

Tardigrades: from genestealers to space marines

Ushakov M. Student of Bionformatic institute. Saint-Petersburg

Tolmachev M. Student of the Bioinformatic institute. Saint-Petersburg

Abstract

Ramazzottius varieornatus have the ability repair or to keep DNA unharmed during affect of UV spectrum radiation. This ability is associated with peptides from chromatin fraction. Knowing genes and proteins that provide this ability possible could play a key role in surviving. Proteins g8100, g11960 and g14472 are the candidates on that role. They regulate apoptosis and intracellular degradation of proteins

Introduction

From all history of living nature, some species were living in extreme conditions defined by wide range of temperature, atmospheric pressure, radiation, UV spectrum of light e.t.c. Back in the days, there were times when all living organisms had to face extreme conditions and prove their adaptability. Also, there is no guarantee (in fact there are some evidences), that we will witness another planetary catastrophe. And humanity need to be prepared for it, unfortunately, we are not the best. The absolute champions in ability to be alive no matter what are Tardigrades. And the key to that ability is hidden in their genes. If we look into the genome of *Ramazzottius varieornatus*, we possibly would find how they can repair their DNA and be tolerant to UV radiation and other stress factors. The key idea of a process is to find proteins encoded by *Ramazzottius varieornatus* DNA which are able to repair or secure DNA from harmful effects.

Materials and methods.

For this project we used sequence of YOKOZUNA-1 strain [1]. Having lack of power we did not assemble genome by ourselves. But if you want to do it you could consider using SOAPdenovo[2], Platanus [3], DISCOVAR[4]. Or you can download assembled file Or only the longest contig [5].

For functional annotations we used AUGUSTUS[6] and predefined gene prediction files [7,8]. For obtaining protein sequence from prediction output we used

getAnnoFasta.pl script[9] After that we got file with proteins and can obtain the total number of them performing grep operation by “>”. Having a list of peptides from chromatin fraction[10], we created database from obtained proteins sequences with makeblastdb command [makeblastdb -in augustus.whole.aa -parse_seqs -dbtype prot] and found matches with peptides by local blast search [blastp -query peptides.fa -db augustus.whole.aa -outfmt 6] . We used custom script with biopython to extract sequences from augustus output by the received names [can be obtained in lab journal]. For prediction of where proteins are found in the cell we used Wolf PSORT [11] on obtained set of proteins. For prediction of subcellular proteins we used TargetP 1.1 [12]. Then we performed blast search against "UniProtKB/Swiss-Prot" database and got result in Hit table format. For obtaining information from it (like query, max score, total score, query cover, e.value e.t.c) we used custom Python 3 script (which can be obtained from lab journal). For prediction of protein function we used HMMER with hmmscan tool and Pfam database [13].

Results

For functional annotation, we used genomic data after processing with *AUGUSTUS* gene predictor. We got 16435 protein sequences. Using blastdb, a local database was created in which a search was performed using blastp. As a result, we got 118 matches with 34 unique proteins.

Since we are interested in the genes involved in the protection and repair of DNA, we must first consider those localized in the nucleus. There are 12 proteins in the WoLF PSORT output with a nucleus-localization signal. Two of them are described as translation initiation factor subunits. The other 3 were not described at all by any of the tools. Information about the remaining 7 is presented in the table 1.

Table 1 – Functional annotation of nucleus-localized proteins of interest

| Protein | Pfam domain | BLAST T e- value | BLAST description |
|---------|-------------|------------------------|----------------------------|
| g5927 | None | 1e-18 | Glucosamine 6-phosphate N- |

| | | | |
|--------|---------------------------------------|-------|---|
| | | | acetyltransferase |
| g7861 | SNF2 family N-terminal domain | 2e-71 | SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A-like protein 1 |
| g8100 | Inositol monophosphatase family | 3e-46 | Inositol monophosphatase 3 |
| g8312 | Region in Clathrin and VPS | 0.0 | Vacuolar protein sorting-associated protein 41 homolog |
| g11960 | Zinc finger, C3HC4 type (RING finger) | 6e-98 | E3 ubiquitin-protein ligase BRE1B |
| g14472 | None | 0.0 | Damage suppressor protein [Ramazzottius varieornatus] |
| g15484 | Vps51/Vps67 | 0.0 | Vacuolar protein sorting-associated protein 51 homolog |

Discussion

the 7 selected proteins, in our opinion, only 3 can be considered as candidate proteins. There is information in the literature that inositol phosphatases play a role in the regulation of apoptosis [14]. Also the ubiquitin-protein ligase protein has been reported to regulate the processes of intracellular degradation of other proteins, as well as to modify their functions [15]. Perhaps these proteins are in some sort of pathway that regulates the destruction of damaged cells or abnormal proteins. And the most important find is the protein from the article we reproduce, found in tardigrades and annotated as damage suppressor.

It is also possible that 3 unannotated proteins may have some important function.

Links

1. <https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=947166>

2. . Luo R, et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience*. 2012;1:18. doi: 10.1186/2047-217X-1-18.
3. Kajitani R, Toshimoto K, Noguchi H, Toyoda A, Ogura Y, Okuno M, Yabana M, Harada M, Nagayasu E, Maruyama H, Kohara Y, Fujiyama A, Hayashi T, Itoh T, “Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads”. *Genome Res*. 2014 Aug;24(8):1384-95. doi: 10.1101/gr.170720.113.
4. <https://software.broadinstitute.org/software/discover/blog/>
5. ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/001/949/185/GCA_001949185.1_Rvar_4.0/GCA_001949185.1_Rvar_4.0_genomic.fna.gz
6. Mario Stanke, Ana Tzvetkova, Burkhard Morgenstern (2006) "AUGUSTUS at EGASP: using EST, protein and genomic alignments for improved gene prediction in the human genome" *BMC Genome Biology*, 7(Suppl 1):S11.
7. <http://public.dobzhanskycenter.ru/mrayko/BIMM185/augustus.whole.gff>
8. https://drive.google.com/file/d/1wBxf6cDgu22NbjAOgTe-8b3Zx60hNKY0/view?usp=drive_web
9. <http://augustus.gobics.de/binaries/scripts/getAnnoFasta.pl>
10. <http://public.dobzhanskycenter.ru/mrayko/BIMM185/peptides.fa>
11. <https://wolfpsort.hgc.jp/>
12. <http://www.cbs.dtu.dk/services/TargetP/>
13. <https://www.ebi.ac.uk/Tools/hmmer/>
14. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2820388/>
15. <http://pfam.xfam.org/family/PF00097.25>