BIOINFORMATICS
INSTITUTE

I was cured all right: how the clockwork pudgy wudgies and the droogs have their supersurvivability

Nikita Vaulin[1, 2,†]

[1]Bioinformatics Institute, Saint-Petersburg, Russia
[2]Skolkovo Institute of Science and Technology, Moscow, Russia

[†]vaulin@ro.ru

Abstract

Tardigrades are a phylum of eight-legged segmented micro-animals, these creatures can be very stress-tolerant, due to prevention of DNA damage. However, molecular mechanism of this damage suppression is not discovered yet. In this work, we explored how they can efficiently reduce damage to their DNA. We figured out which proteins from the *Ramazzottius varieornatus* genome have nuclear localisation, using information about the nuclear peptides obtained using tandem mass spectrometry and found the relevant ones. Then we predicted genes and analyse where encoded proteins are found in the cell, searched for them in the protein databases, trying to do analysis of their functions. Lastly, we found four novel proteins, which are possibly responsible for radiation tolerance in tardigrades. As a result, we propose several approaches to test the effect of these proteins on tolerance to DNA damage based on RNA-seq, ChIP-seq, and heterologous expression.

Key words:
Tardigrades, DNA-damage, gene prediction, functional annotation, protein localisation

Introduction

Tardigrades are tiny invertebrates belonging to the Protostome clade, and they have a unique ability to survive in extreme environmental conditions, such as drying, osmotic stress and freezing [1]. In this situations, tardigrade undergo cryptobiosis, which is a reversible metabolism suppression accompanied by anatomical and physiological changes [2]. This state allows tardigrade to stay alive for many years.

Figure 1. A mechanism of how actually *Ramazzottius varieornatus* enters anhydrobiosis [3]

External toxins also cannot easily kill tardigrade, nor can increasing radiation levels do it [2]. Radiation is a factor that can lead to a serious DNA molecule damage, such as double strand breaks. It was hypothesised that radiation tolerance is caused by the high DNA-repair system activity [4], and in this work we are interested in bioinformatics prediction of proteins causing damage suppression.

Given that radiation affect DNA molecules, we assumed that novel DNA-related protein should be involved in damage tolerance. To find this novel proteins, we used methods of gene prediction, their functional annotation, and also proteins localisation prediction on *Ramazzottius varieornatus* the YOKOZUNA-1 strain sequence and tandem mass spectrometry data. The gene prediction and the search for homologous proteins carry the task of determining what allows tardigrades to be so tolerant to the harmful radiation effects.

Materials and methods

Genomic data

In this work, an assembled genome of the YOKOZUNA-1 strain of *Ramazzottius varieornatus* was used (the data is available *via* the link). The structural annotation of the genome was performed with the AUGUSTUS and the data is available *via* the link [5]. The protein sequences in fasta format were exracted from the structurally annotated genome with the perl script (see Supplemental materials, GitHub repository, *getAnnoFasta.pl*).

Proteins data

Here we used sequences of the proteins, associated with the DNA (the data is available *via* the link). The particular proteins were indentified with the tandem mass spectrometry of the chromatin fraction.

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

Predicted protein ID	BLAST Acc. №	E-value	% Ident.	Description	HMMER Acc. №	Descripton
g10513	-	-	-	-		
g10514	-	-	-	-		
g11806	-	-	-	-		
g11960	Q8CJB9.1	6e-98	27.0%	E3 ubiquitin-protein ligase BRE1B	PF00097.28	Zinc finger, C3HC4 type (RING finger)
g14472	-	-	-	-		
g15484	Q155U0.1	0	45.0%	Vacuolar protein sorting-associated protein 51 homolog	PF08700.14	Vps51/Vps67
g16318	A2VD00.1	4e-08	36.1%	Eukaryotic translation initiation factor 3 subunit A		
g16368	A4II09.1	1e-05	39.3%	Eukaryotic translation initiation factor 3 subunit A		
g5927	Q17427.1	1e-18	38.7%	Glucosamine 6-phosphate N-acetyltransferase		
g7861	B4F769.1	2e-71	37.2%	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A-like protein 1	PF00176.26 PF07443.16	SNF2-related domain, HepA-related protein (HARP)
g8100	Q2YDR3.1	3e-46	36.0%	Inositol monophosphatase 3	PF16540.8	Inositol monophosphatase familyArf6-interacting domain of mitotic kinesin-like protein 1
g8312-t1	Q5KU39.1	0	40.8%	Vacuolar protein sorting-associated protein 41	PF00637.23	Region in Clathrin and VPS

Table 1. BLAST homologs search and HMMER domains search results for the proteins of interest.

Determination of the DNA-associated proteins

DNA-associated proteins were determend based on the predicted protein-coding genes and the tandem mass spectrometry results. A protein database was created from the predicted proteins with the *makeblastdb* tool from the *blast* package v. 2.13.0 [6]. Extracted nuclear peptides were searched against this database with *blastp* with the default parameters. The preteins and their amino acid sequences were extracted with the *faidx* tool from the *samtools* package v.1.16.1 [7].

Localisation of the found proteins was predicted with the WolF PSORT with the *'animal'* organism type selected [8].

Identification of proteins potentially affecting supersurvivability

Found nuclear-localised proteins were searched with the blastp against the UniProtKB/Swiss-Prot database [6]. The protein functional domains were predicted with the HMMER web-application with the Pfam database [9, 10].

Results

In this work a set of potentially affecting supersurvivability nuclear proteins were obtained. From the 16435 predicted proteins and 43 peptides from chromatin fraction the amount of 12 proteins was selected as a set of proteins of interest. The BLAST search against the UniProtKB/Swiss-Prot database and functional domain prediction with HMMER results are presented in the Table 1.

For the BLAST search analysis, the very first best hit is provided. All os the presented entryes were predicted to localise within the nuclear according to the WoLF PSORT analysis.

Discussion

Of the 12 found DNA-associated nuclear proteins several were indentified to be potentially responsible for the supersurvivability. The proteins for which any homologues were found by BLAST have important but quite standard cellular functions. Certain functional domains were also predicted for these proteins.

Nevertheless, surface analysis of these proteins did not reveal anything interesting. One could say that using a search of the available databases is a so-called «lantern search». Based on the results of this work, it is worth looking more closely at those proteins for which no homologues or known functional domains were found. It may be that it is these nuclear proteins that have hitherto unknown functions that are of interest to us.

We found 4 proteins that may be associated with the tolerance of tardigrades to the DNA damage: g10513, g10514, g11806 and g14472 (ID's provided according to the assembly, see). Nothing has been found for these proteins in databases, so it is worth looking into them in more detail.

First of all, it is worth suggesting a more detailed analysis to check whether these proteins are actually produced in the cell. It is worth performing an RNA-seq assay. Experiments can also be performed to isolate these proteins. If succeed, it is worth checking whether these proteins are indeed associated with DNA. To do this, ChIP-seq experiments can be performed. In order to test whether these proteins allow for DNA damage tolerance, the genes of these proteins can be cloned into plasmids to carry out heterologous expression. Increased survival of the recipient organism can be expected in the presence of the necessary protein product under extreme conditions.

Supplemental materials

GitHub repository of the project: Tardigrades_project_BI_2022

References

1. Møbjerg N, Neves RC. New insights into survival strategies of tardigrades. Comparative Biochemistry and Physiology Part A: Molecular &amp Integrative Physiology 2021 Apr;254:110890. https://doi.org/10.1016/j.cbpa.2020.110890.
2. Kasianchuk N, Rzymski P, Kaczmarek Ł. The biomedical potential of tardigrade proteins: A review. Biomedicine &amp Pharmacotherapy 2023 Feb;158:114063. https://doi.org/10.1016/j.biopha.2022.114063.
3. Arakawa K. Examples of Extreme Survival: Tardigrade Genomics and Molecular Anhydrobiology. Annual Review of Animal Biosciences 2022;10:17–37.
4. Hashimoto T, Kunieda T. DNA Protection Protein, a Novel Mechanism of Radiation Tolerance: Lessons from Tardigrades. Life 2017 Jun;7(2):26. https://doi.org/10.3390/life7020026.
5. Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics 2008;24(5):637–644.
6. et al ASF. 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). https://doi.org/10.1093/gigascience/giab008, giab008.
8. Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier C, et al. WoLF PSORT: protein localization predictor. Nucleic acids research 2007;35(suppl_2):W585–W587.
9. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar G, Sonnhammer ELL, et al. Pfam: The protein families database in 2021. Nucleic Acids Research 2020 10;49(D1):D412–D419. https://doi.org/10.1093/nar/gkaa913.
10. Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. HMMER web server: 2018 update. Nucleic acids research 2018;46(W1):W200–W204.