Limnology and Freshwater Biology 2019 (3): 240-244 DOI:10.31951/2658-3518-2019-A-3-240

Original Article

ISSN 2658-3518
LIMNOLOGY FRESHWATER BIOLOGY
www.limnolfwbiol.com

Duplications and subdomain shuffling in diatom metacaspases

Morozov A.A.*

Limnological Institute of the Siberian Branch of the Russian Academy of Sciences, Ulan-Batorskaya 3, 664 033 Irkutsk, Russia.

ABSTRACT. A novel kind of type III metacaspases was described based on published diatom genomes and transcriptomes. These sequences feature multiple copies of the metacaspase domain and, according to the phylogenetic analysis, have been independently created by intragenic duplications in multiple species. Although the phylogeny of diatom metacaspases and its relationships with functional diversity of these enzymes are currently poorly resolved, multiplicated metacaspases are present in most, if not all, major clades of this subfamily.

Keywords: Diatoms, metacaspases, intragenic multiplications

1. Introduction

Diatoms are an essential component of marine and freshwater ecosystems, including Lake Baikal. Although from the ecological point of view they are mostly important as the primary producers, accounting for as much as 40% of the Ocean's production, the most famous feature of these unicellular algae is their ability to build complex species-specific cell walls (called frustules) out of silica. This ability has raised obvious interest because if its potential nanotechnological applications, leading to numerous studies of the processes of silicon assimilation, transport and polymerization. Although the complete pathway has not been described so far, some of its elements are known.

One of these elements is SIT, a silicon transporter protein responsible for the transmembrane import of silicon. It was first described in *Cylindrotheca fusiformis* and its function was confirmed by expression in *Xenopus laevis* oocytes (Hildebrand et al., 1997). SIT homologs were later shown to be ubiquitous in diatoms. The proteins of this family typically contain a single SIT domain featuring one conserved CMLD motif and four conserved GxQ motifs. A later work on a baikalian diatom *Synedra acus* subsp. *radians* (=*Fragillaria radians*) has detected a family of genes and transcripts encoding multiple copies of the SIT domain within a single reading frame. Each of these domain copies contained all the conservative motifs and is presumably functional (Marchenkov et al., 2016). Similar multi-SIT genes were later found in other species of the same genus (Marchenkov et al., 2018) and among the predicted genes from numerous diatom genomic and transcriptomic projects (Durkin et al., 2016), thus confirming that these sequences are not just an artifact, nor a pecularity of this particular species. It's also important to note that multiplications are present in only one of the SIT subfamilies (Durkin et al., 2016).

On the other hand, staining *S. acus* subsp. *radians* total protein with SIT antibodies has detected only a single product whose mass roughly corresponds to a protein with a single SIT domain (Petrova et al., 2007). This, and the presence of conserved aspartate-rich motifs near the domain boundaries (Marchenkov et al., 2018), implies the existence of some proteolytic activation mechanism which acts co-translationally (or very soon after the translation) to cleave multi-SIT precursor into multiple mature SIT proteins. If this is true, it would be reasonable to assume that this mechanism has other targets as well. Although proteolytic activation is widespread in eukaryotes, the production of proteins from multi-copy precursors is usually limited to relatively short peptides like ubiquitin, not large transmembrane transporters. However, diatoms are not a thoroughly studied model object like *H. sapiens* or *S. cerevisiae*. It is not impossible for them to contain some novel regulation mechanisms, and proteolytic processing of multi-SITs can be one of these. To see whether there are other recent duplications, the search was performed on complete sets of predicted proteins from diatom genomes and transcriptomes, returning tens to hundreds of genes per species (Morozov, 2017).

Among the proteins found to be duplicated in multiple diatom species were type III metacaspases. These proteases are related to caspases, although they differ from them in domain organization (p10 subdomain is N-terminal, unlike other classes of caspases and metacaspases where it's C-terminal) and

*Corresponding author.
E-mail address: alexeymorozov1991@gmail.com (Morozov A.A.)

possibly in functions (Choi and Berges, 2013). The exact role of these proteins is unknown: metacaspases were shown to be involved in programmed cell death, like caspases (Thamatrakoln et al., 2012; van Creveld et al., 2019), but they were also expressed outside the cell death conditions suggesting some other roles in cell metabolism, possibly in stress response (Bidle, 2015; Wang et al., 2017). Even their substrate specificity is unclear: on one hand, plant homologs of diatom metacaspases have 20- to 50-fold higher specificity for arginine, rather then aspartate (Bozhkov et al., 2005), and (auto-)cleavage following arginine was experimentally shown for one of the *Phaeodactylum tricornutum* metacaspases. On the other hand, *Skeletonema marinoi* has shown elevated caspase-like activity, as measured by the cleavage of aspartate-containing FITC-v-VAD-FMK, during PCD. Although involvement of metacaspases in aspartate-targeted cleavage was not shown directly, this species (like other diatoms) does not have any caspase-related proteins besides type III metacaspases.

The reason that these proteins are relevant for the study of multiplicated proteins is that caspases are activated by proteolytic cleavage between p10 and p20 domains and proteolytic removal of terminal regulatory elements (Klemencic and Funk, 2018). Therefore their function may not be disrupted by multiplication; more than that, metacaspases may be responsible for the processing of other multiplicated proteins. Their role in SIT processing is questionable because of the aforementioned aspartate-vs-arginine specificity issue, but there is no reason to assume that they wouldn't be able to cleave other multiplicated proteins or at least themselves.

Thus, the goal of this work was to identify multi-metacaspases in diatom algae and study their evolution.

2. Materials and methods

To find metacaspases, a complete collection of predicted proteins from published diatom genomes and transcriptomes (Armbrust et al., 2004; Bowler et al., 2008; Galachyants et al., 2015; Tanaka et al., 2015; Mock et al., 2017) was scanned with hmmsearch (Eddy, 2011) using the PFAM hidden Markov Model of a caspase proteolytic domain (PF00656) as a query. Since this model includes both p20 and p10 subdomains, the subdomain coordinates were determined manually.

The dataset for the phylogenetic analysis included all multiplicated metacaspases, all other metacaspases from the species where these were found, all metacaspases from model diatoms *Thalassiosira pseudonana*, *Phaeodactylum tricornutum* and *Pseudo-nitzschia multiseries*, and all metacaspases from *Skeletonema marinoi*, a species in which metacaspase expression was previously studied (Wang et al., 2017). The complete list of sequence IDs is available in Suppl. 1. Multiplicated metacaspase sequences were split into p10-p20 pairs manually based on HMM alignment coordinates. Resulting sequence set was aligned using clustalo (Sievers et al., 2011) and truncated manually. Maximum likelihood phylogenetic analysis was performed with RAxML (Stamatakis, 2014) using Lee-Gascuel substitution model and gamma rate distribution. To search for potential incongruence between p10 and p20 subunit evolution, the sequences were split programmatically based on HMM alignment coordinates, and then analyzed by a similar pipeline.

3. Results and Discussion

In total, 13 duplicated metacaspases were detected in 7 species (Fig. 1). Not all of them are

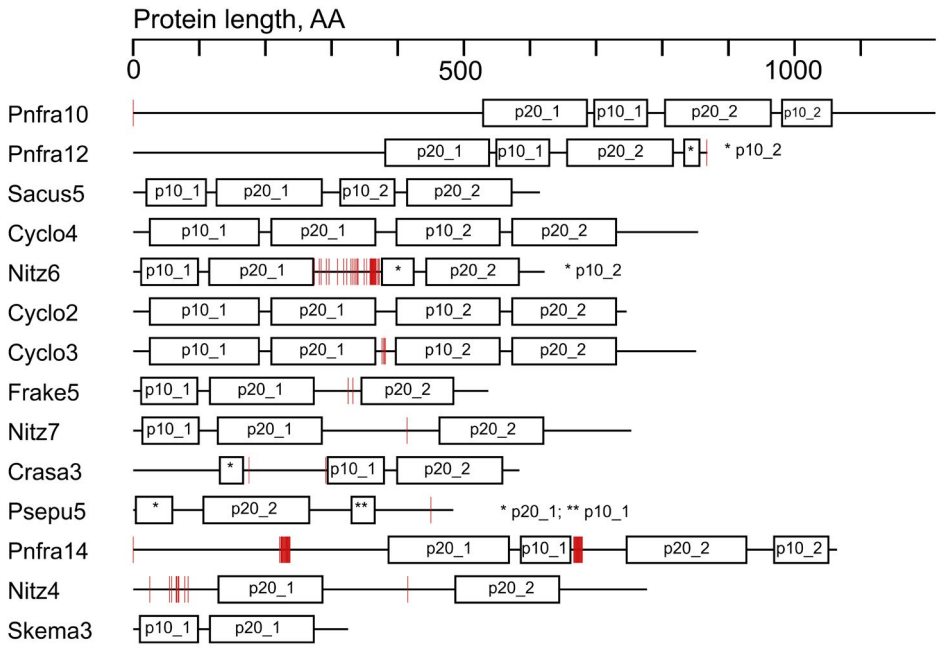

Fig.1. Domain structures of diatom multi-metacaspases. Red lines mark ambiguous aminoacids. A single non-multiplicated type III metacaspase (Skema3) is included for the comparison.

completely duplicated sequences with two p10-p20 subdomain pairs. Some of them merely lack a part of one of the domains, e.g. the first p10 subdomain (and a part of the first p20) in Crasa3. Most likely these sequences are just incomplete, which is possible considering a relatively low coverage of assemblies produced by the MMETSP project (Keeling et al., 2014). Perhaps more interesting is the fact that some sequences from diatoms of the genus *Pseudo-nitzschia* appear to have a p20-p10 order of subdomains similar to metazoan caspases. Hypothetically, they may have undergone a domain rearrangement event that effectively reversed the rearrangement at the root of type III metacaspases.

If these sequences are indeed correct, it would suggest that metacaspases (at least in diatoms) are robust to changes in domain compositions, whether these changes involve copying subdomains or changing their order. However, it's impossible to guarantee that the predicted multi-caspase proteins did not arise as a consequence of some assembly artifact. The chief argument towards this hypothesis is that some of them contain stretches of ambiguous aminoacids near domain boundaries. On the other hand, most of the sequences do not. In addition, the design of MMETSP project involved sequencing and assembly of numerous marine eukaryotic species using exactly the same set of methods, so we could expect to see such artifacts randomly distributed across most diatoms, rather than in multiple subfamilies of the same small set of species. It should also be noted that most of the multiplicated sequences appear functional, retaining the catalytic cysteine residues and not showing significant gaps in the alignment. Finally, it is obviously possible that some of these sequence are valid and others are assembly artifacts.

To see whether these multiplicated metacaspases all fall within a single subfamily like multi-SITs (Durkin et al., 2016), a phylogenetic analysis was performed. The dataset included all multi-metacaspases, all other metacaspase sequences from the same species, all metacaspases from model species *Th. pseudonana* and *Ph. Cylindrus,* and all sequences for which any additional experimental information was available. Multiplicated sequences were manually split into single-domain fragments prior to the alignment. A consensus phylogenetic network of bootstrap replicates for this dataset is shown at Figure 2.

As this network shows, multiplicated type III metacaspases have arisen independently in all species that have them, sometimes more than once per species. To see if the history of multicaspase evolution features some sort of events more complex than the duplication of ancestral sequences (e.g. the exchange of subdomains between paralogs), separate trees were built for p10 and p20 subunits (Suppl. 2 and 3). Although these trees are poorly resolved, they are generally congruent to the network on Fig. 2, indicating that multi-metacaspase origin events are indeed just a series of independent duplications. Unlike multi-SIT, these events are not restricted to one of the subfamilies.

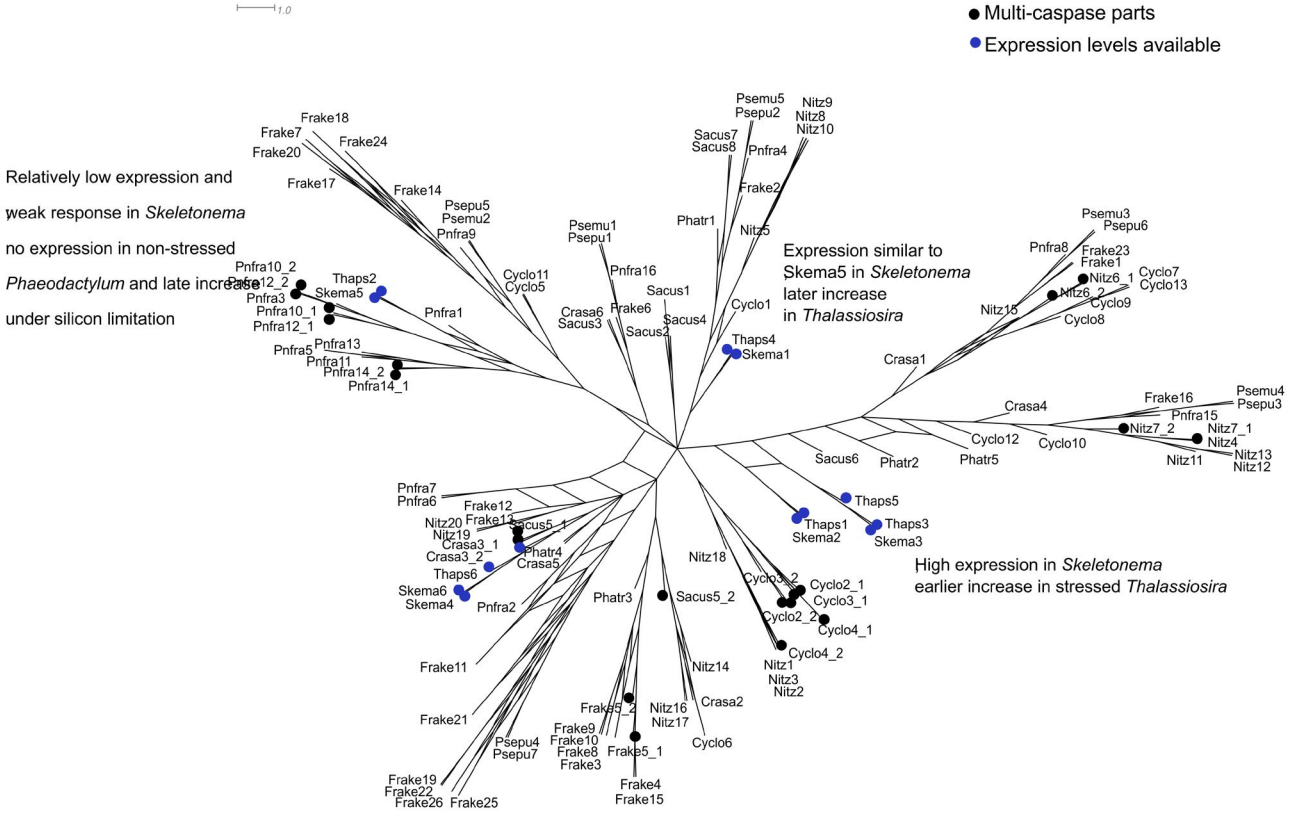

Fig.2. The consensus phylogenetic network of ML bootstrap replicates for diatom metacaspases. Edge length correspond to bootstrap supports of each split; splits with less than 0.33 bootstrap support are not shown. Expression levels under various conditions shown for the sequences for which they are available in literature (blue dots). Multiplicated sequences were split into subsequences

However, diatom metacaspase subfamilies are generally very different to define. For example, although expression patterns of phylogenetically close Skema2/Skema3 and Thaps5/Thaps3 are similar within each species, they differ between *Skeletonema marinoi* and *Thalassiosira pseudonana*. These differences can be explained by a difference in experimental conditions: one species was limited by silicon (Wang et al., 2017) and the other was lacking iron (Bidle and Bender, 2008). Although both conditions were shown to induce programmed cell death, as evidenced by microscopy and direct measurements of caspase-like activity, it's plausible that two different signaling pathways will be employed for two different conditions. But even within each experiment, different expression patterns were found for Skema6/Skema4 and Thaps 5/Thaps1. In each of these two pairs, the sequences form a well-supported clade on the phylogenetic network not shared with any other species. Even if these sequences have diverged very recently, they do not seem to retain a shared ancestral function.

Notably, a similar incongruence between phylogeny and structural features was found for the 2-Cys metacaspases that have a conserved regulatory cysteine residue. This feature was found in numerous metacaspases distributed approximately evenly along the tree (van Creveld et al., 2019). Just like with expression levels, this is clearly a functionally significant feature, but it's unclear how (and if at all) its presence corresponds to the sequences' evolution.

4. Conclusions

A previously undescribed multiplicated metacaspases were found in diatom genomes and proteomes. They are present in most of the clades of the metacaspase tree, although currently there is no consensus regarding the connection between phylogeny of type III metacaspases and their functional diversity. Although it's possible that at least some of these sequences are assembly artifacts, circumstantial evidence points to their existence and functionality. Establishing the functional specificity, if any, of these sequences requires further experimental evidence.

Acknowledgements

The reported study was funded by RFBR according to the research project № 18-34-00441 "Verification and analysis of intragenic duplications in diatom algae".

References

Armbrust E.V., Berges G., Bowler C. et al. 2004. The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. Science 306: 79–86. DOI: 10.1126/science.1101156

Bidle K.D., Bender S.J. 2008. Iron starvation and culture age activate metacaspases and programmed cell death in the marine diatom *Thalassiosira pseudonana*. Eukaryotic Cell 7: 223–236. DOI: 10.1128/EC.00296-07

Bidle K.D. 2015. The molecular ecophysiology of programmed cell death in marine phytoplankton. Annual Review of Marine Science 7: 1–35. DOI: 10.1146/annurev-marine-010213-135014

Bowler C., Allen A.E., Badger J.H. et al. 2008. The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. Nature 456: 239–244. DOI: 10.1038/nature07410

Bozhkov P.V., Suarez M.F., Filonova L.H. et al. 2005. Cystein protease mcII-Pa executes programmed cell death during plant embryogenesis. Proceedings of the National Academy of Sciences 102: 14463–14468. DOI: 10.1073/pnas.0506948102

Choi C.J., Berges J.A. 2013. New types of metacaspases in phytoplankton reveal diverse origins of cell death proteases. Cell Death and Disease 4. DOI: 10.1038/cddis.2013.21

Durkin C.A., Mock Th., Armbrust E.V. 2016. The evolution of silicon transporters in diatoms. Journal of Phycology 52: 716–731. DOI: 10.1111/jpy.12441

Eddy S.R. 2011. Accelerated profile HMM searches. PLOS Computational Biology 7. DOI: 10.1371/journal.pcbi.1002195

Galachyants Yu.P., Zakharova Yu.R., Petrova D.P. et al. 2015. Sequencing of a complete genome of an araphid pennate diatom *Synedra acus* subsp. *radians* from Lake Baikal. Doklady Biochemistry and Biophysics 461: 348–352. DOI: 10.1134/S1607672915020064

Hildebrand M., Volkani B.E., Gassmann W. et al. 1997. A gene family of silicon transporters. Nature 385: 688–689. DOI: 10.1038/385688b0

Keeling P.J., Burki F., Wilcox H.M. et al. 2014. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. PLOS Biology 12. DOI: 10.1371/journal.pbio.1001889

Klemencic M., Funk C. 2018. Structural and functional diversity of caspase homologues in non-metazoan organisms. Protoplasma 255: 387–397. DOI: 10.10007/s00709-017-1145-5

Marchenkov A.M., Petrova D.P., Khabudaev K.V. et al. 2016. Unique configuration of genes of silicon transporter in the freshwater pennate diatom *Synedra acus* subsp. *radians*. Doklady Biochemistry and Biophysics 471: 407–409. DOI: 10.1134/S1607672916060089

Marchenkov A.M., Petrova D.P., Morozov A.A. et al. 2018. A family of silicon transporter structural genes in a pennate diatom *Synedra ulna* subsp. *danica* (Kütz) Skabitsch. PLOS One 13. DOI: 10.1371/journal.pone.0203161

Mock Th., Otillar R.P., Strauss J. et al. 2017. Evolutionary genomics of the cold-adapted diatom *Fragillariopsis cylindrus*. Nature 541: 536–540. DOI: 10.1038/nature20803

Morozov A. 2017. Intragenic multiplications in diatom protein-coding genes. 2017. In: Bioinformatics: from algorithms to applications, pp. 48–49.

Petrova D.P., Bedoshvili Ye.D., Korneva E.S. et al. 2007. Detection of the silicic acid transport protein in the freshwater diatom *Synedra acus* by immunoblotting and immunoelectron microscopy. Doklady Biochemistry and Biophysics 417: 295–298. DOI: 10.1134/S1607672907060014

Sievers F., Wilm A., Gibson T.J. et al. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology 7: 539. DOI: 10.1038/msb.2011.75

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30: 1312–1313. DOI: 10.1093/bioinformatics/btu033

Tanaka T., Maeda Y., Veluchamy A. et al. 2015. Oil accumulation by the oleaginous diatom *Fistulifera solaris* as

revealed by the genome and transcriptome. The Plant Cell 27: 162–176. DOI: 10.1105/tpc.114.135194

Thamatrakoln K., Korenovska O., Niheu A.K. et al. 2012. Whole-genome expression analysis reveals a role for death-related genes in stress acclimation of the diatom Thalassiosira pseudonana. Environmental Microbiology 14: 67–81. DOI: 10.1111/j.1462-2920.2011.02468.x

van Creveld S.G., Ben-Dor S., Mizrachi A. et al. 2019. A redox-regulated type III metacaspase controls cell death in a marine diatom. BioRxiv. DOI: 10.1101/444109

Wang H., Mi T., Zhen Yu. et al. 2017. Metacaspases and programmed cell death in *Skeletonema marinoi* in response to silicate limitation. Journal of Plankton Research 39: 729–743. DOI: 10.1093/plankt/jbw090