Analysis of the tardigrade's genome to determine the causes of stress tolerance

Vinogradova Sofiya, Zherko Nikita
Moscow Institute of Physics and Technology

December 9, 2023

Abstract

Targidrates are well-known organisms that can survive in severe conditions. However, little is known about the reasons for such endurance. In this project we used several prediction methods to identify proteins that could be responsible for this stress resielence. We reveled one protein, Dsup, that binds to DNA in the nucleus and in this way can prevent DNA from UV damage. These finding are of a great importance for further investigation of *Targidrates's* high resistance to stress.

1 Introduction

Targidrates are unique organisms in that they can survive extreme conditions by transitioning between two states: dehydrated and hydrated. Just in the dehydrated stage, when the *Targidrates* are represented by a dense protective gel with genetic information inside, they are resistant to radiation and other extreme conditions.

The reason for this resistance is in the genome, so we are interested in finding the proteins that ensure the survival of the *Targidrates*. Genome analysis shows that there is a small fraction (1.2% or less) of putative foreign genes, gene pathways contributing to stress damage have been lost, and a family of genes associated with damage attenuation has been expanded.

Various protein prediction methods have been used in this work. These methods predict the subcellular localisation of proteins and identify signal peptides that are exported from the cell or based on the predicted presence of any of the N-terminal presequences.

2 Materials and Methods

We work with precomputed AUGUSTUS results in fasta and gff formats. We got the proteins from the gff file in fasta format using getAnnoFasta.pl script [1] and counted the number of proteins using grep v. 2.6.0-FreeBSD with -c flag.

We will use the data that were obtained by mass spectrometry of proteins isolated from chromatin. To find out which proteins correspond to the obtained peptides, we will use classic BLAST+ v 2.6.0+ [2] by means of which did a local alignment-based search. To get the file with extracted proteins we use utility seqtk v.1.4-r122 [3] before sorting, finding unique elements and selecting the column 2 (sseqid) with awk v. 20200816 from the blast results.

Than we predicted protein localisation using two WoLF PSORT [https://wolfpsort.hgc.jp/] and TargetP Server [https://services.healthtech.dtu.dk/services/TargetP-2.0/]. BLAST web-version chttps://blast.ncbi.nlm.nih.gov/Blast.cgi] against the "UniProtKB/Swiss-Prot" database was used to find homologous protein sequences and results were obtained for each protein. Also predicted the functions of proteins using Pfam web-version and seect hmmscan tool [https://www.ebi.ac.uk/Tools/hmmer/].

3 Results

Here we worked with precomputed AUGUSTUS results: augustus.whole.aa and augustus.whole.gff. First, we extracted proteins from augustus.whole.gff and count number of it. There were 16435 proteins present in the file.

We then we interested in only those proteins that interact with the DNA. To do so we took a list of proteins associated with DNA. By doing a local alignment-based search we identified 118 proteins that could interact with DNA. After additional sorting and finding unique peptides we got 34 proteins in total.

To filter the list of 34 proteins we used additional tools. We used WoLF PSORT, which identifies subcellular protein localization based on the presence of protein signaling peptide on th N-terminal end. TargetP also identifies protein localization based on the presence of the N-terminal presequences. To point out the protein function we then used BLAST and Pfam. The results for each program run can be found in the lab journal [https://github.com/rereremin/IB/blob/project4/resluts].

Having taken into account the evidence from the programs listed above we found one protein Dsup (g14472.t1), which could explain the survival rate of the pacifier. Detailed information about this protein can be found in Table 1

Protein	Organism	BLAST best hit	Annotation	e-value	Pfam results	WoLF PSORT results	TargetP results
Dsup	*R. varieornatus*	P0DOW4.1	Damage suppressor	0.0	-	nucle: 28, plas: 2, cyto: 1, cysk: 1	OTHER

Table 1: Info about target protein

4 Discussion

In this project, we tried to identify proteins responsible for the *Targidrates'* high stress resistance. Combining various prediction tools we came up with one candidate protein – Dsup (P0DOW4, g14472.t1). The protein has its cellular localization in the nucleus and, according to BLAST search, is responsible for damage suppression. Since the protein is localized in the nucleus it could save DNA from damage by two different mechanisms. Dsup could be a specific DNA repair protein, or a protein involved in the DNA repair protein cascade or its activation, or a protein that just prevents DNA from degradation. On the other hand, this protein could be a molecule that absorbs radicals within the cell and reduces oxidative stress.

We found several similarities with the paper that first revealed Dsup protein as responsible for the stress resistance of *Targidrates*. The authors of the paper state that Dsup can protect DNA by direct association with it rather than by inducing DNA repair machinery [dsup], which is in line with one of our hypothesis.

To further test Dsup functions one can perform knock out of the Dsup's gene to see for alterations in the cell behavior the the radiation stress. It could be also trasformed on a plasmid to other well-known human cell lines or *E. coli* to test its function.