Deciphering the Genome for Insights into Stress Tolerance Proteins of *Ramazzottius varieornatus* Tardigrade

M. Uzun[a,c], A. Fedorenko[b,c]

[a]*Research Center Biotechnology RAS, Moscow, Russia*
[b]*Skolkovo Institute of Science and Technology, Moscow, Russia*
[c]*Bioinformatics Institute, Saint Petersburg, Russia*

Abstract

Tardigrades, microscopic organisms renowned for their tolerance to extreme conditions, intrigue researchers due to their survival in high or subzero temperatures, high pressure, UV radiation exposure, and even the vacuum of space. The genome sequencing of *Ramazzottius varieornatus*, a stress-tolerant tardigrade species, aimed to unravel the genetic mechanisms responsible for this exceptional resistance. Using computational strategies for novel gene identification, this study delineates stress-related protein candidates from the unannotated genome. The analysis unveils a comprehensive genome of *R. varieornatus*, emphasizing protein sequence annotations and characterizing DNA-binding proteins associated with stress tolerance. These findings present insights into the genetic mechanisms of tardigrades' stress resilience, pinpointing proteins meriting further exploration of stress response mechanisms.

Keywords: Tardigrades, Stress Tolerance, Genome annotation

1. Introduction

Tardigrades, also known as water bears, are unique microscopic organisms that are interesting for their resistance to extreme conditions [1]. Habiting a variety of places such as marine and freshwater or a moist terrestrial environment they endure extreme temperatures ranging from absolute zero to close to boiling [2,3], high pressure [4], exposure to UV radiation[5], and even the vacuum of outer space [6]. The molecular mechanisms of these extraordinary capabilities have intrigued researchers for years hence such harsh conditions typically prove fatal to other organisms.

The genome sequencing of *Ramazzottius varieornatus* may shed light on the potential genetic secrets by using computational identification of novel genes. Applying diverse techniques reliant on comparative genomics or detecting against databases allows to find and annotate novel genes [7]. This approach is still challenging due to the absence of a gold standard and benchmark criteria. Despite this, the examination of gene sequence homology enables the establishment of associative links among genes from distinct genomes [8].

This study delves into the analysis of the unannotated genome of *R. varieornatus*, one of the most stress-tolerant tardigrade species. We detected candidates of the stress-related proteins, and analyzed their molecular functions for further wet-lab investigation.

2. Materials and Methods

2.1. Data Availability

The genome of the *Ramazzottius varieornatus* YOKOZUNA-1 strain can be found in NCBI Genbank (accession number GCA_001949185.1).

The annotation of the assembled genome was obtained in AUGUSTUS and is accessible through this link: https://u.to/mr4qIA.

A list of peptides associated with the DNA-extracted proteins obtained using tandem mass spectrometry can be found here: https://disk.yandex.ru/d/xJqQMGX77Xueqg.

2.2. Used Tools

A local alignment-based search was conducted using BLAST+ v2.15.0 [9] and diamond v2.1.8 [10]. Proteins found after the local blast were also searched against the "UniProtKB/Swiss-Prot" database integrated into the NCBI blast. The prediction of the subcellular localization of proteins was conducted using WoLF PSORT [11] (selected "nucl" localization parameter and score ¿=19), and TargetP [12,13] (selected proteins without signal functions (non-SP)) online tools. The prediction of the protein functions in cases when there is no option to find orthologous sequences in databases with Blast was conducted using HMMER [14], which helps to search protein sequences against a collection of profile-HMMs for different protein domains and motifs. The quality metrics were assessed using QUAST v5.0.2 [15]. All tools were run with default parameters.

3. Results

The analyzed genome of *Ramazzottius varieornatus* YOKOZUNA-1 spanned 55.8 Mbp (200 scaffolds; N50 - 4.74 Mbp; N90 - 1.3 Mbp). The assembly's annotation comprised 16,435 protein sequences. Mass spectrometry analysis of

DNA-binding proteins yielded 43 peptides. A local blast of the annotated genome against the identified peptides resulted in 34 protein sequences putatively associated with tardigrades' DNA, playing a crucial role in stress tolerance. Subsequent to this discovery, these proteins underwent detailed analyses, and those deemed most suitable for stress tolerance roles were selected (Supplementary Table 1).

4. Discussion

Based on the selection strategy and BLAST results we chose 5 candidate proteins potentially involved in stress tolerance of *R. varieornatus*. One of them, Dsup (UniProt accession number P0DOW4), was firstly described in 2016 by T. Hashimoto as a damage suppressor protein localizing near the nuclear DNA [16]. This protein interacts with a free DNA molecule by featuring the nucleosome-binding domain region that facilitates their binding to nucleosomes. *In vitro* experiments also demonstrate chromatin protection from hydroxyl radical-induced cleavage, pointing out their conserved role in preserving chromosomal DNA integrity due to the harsh conditions [17].

Besides, we found two proteins (UniProt accession numbers GAV01971 and GAV01972) previously described in Hashimoto et al., 2016. However, their role and functions remain still unknown. The last two found proteins (UniProt accession numbers GAV09945 and GAV09913) are not characterized by this moment.

The selected candidate proteins must undergo rigorous molecular-genetic validation to ascertain their true involvement in stress tolerance. For instance, employing RNA sequencing (RNA-seq) can unveil dynamic shifts in gene expression profiles in response to stress conditions. Alternatively, a thorough exploration of GFP-fused protein localization under stress conditions can shed light on their spatial dynamics. Moreover, additional investigations into other genetic mechanisms governing tardigrade stress resistance are warranted. For instance, executing RNA interference (RNAi) experiments to selectively suppress the expression of stress-tolerance-associated genes or delving into the realm of epigenetic modifications, such as DNA methylation or histone modifications, under stress conditions, can provide valuable insights.

References

[1] Møbjerg, N. et al. Survival in extreme environments - on the current knowledge of adaptations in tardigrades. *Acta Physiol.* (Oxf). 202, 409–420 (2011).

[2] Neves, R. C., Hvidepil, L. K. B., Sørensen-Hygum, T. L., Stuart, R. M. & Møbjerg, N. Thermotolerance experiments on active and desiccated states of Ramazzottius varieornatus emphasize that tardigrades are sensitive to high temperatures. *Sci. Rep.* 10, 1–12 (2020).

[3] Hengherr, S., Worland, M. R., Reuner, A., Brümmer, F. & Schill, R. O. Freeze tolerance, supercooling points and ice formation: Comparative studies on the subzero temperature survival of limno-terrestrial tardigrades. *J. Exp. Biol.* 212, 802–807 (2009).

[4] Ono, F. et al. Effect of ultra-high pressure on small animals, tardigrades and Artemia. *Cogent Phys.* 3, (2016).

[5] Jönsson, K. I. Radiation tolerance in tardigrades: Current knowledge and potential applications in medicine. *Cancers (Basel)* 11, (2019).

[6] Persson, D. et al. Extreme stress tolerance in tardigrades: Surviving space conditions in low earth orbit. *J. Zool. Syst. Evol. Res.* 49, 90–97 (2011).

[7] Wang, Z., Chen, Y. & Li, Y. A brief review of computational gene prediction methods. *Genomics, proteomics Bioinforma. / Beijing Genomics Inst.* 2, 216–221 (2004).

[8] Klasberg, S., Bitard-Feildel, T. & Mallet, L. Computational identification of novel genes: Current and future perspectives. *Bioinform. Biol. Insights* 10, 121–131 (2016).

[9] Camacho, C. et al. BLAST+: Architecture and applications. *BMC Bioinformatics* 10, 1–9 (2009).

[10] Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60 (2014).

[11] https://wolfpsort.hgc.jp/.

[12] Armenteros, J. J. A. et al. Detecting sequence signals in targeting peptides using deep learning. *Life Sci. Alliance* 2, 1–14 (2019).

[13] https://services.healthtech.dtu.dk/services/TargetP-2.0/.

[14] https://www.ebi.ac.uk/Tools/hmmer/.

[15] Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075 (2013).

[16] Hashimoto, T. et al. Extremotolerant tardigrade genome and improved radiotolerance of human cultured cells by tardigrade-unique protein. *Nat. Commun.* 7, (2016).

[17] Chavez, C., Cruz-Becerra, G., Fei, J., Kassavetis, G. A. & Kadonaga, J. T. The tardigrade damage suppressor protein binds to nucleosomes and protects dna from hydroxyl radicals. *Elife* 8, 1–20 (2019).

Supplementary materials

Supplementary Table 1: Proteins potentially involved in stress tolerance.

Protein	Best blast hit to peptide	e-value	Description	Probable localization(s) in WoLF PSORT	Localization in TargetP
g10513.t1	9	0.003	No significant similarity found	g10513.t1 details nucl: 20, cyto$_n$*ucl* : 14.5, *cyto* : 7, *extr* : 3, *E.R.* : 1, *golg* : 1	OTHER
g10514.t1	38	0.6	No significant similarity found	g10514.t1 details nucl: 19, cyto$_n$*ucl* : 15, *cyto* : 9, *extr* : 3, *mito* : 1	OTHER
g14472.t1	7	0.002	Damage suppressor protein	g14472.t1 details nucl: 28, plas: 2, cyto: 1, cysk: 1	OTHER
g16318.t1	9	4.9	No significant similarity found	g16318.t1 details nucl: 20.5, cyto$_n$*ucl* : 13, *extr* : 5, *cyto* : 4.5, *E.R.* : 1, *golg* : 1	OTHER
g16368.t1	9	5.2	No significant similarity found	g16368.t1 details nucl: 20.5, cyto$_n$*ucl* : 13, *extr* : 5, *cyto* : 4.5, *E.R.* : 1, *golg* : 1	OTHER