBase: a unified resource for functional analysis in agriculture. *Nucleic Acids Research* 2007, *35*. D599-D603.
22. D. L. Wheeler, T. Barrett, D. A. Benson, S. H. Bryant, K. Canese, V. Chetvernin, D. M. Church, M. DiCuccio, R. Edgar, S. Federhen, Database resources of the national center for biotechnology information. *Nucleic acids research* 2007, *35*. D5-D12.
23. T. Dudev, C. Lim, Metal binding and selectivity in zinc proteins. *Journal of the Chinese Chemical Society* 2003, *50*. 1093-1102.
24. B. Rost, Twilight zone of protein sequence alignments. *Protein Engineering* 1999, *12*. 85-94.
25. W. Tian, J. Skolnick, How well is enzyme function conserved as a function of pairwise sequence identity? *Journal of Molecular Biology* 2003, *333*. 863-882.
26. I. Friedberg, Automated protein function prediction-the genomic challenge. *Briefings in Bioinformatics* 2006, *7*. 225-242.
27. I. Bertini, A. Rosato, From genes to metalloproteins: A bioinformatic approach. *European Journal of Inorganic Chemistry* 2007, *2007*. 2546-2555.
28. L. Y. Han, C. Z. Cai, Z. L. Ji, Z. W. Cao, J. Cui, Y. Z. Chen, Predicting functional family of novel enzymes irrespective of sequence similarity: a statistical learning approach. *Nucleic Acids Research* 2004, *32*. 6437-6444.
29. U. Lindh, in *Essentials of Medical Geology*. Springer, 2013, pp 129-177.
30. R. P. Hong Enriquez, T. N. Do, Bioavailability of Metal Ions and Evolutionary Adaptation. *Life* 2012, *2*. 274-285.
31. I. Bertini, L. Decaria, A. Rosato, The annotation of full zinc proteomes. *Journal of Biological Inorganic Chemistry* 2010, *15*. 1071-1078.
32. J. S. C. Gilchrist, M. P. Czubryt, G. N. Pierce, Calcium and calcium-binding proteins in the nucleus. *Molecular and Cellular Biochemistry* 1994, *135*. 79-88.
33. L. S. Field, Y. Furukawa, T. V. O'Halloran, V. C. Culotta, Factors controlling the uptake of yeast copper/zinc superoxide dismutase into mitochondria. *Journal of Biological Chemistry* 2003, *278*. 28052-28059.
34. A. Atkinson, D. R. Winge, Metal acquisition and availability in the mitochondria. *Chemical Reviews* 2009, *109*. 4708-4721.
35. C. Andreini, L. Banci, I. Bertini, S. Elmi, A. Rosato, Non-heme iron through the three domains of life. *Proteins: Structure, Function, and Bioinformatics* 2007, *67*. 317-324.

36. K. Van Baelen, L. Dode, J. Vanoevelen, G. Callewaert, H. De Smedt, L. Missiaen, J. B. Parys, L. Raeymaekers, F. Wuytack, The Ca^{2+}/Mn^{2+} pumps in the Golgi apparatus. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* 2004, *1742*. 103-112.
37. F. Wolf, V. Trapani, Cell (patho) physiology of magnesium. *Clinical Science* 2008, *114*. 27-35.
38. R. H. Kretsinger, Calcium-binding proteins. *Annual Review of Biochemistry* 1976, *45*. 239-266.
39. R. Sitia, I. Braakman, Quality control in the endoplasmic reticulum protein factory. *Nature* 2003, *426*. 891-894.
40. W. Maret, Metalloproteomics, metalloproteomes, and the annotation of metalloproteins. *Metallomics* 2010, *2*. 117-25.
41. K. J. Waldron, J. C. Rutherford, D. Ford, N. J. Robinson, Metalloproteins and metal sensing. *Nature* 2009, *460*. 823-830.
42. D. L. Oliveira, E. S. Nakayasu, L. S. Joffe, A. J. Guimarães, T. J. P. Sobreira, J. D. Nosanchuk, R. J. B. Cordero, S. Frases, A. Casadevall, I. C. Almeida, Biogenesis of extracellular vesicles in yeast. *Communicative & Integrative Biology* 2010, *3*. 533-535.
43. R. J. P. Williams, The Biodistribution of Metal Ions. *Concepts and models in bioinorganic chemistry* 2006.
44. L. Banci, I. Bertini, in *Metallomics and the Cell*. Springer, 2013, pp 1-13.
45. C. L. Dupont, S. Yang, B. Palenik, P. E. Bourne, Modern proteomes contain putative imprints of ancient shifts in trace metal geochemistry. *Proceedings of the National Academy of Sciences* 2006, *103*. 17822-17827.
46. A. L. Zerkle, C. H. House, S. L. Brantley, Biogeochemical signatures through time as inferred from whole microbial genomes. *American Journal of Science* 2005, *305*. 467-502.
47. M. J. Berridge, P. Lipp, M. D. Bootman, The versatility and universality of calcium signalling. *Nature Reviews Molecular Cell Biology* 2000, *1*. 11-21.
48. K. Venkatachalam, D. B. van Rossum, R. L. Patterson, H.-T. Ma, D. L. Gill, The cellular and molecular basis of store-operated calcium entry. *Nature Cell Biology* 2002, *4*. E263-E272.

Figures

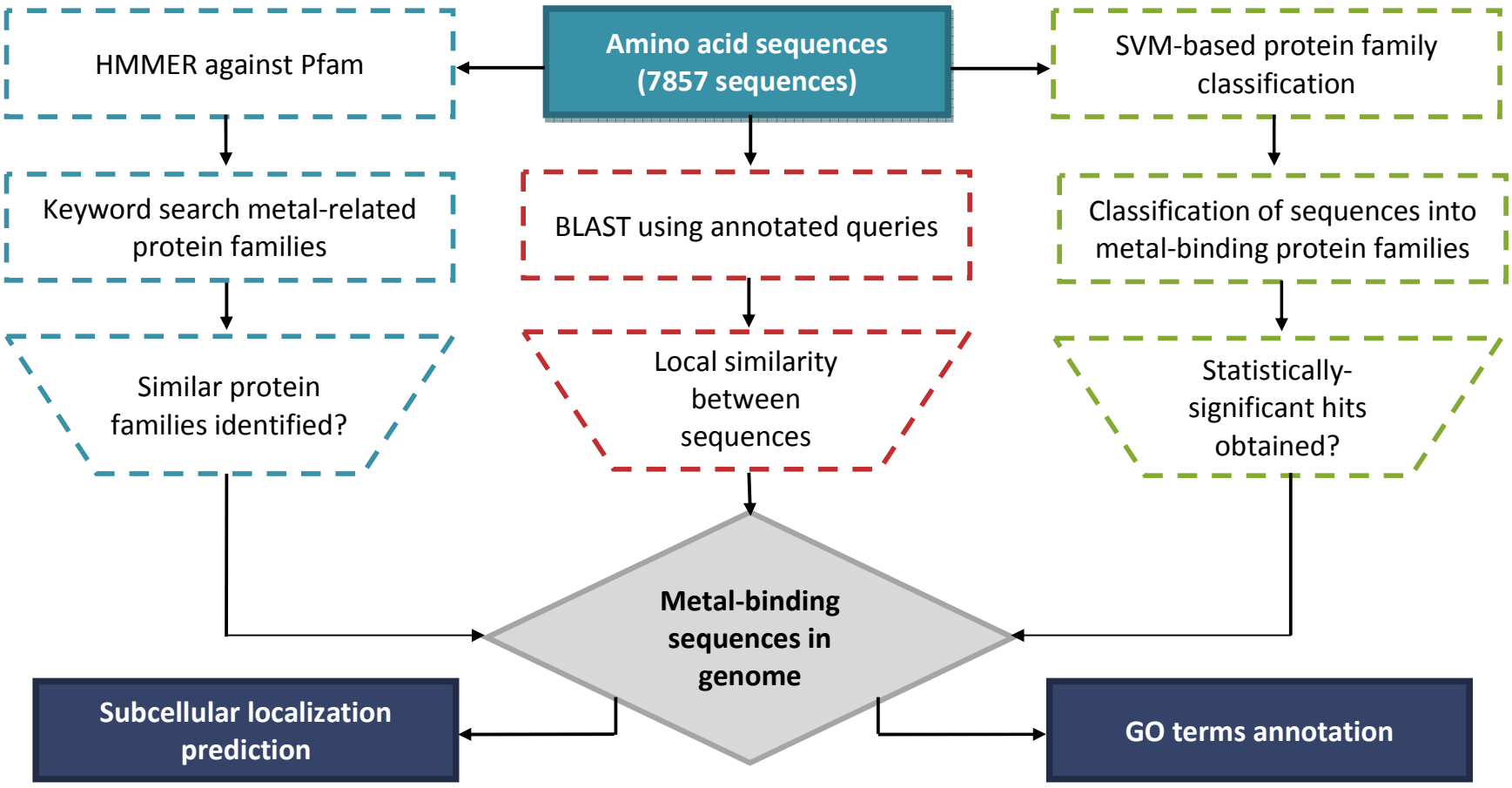


Figure 1. Overall workflow of project.

A flowchart illustrating the steps involved for the identification of putative metal-binding proteins from *G. antarctica* PI12 genome.

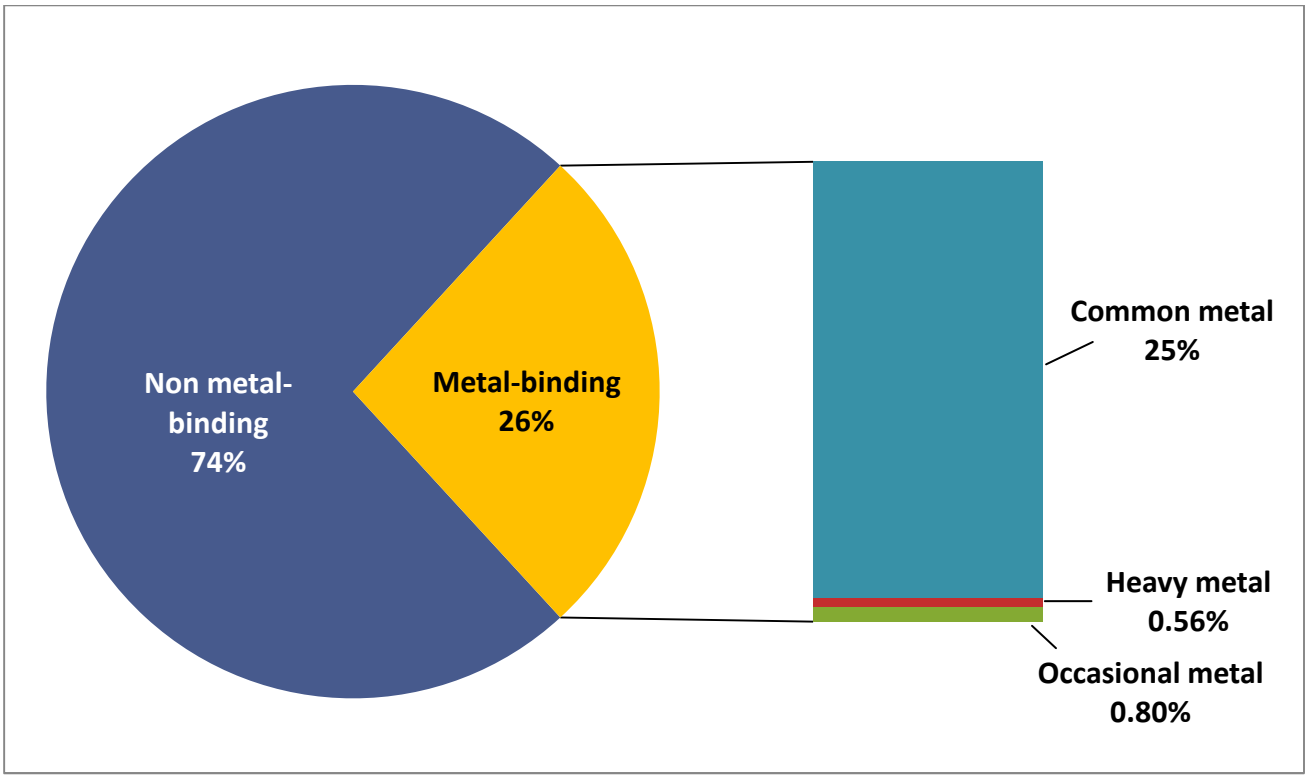
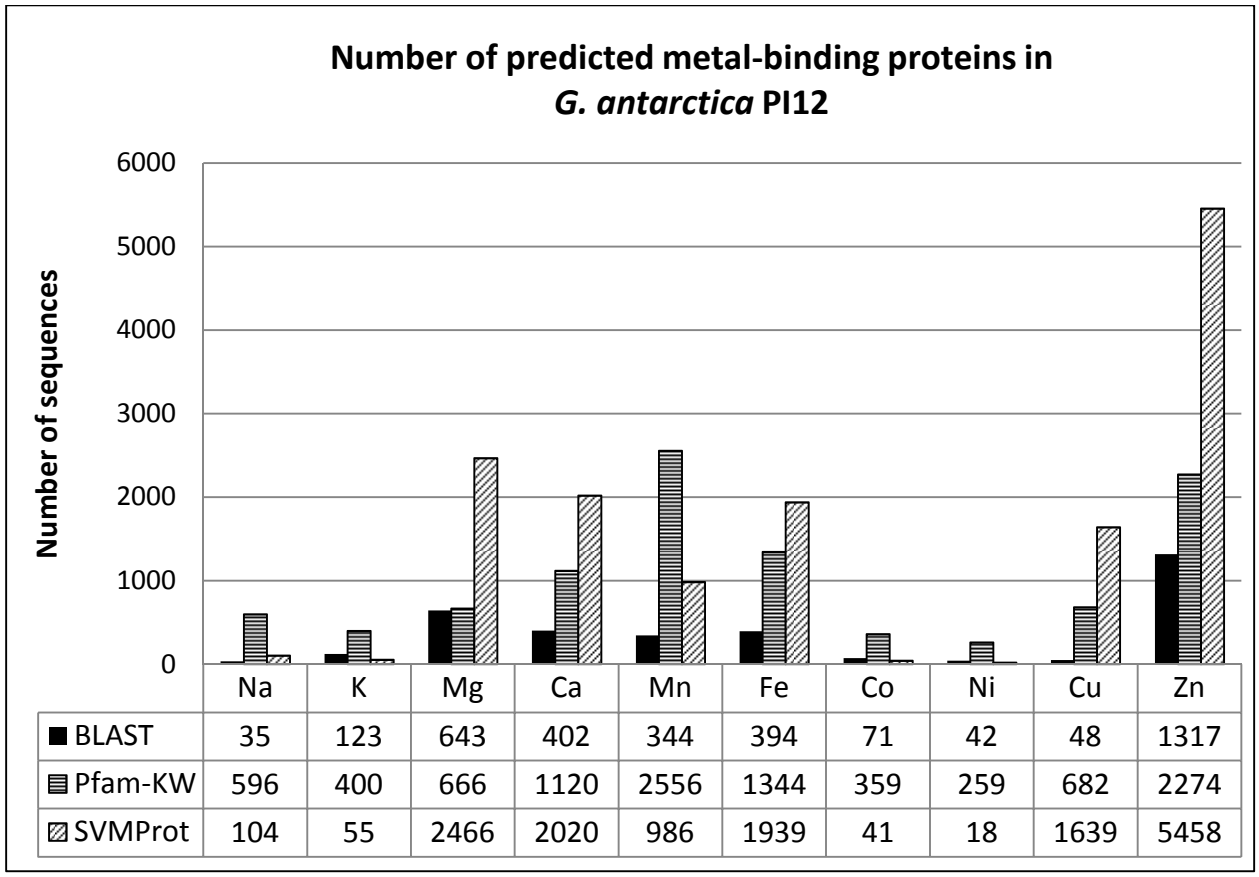


Figure 2. Relative abundance of metal-binding proteins in *G. antarctica* PI12 proteome.

The percentage of putative metal-binding proteins present in the proteome of *G. antarctica* PI12 analyzed using BLASTp. Common metals (ubiquitous to almost all living organisms) include *"Sodium-binding", "Potassium-binding", "Magnesium-binding", "Calcium-binding", "Manganese-binding", "Iron-binding", "Cobalt-binding", "Nickel-binding", "Copper-binding", "Zinc-binding"*; Occasional metals (rarely been found in biological systems but may crucial for particular organisms) include *"Lithium-binding", "Vanadium-binding", Molybdenum-binding", "Tungsten-binding"*; Heavy metals include *"Cadmium-binding", "Mercury-binding" and "Lead-binding"*. About one-quarter (2071 proteins) of the proteome is predicted to be metal-bounded, including 44 sequences which are likely to bind heavy metals.



	Na	K	Mg	Ca	Mn	Fe	Co	Ni	Cu	Zn
BLAST	35	123	643	402	344	394	71	42	48	1317
Pfam-KW	596	400	666	1120	2556	1344	359	259	682	2274
SVMProt	104	55	2466	2020	986	1939	41	18	1639	5458

Figure 3. Comparison for the number of putative metal-binding protein identified using BLAST, Pfam-Keyword (Pfam-KW) and SVMProt.

BLAST and SVMProt results have independently suggested that the *G. antarctica* PI12 proteome is zinc-enriched. However inspecting the domain composition with Pfam terms (denoted as 'Pfam-Keyword') revealed a greater number of manganese proteins compared to zinc.

Figure 4(A)

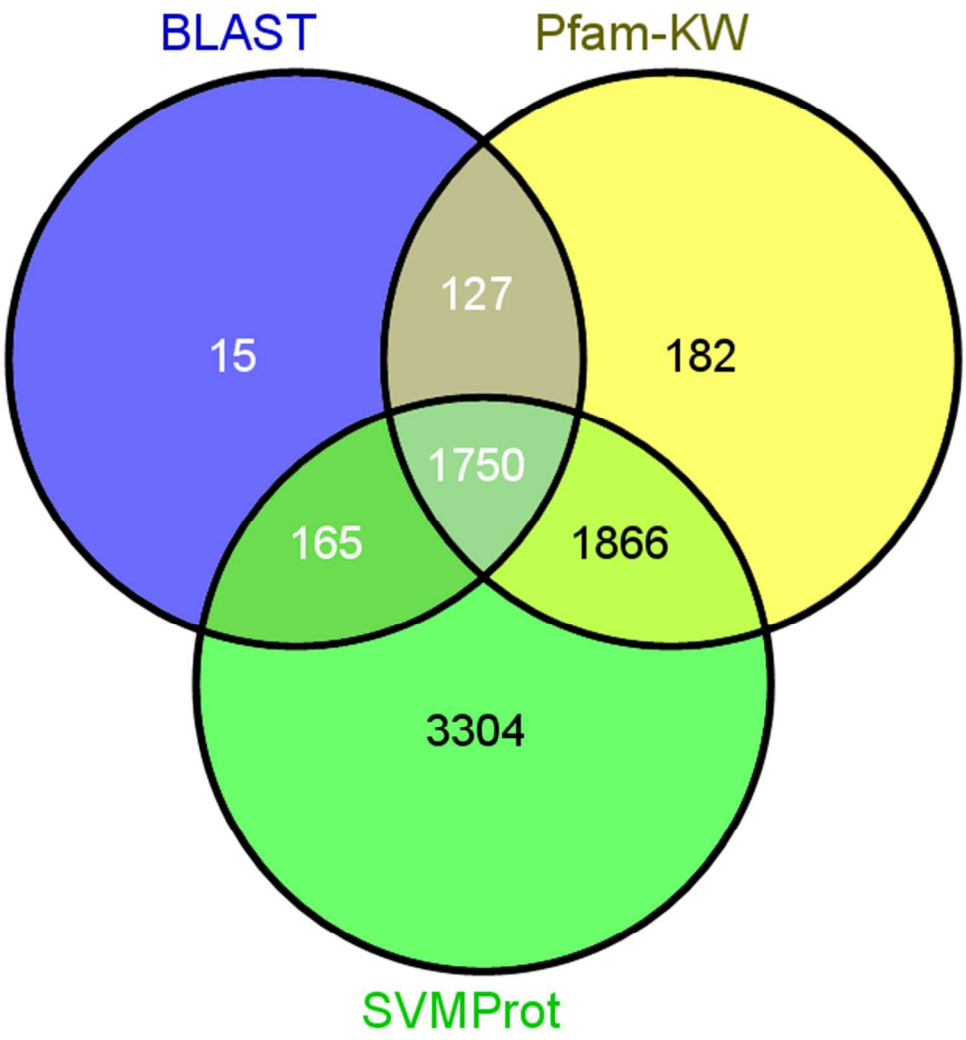
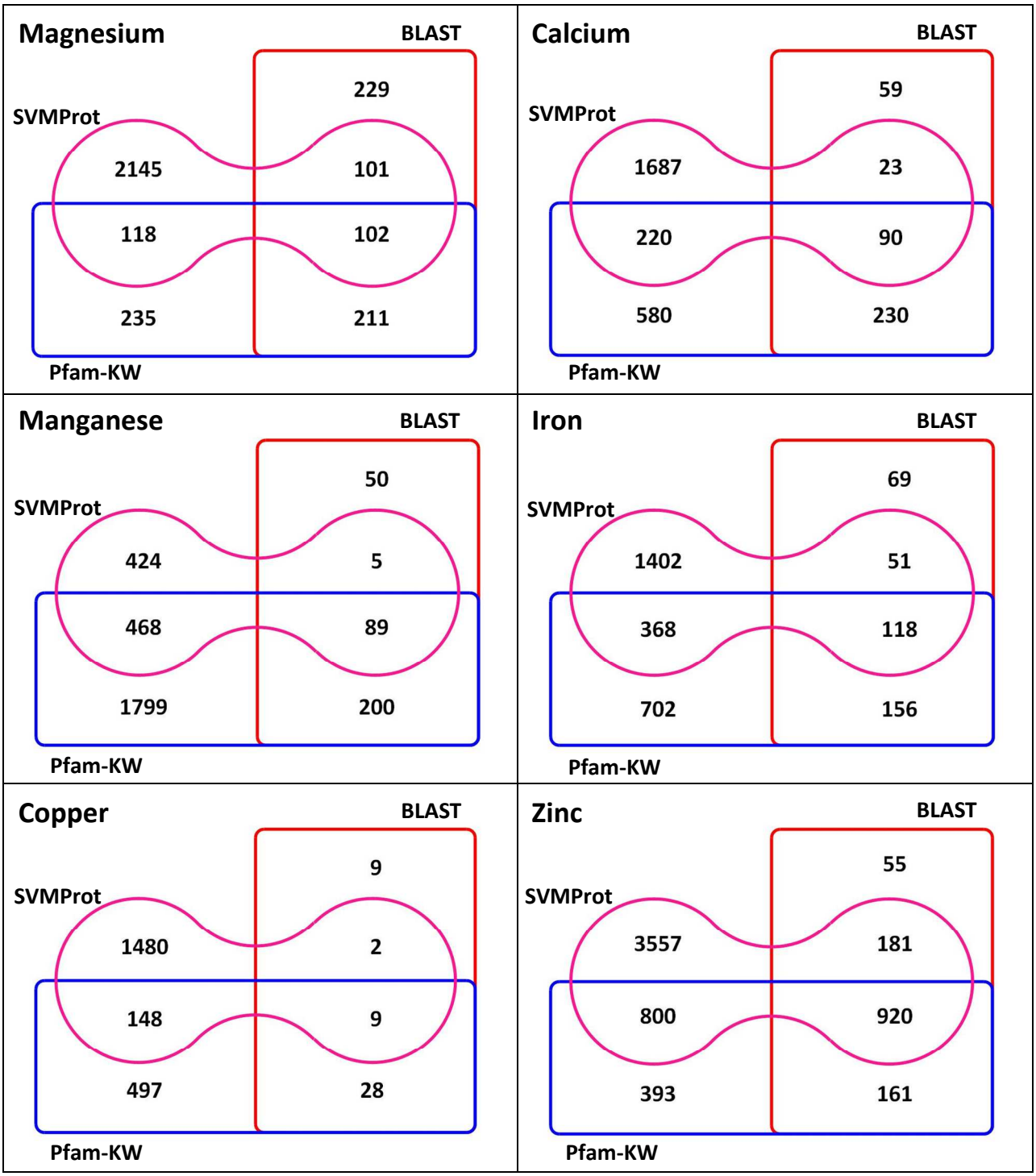


Figure 4. Comparison between the numbers of metal-binding proteins identified.

Venn diagram in Figure 4 (A) represents the overlaps of metal-binding proteins predicted by BLAST, Pfam-Keyword (Pfam-KW), and SVMProt. A total number of 1750 sequences in *G. antarctica* PI12 are predicted to be metal-bounded by all the approaches. The numbers of proteins identified for selected metal-types (magnesium-, calcium-, manganese-, iron-, copper- and zinc-binding) by the mentioned approaches are indicated in Figure 4 (B).

Figure 4(B)



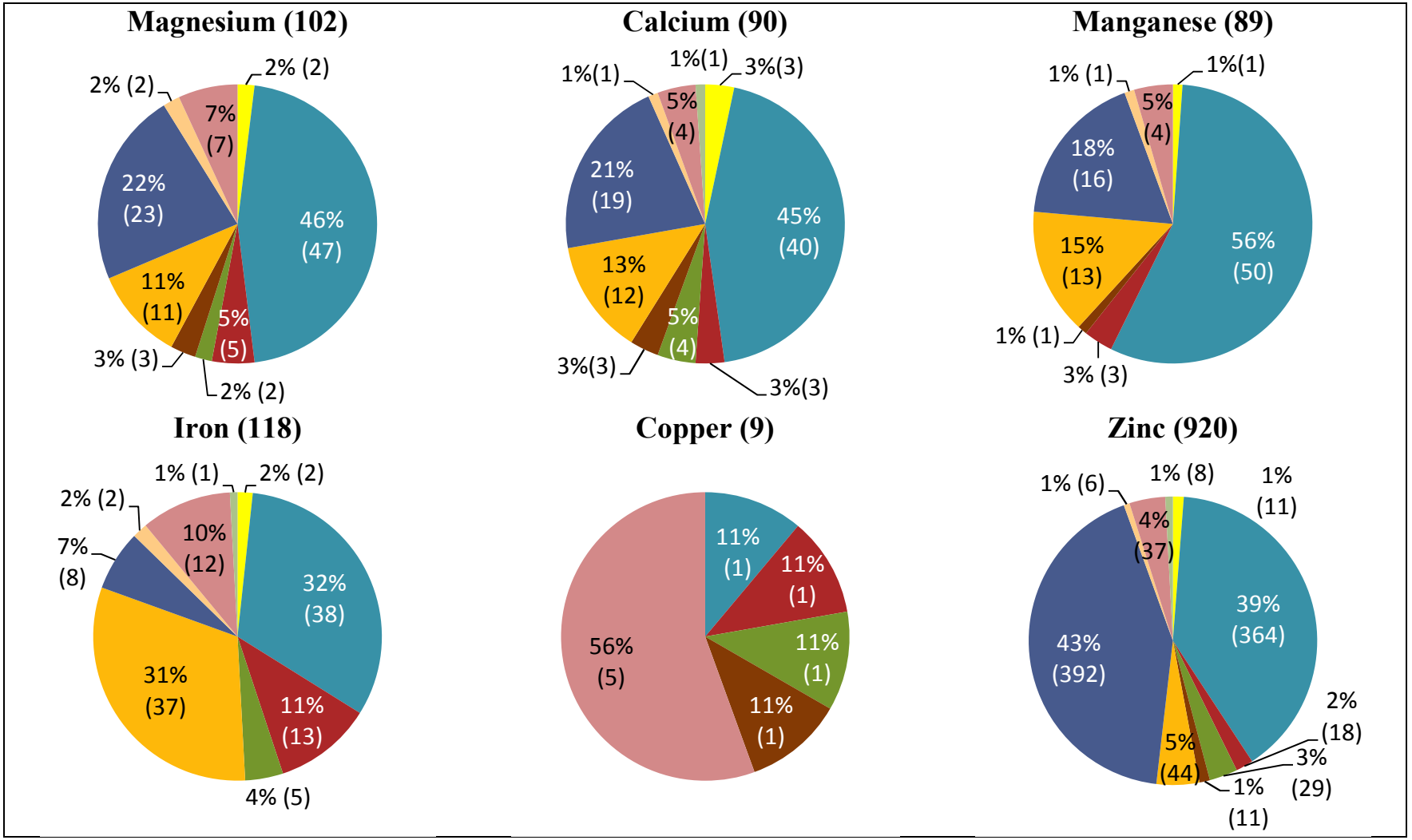


Figure 5. Distribution of the metal-binding proteins in various cellular compartments.

Pie diagrams indicate the relative abundance for the putative magnesium-, calcium-, manganese-, iron-, copper- and zinc-binding proteins distributed in the cellular compartments of *G. antarctica* PI12 . The total numbers of proteins identified (combination of all approaches) for each metal are indicated in parentheses.

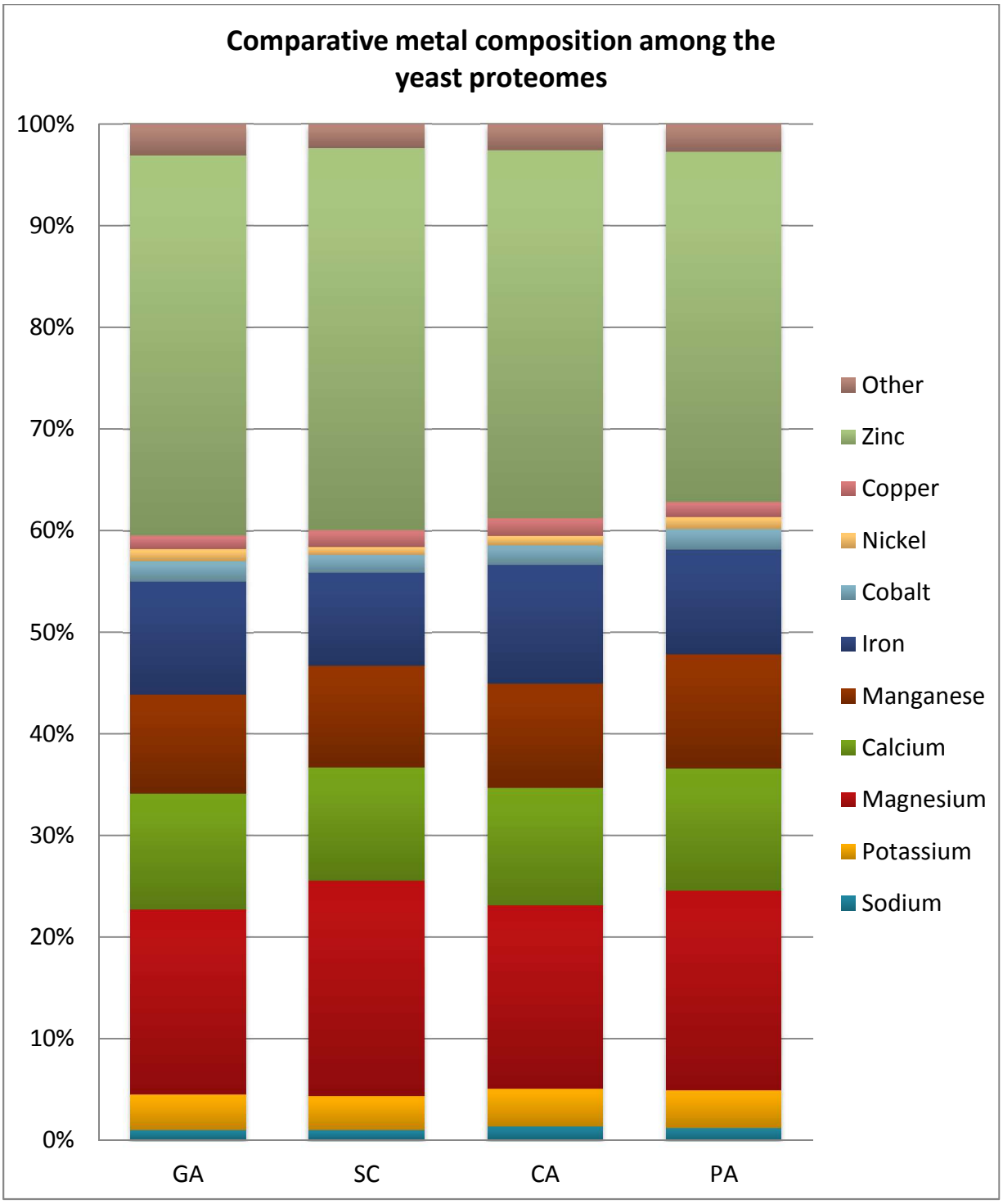


Figure 6. Comparison of metal composition between *G. antarctica* PI12 and warm-adapted yeast counterparts.

Relative abundance of the putative metal-binding proteins in the four yeast proteomes identified using BLASTp. Abbreviation: GA- *Glaciozyma antarctica* PI12; SC- *Saccharomyces cerevisiae* S288c; CA- *Candida albicans* SC5314; PA- *Pichia angusta* ATCC 26012.

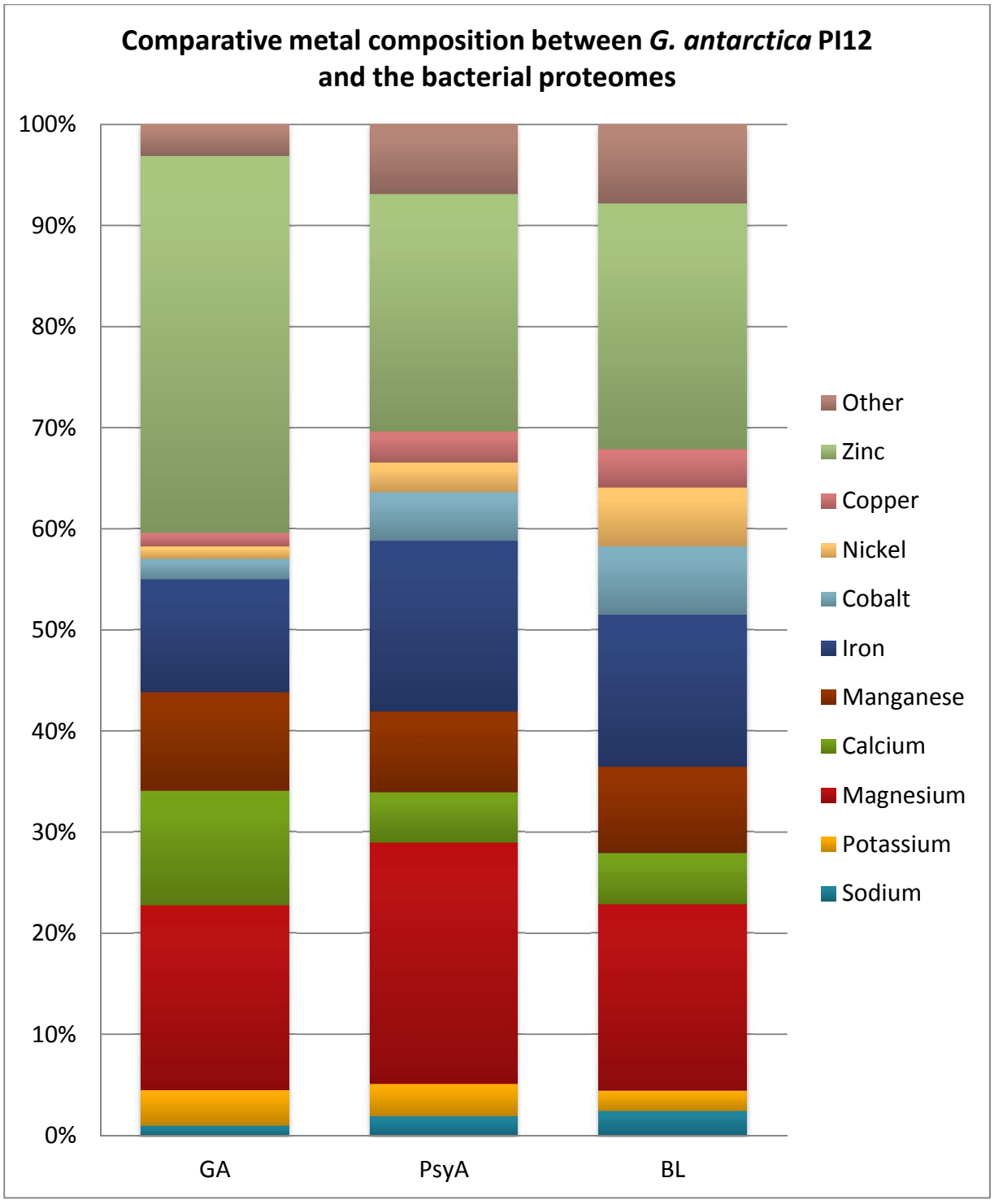


Figure 7. Comparison of metal composition between *G. antarctica* PI12 and the bacteria.

Relative abundance of the putative metal-binding proteins in the two bacterial proteomes identified using BLASTp. Abbreviation: GA- *Glaciozyma antarctica* PI12; PsyA- *Psychrobacter arcticus* 273-4; BL- *Bacillus lehensis* G1.

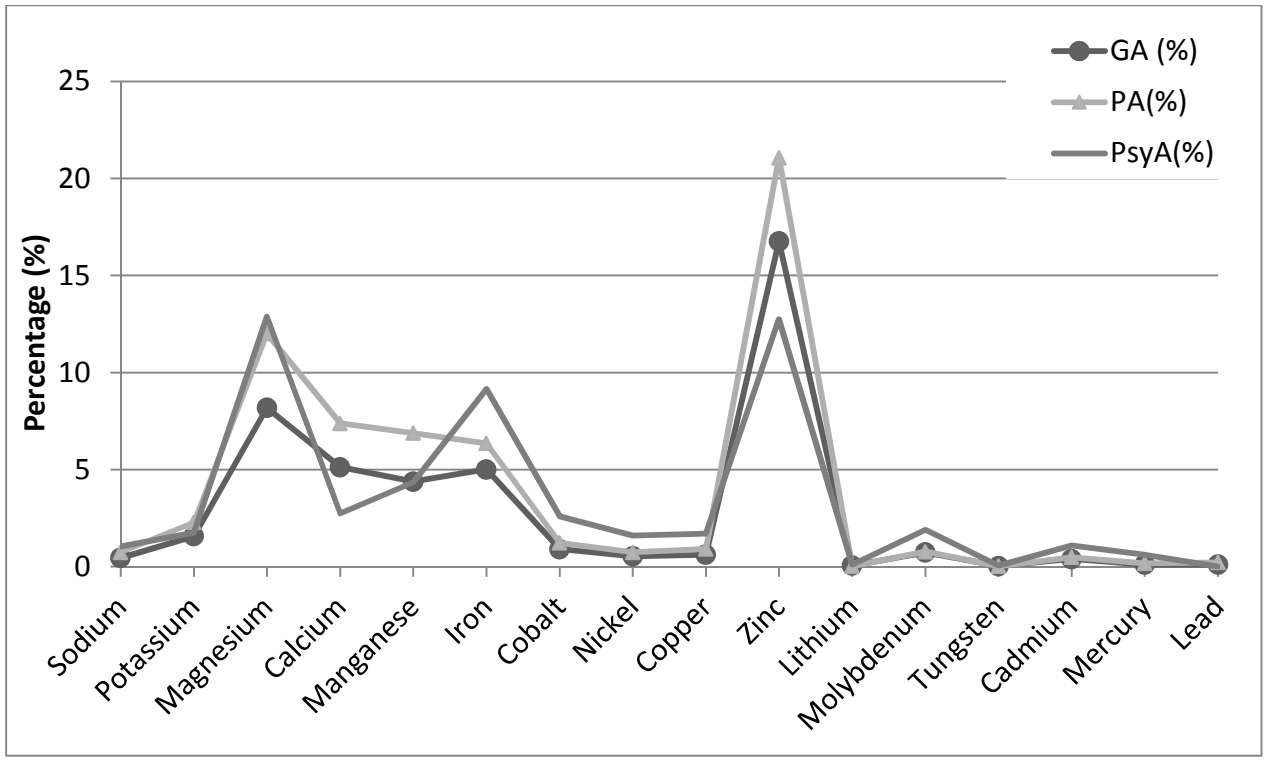


Figure 8. Variation of metal content between the psychrophiles and the thermophile in relative to their proteome size.

The fractional abundance (in %) of the putative metal-binding proteins for the psychrophiles *G. antarctica* PI12 (GA) and *P. arcticus* 273-4 (PsyA) are compared with the thermophilic yeast *P. angusta* ATCC 26012 (PA) in relative to their proteome size.