


RESEARCH ARTICLE

Open Access



Multimomics study of a heterotardigrade, *Echiniscus testudo*, suggests the possibility of convergent evolution of abundant heat-soluble proteins in Tardigrada

Yumi Murai^{1,2}, Maho Yagi-Utsumi^{3,4,5}, Masayuki Fujiwara^{1,2}, Sae Tanaka³, Masaru Tomita^{1,2,6}, Koichi Kato^{3,4,5} and Kazuharu Arakawa^{1,2,3,6*} 

Abstract

Background: Many limno-terrestrial tardigrades can enter an ametabolic state, known as anhydrobiosis, upon desiccation, in which the animals can withstand extreme environments. Through genomics studies, molecular components of anhydrobiosis are beginning to be elucidated, such as the expansion of oxidative stress response genes, loss of stress signaling pathways, and gain of tardigrade-specific heat-soluble protein families designated CAHS and SAHS. However, to date, studies have predominantly investigated the class Eutardigrada, and molecular mechanisms in the remaining class, Heterotardigrada, still remains elusive. To address this gap in the research, we report a multimomics study of the heterotardigrade *Echiniscus testudo*, one of the most desiccation-tolerant species which is not yet culturable in laboratory conditions.

Results: In order to elucidate the molecular basis of anhydrobiosis in *E. testudo*, we employed a multi-omics strategy encompassing genome sequencing, differential transcriptomics, and proteomics. Using ultra-low input library sequencing protocol from a single specimen, we sequenced and assembled the 153.7 Mbp genome annotated using RNA-Seq data. None of the previously identified tardigrade-specific abundant heat-soluble genes was conserved, while the loss and expansion of existing pathways were partly shared. Furthermore, we identified two families novel abundant heat-soluble proteins, which we named *E. testudo* Abundant Heat Soluble (EtAHS), that are predicted to contain large stretches of disordered regions. Likewise the AHS families in eutardigrada, EtAHS shows structural changes from random coil to alpha-helix as the water content was decreased in vitro. These characteristics of EtAHS proteins are analogous to those of CAHS in eutardigrades, while there is no conservation at the sequence level.

Conclusions: Our results suggest that Heterotardigrada have partly shared but distinct anhydrobiosis machinery compared with Eutardigrada, possibly due to convergent evolution within Tardigrada. (276/350).

Keywords: *Echiniscus testudo*, Heterotardigrada, Heat-soluble protein, Anhydrobiosis, Multimomics

* Correspondence: gaou@sfc.keio.ac.jp

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

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

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Water is an essential universal solvent in living organisms that harbors cellular biochemical reactions. However, many limno-terrestrial tardigrades are able to enter an ametabolic state, known as anhydrobiosis, when they face desiccation, and these organisms can subsequently tolerate almost complete water loss while being able to return to their natural active state when rehydrated [1–3]. In anhydrobiosis, tardigrades are capable of tolerating many extreme environments, such as high and low temperatures [4, 5], space vacuum [6], high pressure [7], and high concentrations of organic solvent [8], as well as insults from ultraviolet rays and over 4000 Gy of gamma radiation [9–11].

Early studies on the molecular mechanisms underlying anhydrobiosis initially targeted *Caenorhabditis elegans* [12], *Artemia salina* [13], and *Polypedilum vanderplanki* [14], where the disaccharide trehalose was identified as one of the key compounds, being accumulated at up to 20% of the dry weight during desiccation in *P. vanderplanki*. Trehalose is suggested to protect intracellular components, including membranes and proteins, through water replacement and vitrification [14, 15]. Another of the key factors of the anhydrobiotic machinery is the hydrophilic late embryogenesis abundant (LEA) protein superfamily, which has been identified in both plants and animals [16]. This widespread protein family exhibits little sequence conservation among groups, but they often show similar structural characteristics, where the protein is mostly natively unfolded in the hydrated state and later folds upon desiccation [17, 18]. Another hallmark is the extreme hydrophilic nature of this protein, which maintains solubility even after 10 min of heat treatment at 100 °C [19]. The contribution of LEA proteins in desiccation tolerance is reported to be multifactorial, encompassing protecting membranes, inhibiting the aggregation of proteins, stabilizing cellular components, including vitrified sugar glasses, and serving as a hydration buffer [20].

Trehalose and LEA proteins, however, are not the primary machinery of tardigrade anhydrobiosis. Trehalose levels account for only several percent at maximum, and some species even lack measurable amounts of trehalose [14, 21–23]. LEA proteins are observed in the tardigrade transcriptome but are not among the most abundant group of transcripts, suggesting the existence of unique tardigrade components [24, 25]. Desiccation tolerance and anhydrobiosis is known to be acquired in multiple lineages, with diverse machineries but sometimes with possible convergent evolution [26–28]. With the emerging availability of genomic resources [23, 29], these tardigrade-specific anhydrobiosis-related proteins were first screened using heat solubility assays in the eutardigrade *Ramazzottius varieornatus* [30, 31]. While no sequence similarity to LEA was observed, characteristics of

the cytoplasmic abundant heat soluble (CAHS) protein, such as its high abundance, heat solubility, intrinsically unstructured nature, and structural change to alpha helix upon water loss, mirror that of LEA proteins [30]. These tardigrade-specific abundant heat-soluble genes initially appeared to be highly conserved among different tardigrades, where the CAHS gene was conserved all across the class eutardigrada in Hypsibiidae, Macrobiotidae, and Milnesiidae [23, 25, 30, 32]. 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 anhydrobiosis [33] and prepares homologous proteins by de novo expression during this period [23], similar to the anhydrobiosis induction method observed in the arthropod *P. vanderplanki* [34] and nematode *C. elegans* [35]. 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 machineries [25].

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* [36]; 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 vacuum [37], osmotic pressure [38], and high concentrations of environmental toxicant copper ions, which may increase reactive oxygen species [39]. 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 Tardigrada [40]. 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 methodology [41–43] to elucidate the molecular components of anhydrobiosis in this species.

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 6674 bp, respectively (Table 1, Table S1). 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 visualizations [44], 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% (Fig. S1). 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 (Table S2). Forty-two megabases (27%) of this genome was inferred to consist of unknown repeats. We further predicted 73 ribosomal genes and 1081 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 (Table S3, Fig. S2). 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 (Table S1).

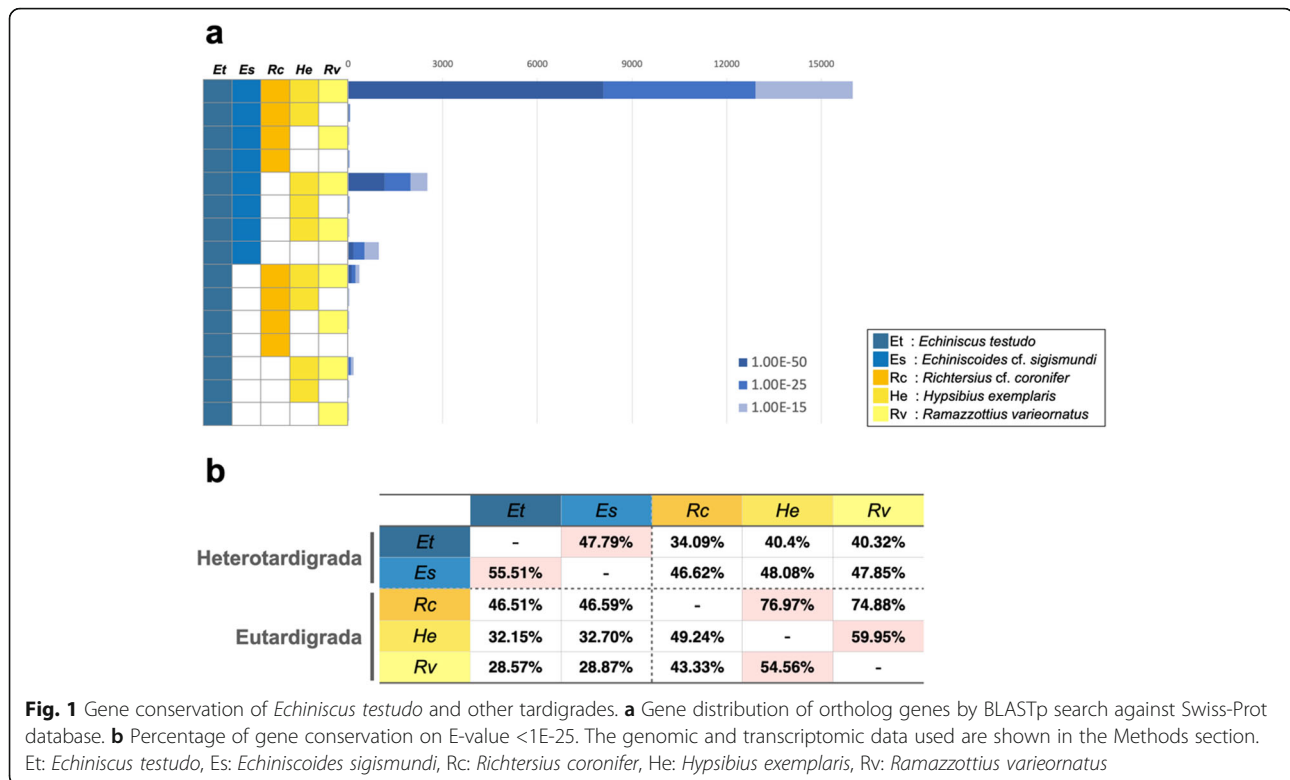
Next, we compared the *E. testudo* gene set with other sequenced tardigrades: transcriptome data of Heterotardigrada *E. cf. sigismundi* and Eutardigrada *Richtersius coronifer*, and genome data of Eutardigrada *R. varieornatus* and *H. exemplaris*. A total of 12,917 genes ($1e-25$) of *E. testudo* were conserved in all 5 species, and approximately 529 genes were shared only among the heterotardigrades *E. testudo* and *E. cf. sigismundi* (Fig. 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 (Fig. 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 tolerance [23]; 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 H_2O_2 , generating stress-responsive pathways, as well as gene duplications in antioxidative stress genes, suggesting intensive genomic adaptations for coping with oxidative stress [23, 29]. 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

Table 1 Statistics of genome assembly

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

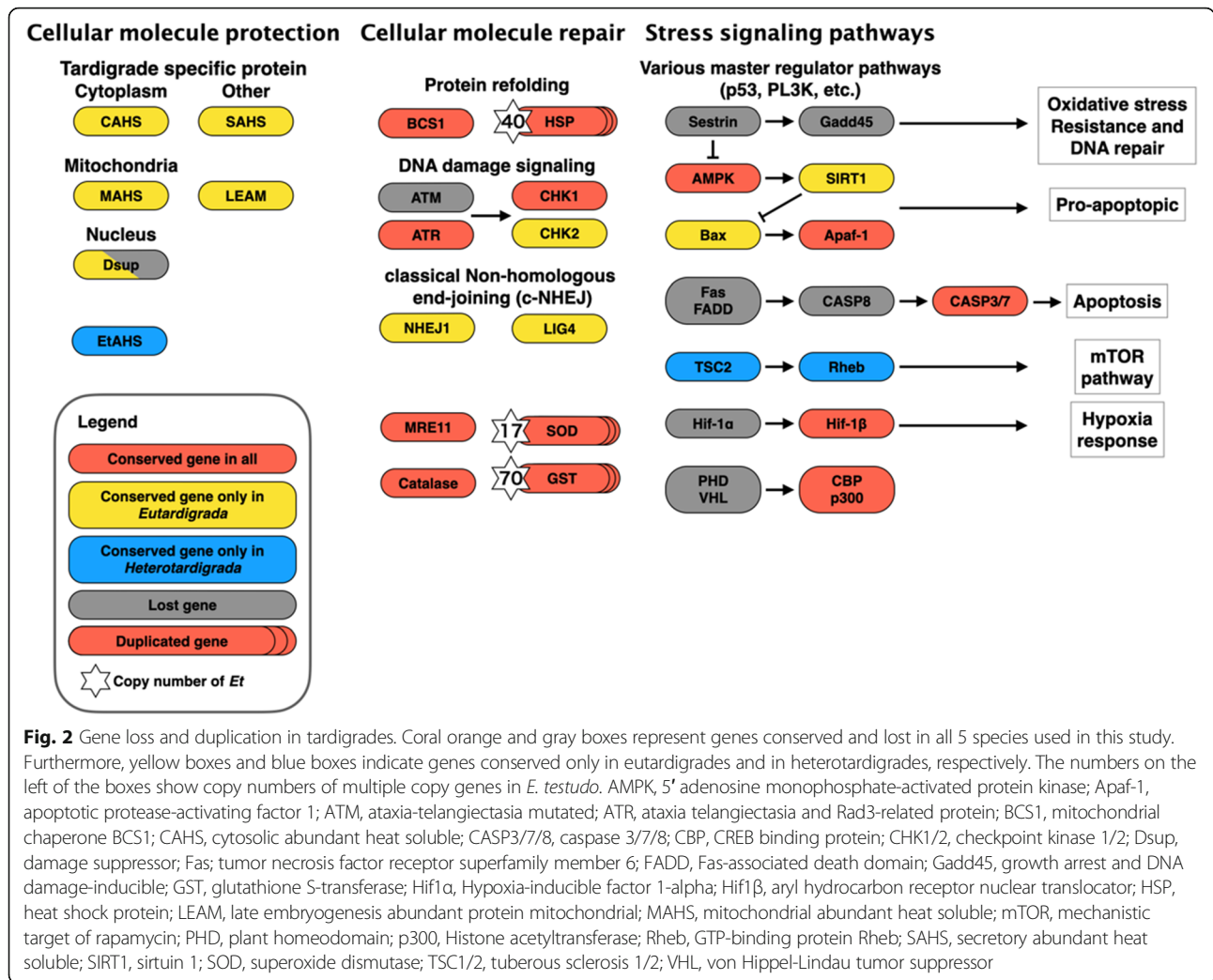


nearly identical orthologs (>99% identity) and possible artifacts (Fig. S3). As a result, we identified 1, 17, 70 and 40 orthologs of catalase, SODs, GST and HSP, respectively (Fig. 2). We also observed that several genes related to stress signaling, *TSC2* and *Rheb*, lost in the Eutardigrada lineage were conserved in *E. testudo* (Fig. 2 and Table S4), 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 abundant heat-soluble 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, Fig. 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 ($q\text{-value} < 0.01$) candidates being identified (Table S5 and Fig. S4). 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* [29].

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 desiccation [30]. Therefore, we considered heat solubility screening to obtain analogous proteins. Following heat-soluble proteomics (Table S6), 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 (Table S7). These



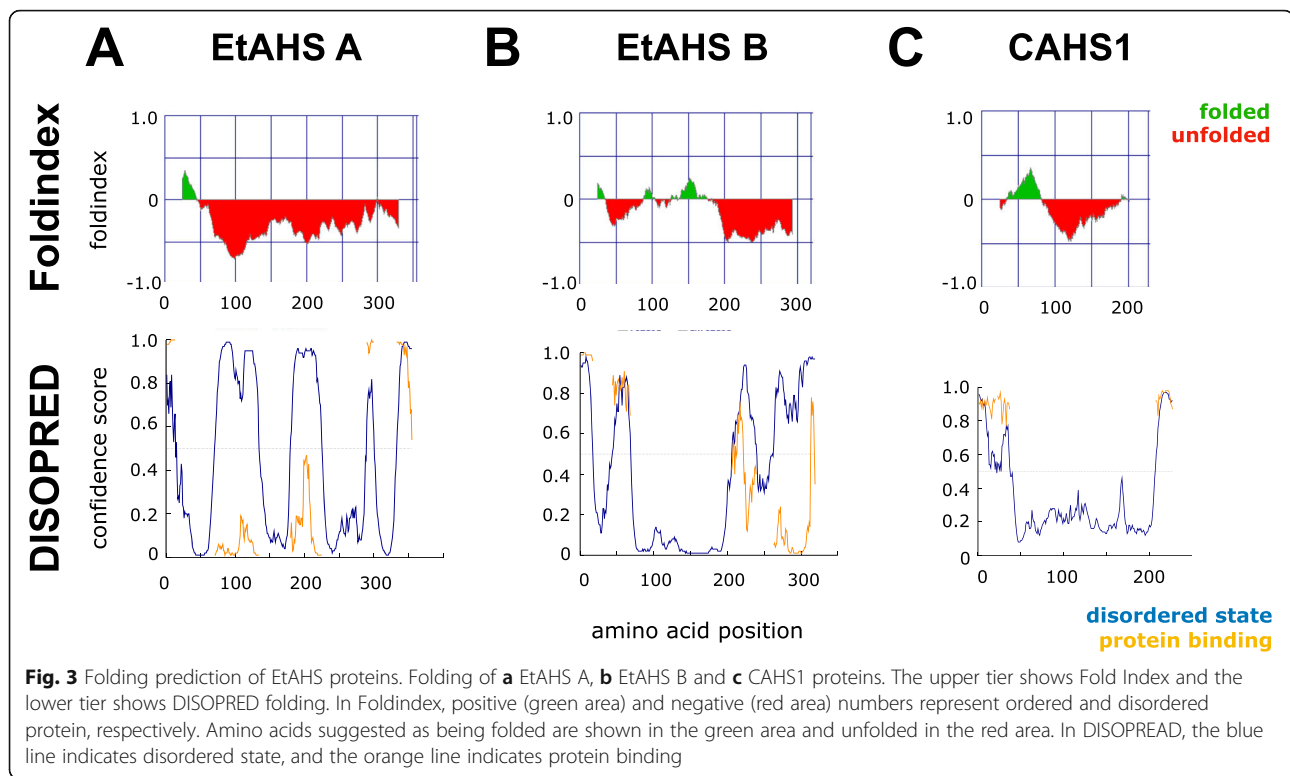
candidates were further screened for their disordered nature with negative fold index values, as these genes are predicted to have large intrinsically unstructured regions. This approach resulted in six genes of two fold index patterns, containing a high proportion of predicted

alpha helical regions, similar to those of CAHS1 (Table 2, Fig. 3). Although there is no sequence similarity (Fig. S5), these *Echiniscus*-specific, heat-soluble, constitutively highly expressed (these genes were not DEGs in the above analysis and are highly expressed both in hydrated

Table 2 Gene expression and Fold Index of EtAHS proteins

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

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



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.

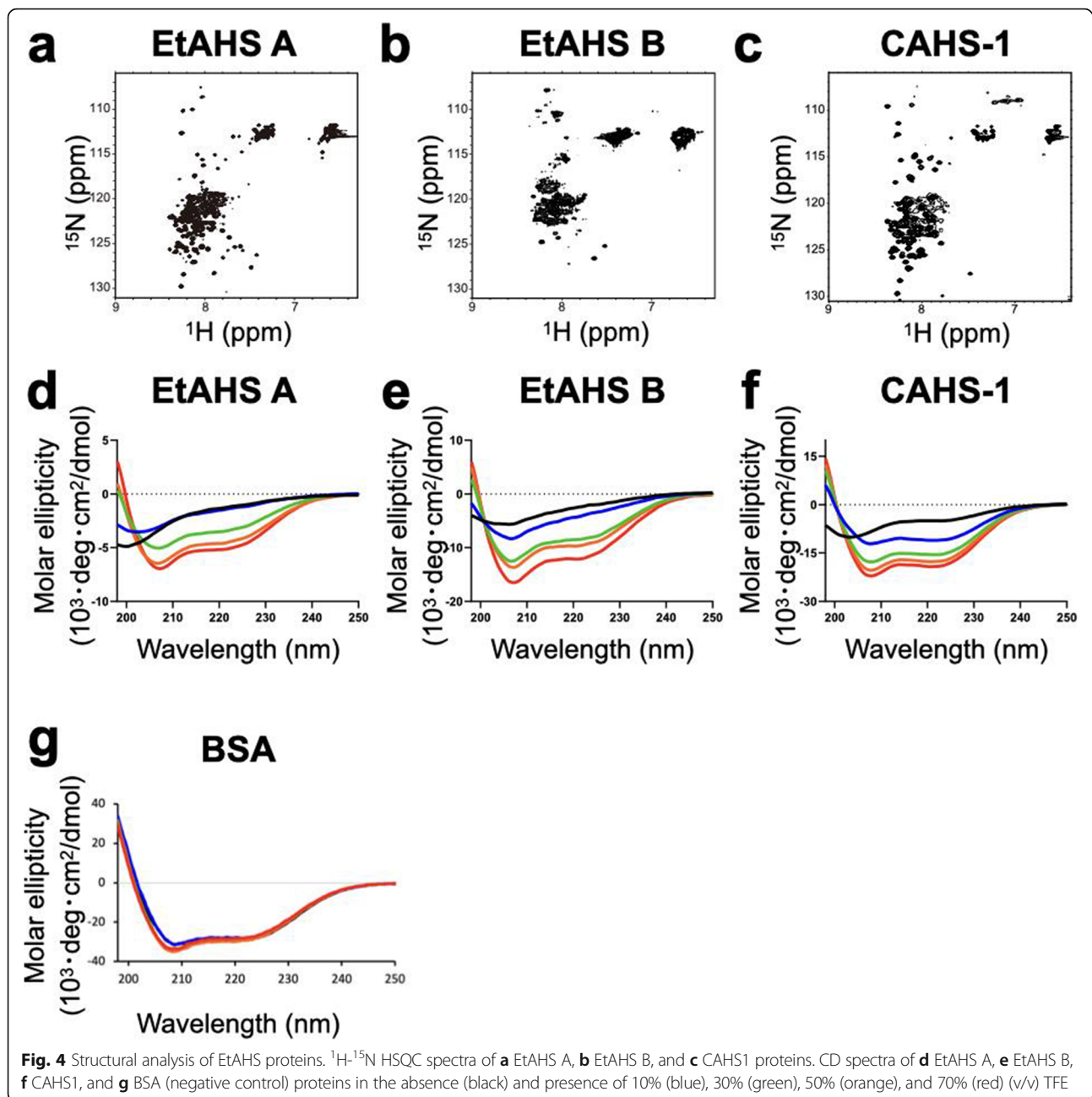
Structural analysis of EtAHS proteins

We next performed structural characterization of the recombinant EtAHS proteins using NMR and CD spectroscopy (Fig. 4). The ^1H - ^{15}N 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 previously [41–43] to achieve whole genome sequencing from a single specimen of tardigrade containing less than 50 pg of DNA collected from the wild. We employed an ultra-low-input library sequencing protocol as our sequencing method instead of pooling large number of specimens, in order to minimize contamination, which has been a problem in tardigrade sequencing, especially for specimens caught in the wild as in this work [45]. Although *E. testudo* is parthenogenetic and thus often found as clonal lineages in the wild, sequencing from a single animal can nevertheless limit natural variation and further eliminate contamination to some extent by manual cleaning and observation under the microscope. Typical low input Illumina sequencing requires around 1 ng of input DNA, which correspond to dozens of specimens, which would increase the possibility of contamination. Long-read sequencing (e.g. Oxford Nanopore) further requires tens of thousands of animals, which is not practical to collect from the wild, and whole-genome amplification has not been successful in our experience. For those reasons, ultra-low-input library sequencing was used in this study, but additional strong precautions should be taken to minimize the possibility of contamination in such field-collected samples [23, 41, 46].



Removal of contamination was only performed after the assembly, because genomic resource for tardigrades is still very much limited, and since this is the first report in the class Heterotardigrada, and identification at the sequenced read level likely result in misidentifications. Moreover, the tardigrade genome is known to have some degree of horizontally transferred genes from prokaryotes. For example, the catalase of the tardigrades is known to be the bacterial type [29], and pre-assembly screening would result in the loss of such genes and subsequently results in an insufficient assembly. 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. Since the genome assembly is quite fragmented in its current form, the evidence from the genome sequence alone may not be definitive to observe gene duplication or loss, even though the BUSCO completeness statistics is sufficiently high. Therefore, when discussing the duplication or loss of genes, we always validated the results in combination with the RNA-Seq data, to test consistencies in genome and transcriptome-based gene copy number detection.

Duplication of genes is further confirmed by multiple alignment and phylogenetic trees, to eliminate possible false positives. However, since we take a conservative methodology gene duplication event could be underestimated when sequence similarity is unusually high; for example, if there is a paralog that is a 99.9% identical in its nucleotide sequence, such as the Mre11 gene copies in *R. varieornatus* [29], there is a possibility of miscounting (assuming one instead of two). 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 reported [25], 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 (ROS) 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 [47, 48]. 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 species [49]. 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 anhydrobiosis [29]. 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 rarely 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 h [50]. 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 anhydrobiosis [23]. Therefore, analogous function is expected in *E. testudo* but without sequence similarity or homology, as in the case of different LEA groups [51]. 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 abundant heat-soluble 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 [36]. 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 abundant heat-soluble proteins are constitutively highly expressed [23].

Therefore, we performed heat solubility proteomics assays to identify hydrophilic proteins, following the procedures established for the identification of LEA and CAHS proteins [19, 30]. 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 analogous [30]. 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 [51, 52]; 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 of this work and is a possible direction for future research. Moreover, the role played by EtAHS or CAHS in anhydrobiosis has not been elucidated to date. Few studies have investigated this topic, and they have obtained ambiguous results [53]. Functional experiments would be required as a future work to fully confirm the contribution of EtAHS in the desiccation tolerance of heterotardigrada. However, direct functional experiments are currently not realistic due to the microscopic size of these species as well as due to the lack of culture methods in laboratory conditions, and knock-in experiments of CAHS and SAHS genes do not immediately show any observable phenotype in human culture cells [30]. Specific molecular mechanisms of LEA and CAHS are yet to be clarified, which also precludes immediate *in vitro* assays. On the other hand, the combined features of EtAHS revealed in this study, namely, high abundance, entirely unstructured nature (not just partially disordered), heat solubility, existence of multiple paralogs, unique-ness in the specific species, mirrors those of LEA, CAHS, and SAHS when these proteins were first identified, and are strongly suggestive of its relationship with anhydrobiosis. Further studies on this subject are also necessary to fully elucidate the molecular mechanisms underlying anhydrobiosis in tardigrades. Nevertheless, the existence of analogous proteins without any sequence similarity suggests the importance of this class of proteins in anhydrobiosis and that tardigrades have independently evolved these protein families at least two times, once in Heterotardigrada and once in Eutardigrada, to enable anhydrobiosis. This phenomenon may represent an example of convergent evolution and highlights the complex evolutionary history of anhydrobiosis within the phylum Tardigrada.

Conclusions

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

Tardigrada, which may represent an example of convergent evolution.

Methods

Animals

We collected *E. testudo* from moss collected from Otsuka-machi, Tsuruoka city, Japan (38°44'23"N, 139°48'28"E). The moss was soaked in water for more than 3 h, and *E. testudo* specimens were picked and collected on an agar-coated plate. The species was identified using assembled contigs corresponding to 18S rRNA, 28S rRNA, COI, ITS1, and ITS2, confirming 99.2 ~ 99.8% identity to previously published sequences available in GenBank. *E. testudo* was allowed to walk around the agar plate for two days, with rigorous washing and renewed plate each day to remove surface and gut contaminants. Tun (anhydrobiotic) state was induced by suspending 20–21 animals in 100 µl of distilled water and by placing it on a 2-cm × 2-cm mesh filter on a paper towel (ASONE). The animals were desiccated at 35% RH for 24 h controlled by 85% glycerol solution [54], followed by 0% RH at room temperature for complete desiccation. Anhydrobiotic success was determined by an over 90% recovery rate after 24 h.

Genome and transcriptome sequencing

We used the ultra-low-input library sequencing protocol established in our previous study [42, 43]. In brief, genomic DNA was extracted using a Quick-gDNA Microprep Kit (Zymo Research) after lysis with two freeze-thaw cycles of –80 °C and 37 °C incubation, and the extracted DNA was sheared to 550-bp target fragments with Covaris M220. The Illumina sequencing library was constructed with a ThruPLEX DNA-Seq kit (Rubicon Genomics), and the purified library was quantified using a Qubit Fluorometer (Life Technologies). The sequencing library was sequenced on the MiSeq platform (Illumina) with 300 bp paired settings. For RNA-Seq, the total RNA was extracted from active and tun samples (3 replicates) using Direct-Zol RNA MicroPrep Kits (Zymo Research), and the sequencing library was constructed using SMARTer v4 Ultra Low Input RNA Kit for Sequencing (Clontech) and KAPA HyperPlus Library Preparation Kit (KAPA BIOSYSTEMS). The RNA-Seq library was multiplexed and sequenced on the NextSeq 500 platform (Illumina) 300 cycles High Output Mode (paired-end). De-multiplexing of RNA-Seq reads was conducted using bcl2fastq v.2 (Illumina). Both DNA-Seq and RNA-Seq reads were submitted to FastQC [55] for quality validation.

Genome assembly and gene finding

Following genome size estimation by GenomeScope [56], the MiSeq reads were assembled with MaSuRCA (v3.1.3)

[57] with the following parameters (PE 350150). Since the specimens were directly collected from the wild, we needed to be extra careful about possible contamination in the sequences. To validate contamination in the genome assembly [23, 41, 46], we first mapped the MiSeq DNA-Seq reads to the assembled genome with BWA (v0.7.17) [58] and converted the output files with SAMtools (v1.7) [59]. Next, we conducted a BLASTn search against the NCBI nr database [60, 61]. These results were submitted to BlobTools (v1.0) [44] 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 [62].

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) [63]. The resulting BAM files were submitted to BRAKER (v1.9) with default settings [64]. To remove possible bacterial contamination, we validated whether these query transcripts were included in the transcriptome assembly produced with Trinity v2.5.1 [65] 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 [66] 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 [67, 68] to assign KEGG orthology IDs. De novo repeat region identification was conducted by RepeatScout [69] and RepeatMasker [70]. We searched tRNA using tRNAscan [71] and rRNA using RNAmmer [72].

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) [73]. Additionally, we mapped the RNA-Seq reads to the coding sequences with BWA MEM and tested for differential expression using DESeq2 [58, 74]. Transcripts with FDR below 0.05 were defined as DEGs.

Gene expansion and lost pathway analysis

To validate abundant heat-soluble genes reported in previous tardigrade genomic studies [23, 29], 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) [75, 76]. Phylogenetic trees were constructed using FastTree (v2.1.10) with default options [77] and visualized with Interactive Tree of Life (iTOL) [78]. Regarding extremely similar sequences, genes with BLASTn identity >99% were regarded as identical artifact copies due to possible misassembly (Table S5, Fig. S1 phylogenetic tree).

Heat soluble proteomics

Heat soluble proteomics was conducted as previously described [30]. Briefly, approximately 3000 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) [79]. 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 Index [80] and DISOPRED [81].

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 EtAHS 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. variegatus*, UniProt ID: J7M799) were performed as described in the literature with slight modifications [30]. 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, CAHS1, and BSA proteins were dissolved at protein concentrations of 5.6 μM, 2.3 μM, 6.8 μM, and 2.0 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.

Abbreviations

CPM: Count per million; CAHS: Cytoplasmic abundant heat soluble; Dsup: Damage suppressor; DEGs: Differentially expressed genes; EtAHS: *Echiniscus testudo* Abundant Heat Soluble; GST: Glutamate-S transferase; HSP: Heat shock protein; HIFs: Hypoxia-inducible factors; LEA: Late embryogenesis abundant; MAHS: Mitochondrial abundant heat soluble; ROS: Reactive oxygen species; SAHS: Secretory abundant heat soluble; SODs: Superoxide dismutases; TFE: Trifluoroethanol; TPM: Transcript per million

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-021-08131-x>.

Additional file 1: Fig. S1. Blobplot analysis of the *E. testudo* genome assembly. The assembled genome was subjected to BlobTools for possible contamination identification and visualize as blobplot. (a) Blobplot of original genome assembly (before screening), (b) blobplot of original genome assembly with RNA-Seq coverage data, and (c) blobplot of after screening genome assembly. (d–j) Blobplot printed for each taxonomic group separately. The upper tier was using DNA-Seq data for mapping and the lower tier blobplot was using RNA-Seq data for mapping. (d) Taxonomic group; Tardigrada, (e) Arthropod, (f) “no-hit”, (g) Bacteroidetes, (h) Proteobacteria, (i) Actinobacteria and (j) other. Scaffolds were submitted to DIAMOND BLASTX analysis against UniProt Reference

Proteomes (2018_09 version) for taxonomy identification, and mapped data by BWA were used for coverage calculation. This information was analyzed by BlobTools. **Fig. S2.** Blobplot analysis of the *E. testudo* transcriptome assembly. The assembled transcriptome was subjected to BlobTools for possible contamination identification and visualize as blobplot. (a) Blobplot of original transcriptome assembly data and (b) after screening transcriptome assembly data. Scaffolds were submitted to DIAMOND BLASTX analysis against UniProt Reference Proteomes (2018_09 version) for taxonomy identification, and mapped data by BWA were used for coverage calculation. This information was analyzed by BlobTools. **Fig. S3.** Phylogenetic tree of duplicated genes. Phylogenetic tree of (a) Catalase, (b) SOD, (c) GST, and (d) HSP. These phylogenetic trees contain each gene of *R. variegatus* and *C. elegans*. Multiple alignments were conducted using MAFFT and phylogenetic trees were constructed by FastTree. Bootstraps are showed under the branch. Red lines indicate highly similar orthologs (> 99% identity). **Fig. S4.** M-A plot of DEGs and plot of fold change-gene expression in anhydrobiosis state. (a) Data represent individual genes and plotted by gray. FRD < 0.05 were defined as DEGs and colored in the red plot. The vertical axis shows log₂ fold-change and the horizontal axis shows log₁₀ baseMean, with positive change indicating up-regulated genes and a negative change indicating the down-regulated genes. (b) All genes were plotted in gray and genes that likely contribute to anhydrobiosis were colored. The vertical axis shows expression level (Transcript per million; TPM) in anhydrobiosis state and the horizontal axis shows fold change between active and anhydrobiosis state.

Additional file 2: Table S1. Screening of genome assembly. **Table S2.** Mapping ration of DNA-Seq and RNA-Seq. **Table S3.** Overview of transcriptome assembly. **Table S4.** Gene conservation of anhydrobiosis related genes. **Table S5.** Annotation, expression level, and fold index of DEGs. **Table S6.** MASCOT output list of heat soluble proteins. **Table S7.** Gene list of tardigrade-specific anhydrobiosis related candidate.

Acknowledgements

The authors are grateful to Yuki Takai (IAB) and Naoko Ishii (IAB) for providing experimental support and to Konosuke M. Ii (IAB) for collecting *E. testudo* on proteome analysis. The authors also thank Yudai Sasaki (NCU) for his contribution during the early stage of this study, Dr. Tadashi Satoh (NCU), Kumiko Hattori (NCU) and Dr. Ean Wai Goh (EXCELLS) for spectroscopic measurements.

Authors' contributions

K.A. conceived and designed the project. Y.M. conducted the RNA sequencing, data analysis and bioinformatics analysis. M.Y.U., S. T and K.K. performed NMR and CD experiments. M.F. conducted proteome analysis. K. A performed genome sequencing. The study was supervised by K.A. and M.T. and both significantly contributed to the interpretation of the data. The manuscript was written by Y. M, M.Y.U and K. A, and all authors reviewed and approved the final manuscripts.

Funding

This research was supported by Grants-in-Aid for Scientific Research KAKENHI of the Japan Society for the Promotion of Science (No. 17H03620 and 21H05279), the Nanotechnology Platform Program (Molecule and Material Synthesis, JPMX09S19MS0051) of MEXT, Japan, the Joint Research of the Exploratory Research Center on Life and Living Systems (EXCELLS program No.18–207, 19–208, 19–501), Instrument Center of Institute for Molecular Science, and research funds from the Yamagata prefectural government and Tsuruoka city. The funding bodies played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Availability of data and materials

The genome/transcriptome sequencing data sets are available at the BioProject accession number PRJNA669587. The assembly and coding sequencing data are available at figshare (<https://doi.org/10.6084/m9.figshare.13060634>). Genome of *R. variegatus* was obtained from GenBank with accession GCA_001949185.1, and that of *H. exemplaris* with GCA_002082055.1.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Institute for Advanced Biosciences, Keio University, Tsuruoka, Yamagata, Japan. ²Systems Biology Program, Graduate School of Media and Governance, Keio University, Fujisawa, Kanagawa, Japan. ³Exploratory Research Center on Life and Living Systems, National Institutes of Natural Sciences, Okazaki, Aichi, Japan. ⁴Institute for Molecular Science, National Institutes of Natural Sciences, Okazaki, Aichi, Japan. ⁵Graduate School of Pharmaceutical Sciences, Nagoya City University, Nagoya, Aichi, Japan. ⁶Faculty of Environment and Information Studies, Keio University, Fujisawa, Kanagawa, Japan.

Received: 20 November 2020 Accepted: 28 October 2021

Published online: 11 November 2021

References

- Keilin D. The problem of anabiosis or latent life: history and current concept. *Proc R Soc Lond B Biol Sci.* 1959;150(939):149–91. <https://doi.org/10.1098/rspb.1959.0013>.
- Crowe J, Hoekstra F, Crowe L. Anhydrobiosis. *Annu Rev Physiol.* 1992;54(1): 579–99. <https://doi.org/10.1146/annurev.ph.54.030192.003051>.
- Moberg N, Halberg KA, Jorgensen A, Persson D, Bjorn M, Ramlov H, et al. Survival in extreme environments - on the current knowledge of adaptations in tardigrades. *Acta Physiol.* 2011;202(3):409–20. <https://doi.org/10.1111/j.1748-1716.2011.02252.x>.
- Hengherr S, Worland MR, Reuner A, Brummer F, Schill RO. High-temperature tolerance in anhydrobiotic tardigrades is limited by glass transition. *Physiol Biochem Zool.* 2009;82(6):749–55. <https://doi.org/10.1086/605954>.
- Hengherr S, Worland MR, Reuner A, Brummer F, Schill RO. Freeze tolerance, supercooling points and ice formation: comparative studies on the subzero temperature survival of limno-terrestrial tardigrades. *J Exp Biol.* 2009;212(Pt 6):802–7. <https://doi.org/10.1242/jeb.025973>.
- Jönsson KI, Rabbow E, Schill RO, Harms-Ringdahl M, Rettger P. Tardigrades survive exposure to space in low earth orbit. *Curr Biol.* 2008;18(17):R729–31. <https://doi.org/10.1016/j.cub.2008.06.048>.
- Ono F, Mori Y, Takarabe K, Fujii A, Saigusa M, Matsushima Y, et al. Effect of ultra-high pressure on small animals, tardigrades and *Artemia*. *Cogent Phys.* 2016;3(1):1167575. <https://doi.org/10.1080/23311940.2016.1167575>.
- Ramløv H, Westh P. Cryptobiosis in the Eutardigrade *Adorybiotus* (Richtersius) coronifer: tolerance to alcohols, temperature and de novo protein synthesis. *Zool Anz.* 2001;240(3–4):517–23. <https://doi.org/10.1078/0044-5231-00062>.
- Horikawa DD, Cumbers J, Sakakibara I, Rogoff D, Leuko S, Harnoto R, et al. Analysis of DNA repair and protection in the tardigrade *Ramazzottius varieornatus* and *Hypsibius dujardini* after exposure to UVC radiation. *PLoS One.* 2013;8(6):e64793. <https://doi.org/10.1371/journal.pone.0064793>.
- Altiero T, Guidetti R, Caselli V, Cesari M, Rebecchi L. Ultraviolet radiation tolerance in hydrated and desiccated eutardigrades. *J Zool Syst Evol Res.* 2011;49(s1):104–10. <https://doi.org/10.1111/j.1439-0469.2010.00607.x>.
- Jönsson KI, Hygum TL, Andersen KN, Clausen LK, Moberg N. Tolerance to gamma radiation in the marine Heterotardigrade, *Echiniscoides sigismundi*. *PLoS One.* 2016;11(12):e0168884. <https://doi.org/10.1371/journal.pone.0168884>.
- Honda Y, Tanaka M, Honda S. Trehalose extends longevity in the nematode *Caenorhabditis elegans*. *Aging Cell.* 2010;9(4):558–69. <https://doi.org/10.1111/j.1474-9726.2010.00582.x>.
- Moore DS, Hansen R, Hand SC. Liposomes with diverse compositions are protected during desiccation by LEA proteins from *Artemia franciscana* and trehalose. *Biochim Biophys Acta.* 2016;1858(1):104–15. <https://doi.org/10.1016/j.bbame.2015.10.019>.
- Sakurai M, Furuki T, Akao K, Tanaka D, Nakahara Y, Kikawada T, et al. Vitrification is essential for anhydrobiosis in an African chironomid, *Polypedilum vanderplanki*. *Proc Natl Acad Sci U S A.* 2008;105(13):5093–8. <https://doi.org/10.1073/pnas.0706197105>.
- Crowe J, Clegg J, Crowe L. Anhydrobiosis: the water replacement hypothesis. Boston: Springer; 1998.
- Caramelo JJ, Lusem ND. When cells lose water: lessons from biophysics and molecular biology. *Prog Biophys Mol Biol.* 2009;99(1):1–6. <https://doi.org/10.1016/j.pbiomolbio.2008.10.001>.
- Goyal K, Tisi L, Basran A, Browne J, Burnell A, Zurdo J, et al. Transition from natively unfolded to folded state induced by desiccation in an anhydrobiotic nematode protein. *J Biol Chem.* 2003;278(15):12977–84. <https://doi.org/10.1074/jbc.M212007200>.
- Goldgur Y, Rom S, Ghirlando R, Shkolnik D, Shadrin N, Konrad Z, et al. Desiccation and zinc binding induce transition of tomato abscisic acid stress ripening 1, a water stress- and salt stress-regulated plant-specific protein, from unfolded to folded state. *Plant Physiol.* 2007;143(2):617–28. <https://doi.org/10.1104/pp.106.092965>.
- Oliveira E, Amara I, Bellido D, Odena MA, Dominguez E, Pages M, et al. LC-MS/MS identification of *Arabidopsis thaliana* heat-stable seed proteins: enriching for LEA-type proteins by acid treatment. *J Mass Spectrom.* 2007; 42(11):1485–95. <https://doi.org/10.1002/jms.1292>.
- Hand SC, Menze MA, Toner M, Boswell L, Moore D. LEA proteins during water stress: not just for plants anymore. *Annu Rev Physiol.* 2011;73(1):115–34. <https://doi.org/10.1146/annurev-physiol-012110-142203>.
- Hengherr S, Heyer AG, Kohler HR, Schill RO. Trehalose and anhydrobiosis in tardigrades—evidence for divergence in responses to dehydration. *FEBS J.* 2008;275(2):281–8. <https://doi.org/10.1111/j.1742-4658.2007.06198.x>.
- Jönsson KI, Persson O. Trehalose in three species of desiccation tolerant tardigrades. *Open Zool J.* 2010;3(1):1–5. <https://doi.org/10.2174/1874336601003010001>.
- Yoshida Y, Koutsovoulos G, Laetsch DR, Stevens L, Kumar S, Horikawa DD, et al. Comparative genomics of the tardigrades *Hypsibius dujardini* and *Ramazzottius varieornatus*. *PLoS Biol.* 2017;15(7):e2002266. <https://doi.org/10.1371/journal.pbio.2002266>.
- Förster F, Liang C, Shkumatov A, Beisser D, Engelmann JC, Schnolzer M, et al. Tardigrade workbench: comparing stress-related proteins, sequence-similar and functional protein clusters as well as RNA elements in tardigrades. *BMC Genomics.* 2009;10(469):1–10. <https://doi.org/10.1186/1471-2164-10-469>.
- Kamilari M, Jorgensen A, Schiott M, Moberg N. Comparative transcriptomics suggest unique molecular adaptations within tardigrade lineages. *BMC Genomics.* 2019;20(1):607. <https://doi.org/10.1186/s12864-019-5912-x>.
- Tunnacliffe A, Lapinski J. Resurrecting Van Leeuwenhoek's rotifers: a reappraisal of the role of disaccharides in anhydrobiosis. *Philos Trans R Soc Lond Ser B Biol Sci.* 2003;358(1438):1755–71. <https://doi.org/10.1098/rstb.2002.1214>.
- Campos F, Cuevas-Velazquez C, Fares MA, Reyes JL, Covarrubias AA. Group 1 LEA proteins, an ancestral plant protein group, are also present in other eukaryotes, and in the archaea and bacteria domains. *Mol Gen Genomics.* 2013;288(10):503–17. <https://doi.org/10.1007/s00438-013-0768-2>.
- Hibshman JD, Clegg JS, Goldstein B. Mechanisms of desiccation tolerance: themes and variations in brine shrimp, roundworms, and tardigrades. *Front Physiol.* 2020;11:592016. <https://doi.org/10.3389/fphys.2020.592016>.
- Hashimoto T, Horikawa D, Saito Y, Kuwahara H, Kozuka-Hata H, Shin I, et al. Extremotolerant tardigrade genome and improved radiotolerance of human cultured cells by tardigrade-unique protein. *Nat Commun.* 2016;7(1):12808. <https://doi.org/10.1038/ncomms12808>.
- Yamaguchi A, Tanaka S, Yamaguchi S, Kuwahara H, Takamura C, Imajoh-Ohmi S, et al. Two novel heat-soluble protein families abundantly expressed in an anhydrobiotic tardigrade. *PLoS One.* 2012;7(8):e44209. <https://doi.org/10.1371/journal.pone.0044209>.
- Tanaka S, Tanaka J, Miwa Y, Horikawa DD, Katayama T, Arakawa K, et al. Novel mitochondria-targeted heat-soluble proteins identified in the anhydrobiotic tardigrade improve osmotic tolerance of human cells. *PLoS One.* 2015;10(2):e0118272. <https://doi.org/10.1371/journal.pone.0118272>.
- Bemm F, Burleigh L, Foerster F, Schmucki R, Ebeling M, Janzen C, et al. Draft genome of the Eutardigrade *Milnesium tardigradum* sheds light on ecdysozoan evolution. *bioRxiv.* 2017; preprint first posted online.
- 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.* 2015;10(12): e0144803. <https://doi.org/10.1371/journal.pone.0144803>.
- 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(Pt 18):2799–802. <https://doi.org/10.1242/jeb.205.18.2799>.
35. Erkut C, Vasilij A, Boland S, Habermann B, Shevchenko A, Kurzchalia TV. Molecular strategies of the *Caenorhabditis elegans* dauer larva to survive extreme desiccation. *PLoS One.* 2013;8(12):e82473. <https://doi.org/10.1371/journal.pone.0082473>.
 36. Wright JC. Desiccation tolerance and water-retentive mechanisms in tardigrades. *J Exp Biol.* 1989;142(1):267–92. <https://doi.org/10.1242/jeb.142.1.267>.
 37. Persson D, Halberg KA, Jorgensen A, Ricci C, Mobjerg N, Kristensen RM. Extreme stress tolerance in tardigrades: surviving space conditions in low earth orbit. *J Zool Syst Evol Res.* 2011;49(s1):90–7. <https://doi.org/10.1111/j.1439-0469.2010.00605.x>.
 38. Heidemann NWT, Smith DK, Hygum TL, Stapano L, Clausen LKB, Jørgensen A, et al. Osmotic stress tolerance in semi-terrestrial tardigrades. *Zool J Linn Soc.* 2016;178(4):912–8. <https://doi.org/10.1111/zoj.12502>.
 39. Hygum TL, Fobian D, Kamiları M, Jorgensen A, Schiott M, Grosell M, et al. Comparative investigation of copper tolerance and identification of putative tolerance related genes in tardigrades. *Front Physiol.* 2017;8:95. <https://doi.org/10.3389/fphys.2017.00095>.
 40. Jorgensen A, Faurby S, Hansen JG, Mobjerg N, Kristensen RM. Molecular phylogeny of Arthrotardigrada (Tardigrada). *Mol Phylogenet Evol.* 2010;54(3):1006–15. <https://doi.org/10.1016/j.ympev.2009.10.006>.
 41. Arakawa K. No evidence for extensive horizontal gene transfer from the draft genome of a tardigrade. *Proc Natl Acad Sci U S A.* 2016;113(22):E3057. <https://doi.org/10.1073/pnas.1602711113>.
 42. Arakawa K, Yoshida Y, Tomita M. Genome sequencing of a single tardigrade *Hypsibius dujardini* individual. *Sci Data.* 2016;3(1):160063. <https://doi.org/10.1038/sdata.2016.63>.
 43. Yoshida Y, Konno S, Nishino R, Murai Y, Tomita M, Arakawa K. Ultralow input genome sequencing library preparation from a single tardigrade specimen. *J Vis Exp.* 2018;137(137). <https://doi.org/10.3791/57615>.
 44. Laetsch D, Blaxter M. BlobTools: Interrogation of genome assemblies [version 1; referees: 2 approved with reservations]. *F1000Research.* 2017;6:1287.
 45. Stec D, Krzywanski L, Arakawa K, Michalczyk L. A new redescription of *Richtersius coronifer*, supported by transcriptome, provides resources for describing concealed species diversity within the monotypic genus *Richtersius* (Eutardigrada). *Zool Lett.* 2020;6(1):2. <https://doi.org/10.1186/s40851-020-0154-y>.
 46. Koutsovoulos G, Kumar S, Laetsch DR, Stevens L, Daub J, Conlon C, et al. No evidence for extensive horizontal gene transfer in the genome of the tardigrade *Hypsibius dujardini*. *Proc Natl Acad Sci U S A.* 2016;113(18):5053–8. <https://doi.org/10.1073/pnas.1600338113>.
 47. Cornette R, Kikawada T. The induction of anhydrobiosis in the sleeping chironomid: current status of our knowledge. *IUBMB Life.* 2011;63(6):419–29. <https://doi.org/10.1002/iub.463>.
 48. Evangelista CCS, Guidelli GV, Borges G, Araujo TF, Souza TAJ, Neves U, et al. Multiple genes contribute to anhydrobiosis (tolerance to extreme desiccation) in the nematode *Panagrolaimus superbus*. *Genet Mol Biol.* 2017;40(4):790–802. <https://doi.org/10.1590/1678-4685-gmb-2017-0030>.
 49. Rytkonen KT, Williams TA, Renshaw GM, Primmer CR, Nikinmaa M. Molecular evolution of the metazoan PHD-HIF oxygen-sensing system. *Mol Biol Evol.* 2011;28(6):1913–26. <https://doi.org/10.1093/molbev/msr012>.
 50. Graham AM, Barreto FS. Loss of the HIF pathway in a widely distributed intertidal crustacean, the copepod *Tigriopus californicus*. *Proc Natl Acad Sci U S A.* 2019;116(26):12913–8. <https://doi.org/10.1073/pnas.1819874116>.
 51. Battaglia M, Covarrubias AA. Late Embryogenesis Abundant (LEA) proteins in legumes. *Front Plant Sci.* 2013;4:190. <https://doi.org/10.3389/fpls.2013.00190>.
 52. Sasaki K, Christov NK, Tsuda S, Imai R. Identification of a novel LEA protein involved in freezing tolerance in wheat. *Plant Cell Physiol.* 2014;55(1):136–47. <https://doi.org/10.1093/pcp/pct164>.
 53. Piszkiwicz S, Gunn KH, Warmuth O, Propst A, Mehta A, Nguyen KH, et al. Protecting activity of desiccated enzymes. *Protein Sci.* 2019;28(5):941–51. <https://doi.org/10.1002/pro.3604>.
 54. Johnson CG. The maintenance of high atmospheric humidities for entomological work with glycerol-water mixtures. *Ann Appl Biol.* 1940;27(2):295–9. <https://doi.org/10.1111/j.1744-7348.1940.tb07499.x>.
 55. Andrews S. FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. 2010. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
 56. Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, et al. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics.* 2017;33(14):2202–4. <https://doi.org/10.1093/bioinformatics/btx153>.
 57. Zimin AV, Marcais G, Puiu D, Roberts M, Salzberg SL, Yorke JA. The MaSuRCA genome assembler. *Bioinformatics.* 2013;29(21):2669–77. <https://doi.org/10.1093/bioinformatics/btt476>.
 58. Jo H, Koh G. Faster single-end alignment generation utilizing multi-thread for BWA. *Biomed Mater Eng.* 2015;26(Suppl 1):S1791–6. <https://doi.org/10.3233/BME-151480>.
 59. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. Genome project data processing subgroup: the sequence alignment/map format and SAMtools. *Bioinformatics.* 2009;25(16):2078–9. <https://doi.org/10.1093/bioinformatics/btp352>.
 60. Altschul S, Madden T, Schaffer A, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997;25(17):3389–402. <https://doi.org/10.1093/nar/25.17.3389>.
 61. Pruitt KD, Tatusova T, Maglott DR. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* 2007;35(Database issue):D61–5. <https://doi.org/10.1093/nar/gkl842>.
 62. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics.* 2015;31(19):3210–2. <https://doi.org/10.1093/bioinformatics/btv351>.
 63. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptsomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 2013;14(4):R36. <https://doi.org/10.1186/gb-2013-14-4-r36>.
 64. Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics.* 2016;32(5):767–9. <https://doi.org/10.1093/bioinformatics/btv661>.
 65. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* 2011;29(7):644–52. <https://doi.org/10.1038/nbt.1883>.
 66. Bairoch A, Boeckmann B, Ferro S, Gasteiger E. Swiss-Prot: juggling between evolution and stability. *Brief Bioinform.* 2004;5(1):39–55. <https://doi.org/10.1093/bib/5.1.39>.
 67. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000;28(1):27–30. <https://doi.org/10.1093/nar/28.1.27>.
 68. Moriya Y, Itoh M, Okuda S, Yoshizawa A, Kanehisa M. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 2007;35(Web Server issue):182–5.
 69. Price AL, Jones NC, Pevzner PA. De novo identification of repeat families in large genomes. *Bioinformatics.* 2005;21(Suppl 1):i351–8. <https://doi.org/10.1093/bioinformatics/bti1018>.
 70. Smit A, Hubley R, Green P: RepeatMasker Open-4.0. <http://www.repeatmasker.org>. 2013–2015.
 71. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 1997;25(5):955–64. <https://doi.org/10.1093/nar/25.5.955>.
 72. Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 2007;35(9):3100–8. <https://doi.org/10.1093/nar/gkm160>.
 73. Bray N, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol.* 2016;34(5):525–7. <https://doi.org/10.1038/nbt.3519>.
 74. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550. <https://doi.org/10.1186/s13059-014-0550-8>.
 75. 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.* 2002;30(14):3059–66. <https://doi.org/10.1093/nar/gkf436>.
 76. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013;30(4):772–80. <https://doi.org/10.1093/molbev/mst010>.
 77. Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol.* 2009;26(7):1641–50. <https://doi.org/10.1093/molbev/msp077>.
 78. Letunic I, Bork P. Interactive tree of life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics.* 2007;23(1):127–8. <https://doi.org/10.1093/bioinformatics/btl529>.

79. Perkins DN, Pappin DJC, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*. 1999;20(18):3551–67. [https://doi.org/10.1002/\(SICI\)1522-2683\(19991201\)20:18<3551::AID-ELPS3551>3.0.CO;2-2](https://doi.org/10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2).
80. Prilusky J, Felder C, Zeev-Ben-Mordehai T, Rydberg E, Man O, Beckmann J, et al. FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics*. 2005;21(16):3435–8. <https://doi.org/10.1093/bioinformatics/bti537>.
81. Ward JJ, McGuffin LJ, Bryson K, Buxton BF, Jones DT. The DISOPRED server for the prediction of protein disorder. *Bioinformatics*. 2004;20(13):2138–9. <https://doi.org/10.1093/bioinformatics/bth195>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

