**Combination of genomic and proteomic data in searching novel Tardigrade proteins responsible for radiation resistance**

Anna Churkina, Eugenia Ivanova (Strebulaeva)

**Abstract**

Some organisms during their evolution developed special mechanisms to sustain severe conditions such as radiation and dehydration. This paper represents the results of defining proteins that could be essential in the radiation resistance mechanism of *Ramazzottius varieornatus*. Using data from tandem mass spectrometry and annotated genome we obtained protein sequences of chromatin fraction. Several databases were exploited to collect information about localization and functional prediction. Evaluating different pieces of evidence, we revealed some proteins with nuclear localization that have no significant similarities with sequences in common protein databases that hypothetically could be involved in resistance mechanisms such as DNA-repairing. Also, we found a protein from different species which has prospective function for our investigation and directly binds DNA. Thus, the specter of the proteins for further experimental verification was narrowed.

**Introduction**

Tardigrades are microscopic (<1 mm) aquatic invertebrates, the hallmark of which is their ability to withstand extreme conditions such as temperature ranges from - 1K to 151°C [2], exposure to radiation [7],[12]; tardigrades can withstand about 1,200 atmospheres [13] and are also capable of life after finding an open space vacuum [9].

Due to the exceptional features of these tiny creatures, geneticists provided studies of their genome in order to find proteins responsible for damage protection. Boothby et al. in 2015 published the sensational information about massive horizontal gene transfer (HGT). They described about one-sixth tardigrade genes as acquired from other species [4]. Unfortunately, this finding was criticized by several letters of researchers. The researchers insisted on bacterial contamination and a significantly lower level of HGT. [1],[3],[11]. Nevertheless, tardigrades obviously possess the properties which may be reflected in their genome as particular genes either newly developed or obtained by HGT.

To study the genetic basis for the superpowers of tardigrades that provide such a high survival rate in this project were examined the genome of *Ramazzottius varieornatus*. For this, in addition to direct sequencing of the tardigrade genome, mass spectrometry of the chromatin fraction was carried out in order to determine the proteins that are associated with DNA. Proteins were identified from the entire list of peptides in the genome by aligning chromatin peptides, which were further investigated as candidate proteins providing superpowers of the tardigrade. Such proteins could participate in the repair of damaged DNA, as well as proteins involved in the organization of chromatin. It is assumed that the altered structure of chromatin, namely its condensed form, contributes to an increase in resistance to ionizing radiation. This is confirmed by the data of studies of the species Deinococcaceae, which indicate that these resistant species are characterized by the presence of a more condensed chromatin structure [14].

Thus, the genomic and proteomic analysis of *R. varieornatus*, in combination with the already accumulated data from the study of the genome of other tardigrade species, will improve the understanding of the mechanisms of resistance of organisms to extreme conditions of the external environment.

**Materials and methods**

To study the genome of *R. varieornatus*, we used sequencing data from R. varieornatus, the YOKOZUNA-1 strain by Illumina and Sanger sequencing data, provided in the University of Tokyo. The assembled genome is available [here](http://kumamushi.org/data/YOKOZUNA-1.scaffolds.fa)**.**

**AUGUSTUS v3.2.3** [15] were used to make functional annotations of proteins, and then prediction. Gff output file was used by the getAnnoFasta.pl to extract protein sequences (fasta). Tandem mass spectrometry has been used to study peptides that were associated with the DNA, the list of peptides can be downloaded [here](http://public.dobzhanskycenter.ru/mrayko/BIMM185/peptides.fa).

To find which proteins from the R. varieornatus genome these peptides correspond to, a database of these peptides was created with **makeblastdb command (BLAST 2.4.0+,** with flag -parse\_seqids**)**, local alignment by **blastp command (BLAST 2.4.0+,** with parameters by default**)** using peptide sequence file as a query. **Samtools v.1.11** [21] and **BioPython v.1.78 (Python 3.8)** were used to extract proteins that have been successfully aligned (*see Supplementary materials for the details of extraction*).

As long as subcellular localization of a protein can be defined by its signal peptides, two services were used. **WolF PSORT** [17] server provides information about cellular localization using features of the genes from several databases and gives results as weights for every defined feature [8]. Another service that can be used for searching signal peptides is **TargetP-2.0** [10], [18]. It spots three types of signal sequences: chloroplast transit peptide (cTP), mitochondrial targeting peptide (mTP) or secretory pathway signal peptide (SP). We used both services with default parameters, choosing only the organism type.

To search for homologous proteins **BLAST search** [16] was provided against the “UniProtKB/Swiss-Prot” database (with default parameters). To predict the function of the proteins **Pfam** [20]databasewas used where the search of the protein sequences performed against a collection of profile-HMMs for different protein domains and motifs.

**Results**

**Functional annotation**

After performing the functional genome annotation, we obtained 16435 protein sequences which is consistent with the average number in multicellular organisms (15-20k).

**Localization prediction**

Combining the data from tandem mass spectrometry and our protein sequences 34 hypothetically nuclear proteins (with 19 proteins with 100% coincidence of alignment) were obtained for further examination. Localization verification by WoLF PSORT and TargetP-2.0 revealed the presence of signal proteins in some of our examined proteins. The results can be seen in table 1 and 2 in Supplementary materials.

**Annotation search with BLAST**

We have found homologs for more than half of the examined proteins but several queries gave the same annotation result and almost all had low identity percent. Detailed results can be found in table 3 of Supplementary materials.

**Function prediction by with Pfam**

Most of the results from the Pfam database were consistent with BLAST search results.

Based on the WoLF PSORT report, namely the intracellular localization of the protein, 5 proteins were proposed as potential components of the mechanism of radiation resistance, the data on which are presented in Table 1.

Table 5 in Supplementary Materials includes the results of each of the queries on each server.

*Table 1 - Proteins-candidates for further investigation.*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Protein** | **BLAST** | **BLAST E-value** | **Pfam** | **WoLF PSORT** | **TargetP** | **% Identity (alignment)** |
| **g10513.t1** |  |  |  | nucl: 20, cyto\_nucl: 14.5, cyto: 7, extr: 3, E.R.: 1, golg: 1 |  | **100** |
| **g10514.t1** |  |  |  | nucl: 19, cyto\_nucl: 15, cyto: 9, extr: 3, mito: 1 |  | **100** |
| **g14472.t1** |  |  |  | nucl: 28, plas: 2, cyto: 1, cysk: 1 |  | **100** |
| **g11806.t1** |  |  |  | nucl: 18, cyto\_nucl: 11.8333, mito: 5, extr: 4, cyto: 3.5, cyto\_pero: 2.66667, cysk\_plas: 1 |  | **67** |
| **g7861.t1** | RecName: Full=SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A-like protein 1; AltName: Full=HepA-related protein; AltName: Full=Sucrose nonfermenting protein 2-like 1 [Rattus norvegicus] | 2e-71 | SNF2 family N-terminal domain | nucl: 16, cyto\_nucl: 14, cyto: 8, plas: 5, pero: 1, cysk: 1, golg: 1 |  | **63** |

**Discussion**

It is assumed that the reason for the resistance of organisms to extreme conditions is the presence of mechanisms of protection and restoration of the DNA structure under the influence of damaging factors, one of which is radiation.

Among the peptides from mass spectrometry that aligned to the amino acid sequence of *R. varieornatus* tardigrade, there are those belonged to the proteins that are involved in the process of organizing chromatin, and are also an essential part of the DNA repair system, which is necessary for the normal functioning of the organism. The presence of these proteins can provide super resistance to radiation and other factors that damage the genetic apparatus of the organism.

Remembering that we explored chromatin fraction of the tardigrade proteins in order to find proteins bound to DNA, in the first place we paid attention to those with nuclear localization. Four sequences had no significant similarities in all used databases and no other information except localization evidence with high weight of nuclear feature. This fact allowed us to presume these proteins as prospective candidates with unknown function that could be protection of the DNA.

In support of this theory we can offer SMARCAL1 (SWI/SNF-related, matrix-associated, actin-dependent regulator of chromatin, subfamily A-like1) as one of the potential proteins that could contribute to radiation resistance. This protein showed consistent E-value in BLAST alignment (2e-71). Despite the fact that the percentage of alignment of this peptide to the amino acid sequence of *R. varieornatus* is only 63%, this protein can be suggested for a more thorough study of its function. This protein maintains the genome integrity during DNA replication - helps to stabilize stalled replication forks and facilitate DNA repair during replication [19]. SMARCAL1 can bind to DNA that contains single- and double-stranded regions such as forks and DNA hairpins which promotes DNA single strand annealing and probably contributes to rapid reparation.

Summarizing the information, we would recommend all five proteins for further analysis because nuclear localization especially together with 100% identity in detection by mass spectrometry specify their possible relation to chromatin and binding to DNA.

As for proteins whose homologues have not been detected by blast, further in-depth study and prediction of the structure and function of this protein may be useful for investigating the causal factors of hyper-resistance to aggressive environmental influences.

Scientists have already performed investigation of the proteins in order to find those responsible for the radiation protection. DNA protection protein - Dsup (Damage suppressor) was described as tardigrade-unique protein, co-localized with nuclear DNA and provided radiation tolerance [5 - 6].

So, our suggestion was consistent with the real studies.

**References**

1. Arakawa K. No evidence for extensive horizontal gene transfer from the draft genome of a tardigrade. Proc Natl Acad Sci U S A. 2016;113(22):E3057. doi:10.1073/pnas.1602711113
2. Becquerel, P. La suspension de la vie au dessous de 1/20 K absolu par demagnetization adiabatique de l’alun de fer dans le vide les plus eléve. Comptes-rendus Hebdomadaires des Seances de l'Académie des Sciences de Paris 231, 261–263 (1950)
3. Bemm F, Weiß CL, Schultz J, Förster F. Genome of a tardigrade: Horizontal gene transfer or bacterial contamination? Proc Natl Acad Sci U S A. 2016;113(22):E3054-E3056. doi:10.1073/pnas.1525116113
4. Boothby TC, Tenlen JR, Smith FW, et al. Evidence for extensive horizontal gene transfer from the draft genome of a tardigrade [published correction appears in Proc Natl Acad Sci U S A. 2016 Sep 6;113(36):E5364]. Proc Natl Acad Sci U S A. 2015;112(52):15976-15981. doi:10.1073/pnas.1510461112.
5. 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
6. Hashimoto T, Kunieda T. DNA Protection Protein, a Novel Mechanism of Radiation Tolerance: Lessons from Tardigrades. Life (Basel). 2017;7(2):26. Published 2017 Jun 15. doi:10.3390/life7020026
7. Horikawa, D. D. et al. Radiation tolerance in the tardigrade Milnesium tardigradum. Int J Radiat Biol 82, 843–848 (2006).
8. Horton P, Park KJ, Obayashi T, et al. WoLF PSORT: protein localization predictor. Nucleic Acids Res. 2007;35(Web Server issue):W585-W587. doi:10.1093/nar/gkm259
9. Jonsson, 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).
10. José Juan Almagro Armenteros, Marco Salvatore, Ole Winther, Olof Emanuelsson, Gunnar von Heijne, Arne Elofsson, and Henrik Nielsen Detecting Sequence Signals in Targeting Peptides Using Deep Learning. Life Science Alliance 2 (5), e201900429. doi:[10.26508/lsa.201900429](https://doi.org/10.26508/lsa.201900429)
11. Koutsovoulos G, Kumar S, Laetsch DR, Stevens L, Daub J, Conlon C, Maroon H, Thomas F, Aboobaker AA, Blaxter M. No evidence for extensive horizontal gene transfer in the genome of the tardigrade Hypsibius dujardini. Proc Natl Acad Sci U S A. 2016 May 3;113(18):5053-8. doi: 10.1073/pnas.1600338113. Epub 2016 Mar 24. PMID: 27035985; PMCID: PMC4983863.
12. May, R.-M., Maria, M. & Guimard, J. Action differentielle des rayons X et ultraviolets sur le tardigrade Macrobiotus areolatus a l'etat actif et desseche. Bulletin biologique de la France et de la Belgique 48, 349–367 (1964).
13. Ono, F. et al. Effect of high hydrostatic pressure on to life of the tiny animal tardigrade. Journal of Physics and Chemistry of Solids 69, 2297–2300 (2008).
14. Zimmerman JM, Battista JR. A ring-like nucleoid is not necessary for radioresistance in the Deinococcaceae. BMC Microbiol. 2005 Mar 31;5:17. doi: 10.1186/1471-2180-5-17. PMID: 15799787; PMCID: PMC1079854.
15. <http://augustus.gobics.de>
16. <https://blast.ncbi.nlm.nih.gov/Blast.cgi>
17. <https://wolfpsort.hgc.jp>
18. <http://www.cbs.dtu.dk/services/TargetP/>
19. <http://www.ebi.ac.uk/interpro/entry/InterPro/IPR010003/>
20. <https://www.ebi.ac.uk/Tools/hmmer/>
21. <http://www.htslib.org>