BIOINFORMATICS INSTITUTE

Bio-informatics Practice

December 9, 2023

Tardigrades genome study: in search of new DNA repair genes

by Artem Vasilev and Tatiana Lisitsa

Abstract

The extraordinary resistance of tardigrades to extreme conditions has been interested researchers for a long time. After the hypothesis about the significant influence of horizontal gene transfer (HGT) was refuted, it became clear that a more thorough analysis of genomic and proteomic data was needed to search for genes responsible for DNA repair. To carry it out, gene prediction and sequence homology search were done using modern bioinformatics approaches. The result was a list of proteins that are potentially associated with DNA damage repair in tardigrades: g11960, g7861 and g14472.

Keywords: Tardigrade, *Ramazzottius varieornatus*, DNA repair, gene prediction, sequence homology

Introduction

Tardigrades, also known as water bears, are microscopic animals found worldwide. Some tardigrade species exhibit extraordinary tolerance to various physical extremes, including almost complete dehydration, extreme levels of ionizing radiation, large fluctuations in external salinity, and the ability to avoid freezing by cooling to below −20 °C. [1] Dehydrated tardigrades can withstand a wide range of physical extremes that would typically be lethal for most organisms. These include extreme temperatures (ranging from −273 °C to nearly 100 °C), high pressures (up to 7.5 GPa), immersion in organic solvents, exposure to high doses of radiation, and even direct exposure to outer space [2]. *R. varieornatus* are among the most extreme-tolerant species of tardigrades.

Such resilience may be attributed to the unique features of the tardigrade genetic apparatus, including the presence of distinct genes responsible for DNA damage repair. While previous studies have indicated that tardigrades can acquire genes through horizontal gene transfer [3], further investigations have not established a significant role for this mechanism in acquiring resistance to extreme environmental conditions [4]. The aim of our work is to identify and annotate tardigrade-specific genes responsible for DNA repair.

Gene prediction, also known as gene annotation, is the process of identifying the location and structure of protein-coding genes within a DNA sequence. This computational process involves using bioinformatics tools and algorithms to analyze genomic data, identifying regions that likely encode functional proteins. Gene prediction considers various features, such as open reading frames (ORFs), codon usage, and splice sites, to distinguish protein-coding genes from non-coding regions in the genome [5].

Sequence homology refers to the similarity between the nucleotide or amino acid sequences of biological molecules in different species or within the same organism. Homologous sequences share a common ancestry and are often functionally related. Sequence homology analysis involves comparing sequences to identify similarities, differences, and evolutionary relationships. Bioinformatics tools, e.g. BLAST, are commonly used for sequence homology searches, helping uncover conserved regions and infer functional relationships between genes or proteins.

Materials and methods

For this project a sequence of the *Ramazzottius varieornatus* (YOKOZUNA-1 strain) was used. We worked on the assembled genome due to technical limitations (all links are available on GitHub in Snakefile). Also we used precomputed AUGUSTUS results to extract protein sequences. For further analysis, we borrowed data from our colleagues who performed tandem mass spectrometry to obtain peptides associated with the DNA (chromatin fraction). To obtain sequences in the *R. varieornatus* genome that match the found peptides, it was decided to use local alignment-based search with BLAST+ [6]. Next, using the samtools faidx utility [7], only the protein sequences of interest to us were selected. To predict the localization of the found

sequences, 2 programs WoLF PSORT [8] and TargetP Server [9] were used. After that, 13 found sequences were analyzed using the BLAST (UniProtKB/Swiss-Prot database)[10] and HMMER (hmmscan tool and the Pfam database) [11] web tools.

Results

The results are summarized in Table 1.

Table 1: Description of found protein sequences

FASTA ID	E-value	Accession	Pfam domains pred.	WoLF loc.	TargetP loc.
g11513	7e-83	Q32PH0.1	TRAPPC9-Trs120	cyto	OTHER
g11960	6e-98	Q8CJB9.1	Zinc finger	nucl	OTHER
g14472	0.0	P0DOW4.1	–	nucl	OTHER
g15484	0.0	Q155U0.1	Vps51/Vps67	cyto_nucl	OTHER
g5927	1e-18	Q17427.1	–	nucl	OTHER
g7861	2e-71	B4F769.1	SNF2-related domain	cyto_nucl	OTHER
g8100	3e-46	Q2YDR3.1	Inositol monophosphatase family	cyto_nucl	OTHER
g8312	0.0	Q5KU39.1	Region in Clathrin and VPS	cyto_nucl	OTHER

Discussion

The g11513 protein, which shares similarity with the NIK- and IKBKB-binding protein from *B. taurus*, *H. sapiens*, *M. musculus*, and others, contains TRAPPC9-Trs120 domains and corresponds to a protein known as TRAPPC9. TRAPPC9 serves as a subunit of the TRAPP II (Transport Protein Particle II) complex, occasionally referred to as Trs120. The TRAPP complexes play a crucial role in membrane trafficking within cells. It's worth noting that TRAPPC9, or Trs120, appears not to be involved in DNA repair processes.

The g15484 protein shows homology with the Vacuolar protein sorting-associated protein 51 homolog from *D. rerio*, *X. laevis*, and *X. tropicalis*, which is part of the GARP (Golgi-associated retrograde protein) complex. This protein contains the Vps51/Vps67 domain and other domains, confirming the assumption that it is involved in protein-protein and protein-membrane interactions but not in DNA repair.

The g8312 protein is similar to the Vacuolar protein sorting-associated protein 41 homolog from *M. musculus*, *H. sapiens*, *D. rerio*, and others. It contains the Region in Clathrin and VPS. Clathrin is a protein that plays a key role in endocytosis, and VPS is associated with the sorting and transport of proteins into cell vacuoles.

The g5927 protein shares similarity with Glucosamine 6-phosphate N-acetyltransferase from *C. elegans*, *D. melanogaster*, *P. abelii*, and *M. musculus*, but only in 14-17% of its sequence. Glucosamine 6-phosphate N-acetyltransferase is an enzyme that facilitates the transfer of an acetyl group from acetyl-CoA to the primary amine in glucosamide-6-phosphate, resulting in the production of free CoA and N-acetyl-D-glucosamine-6-phosphate.

The g8100 protein shows similarity with Inositol monophosphatase 3 from *D. rerio*, *D. pseudoobscura*, *D. melanogaster*, and others in 22-25% of its sequence. This protein contains the Inositol monophosphatase family domain and the Arf6-interacting domain of mitotic kinesin-like protein 1. Proteins of the Inositol monophosphatase 3 family are presumably localized in the Golgi complex.

The g11960 protein shares homology with the E3 ubiquitin protein ligase BRE1B from *R. norvegicus*, *H. sapiens*, *M. musculus*, and others and features a C3HC4-type zinc finger domain. The C3HC4-type zinc finger (RING finger) is a versatile domain engaged in diverse cellular

functions, encompassing ubiquitination, mediation of protein-protein interactions, regulation of gene expression, and participation in DNA repair processes.

The g7861 protein shows homology with SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A-like protein 1 (SMARCAL1) from *R. norvegicus*, *H. sapiens*, *X. tropicalis*, and others. It contains the SNF2-related domain and HepA-related protein motif. SMARCAL1 is a member of the SWI/SNF-like family of matrix-associated, actin-dependent chromatin regulators. This protein plays an important role in regulating the structure and accessibility of chromatin, participating in DNA repair processes, and genome maintenance [12] [13].

Protein g14472 was previously described as a Damage suppressor protein of *R. varieornatus* [2] [14].

So, we can identify three proteins supposed to be involved in the DNA repair process in tardigrades: g11960, g7861 and g14472. Further research is required to confirm our assumptions.

Supplementary materials

GitHub

References

[1] Nadia Møbjerg et al. "Survival in extreme environments–on the current knowledge of adaptations in tardigrades". In: *Acta physiologica* 202.3 (2011), pp. 409–420.

[2] Takuma Hashimoto et al. "Extremotolerant tardigrade genome and improved radiotolerance of human cultured cells by tardigrade-unique protein". In: *Nature communications* 7.1 (2016), p. 12808.

[3] 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 (2015), pp. 15976–15981.

[4] 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 (2016), pp. 5053–5058. DOI: 10.1073/pnas.1600338113. eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.1600338113. URL: https://www.pnas.org/doi/abs/10.1073/pnas.1600338113.

[5] Zhuo Wang, Yazhu Chen, and Yixue Li. "A brief review of computational gene prediction methods". In: *Genomics, proteomics & bioinformatics* 2.4 (2004), pp. 216–221.

[6] Christiam Camacho et al. "BLAST+: architecture and applications". In: *BMC Bioinformatics* 10.1 (Dec. 2009), p. 421.

[7] Petr Danecek et al. "Twelve years of SAMtools and BCFtools". In: *GigaScience* 10.2 (Feb. 2021). giab008. ISSN: 2047-217X. DOI: 10.1093/gigascience/giab008. eprint: https://academic.oup.com/gigascience/article-pdf/10/2/giab008/36332246/giab008.pdf. URL: https://doi.org/10.1093/gigascience/giab008.

[8] Paul Horton et al. "WoLF PSORT: protein localization predictor". en. In: *Nucleic Acids Res* 35.Web Server issue (May 2007), W585–7.

[9] Jose Juan Almagro Armenteros et al. "Detecting sequence signals in targeting peptides using deep learning". en. In: *Life Sci Alliance* 2.5 (Sept. 2019).

[10] Stephen F Altschul et al. "Basic local alignment search tool". In: *Journal of molecular biology* 215.3 (1990), pp. 403–410.

[11] Simon C Potter et al. "HMMER web server: 2018 update". In: *Nucleic Acids Research* 46.W1 (June 2018), W200–W204. ISSN: 0305-1048. DOI: 10.1093/nar/gky448. eprint: https://academic.oup.com/nar/article-pdf/46/W1/W200/25110422/gky448.pdf. URL: https://doi.org/10.1093/nar/gky448.

[12] Matthew A Coleman, Jonathan A Eisen, and Harvey W Mohrenweiser. "Cloning and characterization of HARP/SMARCAL1: a prokaryotic HepA-related SNF2 helicase protein from human and mouse". In: *Genomics* 65.3 (2000), pp. 274–282.

[13] Daniel P Ryan and Tom Owen-Hughes. "Snf2-family proteins: chromatin remodellers for any occasion". In: *Current opinion in chemical biology* 15.5 (2011), pp. 649–656.

[14] Carolina Chavez et al. "The tardigrade damage suppressor protein binds to nucleosomes and protects DNA from hydroxyl radicals". In: *Elife* 8 (2019), e47682.