

Research article

Open Access

Tardigrade workbench: comparing stress-related proteins, sequence-similar and functional protein clusters as well as RNA elements in tardigrades

Frank Förster^{†1}, Chunguang Liang^{†1}, Alexander Shkumatov^{†2}, Daniela Beisser¹, Julia C Engelmann¹, Martina Schnölzer³, Marcus Frohme⁴, Tobias Müller¹, Ralph O Schill⁵ and Thomas Dandekar^{*1}

Address: ¹Dept of Bioinformatics, Biocenter University of Würzburg, 97074 Würzburg, Germany, ²EMBL, Hamburg Outstation, Notkestrasse 85, 22603 Hamburg, Germany, ³Functional Proteome Analysis, German Cancer Research Center, Im Neuenheimer Feld 580, 69120 Heidelberg, Germany, ⁴University of Applied Sciences, Bahnhofstraße 1, 15745 Wildau, Germany and ⁵Dept of Zoology, Institute for Biology, Universität Stuttgart, 70569 Stuttgart, Germany

Email: Frank Förster - frank.foerster@biozentrum.uni-wuerzburg.de; Chunguang Liang - liang@biozentrum.uni-wuerzburg.de; Alexander Shkumatov - ashkumat@embl-hamburg.de; Daniela Beisser - daniela.beisser@biozentrum.uni-wuerzburg.de; Julia C Engelmann - julia.engelmann@klinik.uni-regensburg.de; Martina Schnölzer - m.schnoelzer@dkfz.de; Marcus Frohme - mfrohme@tfh-wildau.de; Tobias Müller - tobias.mueller@biozentrum.uni-wuerzburg.de; Ralph O Schill - ralph.schill@bio.uni-stuttgart.de; Thomas Dandekar* - dandekar@biozentrum.uni-wuerzburg.de

* Corresponding author †Equal contributors

Published: 12 October 2009

Received: 14 April 2009

BMC Genomics 2009, **10**:469 doi:10.1186/1471-2164-10-469

Accepted: 12 October 2009

This article is available from: <http://www.biomedcentral.com/1471-2164/10/469>

© 2009 Förster et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Tardigrades represent an animal phylum with extraordinary resistance to environmental stress.

Results: To gain insights into their stress-specific adaptation potential, major clusters of related and similar proteins are identified, as well as specific functional clusters delineated comparing all tardigrades and individual species (*Milnesium tardigradum*, *Hypsibius dujardini*, *Echiniscus testudo*, *Tulinus stephaniae*, *Richtersius coronifer*) and functional elements in tardigrade mRNAs are analysed. We find that 39.3% of the total sequences clustered in 58 clusters of more than 20 proteins. Among these are ten tardigrade specific as well as a number of stress-specific protein clusters. Tardigrade-specific functional adaptations include strong protein, DNA- and redox protection, maintenance and protein recycling. Specific regulatory elements regulate tardigrade mRNA stability such as lox P DICE elements whereas 14 other RNA elements of higher eukaryotes are not found. Further features of tardigrade specific adaption are rapidly identified by sequence and/or pattern search on the web-tool tardigrade analyzer <http://waterbear.bioapps.biozentrum.uni-wuerzburg.de>. The work-bench offers nucleotide pattern analysis for promotor and regulatory element detection (tardigrade specific; nrdb) as well as rapid COG search for function assignments including species-specific repositories of all analysed data.

Conclusion: Different protein clusters and regulatory elements implicated in tardigrade stress adaptations are analysed including unpublished tardigrade sequences.

Background

Tardigrades are small metazoans resembling microscopic bears ("water-bears", 0.05 mm to 1.5 mm in size) and live in marine, freshwater and terrestrial environments, especially in lichens and mosses [1-3]. They are a phylum of multi-cellular animals capable of reversible suspension of their metabolism and entering a state of cryptobiosis [4,5]. A dehydrated tardigrade, known as anhydrobiotic tun-stage [6,7], can survive for years without water. Moreover, the tun is resistant to extreme pressures and temperatures (low/high), as well as radiation and vacuum [8-13].

Well known species include *Hypsibius dujardini* which is an obligatory parthenogenetic species [14]. The tardigrade *H. dujardini* can be cultured continuously for decades and can be cryopreserved. It has a compact genome, a little smaller than that of *Caenorhabditis elegans* or *Drosophila melanogaster*, and the rate of protein evolution in *H. dujardini* is similar to that of other metazoan taxa [15]. *H. dujardini* has a short generation time, 13-14 days at room temperature. Embryos of *H. dujardini* have a stereotyped cleavage pattern with asymmetric cell divisions, nuclear migrations, and cell migrations occurring in reproducible patterns [15]. Molecular data are sparse but include the purinergic receptor occurring in *H. dujardini* [16].

Milnesium tardigradum is an abundant and ubiquitous terrestrial tardigrade species in Europe and possibly worldwide [17]. It has unique anatomy and motion characteristics compared to other water bears. Most water bears prefer vegetarian food, *M. tardigradum* is more carnivorous, feeding on rotifers and nematodes. The animals are really tough and long-living, one of the reasons why *M. tardigradum* is one of the best-studied species so far.

Questions of general interest are: How related are tardigrade proteins to each other? Which protein families provide tardigrade-specific adaptations? Which regulatory elements influence the mRNA stability? Starting from all published tardigrade sequences as well as 607 unpublished new sequences from *Milnesium tardigradum*, we analyse tardigrade specific clusters of related proteins, functional protein clusters and conserved regulatory elements in mRNA mainly involved in mRNA stability. The different clusters and identified motifs are analysed and discussed, all data are also available as a first anchor to study specific adaptations of tardigrades in more detail (Tardigrade workbench). Furthermore, the tardigrade analyzer, a sequence server to analyse individual tardigrade specific sequences, is made available. It will be regularly updated to include new tardigrade sequences. It has a number of new features for tardigrade analysis not available from standard servers such as the NIH Entrez system [18]: several new species-specific searches (*Echiniscus tes-*

tudo, *Tulinus stephaniae*), additional new sequence information (*M. tardigradum*) and pattern-searches for nucleotide sequences (including pattern search on non-redundant protein database, NRDB). An easy search for clusters of orthologous groups (COG, [19]) different from the COGnitor tool [20] allowing tardigrade specific COG and eukaryotic COG (KOG) searches is also available.

Furthermore, a batch mode allows a rapid analysis of up to 100 sequences simultaneously when uploaded in a file in FASTA format (for tardigrade species or NRDB).

Two fifths of the tardigrade sequences cluster in longer protein families, and we hypothesise for a number of these that they are implicated in the unique stress adaptation potential of tardigrades. We find also ten tardigrade specific clusters. The unique tardigrade adaptations are furthermore indicated by a number of functional COGs and KOGs identified here, showing a particular emphasis on the protection of proteins and DNA. RNA read out is specifically regulated by several motifs for mRNA stability clearly overrepresented in tardigrades.

Results and Discussion

We analysed all publicly available tardigrade sequences (status 9th of April 2009) as well as 607 unpublished *M. tardigradum* sequences from our ongoing transcriptome analysis.

Major tardigrade protein clusters of related sequence-similar proteins

All available tardigrade sequences were clustered by the CLANS algorithm [21]. Interestingly, 39.3 % of the predicted proteins (mainly EST-based predictions) cluster in just 58 major families, each with at least 20 sequences [see additional file 1: Table S1]. These include 4,242 EST sequences from a total of 10,787.

Using these clusters, a number of tardigrade-specific adaptations become apparent (Table 1 [and additional file 1: Table S1]): the clusters include elongation factors (cluster 12), ribosomal RNAs and proteins (cluster 1, 4, 32 and 56) which are part of the transcriptional or translational machinery. Cluster 5 (chitinase binding domain [22]) could provide membrane and structural reorganization or immune protection (e.g. fungi) according to homologous protein sequences characterized in other organisms. Other clusters show protein families related to the tardigrade stress adaptation potential, e.g. ubiquitin-related proteins (cluster 14; maybe stress-induced protein degradation) and cytochrome oxidase-related proteins (cluster 2, suggested to be involved in respiratory chain).

Moreover, proteins responsible for protein degradation (cluster 15) were found as well as proteins regulating

Table 1: CLANS clusters of sequence similar proteins in published tardigrade sequences¹

Number/color	Cluster description	Sequences/percentage
2	Cytochrome c oxidase like (subunit I, EC 1.9.3.1)	425 (3.94%)
3	Uncharacterized protein U88/Glycosyltransferase 8 family	302 (2.80%)
5	Proteins containing a Chitin binding domain	191 (1.77%)
6	Proteins containing an IBR/Neuroparsin/DUF1096 domain	189 (1.75%)
7	Fatty-acid binding protein (FABP) family	127 (1.18%)
8	TSP ² (remote homology to Sericin I)	126 (1.17%)
9	Proteins containing a RNA polymerase Rpb3/Rpb11 dimerisation domain	92 (0.85%)
10	Metallothionein superfamily (Type 15 family./Thioredoxin like)	84 (0.78%)
12	GTP-binding elongation factor family. EF-Tu/EF-1A sub- family	79 (0.72%)
13	GST superfamily. Sigma family	78 (0.70%)
14	Ubiquitin family	75 (0.69%)
15	Cathepsin family (EC 3.4.22.-)	74 (0.67%)
16	Carboxypeptidase A inhibitor like	72 (0.64%)
17	Trichohyalin/Translation initiation factor like	69 (0.60%)
18	TSP ²	65 (0.57%)
19	TSP ²	61 (0.56%)
20	RNA/DNA-binding proteins	60 (0.55%)
	...	
23	Small Heat Shock Protein (HSP20) family	53 (0.47%)
24	Diapause-specific proteins	51 (0.44%)
	...	
38	LEA type 1 family proteins	31 (0.28%)
	...	

¹ Shown are the number of proteins found for the specified cluster, their percentages and the corresponding cluster number in Figure 1. The full Table with all clusters and their color code matching to Figure 1 is given in [additional file 1: Table S1]. Clusters 1, 4 and 11 contain rRNA and are given in [additional file 1].

² Tardigrade specific protein cluster

peptidases (cluster 16). Cluster 23 consists of 53 heat shock proteins which are involved in many stress response reactions [23]. Few diapause specific proteins (cluster 24) are known from other animals. Diapause is a reversible state of developmental suspension. It is observed in diverse taxa, from plants to animals, including marsupials and some other mammals [24] as well as insects (associated molecular function varies but involves calcium channel inhibition [25]) and should here support the tun formation or regulate other (e.g. developmental) metabolic inactive states. Furthermore, proteins involved in storage or transportation of fatty acids also seem to be important (cluster 31, [26]). Late embryogenesis abundant (LEA) protein expression seems to be linked to desiccation stress and the acquisition of desiccation tolerance in organisms [27] e.g. nematodes [28,29] and rotifers [30]. Thirty-one LEA type 1 family proteins were found in cluster 38.

LEA proteins are wide-spread among plants and synthesized in response to certain stresses [31,32]. The LEA type 1 family is well known in higher plants (rice, maize, carrots) to be synthesized during late embryogenesis and in ABA stress response. It includes desiccation-related protein PCC3-06 of *Cratersostigma plantagineum*. LEA type 1 family occurs in bacteria (e.g. *Haemophilus influenzae*, *Dei-*

nococcus radiodurans), but is atypical for animals. However, this is an animal example where LEA family type 1 is well represented and forms a full cluster.

Moreover, ten clusters (8, 18, 19, 30, 33, 35, 37, 42, 51, 55) consist of proteins which seem to be specific for tardigrades. These show no significant homology to known proteins.

Functional clusters of stress-specific adaptations present in tardigrades

To gain a systematic overview of involved tardigrade functions, all available tardigrade sequences were classified species-specific according to COG functional category [19,20] as well as according to COG number and molecular function encoded. Note that in this section "protein" implies one type of protein. A COG or KOG comprises often several sequences from different tardigrades. Prokaryotic (COG) and eukaryotic (KOG) gene clusters were compared (Table 2; details on the WEB <http://waterbear.bioapps.biozentrum.uni-wuerzburg.de/>). Again, several tardigrade-specific adaptations stand out, e.g. highly represented COGs regulate translation elongation factor and sulfate adenylate transferase and a strong ubiquitin system. There are many cysteine proteases (21 proteins). For redox protection there are 14 thioredoxin-domain

Table 2: Highly represented protein functions in Tardigrades (COGs and KOGs).

<i>Information from COG clusters¹:</i>	
Information storage and processing	
75 Translation elongation factor EF-I (COG5256)	
64 GTPases - translation elongation (COG0050)	
58 Peptide chain release factor RF-3 (COG4108)	
Cellular processes and signaling	
31 Ubiquitin (COG5272)	
25 Membrane GTPase LepA (COG0481)	
21 Cysteine protease (COG4870)	
Metabolism	
75 Heme/copper-type cytochrome/quinol oxidases (COG0843)	
67 GTPases - Sulfate adenylate transferase (COG2895)	
Poorly characterized	
11 Dehydrogenases with different specificities (COG1028)	
11 Uncharacterized homolog of Btl101 (COG0401)	
<i>Information from KOG clusters¹:</i>	
Information storage and processing	
77 Translation elongation factor EF-I (KOG0052)	
71 Polypeptide release factor 3 (KOG0459)	
70 Elongation factor I alpha (KOG0458)	
53 Mitochondrial translation elongation factor Tu (KOG0460)	
Cellular processes and signaling	
52 Glutathione S-transferase (KOG1695)	
46 Cysteine proteinase Cathepsin L (KOG1543)	
34 Apolipoprotein D/Lipocalin (KOG4824)	
31 Cysteine proteinase Cathepsin F (KOG1542)	
Metabolism	
78 Cytochrome c oxidase (KOG4769)	
74 Fatty acid-binding protein FABP (KOG4015)	
Poorly characterized	
31 Ubiquitin and ubiquitin-like proteins (KOG0001)	
16 GTPase Rab18, small G protein superfamily (KOG0080)	
15 Ras-related GTPase (KOG0394)	
15 GTPase Rab21, small G protein superfamily (KOG0088)	

¹ Detailed data and all COG/KOG numbers are given on the WEB page http://waterbear.bioapps.biozentrum.uni-wuerzburg.de/cgi-bin/cog_stat.pl. Summarized here are the functions of those clusters of orthologous groups (COGs) occurring particularly often or suggesting tardigrade specific adaptations.

containing proteins and 75 Heme/copper-type cytochrome/quinol-like proteins as well as ubiquinone oxidoreductase subunits (15 proteins). There are ten proteins involved in seleno-cysteine specific translation [33,34]. In eukaryotes, selenoproteins show a mosaic occurrence, with some organisms, such as vertebrates and algae, but notably also tardigrades, having dozens of these proteins, while other organisms, such as higher plants and fungi, having lost all selenoproteins during evolution [34]. Membrane GTPases (25 proteins) are often of Lep A (leader peptidase [35]) type in tardigrades. In general, members of the GTPase superfamily regulate membrane signaling pathways in all cells. However, LepA, as well as NodO, are prokaryotic-type GTPases very similar to protein synthesis elongation factors but apparently have

membrane-related functions [35]. It is interesting to observe this prokaryotic-type GTPase in tardigrades. We suggest that it will have similar function as known in other organisms and thus ensure protein translation (elongation factor) coupled to membrane integrity and possibly cytoskeletal rearrangement which would again boost the tardigrade resistance to stress.

The KOGs show similar highly represented families and adaptations. Abberant proteins are rapidly recognized by ubiquitination-like proteins (220 proteins) and ubiquitin-ligase related enzymes (71 proteins) as well as proteasome regulatory subunits (85 proteins). For protein protection and refolding disulfide isomerases (26 proteins) and cyclophilin type peptidyl-prolyl cis-trans isomerases (43 proteins; KOG 0879-0885) are available. Connected to redox protection are also thirty AAA+type ATPases and three peroxisome assembly factor 2 containing proteins (KOG0736). This broad effort in protein protection is further supported by molecular chaperones (HSP70, mortalins and other; total of 50 proteins) and chaperonin complex components (32 proteins; KOG0356-0364). There are six superoxide dismutases and six copper chaperons for thioredoxins (37 proteins), glutaredoxin-like proteins (nine) and ten thiodisulfide isomerases as well as 52 glutathione-S-transferases. We found 22 hits to helicases. Tardigrade DNA protection is represented by 52 proteins of the molecular chaperone DNA J family: proteins of the DNA J family are classified into 3 types according to their structural domain decomposition. Type I J proteins compose of the J domain, a gly-rich region connecting the J domain and a zinc finger domain, and possibly a C-terminal domain. Type II lacks the Zn-finger domain and type III only contains the J domain [36,37]. The latter two are referred to as DnaJ-like proteins. Analysis of the domains present in tardigrade proteins by SMART [38] and Pfam [39] searches reveals only the J domain and in some cases a transmembrane region, identifying them as type III DnaJ-like proteins. For further information on these COGs/KOGs see Table 3.

Moreover, undesired proteins can be rapidly degraded by cathepsin F-like proteins (31 proteins) or L-like proteins (46 proteins). There are several calcium-dependent protein kinases (25 proteins; KOG0032-0034) and actin-bundling proteins. According to this observation calcium signaling should be implicated in adaptive rearrangement of the cytoskeleton during tardigrade rehydration. The cytoskeleton is a key element in the organisation of eukaryotic cells. It has been described in the literature that the properties of actin are modulated by small heat-shock proteins including a direct actin-small heat-shock protein interaction to inhibit actin polymerization to protect the cytoskeleton [40,41] (compare with the CLANS cluster 24 (Diapause proteins) found in the above analysis).

Table 3: Identified DnaJ-family COGs/KOGs in Tardigrades and *Milnesium tardigradum*¹.

KOG/COG number	COG distribution	COG name	present in	
			Tardigrades	<i>M. tardigradum</i> *
COG0484		DnaJ-class molecular chaperone with C-terminal Zn finger domain	5	
COG2214		DnaJ-class molecular chaperone	8	
KOG0550	A-DH-P-	Molecular chaperone (DnaJ superfamily)	3	2
KOG0691	ACDHYP-	Molecular chaperone (DnaJ su perfamily)	7	2
KOG0712	ACDHYPE	Molecular chaperone (DnaJ su perfamily)	8	2
KOG0713	ACDH---	Molecular chaperone (DnaJ su perfamily)	5	1

¹Shown are the number of proteins found for the specified COG/KOG number, the KOG distribution of the KOG in different eukaryotic species (see abbreviations) and the COG/KOG annotation.

Abbreviations: A *Arabidopsis thaliana*, C *Caenorhabditis elegans*, D *Drosophila melanogaster*, H *Homo sapiens*, Y *Saccharomyces cerevisiae*, P *Schizosaccharomyces pombe*, E *Encephalitozoon cuniculi*. *These include specific unpublished data from ongoing work on *M. tardigradum*

Translation in tardigrades includes polypeptide release factors (71 proteins) and proteins for translation elongation (77 proteins). There are about 80 GTP-binding ADP-ribosylation factors. The secretion system and Rab/Ras GTPases are fully represented (183 proteins). Seventeen tubulin anchor proteins show that the cytoskeleton is well maintained. Finally, we find 14 TNF-associated factors and 34 apolipoprotein D/lipocalin proteins.

Typical motifs in tardigrade mRNAs

The regulatory motif search showed a number of known regulatory RNA elements involved in tardigrade mRNA regulation (Table 4 for *H. dujardini* and *M. tardigradum*). Certainly it can not be formally ruled out that some of these elements work in a tardigrade modified way. Similarly, there are probably further patterns which are tardigrade specific, but not detected with the UTRscan software [42] applied for analysis.

Table 4: Regulatory elements in *Hypsibius dujardini*¹ and *Milnesium tardigradum*² mRNA sequences.

Motiv	<i>Hypsibius dujardini</i>	<i>Milnesium tardigradum</i>
15-LOX-DICE	1528 (1269) ³	46 (45) ³
ADH DRE	60 (58) ³	1 (1) ³
BRE	1 (1) ³	--
Brd-Box	152 (149) ³	28 (22) ³
CPE	21 (21) ³	15 (15) ³
Elastin G3A	1 (1) ³	--
GLUT1	1 (1) ³	--
GY-Box	406 (372) ³	21 (21) ³
IRE	1 (1) ³	--
IRES	1353 (1353) ³	90 (90) ³
K-Box	469 (447) ³	35 (33) ³
SECIS-1	1 (1) ³	--
SECIS-2	6 (6) ³	--
TGE	5 (5) ³	1 (1) ³
TOP	50 (50) ³	1 (1) ³

¹ We considered 5,378 ESTs in *H. dujardini*. ² We considered 607 ESTs in *M. tardigradum*. ³ The number of hits is followed by the number of mRNAs with this hit in brackets to indicate multiple hits.

The RNA elements found include the lox-P DICE element [43] in *H. dujardini* as top hit with as many as 1,269 ESTs (23.6% of all *H. dujardini* EST sequences). The cytidine-rich 15-lipoxygenase differentiation control element (15-LOX DICE, [44]) binds KH domain proteins of the type hnRNP E and K (stronger in multiple copies), mediating mRNA stabilization and translational control [43].

Furthermore, a high number of mRNAs contains K-Boxes (cUGUGAUa, [45]) and brd Boxes (AGCUUUA, [46]). All these elements are involved in mRNA storage and mRNA stability. These two elements are potential targets for miRNAs as shown in *Drosophila melanogaster* [47].

However, in the two tardigrade species compared, only 16 of 30 well known RNA elements are found, suggesting a clear bias in tardigrade mRNA regulation. For example, the widely used AU rich elements in higher organisms [42] such as vertebrates are absent in tardigrades [see additional file 1].

Regulatory elements in tardigrade mRNA are probably important for their adaptation, in particular to support transformation to tun stage and back to active stage again. The list of RNA elements found can be compared for instance to our data on regulatory elements in human anucleate platelets [48] where mRNAs have to be stock-piled for the whole life of the platelet. Due to this comparatively long life, a long mRNA untranslated region is important in these cells. The same should apply to tardigrade mRNAs, since their average UTR is predicted to be long. A different stock-piling scenario occurs in unfertilized eggs, but due to developmental constraints, here localization signals are often in addition important for developmental gradients. We tested for these in tardigrades but did not find a high representation of localization motifs.

Web-tool tardigrade analyzer

We created a convenient platform to allow rapid sequence comparisons of new protein sequences, in particular from new sequencing efforts in tardigrades, to our database by applying rapid heuristic local alignment using BLAST [49] and allowing to search in selected species.

A batch mode allows the analysis of up to 100 sequences simultaneously when uploaded in a file in FASTA format. Output data are displayed according to an enhanced BLAST output format with graphical illustrations. Low expected E-values result for searches using the option of our tardigrade specific databases: a more specific smaller database reduces the probability of false positives. As an alternative for general sequence analysis, a search against the non-redundant database of GenBank can be performed. This takes more computational power and yields higher E-values, however, it identifies functions for most sequences. An additional useful feature is to scan all available data for peptide motifs or PROSITE signatures using a "pattern" module [additional file 1: Fig. S1] or assign potential functions by COGs [19]. The first is helpful to recognize tardigrade proteins in cases where the tardigrade sequence has diverged far, and only critical residues for function are still conserved as motif signatures. It can also be applied to search for regulatory RNA motifs such as polyadenylation sites (e.g. AAUAAA or AAUUA) or recognize promotor modules such as the glucocorticoid receptor element (GRE; palindromic pattern: AGAA CAnnnTGTCT). For this purpose, both, the tardigrade sequences and the non redundant database can be searched (e.g. to look for stress-specific regulatory RNA elements; [additional file 1, Fig. S2]).

Interestingly, this nucleotide (RNA or DNA) specific option is not available on some common servers, e.g. the PHI-BLAST [50] server at NIH. Further options include a user-defined database [additional file 1: Fig. S3] and interactively animated stress clusters (Figure 1).

The tool <http://waterbear.bioapps.biozentrum.uni-wuerzburg.de/> allows rapid searches for tardigrade specific sequences, e.g. molecular adaptations against stress [see additional file 1 for screenshots and a tutorial]. For instance, a search for trehalase sequences shows no trehalase mRNA in the *H. dujardini* sequences. In contrast, there are several heat shock proteins in tardigrades, an example is HSP90 proteins (identified by sequence similarity as well as by a pattern hit based approach using the PROSITE entry PS00298 with the signature Y-x- [NQHD] - [KHR] - [DE] - [IVA] -F- [LM] -R- [ED]; Table 5). Specific COGs are also rapidly assigned for any desired sequence. This includes the option to map the query sequence of interest to any of the known tardigrade specific COGs. Furthermore, nucleotide patterns such as

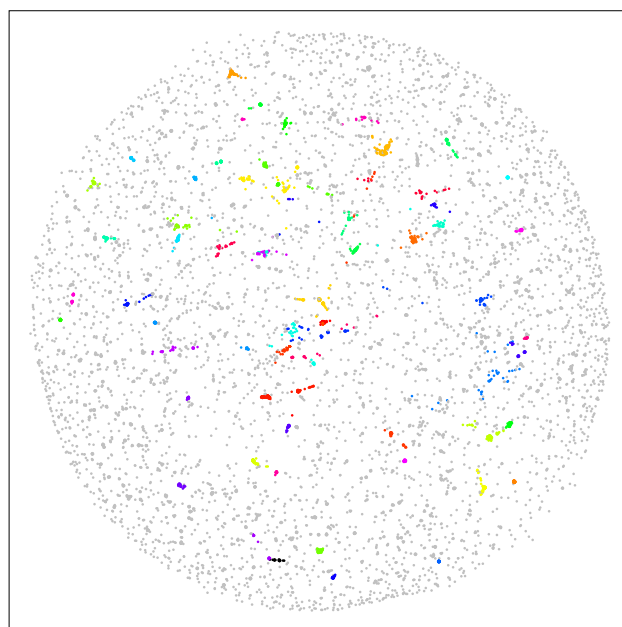


Figure 1
Functional clusters by CLANS of sequence related proteins in tardigrades. All available [see additional file 1: Figure S5] tardigrade protein sequences were clustered in a 3D sphere according to their sequence distance and were projected to the paper plane. Individual protein functions are colored [for color code see additional file 1: Table S1] and all listed in Table 1. Functional clusters appear as patches of an individual color. Color code and clusters can be interactively examined at the Tardigrade workbench <http://waterbear.bioapps.biozentrum.uni-wuerzburg.de> and are given in [additional file 1 Table S1]. figure1.pdf

mRNA polyadenylation sites are rapidly identified e.g. in *H. dujardini* mRNAs [additional file 1: Fig. S4]. Similarly, other mRNA 3'UTR elements can be identified, e.g. AU rich sequences mediating mRNA instability or regulatory K-boxes (motif cUGUGAUa, [45]) in tardigrades.

Implications

Tardigrades show a surprising large amount of related sequences. Certainly, one has to correct for a few genes sequenced from many lineages for phylogenetic studies in tardigrades (cytochrome c, rRNA etc.) However, despite this, a number of tardigrade-specific clusters still remain. Furthermore, Table 1 shows that most of the annotated clusters are stress-related.

Looking at specific protein functions, both COG and KOG proteins show that tardigrades spend an extraordinary effort in protein protection, turnover and recycling as well as redox protection. Some other specific adaptations become apparent also from Table 2, but the complete extent of these adaptations is unclear given the limited

Table 5: HSP90 proteins identified in *Hypsibius dujardini* using the Tardigrade analyzer¹.

Hit	Predicted function/name (Tardigrade analyzer)	Pattern matched	Start position	End position
gi:37213462	hsp90 ²	YSNKEIFLRE	68	77
gi:37213713	hsp90 ²	YSNKEIFLRE	70	79

¹These are hits using the pattern hit option and the heat shock protein PROSITE entry P500298 for pattern generation and recognition. The pattern has the signature of Y-x- [NQHD]- [KHR]- [DE]- [IVA]-F- [LM]-R- [ED]. ²Predicted similarity to Q7PT10 (HSP83 ANOGA) from Swissprot

sampling of available tardigrade sequences. Furthermore, protection of DNA is critical as it has been shown that tardigrade tuns accumulate DNA damage which first has to be repaired before resurrection occurs [51,52]. Taking this into consideration, DNA J proteins were investigated in more detail since proteins of this family are well represented in tardigrades, including several COGs and KOGs. Several data underline the extremely high resistance of tardigrades to temperature, pressure and radiation as well as a high repair potential regarding DNA [11,51]. Thus, we suggest that the high repair potential is also mediated by this well represented protein family. Phylogenetic analysis (Table 3) shows that these proteins are represented by several KOGs as well as the classic COGs in tardigrades. In particular, the first three KOG families are also used in *M. tardigradum*, where extreme stress tolerance requires strong repair mechanisms [17]. Furthermore, all these tardigrade proteins in Table 3 are small, having neither zinc-finger domains nor low complexity regions, but instead consisting of single DNA J domains which would always place them in type I (subfamily A) of DNA-J like proteins. This suggests that the direct interaction with DNA-J like proteins is the key molecular function.

Finally, we could show that there are 16 regulatory elements used in tardigrade mRNA, while a number of other elements known from higher eukaryotic organisms and vertebrates is not used. It is interesting to note that the elements often used in tardigrades are all involved in regulation of mRNA stability. Thus, they may be implicated in stage switching, as presumably in the initial phases of the tun awakening or tun formation, new supply of mRNA is turned off and instead regulation of synthesized mRNA becomes important.

In addition, and for further research we supply the web tool tardigrade analyzer. There are a number of alternative tools available, e.g. from NCBI <http://www.ncbi.nlm.nih.gov/>. However, we offer some species-specific searches not available from these sources as well as RNA and promotor pattern search (not only for tardigrades but also for NRDB; not available from NIH). Furthermore, there are functional COG prediction as well as new, unpublished tardigrade sequences from *M. tardigradum*,

all above reported data including the reported sequences and detailed functional clusterings as well as regular server updates. A better understanding of the survival mechanisms in these organisms will lead to the development of new methods in several areas of biotechnology. For example, preservation of biological materials *in situ*, macromolecules and cells from non-adapted organisms [53]. This is, of course, only a first and very general overview on potential tardigrade specific adaptations, more species-specific data will be considered as more information becomes available.

Conclusion

Tardigrade genomes invest in stress-specific adaptations, this includes major sequence related protein clusters, functional clusters for stress as well as specific regulatory elements in mRNA. For further tardigrade genome analysis we offer the tardigrade workbench as a flexible tool for rapid and efficient analysis of sequence similarity, protein function and clusters, COG membership and regulatory elements.

Methods

Tardigrade-sequences

The cosmopolitan eutardigrade species *M. tardigradum* Doyère 1849 (Apochea, Milnesidae) was cultured. Tardigrades were kept and reared on petri dishes (diameter: 9.4 cm) filled with a small layer of agarose (3 %) (peqGOLD Universal Agarose, peqLAB, Erlangen, Germany) and covered with spring water (Volvic™ water, Danone Waters Deutschland, Wiesbaden, Germany) at 20 ± 2°C and a light/dark cycle of 12 h. Rotifers *Philodina citrina* and nematodes *Panagrellus* sp. were provided as food source, juvenile tardigrades were also fed with green algae *Chlorogonium elongatum*. For all experiments adult animals in good physical condition were taken directly from the culture and starved for three days to avoid preparation of additional RNA originating from not completely digested food in the intestinal system. For an overview of RNAs present both in active and tun stage we used a mixture of the same number of animals.

Total RNA extraction was performed using the QIAGEN RNeasy®Mini kit (Qiagen, Hilden, Germany). The cDNA

synthesis was reversed transcribed using 1 µg total RNA by the Creator™ SMART™ cDNA Library Construction Kit (Clontech-Takara Bio Europe, France). The resulting cDNA was amplified following the manufacturers protocol and cloned into pDNR-Lib cloning vector. The resulting plasmids were used to transform *Escherichia coli* by electroporation. Sequencing of the cDNA-library was done by ABI 3730XL capillary sequencer (GATC Biotech AG, Konstanz, Germany). All obtained EST sequences were deposited with Genbank including dbEST databank.

Nucleotide sequences from other tardigrades were collected from Genbank. For *H. dujardini*, the best represented species, we composed 5,235 ESTs. We stored *H. dujardini* as well as all published sequences of other tardigrade species (e.g. *T. stephaniae*, *E. testudo*, *M. tardigradum*, *R. coronifer*) in a database (10,787 sequences including translated sequences, details in [additional file 1], status on April, 2009).

CLANS clustering

For a systematic overview on tardigrade specific adaptations we first clustered all published tardigrade nucleotide sequences into functional clusters (Figure 1) using the Cluster analysis of sequences (CLANS) algorithm [21]. All sequences were clustered in 3D space using 0.001 as an E-value cut-off for TBLASTX all-against-all searches. [additional file 1: Fig. S4].

Identification of regulatory elements

For this the ESTs of *H. dujardini* and *M. tardigradum* were systematically screened using the software UTRscan [42]. This software screens 30 regulatory elements for RNA regulation with a focus on 3' UTR elements and stability of mRNA. The default settings for batch mode were used and all reported elements were collected.

COG clustering and identification

In order to acquire a systematic overview of the functionalities, we used the latest version of COG/KOG databases <ftp://ftp.ncbi.nih.gov/pub/COG> and the BLAST hits from both nucleotide search and protein search were clustered according to their COG ID. Searches were carried out in parallel on all the tardigrade species including *M. tardigradum*, *H. dujardini*, *E. testudo*, *T. stephaniae* and *R. coronifer*. The results are summarized in a table shown in the tardigrade analyzer, the background color from cold to warm (blue to red) indicates the cluster size, which enables an easy comparison. Moreover, users are allowed to click the COG ID and the hit number. The server then reports the corresponding sequence ID, description, conservation and the homologous entries recorded in the database. The server with its data is automatically updated bi-monthly according to the latest tardigrade databases.

Tardigrade workbench

The tardigrade workbench is implemented in Perl using the Bioperl modules [54]. NCBI BLAST program of 2.2.17 is involved in the software package. A database of PostgreSQL 8.1.9 is applied to manage the tardigrade entries so as to accelerate the searching queried by investigators. The COG cluster information is automatically updated each week and warehoused on the server. In addition, the run of tardigrade workbench requires an Apache server, a linux system of at least 2 GB memory is highly recommended.

Authors' contributions

FF did tardigrade protein data analysis including CLANS clustering and RNA motif analysis. CL established the current version of the tardigrade workbench including programming new routines, data management and nucleotide motif analysis. AS did the initial setup of the server, of the virtual ribosome and the CLANS clustering. DB, JE, MS and MF participated in tardigrade data analysis. TM gave expert advice and input on statistics, RS gave expert advice on tardigrade physiology and zoology. TD led and guided the study including analysis of data and program, supervision, and manuscript writing. All authors participated in the writing of the manuscript and approved the final version.

Additional material

Additional file 1

Additional Tables and Figures. The file contains seven additional figures and two additional tables. One of these tables summarizes annotation and different identifiers for 607 new EST sequences from Milnesium tardigradum.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-469-S1.PDF>]

Acknowledgements

Stylistic corrections by Rosemary Wilson from EMBL Hamburg are gratefully acknowledged. Support by the state Bavaria, DFG (TR34A5) and the German Federal Ministry of Education and Research, BMBF (03I3838A, 03I3838B, 03I3838C, 03I3838D, 03I3838E) is gratefully acknowledged.

References

1. Marcus E, Dahl F: *Spinnentiere oder Arachnoidea IV. Bärtierchen (Tardigrada)* Urban & Fischer Bei Elsevier; 1928.
2. Marcus E: *Zur Ökologie und Physiologie der Tardigraden.* Zool Jahrb Abt Phys 1928, **44**:323-370.
3. Nelson DR: **Current Status of the Tardigrada: Evolution and Ecology.** Integr Comp Biol 2002, **42**:652-659.
4. Keilin D: **The Leeuwenhoek Lecture: The problem of anabiosis or latent life: History and current concept.** Proc R Soc Lond B Biol Sci 1959, **150**:149-191.
5. Ramazzotti G, Maucci W: **The Phylum Tardigrada.** Memorie dell'Istituto Italiano di Idrobiologia, Pallanza 1983, **41**:309-314.

6. Baumann H: **Die Anabiose der Tardigraden.** *Zool Jahrb* 1922, **45**:501-556.
7. Baumann H: **Bemerkungen zur Anabiose von Tardigraden.** *Zool Anz* 1927, **72**:175-179.
8. Horikawa DD, Sakashita T, Katagiri C, Watanabe M, Kikawada T, Nakahara Y, Hamada N, Wada S, Funayama T, Higashi S, Kobayashi Y, Okuda T, Kuwabara M: **Radiation tolerance in the tardigrade *Milnesium tardigradum*.** *Int J Radiat Biol* 2006, **82**:843-848.
9. Hengherr S, Worland MR, Reuner A, Brümmer F, Schill RO: **Freeze tolerance, supercooling points and ice formation: comparative studies on the subzero temperature survival of limno-terrestrial tardigrades.** *J Exp Biol* 2009, **212**:802-807.
10. Hengherr S, Worland MR, Reuner A, Brümmer F, Schill RO: **High-Temperature Tolerance in Anhydrobiotic Tardigrades Is Limited by Glass Transition.** *Physiol Biochem Zool* 2009, **82**(6):749-755.
11. Jönsson KI, Rabbow E, Schill RO, Harms-Ringdahl M, Rettberg P: **Tardigrades survive exposure to space in low Earth orbit.** *Curr Biol* 2008, **18**:R729-R731.
12. Jönsson KI, Schill RO: **Induction of Hsp70 by desiccation, ionising radiation and heat-shock in the eutardigrade *Richtersius coronifer*.** *Comp Biochem Physiol B Biochem Mol Biol* 2007, **146**:456-460.
13. Wright JC: **Cryptobiosis 300 Years on from van Leeuwenhoek: What Have We Learned about Tardigrades?** *Zoologischer Anzeiger - A Journal of Comparative Zoology* 2001, **240**:563-582.
14. Ammermann D: **The cytology of parthenogenesis in the tardigrade *Hypsibius dujardini*.** *Chromosoma* 1967, **23**(2):203-213.
15. Gabriel WN, McNuff R, Patel SK, Gregory TR, Jeck WR, Jones CD, Goldstein B: **The tardigrade *Hypsibius dujardini*, a new model for studying the evolution of development.** *Dev Biol* 2007, **312**:545-559.
16. Bavan S, Straub VA, Blaxter ML, Ennion SJ: **A P2X receptor from the tardigrade species *Hypsibius dujardini* with fast kinetics and sensitivity to zinc and copper.** *BMC Evol Biol* 2009, **9**:17.
17. Kinchin I, Dennis R: *The biology of tardigrades* Portland Press London; 1994.
18. Baxevas AD: **Searching the NCBI databases using Entrez.** *Curr Protoc Hum Genet* 2006, **Chapter 6**: Unit 6.10.
19. Tatusov RL, Galperin MY, Natale DA, Koonin EV: **The COG database: a tool for genome-scale analysis of protein functions and evolution.** *Nucleic Acids Res* 2000, **28**:33-36.
20. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder S, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA: **The COG database: an updated version includes eukaryotes.** *BMC Bioinformatics* 2003, **4**:41.
21. Frickey T, Lupas A: **CLANS: a Java application for visualizing protein families based on pairwise similarity.** *Bioinformatics* 2004, **20**:3702-3704.
22. Tjoelker LW, Gosting L, Frey S, Hunter CL, Trong HL, Steiner B, Brammer H, Gray PV: **Structural and functional definition of the human chitinase chitin-binding domain.** *J Biol Chem* 2000, **275**:514-520.
23. Qian SB, McDonough H, Boellmann F, Cyr DM, Patterson C: **CHIP-mediated stress recovery by sequential ubiquitination of substrates and Hsp70.** *Nature* 2006, **440**:551-555.
24. Chen WH, Ge X, Wang W, Yu J, Hu S: **A gene catalogue for post-diapause development of an anhydrobiotic arthropod *Artemia franciscana*.** *BMC Genomics* 2009, **10**:52.
25. Kim YJ, Nachman RJ, Aimanova K, Gill S, Adams ME: **The pheromone biosynthesis activating neuropeptide (PBAN) receptor of *Heliothis virescens*: identification, functional expression, and structure-activity relationships of ligand analogs.** *Peptides* 2008, **29**:268-275.
26. Alvarez-Ordóñez A, Fernández A, López M, Bernardo A: **Relationship between membrane fatty acid composition and heat resistance of acid and cold stressed *Salmonella* senftenberg CECT 4384.** *Food Microbiol* 2009, **26**:347-353.
27. Tunnacliffe A, Wise MJ: **The continuing conundrum of the LEA proteins.** *Naturwissenschaften* 2007, **94**:791-812.
28. Browne JA, Dolan KM, Tyson T, Goyal K, Tunnacliffe A, Burnell AM: **Dehydration-specific induction of hydrophilic protein genes in the anhydrobiotic nematode *Aphelenchus avenae*.** *Eukaryot Cell* 2004, **3**:966-975.
29. Goyal K, Tisi L, Basran A, Browne J, Burnell A, Zurdo J, Tunnacliffe A: **Transition from natively unfolded to folded state induced by desiccation in an anhydrobiotic nematode protein.** *J Biol Chem* 2003, **278**:12977-12984.
30. Tunnacliffe A, Lapinski J, McGee B: **A putative LEA protein, but no trehalose, is present in anhydrobiotic bdelloid rotifers.** *Hydrobiologia* 2005, **546**:315-321.
31. Kobayashi F, Maeta E, Terashima A, Takumi S: **Positive role of a wheat HvAB15 ortholog in abiotic stress response of seedlings.** *Physiol Plant* 2008, **134**:74-86.
32. Hong-Bo S, Zong-Suo L, Ming-An S: **LEA proteins in higher plants: structure, function, gene expression and regulation.** *Colloids Surf B Biointerfaces* 2005, **45**:131-135.
33. Fagegaltier D, Lescure A, Walczak R, Carbon P, Krol A: **Structural analysis of new local features in SECIS RNA hairpins.** *Nucleic Acids Res* 2000, **28**:2679-2689.
34. Lobanov AV, Hatfield DL, Gladyshev VN: **Eukaryotic selenoproteins and selenoproteomes.** *Biochim Biophys Acta* 2009 in press.
35. March PE: **Membrane-associated GTPases in bacteria.** *Mol Microbiol* 1992, **6**:1253-1257.
36. Walsh P, Bursac D, Law YC, Cyr D, Lithgow T: **The J-protein family: modulating protein assembly, disassembly and translocation.** *EMBO Rep* 2004, **5**:567-571.
37. Cheetham ME, Caplan AJ: **Structure, function and evolution of DnaJ: conservation and adaptation of chaperone function.** *Cell Stress Chaperones* 1998, **3**:28-36.
38. Letunic I, Doerks T, Bork P: **SMART 6: recent updates and new developments.** *Nucleic Acids Res* 2009, **37**:D229-D232.
39. Finn RD, Tate J, Mistry J, Cogill PC, Sammut SJ, Hotz HR, Ceric G, Forslund K, Eddy SR, Sonnhammer ELL, Bateman A: **The Pfam protein families database.** *Nucleic Acids Res* 2008, **36**:D281-D288.
40. Mounier N, Arrigo AP: **Actin cytoskeleton and small heat shock proteins: how do they interact?** *Cell Stress Chaperones* 2002, **7**:167-176.
41. Sun Y, MacRae TH: **Small heat shock proteins: molecular structure and chaperone function.** *Cell Mol Life Sci* 2005, **62**:2460-2476.
42. Pesole G, Liuni S: **Internet resources for the functional analysis of 5' and 3' untranslated regions of eukaryotic mRNAs.** *Trends Genet* 1999, **15**:378.
43. Ostareck-Lederer A, Ostareck DH, Hentze MW: **Cytoplasmic regulatory functions of the KH-domain proteins hnRNPs K and E1/E2.** *Trends Biochem Sci* 1998, **23**:409-411.
44. Ostareck-Lederer A, Ostareck DH, Standart N, Thiele BJ: **Translation of 15-lipoxygenase mRNA is inhibited by a protein that binds to a repeated sequence in the 3' untranslated region.** *EMBO J* 1994, **13**:1476-1481.
45. Lai EC, Burks C, Posakony JW: **The K box, a conserved 3' UTR sequence motif, negatively regulates accumulation of enhancer of split complex transcripts.** *Development* 1998, **125**:4077-4088.
46. Lai E: **Micro RNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation.** *Nat Genet* 2002, **30**:363-364.
47. Lai EC, Tam B, Rubin GM: **Pervasive regulation of *Drosophila* Notch target genes by GY-box-, Brd-box-, and K-box-class microRNAs.** *Genes Dev* 2005, **19**:1067-1080.
48. Dittich M, Birschmann I, Pfrang J, Herterich S, Smolenski A, Walter U, Dandekar T: **Analysis of SAGE data in human platelets: features of the transcriptome in an anucleate cell.** *Thromb Haemost* 2006, **95**:643-651.
49. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25**:3389-3402.
50. Zhang Z, Schäffer AA, Miller W, Madden TL, Lipman DJ, Koonin EV, Altschul SF: **Protein sequence similarity searches using patterns as seeds.** *Nucleic Acids Res* 1998, **26**:3986-3990.
51. Neumann S, Reuner A, Brümmer F, Schill RO: **DNA damage in storage cells of anhydrobiotic tardigrades.** *Comp Biochem Physiol A Mol Integr Physiol* 2009, **153**:425-429.
52. Schill R, Neumann S, Reuner A, Brümmer F, Schill RO: **Detection of DNA damage with single-cell gel electrophoresis in anhydrobiotic tardigrades.** *Comp Biochem Physiol A Mol Integr Physiol* 2008, **151**:32-32.

53. Schill RO, Mali B, Dandekar T, Schnölzer M, Reuter D, Frohme M: **Molecular mechanisms of tolerance in tardigrades: new perspectives for preservation and stabilization of biological material.** *Biotechnol Adv* 2009, **27**:348-352.
54. Stajich JE, Block D, Boulez K, Brenner SE, Chervitz SA, Dagdigian C, Fuellen G, Gilbert JGR, Korf I, Lapp H, Lehtväslaiho H, Matsalla C, Mungall CJ, Osborne BI, Pocock MR, Schattner P, Senger M, Stein LD, Stupka E, Wilkinson MD, Birney E: **The Bioperl toolkit: Perl modules for the life sciences.** *Genome Res* 2002, **12**:1611-1618.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

