# Tardigrades: from genestealers to space marines

Dudkovskaya A., Oshchepkova P.

**Abstract**

A tiny water bear with eight legs and covered in tough cales, which can endure conditions that don't even exist on Earth, such as vacuum, absolute zero and $\gamma$-radiation at doses between 1.0 – 9.0 kGy, is a tardigrade. Their extreme tolerance raise from the ability to dry without critical DNA damage, unlike other animals. The nature of preventing DNA distruction could unveil new DNA protection mechanism. Here we analysed tandem mass spectrometry data obtained from isolated chromatin fraction of Ramazzottius varieornatus to understand how tardigrades' DNA is keeping from radiation-induced damages. We find 5 proteins that can provide this protection function, and one of them is already described in Ramazzottius varieornatus as Damage suppressor protein (Dsup).

## 1 Introduction

Tardigrades are a species of microscopic invertebrates. They have four pairs of legs, their size is about 0,1—1,5 mm. Tardigrades have digestive, excretory, nervous and reproductive systems, however, their respiratory system is dermal. Tardigrades are of the greatest interest due to their incredible endurance. When unfavorable conditions occur, they are able to fall into a state of suspended animation for years, and when favorable conditions occur, they quickly revive. Tardigrades survive mainly due to the called anhydrobiosis, drying. They are able to stay alive after extreme temperature from -273 $C^{\circ}$ to 100 $C^{\circ}$, high pressure (7.5 GPa), exposure to high dose of irradiation and direct exposure to open space.

We are interested in causes of their endurance to high doses of exposure. There are already exist some articles about genome studies of Tardigrades [1], [2]. Most of them provide that there is no extensive Horizontal gene transfer.

We investigate the endurance of *Ramazzottius varieornatus*, one of the most stress-tolerant species of Tardigrades. We are interested in proteins. located in nuclear, because of their location we can suppose that their functions are to protect DNA from damages and repair it. We do functional annotation by looking for homologous proteins and conserved domains to attribute potential functions to the genes and proteins that we found.

## 2 Methods

**Data acquisition**

We use a sequence of the YOKOZUNA-1 strain of *of the Ramazzottius varieornatus*. All data is avaliable here.

**Structural annotation, physical localization prediction and BLAST search**

We use `AUGUSTUS` [3] for functional annotation. Then we extract protein sequences.
To find protein located in nuclear we analyzed the extracted proteins using tandem mass spectrometry. All data is avaliable here. Then we do a local alignment-based search using

`blast` [4]. We create a database and perform the search usind default options. To extract proteins of interest from the initial file we use `samtools` [5], esspesialy `faidx` utility.
For prediction of localization of our proteins from mass-spectrometer data we use `WoLF PSORT` [6]. Then we BLAST proteins that can possible locate in nuclear to find similar proteins in other organisms.

**Pfam prediction**

Also we try to predict the function of proteins even if we can't find orthologous sequences in databases with Blast. To do this we use `HMMER` [7] to search for our protein sequences. This tool use HMM profiles for various protein domains and motifs.

# 3 Results

After qualifying peptides from mass-spectrometry data and structural annotation by using `AUGUSTUS`, we find 34 unique proteins. Then we use `WoLFPSRORT` for localization prediction. Futher we research only proteins that have predicted localisation in nuclear with some score greater then zero. The results are provided in the Table 1.

Table 1: WoLF PSORT results

| Name | Wolf PSORT |
|---|---|
| g10513.t1 | nucl: 20, cyto_nucl: 14.5, cyto: 7, extr: 3, E.R.: 1, golg: 1 |
| g10514.t1 | nucl: 19, cyto_nucl: 15, cyto: 9, extr: 3, mito: 1 |
| g11513.t1 | cyto: 17, cyto_nucl: 12.8333, cyto_mito: 9.83333, nucl: 7.5, E.R.: 3, mito: 1.5, plas: 1, pero: 1, golg: 1 |
| g11806.t1 | nucl: 18, cyto_nucl: 11.8333, mito: 5, extr: 4, cyto: 3.5, cyto_pero: 2.66667, cysk_plas: 1 |
| g11960.t1 | nucl: 32 |
| g13530.t1 | extr: 13, nucl: 6.5, lyso: 5, cyto_nucl: 4.5, plas: 3, E.R.: 3, cyto: 1.5 |
| g14472.t1 | nucl: 28, plas: 2, cyto: 1, cysk: 1 |
| g15484.t1 | nucl: 17.5, cyto_nucl: 15.3333, cyto: 12, cyto_mito: 6.83333, plas: 1, golg: 1 |
| g16318.t1 | nucl: 20.5, cyto_nucl: 13, extr: 5, cyto: 4.5, E.R.: 1, golg: 1 |
| g16368.t1 | nucl: 20.5, cyto_nucl: 13, extr: 5, cyto: 4.5, E.R.: 1, golg: 1 |
| g2203.t1 | plas: 29, nucl: 2, golg: 1 |
| g3428.t1 | mito: 18, cyto: 11, extr: 2, nucl: 1 |
| g5443.t1 | extr: 28, nucl: 3, cyto: 1 |
| g5927.t1 | nucl: 30.5, cyto_nucl: 16.5, cyto: 1.5 |
| g7861.t1 | nucl: 16, cyto_nucl: 14, cyto: 8, plas: 5, pero: 1, cysk: 1, golg: 1 |
| g8100.t1 | nucl: 16.5, cyto_nucl: 12.5, cyto: 7.5, plas: 5, extr: 2, E.R.: 1 |
| g8312.t1 | nucl: 15.5, cyto_nucl: 15.5, cyto: 12.5, mito: 2, plas: 1, golg: 1 |

Further we use blast to find homology proteins and try to predict function of proteins. Here we choose only proteins that have other organisms, not Tardigrades. The results are provided in the Table 2.

Also we try to find some functional motifs, that maybe can explain the endurance of *Ramazzottius varieornatus* by using `HHMER`. `HHMER` found finctional motifs only in 8 proteins. Proteins names and prediction of their functionality are presented in the Table 3.

Table 2: BLAST results

| Name | Description | Scientific Name | Query Cover | E value | Per. Ident |
|---|---|---|---|---|---|
| g10513.t1 | Filaggrin-2 | Homo sapiens | 72% | 4,00E-21 | 43.69% |
| g10514.t1 | TATA-binding protein-associated factor 2N | Homo sapiens | 30% | 0.003 | 57.14% |
| g11513.t1 | Trafficking protein particle complex subunit 9 | Bos taurus | 27% | 3,00E-06 | 36.60% |
| g11806.t1 | RNA-binding protein FUS | Mus musculus | 70% | 3,00E-13 | 44.44% |
| g11960.t1 | E3 ubiquitin-protein ligase Bre1 | Drosophila melan | 47% | 1,00E-31 | 47.33% |
| g13530.t1 | 1,4-beta-N-acetylmuramidase 3 | Dictyostelium dis | 21% | 2,00E-07 | 45.08% |
| g14472.t1 | Non-homologous end joining protein Ku | Frankia casuarina | 14% | 3,00E-04 | 47.06% |
| g15484.t1 | Vacuolar protein sorting-associated protein 51 homolog | Xenopus laevis | 61% | 7,00E-105 | 57.82% |
| g16318.t1 | RNA-binding protein FUS | Mus musculus | 96% | 1,00E-16 | 42.91% |
| g16368.t1 | RNA-binding protein FUS | Mus musculus | 96% | 1,00E-16 | 42.91% |
| g2203.t1 | Myogenesis-regulating glycosidase | Mus musculus | 48% | 7,00E-46 | 48.50% |
| g3428.t1 | Myosin regulatory light chain 12A | Homo sapiens | 92% | 1,00E-70 | 59.63% |
| g5443.t1 | RING1 and YY1-binding protein B | Danio rerio | 28% | 0.38 | 50.00% |
| g5927.t1 | Probable glucosamine 6-phosphate N-acetyltransferase | Drosophila melan | 13% | 5,00E-15 | 43.31% |
| g7861.t1 | SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A-like protein 1 | Xenopus laevis | 53% | 4,00E-61 | 43.96% |
| g8100.t1 | Inositol monophosphatase 3 | Xenopus laevis | 11% | 4,00E-26 | 48.34% |
| g8312.t1 | Vacuolar protein sorting-associated protein 41 homolog | Mus musculus | 86% | 6,00E-135 | 42.05% |

Table 3: HHMER results

| Name | PFAM |
|---|---|
| g11513.t1 | Transport protein Trs120 or TRAPPC9, TRAPP II complex subunit |
| g11960.t1 | Zinc finger, C3HC4 type (RING finger) |
| g15484.t1 | "Vps51/Vps67 ; Exocyst complex component Sec5; Dor1-like family; Vacuolar-sorting protein 54, of GARP complex; COG (conserved oligomeric Golgi) complex component, COG2 |
| g2203.t1 | Glycosyl hydrolases family 31 |
| g3428.t1 | EF-hand domain |
| g7861.t1 | SNF2-related domain; HepA-related protein; Type III restriction enzyme, res subunit |
| g8100.t1 | Inositol monophosphatase family; Arf6-interacting domain of mitotic kinesin-like protein 1 |
| g8312.t1 | Region in Clathrin and VPS |

# 4 Discussion

During our research we find 17 proteins that may locate in nuclear of Tardigrades cells. We try to explore their functions by finding homology genes and functional motifs. So we have some predicted functions of these proteins. The results are provided in the Table 4.

Table 4: Predicted functions of proteins

| Name | Probable function from predictions |
|---|---|
| g10513.t1 | The filaggrin-like protein encoded by this gene is upregulated by calcium, proteolyzed by calpain 1, and is involved in epithelial homeostasis. |
| g10514.t1 | RNA and ssDNA-binding protein that may play specific roles during transcription initiation at distinct promoters. Can enter the preinitiation complex together with the RNA polymerase II (Pol II). |
| g11513.t1 | Component of the TRAPP II complex. TRAPP II seems to play a role in intra-Golgi transport (By similarity). |
| g11806.t1 | DNA/RNA-binding protein that plays a role in various cellular processes such as transcription regulation, RNA splicing, RNA transport, DNA repair and damage response. |
| g11960.t1 | E3 ubiquitin-protein ligase that mediates monoubiquitination of histone H2B to form H2BK123ub1 in association with the E2 enzyme RAD6/UBC2. |
| g13530.t1 | Can provide lysozyme activity. |
| g14472.t1 | Regulates the non-homologous end joining pathway choice of DNA double-strand break repair in human somatic cells. |
| g15484.t1 | Plays a role in vesicle-mediated protein trafficking to lysosomal compartments including the endocytic membrane transport pathways. |
| g16318.t1 | DNA/RNA-binding protein that plays a role in various cellular processes such as transcription regulation, RNA splicing, RNA transport, DNA repair and damage response. |
| g16368.t1 | DNA/RNA-binding protein that plays a role in various cellular processes such as transcription regulation, RNA splicing, RNA transport, DNA repair and damage response. |
| g2203.t1 | Enzyme that hydrolyses the glycosidic bond between two or more carbohydrates, or between a carbohydrate and a non-carbohydrate moiety. |
| g3428.t1 | The EF-hand consists of two alpha helices linked by a short loop region (usually about 12 amino acids) that usually binds calcium ions. EF-hands also appear in each structural domain of the signaling protein calmodulin and in the muscle protein troponin-C. |
| g5443.t1 | Component of a Polycomb group (PcG) multiprotein PRC1-like complex, a complex class required to maintain the transcriptionally repressive state of many genes, including Hox genes, throughout development. PcG PRC1-like complex acts via chromatin remodeling and modification of histones; it mediates monoubiquitination of histone H2A 'Lys-119', rendering chromatin heritably changed in its expressibility. |
| g5927.t1 | Catalytic Activity, acetyl-CoA + D-glucosamine 6-phosphate = CoA + H+ + N-acetyl-D-glucosamine 6-phosphate. |
| g8100.t1 | Catalyzes the hydrolysis of phosphoadenosine phosphate (PAP) to adenosine monophosphate (AMP). |

| | |
|---|---|
| g7861.t1 | ATP-dependent annealing helicase that binds selectively to fork DNA relative to ssDNA or dsDNA and catalyzes the rewinding of the stably unwound DNA. Rewinds single-stranded DNA bubbles that are stably bound by replication protein A (RPA). Acts throughout the genome to reanneal stably unwound DNA, performing the opposite reaction of many enzymes, such as helicases and polymerases, that unwind DNA. |
| g8312.t1 | Plays a role in vesicle-mediated protein trafficking to lysosomal compartments including the endocytic membrane transport pathways. |

Due to these results we have five proteins that are most likely responsible for DNA protection and reparation: g11806.t1, g14472.t1, g16318.t1, g16368.t1, g7861.t1. Thus these proteins are needed to be studied in more details.

Also, when we use `BLAST`, we find that one protein, g14472.t1, can be aligned to Tardigades protein Damage suppressor (Dsup). We found that there is already exists a research showing that it is a unique tardigrade protein that helps repair and protect DNA from radiation [8].

For further studies there is a classical way to test our findingse - gene knockout, but this approach could be tricky due to complex object, there are no established techniques for genetic engineering of tardigrades. Also, to detect that protein is located in the nucleus the ɯreen fluorescent protein may be used as a reporter gene.

You can see the labjournal here: https://docs.google.com/document/d/12lxdHMkuuMftkXd7eVSTjeitIS1J5_GC-XeOCUxC6lc/edit

# References


[1] Arakawa, K., Yoshida, Y. & Tomita, M. Genome sequencing of a single tardigrade Hypsibius dujardini individual. Sci Data 3, 160063 (2016). https://doi.org/10.1038/sdata.2016.63

[2] Kazuharu Arakawa, No evidence for extensive horizontal gene transfer from the draft genome of a tardigrade. BIOLOGICAL SCIENCES, May 12, 2016, 113 (22) E3057, https://doi.org/10.1073/pnas.1602711113

[3] K. J. Hoff and M. Stanke (2013) WebAUGUSTUS - a web service for training AUGUSTUS and predicting genes in eukaryotes Nucleic Acids Res, doi:10.1093/nar/gkt418

[4] Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990) "Basic local alignment search tool." J. Mol. Biol. 215:403-410.

[5] 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 (2021) Twelve years of SAMtools and BCFtools GigaScience 10(2), 2047–217X.

[6] Paul HORTON and Keun-Joon PARK at CBRC, Takeshi OBAYASHI, Kenta NAKAI, WoLF PSORT, Functional Analysis in silico | NAKAI Lab, https://wolfpsort.hgc.jp/

[7] Simon C Potter, Aurélien Luciani, Sean R Eddy, Youngmi Park, Rodrigo Lopez, Robert D Finn, HMMER web server: 2018 update, Nucleic Acids Research, Volume 46, Issue W1, 2 July 2018, Pages W200–W204, https://doi.org/10.1093/nar/gky448

[8] Hashimoto, T., Horikawa, D., Saito, Y. et al. Extremotolerant tardigrade genome and improved radiotolerance of human cultured cells by tardigrade-unique protein. Nat Commun 7, 12808 (2016). https://doi.org/10.1038/ncomms12808