Tardigrades: from genestealers to space marines

Authors: Anna Rybina, Daria Nikanorova

Abstract

Tardigrades are microscopic unique creatures remarkable for their ability to survive extreme conditions, including ionizing radiation, the outer space vacuum, high temperatures and pressure. In this work, we analyzed the genome of one of the most stress-tolerant tardigrade species, *Ramazzottius varieornatus,* particularly focusing on proteins associated with DNA. Using functional annotation approach such as sequence homology search, we predicted several candidate proteins that could be essential for ensuring high stress tolerance in tardigrades due to their involvement in the DNA damage response: SWI/SNF-related protein, E3 ubiquitin ligase with the RING domain, and Dsup protein. These findings could enhance our understanding of unique mechanisms of extraordinary stress tolerance in tardigrades and provide candidate proteins for further investigations.

Introduction

Tardigrades are microscopic animals that exhibit high tolerance towards various extreme environmental conditions, including intense UV and ionizing radiation, high pressure, desiccation, and outer space vacuum.

In 2016, analyzing the draft genome of a tardigrade *Hypsibius dujardini ,* Boothby et al. (Boothby et al. 2015) claimed that about one-sixth of the genes was gained by a tardigrade via horizontal gene transfer (HGT) from bacteria and other sources. This high fraction of foreign genes, not common among animals, could modify gene composition in tardigrades, significantly increasing their stress tolerance. However, independent study made by Koutsovoulos et al. (Koutsovoulos et al. 2016) demonstrated that a high proportion of foreign genes in a tardigrade reported by Boothby et al. (Boothby et al. 2015) was an artifact of contamination during genome sequencing that affected assembly results. Koutsovoulos et al. showed that only about 1-2 % of genes in a tardigrade were obtained due to HGT. Similar estimation of HGT proportion for a tardigrade *Ramazzottius varieornatus* was obtained by Hashimoto et al. (Hashimoto et al. 2016).

Gene prediction is a crucial part of genome annotation after assembly. Coding regions may be defined mathematically (*ab initio* prediction) or with application of previous sequencing results, such as transcriptome, etc. Advanced *ab initio* gene finders typically utilize probabilistic models, such as hidden Markov models (HMMs). Modern tools, such as AUGUSTUS, can combine *ab initio* prediction with incorporation of extrinsic information, e.g. from EST alignments and protein alignments. Functional annotation of predicted genes can be achieved by finding orthologous sequences from other species in databases with Blast (Basic Local Alignment Search Tool). In case proteins show no similarity to any sequences in public databases we can try to predict their function by finding possible conserved protein domains with HMMER. Finally, protein subcellular localization generally reflects its function

and may be defined by presence of special signal peptides in proteins. WoLF PSORT and TargetP are popular tools for protein localization prediction.

In this study, we aimed to clarify the genomic basis for the high stress tolerance in tardigrades, analyzing the genome of *R. varieornatus* - one of the most resistant to extreme conditions tardigrade species. Given that one of the crucial consequences of ionizing radiation is DNA damage, we suggested that key proteins providing tardigrades with high resistance to extreme factors could be associated with DNA. During our work, we focused on analysis and functional prediction of protein sequences extracted from the chromatin fraction.

Methods

Assembled genome of *R. varieornatus* was derived from http://kumamushi.org/data/. Gene prediction was provided via AUGUSTUS (http://bioinf.uni-greifswald.de/augustus/). To extract all protein sequences in fasta format, perl script getAnnoFasta.pl was run on gff annotation. List of peptide sequences of proteins associated with DNA were obtained from tandem mass spectrometry analysis of proteins extracted from the chromatin fraction. To find proteins from the *R. varieornatus* genome corresponding to these peptides, local blast search was performed via BLAST 2.9.0+ (Altschul et al. 1990). First, a local database from the protein fasta file of *R. varieornatus* was created with makeblastdb. Blastp of peptides associated with DNA was run against a local database (output format was specified as -outfmt 6; other parameters were set default). List of identifiers of subject sequences homologous to the query peptides were extracted via bash and saved as a text file. Respective protein sequences of *R. varieornatus* were extracted from protein fasta file using SAMtools v1.11 (samtools faidx) (Li et al. 2009). Localization of resulting proteins was verified using WoLF PSORT (Horton et al. 2007) and TargetP (Emanuelsson et al. 2007). For predicting protein functions based on sequence homology search, BLASTP (https://blast.ncbi.nlm.nih.gov) was used against the “UniProtKB/Swiss-Prot” database and *Ramazzottius varieornatus* was excluded from the search. Other parameters were set default. BLASTP results were cross-validated by hmmscan prediction (https://www.ebi.ac.uk/Tools/hmmer/search/hmmscan) using the Pfam database with default parameters. Data on results of BLASTP search, hmmscan, TargetP, and WoLF PSORT, was summarized into table (See Supplementary table 1).

Results

According to gene prediction results, we obtained 16435 protein sequences. Tardigrades could eliminate fatal damages induced by radiation due to unique mechanisms of DNA reparation, ensuring remarkable survival. As proteins responsible for DNA repair are associated with chromatin, we examined this possibility analyzing protein sequences extracted from the chromatin fraction. Using tandem mass spectrometry data on proteins from the chromatin fraction, we identified 34 proteins associated with DNA and subjected them to subsequent analysis.

Nuclear localization was confirmed only for 12 studied proteins based on maximum number of nearest neighbors to the query (Supplementary table 1). Other proteins could be mostly associated with cytoplasm (1), mitochondria (1), cytosol (1), and plasma membrane (6), or secreted into the extracellular space (13). We predicted signal peptides almost in all extracellular proteins (except for one protein: g5443.t1), and in a single plasma membrane protein (protein id: g11320.t1) (likelihood probability varied from 0.883840 up to 0.999923).

We revealed homologs for 23 studied proteins using BLASTP search, including 8 nuclear proteins (E-value of respective hits << 0.001). Sequence identity varied from about 25 up to 45% and query coverage ranged from 14 to 99%. For almost all proteins which functions were predicted via BLASTP, we could identify the Pfam family domain (Supplementary table 1).

Several proteins (g5467.t1,g5502.t1, g5503.t1, g5616.t1, g5641.t1, g702.t1, g15153.t1, g12562.t1, g1285.t1, and g12388.t1) were predicted as possessing chitinase binding domain. They could function as chitinases (Förster et al. 2012) being responsible for remodelling of chitin-containing structures during change between tun and active stages.

Discussion

Among 34 detected proteins the ones with predicted nuclear localization may be of great importance in terms of DNA repair. We found proteins that are known to be involved in chromatin remodeling during DNA damage response, such as E3 ubiquitin ligase and SWI/SNF-related protein. Moreover, we found a tardigrade-unique protein Dsup (protein id g14472.t1), which safeguards against DNA damage-induced genome instability.

E3 ubiquitin ligase with the RING domain

Protein g11960.t1 was predicted as E3 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 (Brinkmann et al. 2015). Protein g11960.t1 shared about 27 % of similarity with the protein BRE1B from Rattus sp.. The latter is a component of the RNF20/40 E3 ubiquitin-protein ligase complex involved in the DNA damage response (Brinkmann et al. 2015).

SWI/SNF

SWI/SNF complex is known to facilitate repair of double-strand DNA breaks in yeast, *C. elegans* and human cells by homologous recombination and non-homologous end-joining (Ribeiro-Silva, Vermeulen, and Lans 2019). 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 (Förster et al. 2012). We assume that in *Ramazzottius varieornatus* this complex may be involved in DNA breaks thus providing tardigrades survival after exposure to ionizing radiation. This assumption is confirmed by predicted nuclear localization of SWI/SNF-related protein (g7861.t1) (Supplementary table 1).

DSUP

Dsup protein has been previously described as a tardigrade-unique protein associated with DNA to protect DNA damage (Hashimoto et al. 2016). Colocalization of (GFP)-fused Dsup proteins with DNA was shown in Drosophila cell culture and in human cultured HEK 293T cells. Dsup protein suppresses X-ray-induced SSBs in human cultured cells and improves their viability. Protein sequence g14472.t1 from a studied dataset of nuclear proteins aligned to the Dsup sequence (UniProt: P0DOW4) with 100 % identity based on the Needleman-Wunsch algorithm.

Vps51

It was reported that Vps51 is needed for adaptation to the DNA damage checkpoint and could be involved in the autophagy in the yeast cells associated with DNA damage (Dotiwala et al. 2013). Hashimoto et al. (Hashimoto et al. 2016) suggested that in tardigrades autophagy, triggered by severe stress, is repressed leading to decreased destruction of cellular components. We assume that *R. varieornatus* might possess Vps51 protein with specific mutations allowing them to avert excessive amounts of intracellular destructive processes during exposure to high stress conditions.

There are several possible approaches to verify the role of mentioned proteins in tardigrades high tolerance towards ionizing radiation. Firstly differentially expressed genes may be described using RNA-seq data derived from two groups of animals: contros and exposed to radiation. Secondly we can knock out genes of these proteins by CRISPR/Cas approach and find out whether these proteins are indeed required for tardigrades survival.

Supplementary materials

Predicted proteins with additional information are available in Supplementary table 1. Detailed pipeline with code is provided here.

References

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. “Basic Local Alignment Search Tool.” *Journal of Molecular Biology* 215 (3): 403–10.

Boothby, Thomas C., Jennifer R. Tenlen, Frank W. Smith, Jeremy R. Wang, Kiera A. Patanella, Erin Osborne Nishimura, Sophia C. Tintori, et al. 2015. “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* 112 (52): 15976–81.

Brinkmann, Kerstin, Michael Schell, Thorsten Hoppe, and Hamid Kashkar. 2015. “Regulation of the DNA Damage Response by Ubiquitin Conjugation.” *Frontiers in Genetics* 6 (March): 98.

Dotiwala, Farokh, Vinay V. Eapen, Jacob C. Harrison, Ayelet Arbel-Eden, Vikram Ranade, Satoshi Yoshida, and James E. Haber. 2013. “DNA Damage Checkpoint Triggers Autophagy to Regulate the Initiation of Anaphase.” *Proceedings of the National Academy of Sciences of the United States of America* 110 (1): E41–49.

Emanuelsson, Olof, Søren Brunak, Gunnar von Heijne, and Henrik Nielsen. 2007. “Locating Proteins in the Cell Using TargetP, SignalP and Related Tools.” *Nature Protocols* 2 (4): 953–71.

Förster, Frank, Daniela Beisser, Markus A. Grohme, Chunguang Liang, Brahim Mali, Alexander Matthias Siegl, Julia C. Engelmann, et al. 2012. “Transcriptome Analysis in

Tardigrade Species Reveals Specific Molecular Pathways for Stress Adaptations." *Bioinformatics and Biology Insights* 6 (April): 69–96.
Hashimoto, Takuma, Daiki D. Horikawa, Yuki Saito, Hirokazu Kuwahara, Hiroko Kozuka-Hata, Tadasu Shin-I, Yohei Minakuchi, et al. 2016. "Extremotolerant Tardigrade Genome and Improved Radiotolerance of Human Cultured Cells by Tardigrade-Unique Protein." *Nature Communications* 7 (September): 12808.
Horton, Paul, Keun-Joon Park, Takeshi Obayashi, Naoya Fujita, Hajime Harada, C. J. Adams-Collier, and Kenta Nakai. 2007. "WoLF PSORT: Protein Localization Predictor." *Nucleic Acids Research* 35 (Web Server issue): W585–87.
Koutsovoulos, Georgios, Sujai Kumar, Dominik R. Laetsch, Lewis Stevens, Jennifer Daub, Claire Conlon, Habib Maroon, Fran Thomas, Aziz A. Aboobaker, and Mark Blaxter. 2016. "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*.
Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, Richard Durbin, and 1000 Genome Project Data Processing Subgroup. 2009. "The Sequence Alignment/Map Format and SAMtools." *Bioinformatics* 25 (16): 2078–79.
Ribeiro-Silva, Cristina, Wim Vermeulen, and Hannes Lans. 2019. "SWI/SNF: Complex Complexes in Genome Stability and Cancer." *DNA Repair* 77 (May): 87–95.