LIBERTAS ACADEMICA
FREEDOM TO RESEARCH

ORIGINAL RESEARCH

Transcriptome Analysis in Tardigrade Species Reveals Specific Molecular Pathways for Stress Adaptations

Frank Förster[1,§], Daniela Beisser[1,§], Markus A. Grohme[2,§], Chunguang Liang[1,§], Brahim Mali[2], Alexander Matthias Siegl[4], Julia C. Engelmann[5], Alexander V. Shkumatov[6,*], Elham Schokraie[7], Tobias Müller[1], Martina Schnölzer[7], Ralph O. Schill[3], Marcus Frohme[2] and Thomas Dandekar[1]

[1]Dept. of Bioinformatics, Biocenter University of Würzburg, 97074 Würzburg, Germany. [2]Technical University of Applied Sciences, Molecular Biology and Functional Genomics, 15745 Wildau, Germany. [3]Dept. of Zoology, Institute for Biology, Universität Stuttgart, 70569 Stuttgart, Germany. [4]Dept. of Microbial Ecology, University of Vienna, Althanstr. 14, 1090 Vienna, Austria. [5]Inst. for Functional Genomics, University of Regensburg, Josef-Engert-Str. 9, 93053 Regensburg, Germany. [6]EMBL, Hamburg Outstation, Notkestrasse 85, 22603 Hamburg, Germany. [7]Functional Proteome Analysis, German Cancer Research Center, Im Neuenheimer Feld 580, 69120 Heidelberg, Germany. [§]Contributed equally
*Present address: Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 250 Longwood Ave., Boston, MA 02115, USA
Corresponding author email: marcus.frohme@th-wildau.de; dandekar@biozentrum.uni-wuerzburg.de

Abstract: Tardigrades have unique stress-adaptations that allow them to survive extremes of cold, heat, radiation and vacuum. To study this, encoded protein clusters and pathways from an ongoing transcriptome study on the tardigrade *Milnesium tardigradum* were analyzed using bioinformatics tools and compared to expressed sequence tags (ESTs) from *Hypsibius dujardini*, revealing major pathways involved in resistance against extreme environmental conditions.

ESTs are available on the Tardigrade Workbench along with software and databank updates. Our analysis reveals that RNA stability motifs for *M. tardigradum* are different from typical motifs known from higher animals. *M. tardigradum* and *H. dujardini* protein clusters and conserved domains imply metabolic storage pathways for glycogen, glycolipids and specific secondary metabolism as well as stress response pathways (including heat shock proteins, bmh2, and specific repair pathways). Redox-, DNA-, stress- and protein protection pathways complement specific repair capabilities to achieve the strong robustness of *M. tardigradum*.

These pathways are partly conserved in other animals and their manipulation could boost stress adaptation even in human cells. However, the unique combination of resistance and repair pathways make tardigrades and *M. tardigradum* in particular so highly stress resistant.

Keywords: RNA, expressed sequence tag, cluster, protein family, adaptation, tardigrada, transcriptome

This article is available from http://www.la-press.com.

Introduction

Tardigrades represent one of the four main invertebrate taxa in which anhydrobiotic populations are widespread, the other three groups being crustaceans, nematodes and rotifers. In many of these species, the anhydrobiotic state may be induced over the whole life cycle, from the egg to the adult stage.[1–3] The mechanisms by which these anhydrobiotic organisms preserve their cellular integrity under extreme desiccation have remained a mystery to biologists. Previous research on the biochemistry of anhydrobiotic organisms focussed strongly on the role of polyhydroxy compounds, mainly carbohydrates, as membrane dry state stabilizers.[4,5] For instance the eutardigrade *Richtersius coronifer* was shown to accumulate the disaccharide trehalose at about 2.3% dry weight on entering the anhydrobiotic state.[6] In all other eutardigrades and heterotardigrades the accumulation is generally lower, and in *Milnesium tardigradum* trehalose is not detectable.[7]

Furthermore, physiological and molecular studies indicated the role of several stress proteins (heat shock proteins and late embryogenesis abundant proteins) in protection against desiccation damage.[8,9] Stress and late embryogenesis abundant proteins (LEAs) act as molecular chaperones and bind to other proteins—thereby preventing aggregation or unfolding of the protein or promoting protein folding.[10–13] Stress proteins may also protect cells from oxidative damage both in vivo and in vitro.[14–16] Heat-shock protein genes and proteins (Hsp70 family) have been studied in the tardigrade species *M. tardigradum* in the stage of dehydration, rehydration[17,18] and under irradiation.[19] The first verification of LEAs in tardigrades was reported by McGee et al.[20]

However, a more diversified approach is needed to fully capture the different facets of superior stress adaptation in these organisms. Advances in transcriptomics allow one to perform diversified large-scale comparisons exploiting different expressed sequence tag (EST) resources for tardigrades. The present work is based on the analysis of a new large-scale EST dataset generated from the cosmopolitan terrestrial tardigrade *M. tardigradum*, a species that is well known for its remarkable resistance to heat, radiation and cold.[21–23] To identify general tardigrade stress adaptations as well as specific pathways unique to *M. tardigradum* and its extremely high stress resistance, we compared these data to another well-studied tardigrade, *Hypsibius dujardini* (which is well studied for instance regarding development)[24] and, in addition, has a large amount of ESTs that are publicly available. In comparison, it is clearly less resistant to environmental stress than *M. tardigradum*. To make the most of this accumulating EST evidence, all sequences compared are made available on the Tardigrade Workbench[25] including software updates and various comparison options to different tardigrade species and other organisms. The accumulated ESTs allow researchers to identify tardigrade-specific RNA motifs and tardigrade-specific (*M. tardigradum* and/or *H. dujardini*) ESTs as well as ESTs where the implied function occurs in other species, including *Homo sapiens*. The analysis of encoded proteins permits positive predictions on the presence of pathways. However, tardigrade proteome, metabolome, transcriptome, and genome data have to be accumulated further for detailed analysis. Our data establish and show a rich repertoire of tardigrade stress protective pathways (eg, LEAs, heat shock, and bmh2 pathway) including protective metabolism (storage pathways, fatty acid biosynthesis). Furthermore, the repertoire for *M. tardigradum* is clearly richer than in *H. dujardini*, in particular regarding repair functions (oxidative damage, DNA repair, RNA and protein repair). Using EST clustering and annotation of conserved domains, we have annotated tardigrade-specific protein clusters and attempted to determine their involvement in stress protection while verifying suggested pathways by further tardigrade ESTs.

The analyses support the concept that tardigrades combine several strategies to achieve high stress resistance. This does not allow perfect protection and damage is efficiently repaired, particularly in *M. tardigradum*. Such a combination enhances protein protection and turnover, redox protection and specific protection of DNA. In addition to tardigrade-specific adaptations, several of the stress-protective pathways are more generally used and occur in human cells.

To understand the extremely strong tardigrade stress adaptation capabilities we conducted a detailed analysis of their transcriptome in terms of encoded proteins and pathways. This analysis shows that a multitude of repair pathways are critical, including tardigrade-specific ones in combination with classical stress resistance pathways.

Methods

Tardigrade culture

Cultures of the cosmopolitan eutardigrade species *M. tardigradum* Doyère, 1840 (Apochela, Milnesidae) 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 °C ± 2 °C and a light/dark cycle of 12 h. Rotifers of the species *Philodina citrina* and nematodes *Panagrellus* sp. were provided as food source to adults, while, juvenile tardigrades were additionally fed with green algae *Chlorogonium elongatum*. For all experiments adult animals in good condition were taken from the culture and starved for three days. This prevents the preparation of additional RNA originating from incompletely digested food in the intestinal system.

Sequence resources

We obtained all ESTs of *M. tardigradum* from our ongoing sequencing project of *M. tardigradum*. For *H. dujardini* sequences we obtained all available EST sequences from GenBank.[26] Proteins were predicted using a BLASTX[27] search against UniProtKB/SwissProt-, UniProtKB/TrEMBL- and NR-database. The open reading frames (ORFs) for nucleotide sequences, showing significant results (*E*-value <0.001), were extracted. To be more efficient, only sequences without results were searched against more extensive databases (UniProtKB/TrEMBL—and NR-database). For sequences without homology the longest ORF was extracted. All available new and existing sequence information (RNA, encoded proteins) was integrated on the Tardigrade Workbench. For this a new server was hosted at http://waterbear.bioapps.biozentrum. uni-wuerzburg.de. It includes enhanced management capabilities, new databases and rewritten source code. All new EST sequences have been deposited with GenBank and have GenBank accession numbers (refer to supplementary material 2 for the accession numbers of all sequences discussed in this study).

Identification of regulatory elements in *M. tardigradum* mRNAs

Here, all known ESTs of *M. tardigradum* were systematically screened using the software UTRscan.[28] This software screens 30 regulatory elements for RNA regulation (translational elements, stability elements and other well known regulatory elements). The default settings for batch mode were used and the reported elements were collected.

PCR methods

Primers for different heat shock proteins (HSP10: forward 5′-CCACTTCGCTACAAACAG-3′, reverse 5′-GTGATGCCAATAGGAGTG-3′, product length 171 bp; HSP17.2: forward 5′-GTCGATGCAGTCAA ACAG-3′, reverse 5′-GCGTGCTCAGTTACTCTC-3′, product length 99 bp; HSP40: forward 5′-CTCTGGGAAGCCATTACT-3′, reverse 5′-TGAG ACTTGCTTCTCTCC-3′, product length 119 bp; HSP90: forward 5′-GTCGATGCAGTCAAACAG-3′, reverse 5′-GCGTGCTCAGTTACTCTC-3′, product length 99 bp; HSP70: forward 5′-CCAAGTCTTCGA GTGATC-3′, reverse 5′-TCTGCTCAGACAAC AGTC-3′, product length 84 bp) were used to validate EST data of *M. tardigradum* by PCR. The PCR protocol used consists of an 8 min denaturation step at 94 °C, 35 cycles of 30 s denaturation (94 °C) followed by 30 s annealing (56 °C) and 30 s elongation at 72 °C and a final elongation at 72 °C for 10 min in a reaction volume of 20 µl. A total amount of 1 µl of cDNA was added. We used primer3 for the design of the oligonucleotides.[29]

Western blotting

Total protein lysate from *M. tardigradum* (10 µg) was separated on a NuPAGE™ 4%–12% Bis-Tris mini gel (Invitrogen) using MES running buffer. A voltage of 200 V was applied until the bromophenol blue front had reached the bottom of the gel (approx. 40 cm). Separated proteins were electro-transferred onto PVDF membrane for 1.5 h at maximum 50 mA (0.8 cm⁻²) in a semi-dry transfer unit (Hoefer™ TE 77) using the following transfer solution: concentrated anode buffer (300 mM Tris-HCl pH 10.4), anode buffer (25 mM Tris-HCl pH 10.4) and cathode buffer (25 mM Tris-HCl pH 9.4, 40 mM aminohexanoic acid). The PVDF membrane was incubated in blocking buffer containing 5% non-fat milk, 0.1% Tween20 in PBS.

Detection of Hsp70 was carried out using anti-Hsp70 (BD Biosciences Pharmingen) as the primary antibody and a horseradish peroxidase conjugated anti mouse

IgG (GE-Healthcare) as the secondary antibody. For detection of LEA proteins we used Ari LEA Antibody as the primary antibody and a horseradish peroxidise conjugated anti-rabbit IgG (GE-Healthcare) as the secondary antibody. ECL DualVue marker (Western Blotting Detection kit, GE-Healthcare) was used to determine the molecular mass of the target proteins on film. Images were acquired using an Image Scanner Model UTA-1100 (Amersham Biosciences).

Sequence and domain analysis

A number of different iterative sequence alignment procedures and sequence analysis methods were used as previously described[30] including gene context, interaction predictions, domain analysis and phylogenetic tests. Similar sequences were clustered by different methods: CLANS and Inparanoid. While this clustering is described in more detail below, other methods were considered, included COG/KOG clustering, as well as clusters containing specific domains evident by PFAM[31] membership and different domain families according to Interpro[32] and SMART.[33,34]

Clustering using the CLANS method

CLANS performed an all-against-all pairwise sequence comparisons using TBLASTX.[27] Afterwards similar proteins were clustered in 3D applying the CLANS method[35] with a cut-off E-value of 0.001. The program calculates pairwise attraction values based on the HSP P-values obtained from the BLAST run. Finally, clusters were identified using convex clustering and NJ-trees with standard parameters.

Clustering using the Inparanoid method

After all-against-all pairwise sequence comparisons using TBLASTX,[27] the results of the comparison were imported into Inparanoid[36] for prediction of orthologs within the ESTs. As the EST sets cover a substantial fraction of the encoded proteins but do not yet represent the whole protein set of the compared tardigrades, we called the orthologs given by Inparanoid "shared candidate orthologs sequences (sCOS)" and the remaining sequences "candidate single sequence ESTs". The sCOS represent the lower bound of identified orthologs, as further identified proteins will certainly increase the number of orthologs between the two species of tardigrades.

Sequence comparisons to identify tardigrade proteins and stress pathways

We conducted systematic sequence to sequence comparisons starting from verified orthologs of a number of organisms (*Drosophila melanogaster*, *Caenorhabditis elegans*; different plants; fungi; as well as vertebrates) and used these to screen tardigrade proteins for similarity. Good hits to stress proteins and best hits to LEA proteins were further examined for functionality by a number of sequence analysis methods comprising sensitive back-searches,[37] Prosite motifs and domain composition.[30] Furthermore, we conducted a large-scale mapping of all available sequence data from *M. tardigradum* and *H. dujardini* to clusters of orthologous groups found in complete prokaryotic genomes (COGs[38]) or in complete eukaryotic genomes (KOGs[39]). This mapping allowed us to directly predict individual protein functions and protein families by comparing the sequence to well established families from known genome sequences. For this, the latest version (last major update in 2008) of the COG/KOG database was used.[40] Briefly, proteins were classified into the following major COG/KOG classes and functions: (A) RNA processing and modification; (B) chromatin structure and dynamics cycle control and mitosis; (E) amino acid metabolism and transport; (F) nucleotide metabolism and transport; (G) carbohydrate metabolism and transport; (H) coenzyme metabolism; (I) lipid metabolism; (J) translation; (K) transcription; (L) replication and repair; "(M) cell wall/membrane/envelope biogenesis; (N) cell motility; (O) post-translational modification, protein turnover, chaperone functions; (P) inorganic ion transport and metabolism; (Q) secondary structure; (T) signal transduction; (U) intracellular trafficking and secretion; (V) defence mechanisms; (Y) nuclear structure; (Z) cytoskeleton. The classification considers also unresolved cases: (R) general functional prediction only; (S) function unknown.

Phylogenetic comparison of LEA proteins

LEA proteins from *H. dujardini* and *R. coronifer* were collected after the annotation of ESTs. LEA sequences from other organisms were collected from NCBI.[40]

These sequences were aligned and distances were calculated applying ClustalW.[41] Bootstrap supports

for the different branches were computed. The final unrooted tree (Fig. 4) was drawn using the software FigTree (version 1.2.3; http://tree.bio.ed.ac.uk/software/figtree).

Results

Transcriptome generation and analysis platform

H. dujardini ESTs were generated using a directional cDNA library. The library was constructed from adults and juveniles from an established laboratory culture derived from a single parthenogenetic female (culture ED52). As *H. dujardini* releases eggs in the shed cuticle at each post-larval moult, the sample also includes egg and embryo transcripts. A total of 5235 sequences are deposited in dbEST.

M. tardigradum ESTs were generated as part of a research program to study stress adaptations for this species. The ESTs were generated from a directional cDNA library. For these experiments adult animals in a good physical condition were taken directly from the culture and starved for three days to avoid preparation of additional RNA originating from incompletely digested food in the intestinal system. The derived *M. tardigradum* ESTs are a mixture of inactive stage animals and active motile animals.

A total of 9982 *M. tardigradum* sequences were generated in a first round. After filtering out low quality sequences we finally obtained 6593 sequences which were deposited in dbEST. Further newly determined sequences were added in the same way. All GenBank accession identifiers for the sequences analyzed in this manuscript are listed in Additional file 2. Statistical sampling predicts that this represents a substantial fraction of the total transcriptome. In particular, the majority of the highly expressed pathways are already sampled.

Extended and updated tardigrade workbench

The tardigrade workbench server[25] was updated accordingly and now includes the new data mentioned above, as well as all previous public EST data on tardigrades. Furthermore, by adding systematic translation, all encoded proteins and standard protein databases are available for comparison at the workbench. An online tutorial was added at http://waterbear.bioapps.biozentrum.uni-wuerzburg.de/cgi-bin/about.pl.

Moreover, the data structure and the management of the database were improved by changes in database implementation. New software includes a faster application-specific pattern-search, a new version (2.0) of the regular expression and Perl search as well as the possibility to load a fasta-format file with multiple sequences to screen for promoter patterns (including tardigrade-specific or general promoter motifs of interest).

We included novel options for COG statistics (see also below). Information for any COG can be examined and suitable ESTs can be downloaded. The new server is available at http://waterbear.bioapps.biozentrum.uni-wuerzburg.de and serves as an analysis platform for further studies of tardigrade molecular makeup.

Regulatory RNA motifs involved in post transcriptional regulation in *M. tardigradum*

We systematically searched for regulatory RNA motifs, comparing all ESTs (new and old) available (Table 1) within the EST data from *M. tardigradum*. From a screen of 30 well characterized regulatory motifs (see materials and methods) only twelve were detected in *M. tardigradum* sequences. Frequently used regulatory motifs found included mRNA stability signals, such as the LOX 15-DICE element,[42] brd box,[43] and the alcohol dehydrogenase 3′UTR down-regulation control element.[44] In contrast, other elements, such as the widely used AU elements in vertebrates,[45] were not found in the *M. tardigradum* ESTs.

Table 1. Regulatory motifs found by UTR-Scan in 3264 *Milnesium tardigradum* unique genes.

Motif	Number of hits (number of unique genes)
15-LOX-DICE	224 (218)
ADH_DRE	53 (52)
Brd-Box	107 (106)
CPE	37 (37)
GY-Box	98 (96)
IRE	4 (4)
IRES	1029 (1029)
K-Box	315 (295)
SECIS-1	4 (4)
SECIS-2	18 (18)
TGE	4 (4)
TOP	6 (6)

Translational elements such as the IRE were found, but are rare in *M. tardigradum* sequences.

Detailed comparison of different protein clusters in tardigrades

Results from different clustering methods were combined to get maximum information about detected tardigrade EST clusters (details are given in supplementary material).

The software Inparanoid clustered similar sequences in each species (candidate ortholog sequences, COS) with exactly the same domain composition, as detected by Inparanoid.[36] This identified sCOSs present both in the *H. dujardini* and *M. tardigradum* EST data (Fig. 1, top). Furthermore, Inparanoid also identified whether sequences with similar domain composition occur in the same species (paralogues). There were 178 sCOSs, corresponding to 214 *M. tardigradum* and 234 *H. dujardini* EST clusters.

The set of sCOSs is a conservative estimate as only the full genome sequence of both organisms would allow us to identify all orthologs. However, these COSs already show that a substantial fraction of homologous proteins exist in both species, with a major fraction
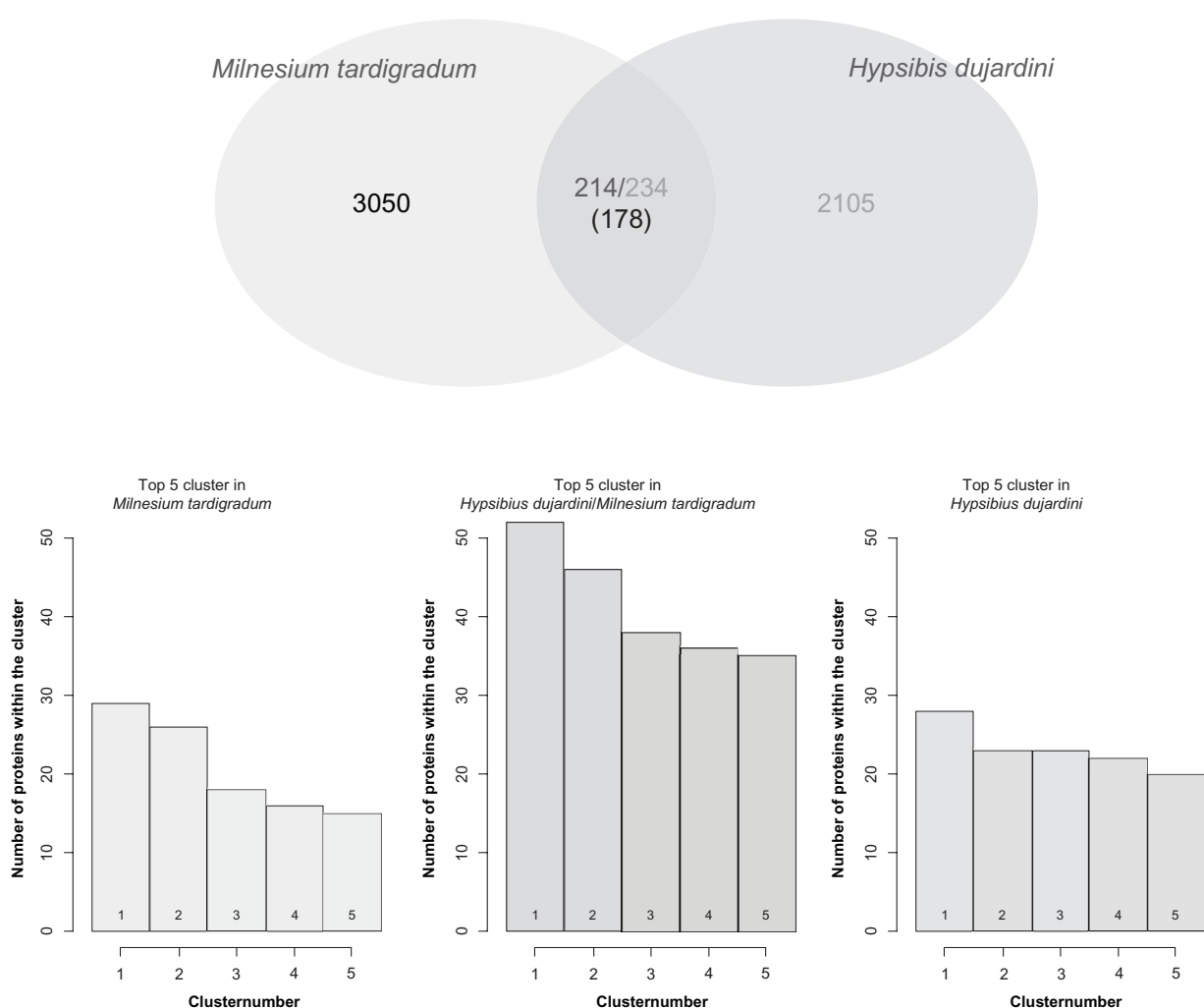
Figure 1. Sequence similar protein clusters in *Milnesium tardigradum* and *Hypsibius dujardini*.
Notes: Top: Candidate ortholog sequences (COS), present in the *H. dujardini* and *M. tardigradum* ESTs. The Venn diagram illustrates the results given by Inparanoid. The left and the right sequence sets show the COS ESTs for *M. tardigradum* and *H. dujardini*. Between the two COS-sets are the shared COS (sCOS) from *M. tardigradum* (left) and *H. dujardini* (right). The number of shared orthologs is given in brackets, above are given the number of sCOS for both species. For a complete listing see Additional File 1, S-Table III. Bottom: Number of clusters containing no less than ten sequence similar proteins according to CLANS clustering. The top 5 clusters for *M. tardigradum* are: 1- Cytochrome p450 like (29 proteins), 2- no homology (26 proteins), 3- Cathepsin like (18 proteins), 4- Rab subfamily like (15 proteins), 5- Ras subfamily like (16 proteins). Top 5 clusters for *H. dujardini* are: 1- no homology (28 proteins), 2- Chitin binding protein like (23 proteins), 3- Ras related proteins (23 proteins), 4- Cystatin like (22 proteins), 5- Glutathione S-transferase like (20 proteins). Top 5 clusters shared by both tardigrades are: 1- no homology (52 proteins), 2- Ras subfamily like (46 proteins), 3- Cathepsin like (38 proteins), 4- Chitin binding protein like (36 proteins), 5- Cystatin like (35 proteins). Details and all found clusters are given in Additional File 1, S-Table Ia–c.

including adaptations against environmental stress (Fig. 1, top).

CLANS clustering,[35] for comparison, identifies de novo protein families in any collection of sequences by sequence similarity (cut-off see materials and methods) alone. CLANS clustering does not demand a sequence to have an ortholog in another species or complete genomes (Fig. 1, bottom).

There are 13 CLANS clusters consisting of 203 proteins for *M. tardigradum*, 16 CLANS clusters consisting of 246 proteins in *H. dujardini*, and 37 CLANS clusters consisting of 684 proteins when all available ESTs from both species are pooled (Additional file 1). In each comparison there are tardigrade-specific protein clusters (a total of 314 proteins in different clusters; twelve clusters are shared). Within the *M. tardigradum* clusters, there are genes related to stress detoxification (cytochrome p450), protein turnover (cathepsin, metalloproteinase-like cluster and three ubiqutin-like clusters), protein secretion (rab-like), redox protection (glutathione S-transferase, cystatin-like) and regulation (ras, calmodulin-like). The second largest cluster is *M. tardigradum* specific. *H. dujardini* also features species-specific clusters (four with no homologies including the top cluster) and, as in *M. tardigradum,* clusters of ras-like proteins, cathepsins, glutathion S-transferase-like and cystatin-like proteins, dehydrogenases and a number of diapause-specific proteins. Diapause is a reversible state of developmental suspension occurring in plants, insects and vertebrates, including marsupials and some other mammals.[46] In tardigrades the presence of diapause-specific proteins may support either the tun form or the regulation of other metabolic inactive states (eg, in development). Cuticula protection is probably supported by chitin binding proteins. As with all further analyses below full details on cluster results are given in Additional File 1.

Specific encoded protein families and those shared with other organisms occurring in *M. tardigradum* and *H. dujardini* ESTs

To predict the repertoire of functions already visible from the currently assembled transcriptome, we compared it with the full transcriptome available from the sequencing effort on *M. tardigradum* ESTs and available data regarding prokaryotic COGs[38] and eukaryotic KOGs.[39]

These are conserved genes grouped across a large number of species (the different classes are explained in material and methods; complete lists in Additional File 1, S-Table II). Functions in *H. dujardini* include (Table 3A): molecular chaperons [category O], uncharacterized proteins [category R] including unusual protein kinases and membrane ATPases [category C]. While ribosomal proteins occur in many organisms [category J], there is COG/KOG evidence that there are specific proteins used here and that there is an increased number of specific ribosomal proteins in *M. tardigradum*. Furthermore, *M. tardigradum* (Table 2B) has even more *Milnesium*-specific, as yet poorly characterized proteins [categories S and R], metabolic enzymes [category C] and secondary metabolic enzymes involved in protein protection and turnover [category O]. Furthermore, there is evidence for specific DNA turnover and protection [category K] and transport processes including, for instance, guanine nucleotide exchange factors and SNARE proteins[47] [category U].

Shared COGs and KOGs are tardigrade-specific reinforcements for protein repair and turnover, proteasome components, cytochrome P450-like, GTPases and GTP-binding proteins (for the cytoskeleton). There are only four (KOG2733-membrane, KOG3098, KOG4604, KOG4431) uncharacterized protein clusters that are shared. The KOG4431 encodes proteins induced by hypoxia that could be involved in tardigrade-specific adaptations.

Experimental validation and phylogenetic analysis of heat shock and LEA proteins

We directly tested the presence of stress proteins in *M. tardigradum* by PCR (heat shock proteins shown in Fig. 2) and by Western blot analysis using antibodies against heat shock protein Hsp70 and LEA protein (Fig. 3). A strong band was detected for Hsp70 slightly below 75 kDa as expected. There are some degradation products visible (fainter bands, particularly in lane 3, showing the LEA protein). Furthermore, the thin upper band visible in lane 2 may indicate an isoform of Hsp70 (for instance, a minor dimer form[83] has been described for Hsp70). When using the antibody against LEA proteins, one prominent band appeared at approx. 60 kDa accompanied by several faint bands in different regions. Thus, the presence of Hsp70 and LEA proteins could be experimentally confirmed. Further details on heat shock

Table 2A. Subset of important COGs/KOGs identified uniquely in *Hybsibius dujardini*.

KOG4582:[R]Uncharacterized conserved protein, contains ZZ-type Zn-finger (9)
KOG0712:[O]Molecular chaperone (DnaJ superfamily) (8)
KOG0714:[O]Molecular chaperone (DnaJ superfamily) (8)
KOG0715:[O]Molecular chaperone (DnaJ superfamily) (8)
KOG0102:[O]Molecular chaperones mortalin/PBP74/GRP75, HSP70 superfamily (7)
KOG0691:[O]Molecular chaperone (DnaJ superfamily) (7)
KOG2835:[F]Phosphoribosylamidoimidazole-succinocarboxamide synthase (6)
COG0152:[F]Phosphoribosylaminoimidazolesuccinocarboxamide (SAICAR) synthase (5)
KOG0187:[J]40S ribosomal protein S17 (5)
COG0089:[J]Ribosomal protein L23 (4)
COG2007:[J]Ribosomal protein S8E (4)
COG0092:[J]Ribosomal protein S3 (3)
KOG1714:[J]60 s ribosomal protein L18 (3)
KOG1768:[J]40 s ribosomal protein S26 (3)
COG0355:[C]F0F1-type ATP synthase, epsilon subunit (mitochondrial delta subunit) (2)
COG1156:[C]Archaeal/vacuolar-type H+-ATPase subunit B (2)
COG1997:[J]Ribosomal protein L37AE/L43A (2)
KOG0300:[S]WD40 repeat-containing protein (2)
KOG0310:[S]Conserved WD40 repeat-containing protein (2)
KOG1235:[R]Predicted unusual protein kinase (2)
KOG3458:[C]NADH:ubiquinone oxidoreductase, NDUFA8/PGIV/19 kDa subunit (2)

Notes: Overview: The entire set contains 44 COGs and 134 KOGs with the following number of COG-classes: 2[A] 1[AD] 26[C] 1[D] 11[E] 4[F] 3[G] 1[GM] 1[GO] 4[H] 7[I] 1[IG] 1[IO] 1[IR] 24[J] 1[K] 4[M] 1[N] 1[NU] 26[O] 2[P] 1[Q] 18[R] 1[RP] 8[S] 11[T] 1[TUZ] 1[TZ] 7[U] 1[V] 2[W] 4[Z]. COGs/KOGs with two or more letters are assigned to several functional classifications. The complete table can be found in the Supplementary Table IIa.

protein distribution and phylogeny are given below and in Additional File 1.

It is known from previous studies that LEA-like proteins[48] occur in tardigrade species. They are known to improve stress resistance in other organisms, particularly in plants. As there are currently no complete sequences available for the *M. tardigradum* LEA proteins, we compared (Fig. 4) all identified LEA proteins from *H. dujardini* (13 sequences, red) and *R. coronifer* (3 sequences, orange). The identified sequences were compared to a number of other bona fide LEA proteins from other organisms (PFAM[37] seed alignment). We used the new classification by Tunnacliffe and Wise,[48] and compared our results to their LEA protein groups 1, 2 and 3. The remaining LEA groups (4–6)[48] have no similarity to the tardigrade proteins. Groups 1, 2 and 3 correspond to the PFAM domains Lea 1, Lea 2, and Lea 4 and are shown in Figure 4 (full seed alignment).

Apart from one protein, all *H. dujardini* proteins cluster in two branches, but have only Lea 4 proteins as relatives from other organisms. Thus all tardigrade sequences including all three *R. coronifer* sequences are found in the area of Lea 4-like proteins, a potential independent LEA protein family is present in *H. dujardini* (Fig. 4; GenBank identifier CO508035.1).

Predicted metabolic enzymes and resulting pathways from *M. tardigradum* and *H. dujardini* EST analysis

The EST sequencing effort on *M. tardigradum* as well as the existing EST sequences of *H. dujardini* provide better insights on central metabolic pathways involved in metabolic protection and stress resistance (Table 3, details in S-Table IV). Thus *H. dujardini* and *M. tardigradum* have key glycolytic enzymes (the list in *M. tardigradum* is even more complete). However, aldolase is found only in *H. dujardini*. Starch and sucrose metabolism is represented by a number of enzymes. On the other hand, basic fatty acid biosynthesis has only one (*H. dujardini*) or two (*M. tardigradum*) general enzymes, but there is clear evidence for fatty acid elongation in mitochondria and fatty acid conversion into different CoA-enzyme intermediates. The latter is again more completely demonstrated in *M. tardigradum* and there is also clear evidence for ketone body synthesis and conversion (which should allow long term adaptation against carbohydrate shortage). Furthermore, linoleic acids synthesis pathways are present and there are several ESTs encoding enzymes of the glycerophospholipid and sphingolipid metabolism.

Table 2B. Important COGs/KOGs identified uniquely in *Milnesium tardigradum*.

KOG0660:[T]Mitogen-activated protein kinase (12)
KOG0666:[K]Cyclin C-dependent kinase CDK8 (10)
KOG3381:[S]Uncharacterized conserved protein (6)
KOG0472:[S]Leucine-rich repeat protein (5)
KOG0473:[S]Leucine-rich repeat protein (5)
COG0271:[T]Stress-induced morphogen (activity unknown) (4)
KOG1042:[D]Germ-line stem cell division protein Hiwi/Piwi; negative developmental regulator (4)
KOG1685:[S]Uncharacterized conserved protein (4)
KOG2443:[S]Uncharacterized conserved protein (4)
KOG2989:[S]Uncharacterized conserved protein (4)
KOG3618:[R]Adenylyl cyclase (4)
KOG3737:[O]Predicted polypeptide N-acetylgalactosaminyltransferase (4)
KOG3738:[O]Predicted polypeptide N-acetylgalactosaminyltransferase (4)
COG0045:[C]Succinyl-CoA synthetase, beta subunit (2)
COG0074:[C]Succinyl-CoA synthetase, alpha subunit (2)
COG0386:[O]Glutathione peroxidase (2)
COG0553:[KL]Superfamily II DNA/RNA helicases, SNF2 family (2)
COG1278:[K]Cold shock proteins (2)
COG1758:[K]DNA-directed RNA polymerase, subunit K/omega (2)
COG1761:[K]DNA-directed RNA polymerase, subunit L (2)
COG2012:[K]DNA-directed RNA polymerase, subunit H, RpoH/RPB5 (2)
COG4973:[L]Site-specific recombinase XerC (2)
COG4974:[L]Site-specific recombinase XerD (2)
KOG0119:[A]Splicing factor 1/branch point binding protein (RRM superfamily) (2)
KOG0272:[A]U4/U6 small nuclear ribonucleoprotein Prp4 (contains WD40 repeats) (2)
KOG0388:[L]SNF2 family DNA-dependent ATPase (2)
KOG0389:[B]SNF2 family DNA-dependent ATPase (2)
KOG0391:[R]SNF2 family DNA-dependent ATPase (2)
KOG0392:[K]SNF2 family DNA-dependent ATPase domain-containing protein (2)
KOG0557:[C]Dihydrolipoamide acetyltransferase (2)
KOG0558:[C]Dihydrolipoamide transacylase (alpha-keto acid dehydrogenase E2 subunit) (2)
KOG0559:[C]Dihydrolipoamide succinyltransferase (2-oxoglutarate dehydrogenase, E2 subunit) (2)
KOG0607:[T]MAP kinase-interacting kinase and related serine/threonine protein kinases (2)
KOG0653:[D]Cyclin B and related kinase-activating proteins (2)
KOG0654:[D]G2/Mitotic-specific cyclin A (2)
KOG0655:[D]G1/S-specific cyclin E (2)
KOG0656:[D]G1/S-specific cyclin D (2)
KOG0698:[T]Serine/threonine protein phosphatase (2)
KOG0699:[T]Serine/threonine protein phosphatase (2)
KOG0870:[K]DNA polymerase epsilon, subunit D (2)
KOG1002:[L]Nucleotide excision repair protein RAD16 (2)
KOG1088:[S]Uncharacterized conserved protein (2)
KOG1435:[IT]Sterol reductase/lamin B receptor (2)
KOG1679:[I]Enoyl-CoA hydratase (2)
KOG1680:[I]Enoyl-CoA hydratase (2)
KOG1681:[I]Enoyl-CoA isomerase (2)
KOG1716:[V]Dual specificity phosphatase (2)
KOG1717:[V]Dual specificity phosphatase (2)
KOG1718:[V]Dual specificity phosphatase (2)
KOG1766:[R]Enhancer of rudimentary (2)
KOG1863:[O]Ubiquitin carboxyl-terminal hydrolase (2)
KOG1864:[O]Ubiquitin-specific protease (2)
KOG1868:[O]Ubiquitin C-terminal hydrolase (2)
KOG1873:[O]Ubiquitin-specific protease (2)
KOG2055:[R]WD40 repeat protein (2)
KOG2103:[S]Uncharacterized conserved protein (2)

(*Continued*)

Table 2B. (*Continued*)

KOG2123:[S]Uncharacterized conserved protein (2)
KOG2166:[D]Cullins (2)
KOG2167:[D]Cullins (2)
KOG2284:[O]E3 ubiquitin ligase, Cullin 2 component (2)
KOG2285:[O]E3 ubiquitin ligase, Cullin 1 component (2)
KOG2442:[R]Uncharacterized conserved protein, contains PA domain (2)
KOG2465:[S]Uncharacterized conserved protein (2)
KOG2514:[S]Uncharacterized conserved protein (2)
KOG2572:[AJ]Ribosome biogenesis protein—Nop58p/Nop5p (2)
KOG2573:[AJ]Ribosome biogenesis protein—Nop56p/Sik1p (2)
KOG3061:[O]Proteasome maturation factor (2)
KOG3150:[S]Uncharacterized conserved protein (2)
KOG3164:[R]Uncharacterized proteins of PilT N-term./Vapc superfamily (2)
KOG3214:[S]Uncharacterized Zn ribbon-containing protein (2)
KOG3228:[S]Uncharacterized conserved protein (2)
KOG3249:[S]Uncharacterized conserved protein (2)
KOG3252:[S]Uncharacterized conserved protein (2)
KOG3277:[S]Uncharacterized conserved protein (2)
KOG3377:[S]Uncharacterized conserved protein (2)
KOG3961:[S]Uncharacterized conserved protein (2)
KOG3992:[S]Uncharacterized conserved protein (2)
KOG3994:[S]Uncharacterized conserved protein (2)
KOG4018:[S]Uncharacterized conserved protein, contains RWD domain (2)
KOG4043:[S]Uncharacterized conserved protein (2)
KOG4069:[S]Uncharacterized conserved protein (2)
KOG4071:[S]Uncharacterized conserved protein (2)
KOG4093:[S]Uncharacterized conserved protein (2)
KOG4095:[R]Uncharacterized conserved protein (tumor-specific protein BCL7 in humans) (2)
KOG4096:[S]Uncharacterized conserved protein (2)
KOG4455:[S]Uncharacterized conserved protein (2)
KOG4523:[S]Uncharacterized conserved protein (2)
KOG4528:[S]Uncharacterized conserved protein (2)
KOG4615:[S]Uncharacterized conserved protein (2)
KOG4681:[S]Uncharacterized conserved protein (2)
KOG4808:[S]Uncharacterized conserved protein (2)
KOG4838:[S]Uncharacterized conserved protein (2)

Notes: *Overview:* The entire set contains 123 COGs and 497 KOGs with the following number of COG-classes: 13[A] 3[AJ] 6[B] 3[BK] 50[C] 1[CIQ] 14[D] 1[DL] 1[DN] 1[DO] 1[DR] 1[DZ] 14[E] 2[EI] 1[ET] 10[F] 34[G] 1[GMO] 10[H] 1[HE] 24[I] 1[IE] 1[IOVE] 1[IT] 62[J] 30[K] 2[KL] 1[KLB] 1[KT] 16[L] 1[LR] 7[M] 1[MJ] 1[N] 46[O] 1[ODR] 1[OK] 1[OKT] 1[OUT] 4[P] 3[PET] 1[PQ] 5[Q] 2[QR] 67[R] 1RD 1[RTU] 63[S] 43[T] 1[TDK] 1[TR] 1[TU] 2[TV] 1[TZ] 38[U] 1[UO] 1[UR] 7[V] 1[WT] 9[Z]. COGs with two or more letters are assigned to several functional classifications. The complete list of all Milnesium specific COGs/KOGs can be found in the Supplementary Table IIb.

We found several metabolite pathways implicated in enhancing stress defence, through the provision of specific metabolites. The trehalose pathway is not fully evident from EST data in *M. tardigradum*. However, *H. dujardini* has an EST encoding trehalase-like protein 1 [GenBank identifier CZ258529.1]. Furthermore, there are clear pathways for glycogen storage and glycolipids (found in both organisms), or specific protein and membrane modifications (several such enzymatic pathways available for both organisms).

Stress pathways predicted from *M. tardigradum* and *H. dujardini* ESTs compared to other organisms

The regulatory stress pathway around 14-3-3 related protein bmh2, known to exist in yeast[49] and man, is also conserved in tardigrades (Fig. 5; Additional File 1: S-Table V and VI). Conserved regulatory input protein Sck1[50] is found in *H. dujardini* as well as in *M. tardigradum*. Function predictions from sequence were verified by additional sequence analysis such as back-searches, Prosite

Table 2C. Subset of important COGs/KOGs identified both in *Hybsibius dujardini* and *Milnesium tardigradum*.

COG5272:[O]Ubiquitin (31/16)
KOG0001:[OR]Ubiquitin and ubiquitin-like proteins (31/16)
KOG0003:[J]Ubiquitin/60 s ribosomal protein L40 fusion (31/16)
KOG0004:[J]Ubiquitin/40S ribosomal protein S27a fusion (31/16)
KOG0005:[DO]Ubiquitin-like protein (31/16)
KOG0881:[O]Cyclophilin type peptidyl-prolyl cis-trans isomerase (6/12)
COG0149:[G]Triosephosphate isomerase (3/2)
KOG0266:[R]WD40 repeat-containing protein (3/5)
KOG0329:[A]ATP-dependent RNA helicase (3/8)
COG0451:[MG]Nucleoside-diphosphate-sugar epimerases (3/4)
KOG0543:[O]FKBP-type peptidyl-prolyl cis-trans isomerase (3/1)
KOG0549:[O]FKBP-type peptidyl-prolyl cis-trans isomerase (3/2)
KOG0730:[O]AAA+-type ATPase (3/10)
KOG0332:[A]ATP-dependent RNA helicase (2/6)
KOG0336:[A]ATP-dependent RNA helicase (2/6)
KOG0727:[O]26S proteasome regulatory complex, ATPase RPT3 (2/10)
KOG0658:[G]Glycogen synthase kinase-3 (1/8)
KOG0731:[O]AAA+-type ATPase containing the peptidase M41 domain (1/9)
KOG0734:[O]AAA+-type ATPase containing the peptidase M41 domain (1/8)
KOG0735:[O]AAA+-type ATPase (1/8)
KOG0736:[O]Peroxisome assembly factor 2 containing the AAA+-type ATPase domain (1/4)
KOG0737:[O]AAA+-type ATPase (1/6)
KOG0739:[O]AAA+-type ATPase (1/4)
KOG0751:[C]Mitochondrial aspartate/glutamate carrier protein Aralar/Citrin (contains EF-hand Ca2+-binding domains) (1/1)
KOG0923:[A]mRNA splicing factor ATP-dependent RNA helicase (1/2)
KOG0924:[A]mRNA splicing factor ATP-dependent RNA helicase (1/2)
KOG0925:[A]mRNA splicing factor ATP-dependent RNA helicase (1/2)
KOG0934:[U]Clathrin adaptor complex, small subunit (1/6)
KOG0935:[U]Clathrin adaptor complex, small subunit (1/6)
KOG0936:[U]Clathrin adaptor complex, small subunit (1/6)
KOG0983:[T]Mitogen-activated protein kinase (MAPK) kinase MKK7/JNKK2 (1/2)
KOG2733:[S]Uncharacterized membrane protein (1/2)
KOG3098:[S]Uncharacterized conserved protein (1/2)
KOG4431:[R]Uncharacterized protein, induced by hypoxia (1/2)
KOG4604:[S]Uncharacterized conserved protein (1/2)

Notes: The entire set contains 129 COGs and 606 KOGs with the following number of COG-classes: 42[A] 2[AJ] 1[AR] 6[B] 1[BK] 1[BL] 51[C] 1[CD] 1[CP] 1[CR] 9[D] 1[DKL] 2[DO] 1[DR] 2[DT] 1[DZ] 9[E] 1[EG] 1[EM] 5[F] 1[FGR] 10[G] 1[GT] 6[H] 1[HC] 19[I] 1[IE] 1[IQR] 1[IT] 1[IU] 127[J] 1[JD] 21[K] 4[L] 1[LKJ] 2[M] 2[MG] 129[O] 3[OC] 1[OE] 1[OR] 1[OT] 1[OUT] 12[P] 1[PT] 15[Q] 1[QI] 72[R] 1[RT] 1[RTKL] 1[RV] 17[S] 57[T] 2[TR] 2[TU] 1[TUZ] 2[TZ] 1[TZR] 38[U] 1[UR] 1[UT] 1[UZ] 6[V] 26[Z] 1[ZD]. COGs/KOGs with two or more letters are assigned to several functional classifications. The complete table can be found in the Supplementary Table IIc.

Motifs and SMART domain composition.[30] Bmh2-dependent metabolic enzymes found in both tardigrades include trehalase and acyltransferases such as LCB1 or LCB2.[49] Furthermore, proteomics data has already identified glycerol-3-phosphate acyltransferases in *M. tardigradum*.[51] The data suggests the presence of a complete stress pathway, including the associated metabolic enzymes in *H. dujardini* as well as key switches in *M. tardigradum*.

Alpert[52] compiled a collection of pathways involved in desiccation tolerance, which was investigated in the present work. For tardigrades, we found no genes indicating stress protection by photopigments.[53,54] This is in accordance with morphological data. Using systematic sequence to sequence comparisons, starting with verified orthologs of a number of organisms (see material and methods), we obtained further results regarding stress pathway proteins using the available EST data from *H. dujardini* and *M. tardigradum* (Table 4, details in Additional File 1, S-Table Va and b). Repair pathways include RNA helicases[55] as well as peroxiredoxin, peroxidase and superoxide dismutase.[56] Several LEA protein[57,58] candidates were found, including very good matches to known LEA proteins (for validation see Fig. 4). Different larger and smaller heat shock proteins[59] could be identified, including Hsp70 in *H. dujardini* and

Table 3. Enzymes in metabolic pathways predicted to be present in *Hybsibius dujardini* and *Milnesium tardigradum* according to the EST data.[1]

Hybsibius dujardini	*Milnesium tardigradum*
Glycolysis/gluconeogenesis	**Glycolysis/gluconeogenesis**
ec:1.2.4.1 pyruvate dehydrogenase (acetyl-transferring)	ec:1.2.1.12 glyceraldehyde-3-phosphate dehydrogenase (phosphorylating)
ec:1.1.1.2 alcohol dehydrogenase (NADP+)	ec:5.3.1.1 triose-phosphate isomerase
ec:2.7.2.3 phosphoglycerate kinase	ec:6.2.1.1 acetate—CoA ligase
ec:5.3.1.1 triose-phosphate isomerase	ec:1.1.1.1 alcohol dehydrogenase
ec:4.1.2.13 fructose-bisphosphate aldolase	ec:5.4.2.2 phosphoglucomutase
ec:1.2.1.12 glyceraldehyde-3-phosphate dehydrogenase (phosphorylating)	ec:1.1.1.27 L-lactate dehydrogenase
ec:1.1.1.27 L-lactate dehydrogenase	ec:1.2.1.3 aldehyde dehydrogenase (NAD+)
	ec:4.2.1.11 phosphopyruvate hydratase
	ec:1.1.1.2 alcohol dehydrogenase (NADP+)
	ec:5.3.1.9 glucose-6-phosphate isomerase
Starch and sucrose metabolism	**Starch and sucrose metabolism**
ec:3.6.1.-	ec:3.6.1.-
ec:2.4.1.11 glycogen(starch) synthase	ec:3.2.1.20 alpha-glucosidase
ec:2.7.7.9 UTP—glucose-1-phosphate uridylyltransferase	ec:3.2.1.39 glucan endo-1,3-beta-D-glucosidase
ec:4.1.1.35 UDP-glucuronate decarboxylase	ec:3.2.1.1 alpha-amylase
ec:3.2.1.4 cellulase	ec:5.4.2.2 phosphoglucomutase
ec:3.2.1.1 alpha-amylase	ec:2.7.1.106 glucose-1,6-bisphosphate synthase
	ec:5.3.1.9 glucose-6-phosphate isomerase
Fatty acid biosynthesis	**Fatty acid biosynthesis**
ec:3.1.2.-	ec:3.1.2.-
	ec:1.3.1.-
Arachidonic acid metabolism	**Arachidonic acid metabolism**
ec:5.3.99.2 prostaglandin-D synthase	ec:1.14.14.1 unspecific monooxygenase
ec:1.14.14.1 unspecific monooxygenase	ec:3.1.1.4 phospholipase A2
ec:3.1.1.4 phospholipase A2	ec:5.3.99.2 prostaglandin-D synthase
ec:1.13.11.-	ec:3.3.2.6 leukotriene-A4 hydrolase
ec:3.3.2.10 soluble epoxide hydrolase	ec:1.11.1.9 glutathione peroxidase
	ec:2.3.2.2 gamma-glutamyltransferase
	ec:5.3.99.5 thromboxane-A synthase
Aminophosphonate metabolism	**Aminophosphonate metabolism**
ec:2.1.1.-	ec:2.1.1.-
Fatty acid elongation in mitochondria	**Fatty acid elongation in mitochondria**
ec:2.3.1.16 acetyl-CoA C-acyltransferase	ec:1.1.1.35 3-hydroxyacyl-CoA dehydrogenase
ec:3.1.2.22 palmitoyl[protein] hydrolase	ec:4.2.1.17 enoyl-CoA hydratase
ec:1.1.1.35 3-hydroxyacyl-CoA dehydrogenase	
Fatty acid metabolism	**Fatty acid metabolism**
ec:1.14.14.1 unspecific monooxygenase	ec:1.14.14.1 unspecific monooxygenase
ec:2.3.1.16 acetyl-CoA C-acyltransferase	ec:1.1.1.35 3-hydroxyacyl-CoA dehydrogenase
ec:1.1.1.35 3-hydroxyacyl-CoA dehydrogenase	ec:1.1.1.1 alcohol dehydrogenase
	ec:5.3.3.8 dodecenoyl-CoA isomerase
	ec:6.2.1.3 long-chain-fatty-acid—CoA ligase
	ec:1.3.99.7 glutaryl-CoA dehydrogenase
	ec:1.3.99.2 butyryl-CoA dehydrogenase
	ec:4.2.1.17 enoyl-CoA hydratase
	ec:1.2.1.3 aldehyde dehydrogenase (NAD+)
	Synthesis and degradation of ketone bodies
	ec:2.3.3.10 hydroxymethylglutaryl-CoA synthase
Linoleic acid metabolism	**Linoleic acid metabolism**
ec:1.1.1.-	ec:1.14.14.1 unspecific monooxygenase
ec:1.14.14.1 unspecific monooxygenase	ec:3.1.1.4 phospholipase A2
ec:3.1.1.4 phospholipase A2	ec:1.1.1.-

(*Continued*)

Table 3. (*Continued*)

Hybsibius dujardini	*Milnesium tardigradum*
Glycerophospholipid metabolism	**Glycerophospholipid metabolism**
ec:2.3.1.51 1-acylglycerol-3-phosphate O-acyltransferase	ec:3.1.3.4 phosphatidate phosphatase
ec:2.7.1.82 ethanolamine kinase	ec:3.1.1.4 phospholipase A2
ec:2.3.1.6 choline O-acetyltransferase	ec:3.1.4.4 phospholipase D
ec:2.3.1.-	ec:1.1.1.8 glycerol-3-phosphate dehydrogenase (NAD+)
ec:3.1.1.7 acetylcholinesterase	ec:2.3.1.-
ec:3.1.1.4 phospholipase A2	ec:2.3.1.51 1-acylglycerol-3-phosphate O-acyltransferase
	ec:2.7.1.82 ethanolamine kinase
Sphingolipid metabolism	**Sphingolipid metabolism**
ec:1.14.-.-	ec:3.1.3.-
ec:3.5.1.23 ceramidase	ec:3.1.3.4 phosphatidate phosphatase
ec:2.4.1.-	ec:3.5.1.23 ceramidase
ec:3.1.3.-	ec:1.14.-.-
	ec:3.2.1.23 beta-galactosidase
	ec:3.1.4.12 sphingomyelin phosphodiesterase
	ec:2.4.1.-
	ec:3.1.3.-
Biosynthesis of steroids	**Biosynthesis of steroids**
ec:5.3.3.2 isopentenyl-diphosphate Delta-isomerase	ec:1.14.21.6 lathosterol oxidase
	ec:5.3.3.2 isopentenyl-diphosphate Delta-isomerase
	ec:1.1.1.170 sterol-4alpha-carboxylate 3-dehydrogenase (decarboxylating)
	ec:1.3.1.-
	ec:2.7.4.2 phosphomevalonate kinase
	ec:1.3.1.21 7-dehydrocholesterol reductase
	ec:1.14.13.72 methylsterol monooxygenase
	ec:2.5.1.10 geranyltranstransferase
	ec:5.3.3.5 cholestenol Delta-isomerase
Terpenoid biosynthesis	**Terpenoid biosynthesis**
ec:5.3.3.2 isopentenyl-diphosphate Delta-isomerase	ec:5.3.3.2 isopentenyl-diphosphate Delta-isomerase
	ec:2.5.1.10 geranyltranstransferase
Carotenoid biosynthesis	**Carotenoid biosynthesis**
ec:1.14.-.-	ec:2.1.1.-
ec:1.14.99.-	ec:1.14.-.-
ec:2.3.1.-	ec:1.-.-.-
ec:2.4.1.-	ec:2.3.1.-
ec:1.-.-.-	ec:2.4.1.-
ec:2.1.1.-	
Ether lipid metabolism	**Ether lipid metabolism**
ec:2.3.1.51 1-acylglycerol-3-phosphate O-acyltransferase	ec:3.1.3.-
ec:2.3.1.-	ec:3.1.3.4 phosphatidate phosphatase
ec:3.1.3.-	ec:3.1.1.4 phospholipase A2
ec:3.1.1.4 phospholipase A2	ec:3.1.1.47 1-alkyl-2-acetylglycerophosphocholine esterase
	ec:3.1.4.4 phospholipase D
	ec:2.3.1.-
	ec:2.3.1.51 1-acylglycerol-3-phosphate O-acyltransferase
	ec:3.1.4.39 alkylglycerophosphoethanol-amine phosphodiesterase
	ec:3.1.3.-
Bile acid biosynthesis	**Bile acid biosynthesis**
ec:1.1.1.-	ec:1.1.1.-
ec:2.3.1.16 acetyl-CoA C-acyltransferase	ec:1.1.1.1 alcohol dehydrogenase

(*Continued*)

Table 3. (*Continued*)

Hybsibius dujardini	Milnesium tardigradum
	ec:3.1.1.13 sterol esterase
	ec:1.3.99.5 3-oxo-5alpha-steroid 4-dehydrogenase
	ec:1.2.1.3 aldehyde dehydrogenase (NAD+)
	ec:1.14.13.15 cholestanetriol 26-monooxygenase
Retinol metabolism	**Retinol metabolism**
ec:1.1.1.-	ec:1.14.14.1 unspecific monooxygenase
ec:1.14.14.1 unspecific monooxygenase	ec:1.1.-.-
ec:3.1.1.-	ec:3.1.1.-
ec:1.1.-.-	ec:1.1.1.-
ec:2.3.1.20 diacylglycerol O-acyltransferase	ec:1.1.1.1 alcohol dehydrogenase
alpha-Linolenic acid metabolism	**alpha-Linolenic acid metabolism**
ec:3.1.2.-	ec:3.1.1.4 phospholipase A2
ec:2.3.1.16 acetyl-CoA C-acyltransferase	ec:1.-.-.-
ec:1.-.-.-	ec:1.1.1.35 3-hydroxyacyl-CoA dehydrogenase
ec:3.1.1.4 phospholipase A2	ec:3.1.2.-
ec:1.1.1.35 3-hydroxyacyl-CoA dehydrogenase	ec:4.2.1.17 enoyl-CoA hydratase

M. tardigradum.[17,19,51] DNA repair appears to be mediated by the MutS pathway.[60] Furthermore, *M. tardigradum* contains specific DNA repair (RAD51, DnaJ family[61]) as well as other DNA protection pathway proteins (DNA helicases). According to the analyzed EST data in tardigrades a specific animal signalling pathway is present, the HOG (high osmolarity glycerole) signalling pathway, including PKA-mediated regulation.[62]

Tardigrade-specific stress adaptations versus general adaptations shared with other animals

There are two other desiccation resistant animal phyla, rotifera and nematodes. To get an insight into
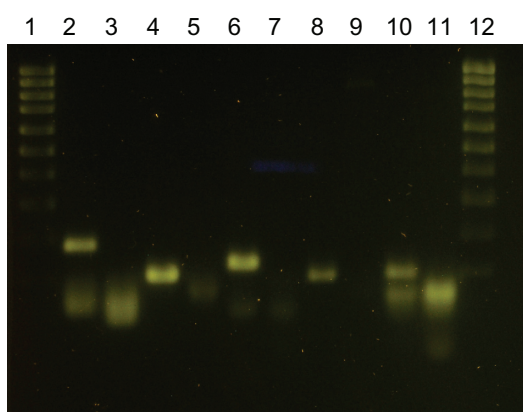

Figure 2. PCR validation of heat shock protein transcripts in *Milnesium tardigradum* and *Hypsibius dujardini*.
Notes: Shown are detected PCR bands using heat shock protein encoding EST primers as described in M&M. Lanes: (1) Hyperladder IV (Bioline); (2) hsp10 (expected: 171 bp); (3) hsp10 Non template control (NTC); (4) hsp20/30 (99 bp); (5) hsp20/30 NTC; (6) hsp40 (119 bp); (7) hsp40 NTC; (8) hsp70 (84 bp); (9) hsp70 NTC; (10) hsp90 (99 bp); (11) hsp90 NTC; (12) Hyperladder IV (Bioline).

the general and tardigrade specific adaptations we compared our inventory of stress pathways found in *M. tardigradum* and/or *H. dujardini* with stress pathways known from rotifera or nematodes. Conservation of key pathways in man is shown in Figure 6. We considered all stress protection proteins found in either *M. tardigradum* or *H. dujardini* and examined whether these proteins are also found in nematodes, rotifers or man. We also examined specific genes involved in nematode or rotifer stress response and whether these are found in tardigrades.

A first overview on this is shown by COGs/KOGs[38,39] and using available genome information regarding protein clusters in nematodes (*C. elegans*) and *H. sapiens*.[40] We considered three major COGs/KOGs[40] categories involved in protective functions by comparison of our tardigrade data and protein clusters to the other two organisms regarding the COGs/KOGs categories V (defence mechanisms), O (posttranslational modification) and L (replication, recombination and repair). The overview (Venn diagram, Fig. 6) shows that there are many shared functions (172), but only 81 tardigrade-"specific" protein clusters (ie, they were neither found in man nor in nematodes). The largest overlap between organism groups is between nematodes and man. To better identify unique tardigrade adaptations we subtracted from these clusters any clusters common for tardigrades and yeast. Compared to rotifers, nematodes, human and yeast these are proteins occurring only in tardigrades (Fig. 7) yielding a list of 46 specific adaptations and COGs/KOGs clusters found only
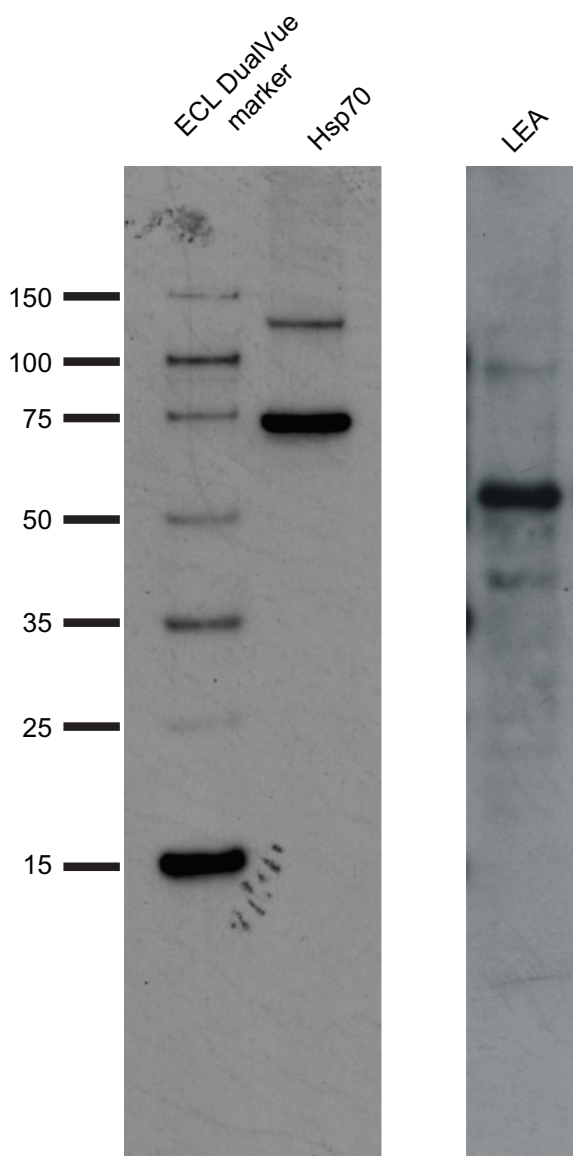
Figure 3. Western blot of heat shock and LEA proteins in *Milnesium tardigradum*.
Notes: The presence of Hsp70 (middle panel) and LEA proteins (right panel) in *M. tardigradum* was confirmed by Western blot analysis using the corresponding antibodies. The ECL DualVue size marker is shown on the left.

in tardigrades within the context of this comparison (Table 5; detailed results in Additional File 1, S-Table VII). They include COGs for glutaredoxin-related proteins and molecular chaperons as well as KOGs involved in molecular chaperoning. This includes the DnaJ superfamily with a large number of different clusters (Milnesium-specific, Table 4C; Additional File, S-Tables II and V). The COG clusters 4973, 4974 supply site-specific recombinases.[63] Collectively, these strongly enhance DNA repair and explain the recently documented resistance of

M. tardigradum against DNA damage.[64,65] Details on all involved functions and protein clusters are given in Additional File 1, including basic COGs/KOGs shared with yeast in tardigrades (Table S7). Figure 8 summarizes all EST predicted adaptations and compares their distribution in other model organisms, focussing on the predicted encoded protein families. Tardigrades combine adaptations found in nematodes, rotifers and vertebrates, with tardigrade-specific adaptations. Moreover, there are a number of tardigrade specific EST clusters in which sequence similarity does not allow the prediction of any specific function. Interestingly, some protein functions (11) are shared only between tardigrades and *H. sapiens* in this comparison (Fig. 8 and Additional File 1, Table 7D): These are pathways including the DNA repair protein RHP57[66] and ubiquitin protein ligase, as well as proteasome maturation factor.[67] Sequence information on rotifers is only limited available. A mitochondrial chaperonin is shared with man, rotifers and *C. elegans*. A Hsp90 family-type molecular chaperone[68,69] is also shared between rotifers and tardigrades (see Additional File 1, S-Table VIIf and g).

Tardigrade-specific EST clusters

Annotation of the tardigrade-specific EST clusters is non-trivial, as major parts of the sequence have only a weak similarity to other organisms. Here, we used a combination of EST sequencing, clustering, annotation of conserved domains and a search for ESTs encoding proteins related to the domain-predicted pathways to make functional predictions of these particular tardigrade-specific EST sequences.

Annotated domains in tardigrade-specific protein clusters (TSPs) are shown in Figure 9. We looked for further ESTs in *M. tardigradum* which supported these predictions using the pathway context (middle panel, at the bottom). In particular, we examined whether some ESTs encoded a regulatory protein for the annotated TSP putative protein or an interacting protein (eg, another enzyme of the predicted pathway). For this, all tardigrade clusters identified by CLANS (including all TSPs published previously[25]) were systematically screened for domains with a known function (applying HMM models from PFAM[31] for different domains, latest version; see Additional File, S-Table VIII). Note that these TSPs include sequences
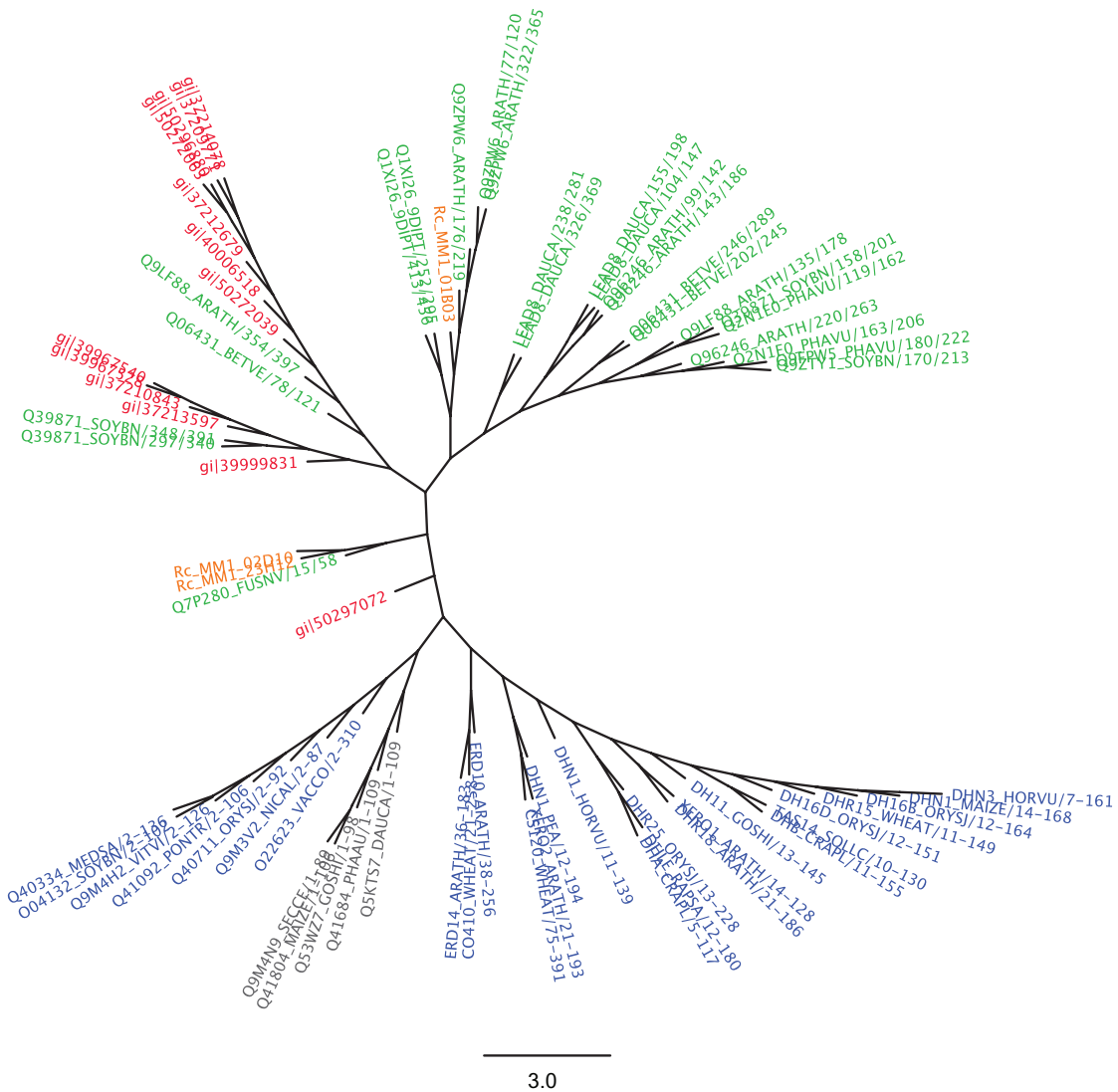
**Figure 4.** Phylogenetic tree of different tardigrade LEA-like proteins.
Notes: Besides several LEA proteins tardigrade sequences from *Hypsibius dujardini* (13 proteins, red) and *Richtersius coronifer* (3 sequences, orange) all sequences from the seed alignments for LEA1 (blue), LEA2 (grey) and LEA4 (green) profile from PFAM are shown. LEA groups 3, 5, 6[48] are even less related. Bootstrap values were tested but are not shown here to prevent cluttering. The tree was generated by Figtree (version 1.2.3).

from *H. dujardini*, *R. coronifer* and *M. tardigradum*. However, a search for PFAM domains contained in the sequences associated with *M. tardigradum* was successful only for several TSP clusters (Table 5). This means that the molecular function conferred by the respective domains is utilized in these clusters and can be predicted to be present in *M. tardigradum*. However, as the rest of the sequences could not be annotated (otherwise these would not be TSPs but shared clusters with other organisms) there are additional molecular functions not yet recognized, that modify the function of the complete protein (eg, regulate it in response to stress or may confer completely novel functions). On the other hand, the domain hits

themselves are all highly significant (*E*-value $<10^{-6}$ and the significant domain hit covers at least 20% of the cluster sequence, see Table 5 TSPs; and Additional File 1, S-Table VIII).

Additional domains, such as regulatory or binding domains, modify these predicted functions encoded in the TSPs. Specific domains identified in TSPs as parts of the complete sequence include similarity to osteopontin (TSP8), lipolysis stimulated receptor (TSP18), sulfite exporter TauE (TSP 19), signal peptidase subunit domain (TSP37), Grb (TSP 51), as well as similarity to the DUF19 domain where we can annotate the DUF domain function (TSP55). TSPs 30, 35 and 42 have similarities to uncharacterized
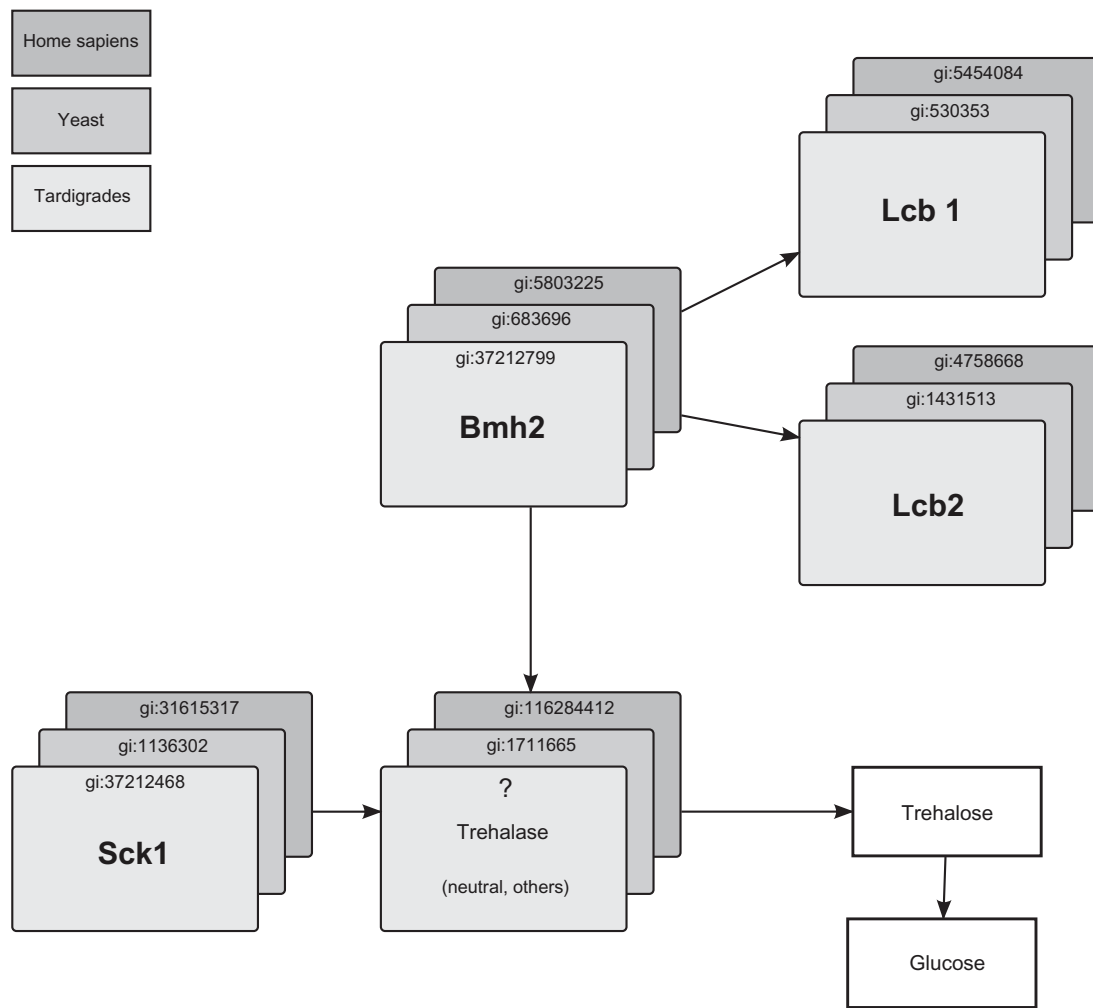
Figure 5. Bmh2 stress pathway is conserved in tardigrades.
Notes: The bmh2 pathway was shown to exist not only in vertebrates but also in yeast. It is also conserved in *Hypsibius dujardini* and *Milnesium tardigradum*. Key regulators (bmh2, Sck1) as well as depending metabolic enzymes are indicated together with GenBank identifiers (top) comparing man, yeast and *H. dujardini*. There are sequence similar candidate sequences for Lcb1 and Lcb2 in *H. dujardini*, but it is not known which is regulated by bmh2 or Sck1. Trehalase has not been found (question mark). In *Milnesium tardigradum* only the two regulators, bmh2 and Sck1 have been identified.

Table 4A. Specific stress pathway proteins predicted to be present in *Hypsibius dujardini* according to the analyzed EST data.[1]

DNA repair	eg, MutS (gb|CD449386.1)
LEA proteins	eg, gb|CK325833.1
DNA protection (RNA helicase)	eg, gb|CD449793.1
Redox protection (peroxidase and superoxid dismutase)	eg, gb|CK326879.1 and eg, gb|CK326506.1
Heat shock proteins	eg, gb|CF544577.1, gb|AAQ94878.1, gb|CD449707.1
Aquaporins	eg, gb|CD449847.1
Not detected Pigments Anti-freeze protein	

Note: [1]Key examples; more information in Additional File 1, S-Table Va.

domain families (DUFs: DUF568, DUF1481 and DUF2543, respectively) and TSP 33 is completely uncharacterized.

Table 5 shows that this partial information is nevertheless enough to predict part of the tentative protein function and to make an initial estimate on potential interactors or pathways that the TSP may be involved in:

In detail, TSP8 encodes domains similar to osteopontin (eg, tardigrade sequence GenBank identifier GE637688.1). Osteopontin enhances secretion, cell motility, and extracellular matrix invasion and is a general cytokine.[82]

TSP18 encodes domains with similarity to acetyl-CoA dehydrogenase C-terminal domain and sequences similar to lipolysis stimulated receptor

Table 4B. Specific stress pathway proteins predicted to be present in *Milnesium tardigradum* according to the translated EST data.[1]

DNA repair	DNA repair protein RAD51, DnaJ Family
DNA protection (RNA helicase)	eg, Helicase_C, DEAD, ATP-dependent RNA helicase, Spliceosome RNA helicase BAT1, Peroxiredoxin-5, Redoxin
Redox protection (superoxid dismutase and peroxidase)	eg, Superoxide dismutase [Cu-Zn], Sod_Cu, AhpC-TSA, Peroxiredoxin-1,-2,-4,-6
Heat shock proteins	eg, Heat shock protein 90, HSP 70, HSP 20, HSP 30
Cold shock protein	Cold shock-like protein cspC
Membrane permeability	eg, MIP (Major Intrinsic Protein), Aquaporin-9,-10
Not detected Pigments Anti-freeze protein	

Note: [1]Key examples; more information in Additional File 1, S-Table Vb.

(eg, GenBank identifier CO741640.1). This indicates a role in lipolysis and fatty acid metabolism is possible, supported by the presence of enzymes for lipid metabolism as noted above.

TSP 19 contains domains similar to taurine E, a sulfite exporter (eg, GenBank identifier CO741254.1). The *E*-value is significant (5.4e-06) and the resulting metabolite, taurine, is known to be a strong antioxidant. However, we did not find pertaining to the pathway such ESTs in *M. tardigradum* (eg. taurine dehydrogenase and sulfite transferase). As the sequences are clearly glycine-rich, these sequence may be implicated in similar tasks as TSP51 (see below).

TSP37 encodes sequences with domains having significant similarity to a signal peptidase subunit domain spc22/23, a well conserved signal peptidase (eg, GenBank identifier CK326237.1). Hence a tentative function prediction would be that the cluster is involved in protein secretion and transport.

TSP51 contains domains related to GRP (Glycine rich protein family; see eg, GenBank identifier CO741993.1). It is known that this domain is found in proteins required for general stress response. These proteins are cytokeratin-like proteins; glycine-rich and RNA-binding proteins. Particular regulators for general stress response are heat shock proteins and a number of them have been identified in *M. tardigradum* (this paper and Reuner et al[17]).

TSP55 could be annotated as it contains the DUF19 domain (match over 126 amino acids in alignment; see eg, GenBank identifier GE637487.1). DUF19 domain is found in one or two copies of several *C. elegans* proteins. It is roughly 130 amino acids long and has 12 conserved cysteines, which form six intradomain disulphide bridges, suggesting that this domain is an extracellular one. The GO annotation for this protein indicates that it has a function in nematode larval development and is a positive regulator of growth rate. Interaction database predictions[70] suggest that the function of the domain is connected to the metabolism of methylations (tRNA methylation, C1-methylation). Further annotation of the DUF19 domain is possible due to the homologs in the genome of the sea-ice bacterium (*Psychromonas ingrahamii*). The neighbouring gene in this genome is known to be involved in tetrahydropterin synthesis. This activity supports amino acids and thymidine biosynthesis and would be important for larval development, as well as growth in general. Furthermore, this neighbour combination is conserved across genomes. We conclude that these genes are probably involved in tetrahydropterin synthesis.

We found several further genes for tetrahydrofolate synthesis in *M. tardigradum* annotated by their homology. These are the genes sepiapterin reductase (including 7,8-dihydrobiopterin:NADP + oxidoreductase [EC:1.1.1.153]) and GTP cyclohydrolase 1 [EC:3.5.4.16] as well as DHFR [EC:1.5.1.3] ([GenBank: GR865643.1], [GenBank: GR865643.1], [GenBank: EZ758936.1], [GenBank: EZ761524.1], respectively). The NO synthase [GenBank: EZ762352.1] also occurs in *M. tardigradum* (long, significant alignment (*E*-value $< 10^{-72}$) over more than 600 residues). Tetrahydropterin synthase regulators in vertebrates are NO as well as NO-Synthase.[71]

Finally, we analyse a new TSP-specific cluster identified from the ongoing sequencing[72] where partial-annotation was possible (see materials and methods). For this TSP a chitin binding domain type 2 could be annotated.[61] Chitinases hydrolyze chitin (N-acetyl-ß-D-glucosamine) in biological structures such as insect exoskeleton, fungal cell wall, nematode egg shell, and other biological structures involved in protection and self-defence. Furthermore, chitinases along with chitin synthetases are essential for remodelling chitin-containing structures during growth and

Table 4C. Specific stress pathways derived from proteins predicted to be present in *Milnesium tardigradum*.[1]

|	
Bmh2 pathway	**Bmh2 (key regulator), Sck1 (key regulator),** Lcb1 (depending enzyme)[+], Lcb2 (depending enzyme)[+] (Fig. 5)[49]
Major DNA repair pathways	
NHEJ (non-homologous end joining):	Rad50 (regulated response)
HRR (homologous recombination):	RuvB, Rad51, Rad50 (regulated responses)
MMR (mismatch repair):	MutS, PCNA (regulated response)
NER (nucleotide excision repair):	Rad23 (regulated responses)
BER (base excision repair):	XRCC1 (regulated response)
Heat- shock response (HSR) pathways	10 kDa GroES/Hsp10 (regulated response) 20–30 kDa GrpE/Hsp27, Hsp20 (Interaction partner of Mef2) 40 kDa DnaJ/Hsp40 (regulated response, stimulates Hsp70) 70 kDa DnaK/Hsp70 (regulated response) 90 kDa HtpG/Hsp90 (regulated response) Mef2 (key regulator), GAPDH (depending enzyme)[81] HSBP1 (regulator of TF HSF1)
LEA pathways	LEA
HOG signal pathway	Hsp12; PBS2, HOG1 and PKA (regulators of Hsp12 in *S. cerevisiae*, found in tardigrades)[62]
No Plant ABA responsive pathway	vp1, vp5 (regulators in *Z. mays, A. thaliana*, not found in tardigrades) B15C (peroxiredoxin), Rab 16 (rice, maize, barley, not found in tardigrades)[56] abi3 (*A. thaliana*, TF: regulates ABA-responsive genes, not found in tardigrades)
Protective pathways	SOD (enzyme), Prx-1,-2,-4,-6 (regulated enzyme by StRE) AP-1 and Nrf2 (redox regulated TFs for StRE, Nrf2 was not found, but its inhibitor INrf2)
RNA helicase	cgh-1, BAT1, BRR2*, FAL1* (regulated responses)
DNA helicase	HelicaseC, RAD5*, TIP49* (regulated responses)

Note: [1]Key examples; more information in Additional File 1, S-Table Vc; *identified in KOGs/COGs, see supplementary table II on our tardigrade website. [+]Lcb was found, but it is not clear which EST is Lcb1 and which Lcb2.

development as well as for defence from chitinous pathogens.[73] In tardigrades we suggest that they will be involved in remodelling chitin during the shift between active and tun stage and that the identified specific tardigrade cluster is also involved in physical protection by chitin. With the identified TSP cluster interactors and regulators of chitinase in the form of binding partners are present in *M. tardigradum*. Chitinase itself is also encoded (GenBank: AAB81847.1; 117 residue alignment, *E*-value < 10[−6]).

Furthermore, regulation[74] may be similar to insects, as there is an ecdyson receptor related EST present (GenBank identifier EZ761204.1). Insect regulation relies upon ecdysone producing enzymes and ecdysone receptor. Supporting evidence to place the tardigrades closer to this super-phylum is provided by extensive steroid synthesis, Table 3).

To combine all different pathways and adaptations analyzed, Figure 9 summarizes general stress pathways for *M. tardigradum* (left), tardigrade-specific pathways (in the sense of the comparison man/rotifers/nematodes/*Drosophila*/tardigrades, middle panel, top) and pathways shared only with human in this comparison (right). In each of these
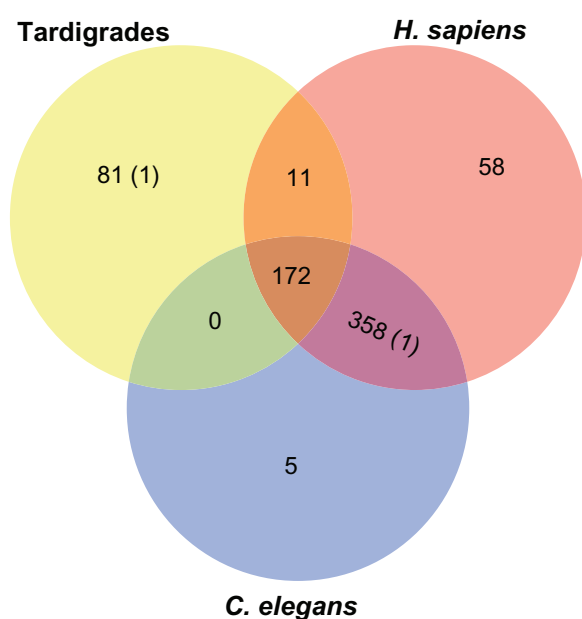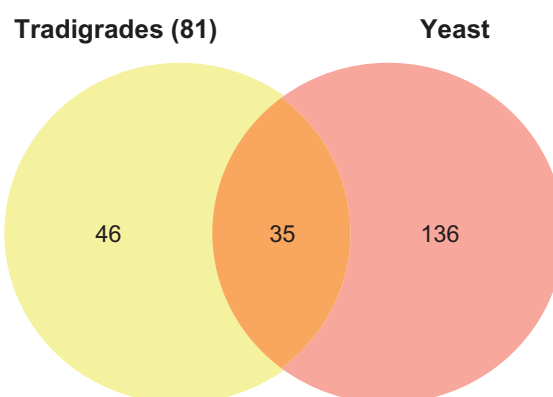
Figure 6. Comparison of COGs/KOGs for *Milnesium tardigradum, Hypsibius dujardini* (Tardigrades), *Homo sapiens* and *Caenorhabditis elegans* (Nematodes).
Notes: The Venn diagram depicts intersections of COGs/KOGs between the two desiccation resistant phyla tardigrades and nematodes (*C. elegans*) versus *H. sapiens*. The numbers in brackets indicate COGs/KOGs that were shown to be also present in rotifers. For the comparison only COGs/KOGs were used that belong to the functional groups [L]: replication, recombination and repair, [V]: defense mechanisms or [O]: posttranslational modification, protein turnover, chaperones. The analyzed COGs/KOGs with descriptions can be found in Additional File 1, S-Table VII.

categories a number of different pathways specifically boost repair and protection of proteins and DNA. In our opinion, it is clearly this multitude and overlap of different proteins that protect *M. tardigradum* and allow its extreme stress tolerance in the tun stage. All human pathways in Figure 9 (right), are found to be shared with tardigrades but not found in the other organisms compared here. These are potential pharmacological targets to enhance stress resistance in human cells (eg, tissue conservation, cryoprotection, surgical interventions).

Discussion

Tardigrades are an animal phylum with a surprising stress-resistance and adaptation potential, using phylum-specific pathways as well as enhancing general pathways to increase stress resistance. As a first step towards a comprehensive transcriptome and full genome analysis this study assesses currently available transcriptome information to probe the tardigrade adaptation potential and addresses a number of specific questions already answerable from the

Figure 7. Comparison of COGs/KOGs specific for *Milnesium tardigradum, Hypsibius dujardini* (both tardigrades) and *Saccharomyces cerevisiae and Saccharomyces pombe* (both yeast).
Notes: The Venn diagram compares the specific COGs/KOGs only found in tardigrades to unicellular clusters of the yeasts *S. cerevisiae* and *S. pombe*. For the comparison only COGs/KOGs were used that belong to the functional groups [L]: replication, recombination and repair, [V]: defence mechanisms or [O]: posttranslational modification, protein turnover, chaperones. The COGs/KOGs with descriptions can be found in Additional File 1, S-Table VII.

current set of EST data. These data cover a substantial fraction of the whole transcriptome with mainly highly expressed ESTs, but also some of the poorly expressed ESTs.

Furthermore, the Tardigrade Workbench was updated with new management options and software added to enhance comparative effort on tardigrade species.

The current tardigrade transcriptome is put into its pathway context looking at all implied levels including RNA motifs, encoded proteins, pathways and implied phenotypic metabolism, signalling and stress protection.

Tardigrade adaptation is strong but there are significant differences between the compared species. The comparison between *H. dujardini* and the particular stress-resistant *M. tardigradum* suggests that *M. tardigradum* disposes of larger protein families for stress-resistance. Importantly, it invests more in repair strategy. This allows recovery from more extensive stress in addition to the direct protective strategy (*H. dujardini* but also *M. tardigradum*).

Comparisons between COGs from important stress tolerant animal groups delineate shared and common COGs between tardigrades and some model organisms (*D. melanogaster*, *C. elegans*). Furthermore, we show that some stress protective pathways from tardigrades occur in human cells although they were not found in the other organisms.

Table 5. Tardigrade specific cluster domain annotation.

| C.	Domain annotation (Pfam)	Loc/Cov	E value	Description
8	Osteopontin	34..89 22.44%	8e-06	Also known as bone sialoprotein, pleiotropic cytokine, enhances uPA secretion, cell motility, and extracellular matrix invasion. Cytosceletal proteins (I; M.t.)
18	Lipolysis stimulated receptor (LSR)	5..49 63.76%	9.6e-06	Fatty acid metabolism stress response; Fatty acid metabolism (I; M.t.)
19	Sulfite exporter TauE/SafE (TauE)	34..91 55.88%	5.4e-06	Taurin strong antioxidant – further taurin metabolism not identified; intrinsic feature: Glycine-rich protein (as TSP 51) Strong redox protective pathways (I; M.t.)
37	Signal peptidase subunit (SPC22)	37..64 28.42%	5.7e-07	Protein secretion, transport; high protein turnover in tardigrades (I; M.t.)
51	Glycine rich protein family (GRP)	39..103 61.53%	1.8e-10	Stress response; regulation by hsp-proteins (R; M.t.)
55	Domain of unknown function (DUF19)	65..213 63.24%	1e-08	NO-Synthase (R; M.t.); Other Interactors
	Chitinase binding			Chitinase (I; M.t.)

Notes: Shown are the encoded and identified domains for annotated tardigrade-specific protein clusters (TSPs). Six tardi-grade specific clusters could be partly annotated and are shown. Left: Cluster numbering is according to Förster et al;[25] TSPs were detected in *Milnesium tardigradum, Richtersius coronifer* and *Hypsibius dujardini*; "new" indicates a new recently identified cluster in *M. tardigradum* from the ongoing EST sequencing effort. regulatory proteins known from other organisms given as detected in *M. tardigradum*: TSP8 Late nodulin protein, TSP18, Iron permease FTR1 family, TSP 19 Tau E sulfite exporter and TSP37 (signal peptidase subunit), TSP51 (Grp domain, small heat shock proteins, hsp 60,90, 17), TSP55 (DUF19 domain; could be annotated to be part of the tetrahyrdopterin synthesis). Right: Known regulators (R) or interactors (I) and their absence (-) or presence (M.t.) in *M. tardigradum* according to the EST data. Further example sequences can be directly accessed from our server, http://waterbear.bioapps.biozentrum.uni-wuerzburg.de.

The function of tardigrade-specific clusters is investigated by analysis of conserved domains and validation of implicated pathways by looking for ESTs encoding further proteins involved in these pathways.

For two protein families (LEA proteins, heat shock proteins), there is also complementary experimental evidence from PCR and Western blotting. By direct comparison of the different clusters, the predicted proteins shared by both species could be identified. Many protective strategies are found and combined in tardigrades. Interestingly, we see clearly in *M. tardigradum* that the protection is not perfect but rather a high degree of damage is tolerated (eg, DNA lesions and oxidation). This damage is then subsequently repaired by a sophisticated combination of pathways, including several tardigrade-specific sequence clusters and pathways. The same is true in protein damage response, controlled by specific protein turnover pathways and boosted by oxidative damage repair as well as heat shock proteins. Both EST libraries are independent efforts and their accessions deposited with GenBank. The EST library of *H. dujardini* was generated in the laboratory of M. Blaxter (Institute of Evolutionary Biology, University of Edinburgh).

The *M. tardigradum* transcriptome is newly generated as a part of a general effort (www.funcrypta. de) to understand the adaptation and physiology of *M. tardigradum*. The aim of this study is to analyze the different levels of the transcriptome, starting from RNA motifs while focussing on different types of sequence clusters and pathways as well as their verification by experiments and further comparisons. Other studies had already looking into other aspects of the adaptation process, such as differential EST libraries between tun and active state in *M. tardigradum* or changes in specific metabolites during adaptation as well as comparison to further tardigrade adaptations.

Our study is a conservative estimate of the differently encoded protein clusters and pathways analyzed. With a full *M. tardigradum* genome sequence and full transcriptome data the description will become more complete. The available data of 11548 EST sequences from *M. tardigradum* and 5235 EST sequences from *H. dujardini* show a number of interesting features in these two different tardigrades. By systematic sequence to sequence comparison it was found that a substantial fraction of the sequences (12.2% for all sequences no matter from which of the two compared tardigrades) comes in sequence-similar families involved
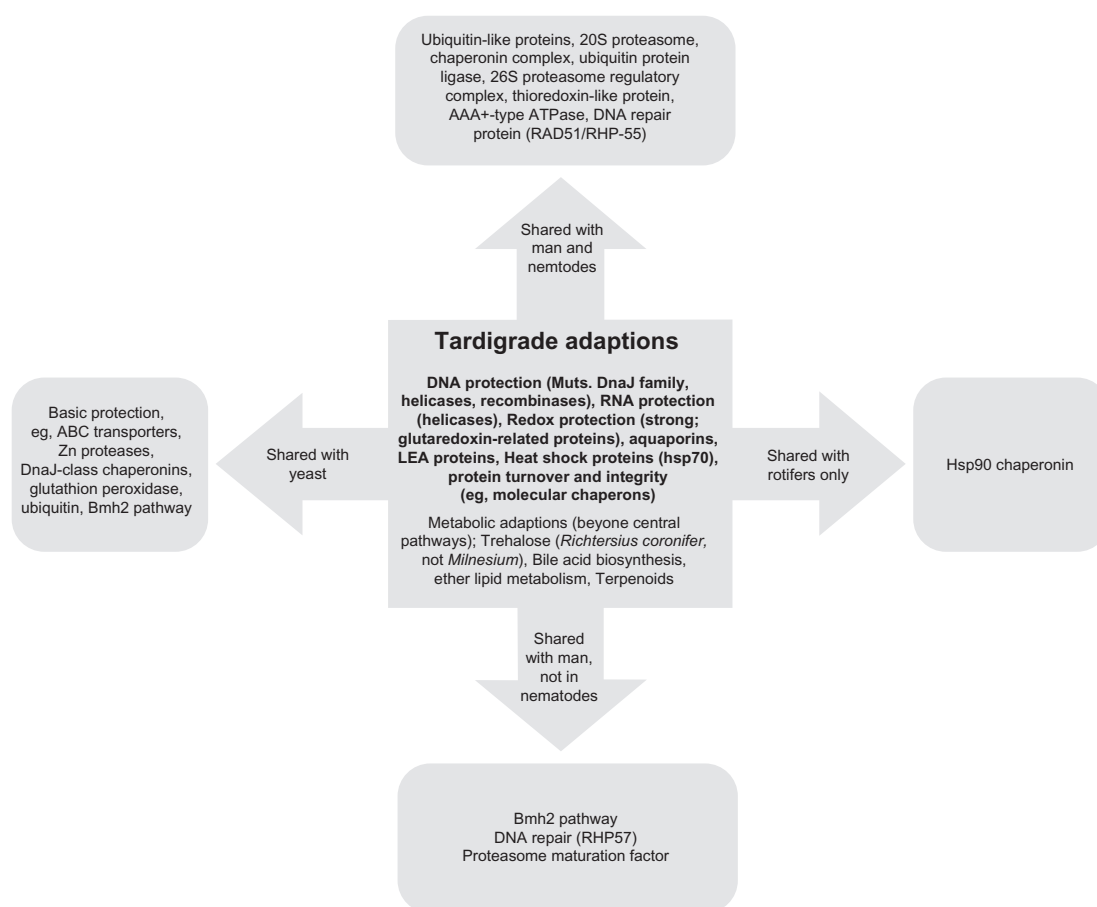
Figure 8. Tardigrade stress adaptation: Which are specific, which are shared?
Notes: Depicted are predicted tardigrade specific protein families in the center of the diagram and protein families shared with other organisms. Top: Man and nematodes; left: yeast (*Saccharomyces pombe, Saccharomyces cerevisiae*); right: rotifers; bottom: man, not in nematodes (14-3-3-like proteins occur in *Caenorhabditis elegans*, but their relation to the bmh2 pathway is not clear).

in protein repair and protection, redox protection, protein turnover, signalling as well as tardigrade-specific unique clusters (Fig. 1, bottom). Interestingly, CLANS clustering of all sequence-similar proteins shows this holds for both *M. tardigradum* (6.2%, 13 clusters) and *H. dujardini* (10.5%, 16 clusters). However, there is stronger stress adaptation potential in *M. tardigradum*.

This general trend is supported by further analyses reported in this paper. Each analysis revealed different specific sequence families and pathways involved in tardigrade-specific adaptations: The predicted metabolic inventory of both tardigrades as well as specific regulatory motifs in *M. tardigradum* mRNA were compared. A number of specific motifs are used to regulate mRNA in *M. tardigradum*. The key mRNA motifs found (lox P DICE,[42] k-box,[75] and brd-box[43]) are involved in stability regulation. Standard mRNA motifs known from vertebrates

were avoided such as the AU-rich element inducing mRNA instability.[45]

Specific metabolic pathways include central primary metabolism and adaptations in lipid and carbohydrate metabolism. Furthermore, there are metabolite connections to stress defence, for instance regarding storage carbohydrates such as glycogen and lipid metabolism to foster membrane protection.

Moreover, a number of specific pathways involved in stress adaptation were identified—including some that also occur in vertebrates. Several major protein families were assembled to pathways in *M. tardigradum* that are implicated in its high resistance to extreme temperatures or vacuum in its tun stage.[7,19,21,23,76] This includes several stress-related clusters of sequence similar proteins, COGs/KOGs as well as metabolic pathways. While there is generally high overlap in the latter, the other two comparisons show strong protection and repair pathways for
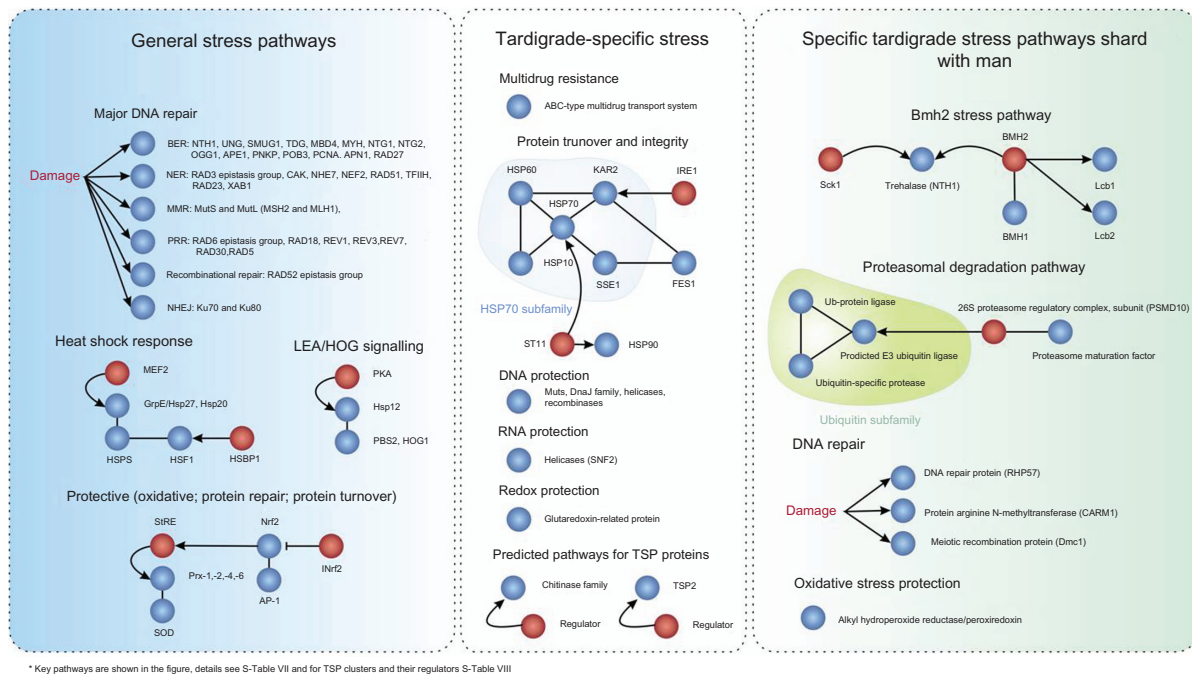
Figure 9. Tardigrade stress adaptation: General and unique adaptations as well as adaptations shared with man.
Notes: Key pathways are shown in the figure, details see S-Table VII and for TSP clusters and their regulators S-Table VIII. Depicted are all categories of different pathways either to more general stress protection (left), unique for tardigrades (middle) or shared with man (right). In this comparison, unique refers to *Milnesium tardigradum* proteins not found in rotifers/nematodes/*Drosophila*/man (middle, top). The term "shared only with man" (right panel) indicates in this comparison tardigrade protein clusters shared only with *Homo sapiens* and not with nematodes, rotifers or yeast. Furthermore, the TSP (tardigrade specific protein clusters) are sequences with no homology outside tardigrades (middle, bottom). Here via domain annotation a tentative function assignment is possible. The categories in the figure try to cover all what is known, protein examples are only key examples. Detailed data are given in Additional File 1, S-Table VII (COG distribution among phyla) and S-Table VIII (TSP tentative domain annotation). Colours: blue, stress protein; red, regulator.

proteins and against redox stress in both organisms but a number of species-specific families not found in the other. In accordance with observations from physiology, *M. tardigradum* has even more specific protection and repair pathways than *H. dujardini* and, accordingly, more protein families are involved.

This includes specific protein families involved in DNA repair.[77] A recent study highlighted the high DNA repair potential in *M. tardigradum* in comet assays[64] but the machinery behind the high repair potential and the numerous DNA breaks endured during tun stage was not clear. The different DNA repair protein pathways analyzed here and compared with *H. dujardini* better explain the protein families involved, in particular of the DnaJ type. The high repair potential is supported by a comparison of protective protein clusters of tardigrades according to the COG classification system[38] to nematodes, man, and yeast as a control. We identify 46 tardigrade-specific protein adaptations including the chaperones also involved in DNA repair.[78] Physiological experiments of several cycles of remitting tardigrades to tun stage and reactivating them show that more than seven cycles, or

a too rapid cycling between stages, is generally not tolerated. These experiments provide additional support that repair processes require time and dominate over direct protection in tardigrades.

Domain annotated tardigrade-specific EST and clusters of related proteins,[25] together with further interacting or regulatory proteins verified to be present in *M. tardigradum,* allow a first look into unique stress-protection pathways in *M. tardigradum*. This includes the chitinase and binding domains for cell wall metabolism, redox protection, NO-synthase and ecdysone receptor as well as glycine-rich proteins.

More than ten protein clusters are actually shared with pathways and protein clusters in humans. These include bmh2 (Figs. 5 and 8), proteasome maturation factors and others. The maturation factor could in principle also be manipulated in man to improve stress tolerance, as proteasome inhibitory drugs are used in cancer therapy.[79,80] However, to improve stress tolerance a suitable activation level is critical. DNA repair protein RHP57 (RAD57 homolog of the fission yeast *Schizosaccharomyces pombe*; Fig. 9) is another conserved and attractive target to improve stress

tolerance in man. It is a recombination repair gene[66] and shown here to be present both in tardigrades and human. DNA stability and repair is considered to be a key factor in cancer[81] From the tardigrade TSP-annotation, though only domain-specific and tentative, several of the identified pathways occur also in man and generally enhance stress-tolerance. These include osteopontin, which in man influences AKT, eNOS and MAP kinases; fatty acid metabolism; taurin metabolism; and finally protein secretion including spc22/23. Taken together, these results identify several similarities between stress pathways in tardigrades and humans. Human stress resistance may be enhanced by influencing key stress pathways homologous to tardigrades by suitable means. Furthermore, besides active protection (stable proteins and DNA; antifreeze proteins), we found various protein and DNA repair pathways. Including ubiquitin-based pathways to remove damaged proteins, repair oxidized and modified proteins as well as sophisticated translational machinery (specific ribosomal proteins).

The significance of the study is two-fold: Firstly, the conservation of tardigrade stress pathways in other organisms, in particular in humans, opens up the possibility to boost stress resistance in these species as well. This is currently explored using peptides generated from tardigrade sequences in cell culture. In principle, further protection is possible by identifying which of the interconnected protection and repair pathways are conserved in human cells and using pharmacological means to activate receptors and regulators of these conserved pathways. However, this will require extensive pharmacological testing.

Secondly, we can now more specifically show strong stress protective pathways in tardigrades. Particularly for *M. tardigradum* there is enhanced protection by a network of protective and repair pathways. This includes a number of tardigrade-specific protein clusters. We show here for the first time that a number of these pathways boost stress protection, at least for those where molecular partial annotation was possible by conserved domains. This was further validated by additional ESTs representing interacting proteins for the predicted pathways.

Conclusions

Starting from an ongoing EST sequencing effort, we first give direct access to the most recent EST sequences via the tardigrade analyzer and then investigate on different levels the implied functions for RNA motifs, encoded protein functions, stress pathways and metabolic networks. Protein clusters are identified that are tardigrade specific (*M. tardigradum, H. dujardini*). Others are shared with distantly related organisms (eg, *C. elegans, D. melanogaster*) and may even boost stress resistance in human cells. There is a high potential for stress protection and even more for repair in tardigrades. *M. tardigradum* excels here by a combination of protection and repair (heat-shock, redox protection, protein and DNA repair).

Description of Additional Files

The following additional data are available with the online version of this paper.

Additional data file 1 gives detailed data on all comparisons made: CLANS clustering of ESTs (S-Table I), COGs/KOGs identified (S-Table II), shared sequences according to Inparanoid (S-Table III), Encoded metabolic pathways (S-Table IV) and stress pathways (S-Table V), bmh2 homolog sequence in *M. tardigradum* (S-Table VI), COGs/KOGs compared to different phyla (S-Table VII) and annotation of tardigrade-specific ESTs (S-Table VIII).

Additional data file 2 is a Table listing all ESTs analyzed in the paper with their GenBank accession number.

Author Contributions

All authors participated in data analysis. FF and MAG assembled ESTs, supervised by MF and TD. FF, DB, MAG, CL and TD conducted different major bioinformatics sequence analyses and annotations. JE did metabolic pathway analyses. EH, MS provided proteomics analyses, ROS provided physiological analyses and TM did statistical analyses. TD led and guided the study. All authors were involved in writing the manuscript and approved its final version.

Acknowledgements

Competing Interests

The authors declare that there are no competing interests.

Funding

Disclosures and Ethics

As a requirement of publication the corresponding author TD confirms to the publisher signed confirmation of compliance with legal and ethical obligations including but not limited to the following: authorship and contributorship, conflicts of interest, privacy and confidentiality. The authors have read and confirmed their agreement with the ICMJE authorship and conflict of interest criteria. The authors have also confirmed that this article is unique and not under consideration or published in any other publication, and that they have permission from rights holders to reproduce any copyrighted material.

References

1. Lubzens E, Cerda J, Clark M, editors. *Dormancy and Resistance in Harsh Environments*. 2010: Springer; New York.
2. Clegg JS. Cryptobiosis—a peculiar state of biological organization. *Comp Biochem Physiol B Biochem Mol Biol*. 2001;128(4):613–24.
3. Crowe JH, Clegg JS, American Institute of Biological Sciences. *Dry biological systems : proceedings of the 1977 symposium of the American Institute of Biological Sciences held in East Lansing, Michigan*. 1978: New York: Academic Press. x, 357 p.
4. Crowe LM. Lessons from nature: the role of sugars in anhydrobiosis. *Comp Biochem Physiol A Mol Integr Physiol*. 2002;131(3):505–13.
5. Hoekstra FA, Wolkers WF, Buitink J, et al. Membrane stabilization in the dry state. *Comparative Biochemistry and Physiology a-Physiology*. 1997;117(3): 335–41.
6. Westh P, Ramlov H. Trehalose Accumulation in the Tardigrade Adorybiotus coronifer during Anhydrobiosis. *Journal of Experimental Zoology*. 1991; 258(3):303–11.
7. Hengherr S, Heyer AG, Köhler HR, Schill RO. Trehalose and anhydrobiosis in tardigrades—evidence for divergence in responses to dehydration. *Febs Journal*. 2008;275(2):281–8.
8. Clegg JS. Desiccation tolerance in encysted embryos of the animal extremophile, Artemia. *Integrative and Comparative Biology*. 2005;45(5): 715–24.
9. Goyal K, Walton LJ, Browne JA, Burnell AM, Tunnacliffe A. Molecular anhydrobiology: Identifying molecules implicated in invertebrate anhydrobiosis. *Integrative and Comparative Biology*. 2005;45(5):702–9.
10. Ellis RJ, Vandervies SM. Molecular Chaperones. *Annual Review of Biochemistry*. 1991;60:321–47.
11. Feder ME, Hofmann GE. Heat-shock proteins, molecular chaperones, and the stress response: Evolutionary and ecological physiology. *Annual Review of Physiology*. 1999;61:243–82.
12. Gething MJ, Sambrook J. Protein Folding in the Cell. *Nature*. 1992; 355(6355):33–45.
13. Kiang JG, Tsokos GC. Heat shock protein 70 kDa: Molecular biology, biochemistry, and physiology. *Pharmacology and Therapeutics*. 1998;80(2): 183–201.
14. Browne J, Tunnacliffe A, Burnell A. Anhydrobiosis—Plant desiccation gene found in a nematode. *Nature*. 2002;416(6876):38–8.
15. Jacquiersarlin MR, Fuller K, Dinhxuan AT, Richard MJ, Polla BS. Protective Effects of Hsp70 in Inflammation. *Experientia*. 1994;50(11–12): 1031–8.
16. Plumier JCL, Ross BM, Currie RW, et al. Transgenic Mice Expressing the Human Heat-Shock Protein-70 Have Improved Postischemic Myocardial Recovery. *Journal of Clinical Investigation*. 1995;95(4):1854–60.
17. Reuner A, Hengherr S, Mali B, et al. Stress response in tardigrades: differential gene expression of molecular chaperones. *Cell Stress Chaperones*. 2010;15(4):423–30.
18. Schill RO, Steinbrück GHB, Köhler HR. Stress gene (hsp70) sequences and quantitative expression in Milnesium tardigradum (Tardigrada) during active and cryptobiotic stages. *Journal of Experimental Biology*. 2004; 207(10):1607–13.
19. 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(4):456–60.
20. McGee BK, Schill RO, Tunnacliffe A. Hydrophilic proteins in invertebrate anhydrobiosis. *Integrative and Comparative Biology*. 2004;44(6): 679–9.
21. 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. *Journal of Experimental Biology*. 2009;212(6):802–7.
22. Horikawa DD, Sakashita T, Katagiri C, et al. Radiation tolerance in the tardigrade Milnesium tardigradum. *International Journal of Radiation Biology*. 2006;82(12):843–8.
23. Rebecchi L, Altiero T, Guidetti R, et al. Tardigrade Resistance to Space Effects: First Results of Experiments on the LIFE-TARSE Mission on FOTON-M3 (September 2007). *Astrobiology*. 2009;9(6):581–91.
24. Gabriel WN, Goldstein B. Segmental expression of Pax3/7 and Engrailed homologs in tardigrade development. *Development Genes and Evolution*. 2007;217(6):421–33.
25. Förster F, Liang C, Shkumatov A, et al. Tardigrade workbench: comparing stress-related proteins, sequence-similar and functional protein clusters as well as RNA elements in tardi grades. *BMC Genomics*. 2009:10.
26. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. Gen Bank. *Nucleic Acids Res*. 2009;37(Database issue):D26–31.
27. Altschul SF, Madden TL, Schaffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997;25(17):3389–402.
28. Mignone F, Grillo G, Licciulli F, et al. UTRdb and UTRsite: a collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs. *Nucleic Acids Res*. 2005;33(Database issue):D141–6.
29. Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol*. 2000;132:365–86.
30. Gaudermann P, Vogl I, Zientz E, et al. Analysis of and function predictions for previously conserved hypothetical or putative proteins in Blochmannia floridanus. *BMC Microbiology*. 2006:6.
31. Finn RD, Mistry J, Tate J, et al. The Pfam protein families database. *Nucleic Acids Res*. 2010;38(Database issue):D211–22.
32. Hunter S, Apweiler R, Attwood TK, et al. Inter Pro: the integrative protein signature database. *Nucleic Acids Res*. 2009;37(Database issue): D211–5.
33. Letunic I, Doerks T, Bork P. SMART 6: recent updates and new developments. *Nucleic Acids Res*. 2009;37(Database issue):D229–32.
34. Schultz J, Milpetz F, Bork P, Ponting CP. SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci U S A*. 1998;95(11):5857–64.
35. Frickey T, Lupas A. CLANS: a Java application for visualizing protein families based on pairwise similarity. *Bio informatics*. 2004;20(18): 3702–4.

36. Remm M, Storm CE, Sonnhammer EL. Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J Mol Biol*. 2001;314(5): 1041–52.

37. Bork P, Gibson TJ. Applying motif and profile searches. *Methods Enzymol*. 1996;266:162–84.

38. Tatusov RL, Koonin EV, Lipman DJ. A genomic perspective on protein families. *Science*. 1997;278(5338):631–7.

39. Tatusov RL, Fedorova ND, Jackson JD, et al. The COG database: an updated version includes eukaryotes. *BMC Bio informatics*. 2003;4:41.

40. Wheeler DL, Barrett T, Benson DA, et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*. 2007;35: D5–12.

41. Larkin MA, Blackshields G, Brown NP, et al. Clustal W and Clustal X version 2.0. *Bio informatics*. 2007;23(21):2947–8.

42. 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 Journal*. 1994;13(6): 1476–81.

43. Lai EC. Micro RNAs are complementary to 3′ UTR sequence motifs that mediate negative post-transcriptional regulation. *Nature Genetics*. 2002; 30(4):363–4.

44. Parsch J, Russell JA, Beerman I, Hartl DL, Stephan W. Deletion of a conserved regulatory element in the Drosophila Adh gene leads to increased alcohol dehydrogenase activity but also delays development. *Genetics*. 2000;156(1):219–27.

45. Chen CY, Shyu AB. AU-rich elements: characterization and importance in mRNA degradation. *Trends Biochem Sci*. 1995;20(11):465–70.

46. 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.

47. Söllner TH. Regulated exocytosis and SNARE function (Review). *Mol Membr Biol*. 2003;20(3):209–20.

48. Tunnacliffe A, Wise MJ. The continuing conundrum of the LEA proteins. *Naturwissenschaften*. 2007;94(10):791–812.

49. Gavin AC, Aloy P, Grandi P, et al. Proteome survey reveals modularity of the yeast cell machinery. *Nature*. 2006;440(7084):631–6.

50. Soto T, Fernandez J, Cansado J, Vicente Soler J, Gacto M. Protein kinase Sck1 is involved in trehalase activation by glucose and nitrogen source in the fission yeast Schizosaccharomyces pombe. *Micro Biology UK*. 1997;143: 2457–63.

51. Schokraie E, Hotz-Wagenblatt A, Warnken U, et al. Proteomic Analysis of Tardigrades: Towards a Better Understanding of Molecular Mechanisms by Anhydrobiotic Organisms. *PLoS ONE*. 2010:5(3).

52. Alpert P. Constraints of tolerance: why are desiccation-tolerant organisms so small or rare? *Journal of Experimental Biology*. 2006;209(9):1575–84.

53. Melgar JC, Guidi L, Remorini D, et al. Antioxidant defences and oxidative damage in salt-treated olive plants under contrasting sunlight irradiance. *Tree Physiol*. 2009;29(9):1187–98.

54. Sanchez-Estudillo L, Freile-Pelegrin Y, Rivera-Madrid R, Robledo D, Narvaez-Zapata JA. Regulation of two photosynthetic pigment-related genes during stress-induced pigment formation in the green alga, Dunaliella salina. *Bio technol Lett*. 2006;28(11):787–91.

55. Ursic D, Chinchilla K, Finkel JS, Culbertson MR. Multiple protein/protein and protein/RNA interactions suggest roles for yeast DNA/RNA helicase Sen1p in transcription, transcription-coupled DNA repair and RNA processing. *Nucleic Acids Res*. 2004;32(8):2441–52.

56. Loprasert S, Vattanaviboon P, Praituan W, Chamnongpol S, Mongkolsuk S. Regulation of the oxidative stress protective enzymes, catalase and superoxide dismutase in Xanthomonas—A review. *Gene*. 1996;179(1):33–7.

57. Liu Y, Zheng YZ. PM2, a group 3 LEA protein from soybean, and its 22-mer repeating region confer salt tolerance in Escherichia coli. *Biochemical and Biophysical Research Communications*. 2005;331(1):325–32.

58. Shao HB, Liang ZS, Shao MA. LEA proteins in higher plants: Structure, function, gene expression and regulation. *Colloids and Surfaces B-Biointerfaces*. 2005;45(3–4):131–5.

59. Morimoto RI, Kline MP, Bimston DN, Cotto JJ. The heat-shock response: regulation and function of heat-shock proteins and molecular chaperones. *Essays in Biochemistry*. 1997;32:17–29.

60. Tessmer I, Yang Y, Zhai J, et al. Mechanism of MutS Searching for DNA Mismatches and Signaling Repair. *Journal of Biological Chemistry*. 2008; 283(52):36646–54.

61. Förster F, Beisser D, Frohme M, Schill RO, Dandekar T. Bioinformatics identifies tardigrade molecular adaptations including the DNA-j family and first steps towards dynamical modelling. *Journal of Zoological Systematics and Evolutionary Research*. 2011;49:120–6.

62. Schüller C, Brewster JL, Alexander MR, Gustin MC, Ruis H. The hog pathway controls osmotic regulation of transcription via the stress-response element (Stre) of the Saccharomyces cerevisiae Ctt1 Gene. *Embo Journal*. 1994;13(18):4382–9.

63. Orrweaver TL, Szostak JW, Rothstein RJ. Yeast Transformation—a Model System for the Study of Recombination. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences*. 1981;78(10):6354–58.

64. Neumann S, Reuner A, Brümmer F, Schill RO. DNA damage in storage cells of anhydrobiotic tardigrades. *Comparative Biochemistry and Physiology a-Molecular and Integrative Physiology*. 2009;153(4):425–9.

65. Schill RO, Neumann S, Reuner A, Brümmer F. Detection of DNA damage with single-cell gel electrophoresis in anhydrobiotic tardigrades. *Comparative Biochemistry and Physiology a-Molecular and Integrative Physiology*. 2008;151(1):S32–2.

66. Tsutsui Y, Morishita T, Iwasaki H, Toh H, Shinagawa H. A recombination repair gene of Schizosaccharomyces pombe, rhp57, is a functional homolog of the Saccharomyces cerevisiae RAD57 gene and is phylogenetically related to the human XRCC3 gene. *Genetics*. 2000;154(4): 1451–61.

67. Fricke B, Heink S, Steffen J, Kloetzel PM, Krueger E. The proteasome maturation protein POMP facilitates major steps of 20S proteasome formation at the endoplasmic reticulum. *Embo Reports*. 2007;8(12):1170–75.

68. Nathan DF, Vos MH, Lindquist S. In vivo functions of the Saccharomyces cerevisiae Hsp90 chaperone. *Proceedings of the National Academy of Sciences of the United States of America*. 1997;94(24):12949–56.

69. Zhao RM, Davey M, Hsu YC, et al. Navigating the chaperone network: An integrative map of physical and genetic interactions mediated by the Hsp90 chaperone. *Cell*. 2005;120(5):715–27.

70. Szklarczyk D, Franceschini A, Kuhn M, et al. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res*. 2011;39:D561–8.

71. Linscheid P, Schaffner A, Blau N, Schoedon G. Regulation of 6-pyruvoyltetrahydropterin synthase activity and messenger RNA abundance in human vascular endothelial cells. *Circulation*. 1998;98(17):1703–6.

72. Mali B, Grohme MA, Förster F, et al. Transcriptome survey of the anhydrobiotic tardigrade Milnesium tardigradum in comparison with Hypsibius dujardini and Richtersius coronifer. *BMC Genomics*. 2010:11.

73. Funkhouser JD, Aronson NN. Chitinase family GH18: evolutionary insights from the genomic history of a diverse protein family. *BMC Evolutionary Biology*. 2007:7.

74. Merzendorfer H, Zimoch L. Chitin metabolism in insects: structure, function and regulation of chitin synthases and chitinases. *Journal of Experimental Biology*. 2003;206(24):4393–412.

75. Lai EC, Tam B, Rubin GM. Pervasive regulation of Drosophila Notch target genes by GY-box-, Brd-box-, and K-box-class microRNAs. *Genes and Development*. 2005;19(9):1067–80.

76. Jönsson KI, Rabbow E, Schill RO, Harms-Ringdahl M, Rettberg P. Tardigrades survive exposure to space in low Earth orbit. *Current Biology*. 2008;18(17):R729–31.

77. Setlow P. I will survive: DNA protection in bacterial spores. *Trends in Microbiology*. 2007;15(4):172–80.

78. Yamamoto T, Mori Y, Ishibashi T, et al. Interaction between proliferating cell nuclear antigen (PCNA) and a DnaJ induced by DNA damage. *Journal of Plant Research*. 2005;118(2):91–7.

79. McConkey DJ, Zhu K. Mechanisms of proteasome inhibitor action and resistance in cancer. *Drug Resist Updat*. 2008;1(4–5):164–79.

80. Yang S, Cui L. *The Action of Aquaporins in Cell Elongation, Salt Stress and Photosynthesis*. Sheng Wu Gong Cheng Xue Bao. 2009;25(3):321–7.

81. Arking R. *The Biology of Aging: Observations and Principles*. 3rd ed. 2006, Oxford, New York. Oxford University Press; xiv, 604 p.

82. Zhou Y, Dai DL, Martinka M, et al. Osteopontin expression correlates with melanoma invasion. *J Invest Dermatol*. 2005;124(5):1044–52.

Supplementary Data Informations

Additional File 1: Additional File 1.doc, word-file with details on all comparisons made.

Additional File 2: Additional File 2.pdf, pdf file listing all ESTs analyzed and their GenBank accession number.