Tardigrades: a Hitchhiker’s Guide to Surviving in Space

Sobolev A., Kupaeva D.

Abstract

Tardigrades are a group of small animals from the Ecdysozoa supergroup. A feature of tardigrades is the ability to survive extremely adverse conditions, including being in outer space, extremely high levels of exposure to ionizing radiation or ultraviolet light. It is assumed that this is provided by a specialized system for repairing DNA damage, but its mechanisms are still not known. In this article, we detect and analyze the DNA-associated proteins of tardigrades using various tools and suggest candidates for the role of damage repair proteins.

Introduction

Tardigrades, also known as “water bear” - a group of land, freshwater and marine small size animals, belonging to the Ecdysozoa superfamily. Tardigrades have the ability to survive adverse conditions in an anhydrobiosis, an ametabolic dry state, reminiscent of the cysta state of a number of other animals. Tardigrades in a dehydrated state also showed unprecedented abilities to survive after high doses of ionizing[1] and UV[2] radiation, and they can survive[3] after a 10-day long exposition in open space.

The harmful effects of radiation on living organisms mainly involve damage to the animal's DNA. The key problem associated with ultraviolet irradiation of tissues is the formation of Thymine cyclobutane dimer. Such damage alters the base pairing ability of the bases involved and can be cytotoxic and block transcription and replication. An important cellular mechanism for preventing potential mutations resulting from damage is structural repair through an excisional nucleotide repair process, in which damage is recognized and removed.[4]

Alternative methods of protection can be the compaction of DNA and the formation of nucleosomes, which protect part of the DNA from possible damage, the spatial localization of chromatin, as well as binding to transcription factors.[5]

Beside direct way of damage DNA by UV-radiation, exsist mediated ways. The most commonly employed means of generating radicals is the Fenton reaction, which produces hydroxyl radicals (•OH) catalytically as a result of the reduction of hydrogen peroxide by Fe(II)-EDTA.[6]

Despite the large amount of data on ways to protect DNA from damage, based on them, we can not guess how exactly tardigrades survive such a high level of DNA exposure. And in

this study, we are trying to find specific proteins of tardigrades, potentially associated with additional DNA protection or repair of damage.

Methods

Object of our study - tardigrade *Ramazzottius varieornatus*. We used a pre-assembled genome assembly. Repeats were masked with the help of RepeatModeller. Gene prediction was performed by web-server AUGUSTUS[7]. For choosing the closest model species from the AUGUSTUS database we used an online tool PhyloT for phylogeny reconstruction. According to results (fig.1) we chose *Acyrthosiphon pisum.* For visualising trees we used other online-tool.

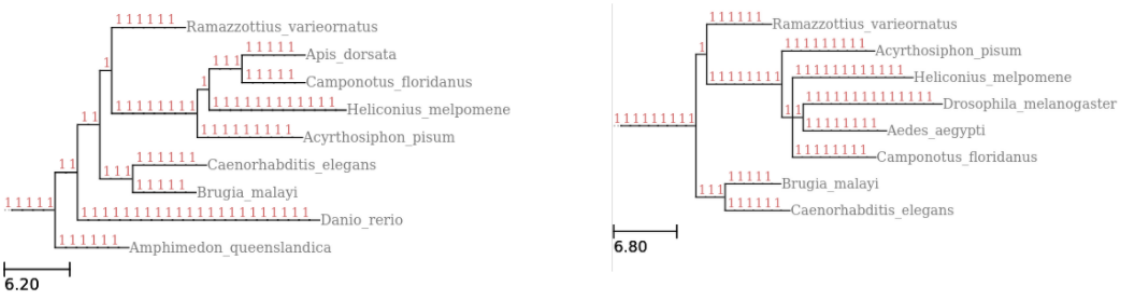

Figure 1. Phylogeny reconstruction for AUSUSTUS database species.

We narrowed our search for proteins by analyzing peptides that form a complex with DNA in the nucleus. Information on peptides detected by mass spectrometry is available at the link. Based on this list of proteins, we made a local blast database using *makeblastdb*. After this, we used the list of peptides as a query for blastp to obtain the proteins that contained peptides' sequences. Using blastp and samtools faidx, we obtained a list of tardigrade proteins containing protein motifs of interest. The functional analysis was carried out using: TargetP 2.0[8] (http://www.cbs.dtu.dk/services/TargetP/) in non-plant organism group, WoLF PSORT[9] (https://wolfpsort.hgc.jp/) (we have chosen "Animal" organism type), BLAST [10](without results for *Ramazzottius varieornatus,* database UniProtKB/Swiss-Prot), and HMMScan[11] using Pfam database. All results were combined into a common database and carefully studied.

All steps of analysis you can find in our supplementary material 1.

Results

Aggregation of data produced by various tools results in five proteins of interest

From the initial *R. varieornatus* assembly, 16435 proteins were obtained. We then As a result, we identified 34 unique proteins which contain target peptides. Among them, 17 were localized in the nucleus according to the WoLF PSORT results. All of them are shown in

supplementary material 2, including annotation. In case of TargetP 2.0, secretory pathway signal peptides were detected in 13 of 34 protein sequences. We then carried out a blastp search, excluding hits from *Ramazzottius varieornatus*. Some hits were found for the 23 of 34 protein sequences. Only top blastp hit (selected by E-value) was taken into account in further analysis. Finally, WoLF PSORT, TargetP 2.0 and blastp results were combined together. We then selected only the proteins for which there is a probable nuclear localization according to WoLF PSORT and filtered out the sequences that contained a secretory pathway signal peptides according to TargetP 2.0. We then considered filtering out the proteins that are more likely localized in the nucleus — if WoLF PSORT score was notably higher for other localizations than the nucleus, these entries were filtered out. As a result, 12 proteins were identified. Finally, we filtered out Pfam entries that were present in our resulting data. We ended up with 5 unique proteins named g11960.t1, g15484.t1, g7861.t1, g8100.t1 and g8312.t1 and moved into research of their structure and the function of identified domains.

Several proteins contain domains that are related to DNA binding and transcription

We first concentrated on studying the functions of proteins that are recognized by Pfam. Protein g11960.t1 contains zinc finger (Znf) RING domain. It is known that proteins with domains are involved in many processes in the cell including gene transcription, translation, mRNA trafficking, cytoskeleton organisation, epithelial development, cell adhesion, protein folding, chromatin remodelling and zinc sensing[12]. Top blastp hit for this protein corresponds to the BRE1B ubiquitin-protein ligase that mediates ubiquitination of H2B histone's Lys120 which plays a central role in histone code and gene regulation. In addition, this gene is also required for Hox genes activation, facilitating its possible role in gene expression regulation in Tardigrades. g15484.t1 protein, along with g8312.t1, is functionally annotated as vacuolar protein-sorting protein, despite the presence of nuclear signal peptide in them. Blastp successfully confirms these findings, corresponding its top hit to vacuolar protein-sorting associated protein 51 homolog of *Danio rerio*. These results suggest that these proteins barely interact with DNA and are not involved in its protection, that is why they did not attract our attention. The opposite situation, however, was observed in case of g7861.t1 protein. There are three Pfam matches in its structure — SNF2 family domain on its N-terminus, HepA-related protein and res subunit of type III restriction enzyme. SNF2 N-terminal domain, the largest of identified three, is involved in various processes involving transcriptional regulation, DNA repair, DNA recombination and chromatin unwinding. HepA-related protein (HARP) is a helicase protein that exhibits single-stranded DNA-dependent ATPase activity, and is ubiquitously expressed in human and mouse tissues[13]. Finally, Res domain is present at the N-terminus of helicases, some endo- and exonucleases. The presence of these domains in g7861.t1 suggest that it can possibly exhibit helicase activity, playing a potential role in DNA repair and recombination. Blastp results facilitate our suggestions — top blastp hit corresponds to SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A-like protein 1, supporting the role of characterized protein in chromatin regulation. As for g8100.t1, two Pfam matches were detected. One of them refers to the inositol monophosphatase family, members of which remove a phosphate group from any mono- or polyphosphorylated inositol. Second match corresponds to the Arf6-interacting domain of mitotic kinesin-like protein 1 which plays a role in cytokinesis and, therefore, cell division. Top blastp hit corresponds to the inositol monophosphatase 3 of *Danio rerio*, however, its query cover and percent identity are weak — 22% and 36.04%, respectively. In

addition, role of the Arf-6 interacting domain in cytokinesis does not involve DNA interactions. Therefore, these controversial findings are not enough to consider this protein potentially involved in DNA damage protection.

Investigation of proteins that do not display Pfam matches

After initial research, we decided to analyze our results more thoroughly and study functions of proteins that have not been characterized by Pfam, 7 proteins in total. Homologs of three of these proteins (g16318.t1, g16368.t1 and g5927.t1) were found using blastp. g16318.t1 and g16368.t1 corresponded to the RNA-binding translation initiation factors 3a (eIF3a) in *Xenopus* species *X. laevis* and *X. tropicalis*, involved in the translation initiation processes. Query cover and percent identity for these proteins, however, are weak, as in case of g8100.t1, and absence of Pfam matches along with these results and lack of DNA relation evidence lead us to the strong suggestion that these proteins should not be considered as proteins of interest (at least in our study). As for the g5927.t1, its top blastp hit corresponds to glucosamine 6-phosphate N-acetyltransferase with weak query cover and percent identity and also is not related to DNA protection. Other four proteins did not display any significant similarity in blastp, therefore, they are rather difficult to characterize.

Discussion

Among all proteins, there are two of interest: g7861.t1 and g11960.t1. It was shown that g7861.t1 contains domains that are responsible for DNA reparation, recombination and gene expression regulation. This protein also seems to exhibit helicase activity. All these findings make this protein a potential candidate for experimental verification of its function. For instance, to study the protein structure more thoroughly, it is possible to use crystallography and X-rays diffraction. Moreover, one can fuse green fluorescent protein (GFP) to the potential candidate to track its subcellular distribution and characterize it further by DNA footprinting analysis[14]. The same situation is observed for the g11960.t1 protein. Its influence on the DNA transcription, albeit more indirect than in case of g7861.t1, was successfully predicted by all tools used in this study. Its zinc finger RING domain can facilitate gene expression via activation of Hox genes and, therefore, can possibly regulate the response of the organism to environmental challenges and constraints. We consider this protein worth studying as well.

There is another question that this study needs to address for sure. There is a category of proteins that we have not characterized at all. These are proteins that were identified as nuclear using WoLF PSORT and TargetP 2.0 but did not show any significant similarities in blastp and Pfam. These proteins exhibit great interest for sure, because they are not characterized at all and may correspond to unique proteins of Tardigrades that prevent their DNA from damage. So, we propose to study these proteins (g10513.t1, g10514.t1, g11806.t1 and g14472.t1) as the main proteins of interest because their structure and functions can shed light on Tardigrades' resistance to the harsh environmental conditions.

Citations

---

1) Horikawa DD, Sakashita T, Katagiri C, Watanabe M, Kikawada T, Nakahara Y, Hamada N, Wada S, Funayama T, Higashi S, Kobayashi Y, Okuda T, Kuwabara M. Radiation tolerance in the tardigrade Milnesium tardigradum. Int J Radiat Biol. 2006 Dec;82(12):843-8. doi: 10.1080/09553000600972956. PMID: 17178624.
2) Altiero, T., Guidetti, R., Caselli, V., Cesari, M., & Rebecchi, L. (2011). Ultraviolet radiation tolerance in hydrated and desiccated eutardigrades. *Journal of Zoological Systematics and Evolutionary Research*, *49*, 104-110.
3) Jönsson, K. I., Rabbow, E., Schill, R. O., Harms-Ringdahl, M., & Rettberg, P. (2008). Tardigrades survive exposure to space in low Earth orbit. *Current biology*, *18*(17), R729-R731.
4) McGregor, W. G. (1999, September). DNA repair, DNA replication, and UV mutagenesis. In *Journal of Investigative Dermatology Symposium Proceedings* (Vol. 4, No. 1, pp. 1-5). Elsevier.
5) Mao, P., Wyrick, J. J., Roberts, S. A., & Smerdon, M. J. (2017). UV-induced DNA damage and mutagenesis in chromatin. *Photochemistry and photobiology*, *93*(1), 216-228.
6) McCue, A. C., Moreau, W. M., & Shell, T. A. (2018). Visible Light-Induced Radical Mediated DNA Damage. *Photochemistry and photobiology*, *94*(3), 545-551.
7) Stanke, M., Diekhans, M., Baertsch, R., & Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics*, *24*(5), 637-644.
8) Armenteros, J. J. A., Salvatore, M., Emanuelsson, O., Winther, O., Von Heijne, G., Elofsson, A., & Nielsen, H. (2019). Detecting sequence signals in targeting peptides using deep learning. *Life science alliance*, *2*(5).
9) Fu, S. C., Imai, K., & Horton, P. (2011). Prediction of leucine-rich nuclear export signal containing proteins with NESsential. *Nucleic acids research*, *39*(16), e111-e111.
10) National Center for Biotechnology Information (NCBI)[Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988] – [cited 2017 Apr 06]. Available from: https://www.ncbi.nlm.nih.gov/
11) Finn, R. D., Clements, J., & Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic acids research*, *39*(suppl_2), W29-W37.
12) Laity JH, Lee BM, Wright PE. Zinc finger proteins: new insights into structural and functional diversity. Curr Opin Struct Biol. 2001 Feb;11(1):39-46. doi: 10.1016/s0959-440x(00)00167-6. PMID: 11179890.
13) Coleman MA, Eisen JA, Mohrenweiser HW. Cloning and characterization of HARP/SMARCAL1: a prokaryotic HepA-related SNF2 helicase protein from human and mouse. Genomics. 2000 May 1;65(3):274-82. doi: 10.1006/geno.2000.6174. PMID: 10857751.
14) Alberts B, Johnson A, Lewis J, et al. Molecular Biology of the Cell. 4th edition. New York: Garland Science; 2002. Analyzing Protein Structure and Function. Available from: https://www.ncbi.nlm.nih.gov/books/NBK26820/

Supplemental resources

---

1) Laboratory journal: https://www.notion.so/project-4-e636180e39fb4f41b5d9d366eadf8860
2) Table with annotation proteins with DNA-associated motifs: https://docs.google.com/spreadsheets/d/1Zn6xATZuOGig6J4hEVUpVWVN6ECz8CWeIuPPF9so0AI/edit#gid=0