Practice

BIOINFORMATICS INSTITUTE

Searching for a tardigrade-unique radiotolerance protein

Balan Anna[1, 3] and Khegai Gleb[1, 2]

[1]Bioinformatics Institute
[2]Moscow Institute of Physics and Technology
[1]RSAU – MTAA named after K.A. Timiryazev

Abstract

Ramazzottius varieornatus can efficiently repair it's DNA. Thanks to this, it can survive in extreme conditions, including the hard vacuum of outer space and solar UV radiation. YOKOZUNA-1 strain genome was sequenced and annotated. After that genome data was combined with proteomic data to find out which proteins might be located in a nucleus and which functions can they have. Tardigrade-unique protein appeared to be a one of the tardigrade's resistance mechanisms. It is a damage suppressor protein, associated with a nuclear DNA.

Keywords: Tardigrade, Waterbear, stress-tolerance, DNA repair

Introduction

Tardigrades have already attracted the attention of the researchers with their amazing endurance. When adverse conditions occur, they are able to fall into a state of suspended animation for years, and when favorable conditions occur, they quickly come to life. Tardigrades survive mainly due to the so-called anhydrobiosis, or drying. When dried, they draw limbs into the body, decrease in volume and take the shape of a barrel. The surface is covered with a wax coating that prevents evaporation. During anabiosis, their metabolism drops to 0.01%, and the water content can reach up to 1% of normal. In a state of suspended animation, tardigrades endure incredible loads (Stone and Vasanthan 2020).

For some time, it was believed that about 17% of the 38,000 genes were "borrowed" from other organisms, including extremophile bacteria. It was assumed that when dried, their DNA breaks up into large fragments, and when they return to living conditions with a normal water content, special proteins "crosslink" and restore damaged DNA. At this moment fragments of foreign DNA were believed to allegedly enter the cells, which are "sewn" into the genome.

However, the reason for the conclusions about the massive borrowing of foreign genes was the contamination (contamination) of DNA samples of tardigrades with foreign bacterial DNA (Koutsovoulos *et al.* 2016).

In order to find the real reasons behind their extreme endurance this project was performed. Gene prediction, the process of determining where a coding gene might be in a genomic sequence was used as well as protein sequence homology.

We analyzed the extracted parts of proteins obtained through tandem mass spectrometry and got a list of peptides that were associated with the DNA. There was a unique one that was not discovered before. It is similar to a protein of Chaetomium thermophilum, eukariotic thermophil (Ștefan *et al.* 2011).

Materials and methods

A genome of the Ramazzottius varieornatus, the YOKOZUNA-1 strain was sequenced and assembled in the University of Tokyo. After that masking with RepeatMasker (Smit *et al.* 2013-2015) was performed. to predict coding regions in the genome we used AUGUSTUS tool. As a result, 16435 coding proteins were found.

Considering that DNA is a major target of UV radiation damage, we assumed that tardigrades might have unique proteins associated with their DNA to protect or effectively repair it. To explore this possibility, we combined genomic and proteomic data.

Proteins from extracted chromatin fraction were analyzed using tandem mass spectrometry. Then we did a local alignment-based search of the obtained protein segments with Classic BLAST+ (Boratyn *et al.* 2019). The whole protein seqiences were found as a result. To find out where in the eukariotic cell these proteins might be located we used two tools and compared the results: WoLF PSORT (Horton *et al.* 2007) and TargetP-2.0 (Horton *et al.* 2019). Six out of twenty proteins were predicted to be found in nucleus. All of them were checked with NCBI BLAST (Boratyn *et al.* 2019). As a result, one protein appeared to be unique and had minimal E-value.

To exclude contamination or sequencing mistakes we predicted it's function with Pfam (Horton *et al.* 2020). It appeared to be similar to a protein of Chaetomium thermophilum, eukariotic thermophil, asssociated with DNA as well.

Results

In order to carry out analysis it is necessary to develop some criteria for the selection of proteins responsible for DNA reparation. As we can see, some proteins have specific functions according to BLAST and Pfam. So we discarded them. Next step we discarded proteins with parameter 'Signal peptide' in Tar-

getP localization. These proteins with this N-terminal extension have to be reported outside the cell nucleus and they can't repair DNA inside the nucleus therefore. Finally, we discarded proteins that were not found in the nucleus by WOLF Psort localization. Thus, our target proteins should fulfill all above criteria in order to be a candidate responsible for DNA reparation (Table 1).
Some protein fragments wasn't aligned on the assembled genome, that probably happened because of the mass-spectrometry analyses problems. This method is hard to perform and get data from, and such a situation is normal.
Some proteins, situated in nucleus were not found by BLAST. They seem to be are freshly discovered. For additional information such as their funcrion the additional analyses has to be performed.

Discussion

Ionizing irradiation and UV-irradiation cause DNA damage. It can induce single-strand breaks(SSBs), double-strand breaks(DSBs), various monomeric base damages and much more. There are three different excision repair pathways for SSBs that are universally present in living organisms: strand-specific mismatch repair, nucleotide excision repair, base excision repair.(Lankinen *et al.* 1996) The DSBs pasticulary hazardous, if they go unrepaired in mammalian cells, they can also cause gene deletion, chromosome loss and other chromosomal aberrations that might ultimately produce cancers. DNA double-strand breaks are repaired by means of two main mechanisms: nonhomologous end joining and homologous recombination.(Lieber 2010)

All these processes occur inside the nucleus and in this context it's reasonable to look for DNA associated proteins with unknown function. In this study we find three proteins that we expect to be unique for tardigrades. We recommend immunohistochemistry for verifying intranuclear localization of selected proteins. To prove DNA protection function can be estimated transfected cell culture with ability to express proteins previously identified tardigrade-unique. After radiation exposure on both cell cultures original and transfected we can evaluate genome alteration rate.

Recent studies show that 1.2% of tardigrade genes are borrowed by horizontal transfer from other kingdoms of living beings (Hashimoto1 *et al.* 2016). The exact mechanism behind HGT in tardigrades is not described though, there is still a chance that such results are false, and the further researches might be performed. The discovered protein was possibly transferred in the tardigrade's genome through HGT, since they are similar, then their functions are similar as well. We have to hold more analyses and experiments about this protein and it's function in order to be sure.

Acknowledgments

This work exists thanks to Mikhail Rayko and his manual :)
For this project we used a sequence of the Ramazzottius varieornatus, the YOKOZUNA-1 strain read in the University of Tokyo.

Literature cited

Boratyn G, Thierry-Mieg J, Thierry-Mieg D. 2019. Magic-blast, an accurate rna-seq aligner for long and short reads.
Hashimoto1 T, Horikawa DD, Saito1 Y, Kuwahara H, Kozuka-Hata5 H, Shin-I6 T, Minakuchi7 Y, Ohishi K, Motoyama A, Aizu T *et al*. 2016. No evidence for extensive horizontal gene transfer in the genome of the tardigrade hypsibius dujardini.
Horton P, Park KJ, Obayashi T, Fujita N, Harada H. 2007. Wolf psort.
Horton P, Park KJ, Obayashi T, Fujita N, Harada H. 2019. Detecting sequence signals in targeting peptides using deep learning.
Horton P, Park KJ, Obayashi T, Fujita N, Harada H. 2020. Pfam: The protein families database in 2021.
Koutsovoulos G, Kumar S, Laetsch DR, Stevens L, Daub. J. 2016. No evidence for extensive horizontal gene transfer in the genome of the tardigrade hypsibius dujardini.
Lankinen M, Vilpo L, JA V. 1996. Uv- and gamma-irradiation-induced dna single-strand breaks and their repair in human blood granulocytes and lymphocytes.
Lieber M. 2010. The mechanism of double-strand dna break repair by the nonhomologous dna end-joining pathway.
Smit A, Hubley R, Green P. 2013-2015. Repeatmasker open - 4.0.
Stefan A, Sarges P, Flemming D, van Noort V, Kunze R, Devos DP, Arumugam M, Bork P, Hurt E. 2011. Insight into structure and assembly of the nuclear pore complex by utilizing the genome of a eukaryotic thermophile. Cell. 146:277–89.
Stone J, Vasanthan T. 2020. Life history traits for the freshwater tardigrade species hypsibius exemplaris reared under laboratory conditions. Journal of Wildlife and Biodiversity. 4:65–72.

Table 1 Integrative table with information about hypotetical genes

Gene	Blast evalue	TargetP classification	WoLF PSORT									
			extr	lyso	nucl	plas	E.R.	cyto	cysk	golg	pero	mito
g12510.t1	1.49E-09	OTHER	-	-	-	+	-	+	-	-	-	-
g4106.t1	2E-06	OTHER	+	+	-	+	+	-	-	+	+	+
g15484.t1	9.59E-06	OTHER	-	-	+	+	-	+	-	+	-	-
g13530.t1	0.000614	SP	+	+	+	+	+	+	-	-	-	-
g14472.t1	0.002	OTHER	-	-	+	+	-	+	+	-	-	-
g10513.t1	0.003	OTHER	+	-	+	-	+	+	-	+	-	-
g5237.t1	0.004	OTHER	-	-	-	+	-	-	-	-	-	+
g5616.t1	0.026	SP	+	-	-	-	-	-	-	-	-	+
g3679.t1	0.046	SP	+	+	-	+	+	-	-	-	-	+
g12562.t1	0.052	SP	+	+	-	-	-	-	-	-	-	-
g5641.t1	0.053	SP	+	+	-	-	-	-	-	-	-	-
g3428.t1	0.055	OTHER	+	-	+	-	-	+	-	-	-	+
g5443.t1	0.094	OTHER	+	-	+	-	-	+	-	-	-	-