Move slowly, live long: tardigrades secret proteins

Bogaichuk Polya[1] and Zhurakovskaya Anya[1]

[1]Bioinformatics Institute

Abstract

It is amazing how small tardigrades are able to resist adverse conditions. To find out which tardigrade proteins are involved in protection against damage, we selected DNA-related proteins localized in the nucleus. We can find out their function by comparing the protein sequence with known proteins in other organisms. We found a previously discovered damage suppressor protein and three potential candidate proteins requiring more research: E3 ubiquitin-protein ligase BRE1B, Inositol monophosphatase 3, and Eukaryotic translation initiation factor 3 subunit A.

Introduction

Tardigrades are tiny invertebrate animals closely related to Arthropoda phylum. They attract the attention of scientists because of their high survivability even under extreme conditions [1]. It is known that tardigrades are incredibly resistant to a wide range of environmental factors when they enter a dehydrated state called anhydrobiosis [2]. But what exactly are the molecular mechanisms underlying their resistance?

The answer to this question can be found in the genome of tardigrades, which was first sequenced in 2015 [3]. It was found that some part of the tardigrade's genome has been horizontally transferred [4]. However, the secret of the tardigrade's success lies in its unique proteins that protect its DNA from damage.

In our work, we will 1) look at the tardigrade's genome and unique protein-coding genes, 2) compare protein sequence homology with other organisms 3) and prioritize candidate proteins.

Methods

Sequence Acquisition

First, we obtained functional annotation of *Ramazzottius varieornatus* genome by running Augustus v3.3.3 with *–species=nasonia* as a parameter [5]. To select proteins physically located in the nucleus tandem mass spectrometry was performed on proteins associated with chromatin. Obtained protein sequences were aligned to *R. varieornatus* genome by running BLAST v2.9.0-2 locally with outfmt6 parameter [6]. For each searched protein best hits were selected and whole sequences of those proteins were extracted from genome with samtools software [7]. After these steps we acquired 34 protein sequences.

Localisation prediction

To predict subcellular localization of proteins we used WoLF PSORT and TargetP 2.0 web-servers and based on the results selected 12 proteins that have major presence in nucleus [8, 9].

To understand protein relations and function we performed BLAST web-server search against UniProtKB/Swiss-Prot databases and HMMER v3.3.2 web-version search against Pfam databases [10, 11]

Results

The present study aims to identify proteins that protect tardigrade DNA from damage. We examined the *R. varieornatus* genome to identify unique proteins and assess their possible function.

In total, we found 16435 protein-coding genes in the tardigrade genome (FIGURE). We compared the sequences of these proteins and the sequences of peptides physically located next to DNA. Only 34 of them are potentially localized in the chromatin.

Using predictors of subcellular localization of WoLF and TargetP proteins, we selected 12 proteins located in the nucleus and made sure that there were no secreted proteins among them (Table 1). All further studies we conducted only with these proteins.

One of the proteins turned out to be similar to the damage suppressor protein previously identified in *Ramazzottius varieornatus*. Some nuclear proteins of tardigrade were also found to be similar to those of *Rattus norvegicus, Danio rerio, Xenopus laevis, Caenorhabditis elegans* and *Mus musculus*. But several proteins showed no sequence similarity to any proteins or motifs in the BLASTP search. For some proteins, the potential function could also be predicted using the Pfam database.

The data obtained on the subcellular localization of proteins, homology with proteins of other organisms, and predicted function allowed us to identify the following proteins:

- g11960 (E3 ubiquitin-protein ligase BRE1B)
- g8100 (Inositol monophosphatase 3)
- g16318 (Eukaryotic translation initiation factor 3 subunit A)
- g14472 (Damage suppressor protein)

Compiled on: December 16, 2022.
Draft manuscript prepared by the author.

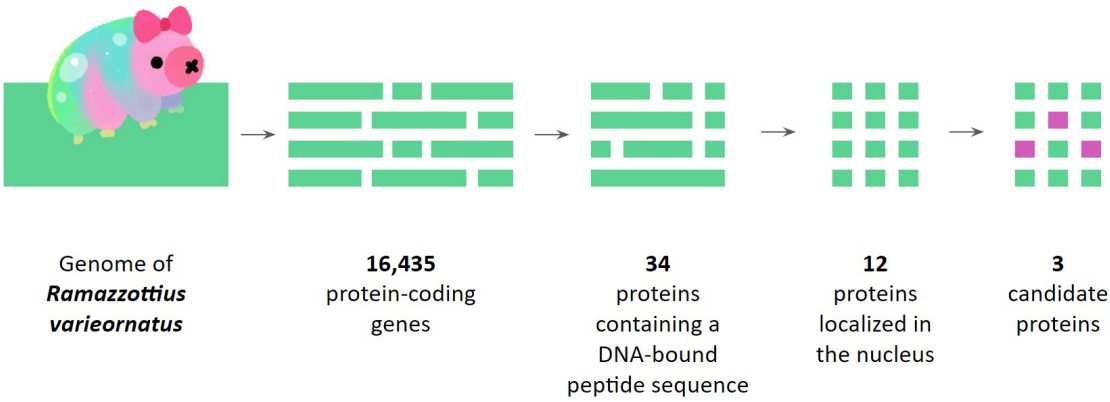

Figure 1. The process of selection of candidate proteins that contribute to the survival of tardigrade.

Table 1. Integrated table

Protein	Best blast hit	Predicted Pfam domain	Probable localization (WoLF PSORT)	Probable localization (TargetP)
g10513	None	None	Nuclear	Other
g10514	None	None	Nuclear	Other
g11806	None	None	Nuclear	Other
g11960	E3 ubiquitin-protein ligase BRE1B	Zinc finger (C3HC4)	Nuclear	Other
g14472	Damage suppressor protein	None	Nuclear	Other
g15484	Vacuolar protein sorting-associated protein 51 homolog	Vps51	Nuclear	Other
g16318	Eukaryotic translation initiation factor 3 subunit A		Nuclear	Other
g16368	Eukaryotic translation initiation factor 3 subunit A		Nuclear	Other
g5927	Glucosamine 6-phosphate N-acetyltransferase		Nuclear	Other
g7861	Inositol monophosphatase 2	SNF2-related domain; HARP	Nuclear	Other
g8100	Inositol monophosphatase 3	Inositol family; Arf6-interacting MKLP1-like protein	Nuclear	Other
g8312	Vacuolar protein sorting-associated protein 41	Region in clathrin	Nuclear	Other

Discussion

Collected data suggest that we found several potential proteins involved in DNA repair and stress response processes in *R. varieornatus* tardigrade. Along with already described Damage supressor protein (Dsup) we hypothesise here about the role of other proteins in damage response in tardigrades [12].

One of the found proteins was identified as an ortholog to E3 ubiquitin ligase (BRE1B) the type of protein that catalyzes the transfer of ubiquitin to the target protein and, what is more interesting in our case, mediates different DNA repair pathways, such as repair of double-strand breaks, nucleotide exclusions and base exclusion all of which are happening in extreme conditions that tardigrades can endure [13]. Moreover, BRE1B is involved in response to ionizing radiation as it has been shown that the knockdown of BRE1 gene in mouse cells increases radiosensitivity [14]

Another protein that stands out is a protein identified as ortholog to Inositol monophosphatase 2 protein. According to Pfam prediction this protein has SNF-2-related domain and HARP domain. HARP domain is especially relevant in our study as it was shown in several studies that it plays important role in DNA repair by stabilizing replication fork facilitating DNA-repair [15]. And, again, the knockdown of this gene leads to replication fork instability and high sensitivity to DNA-damaging agents [16, 17].

Third protein that we want to highlight in our study is a protein that was identified as Eukaryotic translation initiation factor 3 (eIF3) ortholog, which is a protein that also plays important role in cell response to ionizing radiation. Patients with lower expression of eIF3 protein were having poor prognosis for several cancers [18]

Other proteins that showed no sequence similarity to any protein in BLASTP search can also potentially bear unknown functions, as it was with previously described Dsup protein [12]. For further analysis of all those proteins we suggest to perform experiments that could potentially prove nuclear localization (FISH, GFP-fusion), and influence of expression of these genes on extreme conditions (treatment of the cell lines with high temperature, radiation and

with other DNA-damaging factors).

References

1. Jönsson KI, Holm I, Tassidis H. Cell biology of the tardigrades: current knowledge and perspectives. Evo-Devo: Non-model Species in Cell and Developmental Biology 2019;p. 231–249.
2. Carrero D, Pérez-Silva JG, Quesada V, López-Otín C. Differential mechanisms of tolerance to extreme environmental conditions in tardigrades. Scientific reports 2019;9(1):1–11.
3. Boothby TC, Tenlen JR, Smith FW, Wang JR, Patanella KA, Osborne Nishimura E, et al. Evidence for extensive horizontal gene transfer from the draft genome of a tardigrade. Proceedings of the National Academy of Sciences 2015;112(52):15976–15981.
4. Koutsovoulos G, Kumar S, Laetsch DR, Stevens L, Daub J, Conlon C, et al. No evidence for extensive horizontal gene transfer in the genome of the tardigrade Hypsibius dujardini. Proceedings of the National Academy of Sciences 2016;113(18):5053–5058.
5. Stanke M, Morgenstern B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Research 2005 07;33(suppl$_2$) : $W465 - -W467$.
6. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. Journal of molecular biology 1990;215(3):403–410.
7. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. GigaScience 2021 02;10(2).
8. Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, et al. WoLF PSORT: protein localization predictor. Nucleic Acids Research 2007 07;35(suppl$_2$) : $W585 - -W587$.
9. Almagro Armenteros JJ, Salvatore M, Emanuelsson O, Winther O, von Heijne G, Elofsson A, et al. Detecting sequence signals in targeting peptides using deep learning. Life Science Alliance 2019;2(5).
10. Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau D, et al. Database resources of the national center for biotechnology information. Nucleic Acids Research 2021 12;50(D1):D20–D26.
11. Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. HMMER web server: 2018 update. Nucleic Acids Research 2018 06;46(W1):W200–W204.
12. Hashimoto T, Horikawa D, Saito Y, Kuwahara H, Kozuka-Hata H, Shin-I T, et al. Extremotolerant tardigrade genome and improved radiotolerance of human cultured cells by tardigrade-unique protein. Nature Communications 2016 09;7:12808.
13. Natarajan C, Takeda K. Regulation of various DNA repair pathways by E3 ubiquitin ligases. Journal of Cancer Research and Therapeutics 2017;13(2):157–169.
14. Chernikova SB, Dorth JA, Razorenova OV, Game JC, Brown JM. Deficiency in Bre1 Impairs Homologous Recombination Repair and Cell Cycle Checkpoint Response to Radiation Damage in Mammalian Cells. Radiation Research 2010;174(5):558 – 565.
15. Ghosal G, Yuan J, Chen J. The HARP domain dictates the annealing helicase activity of HARP/SMARCAL1. EMBO reports 2011;12(6):574–580.
16. Driscoll R, Cimprich K. HARPing on about the DNA damage response during replication. Genes development 2009 10;23:2359–65.
17. Yuan J, Ghosal G, Chen J. The HARP-like Domain-Containing Protein AH2/ZRANB3 Binds to PCNA and Participates in Cellular Response to Replication Stress. Molecular Cell 2012;47(3):410–421.
18. Tumia R, Wang CJ, Dong T, Ma S, Beebe J, Chen J, et al. eIF3a Regulation of NHEJ Repair Protein Synthesis and Cellular Response to Ionizing Radiation. Frontiers in Cell and Developmental Biology 2020;8.