



BIOINFORMATICS  
INSTITUTE

THE PEPTIDE THAT PROVIDES *R.varieornatus*  
WITH THE RESISTANCE TO UV RADIATION

December 9, 2023

**Project 4**

*by Aiusheeva AJ, Shitik EM*

# 1 Abstract

Tardigrades are of the most interest to scientists because they can withstand extreme environmental conditions such as desiccation, freezing, and radiation. Moreover, they can survive in open space. In this research, we predicted peptides in the assembled genome of *R.varieornatus* using ab initio methods and proposed which of them could provide the micro-animal resistance to UV radiation. As a result, we found 5 unannotated peptides with nuclear localization that could possibly be involved in DNA repair. Moreover, we identify one protein that was previously described to have an affinity to DNA.

# 2 Introduction

Tardigrades are microscopic animals that belong to the phylum Tardigrada within the protostome superclade Ecdysozoa. These eight-legged segmented micro-animals have the length of the body within the range of 50-1200 micrometers. Despite their small size, they are very tenacious and can be found across the whole world in permanent and temporary aquatic environments [1]. Tardigrades are of the most interest to scientists because they can withstand extreme environmental conditions such as desiccation [2], freezing, and radiation [3].

Furthermore, it has been shown that these micro-animals can go into open space. In 2008 tardigrades were confirmed to survive under the space vacuum and in 2011 they were shown to be resilient to microgravity and cosmic radiation [4]. Researchers always tried to find the causes of the extra capabilities of tardigrades by studying their genome. Initially, these micro-animals were proposed to obtain their peculiarities from other foreign organisms through horizontal gene transfer [5]. But, subsequently, this theory was rejected with careful repeating experiments [6].

Since tardigrades do not have phylogenetically close relatives, for the following analysis of their genome it was necessary to apply *ab initio* gene prediction. This approach implies using two types of sequence information: signal sensors and content sensors. The first one includes short sequence motifs, whereas the second one refers to the patterns of codons that are unique for particular species. Ab initio gene prediction is based on different algorithms such as Dynamic Programming, linear discriminant analysis, Linguist methods, Hidden Markov Model, and Neural Networks [7].

In this research, we have analyzed predicted peptides in the assembled genome of *Ramazzottiusvarieornatus* (YOKOZUNA-1 strain) and proposed which of them could provide the micro-animal resistance to UV radiation.

# 3 Materials and Methods

The assembled genome of *Ramazzottiusvarieornatus* (YOKOZUNA-1 strain) has been downloaded from this source:

- [ftp.ncbi.nlm.nih.gov/genomes/all/GCA/001/949/185/GCA01949185.1\\_Rvar4.0/GCA01949185.1\\_Rvar4.0](ftp.ncbi.nlm.nih.gov/genomes/all/GCA/001/949/185/GCA01949185.1_Rvar4.0/GCA01949185.1_Rvar4.0)

Initially, the repeats in the assembled genome were identified and were substituted with *N* ('hardmasking') using RepeatModeller v2.0.1 [8]. Further, the possible encoding genes in the processed genome of *R.varieornatus* were predicted using AUGUSTUS v3.2.0 [9]. The obtained annotation was converted from gff format to fasta with getAnnoFasta.pl script (see Supplementary material).

To restrict the search of peptides that are associated with DNA, the obtained annotation was compared to the list of proteins that were extracted using tandem mass spectrometry. This task was carried two programs: BLAST v2.6.0 [10]. The first step of running these programs was the creation of databases, and the second was the search for matches itself. The peptides that were

predicted and were discovered with tandem mass spectrometry were selected in the original fasta file using python pyfaidx 0.7.2.2 [11].

Subsequently, the selected peptides were analyzed according to three aspects. Firstly, their possible localization in a cell was predicted with WoLF PSORT [12] and TargetP v2.0 [13]. The second aspect included the search for homologous proteins in different foreign organisms with BLAST [14]. Thirdly, the set of selected peptides was analyzed to have similarities with motifs and domains in the Pfam database v36.0 [15] with HmmerWeb v2.43 [16].

## 4 Results

Overall, there were predicted 16435 sequences of peptides in the assembled genome of *R.varieornatus*. Having compared them with the results of tandem mass spectrometry, we selected 34 unique peptide sequences (total: 118) with BLAST [10]. Then we filtered peptide sequences by evalule and pident value, we got 24 (See Supplementary material).

According to TargetP v2.0 [13], 12 peptides had signal peptide motifs. Having analyzed all selected peptides with WoLF PSORT [12], we showed that 12 of them have nuclear localization. It is interesting to note that these 12 peptides correspond to a group 'other' in the results of TargetP v2.0 [13].

The results of searching for homologous proteins and motifs in BLAST [14] and Pfam databases [15] are presented in S\_Table1 (see Supplementary material). This table also contains the results of searching for peptide localization that is described above.

To sum up, 8 peptides were discovered to have nuclear localization without any homologues in the databases. Moreover, one peptide g14472.t1 was shown to be described previously as a Damage suppressor protein in *R.varieornatus*.

## 5 Discussion

With a high probability, the resistance of *R.varieornatus* to UV radiation is tightly connected with proteins that can efficiently repair damage to DNA. This type of peptide must be located in the nucleus and have possible motifs to bind DNA. Among annotated proteins that have relevant localization, none have a function related to DNA repair. They include intra-Golgi transport, vesicle transport, carbohydrate metabolism etc. (see Supplementary material). We would recommend paying attention to 4 unannotated peptides g10513.t1, g3428.t1, g5443.t1, g10514.t1 that have nuclear localization. Moreover, the g14472.t1 peptide attracted our greatest interest. It had also been shown to have no homologs in foreign organisms, but in 2016 scientists confirmed its affinity to DNA [17]. Considering our research and the results described in the mentioned paper, the peptide g14472.t1 takes the most important part in repairing DNA damage that is caused by UV radiation in *R.varieornatus*.

## 6 Supplementary material

- getAnnoFasta.pl
- S\_Table1. Annotation of predicted peptides
- Fasta file with 24 unique peptides that were selected with BLAST and filtering by evalule $\leq$ 10 and pident $\geq$ 70

## References

- [1] Nadja Møbjerg and Ricardo Cardoso Neves. “New insights into survival strategies of tardigrades”. In: *Comparative Biochemistry and Physiology Part A: Molecular amp; Integrative Physiology* 254 (Apr. 2021), p. 110890. ISSN: 1095-6433. DOI: 10.1016/j.cbpa.2020.110890. URL: <http://dx.doi.org/10.1016/j.cbpa.2020.110890>.
- [2] Thomas C. Boothby et al. “Tardigrades Use Intrinsically Disordered Proteins to Survive Desiccation”. In: *Molecular Cell* 65.6 (Mar. 2017), 975–984.e5. ISSN: 1097-2765. DOI: 10.1016/j.molcel.2017.02.018. URL: <http://dx.doi.org/10.1016/j.molcel.2017.02.018>.
- [3] K. Ingemar Jönsson. “Radiation Tolerance in Tardigrades: Current Knowledge and Potential Applications in Medicine”. In: *Cancers* 11.9 (Sept. 2019), p. 1333. ISSN: 2072-6694. DOI: 10.3390/cancers11091333. URL: <http://dx.doi.org/10.3390/cancers11091333>.
- [4] Erdmann Weronika and Kaczmarek Łukasz. “Tardigrades in Space Research - Past and Future”. In: *Origins of Life and Evolution of Biospheres* 47.4 (Oct. 2016), pp. 545–553. ISSN: 1573-0875. DOI: 10.1007/s11084-016-9522-1. URL: <http://dx.doi.org/10.1007/s11084-016-9522-1>.
- [5] Thomas C. Boothby et al. “Evidence for extensive horizontal gene transfer from the draft genome of a tardigrade”. In: *Proceedings of the National Academy of Sciences* 112.52 (Nov. 2015), pp. 15976–15981. ISSN: 1091-6490. DOI: 10.1073/pnas.1510461112. URL: <http://dx.doi.org/10.1073/pnas.1510461112>.
- [6] Georgios Koutsovoulos et al. “No evidence for extensive horizontal gene transfer in the genome of the tardigrade *Hypsibius dujardini*”. In: *Proceedings of the National Academy of Sciences* 113.18 (Mar. 2016), pp. 5053–5058. ISSN: 1091-6490. DOI: 10.1073/pnas.1600338113. URL: <http://dx.doi.org/10.1073/pnas.1600338113>.
- [7] Zhuo Wang, Yazhu Chen, and Yixue Li. “A Brief Review of Computational Gene Prediction Methods”. In: *Genomics, Proteomics amp; Bioinformatics* 2.4 (Nov. 2004), pp. 216–221. ISSN: 1672-0229. DOI: 10.1016/S1672-0229(04)02028-5. URL: [http://dx.doi.org/10.1016/S1672-0229\(04\)02028-5](http://dx.doi.org/10.1016/S1672-0229(04)02028-5).
- [8] *RepeatModeller*. <https://github.com/Dfam-consortium/RepeatModeler>.
- [9] *AUGUSTUS*. <https://github.com/Gaius-Augustus/Augustus>.
- [10] Grzegorz M. Boratyn et al. “BLAST: a more efficient report with usability improvements”. In: *Nucleic Acids Research* 41.W1 (Apr. 2013), W29–W33. ISSN: 0305-1048. DOI: 10.1093/nar/gkt282. URL: <http://dx.doi.org/10.1093/nar/gkt282>.
- [11] *pufaidx package*. <https://pypi.org/project/pyfaidx/>.
- [12] *WoLF PSORT*. <https://wolfpsort.hgc.jp>.
- [13] *TargetP*. <https://services.healthtech.dtu.dk/services/TargetP-2.0/>.
- [14] *BLAST database*. <https://blast.ncbi.nlm.nih.gov/Blast.cgi>.
- [15] *Pfam database*. <https://www.ebi.ac.uk/interpro/>.
- [16] *HmmerWeb*. <https://hmmer-web-docs.readthedocs.io/en/latest/changes.html>.
- [17] Takuma Hashimoto et al. “Extremotolerant tardigrade genome and improved radiotolerance of human cultured cells by tardigrade-unique protein”. In: *Nature Communications* 7.1 (Sept. 2016). ISSN: 2041-1723. DOI: 10.1038/ncomms12808. URL: <http://dx.doi.org/10.1038/ncomms12808>.