BMC
Genomics

RESEARCH ARTICLE

Open Access

Transcriptome survey of the anhydrobiotic tardigrade *Milnesium tardigradum* in comparison with *Hypsibius dujardini* and *Richtersius coronifer*

Brahim Mali[1*], Markus A Grohme[1], Frank Förster[2], Thomas Dandekar[2], Martina Schnölzer[3], Dirk Reuter[4], Weronika Wełnicz[1], Ralph O Schill[5], Marcus Frohme[1]

Abstract

Background: The phenomenon of desiccation tolerance, also called anhydrobiosis, involves the ability of an organism to survive the loss of almost all cellular water without sustaining irreversible damage. Although there are several physiological, morphological and ecological studies on tardigrades, only limited DNA sequence information is available. Therefore, we explored the transcriptome in the active and anhydrobiotic state of the tardigrade *Milnesium tardigradum* which has extraordinary tolerance to desiccation and freezing. In this study, we present the first overview of the transcriptome of *M. tardigradum* and its response to desiccation and discuss potential parallels to stress responses in other organisms.

Results: We sequenced a total of 9984 expressed sequence tags (ESTs) from two cDNA libraries from the eutardigrade *M. tardigradum* in its active and inactive, anhydrobiotic (tun) stage. Assembly of these ESTs resulted in 3283 putative unique transcripts, whereof ~50% showed significant sequence similarity to known genes. The resulting unigenes were functionally annotated using the Gene Ontology (GO) vocabulary. A GO term enrichment analysis revealed several GOs that were significantly underrepresented in the inactive stage. Furthermore we compared the putative unigenes of *M. tardigradum* with ESTs from two other eutardigrade species that are available from public sequence databases, namely *Richtersius coronifer* and *Hypsibius dujardini*. The processed sequences of the three tardigrade species revealed similar functional content and the *M. tardigradum* dataset contained additional sequences from tardigrades not present in the other two.

Conclusions: This study describes novel sequence data from the tardigrade *M. tardigradum*, which significantly contributes to the available tardigrade sequence data and will help to establish this extraordinary tardigrade as a model for studying anhydrobiosis. Functional comparison of active and anhydrobiotic tardigrades revealed a differential distribution of Gene Ontology terms associated with chromatin structure and the translation machinery, which are underrepresented in the inactive animals. These findings imply a widespread metabolic response of the animals on dehydration. The collective tardigrade transcriptome data will serve as a reference for further studies and support the identification and characterization of genes involved in the anhydrobiotic response.

Background

Desiccation tolerance or anhydrobiosis is the ability of an organism to survive almost complete drying without sustaining damage. Anhydrobiosis is observed in certain micro-organisms, plants and animals such as rotifers, brine shrimp cysts, tardigrades and insect larvae,

as for example those of the *Polypedilum vanderplanki* [1-3]. Studying the mechanisms of tolerance against desiccation may lead to development of new methods for preserving biological materials, which is of enormous practical importance in industrial as well as in medical fields [4]. In the dry state, the metabolism is suspended and the duration that anhydrobiotic organisms can survive ranges from years to centuries. Tardigrades are able to survive long periods of desiccation [5-8]. The hitherto longest known observation of an

* Correspondence: bmali@tfh-wildau.de
[1]Molecular Biology and Functional Genomics, University of Applied Sciences Wildau, Bahnhofstraße 1, 15745 Wildau, Germany

BioMed Central

extended lifespan of 20 years has been demonstrated in the anhydrobiotic state of the species *Echiniscus testudo* Doyère 1840 [9]. Anhydrobiosis probably depends on a series of complex morphological, physiological and genetic adaptations that involve the stabilization of macromolecular complexes. As a consequence, a number of components have been identified and appear to be important for protecting these organisms from desiccation damage. Among them are the highly hydrophilic LEA proteins, which have been initially described in plants but have been identified in several invertebrates [10,11,2,12], as well as non-reducing disaccharides like trehalose [13-16]. We are studying anhydrobiosis in the limno-terrestrial tardigrade *Milnesium tardigradum* Doyère 1840 which shows remarkable resistance to adverse environmental conditions in all stages of life [17] - even to extreme levels of ionizing radiation [18] and the vacuum of space in low earth orbit [19]. *M. tardigradum* outperforms several other tardigrade species in tolerance e.g. survival of extreme temperatures above 100°C [20] as well as freezing [21,22]. Similar anhydrobiotic resistance to extreme environmental stress has been observed in other animals such as bdelloid rotifers or chironomid larvae [23,24] suggesting common mechanisms that allow anhydrobiotic survival and conferring radiation tolerance. The tardigrade phylum currently includes more than 1000 species living in the sea, in fresh water and on land. These last, needing at least a film of water to be active, are called limno-terrestrial and include most of the anhydrobiotic species [25]. They have been studied for their fascinating ability to perform anhydrobiosis and consequently serve as a potential model for studying tolerance and survival of multicellular organisms to a variety of extreme environmental conditions. Although there are several physiological, morphological and ecological studies on anhydrobiotic tardigrades [26-30], only limited DNA sequence information from molecular phylogenetic studies is available [30-34]. However, some sequence resources are only available from the species *Hypsibius dujardini* Doyère 1840 [Daub et al. Unpublished data 2003] and *Richtersius coronifer* (Richters 1903) [35]. Studies of *H. dujardini* have been focused mainly on developmental and evolutionary biology [36-38]. In this study we generated 9984 ESTs of *M. tardigradum* from active and inactive (anhydrobiotic/tun) stages, thereby establishing *M. tardigradum* as a model for anhydrobiosis research. These ESTs and the resulting unigenes were functionally annotated using Gene Ontology vocabulary. Furthermore, a cross-species comparison between *M. tardigradum*, *H. dujardini* and *R. coronifer* has been performed.

Results and discussion

cDNA libraries and sequence datasets

We have generated two directionally cloned cDNA libraries from active and inactive stages of *M. tardigradum* and subjected them to single pass Sanger sequencing. Furthermore we retrieved two EST datasets from public sequence databases (see Table 1 and 2). The datasets used in this study consisted of EST sequences from, *M. tardigradum* active and inactive stages, *H. dujardini* [Daub et al. Unpublished data 2003] and *R. coronifer* [35] of which the latter two were retrieved from NCBI (National Center for Biotechnology Information) dbEST and the NCBI Trace Archive. The source of all tardigrade samples consisted of whole adult animals except for the *H. dujardini* sample where adults and juveniles had been pooled.
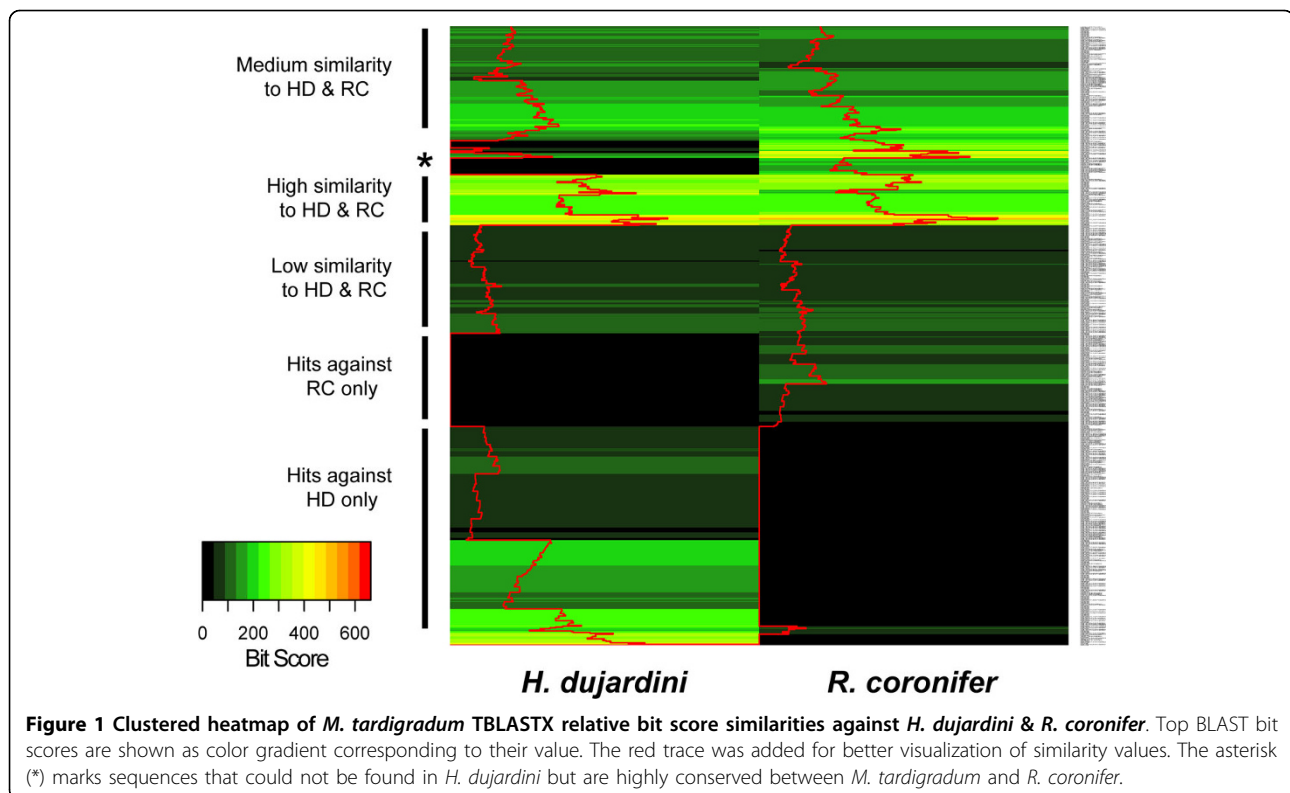
Analysis of the *M. tardigradum* cDNA library

As summarized in Table 1, a total of 9979 clones were sequenced from the *M. tardigradum* library generated from two stages, active and inactive, in order to obtain various transcripts and to extract putative anhydrobiotic candidate genes. Assembly of the ESTs allowed the identification of 1997 and 1858 non-redundant sequences for active and inactive stages, respectively. The average unigene length was 579 nucleotides. Homology search (BLASTX) using *M. tardigradum* unigenes against the NCBI database showed that nearly 50% of the ESTs had no corresponding entry in GenBank. All ESTs were deposited in GenBank (See accession numbers in the additional file 1). The three available tardigade datasets were processed and compared (Table 2, Figure 1) in order to get an overview of the similarity and redundancy between our library and the other two EST resources.

GO enrichment analysis of *M. tardigradum* ESTs

To study the functional differences between active and inactive stages of *M. tardigradum*, we performed a GO

Table 1 Summary of the expressed sequence tag (EST) analysis of the *M. tardigradum* stages (active and inactive).

Description	Active	Inactive
Total number of raw sequences	4992	4987
Total number of quality ESTs	3617	3498
Number of contigs	466	431
Number of ESTs in contigs	2103	2106
Average clone per contig	4.5	4.8
Number of singletons	1540	1437
Total non-redundant sequences	1997	1858
Blast hits (%)	52.83	51.18
No blast hits (%)	47.17	48.82

Figure 1 Clustered heatmap of *M. tardigradum* TBLASTX relative bit score similarities against *H. dujardini* & *R. coronifer*. Top BLAST bit scores are shown as color gradient corresponding to their value. The red trace was added for better visualization of similarity values. The asterisk (*) marks sequences that could not be found in *H. dujardini* but are highly conserved between *M. tardigradum* and *R. coronifer*.

enrichment analysis between the two datasets (Figure 2; additional file 2). Studying functional differences give insight into global mechanisms that are at work in the desiccating animals. Comparing the datasets revealed that 24 GO terms were significantly underrepresented in the inactive stage. The underrepresented GO-terms which were mapped to "nucleosome", "nucleosome assembly", "chromatin assembly or disassembly" and "chromatin assembly" (GO:0000786, GO:0006334, GO:0006333, GO:0031497) consist exclusively of transcripts coding for histones. The cellular component (CC) subset of differential terms is also solely associated with structural components of the genome, such as "nucleosome" (GO:0000786), "chromatin" (GO:0005694), "chromosome" (GO:0000785), and "chromosomal part" (GO:0044427). Finding only underrepresented terms is consistent with the global metabolic arrest of animals
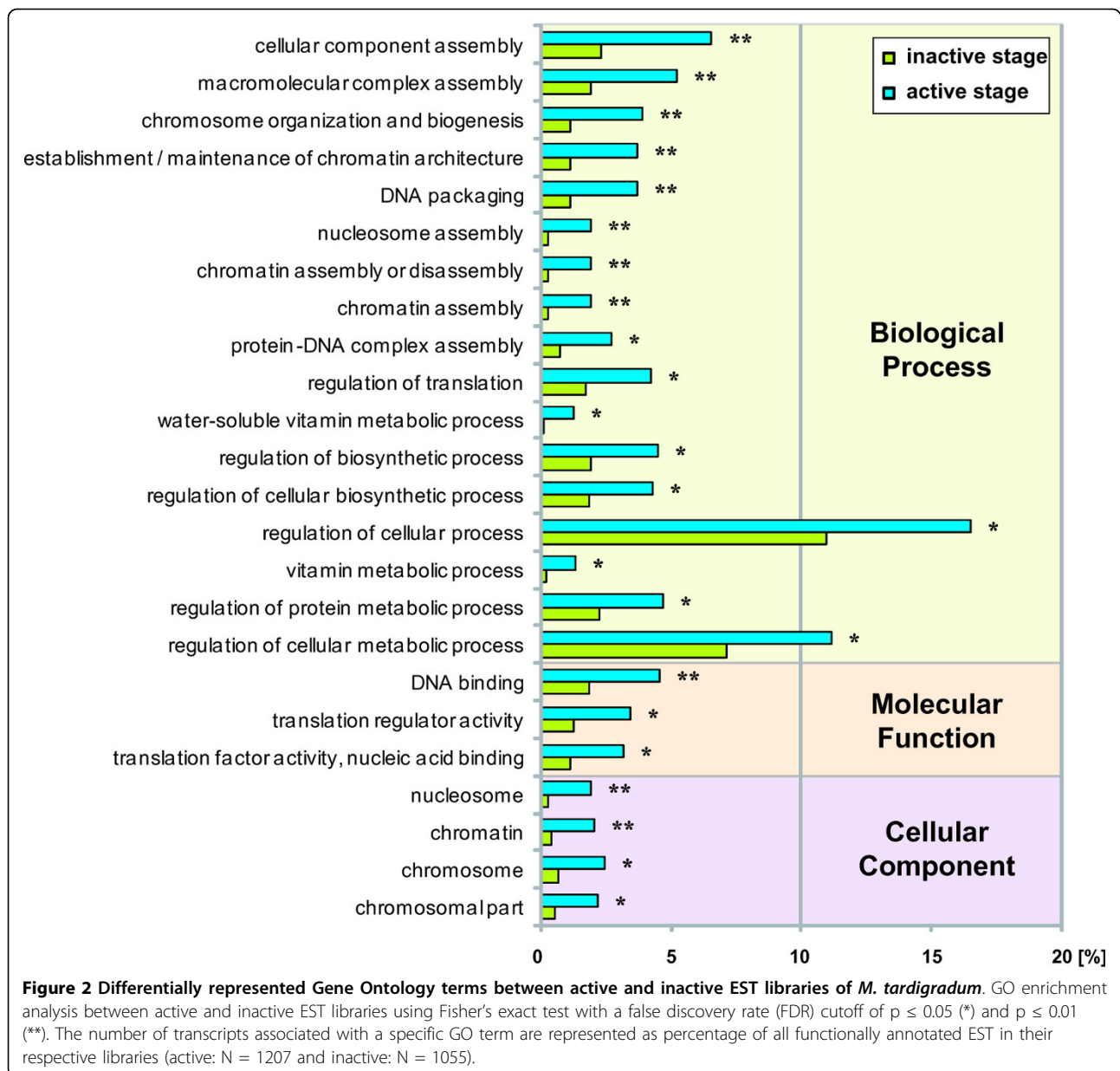
undergoing cryptobiosis. Histone mRNA expression is tightly linked to DNA replication and regulated by the cell cycle [39]. A study in *Caenorhabditis elegans* under anoxia showed similar adaptations such as cell cycle arrest, dephosphorylation of the histone H3 and morphological changes in the chromatin distribution [40]. A metabolic suppression could limit cellular and genomic damage by reducing the energy turnover to a minimum making the organism less susceptible to stress and therefore ensuring cell survival e.g. by decreased production of free radicals. Also GO-terms involved in translation regulation seem to be affected e.g. "regulation of translation" (GO:0006417), "translation regulator activity" (GO:0045182) and "translation factor activity, nucleic acid binding" (GO:0008135), implying modulation of translational activity as a response to desiccation.

The most abundant ESTs in active and inactive libraries of *M. tardigradum*

The total EST count obtained by comparing the active against the inactive dataset of *M. tardigradum* is summarized in the Table 3. The relative abundance of some transcripts in the inactive stage may indicate that they have been transcribed during the desiccation process or have been stored to be translated on rehydration. Biologically, high survival rates in *M. tardigradum* are accomplished only when drying slowly at high relative

Table 2 Summary of number of EST sequences, contigs, and singletons in tree tardigrade cDNA libraries.

	M. tardigradum	*H. dujardini*	*R. coronifer*
# of raw sequences	9984	5235	3360
# of quality ESTs	7209	5221	2819
Singleton	2419	1640	1083
Contigs	864	707	373
unigene	**3283**	**2347**	**1456**

Figure 2 Differentially represented Gene Ontology terms between active and inactive EST libraries of *M. tardigradum*. GO enrichment analysis between active and inactive EST libraries using Fisher's exact test with a false discovery rate (FDR) cutoff of p ≤ 0.05 (*) and p ≤ 0.01 (**). The number of transcripts associated with a specific GO term are represented as percentage of all functionally annotated EST in their respective libraries (active: N = 1207 and inactive: N = 1055).

humidity [16,41], suggesting that anhydrobiotic tardigrades like rotifers [42], need time to activate certain mechanisms for optimal anhydrobiosis. Probably this is because the transcription of RNAs coding for protection components has to take place. Among the genes that have a higher relative representation in the inactive stage are as follows:

Lipid-related transcripts
Lipid-related transcripts are represented mainly by intracellular fatty acid binding protein (FABP). FABPs have a low molecular mass and bind with high affinity to hydrophobic ligands such as saturated and unsaturated long-chain fatty acids. Various functions have been proposed for FABPs such as the uptake, transport, and delivery of fatty acids to beta-oxidation [43,44]. FABPs are also thought to be active fatty acid chaperones by protecting and shuttling fatty acids within the cell [45,46]. However the biological role and mechanisms of action of FABPs remain poorly understood. The transcript level of FABP was identified by cDNA array and Northern blot analysis as being up-regulated during hibernation of ground squirrels [43,47]. Members of the FABPs family have recently been identified and reported to increase in the monogonont rotifer *Brachionus plicatilis* during dormancy

Table 3 The most abundantly represented transcripts in the *M. tardigradum* active and inactive libraries.

Gene family	EST count		E-value
	Active	Inactive	
unknown	65	113	
cytochrome b	43	37	1E-080
intracellular fatty acid binding protein	30	43	7E-015
kazal-type serine proteinase inhibitor	24	38	7E-06
unknown	38	32	
ATP synthase F0 subunit 6	28	28	7E-017
unknown	30	21	
unknown	21	28	
unknown	19	23	
cytochrome c oxidase subunit III	21	20	4E-036
cytochrome c oxidase subunit I	8	29	5E-144
40S ribosomal protein S27	13	15	2E-038
40S ribosomal protein S25	19	13	4E-023
40S ribosomal protein S21	10	20	7E-024
NADH dehydrogenase subunit 4	13	15	2E-021
unknown	16	10	
vitellogenin	9	16	2E-013
similar to Actin-5C isoform 2	3	16	1E-015
cytochrome oxidase subunit II	8	10	2E-034
unknown	5	12	
cystatin B	5	10	5E-014
NADH dehydrogenase subunit 5	4	10	3E-021
elongation factor 1 alpha	10	2	4E-124

[48]. The presence of FAPB in inactive stage of *M. tardigradum* may imply conserved mechanisms shared between rotifer dormancy and anhydrobiosis in tardigrades and presumably other organisms as well. FAPB may protect membranes and ensure fatty acids as energy saving storage during anhydrobiosis.

Protease inhibitors

To date, little is known about the possible mechanisms of proteolytic inhibition or suppression in anhydrobiotic organisms. Protease inhibitors are candidate genes which would offer protection against protein degradation during anhydrobiosis. Among the abundant protease inhibitors transcripts in inactive stages of *M. tardigradum* are Kazal-type serine proteinase inhibitor and Cystatin B. Overexpression of Cystatin B (an intracellular cysteine proteinase inhibitor) in transgenic yeast and *Arabidopsis* showed an increase in the resistance to high salt, drought, oxidative, and cold stresses [49]. Elevated levels of transcripts coding for protease inhibitors such as Cystatin B have also been found in brine shrimp cysts [50]. The abundance of protease inhibitors may inhibit proteolytic reactions of proteases that could damage tissues during the desiccation process or as a response to induction of proteases as a result of

aggregated proteins. Also a protection against microbial degradation could be possible as this can occur at humidity levels at which tardigrades can't rehydrate and actively mobilize any cellular defence mechanisms.

Cytochrome c oxidase subunit I

Cytochrome c oxidase subunit I (COXI) is a mitochondrial gene that encodes the cytochrome c oxidase subunit I, a crucial enzyme involved in oxidative phosphorylation and thus energy production. COXI was over threefold more represented in the inactive state. Transcripts encoding COXI were also abundantly expressed during dehydration stress in the antarctic nematode *Plectus murrayi* [51] and up regulated by temperature increase in the yeast-like fungus *Cryptococcus neoformans* [52]. The mitochondrial COXI upregulation may serve to prevent the damage to the electron transport chain caused by desiccation and to keep an increased energy production for the survival of the tardigrades.

The sequences, which could not be assigned any function based on homology search in NCBI, were searched for conserved domains in ProDom [53] and Swiss-Prot databases [54] but did not show any hits. Since these are not all full-length sequences, it is possible that they may have missed characteristic motifs or domains for classification. A detailed investigation of their function as well as other identified transcripts presented in Table 3 will be a task in the future.

Transcripts with putative functions in desiccation resistance identified in all three tardigrade species datasets

In a cross-search over the four tardigrade EST resources (active and inactive libraries of *M. tardigradum*, *R. coronifer* and *H. dujardini*), transcripts which are potentially associated with desiccation tolerance during anhydrobiosis in other organisms were identified (see additional files 3, 4 and 5).

Detoxification-related genes

Oxidative stress proteins have been shown to be an important component in many biological processes [55]. They mediate detoxification and have putative roles as antioxidants such as glutathione S-transferase (GST), thioredoxin, superoxide dismutase (SOD), glutathione peroxidases and peroxiredoxin. It was shown that overexpression of GST/glutathione peroxidase increased the resistance to oxidative and water stress in transgenic tobacco plants [56]. GSTs are a diverse superfamily of multifunctional proteins that are reported to play a prominent role in the detoxification metabolism in nematodes [57]. In particular the up-regulation of detoxifying enzymes GST and SOD in *Plectus murrayi* [51] suggests

an efficient role of reactive oxygen species (ROS) scavenging mechanisms under desiccation stress. These observations led us to postulate that the tardigrade GST and SOD are likely to deal with oxidatively damaged cellular components during desiccation. These enzymes that help in the removal of these compounds contribute to cellular survival after oxidative damage.

Aquaporins

Many organisms adapt to desiccation stress by the activation of various water-channel proteins, called aquaporins (AQP) [58,59]. Data from *Polypedilum vanderplanki* indicates that of the two aquaporins isolated from this organism, one is involved in anhydrobiosis, whereas the other controls water homeostasis of the fat body during normal conditions [60]. Similarly, the aquaporins in larvae of the goldenrod gall fly, *Eurosta solidaginis* were either upregulated (AQP3) or downregulated (AQP2 and AQP4) following desiccation [61]. The upregulated AQP3 is especially intriguing because it is permeable to water and glycerol across the cell membrane as larvae prepare for the osmotic stress associated with desiccation. In our study aquaporin transcripts have been identified in all tardigrade datasets. These AQPs may act in concert with other transmembrane proteins to mediate the rapid transport of water across the plasma membrane during anhydrobiosis when its diffusion through the phospholipid layer of the membrane is limited.

Molecular chaperones

In the four tardigrade datasets we have identified some putative heat shock protein (HSP) encoding genes. HSPs are highly conserved throughout evolution and they function as molecular chaperones and play primary roles in protein biosynthesis and folding [62]. In tardigrades, there is considerable debate concerning the role of HSPs under desiccation stress. In the *R. coronifer*, a lower level of Hsp70 protein was found in desiccated animals when compared with active ones [63]. In *M. tardigradum*, one isoform of the *hsp70* tarnscripts showed up-regulation during the transition from active to the inactive state [64,65], while the other *hsp70* isoforms are downregulated and seem not to be directly involved in anhydrobiosis. Using the same model *M. tardigradum*, Reuner et al. [65] found an upregualtion of hsp90 in the inactive state. Certainly Hsp70 isoforms and hsp90 are involved in tardigrade desiccation, but further studies are necessary to understand how these proteins work to protect anhydrobiotic organisms.

Much attention was recently paid on the chaperone-like LEA (late embryogenesis abundant) proteins in anhydrobiotic animals [66,67]. LEA proteins are mainly low molecular weight (10-30 kDa) proteins associated with tolerance to water stress resulting from desiccation and cold shock [68,69]. Genes encoding LEA-like proteins have been identified in the nematode *Aphelenchus avenae* under dehydration condition [70-72]. A similar gene was identified and upregulated in the larvae of *P. vanderplanki* by water stress imposed by either desiccation or hypersalinity [73]. Recently, LEA have also been identified and shown to be induced under dehydration in the springtail *Megaphorura arctica* [74]. In the tardigrade EST libraries, LEA transcripts have been found in the *H. dujardini* library (the less tolerant tardigrade) and also in the proteome map of *M. tardigardum* [75]. These data suggest that LEA-like proteins could be widespread in anhydrobiotic organisms and serve important functions during desiccation.

The translationally controlled tumor protein (TCTP) found in all tardigrada datasets is often designated as a stress-related protein because of its highly regulated expression in stress conditions and its close relation to a family of small chaperone proteins [76]. Importantly, TCTP can bind to native proteins and protect them from thermal denaturation [77].

Trehalose synthesis-related gene

Trehalose, which accumulates in many anhydrobiotic organisms during desiccation is proposed to act as a common water replacement molecules and stabilizer of biological structures [78-80]. The accumulation of trehalose has been reported in the cysts of the crustacean *Artemia franciscana* [81], in the nematode *Aphelenchus avenae* [82] and in the insect larvae of the *P. vanderplanki* [83]. However, anhydrobiotic Bdelloid rotifers are unable to produce trehalose [41,84]. In addition, the trehalose-6-phosphate synthase genes (*tps*) have not been found in rotifer genomes [41]. Although trehalose accumulates substantially in the eutardigrade *Adorybiotus coronifer* [85], it was surprisingly immeasurable in *M. tardigardum* [16] and we could not find transcripts of *tps* in *M. tardigradum* ESTs. Nevertheless, transcripts coding for trehalases have been described in *M. tardigradum* [86] but we propose that its function is probably limited to the catabolism of trehalose taken up from food sources. The hypothesis of trehalose as a protective agent during desiccation may not be applicable to all anhydrobiotic organisms and in *M. tardigradum* other strategies are probably employed.

Comparative ESTs analysis between the three tardigrade species

The datasets analysed in this study represent most of the available transcriptome data from tardigrades, and until now there is little information on tardigrade genome and transcriptome structure. The genome sizes range from very compact genomes, ~75 Mb for *H.*

dujardini, considered as one of the smallest tardigrade genomes [36], up to 800 Mb for other species http://www.genomesize.com. Our dataset adds a substantial part towards the complete gene content in the tardigrada species.

To investigate the complementation of the three tardigrade datasets (*M. tardigradum*, *H. dujardini* and *R. coronifer*) we searched for putative orthologous sequences across all three datasets. Using a TBLASTX search with an e-value threshold of 10^{-5} we compared the *M. tardigradum* unigenes against the other two datasets. The BLAST bit-score of each top-scoring hit was extracted and *M. tardigradum* sequences that exhibited sequence similarity against at least one other tardigrade species are presented as a clustered heatmap in Figure 1 (see also additional file 6). *M. tardigradum* unigenes show similarities against both other species with some hits only present in either one of them (N = 785). A higher coverage of *M. tardigradum* transcripts can be seen in the *H. dujardini* dataset compared to *R. coronifer* which is likely due to the smaller *R. coronifer* dataset. This cross species comparison implies that the remaining 2498 unigenes contained in the *M. tardigradum* dataset represent further yet unknown tardigrade transcripts and expands the known tardigrade sequence data. These might be very interesting for studying the evolutionary relationships of protein families.

To calculate the average relative transcriptome sequence similarity between *M. tardigradum* and the other two tardigrade species we included only sequences that were common to all three tardigrade species (N = 368). These contained mainly abundant transcripts e.g. ribosomal proteins, ADP-ribosylation factor, ubiquitin, glyceraldehyde-3-phosphate dehydrogenase and heat shock proteins. The resulting average transcript similarity for *M. tardigradum* against *H. dujardini* was 147.66 +/- 88.27 and *M. tardigradum* against *R. coronifer* 150.95 +/- 93.74. This is reflected in the phylogenetic distance calculated using 18S rRNA sequences (see additional file 7), which positions *R. coronifer* closer to *M. tardigradum*.

Conclusions

This study describes novel sequence data from the tardigrade *M. tardigradum* that identified a set of 3283 unigenes, which significantly contributes to the available tardigrade sequence data and will help to establish this tardigrade as a model for studying desiccation tolerance. The comparison of active and inactive stage EST libraries by performing an exploratory GO enrichment analysis suggests a metabolic suppression in terms of replication and translation during desiccation. The tardigrade-EST resource generated from this study will serve as a reference for future global gene expression

experiments, aiming at the identification of key regulators of desiccation resistance during anhydrobiosis. Furthermore the datasets of *H. dujardini* and *R. coronifer* will serve as additional resources that could give clues about the evolutionary conservation of these regulators between tardigrade species of different anhydrobiotic capabilities.

Methods
Animal culture and sampling

M. tardigradum was reared in a laboratory culture on 3% agar plates covered with Volvic® mineral water at 20 ± 2°C and a light/dark cycle of 12 h as previously published [8]. For all experiments, adult animals (eight weeks after hatching) in good condition were collected directly from the agar plate using a pipette and a stereomicroscope. Tardigrades were starved for 3 days, and washed for several times with Volvic® mineral water before being processed to avoid contaminations. A total of 1000 animals were collected into 1.5 ml Eppendorf tubes in aliquots of 200 animals each. Animals representing the active state were frozen directly in liquid nitrogen. Anhydrobiotic stages of *M. tardigradum* were generated by a previously published protocol [8]. Briefly, *M. tardigradum* (200 animals) were placed in 1.5 ml Eppendorf tube and desiccated at room-temperature at 85% relative humidity (RH) for 12 to 16 hours (till they have completed the tun formation) and then at 35% RH for further 48 hours. The animals were frozen at -80°C until their experimental use.

Library construction

Total RNA extraction was performed by following the instructions of QIAGEN RNeasy R Mini kit (Qiagen, Hilden, Germany). Complete lysis of the tardigrades and especially disruption of their harsh cuticle was achieved by sonication on ice for 1 min (duty cycle 0.5s) by using a microsonicator (Probe 73, Sonopuls; Bandelin). For cDNA synthesis 1 μg total RNA was reverse transcribed using the Creator™ SMART™ cDNA Library Construction Kit following the manufacturers recommendations (Clontech-TaKaRa Bio Europe, France). The resulting first strand cDNA was amplified by LD-PCR for 18 cycles according to the manufacturers protocol using the 5' PCR primer (5'-AAGCAGTGGTATCAACGCA-GAGT-3') as the forward and the CDSIII/3'PCR Primer (5'-ATTCTAGAGGCCGAGGCGGCCGACATG-d(T) 30N-1N-3') as reverse primer. The amplified PCR products were then analyzed by agarose gel electrophoresis. After digestion of the amplified cDNA with the *Sfi*I restriction enzyme, products smaller than 300 bp were removed using the Chroma Spin-400 column as described in the Creator SMART™ protocol and cloned into pDNR-Lib cloning vector. This procedure was

chosen because of the low amount of starting material. Plasmids were transferred via electroporation to *Escherichia coli* (strain DH10B, Invitrogen, Karlsruhe, Germany).

cDNA sequencing
In total, 9984 cDNA clones were either picked by hand or automatically using a QPix robot (Genetix, UK) into 384 well LB-agar culture plates containing chloramphenicol. Sequencing was mostly from the 5' end using standard M13 forward sequencing primer. The sequencing of the cDNA library was sequenced on a ABI 3730XL capillary sequencer by GATC Biotech AG (Konstanz, Germany).

Sequence analysis and annotation
The EST analysis pipeline (Figure 3) includes typically, EST pre-processing, EST assembly and annotation of the resulting unigenes. The result is the generation of a clean, high-quality EST sequence set. Both chromatogram (*M. tardigradum*) and FASTA sequences (*H. dujardini* and *R. coronifer*) files are accepted as entry point to the analysis. Base calling was performed with phred [87,88] using a score threshold of 20. Low quality sequences, cloning vector, poly A or T tails, adaptors, and short sequences (<100 base pairs) are removed from the sequences with SeqClean [89]. Repetitive elements are masked with RepeatMasker [90]. Sequences that can be considered contaminants and unexpected vector sequences are also removed with SeqClean, using NCBI's UniVec database (v5.1) [91].

TIGR Gene Indices clustering tools (TGICL) with standard parameters [92] and CAP3 [93] have been used for the assembly step. For functional annotation, processed putative unique transcripts were loaded into the Blast2GO software [94]. Blasting was done with BLASTX algorithm using Blast2GO (v2.3.5) standard parameters. Unigenes were annotated with GO terms using standard evidence GO weight parameters. The 'Augment Annotation by ANNEX' function was used to refine annotations. Subsequently, Inter-ProScan [95] was performed to find conserved functional domains. GO terms derived from domains were merged into the existing GO annotation of the respective unigenes.

GO enrichment analysis
Identification of GO terms differentially enriched between the active and inactive *M. tardigradum* datasets

Figure 3 Overview of the tardigrade EST processing pipeline. Software used by each process in this flowchart are shown in brackets.
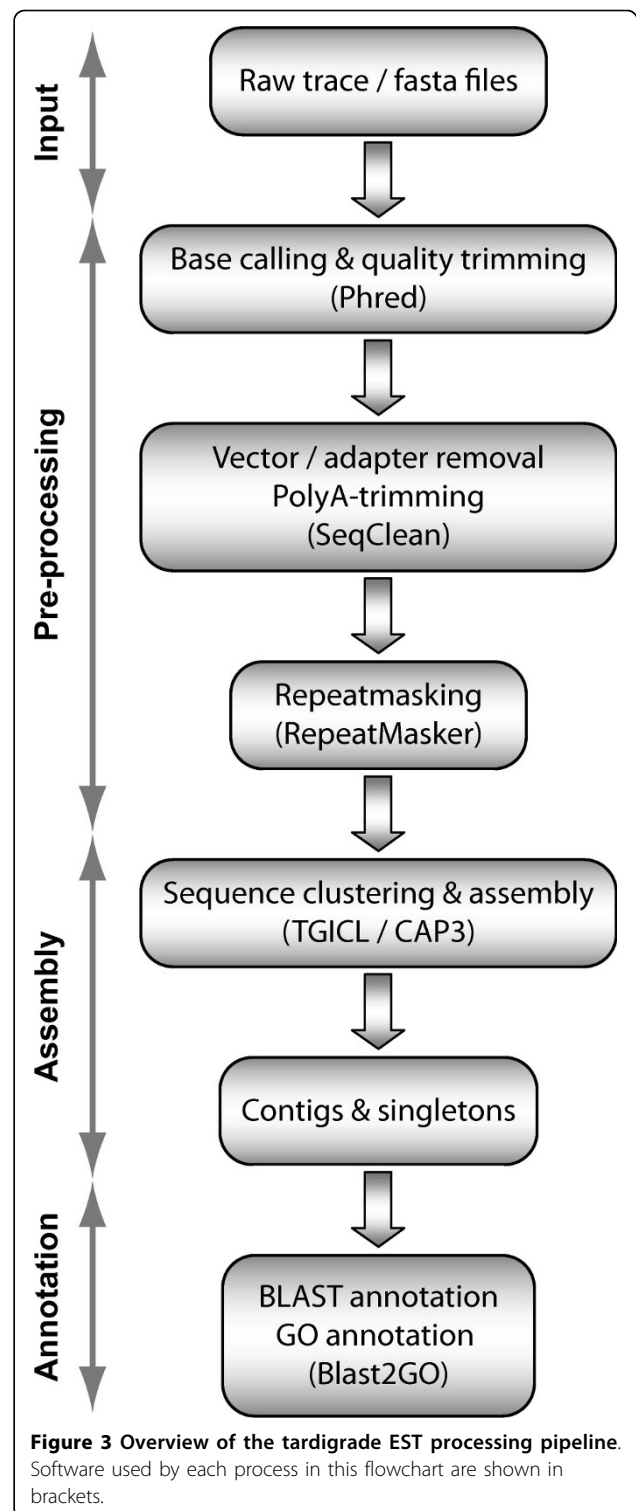
was performed using the GOSSIP statistical framework [96] webservice via the BLAST2GO software. GOSSIP employs 2 × 2 contingency tables of annotation frequencies for each GO term and computes p-values using Fisher's exact test. The statistical framework accounts for false positives (type-I-errors) that arise from multiple testing by calculating adjusted p-values. We screened for significantly enriched GO-terms by controlling the false discovery rate (FDR), setting a cut-off threshold of pFDR(p) ≤ 0.05. GO terms fulfilling this criterion were considered differentially enriched between the two *M. tardigradum* datasets.

Additional file 1: List of *M. tardigradum* ESTs and their GenBank accession numbers. This file provides a list of dbEST ID, User ID and GenBank accession numbers of all *M. tardigradum* ESTs.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-11-168-S1.XLS]

Additional file 2: GO-enrichment analysis statistics of *M. tardigradum* datasets. This file contains details about the GO-enrichment analysis between the active and inactive stage of *M. tardigradum* using GOSSIP. A list of enriched ESTs is provided.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-11-168-S2.XLS]

Additional file 3: Putative anhydrobiotic transcripts identified in *M. tardigradum* dataset. This file provides a list of ESTs in active and inactive stages of *M. tardigradum* that are potentially associated with desiccation tolerance.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-11-168-S3.XLS]

Additional file 4: Putative anhydrobiotic transcripts identified in *H. dujardini* dataset. This file provides a list of *H. dujardini* sequences that are potentially associated with desiccation tolerance.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-11-168-S4.XLS]

Additional file 5: Putative anhydrobiotic transcripts identified in *R. coronifer* dataset. This file provides a list of *R. coronifer* sequences that are potentially associated with desiccation tolerance.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-11-168-S5.XLS]

Additional file 6: Putative orthologous sequences of *M. tardigradum* against *H. dujardini* and *R. coronifer*. This file provides a list of the putative orthologues shared by tardigrade species investigated in this study.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-11-168-S6.XLS]

Additional file 7: Phylogenetic tree based on tardigrade 18S rRNA sequences. Displays a phylogenetic tree constructed from *E. testudo, M. tardigradum, R. coronifer* and *H. dujardini* 18S rRNA sequences.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-11-168-S7.PDF]

Author details
[1]Molecular Biology and Functional Genomics, University of Applied Sciences Wildau, Bahnhofstraße 1, 15745 Wildau, Germany. [2]Department of Bioinformatics, Biocenter, University of Würzburg, Am Hubland, 97074 Würzburg, Germany. [3]German Cancer Research Center (DKFZ), Protein Analysis Facility, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany. [4]Oncoscience AG, Wedel, Germany. [5]Biological Institute, Zoology, Universität Stuttgart, Pfaffenwaldring 57, 70569 Stuttgart, Germany.

References
1. Crowe JH, Hoekstra FA, Crowe LM: **Anhydrobiosis.** *Annu Rev Physiol* 1992, **54**:579-599.
2. Tunnacliffe A, Lapinski J: **Resurrecting van Leeuwenhoek's rotifers: a reappraisal of the role of disaccharides in anhydrobiosis.** *Philos Trans R Soc Lond B Biol Sci* 2003, **358**:1755-1771.
3. Watanabe M: **Anhydrobiosis in invertebrates.** *Appl Entomol Zool* 2006, **41**:15-31.
4. Schill RO, Mali B, Dandekar T, Schnölzer M, Reuter D, Frohme M: **Molecular mechanisms of tolerance in tardigrades: New perspectives for preservation and stabilization of biological material.** *Biotech Adv* 2009, **27**:348-352.
5. Baumann H: **Bemerkungen zur Anabiose der Tardigraden.** *Zool Anz* 1927, **72**:1-4.
6. Guidetti R, Jönsson KI: **Long-term anhydrobiotic survival in semi-terrestrial micrometazoans.** *J Zool* 2002, **257**:181-187.
7. Bertolani R, Guidetti R, Jönsson KI, Altiero T, Boschini D, Rebecchi L: **Experiences with dormancy in tardigrades.** *J Limnol* 2004, **63**:16-25.
8. Hengherr S, Brümmer F, Schill RO: **Anhydrobiosis in tardigrades and its effects on longevity traits.** *J Zool* 2008, **275**:216-220.
9. Jørgensen A, Møbjerg N, Kristensen RM: **A molecular study of the tardigrade *Echiniscus testudo* (Echiniscidae) reveals low DNA sequence diversity over a large geographical area.** *J Limnol* 2007, **66(Suppl 1)**:77-83.
10. Browne J, Tunnacliffe A, Burnell A: **Anhydrobiosis: plant desiccation gene found in a nematode.** *Nature* 2002, **416**:38.
11. Gal TZ, Glazer I, Koltai H: **An LEA group 3 family member is involved in survival of *C. elegans* during exposure to stress.** *FEBS Lett* 2004, **577**:21-26.
12. Kikawada T, Nakahara Y, Kanamori Y, Iwata K, Watanabe M, McGee B, Tunnacliffe A, Okuda T: **Dehydration-induced expression of LEA proteins in an anhydrobiotic chironomid.** *Biochem Biophys Res Commun* 2006, **348**:56-61.
13. Oliver AE, Leprince O, Wolkers WF, Hincha DK, Heyer AG, Crowe JH: **Non-disaccharidebased mechanisms of protection during drying.** *Cryobiology* 2001, **43**:151-167.
14. Crowe LM: **Lessons from nature: the role of sugars in anhydrobiosis.** *Comp Biochem Physiol Part A Mol Integr Physiol* 2002, **131**:505-513.
15. Crowe JH: **Trehalose as a "chemical chaperone": fact and fantasy.** *Adv Exp Med Biol* 2007, **594**:143-158.
16. Hengherr S, Heyer AG, Köhler HR, Schill RO: **Trehalose and anhydrobiosis in tardigrades-evidence for divergence in responses to dehydration.** *FEBS J* 2008, **275**:281-288.
17. Schill RO, Fritz GB: **Desiccation tolerance in embryonic stages of the tardigrade *Milnesium tardigradum*.** *J Zool* 2008, **276**:103-107.

18. Horikawa DD, Sakashita T, Katagiri C, Watanabe M, Kikawada T, Nakahara Y, Hamada N, Wada S, Funayama T, Higashi S, Kobayashi Y, Okuda T, Kuwabara M: Radiation tolerance in the tardigrade *Milnesium tardigradum*. *Int J Radiat Biol* 2006, **82**:843-848.

19. Jönsson KI, Rabbow E, Schill RO, Harms-Ringdahl M, Rettberg P: Tardigrades survive exposure to space in low Earth orbit. *Curr Biol* 2008, **18**:R729-R731.

20. Hengherr S, Worland MR, Reuner A, Brümmer F, Schill RO: High-temperature tolerance in anhydrobiotic tardigrades is limited by glass transition. *Physiol Biochem Zool* 2009, **82**:749-755.

21. Hengherr S, Reuner A, Worland R, Brümmer F, Schill RO: Freeze tolerance, super cooling points and ice formation: comparative studies on the subzero temperature survival of limno-terrestrial tardigrades. *J Exp Biol* 2009, **212**:802-807.

22. Hengherr S, Reuner A, Brümmer F, Schill R: Ice crystallization and freeze tolerance in embryonic stages of tardigrades. *Comp Biochem Physiol Part A Mol Integr Physiol* 2010.

23. Gladyshev E, Meselson M: Extreme resistance of bdelloid rotifers to ionizing radiation. *Proc Natl Acad Sci USA* 2008, **105**:5139-5144.

24. Watanabe M, Nakahara Y, Sakashita T, Kikawada T, Fujita A, Hamada N, Horikawa DD, Wada S, Kobayashi Y, Okuda T: Physiological changes leading to anhydrobiosis improve radiation tolerance in *Polypedilum vanderplanki* larvae. *J Insect Physiol* 2007, **53**:573-579.

25. Guidetti R, Bertolani R: Tardigrade taxonomy: an updated check list of the taxa and a list of characters for their identification. *Zootaxa* 2005, **845**:1-46.

26. Ramløv H, Westh P: Survival of the cryptobiotic eutardigrade *Adorybiotus coronifer* during cooling to -196°C : effect of cooling rate, trehalose level, and short-term acclimation. *Cryobiology* 1992, **29**:125-130.

27. Ramløv H, Westh P: Cryptobiosis in the eutardigrade *Adorybiotus coronifer* : tolerance to alcohols, temperature and de novo protein synthesis. *Zool Anz* 2001, **240**:517-523.

28. Wright JC: The significance of four xeric parameters in the ecology of terrestrial Tardigrada. *J Zool* 1991, **224**:59-77.

29. Jönsson KL, Borsari S, Rebecchi L: Anhydrobiotic Survival in Populations of the Tardigrades *Richtersius coronifer* and *Ramazzottius oberhaeuseri* from Italy and Sweden. *Zool Anz* 2001, **240**:419-423.

30. Guil N, Giribet G: Fine scale population structure in the *Echiniscus blumi-canadensis* series (Heterotardigrada, Tardigrada) in an Iberian mountain range–When morphology fails to explain genetic structure. *Mol Phylogenet Evol* 2009, **51**:606-613.

31. Nichols PB, Nelson DR, Garey1 JR: A family level analysis of tardigrade phylogeny. *Hydrobiologia* 2006, **558**:53-60.

32. Sands CJ, Convey P, Linse K, McInnes SJ: Assessing meiofaunal variation among individuals utilising morphological and molecular approaches: an example using the Tardigrada. *BMC Ecol* 2008, **8**:7-18.

33. Schill RO, Steinbrück G: Identification and differentiation of heterotardigrada and eutardigrada species by riboprinting. *J Zool Syst Evol Res* 2007, **45**:184-190.

34. Schill RO, Förster F, Dandekar T, Wolf M: Distinguishing species in Paramacrobiotus (Tardigrada) via compensatory base change analysis of internal transcribed spacer 2 secondary structures, with the description of three new species. *Org Divers Evol* .

35. Dunn CW, Hejnol A, Matus DQ, Pang K, Browne WE, Smith SA, Seaver E, Rouse GW, Obst M, Edgecombe GD, Sørensen MV, Haddock SHD, Schmidt-Rhaesa A, Okusu A, Kristensen RM, Wheeler WC, Martindale MQ, Giribet G: Broad hylogenomic sampling improves resolution of the animal tree of life. *Nature* 2008, **452**:745-749.

36. Gabriel WN, McNuff R, Patel SK, Gregory TR, Jeck WR, Jones CD, Goldstein B: The tardigrade *Hypsibius dujardini*, a new model for studying the evolution of development. *Dev Biol* 2007, **312**:545-559.

37. Gabriel WN, Goldstein B: Segmental expression of Pax3/7 and engrailed homologs in tardigrade development. *Dev Genes Evol* 2007, **217**:421-433.

38. Bavan S, Straub VA, Blaxter ML, Ennion SJ: A P2X receptor from the tardigrade species *Hypsibius dujardini* with fast kinetics and sensitivity to zinc and copper. *BMC Evol Biol* 2009, **9**:17.

39. Marzluff WF, Duronio RJ: Histone mRNA expression: multiple levels of cell cycle regulation and important developmental consequences. *Curr Opin Cell Biol* 2002, **14**:692-699.

40. Padilla PA, Nystul TG, Zager RA, Johnson ACM, Roth MB: Dephosphorylation of cell cycle-regulated proteins correlates with anoxia-induced suspended animation in *Caenorhabditis elegans*. *Mol Biol Cell* 2002, **13**:1473-1483.

41. Horikawa DD, Higashi S: Desiccation tolerance of the tardigrade *Milnesium tardigradum* collected in Sapporo, Japan, and Bogor, Indonesia. *Zoolog Sci* 2004, **21**:813-816.

42. Lapinski J, Tunnacliffe A: Anhydrobiosis without trehalose in bdelloid rotifers. *FEBS Letters* 2003, **553**:387-390.

43. Hittel D, Storey KB: Differential expression of adipose - and heart-type fatty acid binding proteins in hibernating ground squirrels. *Biochim Biophys Acta* 2001, **1522**:238-243.

44. Haunerland NH, Spener F: Fatty acid-binding proteins–insights from genetic manipulations. *Prog Lipid Res* 2004, **43**:328-349.

45. Coe NR, Bernlohr DA: Physiological properties and functions of intracellular fatty acid-binding proteins. *Biochim Biophys Acta* 1998, **1391**:287-306.

46. Makowski L, Hotamisligil GS: Fatty Acid Binding Proteins - The Evolutionary Crossroads of Inflammatory and Metabolic Responses. *J Nutr* 2004, **134**:2464S-2468S.

47. Hittel D, Storey KB: The translation state of differentially expressed mRNAs in the hibernating 13-lined ground squirrel (*Spermophilus tridecemlineatus*). *Arch Biochem Biophys* 2002, **401**:244-254.

48. Denekamp NY, Thorne MAS, Clark MS, Kube M, Reinhardt R, Lubzens E: Discovering genes associated with dormancy in the monogonont rotifer *Brachionus plicatilis*. *BMC Genomics* 2009, **10**:108.

49. Zhang X, Liu S, Takano T: Two cysteine proteinaseinhibitors from *Arabidopsis thaliana*, AtCYSa and AtCYSb, increasing the salt, drought, oxidation and cold tolerance. *Plant Mol Biol* 2008, **68**:131-143.

50. 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.

51. Adhikari BN, Wall DH, Adams BJ: Desiccation survival in an Antarctic nematode: molecular analysis using expressed sequenced tags. *BMC Genomics* 2009, **10**:69.

52. Toffaletti DL, Del Poeta M, Rude TH, Dietrich F, Perfect JR: Regulation of cytochrome c oxidase subunit 1 (COX1) expression in *Cryptococcus neoformans* by temperature and host environment. *Microbiology* 2003, **149**:1041-1049.

53. Corpet F, Servant F, Gouzy J, Kahn D: ProDom and ProDom-CG: tools for protein domain analysis and whole genome comparisons. *Nucleic Acids Res* 2000, **28**:267-269.

54. Bairoch A, Apweiler R: The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1998. *Nucleic Acids Res* 1998, **26**:38-42.

55. Françaa MB, Paneka AD, Eleutherio ECA: Oxidative stress and its effects during dehydration. *Comp Biochem Physiol* 2007, **146**:621-631.

56. Roxas VP, Smith RK, Allen ER, Allen RD: Overexpression of glutathione S-transferase/glutathione peroxidase enhances the growth of transgenic tobacco seedlings during stress. *Nat Biotechnol* 1997, **15**:988-991.

57. Lindblom TH, Dodd AK: Xenobiotic detoxification in thenematode *Caenorhabditis elegans*. *J Exp Zool* 2006, **305**:720-730.

58. Izumi Y, Sonoda S, Yoshida H, Danks HV, Tsumuki H: Roleof membrane transport of water and glycerol in the freeze toleranceof the rice stem borer, *Chilo suppressalis* Walker (Lepidoptera: Pyralidae). *J Insect Physiol* 2006, **52**:215-220.

59. Philip BN, Yi SX, Elnitsky MA, Lee RE Jr: Aquaporins play a role in desiccation and freeze tolerance in larvae of the goldenrod gall fly, *Eurosta solidaginis*. *J Exp Biol* 2008, **211**:1114-1119.

60. Kikawada T, Saito A, Kanamori Y, Fujita M, Snigórska K, Watanabe M, Okuda T: Dehydration-inducible changes in expression of twoaquaporins in the sleeping chironomid, *Polypedilum vanderplanki*. *Biochim Biophys Acta* 2008, **1778**:514-520.

61. Philip BN, Yi SX, Elnitsky MA, Lee RE Jr: Aquaporins play a role in desiccation and freeze tolerance in larvae of the goldenrod gall fly, *Eurosta solidaginis*. *J Exp Biol* 2008, **211**:1114-1119.

62. Feder ME, Hofmann GE: Heat shock proteins, molecular chaperones, and the stress response: evolutionary and ecological physiology. *Annu Rev Phys* 1999, **61**:243-282.

63. Jönsson KI, Schill RO: Induction of Hsp70 by desiccation, ionising radiation and heatshock in the eutardigrade *Richtersius coronifer*. *Comp Biochem Physiol B Comp Physiol* 2007, **146**:456-460.

64. Schill RO, Steinbrück GH, Köhler H-R: **Stress gene (hsp70) sequences and quantitative expression in** *Milnesium tardigradum* **(Tardigrada) during active and cryptobiotic stages.** *J Exp Biol* 2004, **207**:1607-1613.
65. Reuner A, Hengherr S, Mali B, Förster F, Arndt D, Reinhardt R, Dandekar T, Frohme M, Brümmer F, Schill RO: **Stress-response in tardigrades: Differential gene expression of molecular chaperones.** *Cell Stress & Chaperones* 2009.
66. Tunnacliffe A, Wise MJ: **The continuing conundrum of the LEA proteins.** *Naturwissenschaften* 2007, **94**:791-812.
67. McGee B, Schill RO, Tunnacliffe A: **Hydrophilic proteins in invertebrate anhydrobiosis.** *Integrative and Comparative Biology* 2004, **44**:679-679.
68. Ingram J, Bartels D: **The molecular basis of dehydration tolerance in plants.** *Annu Rev Plant Physiol Plant Mol Biol* 1994, **47**:377-403.
69. Hoekstra FA, Golovina EA, Buitink J: **Mechanisms of plant desiccation tolerance.** *Trends Plant Sci* 2001, **6**:431-438.
70. Browne JA, Tunnacliffe A, Burnell A: **Anhydrobiosis - plant desiccation gene found in a nematode.** *Nature* 2002, **416**:38.
71. Browne JA, Dolan KM, Tyson T, Goyal K, Tunnacliffe A, Burnell AM: **Dehydration-specific induction of hydrophilic protein genes in the anhydrobiotic nematode** *Aphelenchus avenae*. *Eukaryot Cell* 2004, **3**:966-975.
72. Goyal K, Walton LJ, Browne JA, Burnell AM, Tunnacliffe A: **Molecular anhydrobiosis: identifying molecules implicated in invertebrate anhydrobiosis.** *Integr Comp Biol* 2005, **45**:702-709.
73. Kikawada T, Nakahara Y, Kanamori Y, Iwata K, Watanabe M, McGee B, Tunnacliffe A, Okuda T: **Dehydration-induced expression of LEA proteins in an anhydrobiotic chironomid.** *Biochem Biophys Res Commun* 2006, **348**:56-61.
74. Bahrndorff S, Tunnacliffe A, Wise MJ, McGee B, Holmstrup M, Loeschcke V: **Bioinformatics and protein expression analyses implicate LEA proteins in the drought response of Collembola.** *J Insect Physiol* 2009, **55**:210-217.
75. Schokraie E, Hotz-Wagenblatt A, Warnken U, Mali B, Frohme M, Förster F, Dandekar D, Hengherr S, Schill RO, Schnoelzer M: **Proteomic analysis of tardigrades: towards a better understanding of molecular mechanisms by anhydrobiotic organisms.** *PLoS ONE* 2010.
76. Bommer UA, Thiele BJ: **The translationally controlled tumour protein (TCTP).** *Int J Biochem Cell Biol* 2004, **36**:379-385.
77. Gnanasekar M, Dakshinamoorthy G, Ramaswamy K: **Translationally controlled tumor protein is a novel heat shock protein with chaperone-like activity.** *Biochem Biophys Res Commun* 2009, **386**:333-337.
78. Wolkers WF, Tablin F, JH Crowe: **From anhydrobiosis to freeze-drying of eukaryotic cells.** *Comp Biochem Physiol* 2002, **131**:535-543.
79. Shirkey B, McMaster NJ, Smith SC, Wright DJ, Rodriguez H, Jaruga P, Birincioglu M, Helm R, Potts M: **Genomic DNA of** *Nostoc commune* **(Cyanobacteria) becomes covalently modified during long-term (decades) desiccation but is protected from oxidative damage and degradation.** *Nucleic Acids Res* 2003, **31**:2995-3005.
80. Crowe JH, Crowea LM, Wolkers WF, Oliver AE, Ma X, Auh J-H, Tang M, Zhu S, Norris J, Tablin F: **Stabilization of Dry Mammalian Cells: Lessons from Nature.** *Integr Comp Biol* 2005, **45**:810-820.
81. Clegg JS: **The origin of trehalose and its significance during emergence of encysted dormant embryos of** *Artemia salina*. *Comp Biochem Physiol* 1965, **14**:135-143.
82. Crowe JH, Madin KAC: **Anhydrobiosis in nematodes: evaporative water loss survival.** *J Exp Zool* 1975, **193**:323-334.
83. Watanabe M, Kikawada T, Minagawa N, Yukuhiro F, Okuda T: **Mechanism allowing an insect to survive complete dehydration and extreme temperatures.** *J Exp Biol* 2002, **205**:2799-22802.
84. Caprioli M, Kathiolm AK, Melno G, Ramløv H, Ricci C, Santo N: **Trehalose in desiccated rotifers: a comparison between a bdelloid and a monogonont species.** *Comp Biochem Physiol* 2004, **139**:527-532.
85. Westh P, Ramløv H: **Trehalose accumulation in the tardigrade** *Adorybiotus coronifer* **during anhydrobiosis.** *J Exp Biol* 1991, **258**:303-311.
86. Takekazu K, Takeo K: **Identification and analysis of trehalase from tardigrades,** *Milnesium tardigradum*. *Zool Sci* 2005, **22**:1493.
87. Ewing B, Hillier L, Wendl MC, Green P: **Base-calling of automated sequencer traces using phred. I. Accuracy assessment.** *Genome Res* 1998, **8**:175-185.
88. Ewing B, Green P: **Base-calling of automated sequencer traces using phred. II. Error probabilities.** *Genome Res* 1998, **8**:186-194.
89. DFCI Gene Indices Software Tools. [http://compbio.dfci.harvard.edu/tgi/software/].
90. RepeatMasker homepage. [http://www.repeatmasker.org].
91. UniVec database at National Center for Biotechnology Information. [http://www.ncbi.nlm.nih.gov/VecScreen/UniVec.html].
92. Pertea G, Huang X, Liang F, Antonescu V, Sultana R, Karamycheva S, Lee Y, White J, Cheung F, Parvizi B, Tsai J, Quackenbush J: **TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets.** *Bioinformatics* 2003, **19**:651-652.
93. Huang X, Madan A: **CAP3: A DNA sequence assembly program.** *Genome Res* 1999, **9**:868-877.
94. Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A: **High-throughput functional annotation and data mining with the Blast2GO suite.** *Nucleic Acids Res* 2008, **36**:3420-3435.
95. Zdobnov EM, Apweiler R: **InterProScan-an integration platform for the signature-recognition methods in InterPro.** *Bioinformatics* 2001, **17**:847-848.
96. Blüthgen N, Brand K, Cajavec B, Swat M, Herzel H, Beule D: **Biological profiling of gene groups utilizing Gene Ontology.** *Genome Inform* 2005, **16**:106-15.