Revealing genetic mechanisms behind *Ramazzottius varieornatus* resilience to environmental stresses.

Michael Greenberbg, Polina Vaganova
Bioinformatics Institute, Saint Petersburg, Russia

Abstract. Tardigrades exhibit exceptional resilience to extreme environmental conditions, yet the genetic basis of this resilience remains elusive. In this study we explore the genome of *Ramazzottius varieornatus* (YOKOZUNA-1 strain) to identify genes involved in DNA repair. Utilizing gene prediction, functional annotation and protein localization tools we identified several potential nuclear proteins, possibly involved in DNA reparation and protection. Among these proteins we found SWI/SNF ortholog, alongside unique Tardigrada-specific proteins like Dsup, and 3 unidentified novel proteins that require further investigation.

Keywords: *Ramazzottius varieornatus*, gene prediction, functional annotation, protein localization, radiation tolerance.

1 Introduction

Tardigrades are microscopic animals from Ecdysosoa group. They are mostly known for some of their representatives, like *Ramazzottius varieornatus* or *Hypsibius dujardini* which are showing exceptional resilience because of their cryptobiosis ability. These animals are able to resist various extreme environmental factors, including low oxygen levels, desiccation,[1] both high and low temperature,[2] intense pressure[3] and high radiation.[4] Even when tardigrades are not in their cryptobiotic states they are able to withstand extreme UV levels.[5]

Several full genome studies were conducted to identify reasons behind Tardigrada unique survivability. In some studies, a high level of horizontal gene transfer in the genome of tardigrades was shown,[6] however, this was later refuted.[7] Analyzing the genome of tardigrades, with a focus on gene orthologs associated with DNA repair and other cellular mechanisms ensuring their resilience, can offer insights into their exceptional survival abilities. Additionally, finding novel unique genes specific to Tardigrada using gene prediction approaches may reveal key elements contributing to their resistance.

The objective of this study is the identification of candidate genes within the *Ramazzottius varieornatus* genome (YOKOZUNA-1 strain) that potentially play a role in DNA reparation.

2 Materials and methods

For this study we used previously assembled genome of the *Ramazzottius varieornatus* (YOKOZUNA-1 strain).[8] To focus on region of interest we used list of peptides that were associated with the DNA of *Ramazzottius varieornatus*. Peptides from this list were obtained via tandem mass-spectrometry.

Functional annotation of genome was made using `AUGUSTUS`[9] software. To extract the list of proteins from the `AUGUSTUS` output the script provided by `AUGUSTUS` developers was used.

For the alignment of chromatin-associated peptides from the list to the extracted proteins `BLAST` command line tool[10] (version 2.12.0) and `samtools`[11] (version 1.3.1) software were used.

For aligned proteins localization prediction we used web-services `WoLF PSORT`[12] and `TargetP`.[13] We performed search of aligned proteins in the UniProtKB/Swiss-Prot database among `Metazoa (taxid:33208)` data with `blastp`[14] web service. We also used `HMMER` web-service[15] to reveal functional domains and motifs in aligned proteins.

Programs listed here were launched with default settings.

3 Results

The assembled genome of R. varieornatus was annotated using the `AUGUSTUS`[9] software, resulting in the identification of **16435** coding DNA sequences. Subsequently, a local database was constructed for aligning tandem mass spectrometry-identified peptides to pinpoint genes encoding potential chromatin-associated proteins, leading to the identification of **34** candidate genes. To determine the subcellular localization of these candidates, `WoLF PSORT`[12] and `TargetP`[13] tools were employed to detect signal peptides within the sequences. Furthermore, a BLAST search was conducted to find orthologs of the potential candidate genes. Finally, domains within the translated candidate genes were predicted.

Table 1: *Number of predicted proteins*

Predicted proteins according to `AUGUSTUS`	16435
Proteins aligned to chromatin-associated peptides	34
Nuclear proteins accroding to `WoLF PSORT` and `TargetP`	12
Proteins with identified orthologues from UniProtKB/Swiss-Prot database	25
Proteins with identified domains according to `HMMER`	20

Based on our analysis, we discovered **12** genes whose protein products are potentially localized within the nucleus. However, among these, **3** genes lacked annotation. Summary table for all 34 candidate genes can be found in supplementary materials.

4 Discussion

We obtained several DNA-associated proteins of *Ramazzottius varieornatus* which potentially related to the DNA repair and protection resulting in its high resilience to environmental stresses.

First of all, we identified Damage suppressor protein (Dsup) that were previously found in *R. varieornatus* by Hashimoto et al.[4]This specific protein unique to tardigrades has been suggested to suppress X-ray-induced DNA damage and enhance resistance to radiation. These findings were demonstrated through experiments conducted using transgenic HEK293 cell

cultures expressing Dsup, enabling them to withstand higher doses of radiation.[4] Nevertheless, the level of resistance exhibited by tardigrade cells was significantly greater, suggesting the involvement of additional mechanisms in this process.

Another annotated protein was found to be from the family of SWI/SNF chromatin remodelling complexes. SWI/SNF complexes contribute to successful DNA repair via both the homologous recombination and non-homologous end joining.[16] Moreover, SWI/SNF activity contributes to the repressing transcription near double strand breaks thus can affect the reparation process.[16]

Finally, we have identified 3 nuclear proteins lacking any annotation. It's plausible to speculate that these proteins might also contribute to the unique resilience observed in *R. varieornatus*, however further investigation are needed. For example, labeling the protein with fluorescence tag can reveal its localisation, confirming its association with the nucleus. To assess the protein function linked to DNA repair, an experiment involving exposing transgenic cell lines to a stressful environment can be used.

References

1 J. F. Fleming, D. Pisani, and K. Arakawa, "The evolution of temperature and desiccation-related protein families in tardigrada reveals the complex acquisition of extremotolerance," *Genome Biology and Evolution* (2023).

2 S. Hengherr, M. Worland, A. Reuner, *et al.*, "High-temperature tolerance in anhydrobiotic tardigrades is limited by glass transition," *Physiological and Biochemical Zoology* **82**, 749–755 (2009).

3 K. Seki and M. Toyoshima, "Preserving tardigrades under pressure," *Nature* **395**, 853–854 (1998).

4 T. Hashimoto, D. D. Horikawa, Y. Saito, *et al.*, "Extremotolerant tardigrade genome and improved radiotolerance of human cultured cells by tardigrade-unique protein," *Nature Communications* **7** (2016).

5 T. Altiero, R. Guidetti, V. Caselli, *et al.*, "Ultraviolet radiation tolerance in hydrated and desiccated eutardigrades," *Journal of Zoological Systematics and Evolutionary Research* **49**, 104–110 (2011).

6 T. C. Boothby, J. R. Tenlen, F. W. Smith, *et al.*, "Evidence for extensive horizontal gene transfer from the draft genome of a tardigrade," *Proceedings of the National Academy of Sciences* **112**, 15976–15981 (2015).

7 G. Koutsovoulos, S. Kumar, D. R. Laetsch, *et al.*, "No evidence for extensive horizontal gene transfer in the genome of the tardigrade hypsibius dujardini," *Proceedings of the National Academy of Sciences* **113**, 5053–5058 (2016).

8 R. Bertolani and I. M. Kinchin, "A new species of ramazzottius (tardigrada, hypsibiidae) in a rain gutter sediment from england," *Zoological Journal of the Linnean Society* **109**, 327–333 (1993).

9 M. Stanke and B. Morgenstern, "Augustus: a web server for gene prediction in eukaryotes that allows user-defined constraints," *Nucleic Acids Research* **33**, W465–W467 (2005).

10 G. M. Boratyn, C. Camacho, P. S. Cooper, *et al.*, "Blast: a more efficient report with usability improvements," *Nucleic Acids Research* **41**, W29–W33 (2013).

11 P. Danecek, J. K. Bonfield, J. Liddle, *et al.*, "Twelve years of SAMtools and BCFtools," *GigaScience* **10** (2021).

12 P. Horton, K.-J. Park, T. Obayashi, *et al.*, "Wolf psort: protein localization predictor," *Nucleic Acids Research* **35**, W585–W587 (2007).

13 J. J. Almagro Armenteros, M. Salvatore, O. Emanuelsson, *et al.*, "Detecting sequence signals in targeting peptides using deep learning," *Life Science Alliance* **2**, e201900429 (2019).

14 "Database resources of the national center for biotechnology information," *Nucleic Acids Research* **41**, D8–D20 (2012).

15 R. D. Finn, J. Clements, and S. R. Eddy, "Hmmer web server: interactive sequence similarity searching," *Nucleic Acids Research* **39**, W29–W37 (2011).

16 A. Harrod, K. A. Lane, and J. A. Downs, "The role of the swi/snf chromatin remodelling complex in the response to dna double strand breaks," *DNA Repair* **93**, 102919 (2020).

5 Supplementary materials

Data availability

- Table with all predicted proteins and their annotations can be found here.