Practice

BIOINFORMATICS INSTITUTE

Functional annotation of Tardigrade genome assembly reveals potential radiation-resistance genes

Valeriya Rubinova, Anna Chechenina

Abstract

It is well-known that high-dose of radiation could cause severe DNA lesions, including double-strand breaks, and leads to genome instability and less viability. However, microscopic eukaryotes from the *Tardigrade* group have a surprisingly high level of the radio-resistance and can withstand several thousand Gray units of ionizing radiation when the lethal dose for mammals is less than ten Hashimoto and Kunieda (2017). In this study, we performed the functional analysis of the annotated genome assembly of *Ramazzottius varieornatus* to find genes that could play important roles in the radio-resistance. We looked for the DNA-biding proteins that are localized in the nucleus and found several potentially essential proteins. One of them was the Damage suppressor protein, which was already described by multiple studies as a key regulator of DNA radiation resistance.
Supplementary materials can be found via the following link: https://github.com/checheanya/BI_git/tree/main/HW4.

Keywords: Tardigrades; radiation; DNA reparation; functional annotation

Introduction

Tardigrades are microscopic animals found worldwide in aquatic as well as terrestrial ecosystems. They are renowned for their ability to tolerate a variety of extreme conditions, including desiccation, severe osmotic shock, freezing in liquid nitrogen, and even exposure to space vacuum, cosmic radiation and complete desiccation for up to a decade Møbjerg *et al.* (2011); Møbjerg and Neves (2021). Cryptobiosis seems to provide animals with the potential to survive such conditions that are far beyond any constraints set by their normal environment Møbjerg and Neves (2021).

Cryptobiosis challenges our perception of the transition between life and death of an organism. Understanding the mechanisms that underlie the ability to stabilize biological structures, from macromolecules across cellular, tissue and organ levels to the whole animal, and subsequently restart life after years of metabolic suspension has great potential for translational and applied sciences. As new molecular tools allow for increasingly detailed investigations, tardigrades are indeed gaining attention by physiologists and biochemists Møbjerg and Neves (2021).

There are many different studies that predict the proteins involved in stress tolerance. For example, in the study Boothby *et al.* (2017) it was found that tardigrades express many tardigrade intrinsically disordered proteins (TDPs) in response to drying. Other studies Hashimoto *et al.* (2016); Hashimoto and Kunieda (2017) have found that the tardigrade unique DNA-associated protein, *Dsup*, improves radiotolerance. The reduced DNA fragmentation in Dsup-expressing cells was likely due to the reduced occurrence of DNA breaks rather than facilitation of the DNA repair process.

In our study, we also predicted several proteins most likely involved in DNA protection and repair in tardigrades. In order to identify these proteins, we first assembled their genome, then we were looking for homologous proteins and conserved domains to assign potential functions to the genes and proteins we find. In the next step, we selected only those proteins that make up the chromatin fraction. And they have only studied a group of proteins that are localized in the nucleus.

Methods

In our study we used a genome assembly for the *Ramazzottius varieornatus*, the YOKOZUNA-1 strain obtained from the NCBI library here. Functional annotation was made using the AUGUSTUS tool Stanke and Morgenstern (2005). To analyze protein sequences found in the annotation, we converted the gtf file to the fasta format.

In the previously done experimental work, our colleagues performed nucleic proteins isolation and extraction assay. Then these samples were analyzed the using tandem mass spectrometry method. This method allows to get peptides as parts of the extracted proteins involved in the interaction with chromatin. Using the list of these proteins we derived full sequences of these proteins of interest from the list of annotated proteins from the previous step. In order to do this we used blastp Altschul *et al.* (1990) and Diamond Buchfink *et al.* (2021) tools that build local alignments using the entry sequences and the provided library of sequences. Thus, in the first step, we built the database using our full annotation file and then ran local alignments. However, it turned out that the Diamond was not capable of handling sequences shorter than 12 amino acids, so we got only four hits compared to the 118 from blastp.

To identify which proteins among the ones we found are known to localize in the nucleus and possibly interact with DNA we used WoLF PSORT Horton *et al.* (2007) and TargetP 2.0 Armenteros *et al.* (2019) servers.

In the end, we performed searches in the BLAST and HMMER servers to find any orthologs to the proteins of interest. Results could be found in the Tables 1-2 in the Supplementary

section.

Results

Since the aim was to identify which proteins could play key roles in radiation resistance, we were mainly interested in the DNA-interacting proteins. There are 16435 proteins in total in the annotation of the *Ramazzottius varieornatus* genome assembly. At the same time, 43 peptides were obtained using tandem mass spectrometry. The local alignment results showed that there are 97 unique genes that could have significant similarities with these peptides.

We showed that among found proteins there are 22 proteins that were classified by TargetP 2.0 as a signal, while others were neither signal nor mitochondrial. WoLF PSORT annotated 15 proteins as possibly located in the nucleus (for the full list see Tables 1-2 in the Supplementary section). We took these proteins for further investigation.

Four of these 15 proteins did not have any significant orthologs in the UniprotKB/Swiss-Prot databases, possibly because of their very short sequence with many unidentified positions. However, we found some proteins with orthologues that might be related to the DNA reparation: first of all, Damage suppressor protein (gene *g*14472.*t*1), Eukaryotic translation initiation factor 3 subunit A (eIF3a, gene *g*16318.*t*1), and an E3 ubiquitin-protein ligase BRE1B (gene *g*11960.*t*1) which has ZINC-finger in the structure so it can bind nucleotides.

Moreover, we found out that some of the found proteins are very similar to plant proteins. For example, g8312.t1 has more than 30% identity with more than 80% coverage with vacuolar protein sorting-associated protein 41 homologs in different plants (*Malus sp.*, *Gossypium sp.*, *Carex sp.*). Or g16368.t1, g16318.t1 has high similarity to some proteins of *Punica granatum*. However, these findings mostly support just the common organization of the eukaryotic cells and do not make any significant insights into the horizontal gene transfer discussion Koutsovoulos *et al.* (2016).

Discussion

We have identified a protein that most likely plays a major role in DNA repair in tardigrades - Damage suppressor protein (*Dupp*), *g*14472.*t*1. There is a study that also confirms the role of this protein in DNA repair Hashimoto *et al.* (2016). One of the proteins they studied (*Dsup*), was co-localized with nuclear DNA, and similar co-localization was also observed in human cultured HEK 293T cells. Their transcriptome data revealed abundant expression of Dsup in an early embryonic stage, which is consistent because nuclear DNA extensively replicates in the embryonic stage. To verify the localization of *Dsup* protein in tardigrade cells, they performed immunohistochemistry with frozen sections of tardigrade embryos. In almost all tardigrade cells expressing *Dsup*, *Dsup* proteins co-localized with nuclear DNA Hashimoto *et al.* (2016). Also using human cultured cells, it was demonstrated that a tardigrade-unique DNA-associating protein suppresses X-ray-induced DNA damage by $\approx$ 40 and improves radiotolerance Hashimoto *et al.* (2016).

Both ionizing radiation and hydrogen peroxide generate hydroxyl radicals as a major ROS product in cells. It was shown that apart from the ability to non-specifically interact with free DNA Hashimoto *et al.* (2016) Dsup and Dsup-like proteins can also protect chromatin from cleavage by hydroxyl radicals Chavez *et al.* (2019). Dsup is a highly basic protein, especially in its C-terminus, so it is highly prone to DNA binding. Dsup has a structure with high level of intrinsic disorder, which combined with strong electrostatic attraction to DNA, promotes the formation if the flexible aggregates of Dsup and DNA Mínguez-Toral *et al.* (2020).

Component of the RNF20/40 E3 ubiquitin-protein ligase complex mediates monoubiquitination of 'Lys-120' of histone H2B (H2BK120ub1). H2BK120ub1 gives a specific tag for epigenetic transcriptional activation and is also prerequisite for histone H3 'Lys-4' and 'Lys-79' methylation (H3K4me and H3K79me, respectively). It thereby plays a central role in histone code and gene regulation www.uniprot.org.

RNA-binding component of the eukaryotic translation initiation factor 3 (eIF-3) complex is required for several steps in the initiation of protein synthesis. The eIF-3 complex specifically targets and initiates translation of a subset of mRNAs involved in cell proliferation, including cell cycling, differentiation and apoptosis, and uses different modes of RNA stem-loop binding to exert either translational activation or repression www.uniprot.org. It was shown to participate in a complex that binds near the mRNA entry channel and regulates the transition between scanning-conducive and initiation-competent conformations of the preinitiation complex Wagner *et al.* (2014); Chiu *et al.* (2010).

These two proteins are most likely not involved in the development of resistance to stress conditions, DNA protection and reparation.

Taking into account all our results, we can suggest verifying the role of Dsup and other proteins in the DNA radiation-resistance mechanisms. This might be done by making knockout or knockdown experiments on Tardigrades and checking for the differences in the DNA damage in the different conditions. Also, we can perform differential expression analysis of Tardigrades exposed to radiation and in normal conditions. According to our hypothesis, the expression of Dsup and other related genes should increase.

However, since Tardigrades are hard to cultivate in laboratory conditions, experiments on the bacteria or other eukaryotic cell lines might be more reliable. Thus to experimentally test whether the damage suppressor protein is actually involved in the development of resistance to stress conditions, we can insert the gene encoding this protein into bacteria or cell lines, subject them to stress conditions (irradiation, high or low temperatures, abnormal pressure) and see if they become more resistant to such an impact Hashimoto *et al.* (2016).

Literature cited

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. Journal of Molecular Biology. 215:403–410.

Armenteros JJA, Salvatore M, Emanuelsson O, Winther O, von Heijne G, Elofsson A, Nielsen H. 2019. Detecting sequence signals in targeting peptides using deep learning. Life Science Alliance. 2:e201900429.

Boothby TC, Tapia H, Brozena AH, Piszkiewicz S, Smith AE, Giovannini I, Rebecchi L, Pielak GJ, Koshland D, Goldstein B. 2017. Tardigrades use intrinsically disordered proteins to survive desiccation. Molecular Cell. 65:975–984.e5.

Buchfink B, Reuter K, Drost HG. 2021. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nature Methods. 18:366–368.

Chavez C, Cruz-Becerra G, Fei J, Kassavetis GA, Kadonaga JT. 2019. The tardigrade damage suppressor protein binds to nucleosomes and protects DNA from hydroxyl radicals. eLife. 8.

Chiu WL, Wagner S, Herrmannova
A, Burela L, Zhang F, Saini AK, Vala
s
ek L, Hinnebusch AG. 2010. The c-terminal region of eukaryotic translation initiation factor 3a (eIF3a) promotes mRNA recruitment, scanning, and, together with eIF3j and the eIF3b RNA recognition motif, selection of AUG start codons. Molecular and Cellular Biology. 30:4415–4434.

Hashimoto T, Horikawa DD, Saito Y, Kuwahara H, Kozuka-Hata H, Shin-I T, Minakuchi Y, Ohishi K, Motoyama A, Aizu T *et al*. 2016. Extremotolerant tardigrade genome and improved radiotolerance of human cultured cells by tardigrade-unique protein. Nature Communications. 7.

Hashimoto T, Kunieda T. 2017. DNA protection protein, a novel mechanism of radiation tolerance: Lessons from tardigrades. Life. 7:26.

Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier C, Nakai K. 2007. WoLF PSORT: protein localization predictor. Nucleic Acids Research. 35:W585–W587.

Koutsovoulos G, Kumar S, Laetsch DR, Stevens L, Daub J, Conlon C, Maroon H, Thomas F, Aboobaker AA, Blaxter M. 2016. No evidence for extensive horizontal gene transfer in the genome of the tardigrade ihypsibius dujardini/i. Proceedings of the National Academy of Sciences. 113:5053–5058.

Mínguez-Toral M, Cuevas-Zuviría B, Garrido-Arandia M, Pacios LF. 2020. A computational structural study on the DNA-protecting role of the tardigrade-unique dsup protein. Scientific Reports. 10.

Møbjerg N, Halberg KA, Jørgensen A, Persson D, Bjørn M, Ramløv H, Kristensen RM. 2011. Survival in extreme environments - on the current knowledge of adaptations in tardigrades. Acta Physiologica. 202:409–420.

Møbjerg N, Neves RC. 2021. New insights into survival strategies of tardigrades. Comparative Biochemistry and Physiology Part A: Molecular &amp Integrative Physiology. 254:110890.

Stanke M, Morgenstern B. 2005. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Research. 33:W465–W467.

Wagner S, Herrmannová A, Malík R, Peclinovská L, Valášek LS. 2014. Functional and biochemical characterization of human eukaryotic translation initiation factor 3 in living cells. Molecular and Cellular Biology. 34:3041–3052.

Supplementary

Table 1 BLAST hits for the genes located in the nucleus

Gene ID	Accession Number	E-value	% Ident	% Query coverage	Annotation
g14472.t1	P0DOW4.1	0.0	100.00	100	Damage suppressor protein [Ramazzottius varieornatus]
g7861.t1	B4F769.1	$2e-71$	37.21	99	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A-like protein 1 [Rattus norvegicus]
g10513.t1	-	-	-	-	-
g11806.t1	-	-	-	-	-
g16318.t1	A4II09.1	$9e-08$	36.54	40	Eukaryotic translation initiation factor 3 subunit A (eIF3a) [Xenopus tropicalis]
g16368.t1	A4II09.1	$1e-05$	39.29	35	Eukaryotic translation initiation factor 3 subunit A (eIF3a) [Xenopus tropicalis]
g3428.t1	P19105.2	$1e-63$	58.44	89	Myosin regulatory light chain 12A [Homo sapiens]
g11513.t1	Q6PA97.1	$3e-77$	29.08	66	Trafficking protein particle complex subunit 9 [Xenopus laevis]
g5443.t1	-	-	-	-	-
g8312.t1	Q5KU39.1	0.0	40.84	84	Vacuolar protein sorting-associated protein 41 homolog [Mus musculus]
g5927.t1	Q17427.1	$1e-18$	38.64	14	Glucosamine 6-phosphate N-acetyltransferase [Caenorhabditis elegans]
g8100.t1	Q2YDR3.1	$3e-46$	36.04	22	Inositol monophosphatase 3 [Danio rerio]
g11960.t1	Q8CJB9.1	$6e-98$	26.96	96	E3 ubiquitin-protein ligase BRE1B [Rattus norvegicus]
g10514.t1	-	-	-	-	-
g2203.t1	Q69ZQ1.2	$2e-126$	35.93	75	Myogenesis-regulating glycosidase [Mus musculus]

Table 2 Cumulative information on the found genes

Gene ID	Blast: annotation	Blast: e-value	Predicted Pfam domains	Localization (WoLF)	Localization (TargetP)
g14472.t1	Damage suppressor protein	0.0	no results	nucl: 28, plas: 2, cyto: 1, cysk: 1	other
g7861.t1	SWISNF-related matrix-associated actin-dependent regulator of chromatin subfamily A-like protein 1	$2e-71$	SNF2-related domain, HepA-related protein (HARP), Type III restriction enzyme, res subunit	nucl: 16, cyto-nucl: 14, cyto: 8, plas: 5, pero: 1, cysk: 1, golg: 1	other
g10513.t1	-	-	no hits	nucl: 20, cyto-nucl: 14.5, cyto: 7, extr: 3, E.R.: 1, golg: 1	other
g11806.t1	-	-	no hits	nucl: 18, cyto-nucl: 11.8333, mito: 5, extr: 4, cyto: 3.5, cyto-pero: 2.66667, cysk-plas: 1	other
g16318.t1	Eukaryotic translation initiation factor 3 subunit A (eIF3a)	$9e-08$	no hits	nucl: 20.5, cyto-nucl: 13, extr: 5, cyto: 4.5, E.R.: 1, golg: 1	other
g16368.t1	Eukaryotic translation initiation factor 3 subunit A (eIF3a)	$1e-05$	no hits	nucl: 20.5, cyto-nucl: 13, extr: 5, cyto: 4.5, E.R.: 1, golg: 1	other
g3428.t1	Myosin regulatory light chain 12A	$1e-63$	EF hand, EF hand domain	mito: 18, cyto: 11, extr: 2, nucl: 1	other
g11513.t1	Trafficking protein particle complex subunit 9	$3e-77$	Transport protein Trs120 or TRAPPC9, TRAPP II complex subunit	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	other
g5443.t1	-	-	Chitin binding Peritrophin-A domain	extr: 28, nucl: 3, cyto: 1	other, signal (0.0438)
g8312.t1	Vacuolar protein sorting-associated protein 41 homolog	0.0	Region in Clathrin and VPS	nucl: 15.5, cyto-nucl: 15.5, cyto: 12.5, mito: 2, plas: 1, golg: 1	other
g5927.t1	Glucosamine 6-phosphate N-acetyltransferase	$1e-18$	no hits	nucl: 30.5, cyto-nucl: 16.5, cyto: 1.5	other
g8100.t1	Inositol monophosphatase 3	$3e-46$	Inositol monophosphatase family, Arf6-interacting domain of mitotic kinesin-like protein 1	nucl: 16.5, cyto-nucl: 12.5, cyto: 7.5, plas: 5, extr: 2, E.R.: 1	other
g11960.t1	E3 ubiquitin-protein ligase BRE1B	$6e-98$	Zinc finger, C3HC4 type (RING finger)	nucl: 32	other
g10514.t1	-	-	ho hits	nucl: 19, cyto-nucl: 15, cyto: 9, extr: 3, mito: 1	other
g2203.t1	Myogenesis-regulating glycosidase	$2e-126$	Glycosyl hydrolases family 31	plas: 29, nucl: 2, golg: 1	other