BIOINFORMATICS INSTITUTE

Decoding Tardigrade Resilience: Unraveling DNA Repair Mechanisms in Extreme Environments

December 9, 2023

Homework number 4

by Ilia Popov and Natalia Erofeeva

This study explores tardigrade genomics, focusing on DNA repair-related proteins to decipher stress resilience mechanisms. Computational methods identified 12 proteins with nuclear localization, including *g14472.t1*, a chromatin-associating protein. BLAST revealed *g14472.t1* as a 'Damage suppressor protein', while Pfam prediction found no domains. The discussion highlights *g14472.t1's* role in shielding DNA from radiation and oxidative stress, aligning with literature on tardigrade DNA-associated proteins. Our findings underscore *g14472.t1* as a crucial player in tardigrade stress tolerance, paving the way for deeper molecular investigations.

Introduction
Tardigrades, commonly known as water bears, represent microscopic, water-dwelling organisms within the phylum Tardigrada. These resilient creatures have captivated scientific curiosity due to their remarkable survival skills in extreme environments, enduring challenges such as desiccation, radiation, and fluctuating temperatures [1,2]. The exploration of tardigrade genomes has emerged as a pivotal avenue for unraveling the genetic intricacies that underscore their exceptional adaptability to harsh conditions [3]. Insight into the genetic landscape of tardigrades holds promise for identifying key genes and pathways associated with stress resistance, opening new avenues for biotechnological and medical advancements [1,2].

This study employs computational methodologies, focusing on gene prediction and sequence homology analysis, to probe the genomes of tardigrades [4]. These computational tools serve as powerful instruments for the identification and annotation of genes, particularly those integral to DNA repair mechanisms. The investigation into tardigrade DNA repair capabilities is paramount for comprehending their resilience against environmental stressors, such as radiation [1].

Materials and methods
Genome Retrieval:
We obtained the genomic data for *Ramazzottius varieornatus* (strain YOKOZUNA-1) from the National Center for Biotechnology Information (NCBI) database, specifically the assembly with accession number GCA_001949185.1 [5].

Structural Annotation:
To functionally annotate the genomic regions, we utilized precomputed AUGUSTUS results, including protein fasta and GFF files. The *getAnnoFasta.pl* script was employed to extract protein sequences from the AUGUSTUS prediction output [6].

Physical Localization:
We investigated the physical localization of proteins associated with DNA by performing a local alignment-based search. A local BLAST database was created from the protein fasta file obtained in the previous step. The blastp command was then used to search for peptides associated with DNA, as extracted from experimental data.

Localization Prediction:
Subcellular localization predictions were made using tools like WoLF PSORT and TargetP [7,8]. WoLF PSORT analyzes the presence of signal peptides, providing insights into the cellular localization of proteins. TargetP predicts localization based on the presence of N-terminal presequences. Results were obtained from the respective online servers.

BLAST Search:
BLAST searches against the UniProtKB/Swiss-Prot database were conducted to identify homologous sequences. Accession numbers, E-values, percent identity, and query coverage for each best hit were recorded [9].

Pfam Prediction:
Pfam predictions were performed using HMMER and the Pfam database [10,11]. This approach enabled us to identify protein domains and motifs in the selected protein sequences.
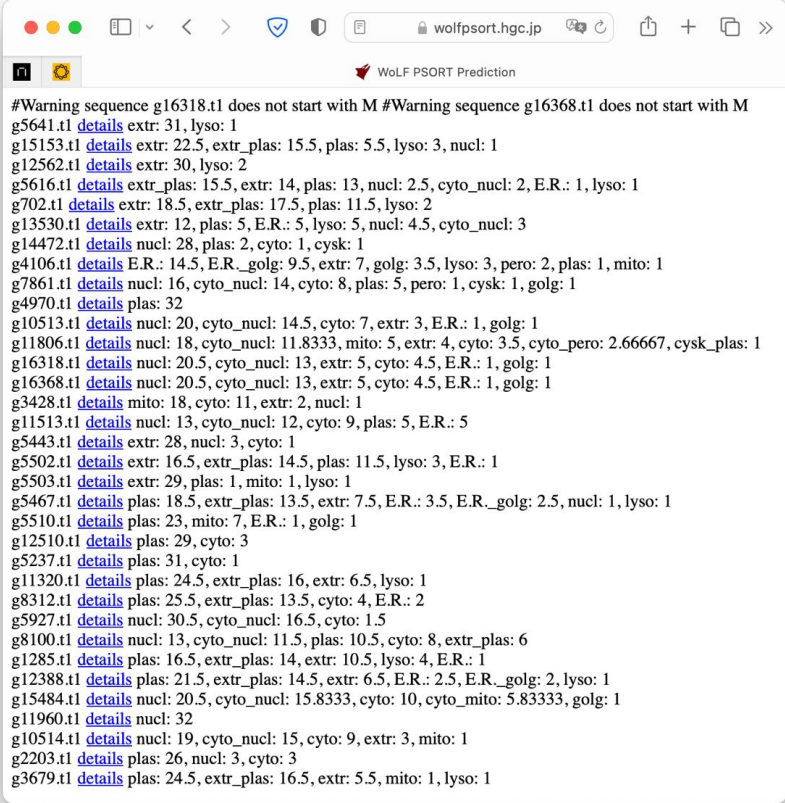
Results

Physical Localization:

During the Physical Localization phase, we initially obtained 118 protein sequences. To refine our dataset, we applied a sorting procedure to retain only unique protein sequences based on their respective IDs. As a result, we identified a subset of 34 unique protein sequences, ensuring a focused selection for further analyses.

Localization Prediction:

After applying WoLF PSORT for subcellular localization prediction, we strategically narrowed down our focus to proteins exhibiting nuclear localization. Consequently, we retained a subset of 12 protein sequences for further analysis, emphasizing their potential roles in nuclear processes.

Figure 1 illustrates the results of the initial WoLF PSORT analysis conducted on the set of 34 protein sequences. The figure provides a comprehensive overview of the diverse subcellular localization predictions for these proteins, highlighting the various cellular compartments they may be associated with. This initial analysis served as a crucial step in the selection process, guiding subsequent decisions to focus on proteins with nuclear localization for further investigation.

Figure 1. Initial WoLF PSORT results.

Figure 2 displays the refined WoLF PSORT results after meticulous sorting to include only proteins exhibiting nuclear localization. The figure underscores our focused approach, retaining 12 protein sequences that specifically localize to the cell nucleus. This stringent selection aligns with our overarching objective of scrutinizing genes related to DNA repair mechanisms.

Figure 2. WoLF PSORT results after sorting nuclear localization only.
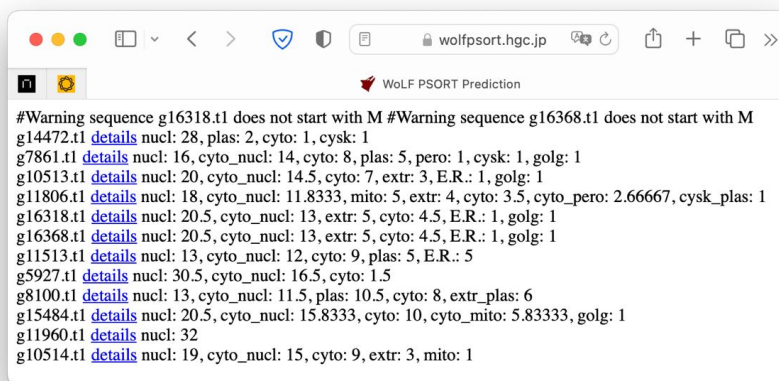
In addition to WoLF PSORT, we employed TargetP for further localization prediction, generating a comprehensive summary of the 12 selected protein sequences. The results indicate that all proteins fall into the 'Other' category, as none of them are associated with a 'Signal peptide' or 'Mitochondrial transfer peptide'. This outcome refines our focus, homing in on proteins with diverse cellular localizations within the 'Other' category. The absence of Signal peptides or Mitochondrial transfer peptides suggests that the identified proteins may have distinct roles within the cell, warranting a closer examination of their functions in subsequent analyses.

BLAST Search:

The BLAST search provided insights into the homology of our protein sequences. Out of the initial set of 12 protein sequences, BLAST successfully identified hits for 7 of them. Interestingly, only one protein, *g14472.t1*, obtained a best hit. The results for this protein are particularly significant, revealing a perfect E-value of 0.0 and annotation as a 'Damage suppressor protein' [12]. This emphasizes the potential importance of *g14472.t1* in the context of DNA repair mechanisms, indicating its role as a putative damage suppressor protein in tardigrades.

Figure 3. Graphic Summary on top 1 Blast Hit.
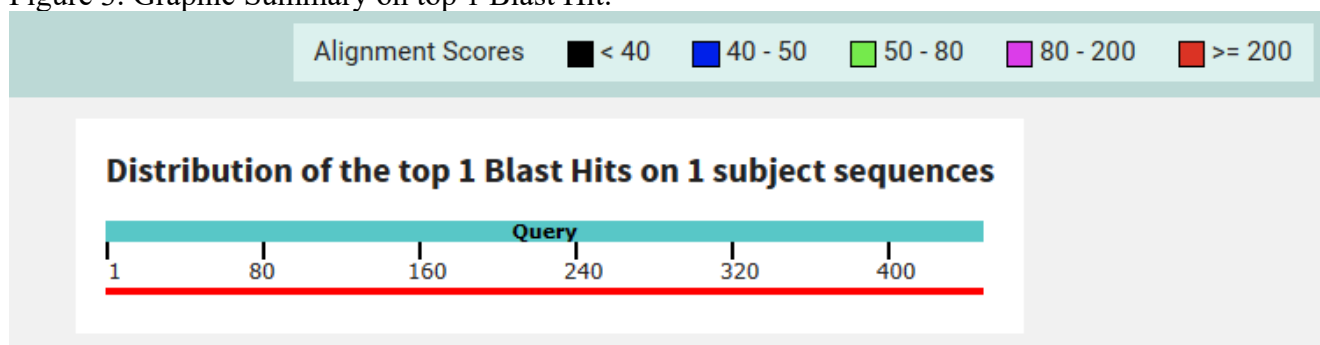
Pfam Prediction:

The Pfam prediction analysis aimed to identify potential protein domains within our set of sequences. However, the results were distinctive, revealing predicted Pfam domains for only 5 out of the 12 proteins, while the protein of interest, *g14472.t1*, showed no predicted Pfam domains.

Figure 4. Summary of Pfam prediction results.

Discussion

Our exploration of tardigrade genomes, specifically focusing on DNA repair-related proteins, has revealed intriguing candidates with potential implications for understanding the molecular mechanisms underlying tardigrades' exceptional stress resistance. In particular, the analysis of the chromatin-associating protein encoded by *g14472.t1* presents compelling evidence of its significant role in the organism's tolerance to harsh environmental stresses.

The unique chromatin-associating properties of *g14472.t1* distinguish it as a key player in the cellular response to environmental challenges. Previous studies have highlighted its remarkable affinity for nucleosomes compared to free DNA, suggesting a specific role in chromatin protection [2,13]. This property becomes especially pertinent in the context of our findings, as g14472.t1 demonstrates a higher binding affinity to nucleosomes, acting as a safeguard against damage caused by hydroxyl radical-mediated cleavage induced by X-rays or hydrogen peroxide treatment.

The observed suppression of X-ray-induced DNA damage, encompassing both single-strand breaks and more hazardous double-strand breaks, underscores the importance of *g14472.t1* in maintaining genomic integrity under genotoxic stress. This protective role aligns with our broader objective of identifying genes associated with DNA repair mechanisms in tardigrades. Furthermore, the improved radiotolerance attributed to *g14472.t1* signifies its potential as a key determinant of tardigrades' ability to withstand radiation, a notable environmental stressor.

The shielding effect of *g14472.t1* extends beyond X-ray-induced damage, encompassing protection against reactive oxygen species. This multifaceted defence mechanism positions *g14472.t1* as a central component in safeguarding the tardigrade genome from various sources of environmental stress.

To validate and extend our findings, future experiments could focus on elucidating the specific molecular interactions between *g14472.t1* and nucleosomes. Additionally, targeted studies could explore the downstream effects of *g14472.t1* in the context of other genetic mechanisms contributing to tardigrade stress resistance.

In conclusion, our study sheds light on a novel candidate, *g14472.t1*, as a potential key player in tardigrade stress resilience. The chromatin-associating protein exhibits a unique set of protective functions, emphasizing its significance in maintaining genomic stability under challenging environmental conditions. Further research into the precise molecular mechanisms and interactions involving *g14472.t1* will contribute to a more comprehensive understanding of tardigrade stress response strategies.

References

1. Boothby, T.C.; Tapia, H.; Brozena, A.H.; Piszkiewicz, S.; Smith, A.E.; Giovannini, I.; Rebecchi, L.; Pielak, G.J.; Koshland, D.; Goldstein, B. Tardigrades Use Intrinsically Disordered Proteins to Survive Desiccation. *Mol Cell* **2017**, *65*, 975-984.e5, doi:10.1016/j.molcel.2017.02.018.
2. Hashimoto, T.; Horikawa, D.D.; Saito, Y.; Kuwahara, H.; Kozuka-Hata, H.; Shin-I, T.; Minakuchi, Y.; Ohishi, K.; Motoyama, A.; Aizu, T.; et al. Extremotolerant Tardigrade Genome and Improved Radiotolerance of Human Cultured Cells by Tardigrade-Unique Protein. *Nat Commun* **2016**, *7*, 12808, doi:10.1038/ncomms12808.
3. Yoshida, Y.; Koutsovoulos, G.; Laetsch, D.R.; Stevens, L.; Kumar, S.; Horikawa, D.D.; Ishino, K.; Komine, S.; Kunieda, T.; Tomita, M.; et al. Comparative Genomics of the Tardigrades Hypsibius Dujardini and Ramazzottius Varieornatus. *PLOS Biology* **2017**, *15*, e2002266, doi:10.1371/journal.pbio.2002266.
4. Koutsovoulos, G.; Kumar, S.; Laetsch, D.R.; Stevens, L.; Daub, J.; Conlon, C.; Maroon, H.; Thomas, F.; Aboobaker, A.A.; Blaxter, M. No Evidence for Extensive Horizontal Gene Transfer in the Genome of the Tardigrade Hypsibius Dujardini. *Proceedings of the National Academy of Sciences* **2016**, *113*, 5053–5058, doi:10.1073/pnas.1600338113.
5. Index of /Genomes/All/GCA/001/949/185/GCA_001949185.1_Rvar_4.0 Available online: https://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/001/949/185/GCA_001949185.1_Rvar_4.0/ (accessed on 9 December 2023).
6. Augustus/Scripts/getAnnoFasta.Pl at Master · Nextgenusfs/Augustus Available online: https://github.com/nextgenusfs/augustus/blob/master/scripts/getAnnoFasta.pl (accessed on 9 December 2023).
7. WoLF PSORT: Protein Subcellular Localization Prediction Available online: https://wolfpsort.hgc.jp/ (accessed on 9 December 2023).
8. Almagro Armenteros, J.J.; Salvatore, M.; Emanuelsson, O.; Winther, O.; Von Heijne, G.; Elofsson, A.; Nielsen, H. Detecting Sequence Signals in Targeting Peptides Using Deep Learning. *Life Sci. Alliance* **2019**, *2*, e201900429, doi:10.26508/lsa.201900429.
9. BLAST: Basic Local Alignment Search Tool Available online: https://blast.ncbi.nlm.nih.gov/Blast.cgi (accessed on 9 December 2023).
10. HMMER Available online: http://hmmer.org/ (accessed on 9 December 2023).
11. Mistry, J.; Chuguransky, S.; Williams, L.; Qureshi, M.; Salazar, G.A.; Sonnhammer, E.L.L.; Tosatto, S.C.E.; Paladin, L.; Raj, S.; Richardson, L.J.; et al. Pfam: The Protein Families Database in 2021. *Nucleic Acids Research* **2021**, *49*, D412–D419, doi:10.1093/nar/gkaa913.
12. RecName: Full=Damage Suppressor Protein - Protein - NCBI Available online: https://www.ncbi.nlm.nih.gov/protein/P0DOW4.1?report=genbank&log$=prottop&blast_rank=1&RID=R86XC0MF013 (accessed on 9 December 2023).
13. Chavez, C.; Cruz-Becerra, G.; Fei, J.; Kassavetis, G.A.; Kadonaga, J.T. The Tardigrade Damage Suppressor Protein Binds to Nucleosomes and Protects DNA from Hydroxyl Radicals. *Elife* **2019**, *8*, e47682, doi:10.7554/eLife.47682.