BIOINFORMATICS INSTITUTE

Identification of the possible genetic background of the *Ramazzottius varieornatus* unique radiation resistance

Anna Malykhina
Bioinformatics Institute, Saint-Petersburg, Russia

Sidorenko Oksana
Bioinformatics Institute, Saint-Petersburg, Russia

December 14, 2022

Abstract

Tardigrades are small multicellular eukaryotes interesting for their unique stress resistance to various physical extremes, *e.g.* radiation-tolerance. The mechanisms of this phenomenon were not clear until 2016. We set the task of modeling a study that could provide materials for the discoveries of 2016. For this purpose we used the genome sequence of *Ramazzottius varieornatus*, the YOKOZUNA-1 strain, sequenced in 2015 in the University of Tokyo to analyze and identify genes that may be responsible for effective DNA protection and repair. A number of chromatin fraction proteins were found that had no orthologous sequences except sequences of *R. varieornatus* which were unknown before 2016. Subsequent studies by laboratory and bioinformatic methods confirmed that the proteins we proposed for further study were envolved in the mechanisms of radiation-tolerance. Moreover, the Damage suppressor (Dsup), a tardigrade-unique protein that suppresses radiation-induced DNA damage was found among them.

Keywords: *R. varieornatus*, stress-tolerance, X-ray-induced DNA damage, DNA repair, gene prediction, functional annotation, structural annotation

Introduction

Invincible tardigrades

Tardigrades (*Tardigrada*), also known as water bears or moss piglets, are a phylum of small invertebrates. [1] More than 1,200 species of tardigrades have been reported to inhabit all kinds of water environments. [2] All tardigrades require surrounding water to grow and reproduce, but some species-typically those living in the limno-terrestrial environments—have the ability to tolerate almost complete dehydration. When encountering desiccation, tolerant tardigrades lose body water and enter a contracted dehydrated state called anhydrobiosis, which is a reversible ametabolic state. The dehydrated tardigrades withstand a wide range of physical extremes that normally disallow the survival of most organisms, such as extreme temperatures (from -273°C [3] to nearly 100°C [4] [5]), high pressure (more than 1,200 times atmospheric pressure; some species can also withstand pressure of 6,000 atmospheres, which is nearly six times the pressure of water in the deepest ocean trench, the Mariana Trench [6]), immersion in organic solvent [5] [7], exposure to high dose of irradiation (1,000 times more radiation than other animals [3]). [8] Tardigrades are the first known animal [11] to survive after exposure to outer space. [9] [10].

Background and design of the research

In this work, we focused on the radiation-tolerance of tardigrades. It has been shown that their ability to enter the dehydrated state does not explain their radiation-tolerance: when hydrated, they still remain very resistant to shortwave UV radiation. Scientists have suggested the tardigrades' outstanding capacities of DNA damage repair and DNA protection. [12] [13] [14] [15]
Previous tardigrade genome studies were often based on the analysis of the *Drosophila melanogaster* and *Caenorhabditis elegans* genomes as well-studied and taxonomically close enough species to tardigrades [Figure 1], and on the analysis of the *Hypsibius dujardini* - another species of the *Tardigrada*.

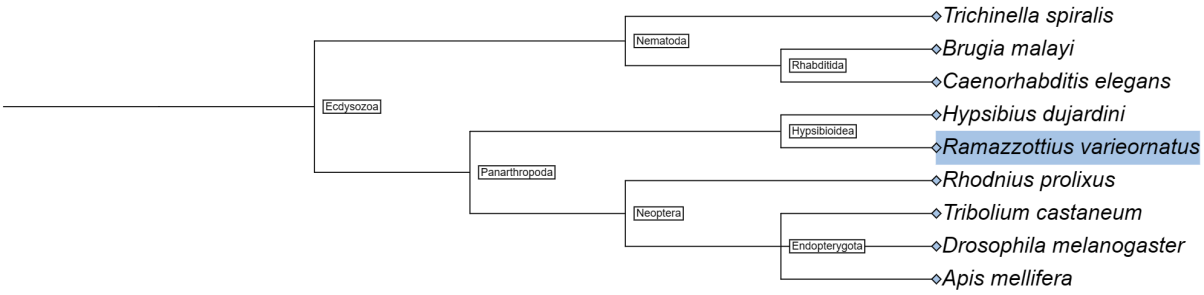

Figure 1: Simplified phylogenetic tree of *Ecdysozoa*.
Tardigrades have been placed in *Panarthropoda* using a plethora of morphological characters, as well as in several molecular analyses designed to counter long branch attraction, and based on a novel microRNA [18], but they are also often drawn to *Nematoida* in molecular analyses [19] [20]
The image was build using PhyloT software [22].

H. dujardini was the only non-limno-terrestrial tardigrade species that have been used in radiation studies until 2015. [16] However, the genome of *R. varieornatus* was sequenced by a team of researchers from the University of Tokyo in 2015. Even before this date it was documented [14] a much higher radiation tolerance in *R. varieornatus*. There was also provided evidence of a lower tolerance to desiccation in *H. dujardini*. Also the lower desiccation tolerance in *H. dujardini* compared to several other tardigrades that inhabit limno-terrestrial microhabitats was founded. [16] [17] Based on such a high stress resistance of *R. varieornatus*, even in comparison with a related species (*H. dujardini*), we suggested the presence of

some unique proteins in *R. varieornatus*. Thus, it made sense to analyze the *R. varieornatus* genome that was computed with modeling tool which has been trained on tardigrades cDNA data, without relying on parallels with genome sequences of *C. elegans* and *D. melanogaster* since such a formulation of the problem makes their genomes unsuitable as model ones. Considering DNA as a major target of radiation damage, we hypothesized that proteins that associate with DNA to protect and/or to effectively repair DNA in the tardigrade should be searched. [8] To study this possibility, we have analysed the list of peptides that were presumably associated with the DNA, found the relevant proteins in the tardigrade genome and focused only on those that predicted to be localized in the nucleus. Then we looked for the annotation in number of different databases.

Methods

Sequence source

Assembled genome of *R. varieornatus* was downloaded from: ftp.ncbi.nlm.nih.gov/genomes/all/GCA/001/949/185/GCA_001949185.1_Rvar_4.0/GCA_001949185.1_Rvar_4.0_genomic.fna.gz
To obtain information about genes and proteins, tardigrade genome was computed with AUGUSTUS modeling tool [21] which has been trained on tardigrade cDNA data. The result was downloaded from: https://drive.google.com/file/d/1wBxf6cDgu22NbjAOgTe-8b3Zx60hNKY0/view?usp=drive_web
List of peptides from nuclear fraction was downloaded from: https://disk.yandex.ru/d/xJqQMGX77Xueqg

Data analysis

Phylogenetic tree was build using PhyloT [22]. Repeats were identified by RepeatModeler(v2.0.3) [23]. Information about genes based on relative species was obtained using AUGUSTUS prediction tool [21]. For protein sequence extraction getAnnoFasta.pl(augustus.gobics.de/binaries/scripts/getAnnoFasta.pl) script was used. Local alignment-based search was done by BLAST program (v 2.11.0) with default parameters. Samtools program (v 1.16.1) [24] [25] was used for indexing. Self-written script was used for protein sequence extraction (see Supplementary). For protein subcellular localization prediction two website programs were used: WolF PSORT [26] and TargetP - 2.0 [27] with default parameters. For proteins annotation BLAST search [28] and Pfam [29] prediction were used.

Results

Data preprocessing

First, *R. varieornatus* genome was masked for repeats before alignment-based search and protein analysis. There were about 600 different repeats reported by RepeatModeler (575, 566, 557 – depending on program launch). AUGUSTUS gene prediction based on annotated genomes of *D. melanogaster* and *C. elegans* outputed 11300 and 17769 genes accordingly. This result shows that *R. varieornatus* is closer to *C. elegans* than to *D. melanogaster* in terms of genes number and is consistent with some doubt about the membership of *Tardigrada* in *Panarthropoda* in a series of phylogenomic analyses. Several studies have found a relationship of *Tardigrada* to *Nematoida* (this debate is often discussed as a nematode–tardigrade relationship). [20] Final dataset was obtained from AUGUSTUS trained on tardigrades cDNA and contained 16435 genes.

Gene filtration

According to the planned study design we assumed that tardigrades' radiation-tolerance may be linked to some nuclear proteins. To check this hypothesis, our collaborators obtained nuclear fraction of *R. varieornatus* and analyzed peptides by mass spectrometry. The resulting file contained 43 peptides. We got 34 matches after alignment-based search: with a local database from annotated *R. varieornatus* genome and peptide sequence file as a query. However, not all these proteins could be of interest for our purpose as a contamination during nuclear fraction preparation is possible. Thus, two additional programs (WolF PSORT and TargetP) were used to find signal sequences corresponding to nuclear localization. After such filtration we received 12 proteins which were consequently annotated with BLAST and Pfam [Figure 2]. It worth mentioning that 4 of these 12 proteins were marked as "uncharacterized" which means that they were not described previously for other species and are unique for *R. varieornatus*. We could propose these 12 proteins for further studies to confirm or refute our assumption of their role in tardigrades' radiation-tolerance mechanisms.

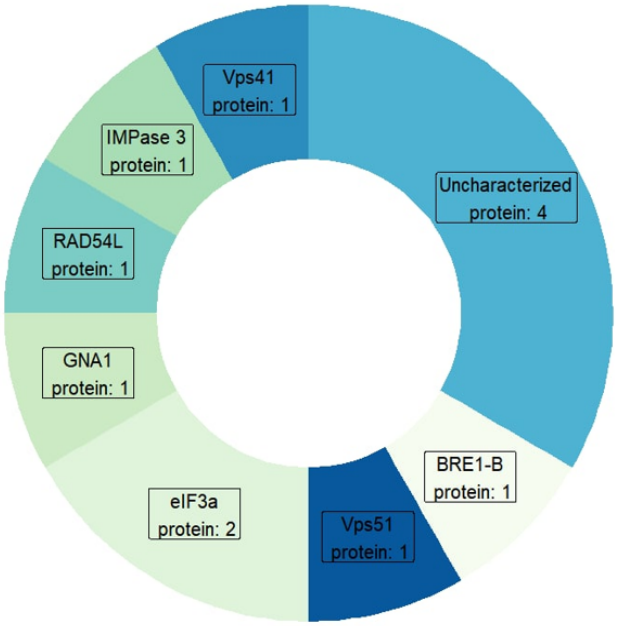

Figure 2: Proteins of interest presented in nucleus of *R. varieornatus* detected after genome filtration and annotation.

Discussion

Dsup

Based on current data obtained in recent years, *e.g.* Hashimoto et al. "Extremotolerant tardigrade genome and improved radiotolerance of human cultured cells by tardigrade-unique protein." [8] we can conclude that all the detected genes are classified as tardigrade-unique genes that exhibited no or low similarity to non-tardigrade proteins and were designated by the authors as *RvY*. One of these 12 proteins, termed Damage suppressor (**Dsup**), co-localized with nuclear DNA protein, turned out to be a a

tardigrade-unique DNA-associating protein that suppresses X-ray-induced DNA damage by $\sim 40\%$ and improves radiotolerance. [2] [8] Later it was shown that *R. varieornatus* Dsup is a nucleosome-binding protein that protects chromatin from hydroxyl radicals. [30]

Other detected proteins

We have found several indications that other identified proteins may be involved in DNA reparation processes.

DNA repair and recombination protein RAD54-like or **SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A-like protein 1**
This protein was predicted by BLAST globally in two ways: e-value was 2e-71 for *Rattus norvegicus* and 5e-14 for *Drosophila pseudoobscura* accordingly. We left both variants for consideration and proposals for further studies. SNF2-related domain is found in proteins involved in a variety of processes including transcription regulation, DNA repair (e.g., ERCC6, RAD16, RAD5), DNA recombination (e.g., RAD54), and chromatin unwinding (e.g., ISWI) as well as a variety of other proteins with little functional information (e.g., lodestar, ETL1). SNF2 functions as the ATPase component of the SNF2/SWI multisubunit complex, which utilises energy derived from ATP hydrolysis to disrupt histone-DNA interactions, resulting in the increased accessibility of DNA to transcription factors. [29]

E3 ubiquitin-protein ligase BRE1B
The cellular response to external stress signals and DNA damage depends on the activity of ubiquitin ligases (E3s), which regulate numerous cellular processes, including homeostasis, metabolism and cell cycle progression. E3s recognize, interact with and ubiquitylate protein substrates in a temporally and spatially regulated manner. Ubiquitin ligases coordinate the cell cycle and DNA damage repair to maintain genome integrity. Ubiquitin-dependent deregulation of specific signalling hubs in turn alters cellular processes as diverse as cell growth, metabolism, DNA repair, transcription, translation and survival. [31]

Eukaryotic translation initiation factor 3 subunit A eIF3a
It was shown that eIF3a upregulation sensitized cellular response to ionizing radiation (IR) while its downregulation caused resistance to IR. eIF3a increases IR-induced DNA damages and decreases non-homologous end joining (NHEJ) activity by suppressing the synthesis of NHEJ repair proteins. Furthermore, analysis of existing patient database shows that eIF3a expression associates with better overall survival of breast, gastric, lung, and ovarian cancer patients. These findings together suggest that eIF3a plays an important role in cellular response to DNA-damaging treatments by regulating the synthesis of DNA repair proteins. [32]

Glucosamine 6-phosphate N-acetyltransferase
As part of the study of potential role of glucosamine-phosphate N-acetyltransferase 1 in the development of lung adenocarcinoma (LUAD) a network of kinases related to glucosamine-phosphate N-acetyltransferase 1 in LUAD, including CDK1, PLK1, AURKB, CDK2, and AURKA, was revealed. PLK1, AURKB, and AURKA are crucial factors not only in mitosis, but also in non-mitosis function and DNA damage response. In LUAD, glucosamine-phosphate N-acetyltransferase 1 deficiency results in cell cycle arrest, DNA damage, and repair response dysfunction, which might be due to the synergistic effects of these kinases. [33]

Further research

There is a wide field of activity for researchers in establishing the specific functions of proteins that may be responsible for the tardigrades' radiation-tolerance and stress-tolerance in general. Within our study we have not found an objective evidence of the effect of two Vps proteins and IMPase 3 on DNA protection or repair. In addition, 3 more discovered proteins remain uncharacterized and the most interesting, being unique to *R. varieornatus*. It is possible that Vps proteins could acquire a new function, such as sorting out damaged nuclear proteins, and IMPase 3 may be an enzyme that affects DNA repair or protection, since it is involved in the phosphatidylinositol signaling pathway which takes part in a number of cellular functions. We can offer to study protein - protein interaction by the co-immunoprecipitation method, DNA-protein interaction by the Chip-seq method or gene knockdown to check the version of the signaling pathway, as well as to study mutations and protein structure to verify our findings and for further experiments to explore other genetic mechanisms of the stress resistance of these small resilient creatures.

Supplementary

The lab notebook with the detailed pipeline, settings and details can be found at the link https://docs.google.com/document/d/15tqpawUbcyQVrQk2ROyagn1RDmKnNv8r0vcZhsItCIg/edit?usp=sharing.

Acknowledgements

We are grateful to our teacher Mikhail Rayko for the idea of the research, provided materials, generous support, constructive comments and helpful discussions.

References

[1] Tardigrades. (Sarah Bordenstein). Retrieved December 14, 2022, from https://serc.carleton.edu/microbelife/topics/tardigrade/index.html

[2] Carrero D, Pérez-Silva JG, Quesada V, López-Otín C. Differential mechanisms of tolerance to extreme environmental conditions in tardigrades. Sci Rep. 2019 Oct 17;9(1):14938. doi: 10.1038/s41598-019-51471-8. PMID: 31624306; PMCID: PMC6797769.

[3] Becquerel P. La suspension de la vie au-dessous de 1/20K absolu par démagnétisation adiabatique de l'alun de fer dans le vide le plus élevé. Compt. Rend. Acad. Sci. 231, 261–263 (1950).

[4] Hengherr S., Worland M. R., Reuner A., Brümmer F. & Schill R. O. High-temperature tolerance in anhydrobiotic tardigrades is limited by glass transition. Physiol. Biochem. Zool. 82, 749–755 (2009).

[5] Horikawa D. D. et al. Establishment of a rearing system of the extremotolerant tardigrade Ramazzottius varieornatus: a new model animal for astrobiology. Astrobiology 8, 549–556 (2008).

[6] Seki, Kunihiro; Toyoshima, Masato (1998). "Preserving tardigrades under pressure". Nature. 395 (6705): 853–54. doi: 10.1038/27576

[7] Ramløv H. & Westh P. Cryptobiosis in the eutardigrade Adorybiotus (Richtersius) coronifer: tolerance to alcohols, temperature and de novo protein synthesis. Zool. Anz. 240, 517–523 (2001).

[8] Hashimoto T, Horikawa DD, Saito Y, Kuwahara H, Kozuka-Hata H, Shin-I T, Minakuchi Y, Ohishi K, Motoyama A, Aizu T, Enomoto A, Kondo K, Tanaka S, Hara Y, Koshikawa S, Sagara H, Miura T, Yokobori SI, Miyagawa K, Suzuki Y, Kubo T, Oyama M, Kohara Y, Fujiyama A, Arakawa K, Katayama T, Toyoda A, Kunieda T. Extremotolerant tardigrade genome and improved radiotolerance of human cultured cells by tardigrade-unique protein. Nat Commun. 2016 Sep 20;7:12808. doi: 10.1038/ncomms12808. PMID: 27649274; PMCID: PMC5034306.

[9] Retrieved December 14, 2022, from http://tardigradesinspace.blogspot.com/

[10] Jönsson K. I., Rabbow E., Schill R. O., Harms-Ringdahl M. & Rettberg P. Tardigrades survive exposure to space in low Earth orbit. Curr. Biol. 18, R729–R731 (2008).

[11] Courtland, Rachel (8 September 2008). "'Water bears' are first animal to survive space vacuum". New Scientist. Retrieved 2011-05-22. https://www.newscientist.com/article/dn14690-water-bears-are-first-animal-to-survive-space-vacuum/

[12] Yoshida Y, Tanaka S. Deciphering the Biological Enigma-Genomic Evolution Underlying Anhydrobiosis in the Phylum Tardigrada and the Chironomid Polypedilum vanderplanki. Insects. 2022 Jun 19;13(6):557. doi: 10.3390/insects13060557. PMID: 35735894; PMCID: PMC9224920.

[13] 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.

[14] Horikawa DD, Cumbers J, Sakakibara I, Rogoff D, Leuko S, Harnoto R, Arakawa K, Katayama T, Kunieda T, Toyoda A, Fujiyama A, Rothschild LJ. Analysis of DNA repair and protection in the Tardigrade Ramazzottius varieornatus and Hypsibius dujardini after exposure to UVC radiation. PLoS One. 2013 Jun 6;8(6):e64793. doi: 10.1371/journal.pone.0064793. PMID: 23762256; PMCID: PMC3675078.

[15] Møbjerg N, Halberg KA, Jørgensen A, Persson D, Bjørn M, Ramløv H, Kristensen RM. Survival in extreme environments - on the current knowledge of adaptations in tardigrades. Acta Physiol (Oxf). 2011 Jul;202(3):409-20. doi: 10.1111/j.1748-1716.2011.02252.x. Epub 2011 Mar 22. PMID: 21251237.

[16] Beltrán-Pardo E, Jönsson KI, Harms-Ringdahl M, Haghdoost S, Wojcik A. Tolerance to Gamma Radiation in the Tardigrade Hypsibius dujardini from Embryo to Adult Correlate Inversely with Cellular Proliferation. PLoS One. 2015 Jul 24;10(7):e0133658. doi: 10.1371/journal.pone.0133658. PMID: 26208275; PMCID: PMC4514856.

[17] Wright JC. Desiccation tolerance and water-retentive mechanisms in tardigrades. J Exp Biol. 1989; 142: 267–292.

[18] Campbell LI, Rota-Stabelli O, Edgecombe GD, Marchioro T, Longhorn SJ, Telford MJ, Philippe H, Rebecchi L, Peterson KJ, Pisani D. 2011. MicroRNAs and phylogenomics resolve the relationships of Tardigrada and suggest that velvet worms are the sister group of Arthropoda. Proc Natl Acad Sci U S A 108:15920–4.

[19] Yoshida Y, Koutsovoulos G, Laetsch DR, Stevens L, Kumar S, Horikawa DD, Ishino K, Komine S, Kunieda T, Tomita M, et al. 2017. Comparative genomics of the tardigrades Hypsibius dujardini and Ramazzottius varieornatus. BioRxiv (doi:10.1101/112664).

[20] Gonzalo Giribet, Gregory D. Edgecombe, Current Understanding of Ecdysozoa and its Internal Phylogenetic Relationships, Integrative and Comparative Biology, Volume 57, Issue 3, September 2017, Pages 455–466, https://doi.org/10.1093/icb/icx072

[21] Hoff KJ, Stanke M. WebAUGUSTUS–a web service for training AUGUSTUS and predicting genes in eukaryotes. Nucleic Acids Res. 2013 Jul;41(Web Server issue):W123-8. doi: 10.1093/nar/gkt418. Epub 2013 May 21. PMID: 23700307; PMCID: PMC3692069.

[22] PhyloT: a phylogenetic tree generator [Internet]. [cited 2022 Dec 15]. Available from: phylot.biobyte.de

[23] RepeatModeler Download Page [Internet]. [cited 2022 Dec 15]. Available from: http://www.repeatmasker.org/RepeatModeler/

[24] Samtools [Internet]. [cited 2022 Dec 15]. Available from: https://www.htslib.org/

[25] Twelve years of SAMtools and BCFtools Petr Danecek, James K Bonfield, Jennifer Liddle, John Marshall, Valeriu Ohan, Martin O Pollard, Andrew Whitwham, Thomas Keane, Shane A McCarthy, Robert M Davies, Heng Li GigaScience, Volume 10, Issue 2, February 2021, giab008, https://doi.org/10.1093/gigascience/giab008

[26] WoLF PSORT: Protein Subcellular Localization Prediction [Internet]. [cited 2022 Dec 15]. Available from: https://wolfpsort.hgc.jp/

[27] TargetP - 2.0 - Services - DTU Health Tech [Internet]. [cited 2022 Dec 15]. Available from: https://services.healthtech.dtu.dk/service.php?TargetP-2.0

[28] BLAST: Basic Local Alignment Search Tool [Internet]. [cited 2022 Dec 15]. Available from: https://blast.ncbi.nlm.nih.gov/Blast.cgi

[29] Biosequence analysis using profile hidden Markov Models — HMMER [Internet]. [cited 2022 Dec 15]. Available from: https://www.ebi.ac.uk/Tools/hmmer/

[30] Chavez C, Cruz-Becerra G, Fei J, Kassavetis GA, Kadonaga JT. The tardigrade damage suppressor protein binds to nucleosomes and protects DNA from hydroxyl radicals. Elife. 2019 Oct 1;8:e47682. doi: 10.7554/eLife.47682. PMID: 31571581; PMCID: PMC6773438.

[31] Senft D, Qi J, Ronai ZA. Ubiquitin ligases in oncogenic transformation and cancer therapy. Nat Rev Cancer. 2018 Feb;18(2):69-88. doi: 10.1038/nrc.2017.105. Epub 2017 Dec 15. PMID: 29242641; PMCID: PMC6054770.

[32] Tumia R, Wang CJ, Dong T, Ma S, Beebe J, Chen J, Dong Z, Liu JY, Zhang JT. eIF3a Regulation of NHEJ Repair Protein Synthesis and Cellular Response to Ionizing Radiation. Front Cell Dev Biol. 2020 Aug 19;8:753. doi: 10.3389/fcell.2020.00753. Erratum in: Front Cell Dev Biol. 2021 Jan 07;8:629218. PMID: 32974334; PMCID: PMC7466773.

[33] Zhang S, Zhang H, Li H, Guo J, Wang J, Zhang L. Potential role of glucosamine-phosphate N-acetyltransferase 1 in the development of lung adenocarcinoma. Aging (Albany NY). 2021 Mar 3;13(5):7430-7453. doi: 10.18632/aging.202604. Epub 2021 Mar 3. PMID: 33686019; PMCID: PMC7993716.