Title:

Multimomics study of a heterotardigrade, *Echiniscus testudo*, suggests convergent evolution of anhydrobiosis-related proteins in Tardigrada

Yumi Murai^{1,2}, Maho Yagi-Utsumi^{3,4,5}, Masayuki Fujiwara^{1,2}, Masaru Tomita^{1,2,6}, Koichi Kato^{3,4,5}, Kazuharu Arakawa^{1,2,3,6*}

¹ Institute for Advanced Biosciences, Keio University, Tsuruoka, Yamagata, Japan

² Systems Biology Program, Graduate School of Media and Governance, Keio University, Fujisawa, Kanagawa, Japan

³ Exploratory Research Center on Life and Living Systems, National Institutes of Natural Sciences, 5-1 Higashiyama, Myodaiji, Okazaki 444-8787, Japan

⁴ Institute for Molecular Science, National Institutes of Natural Sciences, 5-1 Higashiyama, Myodaiji, Okazaki 444-8787, Japan

⁵ Graduate School of Pharmaceutical Sciences, Nagoya City University, Tanabe-dori 3-1, Mizuho-ku, Nagoya 467-8603, Japan

⁶ Faculty of Environment and Information Studies, Keio University, Fujisawa, Kanagawa, Japan

*Corresponding author

E-mail: gaou@sfc.keio.ac.jp

Abstract

Many limno-terrestrial tardigrades can enter an ametabolic state upon desiccation, in which the animals can withstand extreme environments. To date, studies of the molecular mechanism have predominantly investigated the class Eutardigrada, and that in the Heterotardigrada, remains elusive. To this end, we report a multiomics study of the heterotardigrade *Echiniscus testudo*, which is one of the most desiccation-tolerant species. None of the previously identified tardigrade-specific anhydrobiosis-related genes was conserved, while the loss and expansion of existing pathways were partly shared. Furthermore, we identified two families of novel abundant heat-soluble proteins and the proteins exhibited structural changes from random coil to α -helix as the water content decreased *in vitro*. These characteristics are analogous to those of anhydrobiosis-related protein in eutardigrades, while there is no conservation at the sequence level. Our results suggest that Heterotardigrada have partly shared but distinct anhydrobiosis machinery compared with Eutardigrada, possibly due to convergent evolution within Tardigrada.

(150/150 words)

41 Introduction

42 Water is an essential universal solvent in living organisms that harbors cellular biochemical
43 reactions. However, many limno-terrestrial tardigrades are able to enter an ametabolic state, known
44 as anhydrobiosis, when they face desiccation, and these organisms can subsequently tolerate almost
45 complete water loss while being able to return to their natural active state when rehydrated¹⁻³. In
46 anhydrobiosis, tardigrades are capable of tolerating many extreme environments, such as high and
47 low temperatures^{4,5}, space vacuum⁶, high pressure⁷, and high concentrations of organic solvent⁸,
48 as well as insults from ultraviolet rays and over 4,000 Gy of gamma radiation⁹⁻¹¹.

49 Early studies on the molecular mechanisms underlying anhydrobiosis initially targeted
50 *Caenorhabditis elegans*¹², *Artemia salina*¹³, and *Polypedilum vanderplanki*¹⁴, where the
51 disaccharide trehalose was identified as one of the key compounds, being accumulated at up to 20%
52 of the dry weight during desiccation in *P. vanderplanki*. Trehalose is suggested to protect intracellular
53 components, including membranes and proteins, through water replacement and vitrification^{14,15}.
54 Another of the key factors of the anhydrobiotic machinery is the hydrophilic late embryogenesis
55 abundant (LEA) protein superfamily, which has been identified in both plants and animals¹⁶. This
56 widespread protein family exhibits little sequence conservation among groups, but they often show
57 similar structural characteristics, where the protein is mostly natively unfolded in the hydrated state
58 and later folds upon desiccation^{17,18}. Another hallmark is the extreme hydrophilic nature of this
59 protein, which maintains solubility even after ten minutes of heat treatment at 100 °C¹⁹. The
60 contribution of LEA proteins in desiccation tolerance is reported to be multifactorial, encompassing
61 protecting membranes, inhibiting the aggregation of proteins, stabilizing cellular components,
62 including vitrified sugar glasses, and serving as a hydration buffer²⁰.

63 Trehalose and LEA proteins, however, are not the primary machinery of tardigrade
64 anhydrobiosis. Trehalose levels account for only several percent at maximum, and some species even
65 lack measurable amounts of trehalose^{14,21-23}. LEA proteins are observed in the tardigrade
66 transcriptome but are not among the most abundant group of transcripts, suggesting the existence of
67 unique tardigrade components^{24,25}. With the emerging availability of genomic resources^{23,26}, these
68 tardigrade-specific anhydrobiosis-related proteins were first screened using heat solubility assays in
69 the eutardigrade *Ramazzottius varieornatus*^{27,28}. While no sequence similarity to LEA was observed,
70 characteristics of the cytoplasmic abundant heat soluble (CAHS) protein, such as its high abundance,
71 heat solubility, intrinsically unstructured nature, and structural change to alpha helix upon water loss,
72 mirror that of LEA proteins²⁷. These tardigrade-specific anhydrobiosis-related genes initially
73 appeared to be highly conserved among different tardigrades, where the CAHS gene was conserved
74 all across the class eutardigrada in Hypsibiidae, Macrobiotidae, and Milnesiidae^{23,25,27,29}.
75 Intriguingly, eutardigrades appear to adopt to environments with differing desiccation conditions by

adjusting the expression of these genes while maintaining the gene sets. For example, the strong anhydrobiote *R. varieornatus* can immediately enter anhydrobiosis upon desiccation with constitutively high expression of CAHS and other proteins required for anhydrobiosis. On the other hand, semiaquatic *Hypsibius exemplaris* requires 24-48 h of preconditioning before entering anhydrobiosis³⁰ and prepares homologous proteins by de novo expression during this period²³, similar to the anhydrobiosis induction method observed in the arthropod *P. vanderplanki*³¹ and nematode *C. elegans*³². However, conservation in the other class of Heterotardigrada was not observed from transcriptome analysis of a marine species, suggesting the possibility of an alternative set of machineries²⁵.

To date, molecular analysis of Heterotardigrada has been predominantly limited due to the lack of sustainable laboratory rearing systems. Several species of terrestrial heterotardigrades (*i.e.*, *Echiniscus testudo*) are reported to have comparable or even higher desiccation tolerance than *R. varieornatus*³³; therefore, a certain specialized machinery should also exist in these species. Similarly, series of studies have reported the tolerance of heterotardigrades to various extreme environments, including exposure to vacuum³⁴, osmotic pressure³⁵, and high concentrations of environmental toxicant copper ions, which may increase reactive oxygen species³⁶. Evolutionary perspectives on such genetic components would be especially intriguing in Heterotardigrada, since the order Arthrotardigrada predominantly harbors marine species and is considered to be an ancestral group within the phylum Tardigrada³⁷. The existence of LEA proteins in a wide variety of organisms without notable sequence similarity suggests an analogous mechanism in heterotardigrades.

To this end, we conducted a multiomic study of the heterotardigrade *E. testudo* using ultrasensitive methodology³⁸⁻⁴⁰ to elucidate the molecular components of anhydrobiosis in this species. This is the first draft genome of a heterotardigrade, in which the tardigrade-unique proteins identified in eutardigrades are not conserved. We identified two families of novel abundant heat-soluble proteins in this genome, which we named *E. testudo* Abundant Heat Soluble (EtAHS), and their structural analogy to the eutardigrade-specific proteins while lacking any sequence similarity suggests convergent evolution of anhydrobiosis machinery in the two classes of tardigrades.

Results

Genome of *Echiniscus testudo*

We obtained 33 Gb of DNA sequencing data, corresponding to 85X coverage of the estimated genome, from the K-mer distribution using GenomeScope (110 Mb). The final genome assembly spanned 153.7 Mbp (30,095 scaffolds), and the longest contig and N50 length were 41,023 bp and 6,674 bp, respectively (Table 1, Supplementary Table 1). This sequencing was performed after extensive screening for contamination, which is especially critical in this work, since the specimens were collected from the wild. The initial step of the screening was based on BlobTools visualizations⁴¹, and scaffolds were removed based on similarity to bacterial sequences with low RNA-Seq coverage (most scaffolds $< 10^{-1}$) and GC contents lower than 40% or higher than 60% (Supplementary Fig. 1). The BUSCO completeness score of the genome assembly was 92.7% (eukaryote dataset), and we observed high mapping ratios for DNA-Seq and RNA-Seq data (Supplementary Table 2). Forty-two megabases (27%) of this genome was inferred to consist of unknown repeats. We further predicted 73 ribosomal genes and 1,081 transfer RNA genes. A total of 42,608 protein-coding genes (maximum TPM > 1) were predicted, of which 57.6% of the genes showed BLASTp homology against the Swiss-Prot database (E-value $< 1e-5$). Since the assembly length exceeds the expected genome size, this assembly and gene annotations presumably contain duplicate assemblies. Transcriptome assembly was also filtered by CPM (count per million; CPM > 1) to eliminate possible contamination and was further verified by BlobPlot (Supplementary Table 3, Supplementary Fig. 2.). A total of 98.6% of the transcriptome assembly was determined to match the genome assembly by a BLASTn search at an e-value threshold of $1e-50$. The genome assembly is not very contiguous, which is a limitation of the ultra-low-input protocol from a single specimen that only contains less than 50 pg of DNA. On the other hand, agreement between the genome and transcriptome assembly, as well as the BUSCO completeness, shows sufficient completeness of the gene repertoire (Supplementary Table 1).

Next, we compared the *E. testudo* gene set with other sequenced tardigrades: transcriptome data of Heterotardigrada *E. cf. sigismundi* and Eutardigrada *Richtersius coronifer*, and genome data of Eutardigrada *R. varieornatus* and *H. exemplaris*. A total of 12,917 genes ($1e-25$) of *E. testudo* were conserved in all 5 species, and approximately 529 genes were shared only among the heterotardigrades *E. testudo* and *E. cf. sigismundi* (Figure 1a). The ratio of conserved genes between each pair of five species showed that the two classes of Tardigrada, Heterotardigrada and Eutardigrada, possess class-specific conservation patterns (Figure 1b), supporting the possibility of distinct anhydrobiosis machineries.

Comparative genomics of anhydrobiosis in tardigrades

A previous comparative study on the anhydrobiotic machinery between *R. varieornatus* and *H. exemplaris* reported several pathways that contribute to desiccation tolerance²³; however, the conservation of these pathways (or genes) has not been identified to date in *Echiniscus*. We first focused on antioxidative stress-related pathways, since previous species have undergone extensive gene loss in H₂O₂, generating stress-responsive pathways, as well as gene duplications in antioxidative stress genes, suggesting intensive genomic adaptations for coping with oxidative stress^{23,26}. Naive screening of catalase, superoxide dismutases (SODs), glutamate-S transferase (GST), and heat shock protein (HSP) resulted in 5, 31, 94 and 57 copies, respectively. However, since the genome assembly likely contained duplicated assembly artifacts, we further confirmed these candidates by constructing phylogenetic trees of each gene from genomic and transcriptomic sequences, thereby removing nearly identical orthologs (> 99% identity) and possible artifacts (Supplementary Fig. 3). As a result, we identified 1, 17, 70 and 40 orthologs of catalase, SODs, GST and HSP, respectively (Figure 2). We also observed that several genes related to stress signaling, *TSC2* and *Rheb*, lost in the Eutardigrade lineage were conserved in *E. testudo* (Figure 2 and Supplementary Table 4), and this conservation pattern was in line with findings observed for *E. cf. sigismundi*, suggesting gene loss in Eutardigrada after divergence from Heterotardigrada. The majority of the genes associated with the stress response (*i.e.*, loss of Sestrin and HIF-1 α) and antioxidative stress genes (*i.e.*, duplication of SOD and GST) showed conservation patterns in *E. testudo* similar to those observed in *R. varieornatus* and *H. exemplaris*, suggesting a very early trait of the common ancestor of Tardigrada. We subsequently focused on tardigrade-specific anhydrobiosis-related genes that have been identified eutardigrades (*i.e.*, CAHS, secretory abundant heat soluble (SAHS), mitochondrial abundant heat soluble (MAHS), LEAM and damage suppressor (Dsup)). Similar to *E. cf. sigismundi*, these genes were not confirmed in the *E. testudo* genome (BLASTp search, E-value < 1e-3, Figure 2). These results suggest that while there are conserved core mechanisms in the oxidative stress response and stress signaling within the phylum Tardigrada, distinct mechanisms exist in heterotardigrades and eutardigrades.

To screen for heterotardigrade-specific anhydrobiotic machinery, we subsequently analyzed differential gene expression between fully hydrated and desiccated individuals. Differentially expressed genes (DEGs) were extremely limited, with only 21 (q < 0.05) or 13 (q-value < 0.01) candidates being identified (Supplementary Table 5 and Supplementary Fig. 4). The fold change of DEGs was moderately high (median 6.88), but most of these genes had very low expression levels (hydrated TPM < 10 is 16/21 at q < 0.05, 11/13 at q < 0.01), thus resulting in high fold changes. Overall, there is only minimal change in expression between the hydrated and anhydrobiotic states,

and screening from differential expression does not seem feasible, as is observed for the strong anhydrobiote *R. varieornatus*²⁶.

Novel heat soluble protein in heterotardigrades

The results obtained to date suggest the presence of heterotardigrade-specific machinery of anhydrobiosis, which complements the functions of the LEA or CAHS proteins observed in other anhydrobiotes, presumably contributing to cellular molecule protection upon desiccation²⁷. Therefore, we considered heat solubility screening to obtain analogous proteins. Following heat-soluble proteomics (Supplementary Table 6), filtering for highly expressed genes (maximum TPM > 100) from transcriptome data and lack of sequence similarity in Swiss-Prot yielded 32 genes as final candidates (Supplementary Table 7). These candidates were further screened for their disordered nature with negative fold index values, as these genes are predicted to have large intrinsically unstructured regions. This approach resulted in six genes of two fold index patterns, containing a high proportion of predicted alpha helical regions, similar to those of CAHS1 (Table 2, Figure 3). Although there is no sequence similarity, these *Echiniscus*-specific, heat-soluble, constitutively highly expressed (these genes were not DEGs in the above analysis and are highly expressed both in hydrated and tun states), alpha-helical proteins with intrinsically unstructured characteristics are all analogous to the CAHS protein identified in *R. varieornatus*; therefore, we named these genes EtAHS (*Echiniscus testudo* Abundant Heat Soluble). A BLAST search also identified EtAHS in *E. cf. sigismundi*, but it was not found in all eutardigrades, that is, *R. varieornatus*, *H. exemplaris* and *R. coronifer*, suggesting heterotardigrade-specific conservation.

Structural analysis of EtAHS proteins

We next performed structural characterization of the recombinant EtAHS proteins using NMR and CD spectroscopy (Figure 4). The ¹H-¹⁵N HSQC peaks were observed within a narrow spectral region (7.4-8.4 ppm for ¹H chemical shift), indicating that the EtAHS proteins were largely unstructured in solution. The unstructured properties of EtAHS proteins were confirmed by CD spectral data. Moreover, the CD data demonstrated that the desolvating agent of trifluoroethanol (TFE) induced alpha-helical conformation in the EtAHS proteins. These properties are similar to those of CAHS, which also undergoes a conformational transition from a largely unstructured state to an alpha-helical state in water-deficient conditions.

Discussion

Molecular and genomic studies of tardigrades have mostly been limited to the class Eutardigrada, where numerous species are culturable in laboratory conditions. On the other hand, no sustainable culture of heterotardigrades has been established to date. Therefore, we employed an ultra-low-input genome sequencing protocol that we developed previously³⁸⁻⁴⁰ to achieve whole genome sequencing from a single specimen of tardigrade containing less than 50 pg of DNA collected from the wild. However, such a sample requires strong precautions to minimize the possibility of contamination^{23,38,42}. After rigorous filtering of possible contaminants as described above, we generated a 153.7 Mbp draft genome assembly of *E. testudo* with high genome completeness (BUSCO score; 92.7%) and a 98.6% match of transcriptome assembly to the genome. This report is the first to describe a heterotardigrade genome, and this novel resource may facilitate future comparative genomics and evolutionary studies of the phylum Tardigrada. Transcriptome assembly of a heterotardigrade, *E. cf. sigismundi*, has been reported²⁵, and while such data are valuable in identifying the presence of a gene, genomic data are required to discuss the absence of a gene and to perform further analysis of other noncoding features.

Comparing the gene repertoire of *E. testudo* with other tardigrades, namely, *E. cf. sigismundi*, *R. varieornatus*, *H. exemplaris* and *R. coronifer*, duplication of stress-response genes, such as GST, SODs and HSP, was observed as a common characteristic of anhydrobiotic tardigrades. Reactive oxygen species and the oxidative damage induced by them are among the most critical sources of cellular damage upon desiccation; therefore, the expansion of the antioxidant defense repertoire appears to be a common mechanism in anhydrobiosis, not one that is limited to tardigrades^{43,44}. In addition, we identified several genes in stress signaling pathways, i.e., genes involved in stress signaling pathways. Sestrin, Gadd45, FADD, and Caspase 8 were lost *E. testudo*, similar to the eutardigrades *R. varieornatus* and *H. exemplaris*, suggesting that these orthologs were lost prior to the diversion of Eutardigrada and Heterotardigrada. Hypoxia-inducible factors (HIFs) are transcription factors that serve as master regulators of oxygen homeostasis and are highly conserved in most metazoans, including desiccation tolerance species⁴⁵. Initially, the loss of this modulator of apoptosis induction was suggested to be associated with anhydrobiosis to prevent apoptosis and to enable cellular repair upon rehydration from anhydrobiosis²⁶. However, if the loss of this gene occurred before the divergence of Heterotardigrada and Eutardigrada, this would imply that this gene loss is also present in the marine tardigrades of Arthrotardigrada, which do not possess anhydrobiotic capability, thereby weakening the link between these gene losses and anhydrobiosis. Recently, it was reported that an abundant intertidal crustacean, the copepod *Tigriopus californicus*, also lacks HIF- α and EGLN, but this species can tolerate extremely low levels of available O₂ for at least 24 h⁴⁶. This

finding suggests that the loss of HIF- α /EGLN is not specific to tardigrades and may not be related to anhydrobiosis.

While these core components are conserved, none of the tardigrade-specific proteins related to anhydrobiosis identified in *R. varieornatus*, namely, CAHS, SAHS, MAHS, LEAM, and Dsup, is observed in *E. testudo*. CAHS protein is highly expressed in *R. varieornatus*, and the presence of a large number of paralogs within the genome suggests an important role played by this protein during anhydrobiosis. While CAHS is constitutively highly expressed in the strong anhydrobiote *R. varieornatus*, expression of this protein is induced upon desiccation in the weak anhydrobiote *H. exemplaris*, indicating a direct functional link in anhydrobiosis²³. Therefore, analogous function is expected in *E. testudo* but without sequence similarity or homology, as in the case of different LEA groups⁴⁷. The rate of gene conservation is higher within the classes of Tardigrada, and 529 genes are only shared by heterotardigrades. Eighty-nine of these proteins have a homolog in Swiss-Prot (1e-25); therefore, candidates for heterotardigrade-specific anhydrobiosis-related proteins should be among these 440 genes. Screening of these candidates by differential expression in active and tun states was not possible due to the lack of overall expression change. This result is not surprising because strong anhydrobiotes, such as *R. varieornatus* and *E. testudo*, can enter anhydrobiosis immediately, forming tuns and losing most of their body water within 30 min. This time interval is not sufficient to conduct de novo gene transcription and protein translation, and a lack of expression changes has been reported for *R. varieornatus*, where anhydrobiosis-related proteins are constitutively highly expressed²³.

Therefore, we performed heat solubility proteomics assays to identify hydrophilic proteins, following the procedures established for the identification of LEA and CAHS proteins^{19,27}. These classes of proteins are notably hydrophilic and are often comprised of intrinsically unstructured domains that turn into alpha-helical structures upon water loss. These properties enable the protein to remain soluble even after heat treatment at near-boiling temperatures. As a result, we identified two families of six genes that we designated EtAHS that were conserved only in heterotardigrades. These proteins were predicted to contain unstructured regions similar to those observed in CAHS and SAHS, suggesting that these genes may be analogous²⁷. Further confirmation of the EtAHS structure using recombinant proteins expressed in *E. coli* showed a disordered nature by NMR spectroscopy and a change in structure from random coil to alpha helix upon water loss, which was emulated by replacing water with TFE, and both structural features resembled those of CAHS. The abundance, heat solubility, hydrophilic nature, and intrinsically disordered structure that forms an alpha-helical structure upon water loss are all characteristic of the LEA protein family. LEA proteins do not exhibit sequence similarity among groups^{47,48}; therefore, it is possible to consider that tardigrade heat-soluble protein families may be a new LEA protein family group, but such generalization is beyond the scope

272 of this work and is a possible direction for future research. Moreover, the role played by EtAHS or
 273 CAHS in anhydrobiosis has not been elucidated to date. Few studies have investigated this topic, and
 274 they have obtained ambiguous results ⁴⁹. Further studies on this subject are also necessary to fully
 275 elucidate the molecular mechanisms underlying anhydrobiosis in tardigrades. Nevertheless, the
 276 existence of analogous proteins without any sequence similarity suggests the importance of this class
 277 of proteins in anhydrobiosis and that tardigrades have independently evolved these protein families
 278 at least two times, once in Heterotardigrada and once in Eutardigrada, to enable anhydrobiosis. This
 279 phenomenon may represent an example of convergent evolution and highlights the complex
 280 evolutionary history of anhydrobiosis within the phylum Tardigrada.

281

282

283 **Conclusions**

284 In this study, we performed genomic and transcriptomic sequencing of the heterotardigrade *E. testudo*
 285 by using an ultra-low-input library sequencing method, and by combining these results with heat-
 286 soluble proteome data, we identified a novel heterotardigrade-specific anhydrobiosis-related gene
 287 family designated EtAHS, which shares multiple analogous characteristics, including heat solubility,
 288 constitutive high expression, intrinsically unstructured nature, and shift to alpha-helical structure
 289 upon water loss, with the eutardigrade analog CAHS. Structural features were further confirmed using
 290 NMR and CD analyses, which also indicated similar characteristics to those of CAHS. These results
 291 suggest the existence of two distinct, independently acquired mechanisms of anhydrobiosis in the two
 292 classes of Tardigrada, which may represent an example of convergent evolution.

293 **Materials and Methods**

294 **Animals**

295 We collected *E. testudo* from moss collected from Otsuka-machi, Tsuruoka city, Japan
 296 (38°44'23"N, 139°48'28"E). The moss was soaked in water for more than 3 h, and *E. testudo*
 297 specimens were picked and collected on an agar-coated plate. *E. testudo* was allowed to walk around
 298 the agar plate for two days, with rigorous washing and renewed plate each day to remove surface and
 299 gut contaminants. Tun (anhydrobiotic) state was induced by suspending 20-21 animals in 100 µl of
 300 distilled water and by placing it on a 2-cm x 2-cm mesh filter on a paper towel (ASONE). The animals
 301 were desiccated at 35% RH for 24 h controlled by 85% glycerol solution⁵⁰, followed by 0% RH at
 302 room temperature for complete desiccation. Anhydrobiotic success was determined by an over 90%
 303 recovery rate after 24 h.

305 **Genome and transcriptome sequencing**

306 We used the ultra-low-input library sequencing protocol established in our previous study
 307 ^{39,40}. In brief, genomic DNA was extracted using a Quick-gDNA Microprep Kit (Zymo Research)
 308 after lysis with two freeze-thaw cycles of -80°C and 37°C incubation, and the extracted DNA was
 309 sheared to 550-bp target fragments with Covaris M220. The Illumina sequencing library was
 310 constructed with a ThruPLEX DNA-Seq kit (Rubicon Genomics), and the purified library was
 311 quantified using a Qubit Fluorometer (Life Technologies). The sequencing library was sequenced on
 312 the MiSeq platform (Illumina) with 300 bp paired settings. For RNA-Seq, the total RNA was
 313 extracted from active and tun samples (3 replicates) using Direct-Zol RNA MicroPrep Kits (Zymo
 314 Research), and the sequencing library was constructed using SMARTer v4 Ultra Low Input RNA Kit
 315 for Sequencing (Clontech) and KAPA HyperPlus Library Preparation Kit (KAPA BIOSYSTEMS).
 316 The RNA-Seq library was multiplexed and sequenced on the NextSeq 500 platform (Illumina) 300
 317 cycles High Output Mode (paired-end). De-multiplexing of RNA-Seq reads was conducted using
 318 bcl2fastq v.2 (Illumina). Both DNA-Seq and RNA-Seq reads were submitted to FastQC⁵¹ for quality
 319 validation.

321 **Genome assembly and gene finding**

322 Following genome size estimation by GenomeScope⁵², the MiSeq reads were assembled with
 323 MaSuRCA (v3.1.3)⁵³ with the following parameters (PE 350 150). Since the specimens were directly
 324 collected from the wild, we needed to be extra careful about possible contamination in the sequences.
 325 To validate contamination in the genome assembly^{23,38,42}, we first mapped the MiSeq DNA-Seq reads
 326 to the assembled genome with BWA (v0.7.17)⁵⁴ and converted the output files with SAMtools (v1.7)

⁵⁵. Next, we conducted a BLASTn search against the NCBI nr database ^{56,57}. These results were submitted to BlobTools (v1.0) ⁴¹ for visualization. Contigs with GC content below 40% and above 60% and contigs identified to be bacterial were determined as contaminants and removed from the assembly. We validated the genome completeness with BUSCO (v3.0.2) using the eukaryote lineage gene set ⁵⁸.

For identifying genes, we first mapped the RNA-Seq data (active and tuned three replicates each) to the genome assembly with TopHat2 (v2.1.1) ⁵⁹. The resulting BAM files were submitted to BRAKER (v1.9) with default settings ⁶⁰. To remove possible bacterial contamination, we validated whether these query transcripts were included in the transcriptome assembly produced with Trinity v.2.5.1 ⁶¹ since polyA selection during the sequencing library construction process would filter non-polyA transcripts. Furthermore, we performed tBLASTn against the nr database and excluded nonmetazoan hits to remove possible eukaryotic contamination. We submitted these query sequences to a tBLASTn search against the transcriptome assembly and excluded transcripts that had identity below 99%. To annotate the predicted gene models, we performed similarity searches using BLASTp (E-value < 1e-5) against Swiss-Prot ⁶² and against tardigrade-predicted proteome sequences of *R. varieornatus* and *H. exemplaris* genomes (E-value < 1e-5). We also submitted the amino acid sequences to KEGG Automatic Annotation Server ^{63,64} to assign KEGG orthology IDs. De novo repeat region identification was conducted by RepeatScout ⁶⁵ and RepeatMasker ⁶⁶. We searched tRNA using tRNAscan ⁶⁷ and rRNA using RNAmmer ⁶⁸.

This final assembly data contained 42,608 predicted protein-coding genes with expression levels of > 1 transcript per million (TPM) in at least one of six RNA-Seq samples. A total of 57.6% of the genes showed BLASTp similarity against the Swiss-Prot database (E-value < 1e-5). To identify genes that were induced during anhydrobiosis, we quantified the gene expression (TPM) with Kallisto (v0.42.1) ⁶⁹. Additionally, we mapped the RNA-Seq reads to the coding sequences with BWA MEM and tested for differential expression using DESeq2 ^{54,70}. Transcripts with FDR below 0.05 were defined as DEGs.

Gene expansion and lost pathway analysis

To validate anhydrobiosis-related genes reported in previous tardigrade genomic studies ^{23,26}, KAAS orthologous annotation was used to initially screen for possible gene losses. For orthologs that were initially found to be missing in the *E. testudo* genome, we further confirmed gene loss by using orthologs in *R. varieornatus*, *H. exemplaris*, and Swiss-Prot for tBLASTn searches to the *E. testudo* genome.

To identify multicopy genes, we conducted multiple alignment and counted the copy number. Multiple alignment with *R. varieornatus* and *C. elegans* was carried out using MAFFT (v7.221) ^{71,72}.

Phylogenetic trees were constructed using FastTree (v2.1.10) with default options⁷³ and visualized with Interactive Tree of Life (iTOL)⁷⁴. Regarding extremely similar sequences, genes with BLASTn identity > 99% were regarded as identical artifact copies due to possible misassembly (Supplementary Table 5, Supplementary Fig. 1 phylogenetic tree).

Heat soluble proteomics

Heat soluble proteomics was conducted as previously described²⁷. Briefly, approximately 3,000 individuals were collected from the wild and cleaned as described above and homogenized using BioMasher II (Nippi) in PBS (Nippon Gene) on ice with protease inhibitors (Roche). The lysate was heated at 92°C for 20 min, and the soluble fraction was collected by taking the supernatant after centrifugation at 12,000 rpm for 20 min. Proteins were digested with trypsin, and tryptic peptides were separated and identified with an UltiMate 3000 nanoLC pump (Dionex Co., Sunnyvale, CA, USA) and an LTQ Orbitrap XL ETD (Thermo Electron, Waltham, MA, USA). Corresponding peptide sequences were retrieved from six frame translation data of our initial genome assembly using MASCOT software (Matrix Science)⁷⁵. Candidates were further screened with the following conditions: (1) lack of conservation in other tardigrades or metazoans and (2) high mRNA expression (TPM >100) in the tun state. We then predicted the structural features using the Fold Index⁷⁶ and DISOPRED⁷⁷.

Structural analysis of EtAHS proteins

The gene derived from g97955.t1, which encodes residues Thr29-Gln355 of EtAHS A protein, was cloned into pET28a (Novagen). The gene derived from g8031.t1, which encodes residues Gln20-Lys319 of EtAHS B protein, was cloned into modified pCold-1 (Takara Bio Inc.), in which the factor Xa cleavage site was replaced with tobacco etch virus (TEV) protease cleavage. The recombinant EtAHS proteins were expressed in the *E. coli* BL21(DE3) CodonPlus strain cultured in LB medium. The His₆-tagged fusion EtAHS A protein was purified from cell lysates with a Ni²⁺-nitrilotriacetic acid column (GE Healthcare). The fusion protein was cleaved by incubation with thrombin protease (Sigma Aldrich) and then purified with Superdex 200 pg (GE Healthcare). For preparation of EtAHS B protein, after lysis by sonication, the insoluble inclusion bodies were extensively washed with 20 mM Tris-HCl (pH 8.0) containing 150 mM NaCl and 2% Triton X-100 and then solubilized with 6 M guanidinium chloride, 50 mM Tris-HCl (pH 8.0), and 1 mM dithiothreitol. The solubilized proteins were refolded by dilution (0.1 mg/mL) in 20 mM Tris-HCl (pH 7.5), 100 mM NaCl, 3 mM reduced glutathione, and 0.3 mM oxidized glutathione at 4°C for 12 h. The His₆-tagged fusion protein was purified with a Ni²⁺-nitrilotriacetic acid column (GE Healthcare). The EtHAS B protein, from which

the N-terminal His₆-tag peptide was removed by TEV protease digestion, was further purified with a HiLoad Superdex 75 pg (GE Healthcare). The expression and purification of recombinant CAHS1 protein (*R. varieornatus*, UniProt ID: J7M799) were performed as described in the literature with slight modifications²⁷. For the production of ¹⁵N-labeled EtAHS A, EtAHS B, and CAHS1 proteins, cells were grown in M9 minimal media containing [¹⁵N]NH₄Cl (1 g/L). CD spectra were measured at room temperature on a JASCO J-720WI apparatus using a 1.0-mm path length quartz cell. The EtAHS A, EtAHS B, and CAHS1 proteins were dissolved at protein concentrations of 5.6 μM, 2.3 μM, and 6.8 μM, respectively, in 10 mM potassium phosphate buffer (pH 7.4) in the absence and presence of TFE.

NMR spectral measurements were made on a Bruker DMX-500 spectrometer equipped with a cryogenic probe. The probe temperature was set to 5°C. ¹⁵N-labeled EtAHS A, EtAHS B, and CAHS1 proteins were dissolved at a protein concentration of 0.15 mM in 10 mM potassium phosphate buffer (pH 7.4) containing 5% (v/v) ²H₂O.

Data availability

The genome/transcriptome sequencing data sets are available at the BioProject accession number PRJNA669587. The assembly and coding sequencing data are available at <https://doi.org/10.6084/m9.figshare.13060634>.

417 Reference

- 418 1 Keilin, D. The problem of anabiosis or latent life: history and current concept. *Proc. R. Soc. Lond. B Biol. Sci.*
419 **150**, 149-191 (1959).
- 420 2 Crowe, J., Hoekstra, F. & Crowe, L. Anhydrobiosis. *Annu. Rev. Physiol.* **54**, 579-599 (1992).
- 421 3 Mobjerg, N. *et al.* Survival in extreme environments - on the current knowledge of adaptations in tardigrades.
422 *Acta. Physiol. (Oxf)* **202**, 409-420 (2011).
- 423 4 Hengherr, S., Worland, M. R., Reuner, A., Brummer, F. & Schill, R. O. High-temperature tolerance in
424 anhydrobiotic tardigrades is limited by glass transition. *Physiol. Biochem. Zool.* **82**, 749-755 (2009).
- 425 5 Hengherr, S., Worland, M. R., Reuner, A., Brummer, F. & Schill, R. O. Freeze tolerance, supercooling points
426 and ice formation: comparative studies on the subzero temperature survival of limno-terrestrial tardigrades. *J.*
427 *Exp. Biol.* **212**, 802-807 (2009).
- 428 6 Jonsson, K. I., Rabbow, E., Schill, R. O., Harms-Ringdahl, M. & Rettberg, P. Tardigrades survive exposure to
429 space in low Earth orbit. *Curr. Biol.* **18**, R729-R731 (2008).
- 430 7 Ono, F. *et al.* Effect of ultra-high pressure on small animals, tardigrades and *Artemia*. *Cogent. Phys.* **3**,
431 1167575 (2016).
- 432 8 Ramløv, H. & Westh, P. Cryptobiosis in the Eutardigrade *Adorybiotus* (Richtersius) *coronifer*: Tolerance to
433 Alcohols, Temperature and de novo Protein Synthesis. *Zool. Anz.* **240**, 517-523 (2001).
- 434 9 Horikawa, D. *et al.* Analysis of DNA repair and protection in the Tardigrade *Ramazzottius varieornatus* and
435 *Hypsibius dujardini* after exposure to UVC radiation. *PLoS One* **8**, e64793 (2013).
- 436 10 Altiero, T., Guidetti, R., Caselli, V., Cesari, M. & Rebecchi, L. Ultraviolet radiation tolerance in hydrated and
437 desiccated eutardigrades. *J. Zool. Syst. Evol. Res.* **49**, 104-110 (2011).
- 438 11 Jonsson, K. I., Hygum, T. L., Andersen, K. N., Clausen, L. K. & Mobjerg, N. Tolerance to Gamma Radiation
439 in the Marine Heterotardigrade, *Echiniscoides sigismundi*. *PLoS One* **11**, e0168884 (2016).
- 440 12 Honda, Y., Tanaka, M. & Honda, S. Trehalose extends longevity in the nematode *Caenorhabditis elegans*.
441 *Aging Cell* **9**, 558-569 (2010).
- 442 13 Moore, D. S., Hansen, R. & Hand, S. C. Liposomes with diverse compositions are protected during desiccation
443 by LEA proteins from *Artemia franciscana* and trehalose. *Biochim. Biophys. Acta.* **1858**, 104-115 (2016).
- 444 14 Sakurai, M. *et al.* Vittrification is essential for anhydrobiosis in an African chironomid, *Polypedium*
445 *vanderplanki*. *Proc. Natl. Acad. Sci. USA* **105**, 5093-5098 (2008).
- 446 15 Crowe, J., Clegg, J. & Crowe, L. *Anhydrobiosis: the water replacement hypothesis.*, (Springer, Boston, MA,
447 1998).
- 448 16 Caramelo, J. J. & Iusem, N. D. When cells lose water: Lessons from biophysics and molecular biology. *Prog.*
449 *Biophys. Mol. Biol.* **99**, 1-6 (2009).
- 450 17 Goyal, K. *et al.* Transition from natively unfolded to folded state induced by desiccation in an anhydrobiotic
451 nematode protein. *J. Biol. Chem.* **278**, 12977-12984 (2003).
- 452 18 Goldgur, Y. *et al.* Desiccation and zinc binding induce transition of tomato abscisic acid stress ripening 1, a
453 water stress- and salt stress-regulated plant-specific protein, from unfolded to folded state. *Plant Physiol.* **143**,
454 617-628 (2007).
- 455 19 Oliveira, E. *et al.* LC-MSMS identification of *Arabidopsis thaliana* heat-stable seed proteins: enriching for
456 LEA-type proteins by acid treatment. *J. Mass Spectrom.* **42**, 1485-1495 (2007).
- 457 20 Hand, S. C., Menze, M. A., Toner, M., Boswell, L. & Moore, D. LEA proteins during water stress: not just for
458 plants anymore. *Annu. Rev. Physiol.* **73**, 115-134 (2011).
- 459 21 Hengherr, S., Heyer, A. G., Kohler, H. R. & Schill, R. O. Trehalose and anhydrobiosis in tardigrades—
460 evidence for divergence in responses to dehydration. *FEBS J.* **275**, 281-288 (2008).
- 461 22 Jönsson, K. & Persson, O. Trehalose in Three Species of Desiccation Tolerant Tardigrades. *Open Zool. J.* **3**, 1-
462 5 (2010).
- 463 23 Yoshida, Y. *et al.* Comparative genomics of the tardigrades *Hypsibius dujardini* and *Ramazzottius*
464 *varieornatus*. *PLoS Biol.* **15**, e2002266 (2017).
- 465 24 Förster, F. *et al.* Tardigrade workbench: comparing stress-related proteins, sequence-similar and functional
466 protein clusters as well as RNA elements in tardigrades. *BMC Genomics* **10**, 1-10 (2009).
- 467 25 Kamilari, M., Jorgensen, A., Schiott, M. & Mobjerg, N. Comparative transcriptomics suggest unique
468 molecular adaptations within tardigrade lineages. *BMC Genomics* **20**, 607 (2019).
- 469 26 Hashimoto, T. *et al.* Extremotolerant tardigrade genome and improved radiotolerance of human cultured cells
470 by tardigrade-unique protein. *Nat. Commun.* **7**, 12808 (2016).
- 471 27 Yamaguchi, A. *et al.* Two novel heat-soluble protein families abundantly expressed in an anhydrobiotic
472 tardigrade. *PLoS One* **7**, e44209 (2012).
- 473 28 Tanaka, S. *et al.* Novel mitochondria-targeted heat-soluble proteins identified in the anhydrobiotic Tardigrade
474 improve osmotic tolerance of human cells. *PLoS One* **10**, e0118272 (2015).
- 475 29 Bemm, F. *et al.* Draft genome of the Eutardigrade *Milnesium tardigradum* sheds light on ecdysozoan
476 evolution. *bioRxiv* (2017).

- 30 Kondo, K., Kubo, T. & Kunieda, T. Suggested Involvement of PP1/PP2A Activity and De Novo Gene Expression in Anhydrobiotic Survival in a Tardigrade, *Hypsibius dujardini*, by Chemical Genetic Approach. *PLoS One* **10**, e0144803 (2015).
- 31 Watanabe, M., Kikawada, T., Minagawa, N., Yukuhiro, F. & Okuda, T. Mechanism allowing an insect to survive complete dehydration and extreme temperatures. *J. Exp. Biol.* **205**, 2799-2802 (2002).
- 32 Erkut, C. *et al.* Molecular strategies of the *Caenorhabditis elegans* dauer larva to survive extreme desiccation. *PLoS One* **8**, e82473 (2013).
- 33 Wright, J. Desiccation Tolerance and Water-Retentive Mechanisms in Tardigrades. *J. Exp. Biol.*, 267-292 (1989).
- 34 Persson, D. *et al.* Extreme stress tolerance in tardigrades: surviving space conditions in low earth orbit. *J. Zool. Syst. Evol. Res.* **49**, 90-97 (2011).
- 35 Heidemann, N. W. T. *et al.* Osmotic stress tolerance in semi-terrestrial tardigrades. *Zool. J. Linn. Soc.* **178**, 912-918 (2016).
- 36 Hygum, T. L. *et al.* Comparative Investigation of Copper Tolerance and Identification of Putative Tolerance Related Genes in Tardigrades. *Front. Physiol.* **8**, 95 (2017).
- 37 Jorgensen, A., Faurby, S., Hansen, J. G., Mobjerg, N. & Kristensen, R. M. Molecular phylogeny of Arthrotardigrada (Tardigrada). *Mol. Phylogenet. Evol.* **54**, 1006-1015 (2010).
- 38 Arakawa, K. No evidence for extensive horizontal gene transfer from the draft genome of a tardigrade. *Proc. Natl. Acad. Sci. USA* **113**, E3057 (2016).
- 39 Arakawa, K., Yoshida, Y. & Tomita, M. Genome sequencing of a single tardigrade *Hypsibius dujardini* individual. *Sci. Data* **3**, 160063 (2016).
- 40 Yoshida, Y. *et al.* Ultralow Input Genome Sequencing Library Preparation from a Single Tardigrade Specimen. *J. Vis. Exp.* (2018).
- 41 Laetsch, D. & Blaxter, M. BlobTools: Interrogation of genome assemblies [version 1; referees: 2 approved with reservations]. *F1000Research* **6** (2017).
- 42 Koutsovoulos, G. *et al.* No evidence for extensive horizontal gene transfer in the genome of the tardigrade *Hypsibius dujardini*. *Proc. Natl. Acad. Sci. USA* **113**, 5053-5058 (2016).
- 43 Cornette, R. & Kikawada, T. The induction of anhydrobiosis in the sleeping chironomid: current status of our knowledge. *IUBMB Life* **63**, 419-429 (2011).
- 44 Evangelista, C. C. S. *et al.* Multiple genes contribute to anhydrobiosis (tolerance to extreme desiccation) in the nematode *Panagrolaimus superbus*. *Genet. Mol. Biol.* **40**, 790-802 (2017).
- 45 Rytönen, K. T., Williams, T. A., Renshaw, G. M., Primmer, C. R. & Nikinmaa, M. Molecular evolution of the metazoan PHD-HIF oxygen-sensing system. *Mol. Biol. Evol.* **28**, 1913-1926 (2011).
- 46 Graham, A. M. & Barreto, F. S. Loss of the HIF pathway in a widely distributed intertidal crustacean, the copepod *Tigriopus californicus*. *Proc. Natl. Acad. Sci. USA* **116**, 12913-12918 (2019).
- 47 Sasaki, K., Christov, N. K., Tsuda, S. & Imai, R. Identification of a novel LEA protein involved in freezing tolerance in wheat. *Plant Cell Physiol.* **55**, 136-147 (2014).
- 48 Battaglia, M. & Covarrubias, A. A. Late Embryogenesis Abundant (LEA) proteins in legumes. *Front. Plant Sci.* **4**, 190 (2013).
- 49 Piszkiwicz, S. *et al.* Protecting activity of desiccated enzymes. *Protein Sci.* **28**, 941-951 (2019).
- 50 Johnson, C. G. The maintenance of high atmospheric humidities for entomological work with glycerol-water mixtures. *Ann Appl Biol* **27**, 295-299 (1940).
- 51 Simon, A. *FastQC a quality tool for high-throughput sequence data.*, 2010).
- 52 Vurture, G. W. *et al.* GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **33**, 2202-2204 (2017).
- 53 Zimin, A. V. *et al.* The MaSuRCA genome assembler. *Bioinformatics* **29**, 2669-2677 (2013).
- 54 Jo, H. & Koh, G. Faster single-end alignment generation utilizing multi-thread for BWA. *Biomed. Mater. Eng.* **26 Suppl 1**, S1791-1796 (2015).
- 55 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
- 56 Altschul, S. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389-3402 (1997).
- 57 Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **35**, D61-65 (2007).
- 58 Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210-3212 (2015).
- 59 Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
- 60 Hoff, K. J., Lange, S., Lomsadze, A., Borodovsky, M. & Stanke, M. BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* **32**, 767-769 (2016).
- 61 Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644-652 (2011).

- 62 Bairoch, A., Boeckmann, B., Ferro, S. & Gasteiger, E. Swiss-Prot: juggling between evolution and stability. *Brief. Bioinform.* **5**, 39-55 (2004).
- 63 Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27-30 (2000).
- 64 Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. & Kanehisa, M. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* **35**, 182-185 (2007).
- 65 Price, A. L., Jones, N. C. & Pevzner, P. A. De novo identification of repeat families in large genomes. *Bioinformatics* **21 Suppl 1**, i351-358 (2005).
- 66 Smit, A., Hubley, R. & Green, P. RepeatMasker Open-4.0. <http://www.repeatmasker.org>. (2013–2015).
- 67 Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 955-964 (1997).
- 68 Lagesen, K. *et al.* RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* **35**, 3100-3108 (2007).
- 69 Bray, N., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525-527 (2016).
- 70 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
- 71 Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059-3066 (2002).
- 72 Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772-780 (2013).
- 73 Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26**, 1641-1650 (2009).
- 74 Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**, 127-128 (2007).
- 75 Perkins, D. N., Pappin, D. J. C., Creasy, D. M. & Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551-3567 (1999).
- 76 Prilusky, J. *et al.* FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics* **21**, 3435-3438 (2005).
- 77 Ward, J. J., McGuffin, L. J., Bryson, K., Buxton, B. F. & Jones, D. T. The DISOPRED server for the prediction of protein disorder. *Bioinformatics* **20**, 2138-2139 (2004).

573 Acknowledgments

574 The authors are grateful to Yuki Takai (IAB) and Naoko Ishii (IAB) for providing experimental
 575 support and to Konosuke M. Ii (IAB) for collecting *E. testudo* on proteome analysis. The authors also
 576 thank Yudai Sasaki (NCU) for his contribution during the early stage of this study, Dr Tadashi Satoh,
 577 Kumiko Hattori (NCU) and Dr Ean Wai Goh (ExCELLS) for spectroscopic measurements. This
 578 research was supported by Grants-in-Aid for Scientific Research KAKENHI of the Japan Society for
 579 the Promotion of Science (No. 17H03620), the Nanotechnology Platform Program (Molecule and
 580 Material Synthesis, JPMX09S19MS0051) of MEXT, Japan, the Joint Research of the Exploratory
 581 Research Center on Life and Living Systems (ExCELLS program No.18-207, 19-208, 19-501), and
 582 research funds from the Yamagata prefectural government and Tsuruoka city.

585 Contributions

586 K.A. conceived and designed the project. Y.M. conducted the RNA sequencing, data analysis and
 587 bioinformatics analysis. M.Y.U. and K.K. performed experiments of NMR and CD spectrum. M.F.
 588 conducted proteome analysis. K.A performed genome sequencing. The study was supervised by
 589 K.A. and M.T. The manuscript was drafted by Y.M, M.Y.U and K.A, and all authors reviewed and
 590 approved the final manuscripts.

592 Competing interests

593 The authors declare no competing interests.

594

595

Figures

Figure 1. Gene conservation of *Echiniscus testudo* and other tardigrades.

(a) Gene distribution of ortholog genes by BLASTp search against Swiss-Prot database. (b) Percentage of gene conservation on E-value < 1E-25. The genomic and transcriptomic data used are shown in the Methods section. Et: *Echiniscus testudo*, Es: *Echiniscoides sigismundi*, Rc: *Richtersius coronifer*, He: *Hypsibius exemplaris*, Rv: *Ramazzottius varieornatus*.

Figure 2. Gene loss and duplication in tardigrades.

Coral orange and gray boxes represent genes conserved and lost in all 5 species used in this study. Furthermore, yellow boxes and blue boxes indicate genes conserved only in eutardigrades and in heterotardigrades, respectively. The numbers on the left of the boxes show copy numbers of multiple copy genes in *E. testudo*. AMPK, 5' adenosine monophosphate-activated protein kinase; Apaf-1, apoptotic protease-activating factor 1; ATM, ataxia-telangiectasia mutated; ATR, ataxia telangiectasia and Rad3-related protein; BCS1, mitochondrial chaperone BCS1; CAHS, cytosolic abundant heat soluble; CASP3/7/8, caspase 3/7/8; CBP, CREB binding protein; CHK1/2, checkpoint kinase 1/2; Dsup, damage suppressor; Fas, tumor necrosis factor receptor superfamily member 6; FADD, Fas-associated death domain; Gadd45, growth arrest and DNA damage-inducible; GST, glutathione S-transferase; Hif1α, Hypoxia-inducible factor 1-alpha; Hif1β, aryl hydrocarbon receptor nuclear translocator; HSP, heat shock protein; LEAM, late embryogenesis abundant protein mitochondrial; MAHS, mitochondrial abundant heat soluble; mTOR, mechanistic target of rapamycin; PHD, plant homeodomain; p300, Histone acetyltransferase; Rheb, GTP-binding protein Rheb; SAHS, secretory abundant heat soluble; SIRT1, sirtuin 1; SOD, superoxide dismutase; TSC1/2, tuberous sclerosis 1/2; VHL, von Hippel-Lindau tumor suppressor.

Figure 3. Folding prediction of EtAHS proteins.

Folding of (a) EtAHS A, (b) EtAHS B and (c) CAHS1 proteins. The upper tier shows Fold Index and the lower tier shows DISOPRED folding. In Foldindex, positive (green area) and negative (red area) numbers represent ordered and disordered protein, respectively. Amino acids suggested as being folded are shown in the green area and unfolded in the red area. In DISOPRED, the blue line indicates disordered state, and the orange line indicates protein binding.

Figure 4. Structural analysis of EtAHS proteins

^1H - ^{15}N HSQC spectra of (a) EtAHS A, (b) EtAHS B, and (c) CAHS1 proteins. CD spectra of (d) EtAHS A, (e) EtAHS B, and (f) CAHS1 proteins in the absence (black) and presence of 10% (blue), 30% (green), 50% (orange), and 70% (red) (v/v) TFE.

Tables

Genome Statistics	Before screening (Original)	After screening
Sequencer	MiSeq	
Scaffold number	107,804	30,095
Total scaffold length (bp)	388,783,196	153,718,311
Average scaffold length (bp)	3,606	5,107
Longest scaffold length (bp)	326,619	41,023
Shortest scaffold length (bp)	1,000	1,001
N50	4,891	6,674
BUSCO completeness (%)	89.4	87.8

Table 1. Statistics of genome assembly.

Gene ID	TPM (average)		Fold Index	Folding type
	Active	Tun		
g97955.t1	583	400	-0.210	EtAHS A
g110942.t1	504	569	-0.204	
g158834.t1	715	825	-0.215	
g8031.t1	4,146	2,538	-0.092	EtAHS B
g94762.t1	4,171	2,460	-0.101	
g184550.t1	3,749	2,258	-0.082	

Table 1. Gene expression and Fold Index of EtAHS proteins.

Shown TPM values in Active and Tun are an average of 3 replicates, respectively.