Tardigrades: from genestealers to space marines

Author: Kristina Zheltova

Abstract

Tardigrades are microscopic animals capable of withstanding some of the most severe environmental conditions. They can survive freezing (up to 1°K), total dehydration, pressure (more than 1,200 atmospheres) and radiation (1,000 times more radiation than other animals). In this work I analyzed genome of Ramazzottius varieornatus. Considering that DNA is a major target of radiation damage, I found proteins associated with their DNA and predict proteins that could be involved to reparation process: SNF2-related protein, E3 ubiquitin ligase with the RING finger domain, and Dsup protein. These findings could help us to understand stress tolerance mechanisms of tardigrades.

Introduction

Tardigrades are among the most resilient animals known, with individual species able to survive extreme conditions – such as exposure to extreme temperatures, extreme pressures (both high and low), air deprivation, radiation, dehydration, and starvation – that would quickly kill most other known forms of life. Tardigrades have survived exposure to outer space [1].

Boothby et al. in [2] reported evidence for an unprecedented degree of HGT into an animal genome, based on a draft genome of a tardigrade, Hypsibius dujardini. They estimate that approximately one-sixth of tardigrade genes entered by HGT, nearly double the fraction found in the most extreme cases of HGT into animals known to date. Animals that can survive extremes may be particularly prone to acquiring foreign genes.

After that in [3] Koutsovoulo et al. found no support for extensive fHGT. Cross-comparison of assemblies showed that the overwhelming majority of HGT candidates in the Boothby et al. genome derived from contaminants. They conclude that fHGT into H. dujardini accounts for at most 1–2% of genes and that the proposal that one-sixth of tardigrade genes originate from functional HGT events is an artifact of undetected contamination.

Gene prediction is one of the most important and alluring problems in computational biology. Its importance comes from the inherent value of the set of protein-coding genes for other analysis.[4]. There are many tools for gene prediction, for example, AUGUSTUS which is a gene-prediction tool that utilizes either a generalized hidden Markov model (generalized HMM) or semi-Markov conditional random field (CRF) to identify protein-coding genes and their structures in a given genome sequence. Functional annotation of genes can be found with Blast of HMMER and protein localization prediction can be done with WoLF PSORT and TargetP.

In this work, I aimed to find the genomic basis for the tardigrades high stress tolerance, using a sequence of the Ramazzottius varieornatus. One of the consequences of UV radiation is DNA damage, so we can suggest, that key proteins providing tardigrades with high stress tolerance could be associated with DNA.

Methods

Genome sequence of the Ramazzottius varieornatus was taken from here.

Structural annotation performed with AUGUSTUS [5]. To extract protein sequences from the prediction output, script getAnnoFasta.pl was used. List of peptides that were associated with the DNA was obtained using using tandem mass spectrometry and taken from here. To find proteins from the R. varieornatus genome corresponding to these peptides, local alignment-based search was performed with BLAST+ [6].

Localization prediction was verified using WoLF PSORT [7] and TargetP [8]. Protein functions predictions were obtained with BLASTP search against the "UniProtKB/Swiss-Prot" database. Also search protein sequences against a collection of profile-HMMs for different protein domains and motifs was performed with HMMER [11].

Results

From gene prediction results, I obtained 16435 protein sequences. Considering that DNA is a major target of radiation damage, it's possible to hypothesize that tardigrades might have unique proteins associated with their DNA to protect and/or effectively repair it. To explore this possibility, I combined genomic and proteomic data. Using tandem mass spectrometry data on proteins, I identified 34 proteins associated with DNA. Nuclear localization was confirmed only for 12 studied proteins.

For almost in all extracellular proteins TargetP type was predicted as signal peptide (except g5443.t1). I revealed homologs for 23 studied proteins using BLASTP search. For most proteins predicted with BLASTP, it was possible to identify the Pfam family domain.

Several proteins (g5467.t1, g5616.t1, g15153.t1, g5503.t1, g5502.t1, g5641.t1, g12562.t1, g702.t1, g1285.t1, g12388.t1) were predicted as Chitin binding Peritrophin-A domain. They could function as chitinases [9] and participate in chitin-containing structures remodelling.

Discussion

Among the 34 obtained proteins, those that are thought to have nuclear localization may be of great importance in terms of DNA repair.

Protein g11960.t1 was predicted as ubiquitin-protein ligase with the RING finger domain. This type of ubiquitin-protein ligase has been reported to play an important role in DNA damage response, including recognition and reparation [10].

SNF2-related protein (g7861.t1) was found. Transcriptome analysis of tardigrade species Milnesium tardigradum reveals several unique proteins from the SNF2 family considered to be involved in resistance against extreme environmental conditions [9]. It is possible that in Ramazzottius varieornatus this SWI/SNF complex may be involved in DNA breaks thus providing tardigrades survival after exposure to ionizing radiation.

Also, a Dsup protein specific to tardigrades (g14472.t1) was found. This protein protects against genomic instability caused by DNA damage.

References

1. Wikipedia - Tardigrade
2. Boothby, Thomas C et al. "Evidence for extensive horizontal gene transfer from the draft genome of a tardigrade." Proceedings of the National Academy of Sciences of the United States of America vol. 112,52 (2015): 15976-81. doi:10.1073/pnas.1510461112
3. Koutsovoulos, Georgios et al. "No evidence for extensive horizontal gene transfer in the genome of the tardigrade Hypsibius dujardini." Proceedings of the National Academy of

Sciences of the United States of America vol. 113,18 (2016): 5053-8. doi:10.1073/pnas.1600338113

4. Flicek, Paul. "Gene prediction: compare and CONTRAST." Genome biology vol. 8,12 (2007): 233. doi:10.1186/gb-2007-8-12-233
5. Stanke, M., & Morgenstern, B. (2005). AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic acids research
6. Altschul, S F et al. "Basic local alignment search tool." Journal of molecular biology vol. 215,3 (1990): 403-10. doi:10.1016/S0022-2836(05)80360-2
7. Horton, Paul et al. "WoLF PSORT: protein localization predictor." Nucleic acids research vol. 35,Web Server issue (2007): W585-7. doi:10.1093/nar/gkm259
8. Detecting Sequence Signals in Targeting Peptides Using Deep Learning José Juan Almagro Armenteros, Marco Salvatore, Ole Winther, Olof Emanuelsson, Gunnar von Heijne, Arne Elofsson, and Henrik Nielsen Life Science Alliance 2 (5), e201900429. doi:10.26508/lsa.201900429
9. Förster, Frank et al. "Transcriptome analysis in tardigrade species reveals specific molecular pathways for stress adaptations." Bioinformatics and biology insights vol. 6 (2012): 69-96. doi:10.4137/BBI.S9150
10. Brinkmann, Kerstin et al. "Regulation of the DNA damage response by ubiquitin conjugation." Frontiers in genetics vol. 6 98. 10 Mar. 2015, doi:10.3389/fgene.2015.00098
11. https://www.ebi.ac.uk/Tools/hmmer/