Practice

BIOINFORMATICS INSTITUTE

Research of the unique DNA repair protein in the genome of a tardigrade *Ramazzottius varieornatus*

Ekaterina Grigorieva[1, 2] and Mikhail Filippov[1, 3]

[1]Bioinformatics Institute
[2]Moscow Institute of Physics and Technology
[3]Herzen University

Abstract

In this study we take a close look at the genome of Ramazzottius varieornatus, one of the most persistent tardigrade, and review its unique proteins. For that purpose we predicted R. varieornatus genes and compared their products with proteins of SwssProt database. This resulted in number of revealed proteins with no known homology, which structure and function was predicted based on their amino-acid sequence. That approach revealed an unique protein that is not known in other organisms, which can be linked to DNA repair system.

Keywords: tardigrade,

Introduction

Tardigrada is a Ecdysozoan phylum sister to *Arthropoda* , *Nematoida* and *Onychophora* (figure 1) [Campbell et al., 2011] that has some unique features. Notably, tardigrades are one of the most persistent eukaryotes, capable to survive rough radioactive exposure, pressure drops, hypothermia, dehydration and other impacts, lethal to other beings. Researchers looked for the source of such abilities in this animal's genome, and one of the tardigrade species, *Hypsibius dujardini* was chosen to perform de novo assembly of genome. Strikingly, very large of alien genes was observed (17,5%) and this fact was initially interpreted as the consequence of extremely high rate of *horizontal gene transfer* (HGT) [Boothby et al., 2015]. Later this assumption was refuted and such a large amount of foreign genes observed in original study was thought to be a result of contamination [Koutsovoulos et al., 2016]. Still, further research on the one of the most durable tardigrades, *Ramazzottius varieornatus*, have shown another unique feature of tardigrade genome: this creatures have unique proteins linked to the DNA repair system, which allow them to survive severe DNA damage [Hashimoto et al., 2016].

In this study we reproduced the basic workflow oh the researchers that originally studied *R. varieornatus* unique DNA repair proteins. We have predicted tardigrade's genes *ab initio*, obtained amino-acid sequences of genes' protein products and predicted their functions based on sequence homology, where the main idea is that if proteins or parts of proteins have similar sequence, they functions may also be similar.

Materials and methods

R. varieornatus is avaiable for download in GenBank database:

URL: https://www.ncbi.nlm.nih.gov/assembly/GCA_001949185.1/

Assembly is 55,842,812 b.p. long with 200 scaffolds. **N50** = 4,740,345, **L50** = 4.

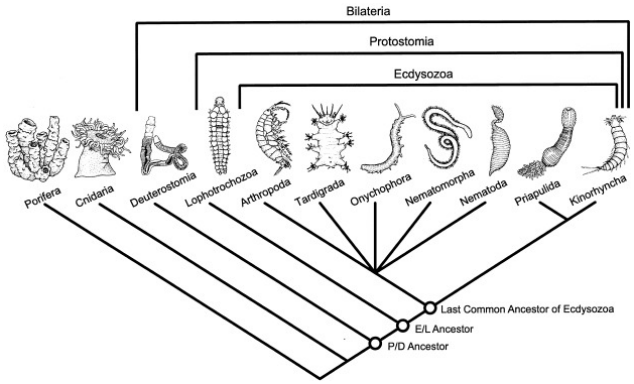

Figure 1 *Ecdysozoa* phylogeny according to Valentine et al.

Gene prediction and annotation was performed with **AUGUSTUS 3.3.3** [Stanke et al., 2008]. *Caenorhabditis elegans* was chosen for prediction training as an organism of sister for tardigrades group - Additional information of protein localization was obtained using tandem mass spectrometry. Peptides that was shown to bind DNA was aligned on the proteome with **BLAST 2.5.0** (blastp) [Altschul et al., 1999]. Subcellular localization of the proteins that appeared to bind DNA was predicted with **WoLF PSORT** [Horton et al., 2007] and **TargetP 2.0** [Almagro Armenteros et al., 2019] based on signal peptides. We predicted functions of the proteins using **HMMER** [Simon C Potter et al., 2018].

Results

AUGUSTUS run trained on *C. elegans* resulted in 17769 predicted proteins. Tandem mass spectrometry revealed 43 short peptides linked to DNA, which were aligned with BLASTp on 34 proteins of the reference proteome. Annotation provided by BLASTp and

predicted with WoLF PSORT and TargetP subcelluelar localization shown in the *table 1* at the very end of this document.

Discussion

As none of the proteins with known homology or predicted functions cannot be linked to DNA repair sequence, we must be interested in some protein that has not shown significant result in BLAST. Our candidate should have predicted nuclear localization and should not have predicted functional domains or known homology revealed with BLAST. Such proteins are:

- g6469.t1
- g10817.t1
- g11439.t1
- g12813.t1

But one of discovered proteins, g15612.t1 of referred table 1, is already annotated because of original research results was submitted in SwissProt database. Let's focus on the situation when for obvious reasons we have not that ready-to-go annotation available.

Information about exact function of the protein now annotated as Damage suppressor protein (*Dsup*) of *R. varieornatus* was totally unknown for the authors of original research. Obviously, there is no available methods for *in silico* prediction of functions of totally novel proteins when only their sequence is known. So, first step in function clarification was to confirm nuclear localization, that authors, as do we, predicted based on C-terminus sequence. Co-expression of GFP-marked *Dsup* with definitely nuclear proteins showed that unlike other DNA-binding proteins reviewed in table 1, *Dsup* for sure localizes in the nuclei. Second step is to link *Dsup* expression to the life stages of the tardirade where DNA repair is critical, like early embryonic period. That can be determined immunohistochemically and with transcriptome analysis on the tardigrade embryos. It was found that almost all tardigrade embryos are highly positive for *Dsup*. Next step was to research the ability of Dsup to bind with DNA, and that can be performed with gel-shift assay. Finally, when it is known that *Dsup* is for sure linked with DNA, authors confirmed its ability to enhance DNA repair and viability of cells with series of tests on HEK293 cell cultures [Hashimoto et al., 2016]. At this moment it is known that *Dsup* both tightly co-localized with the DNA and enhances DNA repair even when expressed in plants [Kirke et al., 2020], but exact mechanisms of that abilities is still not known. Research of structure-function relation is a probable field of great interest. One of the latest attempts to determine structure of the *Dsup*-DNA complex resulted in assumption that it as flexible system and *Dsup* adapts to the 3D-structure of the DNA (figure 2) [Mínguez-Toral et al., 2020]. Still, it is totally unclear how exactly it repairs DNA and which DNA breaks this protein tends to repair. Overall, there is a lot more to know!

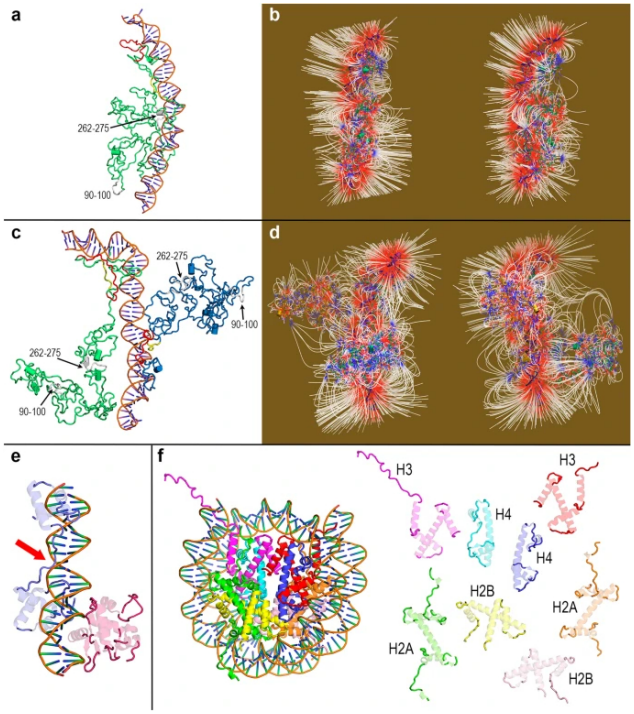

Figure 2 *Dsup* 3D structure according to Mínguez-Toral et al.

Supplementary data

Here below you can find supplementary materials. The code, raw data and results are available at

https://github.com/grigorievaekaterina/BI_Project_4

https://github.com/quasicephalus/Bioinformatics-workshop

References

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) "Basic local alignment search tool." J. Mol. Biol. 215:403-410
- Almagro Armenteros JJ, 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 Sep 30;2(5):e201900429. doi: 10.26508/lsa.201900429. PMID: 31570514; PMCID: PMC6769257.
- Simon C Potter, Aurélien Luciani, Sean R Eddy, Youngmi Park, Rodrigo Lopez, Robert D Finn, HMMER web server: 2018 update, Nucleic Acids Research, Volume 46, Issue W1, 2 July 2018, Pages W200–W204, https://doi.org/10.1093/nar/gky448
- Boothby TC, Tenlen JR, Smith FW, Wang JR, Patanella KA, Nishimura EO, Tintori SC, Li Q, Jones CD, Yandell M, Messina DN, Glasscock J, Goldstein B. Evidence for extensive horizontal gene transfer from the draft genome of a tardigrade. Proc Natl Acad Sci U S A. 2015 Dec 29;112(52):15976-81. doi: 10.1073/pnas.1510461112. Epub 2015 Nov 23. Erratum in: Proc Natl Acad Sci U S A. 2016 Sep 6;113(36):E5364. PMID: 26598659; PMCID: PMC4702960.
- Campbell, Lahcen I.; Rota-Stabelli, Omar; Edgecombe, Gregory D.; Marchioro, Trevor; Longhorn, Stuart J.; Telford, Maximilian J.; Philippe, Hervé; Rebecchi, Lorena; Peterson, Kevin J.; Pisani, Davide (2011). "MicroRNAs and phylogenomics resolve the relationships of Tardigrada and suggest that velvet worms are the sister group of Arthropoda". Proceedings of the National Academy of Sciences. 108 (38): 15920–24. Bibcode:2011PNAS..10815920C. doi:10.1073/pnas.1105499108. PMC 3179045. PMID 21896763.
- Hashimoto T, Horikawa DD, Saito Y, Kuwahara H, Kozuka-Hata H, Shin-I T, Minakuchi Y, Ohishi K, Motoyama A, Aizu T, Enomoto A, Kondo K, Tanaka S, Hara Y, Koshikawa S, Sagara H, Miura T, Yokobori SI, Miyagawa K, Suzuki Y, Kubo T, Oyama M, Kohara Y, Fujiyama A, Arakawa K, Katayama T, Toyoda A, Kunieda T. Extremotolerant tardi-

grade genome and improved radiotolerance of human cultured cells by tardigrade-unique protein. Nat Commun. 2016 Sep 20;7:12808. doi: 10.1038/ncomms12808. PMID: 27649274; PMCID: PMC5034306.

- Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, Nakai K. WoLF PSORT: protein localization predictor. Nucleic Acids Res. 2007 Jul;35(Web Server issue):W585-7. doi: 10.1093/nar/gkm259. Epub 2007 May 21. PMID: 17517783; PMCID: PMC1933216.
- Kirke J, Jin XL, Zhang XH. Expression of a Tardigrade Dsup Gene Enhances Genome Protection in Plants. Mol Biotechnol. 2020 Dec;62(11-12):563-571. doi: 10.1007/s12033-020-00273-9. Epub 2020 Sep 21. PMID: 32955680.
- Koutsovoulos G, Kumar S, Laetsch DR, Stevens L, Daub J, Conlon C, Maroon H, Thomas F, Aboobaker AA, Blaxter M. No evidence for extensive horizontal gene transfer in the genome of the tardigrade Hypsibius dujardini. Proc Natl Acad Sci U S A. 2016 May 3;113(18):5053-8. doi: 10.1073/pnas.1600338113. Epub 2016 Mar 24. PMID: 27035985; PMCID: PMC4983863.
- Mario Stanke, Mark Diekhans, Robert Baertsch, David Haussler (2008) Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics doi: 10.1093/bioinformatics/btn013
- Mínguez-Toral, M., Cuevas-Zuviría, B., Garrido-Arandia, M. et al. A computational structural study on the DNA-protecting role of the tardigrade-unique Dsup protein. Sci Rep 10, 13424 (2020). https://doi.org/10.1038/s41598-020-70431-1
- Valentine JW, Collins AG. The significance of moulting in Ecdysozoan evolution. Evol Dev. 2000 May-Jun;2(3):152-6. doi: 10.1046/j.1525-142x.2000.00043.x. PMID: 11252571.

Acknowledgments

We thank Mike Rayko for supervising and supporting us.

Table 1 List of proteins with predictions

Protein ID	BLAST annotation	Species	e-value	predicted Pfam domains	Localization (WoLF PSORT)	Localization (TargetP)
g763.t1	U-scoloptoxin(01)-Er1a	Ethmostigmus rubripes	6E-11	Chitin binding Peritrophin-A domain	plas	Other
g2369.t1	Myogenesis-regulating glycosidase	Mus musculus	6E-115	Glycosyl hydrolases family 1	plas	Other
g3715.t1	Myosin regulatory light chain sqh	Drosophila melanogaster	4E-59		nucl	Other
g3972.t1	Nucleotide exchange factor SIL1	Homo sapiens	6E-28	Astacin (Peptidase family M12A)	plas	Other
g4276.t1					plas	Other
g4446.t1					plas	Other
g5690.t1					cyto	Other
g5724.t1					plas	Signal peptide
g5943.t1					extr	Other
g5953.t1	Peptidyl-glycine alpha-amidating monooxygenase	Bos taurus	8E-129	Copper type II ascorbate-dependent monooxygenase, C-terminal domain	plas	Signal peptide
g6016.t1	rRNA 2′-O-methyltransferase fibrillarin	Drosophila erecta	3E-121	Fibrillarin	nucl	Other
g6469.t1					nucl	Other
g8583.t1	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A-like protein 1	Rattus norvegicus	8E-70	SNF2-related domain	nucl	Other
g9086.t1	Vacuolar protein sorting-associated protein 41 homolog	Homo sapiens	0	Region in Clathrin and VPS	cyto_nucl	Other
g9432.t1	Probable aminopeptidase NPEPL1	Homo sapiens	3E-146	Cytosol aminopeptidase family, catalytic domain	plas	Other
g10655.t1					mito	Other
g10817.t1					nucl	Other
g11439.t1					nucl	Other
g12290.t1	Chondroitin proteoglycan 1	Caenorhabditis briggsae	0.025		plas	Signal peptide
g12492.t1	Trafficking protein particle complex subunit 9	Bos taurus	2E-58	Transport protein Trs120 or TRAPPC9, TRAPP II complex subunit	E.R.	Other
g12813.t1					nucl	Other
g12974.t1	E3 ubiquitin-protein ligase BRE1B	Rattus norvegicus	1E-51		nucl	Other
g13562.t1					plas	Other
g14655.t1					extr	Signal peptide
g15612.t1	**Damage suppressor protein**	**Ramazzottius varieornatus**	**0**		**nucl**	**Other**
g16734.t1	Vacuolar protein sorting-associated protein 51 homolog	Xenopus laevis	8E-147	Vps51/Vps67	cyto	Other
g17688.t1	Eukaryotic translation initiation factor 3 subunit A	Xenopus tropicalis	1E-05		nucl	Other