BIOINFORMATICS
INSTITUTE

Tardigrades: piglets that bear a unique DNA protection

Polina Guseva[1, †] and Nikita Vyatkin[1, 2, †]

[1]Bioinformatics Institute, Saint-Petersburg, Russia
[2]Saint-Petersburg State University, Russia

[†]Contributed equally.

Abstract

Being such small creatures tardigrades possess a unique ability to survive extreme conditions no matter what. Unrevealing their pathways to prevent excessive DNA damage could have a potentially astonishing impact. In this project, proteins from chromatin of *Ramazzottius varieornatus* are cross-linked to its annotated genome. The matched reads are filtered by location in the nucleus and functional properties based on homology and conservative domains. Among candidates with available description, none is suitable for DNA repair or damage prevention. Further research on the rest is needed as well as higher-quality sequencing.

Key words: Tardigrada, DNA damage response, stress tolerance, extremotolerance

Introduction

Tardigrades are meiofaunal aquatic Ecdyzozoa used to be extremotolerant since their environment could be especially harsh [1]. Therefore, scientists all over the world are curious how these water bears got through a such amount of stress and survived all five mass extinctions and open space [2]. Firstly, this adaptability was suggested to be due to horizontal gene transfer [3], yet lately this hypothesis was rejected because of contamination [4]. Hence unique tardigradian features are tried to be deciphered by searching for sequence or domain homology among other eukaryotic genes *in silico* with a potential *in vivo* conformation. The most recent update about tardigrade genome studies is written by K. Arakawa [5].

DNA damage could happen due to cellular metabolism, exposure to physical (UV light, radiation, dehydration, pressure, temperature) and chemical (oxidation (ROS), hypoxia, genotoxins) stresses, replication errors, or spontaneously [6], [7]. Therefore, the vast repair mechanism (DDR, or DNA Damage Response) includes cell cycle checkpoint activation, transcriptional program activation, direct DNA repair, or apoptosis as a last option [8], [9]. Yet it is better to prevent such damage by conserving and protecting DNA structure as the Dsup protein from tardigrades [10]. All this DNA-associated machinery has distinctive conservative domains that recognise and/or bind DNA [11]. For example, helix-turn-helix (HTH including winged HTH) is the most common in regulation factors [12], zinc-coordinating proteins (zinc fingers) are usually found in transcription factors [13], zipper-type proteins (including leucine zipper and helix-loop-helix family) position in a major groove and regulate too [14], high-mobility group box (HMG or HMG-box) are more flexible and involve in replication and transcription [15], and so many others [16] including RNA-containing structures [17]. These motives could be found in various different enzymes since they just perform DNA-binding functions.

In this paper, we hypothesise that some proteins from chromatin fraction of *Ramazzottius varieornatus* could be recognised as potential targets for experimental selection to identify the source of unique tardigradian DNA stress resistance. To prove that we 1) filter from the annotated genome only protein from chromatin, 2) the rest is additionally cleaned by signal peptide and N-terminal presequences to get rid of non-nuclear ones, 3) last but not least a function of potential candidates is suggested by homologous search and conservative domain exploration.

Methods

Hereby the already assembled genome of a tardigrade *Ramazzottius varieornatus* (Bertolani and Kinchin, 1993), the YOKOZUNA-1 strain (GenBank ID #947166) is used as a model species. Moreover, a chromatin fraction is extracted and obtained proteins are analysed by tandem mass spectrometry.

First of all, the genome is functionally annotated using homology patterns and conservative domains by AUGUSTUS v3.2.3 [18]. Solely coding sequences are chosen for further analysis.

To target only proteins connected to DNA maintenance, several approaches are implemented to filter only relevant ones. Thus, reads from a chromatin fracture are cross-linked to the annotated genes by Python v3.7, Protein-Protein BLAST v2.12.0 [19], and Diamond v2.0.15.153 [20] (`-very-sensitive`). As BLAST works both with negligible mutations via BLOSUM and with short reads, hereinafter results from it are used since they have more biological sense. To locate only nuclear proteins, WoLF PSORT [21] searches via signal peptides throughout all groups of live organisms (animals, plants, and fungi) to catch any outliers. Whereas, TargetP v2.0 [22] predicts based on N-terminal presequences with plants and non-plants options for extra comparison. Filtered out by these estimations genes are compared based on homology by BLAST to those described in the UniProtKB/Swiss-Prot database [23]. Only proteins with identity more than 20% are chosen. Furthermore, to identify functions of not discovered yet proteins, HmmerWeb v2.43

Compiled on: December 17, 2022.
Draft manuscript prepared by the author.

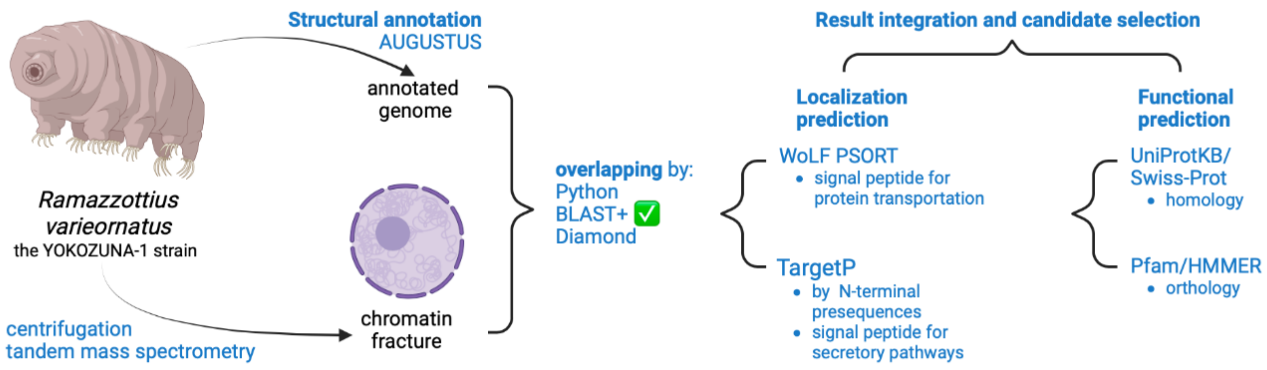

Figure 1. Methods: graphic summary

[24] scans through domains by looking at orthologous sequences. If the opposite is not stated, the default settings are applied. After all, the most suitable proteins are chosen with the best hits among acquired outputs from all tests.

Results

Overall, 29 peptide reads from tandem mass spectrometry are overlaid to 16435 annotated genes within the 56.5 Mbp genome. Among unique intersecting ones, there are 35 found by Python merging, 4 by Diamond, and 34 by BLAST.

For narrowing down potential targets by location, only 15 (12 for animals) proteins most commonly for a cell nucleus are selected from WoLF PSORT (tab.2). While TargetP allocates 21 proteins outside of any transit pathways (tab.2). Overlapping the gained lists there are 12 shared proteins (tab. 3). Also, to filter specifically DNA-associated proteins, there are 7 suitable homologous sequences in BLAST (tab.4) and 5 reads fitting the description domain by Pfam (tab.5). In the end, at least some functional information is retrieved for 7 proteins (tab.1).

After all filtering parameters are combined just two proteins could potentially interact with DNA. However, none of them involves in DNA stress resistance and DDR pathways (tab.1).

Discussion

Such a low coverage of domain and homology recognition (7/12 reads, tab.1) could be explained by unavailable sequence parts within several reads. Thus, proteins g10513.t1 and g10514.t1 have continuous regions of uncertain content (X). Therefore, further research should be conducted with increased quality requirements or genome/transcriptome investigation with sufficient coverage.

As for a functional prediction, homologous search does not reveal much for target verification due to a low identity score, whereas domain description could already be more helpful. Possibly homology acquired by BLAST contains such a low similarity rate due to insufficient information about invertebrates and their proteins. Nevertheless, the g7861.t1 protein is likely involved in the nucleosome rearrangement and chromatin remodelling leading to regulating of gene expression [25]. Moreover, it has a sequence similar to a HARP domain that works as a helicase [26] and less feasible (e-value= $6.1 * 10^{-6}$) has an endonuclease activity (tab.1). That could be important for an effective response to environmental stressors. Furthermore, g11960.t1 also presumably interacts with DNA via zinc (RING) fingers and regulates suppressors via ubiquitination [27]. While other candidates are highly unlikely to have any interactions with DNA since the loss of any DNA-binding domains but a machinery for signalling, vacuolar transport, or translation (g8100.t1 [28], g8312.t1, g15484.t1, g16318.t1, and g16368.t1).

Since none of discussed functionally annotated proteins performs any DNA repair or protection, further research could focus on other proteins from the list (g5927.t1, g10513.t1, g10514.t1, g11806.t1, g14472.t1). Actually, this approach was implemented and the g14472.t1 protein turned out to reveal a unique radiotolerance ability [29]. The potential next step in protein detection could be RNA interference to silence a protein and monitor its impact.

Table 1. Summary table for filtered proteins from chromatin fraction

Query Name	Best BLAST Hit				Predicted Pfam Domains			Nuclear localization probability (WoLF PSORT), %
	Subject Name	Organism	Identity, %	E-value	Family Id	Description	E-value	
g5927.t1	-	-		-	-	-	-	95.31
g7861.t1	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A-like protein 1	*Rattus norvegicus*	37.21	1.55e-71	SNF2-rel_dom HARP ResIII	SNF2-related domain HepA-related protein (HARP) Type III restriction enzyme, res subunit	1.2e-28 2.6e-10 6.1e-06	50.0
g8100.t1	Inositol monophosphatase 3	*Danio rerio*	36.04	2.96e-46	Inositol_P MKLP1_Arf_bdg	Inositol monophosphatase family Arf6-interacting domain of mitotic kinesin-like protein 1	1.9e-37 5.1e-27	51.56
g8312.t1	Vacuolar protein sorting-associated protein 41homolog	*Mus musculus*	40.84	0.0	Clathrin	Region in Clathrin and VPS	5.4e-23	48.44
g10513.t1	-	-		-	-	-	-	62.5
g10514.t1	-	-		-	-	-	-	59.38
g11806.t1	-	-		-	-	-	-	56.25
g11960.t1	E3 ubiquitin-protein ligase BRE1B	*Rattus norvegicus*	26.96	6.13e-98	zf-C3HC4	Zinc finger, C3HC4 type (RING finger)	4.2e-05	100.0
g14472.t1	-	-		-	-	-	-	87.5
g15484.t1	Vacuolar protein sorting-associated protein 51 homolog	*Danio rerio*	45.03	0.0	Vps51 Sec5 Dor1 Vps54_N COG2	Vps51/Vps67 Exocyst complex component Sec5 Dor1-like family Vacuolar-sorting protein 54, of GARP complex COG (conserved oligomeric Golgi) complex component, COG2	1.3e-23 3.4e-23 1.2e-11 2.4e-10 2.5e-06	54.69
g16318.t1	Eukaryotic translation initiation factor 3 subunit A	*Xenopus laevis*	36.11	4.09e-08	-	-	-	64.06
g16368.t1	Eukaryotic translation initiation factor 3 subunit A	*Xenopus tropicalis*	39.29	1.20e-05	-	-	-	64.06

All listed above proteins have the **other** signal in the TargetP output

References

1. Møbjerg N, Halberg K, Jørgensen A, Persson D, Bjørn M, Ramløv H, et al. Survival in extreme environments–on the current knowledge of adaptations in tardigrades. Acta physiologica 2011;202(3):409–420. Doi:10.1111/j.1748-1716.2011.02252.x.
2. Lațici T. From one master of survival to another: a tardigrade's plea for NATO2030. Egmont Royal Institute for International Relations 2021;Doi:10.1111/j.1748-1716.2011.02252.x.
3. Boothby TC, Tenlen JR, Smith FW, Wang JR, Patanella KA, Osborne Nishimura E, et al. Evidence for extensive horizontal gene transfer from the draft genome of a tardigrade. Proceedings of the National Academy of Sciences 2015;112(52):15976–15981. Doi:10.1073/pnas.1510461112.
4. Koutsovoulos G, Kumar S, Laetsch DR, Stevens L, Daub J, Conlon C, et al. No evidence for extensive horizontal gene transfer in the genome of the tardigrade Hypsibius dujardini. Proceedings of the National Academy of Sciences 2016;113(18):5053–5058. Doi:10.1073/pnas.1600338113.
5. Arakawa K. Examples of Extreme Survival: Tardigrade Genomics and Molecular Anhydrobiology. Annual Review of Animal Biosciences 2022;10:17–37. Doi:10.1146/annurev-animal-021419-083711.
6. Lindahl T. Instability and decay of the primary structure of DNA. nature 1993;362(6422):709–715. Doi:10.1038/362709a0.
7. Friedberg EC, Walker GC, Siede W, Wood RD. DNA repair and mutagenesis. American Society for Microbiology Press; 2005. Doi:10.1128/9781555816704.
8. Zhou BBS, Elledge SJ. The DNA damage response: putting checkpoints in perspective. Nature 2000;408(6811):433–439. Doi:10.1038/35044005.
9. Giglia-Mari G, Zotter A, Vermeulen W. DNA damage response. Cold Spring Harbor perspectives in biology 2011;3(1):a000745. Doi:10.1101/cshperspect.a000745.
10. Hashimoto T, Kunieda T. DNA protection protein, a novel mechanism of radiation tolerance: lessons from tardigrades. Life 2017;7(2):26. Doi:10.3390/life7020026.
11. Luscombe NM, Austin SE, Berman HM, Thornton JM. An overview of the structures of protein-DNA complexes. Genome biology 2000;1(1):1–37. Doi:10.1186/gb-2000-1-1-reviews001.
12. Yesudhas D, Batool M, Anwar MA, Panneerselvam S, Choi S. Proteins recognizing DNA: Structural uniqueness and versatility of DNA-binding domains in stem cell transcription factors. Genes 2017;8(8):192. Doi:10.3390/genes8080192.
13. Rakhra G, Rakhra G. Zinc finger proteins: insights into the transcriptional and post transcriptional regulation of immune response. Molecular Biology Reports 2021;48(7):5735–5743. Doi:10.1007/s11033-021-06556-x.
14. Hakoshima T. Leucine zippers. eLS 2014; Doi:10.1002/9780470015902.a0005049.pub2.
15. Malarkey CS, Churchill ME. The high mobility group box: the ultimate utility player of a cell. Trends in biochemical sciences 2012;37(12):553–562. Doi:10.1016/j.tibs.2012.09.003.
16. Prabakaran P, Siebers JG, Ahmad S, Gromiha MM, Singarayan MG, Sarai A. Classification of protein-DNA complexes based on structural descriptors. Structure 2006;14(9):1355–1367. Doi:10.1016/j.str.2006.06.018.
17. Santoro SW, Joyce GF. A general purpose RNA-cleaving DNA enzyme. Proceedings of the national academy of sciences 1997;94(9):4262–4266. Doi:10.1073/pnas.94.9.4262.
18. Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics 2008;24:637–644. Doi:10.1093/bioinformatics/btn013.
19. Gish W, States DJ. Identification of protein coding regions by database similarity search. Nature genetics 1993;3(3):266–272. Doi:10.1038/ng0393-266.
20. Buchfink RKDH B. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat Methods 2021;18:366–368. Doi:10.1038/s41592-021-01101-x.
21. Horton P OTFNHHACCNK Park KJ. WoLF PSORT: protein localization predictor. Nucleic acids research 2007;35:585–587. Doi:10.1093/nar/gkm259.
22. Almagro Armenteros JJ, Salvatore M, Winther O, Emanuelsson O, von Heijne G, Elofsson A, et al. Detecting Sequence Signals in Targeting Peptides Using Deep Learning. Life Science Alliance;2.
23. Boutet E, Lieberherr D, Tognolli M, Schneider M, Bairoch A. Uniprotkb/swiss-prot. In: Plant bioinformatics Springer; 2007.p. 89–112. Doi:10.1007/978-1-59745-535-0_4.
24. Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. HMMER web server: 2018 update. Nucleic acids research 2018;46(W1):W200–W204. Doi:10.1093/nar/gky448.
25. Peterson CL, Tamkun JW. The SWI-SNF complex: a chromatin remodeling machine? Trends in biochemical sciences 1995;20(4):143–146. Doi:10.1016/S0968-0004(00)88990-2.
26. Yusufzai T, Kadonaga JT. HARP is an ATP-driven annealing helicase. Science 2008;322(5902):748–750. Doi:10.1126/science.1161233.
27. Lee YJ, Lee JE, Choi HJ, Lim JS, Jung HJ, Baek MC, et al. E3 ubiquitin-protein ligases in rat kidney collecting duct: response to vasopressin stimulation and withdrawal. American Journal of Physiology-Renal Physiology 2011;301(4):F883–F896. Doi:10.1152/ajprenal.00117.2011.
28. Atack JR, Broughton HB, Pollack SJ. Structure and mechanism of inositol monophosphatase. Febs Letters 1995;361(1):1–7. Doi:10.1016/0014-5793(95)00063-F.
29. Hashimoto T, Horikawa DD, Saito Y, Kuwahara H, Kozuka-Hata H, Shin-i T, et al. Extremotolerant tardigrade genome and improved radiotolerance of human cultured cells by tardigrade-unique protein. Nature communications 2016;7(1):1–14. Doi:10.1038/ncomms12808.

Supplementary materials

Table 2. Location prediction of found proteins by WoLF PSORT and TargetP

protein id	WoLF PSORT predictions			TargetP predictions	
	for animals	for plants	for fungi	for nonplants	for plants
g702.t1	extr: 29, plas: 2, lyso: 1	extr: 6, vacu: 5, golg: 2, E.R.: 1	extr: 26	SP	SP
g1285.t1	extr: 25, plas: 5, mito: 1, lyso: 1	vacu: 5, chlo: 3, extr: 3, golg: 2, nucl: 1	extr: 24, golg: 2	SP	SP
g2203.t1	plas: 29, nucl: 2, golg: 1	cyto: 5, nucl: 4, E.R.: 3, vacu: 2	plas: 9, cyto: 8.5, cyto_nucl: 5, mito: 4, extr: 2, pero: 1, E.R.: 1, golg: 1	OTHER	OTHER
g3428.t1	mito: 18, cyto: 11, extr: 2, nucl: 1	chlo: 6, mito: 6, nucl: 1, cyto: 1	cyto: 12.5, cyto_nucl: 11, nucl: 8.5, mito: 5, pero: 1	OTHER	OTHER
g3679.t1	extr: 26, mito: 2, lyso: 2, plas: 1, E.R.: 1	chlo: 6, nucl: 3, mito: 2, vacu: 2, extr: 1	extr: 26	SP	SP
g4106.t1	E.R.: 14.5, E.R._golg: 9.5, extr: 7, golg: 3.5, lyso: 3, pero: 2, plas: 1, mito: 1	chlo: 6, E.R.: 3, plas: 2, pero: 2, vacu: 1	pero: 10, cyto: 5.5, E.R.: 5, cyto_nucl: 3.5, mito: 3, plas: 3	OTHER	SP
g4970.t1	plas: 32	nucl: 4, cyto: 3, E.R.: 3, vacu: 2, mito: 1, plas: 1	plas: 16, cyto: 3, nucl: 2, mito: 2, pero: 2, golg: 1, vacu: 1	OTHER	OTHER
g5237.t1	plas: 24, mito: 8	chlo: 5, golg: 4, nucl: 2, extr: 2, plas: 1	extr: 10, nucl: 4, plas: 4, mito: 3, cyto: 3, pero: 2, golg: 1	OTHER	OTHER
g5443.t1	extr: 28, nucl: 3, cyto: 1	chlo: 9, nucl: 2, cyto: 1, extr: 1, E.R.: 1	extr: 17, nucl: 5, cyto: 2, mito: 1, E.R.: 1, vacu: 1	OTHER	OTHER
g5467.t1	extr: 27, plas: 4, mito: 1	extr: 6, vacu: 5, chlo: 2, golg: 1	extr: 25	SP	SP
g5502.t1	extr: 31, lyso: 1	chlo: 9, extr: 5	extr: 27	SP	SP
g5503.t1	extr: 29, plas: 1, mito: 1, lyso: 1	chlo: 10, extr: 3, cyto: 1	extr: 19, mito: 4, golg: 2, pero: 1, E.R.: 1	SP	SP
g5510.t1	plas: 23, mito: 7, E.R.: 1, golg: 1	chlo: 5, plas: 3.5, cyto_plas: 2.5, E.R.: 2, mito: 1, pero: 1, golg: 1	plas: 21, mito: 3, E.R.: 2, pero: 1	OTHER	OTHER
g5616.t1	extr: 31, mito: 1	extr: 12, vacu: 2	extr: 25	SP	SP
g5641.t1	extr: 31, lyso: 1	extr: 8, vacu: 4, golg: 2	extr: 25	SP	SP
g5927.t1	nucl: 30.5, cyto_nucl: 16.5, cyto: 1.5	nucl: 14	nucl: 22.5, cyto_nucl: 14, cyto: 4.5	OTHER	OTHER
g7861.t1	nucl: 16, cyto_nucl: 14, cyto: 8, plas: 5, pero: 1, cysk: 1, golg: 1	nucl: 4, vacu: 4, cyto: 3, E.R.: 2, chlo: 1	nucl: 11, plas: 9, cyto: 4, mito: 2, pero: 1	OTHER	OTHER
g8100.t1	nucl: 16.5, cyto_nucl: 12.5, cyto: 7.5, plas: 5, extr: 2, E.R.: 1	nucl: 4, E.R.: 4, cyto: 2, mito: 2, vacu: 2	cyto: 11.5, cyto_nucl: 8.5, mito: 5, nucl: 4.5, pero: 2, plas: 1, E.R.: 1, golg: 1, vacu: 1	OTHER	OTHER
g8312.t1	nucl: 15.5, cyto_nucl: 15.5, cyto: 12.5, mito: 2, plas: 1, golg: 1	nucl: 7, cyto: 5, chlo: 1, vacu: 1	nucl: 15.5, cyto_nucl: 12, cyto: 5.5, pero: 5, golg: 1	OTHER	OTHER
g10513.t1	nucl: 20, cyto_nucl: 14.5, cyto: 7, extr: 3, E.R.: 1, golg: 1	nucl: 13, chlo: 1	nucl: 18.5, cyto_nucl: 15, cyto: 6.5, mito: 2	OTHER	OTHER
g10514.t1	nucl: 19, cyto_nucl: 15, cyto: 9, extr: 3, mito: 1	nucl: 8, cyto: 5, plas: 1	nucl: 20.5, cyto_nucl: 14.5, cyto: 5.5, mito: 1	OTHER	OTHER
g11320.t1	plas: 24.5, extr_plas: 16, extr: 6.5, lyso: 1	extr: 11, golg: 2, E.R.: 1	extr: 27	SP	SP
g11513.t1	cyto: 17, cyto_nucl: 12.83, cyto_mito: 9.83, nucl: 7.5, E.R.: 3, mito: 1.5, plas: 1, pero: 1, golg: 1	nucl: 4, plas: 3, cyto: 2, E.R.: 2, chlo: 1, vacu: 1, golg: 1	cyto: 13, cyto_nucl: 9.5, mito: 6, nucl: 4, pero: 2, plas: 1, E.R.: 1	OTHER	OTHER
g11806.t1	nucl: 18, cyto_nucl: 11.83, mito: 5, extr: 4, cyto: 3.5, cyto_pero: 2.67, cysk_plas: 1	nucl: 11.5, cyto_nucl: 6.5, plas: 1, extr: 1	nucl: 18, cyto_nucl: 14, cyto: 6, mito: 3	OTHER	OTHER
g11960.t1	nucl: 32	nucl: 13.5, cyto_nucl: 7.5	nucl: 25, cyto_nucl: 15.5	OTHER	OTHER
g12388.t1	extr: 25, plas: 4, mito: 2, lyso: 1	chlo: 10, vacu: 2, cyto: 1, mito: 1	extr: 20, mito: 3, E.R.: 2, cyto: 1, pero: 1	SP	SP
g12510.t1	plas: 29, cyto: 3	plas: 5, vacu: 3, E.R.: 3, cyto: 2, golg: 1	mito: 13, plas: 8, cyto: 5, vacu: 1	OTHER	OTHER
g12562.t1	extr: 30, lyso: 2	extr: 9, vacu: 3, golg: 2	extr: 25	SP	SP
g13530.t1	extr: 13, nucl: 6.5, lyso: 5, cyto_nucl: 4.5, plas: 3, E.R.: 3, cyto: 1.5	nucl: 11, chlo: 1, cyto: 1, cysk: 1	extr: 24, nucl: 1, mito: 1, cyto: 1	SP	SP
g14472.t1	nucl: 28, plas: 2, cyto: 1, cysk: 1	nucl: 14	nucl: 14, cyto_nucl: 11.5, mito: 5, cyto: 5, extr: 1, pero: 1, cysk: 1	OTHER	OTHER
g15153.t1	extr: 32	extr: 11, vacu: 2, E.R.: 1	extr: 27	SP	SP
g15484.t1	nucl: 17.5, cyto_nucl: 15.33, cyto: 12, cyto_mito: 6.83, plas: 1, golg: 1	nucl: 11, plas: 1, vacu: 1, golg: 1	nucl: 16.5, cyto_nucl: 12.5, cyto: 7.5, pero: 3	OTHER	OTHER
g16318.t1	nucl: 20.5, cyto_nucl: 13, extr: 5, cyto: 4.5, E.R.: 1, golg: 1	nucl: 13, chlo: 1	nucl: 17, cyto_nucl: 14, cyto: 7, mito: 1, extr: 1, golg: 1	OTHER	OTHER
g16368.t1	nucl: 20.5, cyto_nucl: 13, extr: 5, cyto: 4.5, E.R.: 1, golg: 1	nucl: 13, chlo: 1	nucl: 17.5, cyto_nucl: 13.5, cyto: 6.5, mito: 1, extr: 1, golg: 1	OTHER	OTHER

Table 3. Probability of prediction by WoLF PSORT and TargetP

protein id	Nuclear localization probability (WoLF PSORT predictions), %	Secretory pathway signal peptide
g702.t1	0	Yes
g1285.t1	0	Yes
g2203.t1	6.25	No
g3428.t1	3.13	No
g3679.t1	0	Yes
g4106.t1	0	No
g4970.t1	0	No
g5237.t1	0	No
g5443.t1	9.38	No
g5467.t1	0	Yes
g5502.t1	0	Yes
g5503.t1	0	Yes
g5510.t1	0	No
g5616.t1	0	Yes
g5641.t1	0	Yes
g5927.t1	95.31	No
g7861.t1	50.0	No
g8100.t1	51.56	No
g8312.t1	48.44	No
g10513.t1	62.5	No
g10514.t1	59.38	No
g11320.t1	0	Yes
g11513.t1	23.44	No
g11806.t1	56.25	No
g11960.t1	100	No
g12388.t1	0	Yes
g12510.t1	0	No
g12562.t1	0	Yes
g13530.t1	20.31	Yes
g14472.t1	87.5	No
g15153.t1	0	Yes
g15484.t1	54.69	No
g16318.t1	64.06	No
g16368.t1	64.06	No

Table 4. Closest homologous sequences by BLAST for only nuclear-located proteins: the threshold of 20% identity is implemented

Protein id	Subject	Subject name	Organism	% identity	evalue
g5927.t1	-				
g7861.t1	B4F769.1	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A-like protein 1	*Rattus norvegicus*	37.209	1.55e-71
g8100.t1	Q2YDR3.1	Inositol monophosphatase 3	*Danio rerio*	36.039	2.96e-46
g8312.t1	Q5KU39.1	Vacuolar protein sorting-associated protein 41homolog	*Mus musculus*	40.843	0.0
g10513.t1	-				
g10514.t1	-				
g11806.t1	-				
g11960.t1	Q8CJB9.1	E3 ubiquitin-protein ligase BRE1B	*Rattus norvegicus*	26.956	6.13e-98
g14472.t1	-				
g15484.t1	Q155U0.1	Vacuolar protein sorting-associated protein 51 homolog	*Danio rerio*	45.026	0.0
g16318.t1	A2VD00.1	Eukaryotic translation initiation factor 3 subunit A	*Xenopus laevis*	36.111	4.09e-08
g16368.t1	A4II09.1	Eukaryotic translation initiation factor 3 subunit A	*Xenopus tropicalis*	39.286	1.20e-05

Table 5. Domain findings by Pfam

Query Name	Family Id	Description	Start	End	E-value
g5927.t1	-				
g7861.t1	SNF2-rel_dom	SNF2-related domain	269	566	1.2e-28
	HARP	HepA-related protein (HARP)	173	228	2.6e-10
	ResIII	Type III restriction enzyme, res subunit	251	413	6.1e-06
g8100.t1	Inositol_P	Inositol monophosphatase family	449	788	1.9e-37
	MKLP1_Arf_bdg	Arf6-interacting domain of mitotic kinesin-like protein 1	1183	1287	5.1e-27
g8312.t1	Clathrin	Region in Clathrin and VPS	652	792	5.4e-23
g10513.t1	-				
g10514.t1	-				
g11806.t1	-				
g11960.t1	zf-C3HC4	Zinc finger, C3HC4 type (RING finger)	927	965	4.2e-05
g14472.t1	-				
g15484.t1	Vps51	Vps51/Vps67	10	96	1.3e-23
	Sec5	Exocyst complex component Sec5	3	476	3.4e-23
	Dor1	Dor1-like family	23	242	1.2e-11
	Vps54_N	Vacuolar-sorting protein 54, of GARP complex	11	198	2.4e-10
	COG2	COG (conserved oligomeric Golgi) complex component, COG2	6	137	2.5e-06
g16318.t1	-				
g16368.t1	-				