Cooption of heat shock regulatory system for anhydrobiosis in the sleeping chironomid *Polypedilum vanderplanki*

Pavel V. Mazin[a,b,c,1], Elena Shagimardanova[d], Olga Kozlova[d], Alexander Cherkasov[d], Roman Sutormin[e], Vita V. Stepanova[a,b], Alexey Stupnikov[f], Maria Logacheva[a,b,g], Aleksey Penin[b,g,h], Yoichiro Sogame[i,j,k], Richard Cornette[i], Shoko Tokumoto[l], Yugo Miyata[i,m], Takahiro Kikawada (黄川田　隆洋)[i,l,1], Mikhail S. Gelfand[a,b,c,n,1], and Oleg Gusev[d,o,p,1]

[a]Center for Data-Intensive Biomedicine and Biotechnology, Skolkovo Institute of Science and Technology, Moscow, 143028, Russia; [b]Institute for Information Transmission Problems (Kharkevich Institute) RAS, Moscow, 127051, Russian Federation; [c]Faculty of Computer Science, Higher School of Economics, Moscow, 119991, Russian Federation; [d]Institute of Fundamental Medicine and Biology, Kazan Federal University, Kazan, 420012, Russian Federation; [e]Lawrence Berkeley National Laboratory, University of California, Berkeley, CA 94710; [f]Department of Oncology, School of Medicine, Johns Hopkins University, Baltimore, MD 21287; [g]A. N. Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, 119991, Russian Federation; [h]Department of Genetics, Faculty of Biology, Lomonosov Moscow State University, Moscow, 119991, Russian Federation; [i]Anhydrobiosis Research Group, Molecular Biomimetics Research Unit, Institute of Agrobiological Sciences, National Institute of Agriculture and Food Research Organization, Tsukuba, 305-0851, Japan; [j]Japan Society for the Promotion of Science, Tokyo, 102-0083, Japan; [k]Department of Applied Chemistry and Biochemistry, National Institute of Technology, Fukushima College, 970-8034 Iwaki, Japan; [l]Department of Integrated Biosciences, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Chiba 277-8562, Japan; [m]Center for Biological Resources and Informatics, Tokyo Institute of Technology, Nagatsuta-cho, Midori-ku, 226-8501 Yokohama, Japan; [n]Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, 119991, Russian Federation; [o]RIKEN Innovation Center, RIKEN, Yokohama, 650-0047, Japan; and [p]Center for Life Science Technologies, RIKEN, Yokohama, 650-0047, Japan

Edited by David L. Denlinger, The Ohio State University, Columbus, OH, and approved January 24, 2018 (received for review November 8, 2017)

***Polypedilum vanderplanki* is a striking and unique example of an insect that can survive almost complete desiccation. Its genome and a set of dehydration–rehydration transcriptomes, together with the genome of *Polypedilum nubifer* (a congeneric desiccation-sensitive midge), were recently released. Here, using published and newly generated datasets reflecting detailed transcriptome changes during anhydrobiosis, as well as a developmental series, we show that the TCTAGAA DNA motif, which closely resembles the binding motif of the *Drosophila melanogaster* heat shock transcription activator (Hsf), is significantly enriched in the promoter regions of desiccation-induced genes in *P. vanderplanki*, such as genes encoding late embryogenesis abundant (LEA) proteins, thioredoxins, or trehalose metabolism-related genes, but not in *P. nubifer*. Unlike *P. nubifer*, *P. vanderplanki* has double TCTAGAA sites upstream of the Hsf gene itself, which is probably responsible for the stronger activation of Hsf in *P. vanderplanki* during desiccation compared with *P. nubifer*. To confirm the role of Hsf in desiccation-induced gene activation, we used the Pv11 cell line, derived from *P. vanderplanki* embryo. After preincubation with trehalose, Pv11 cells can enter anhydrobiosis and survive desiccation. We showed that Hsf knockdown suppresses trehalose-induced activation of multiple predicted Hsf targets (including *P. vanderplanki*-specific LEA protein genes) and reduces the desiccation survival rate of Pv11 cells fivefold. Thus, cooption of the heat shock regulatory system has been an important evolutionary mechanism for adaptation to desiccation in *P. vanderplanki*.**

desiccation tolerance | anhydrobiosis | RNA-seq | *Polypedilum vanderplanki* | heat shock

The larva of the sleeping chironomid *Polypedilum vanderplanki* is the largest and the most complex animal that can enter anhydrobiosis, an ametabolic state, to survive drought (1). During desiccation, the *P. vanderplanki* larva loses 97% of its body water, resulting in the shutdown of all metabolic and physiological processes. Once rehydrated, the larva revives in less than 1 h and becomes almost indistinguishable from nondehydrated individuals (2). Previous studies resulted in the identification of several major molecular components involved in desiccation tolerance (1). Body water is mostly replaced by the disaccharide trehalose, which forms a solid glass matrix in the dry state and preserves molecular and cellular structures (3). Late embryogenesis abundant (LEA) proteins, likely acquired by horizontal gene transfer (4), together with molecular chaperones and trehalose, help to maintain proteins in the native state (5). The expanded group of thioredoxins and other antioxidant enzymes protect proteins against dehydration-induced oxidative stress (4, 6).

The recent release of the genome sequences and several RNA-sequencing (RNA-seq) transcriptomes of *P. vanderplanki* and *Polypedilum nubifer* (a congeneric, desiccation-sensitive midge) opens the door for a genome-wide study of the molecular mechanisms of anhydrobiosis and the underlying regulation networks (4). In contrast to other anhydrobiotic organisms, for

Significance

Anhydrobiosis is an ametabolic state found in several organisms that can survive extreme desiccation. It is of practical interest because its application to other systems might allow room temperature preservation of cells, tissues, or organs in the dry state. The insect *Polypedilum vanderplanki* is the most complex animal that can enter anhydrobiosis. Proteins responsible for desiccation tolerance in *P. vanderplanki* are relatively well studied, but little is known about mechanisms underlying their induction during desiccation. Here, we show that the heat shock transcription factor regulatory network was coopted during the evolution of *P. vanderplanki* to activate many known desiccation-protective genes, including genes encoding late embryogenesis abundant (LEA) proteins.

Author contributions: E.S., R.C., T.K., M.S.G., and O.G. designed research; P.V.M., A.C., M.L., A.P., Y.S., R.C., S.T., and Y.M. performed research; P.V.M., O.K., R.S., V.V.S., and A.S. analyzed data; and P.V.M., R.C., T.K., M.S.G., and O.G. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The data reported in this paper have been deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) database, https://www.ncbi.nlm.nih.gov/geo (accession no. SRP070984).

[1]To whom correspondence may be addressed. Email: iaa.aka@gmail.com, kikawada@affrc.go.jp, gelfand@iitp.ru, or oleg.gusev@riken.jp.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1719493115/-/DCSupplemental.

Published online February 20, 2018.

example tardigrades, where desiccation causes only minor changes in gene expression (7), in the sleeping chironomid, the gene expression profiles are strongly affected by desiccation, and the activity of at least some genes in specific regions of the genome (namely, ARId regions) contributes significantly to survival after anhydrobiosis (4). However, the underlying regulatory mechanisms are mostly unknown.

Recently, the cell line Pv11, derived from *P. vanderplanki* embryos, was shown to be a powerful tool for the investigation of desiccation tolerance (8, 9). It is the only known animal cell line that can survive complete desiccation. To successfully enter anhydrobiosis, Pv11 cells require incubation in a 0.6 M trehalose solution. Dried Pv11 cells can be preserved at room temperature for several months, with cell proliferation resuming after rehydration. While trehalose treatment is necessary for Pv11 cells to survive desiccation, it does not induce anhydrobiosis in other insect cell lines, which indicates the importance of specific factors, intrinsic to Pv11 cells, that promote desiccation tolerance. Thus, an understanding of these functions in Pv11 could potentially lead to their application to other systems, allowing the development of ambient-temperature preservation of animal cell cultures more generally.

In this study, using a detailed dataset from larvae at different stages of anhydrobiosis, we aimed to identify transcriptomic signatures linked to *P. vanderplanki* desiccation tolerance. We show that the heat shock transcription factor (Hsf) is likely to be responsible for desiccation-induced activation of hundreds of genes, including those playing key roles in anhydrobiosis. According to our results, several of these genes are also activated by trehalose treatment in Pv11 cells. Hsf knockdown experiments allowed us to directly implicate Hsf in desiccation-induced activation of gene expression. Moreover, Hsf knockdown results in a fivefold lower survival rate of Pv11 cells subjected to desiccation, thereby demonstrating an indispensable role for Hsf in anhydrobiosis.

Results

We used one previously published (20 samples) and one new (33 samples) RNA-seq dataset (DS1 and DS2, respectively; *SI Appendix*, Tables S1 and S2). The former encompasses the major developmental stages and the larval dehydration–rehydration cycle of *P. vanderplanki* and, partially, *P. nubifer*. DS2 covers *P. vanderplanki* dehydration–rehydration cycle with higher time resolution. We used these data to extend and improve the existing genome annotation (*SI Appendix*, Tables S3–S5 and Notes) and to quantify gene expression and splicing. While multidimensional scaling (MDS) analysis (*Methods*) revealed a good agreement between the datasets (Fig. 1*A* and *SI Appendix*, Fig. S1 and Notes), the interreplicate variability of gene expression is noticeably lower in DS1 than in DS2 (3.1- to 4.3-fold for the trended dispersion in dependence on the considered conditions; *SI Appendix*, Fig. S2). It could be explained by differences in the library preparation. Particularly, DS2 exhibit a higher level of duplicated reads (90% vs. 60% in DS1)—a sign overamplification. While these problems do not disrupt the overall agreement between the datasets, the relatively high variability in DS2 complicates the analysis of differential expression. Therefore, all statistical tests below are based on DS1, and DS2 is used only for confirmation. To evaluate gene expression changes under dehydration–rehydration cycle, we compared control predesiccation hydrated and active larvae (d0) to larvae after 24 and 48 h of desiccation (d24 and d48) and to larvae after 3 and 24 h of rehydration (r3 and r24). In *P. vanderplanki*, the number of suppressed genes [edgeR, Benjamini–Hochberg (BH)-corrected, $P < 0.05$; fold change, >3] exceeds the number of activated genes during desiccation, but the reverse trend is observed during rehydration. Long noncoding RNAs are significantly enriched among desiccation-activated genes (Fisher test, $P = 5 \times 10^{-24}$; odds ratio, 1.9; Fig. 1*B*). These effects are reproduced in DS2 (*SI Appendix*, Figs. S3 and S4 and Notes). Gene expression changes observed during the egg-to-larva transition are reversed during pupa formation in *P. vanderplanki* (log fold-change Pearson correlation coefficient rho = −0.83; Fig. 1*D*) in agreement with results obtained for *Drosophila melanogaster* (10). Noticeably, gene expression changes observed during pupa formation in *P. vanderplanki* exhibit correlation with gene expression changes observed under desiccation conditions (rho = 0.33; Fig. 1*E*). Hundreds of genes, including genes known to encode proteins involved in desiccation tolerance, such as heat shock proteins (HSPs), thioredoxins, and LEA proteins, are activated after 24- and/or 48-h desiccation. While the expression of most genes was restored to predesiccation levels after 24-h desiccation, several genes, such as some globins and cytochromes, transforming growth factor-β and genes related to fatty acid synthesis, maintained their altered expression patterns even in rehydrated larvae. Histone deacetylases and nuclear hormone receptors were significantly enriched among genes differentially expressed during the dehydration–rehydration cycle (for details, see *SI Appendix*, Tables S6 and S7 and Notes).

Hsf Activates Desiccation-Related Genes. To uncover the regulatory mechanisms underlying the observed gene expression changes, we compared the 2,000-nt upstream regions of up- and down-regulated genes at the 24-h (in both species) and 48-h (in *P. vanderplanki*) desiccation time points. We found five hexamers that were significantly (two-sided Fisher test, BH-corrected, $P < 0.05$) overrepresented or underrepresented in up-regulated genes compared with down-regulated genes. Three sequences were enriched in the upstream regions of up-regulated *P. vanderplanki* genes after 48-h desiccation: TCTAGA, CTAGAA, and TTCTCG (odds ratios, 1.58, 1.43, and 1.37, respectively). TCTAGA was also significantly enriched upstream of genes up-regulated after 24-h desiccation. Two hexamers (CTTATC and GTTGGC) were significantly enriched in upstream regions of genes suppressed after 24-h desiccation in *P. vanderplanki*. No hexamer was significantly enriched in *P. nubifer*. Two hexamers associated with up-regulated genes overlap and form a near-palindromic motif TCTAGAA that closely resembles the binding motif of the *D. melanogaster* heat shock transcription factor Dmel\Hsf [TOMTOM (11) search in the FlyFactorSurvey database (12); $P < 0.0004$]. We also found that one of the remaining hexamers, GTTGGC, is similar ($P < 0.0001$) to the binding motif of *D. melanogaster* zinc finger protein Dmel\CG4854. Unlike *P. vanderplanki*, in *P. nubifer* the TCTAGAA motif is not associated with desiccation-induced genes (Fisher test, $P = 0.59$; odds ratio, 0.87; Fig. 2*A*).

Association of the TCTAGAA motif with desiccation-activated genes was reproduced in DS2. TCTAGA and CTAGAA have the lowest (among all possible hexamers) P values for the association with genes that are significantly activated in DS2 after both 24-h (P values are 6.7×10^{-5} and 3.8×10^{-4}, respectively) and 48-h desiccation (P values are 1.0×10^{-5} and 6.0×10^{-5}, respectively).

Using the position weight matrix (PWM)-based approach implemented in motif analysis using the hypergeometric optimization of motif enrichment (HOMER) (13) yielded similar results, with the most significant motif having the consensus sequence TCTAGAAAGA and being observed in upstream regions of 33% of the genes that were significantly activated after 48-h desiccation in *P. vanderplanki* (*SI Appendix*, Table S8).

The position analysis revealed that the TCTAGAA sequence is enriched within 1,000 nt upstream or downstream of transcription starts of desiccation-induced genes in *P. vanderplanki* (Fig. 2*B*). This effect is not observed for down-regulated *P. vanderplanki* genes or desiccation-induced *P. nubifer* genes. In particular, genes that encode LEA proteins, whose expression was shown previously to respond to desiccation (5), thioredoxins, and chaperones have the TCTAGAA motif in their promoter regions in *P. vanderplanki* significantly more frequently than expected (Fisher test, $P < 0.007$, $P < 0.01$, and $P < 0.0008$, respectively). This could be a secondary effect explained by the fact that both the genes

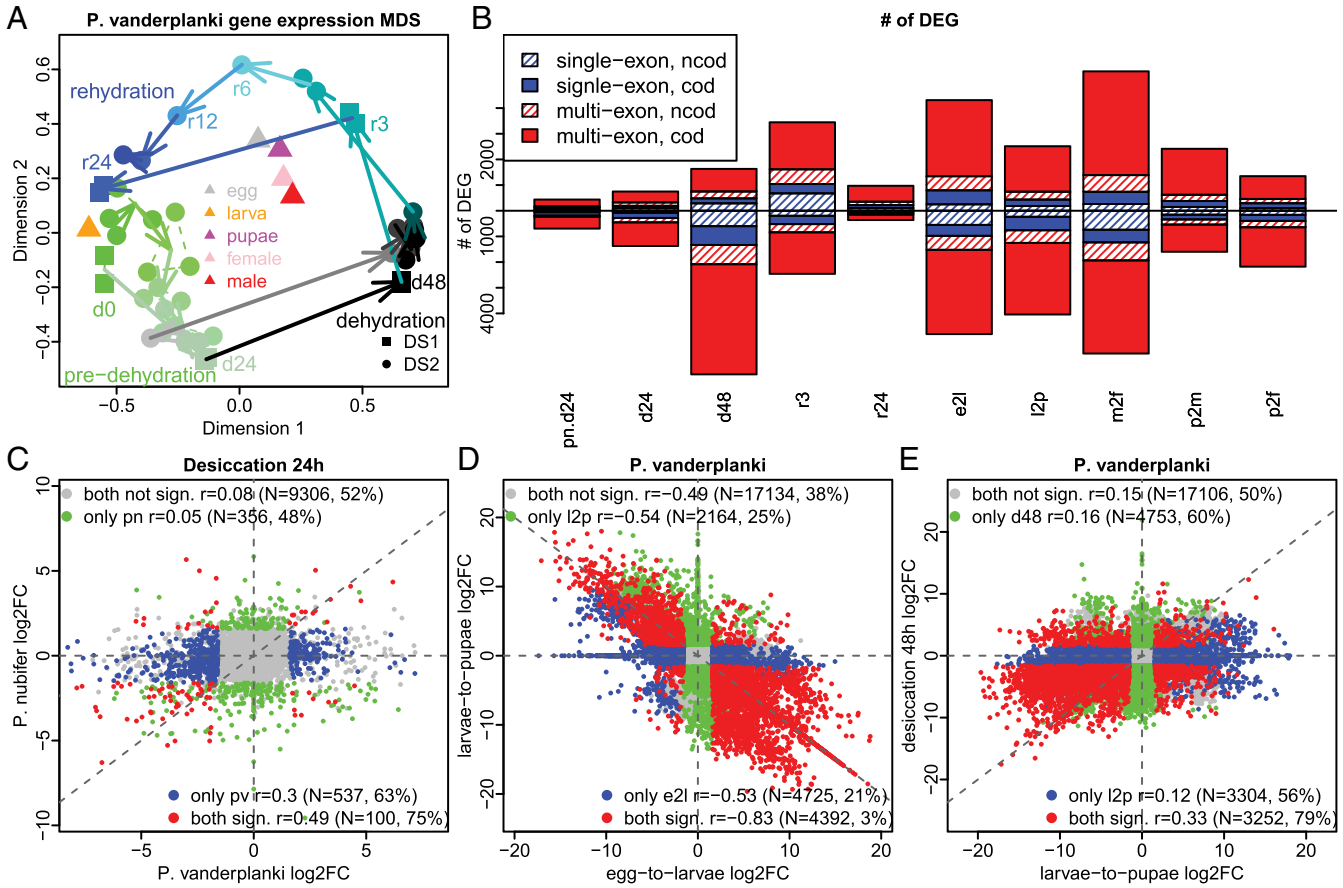

Fig. 1. Differential expression analysis. (*A*) MDS (two components) *P. vanderplanki* samples from DS1 (rectangles and triangles) and DS2 (circles). Centers of mass of sequential groups of replicates within each dataset are connected by arrows. One minus the Pearson correlation coefficient between normalized (within each dataset) RPKM values was used as the distance. Colors denote duration of desiccation (from green to black) and rehydration (from black to blue). (*B*) The number of differentially expressed genes in *P. nubifer* (predesiccation control vs. 24-h desiccation; pn.d24) and in *P. vanderplanki* [predesiccation control vs. 24-h (d24) and 48-h (d48) desiccation and 3-h (r3) and 24-h (r24) rehydration; comparisons of egg vs. larva (e2l), larva vs. pupa (l2p), male vs. female (m2f), pupa vs. male (p2m), and pupa vs. female (p2f) are also shown]. cod, coding; ncod, noncoding. (*C*–*E*) The comparison of logtwofold changes between *P. nubifer* and *P. vanderplanki* after 24-h desiccation (*C*), egg-to-larva vs. larva-to-pupa in *P. vanderplanki* (*D*), and larva-to-pupa vs. 48-h desiccation (*E*). Each point represents a single gene; genes that are significantly differentially expressed in both comparisons are shown in red. Genes with significant expression changes only in *P. nubifer* (*C*), during larvae-to-pupae transition (*D*), and after 48-h desiccation (*E*) are shown in green. Genes with significant expression changes only in *P. vanderplanki* (*C*), during egg-to-larvae transition (*D*), and during larvae-to-pupae transition (*E*) are shown in blue.

functioning in desiccation tolerance and the genes with the TCTAGAA motif are frequently activated under dehydration conditions. However, even among genes significantly activated after 48-h desiccation, chaperones and LEA protein genes are still significantly associated with the motif ($P < 0.0003$ and $P < 0.04$, respectively).

Three of the four genes involved in the metabolism of trehalose [a disaccharide of glucose that stabilizes biomolecules and cells in the dry state (3)] have TCTAGAA sites in their promoter regions in *P. vanderplanki* but not in *P. nubifer*; the exception is the gene encoding trehalase (TREH), which hydrolyses trehalose. While trehalose-6-phosphate phosphatase (TPP) and trehalose transporter TRET1 are weakly activated by desiccation in both species, trehalose-6-phosphate synthase (TPS) is highly (more than fourfold) activated, but only in *P. vanderplanki* (Fig. 2*C*).

Regulation of the trehalose metabolism genes by Hsf in *P. vanderplanki* is just one example of the dramatic differences in the Hsf regulon in the two *Polypedilum* species. Among 7,979 one-to-one orthologs that have assembled upstream regions, 993 and 833 genes in *P. vanderplanki* and *P. nubifer*, respectively, have TCTAGAA sites in the (−1,000, +1,000) promoter region, and 180 of them have the motif in both species. This overlap is significantly greater than expected by chance (Fisher test, $P < 5 \times 10^{-15}$; odds ratio, >2.1), but equal or higher odds ratios are observed for 17% of all possible permutations of the TCTAGAA motif. Thus, TCTAGAA motifs are only modestly more conserved than random sequences. Hence, the motif is associated with essentially different gene sets in these two species.

For each InterPro domain, we calculated the number of genes with and without a TCTAGAA site in *P. vanderplanki* and *P. nubifer*. While these distributions are not significantly different (Pearson's χ^2 test, $P > 0.8$), top InterPro families enriched in genes with TCTAGAA do not overlap in the two species. In *P. vanderplanki*, the top enriched domains ($P < 0.01$) are GroEL, the galanin receptor family, transcription factor IIS, glutaredoxins, and LEA protein genes. In *P. nubifer*, the top domains are HSP20, erythrocruorin, and globins (Fig. 2*D*).

The closest homologs of *D. melanogaster* Hsf in *P. vanderplanki* and *P. nubifer* are PvG018882 (Pv.Hsf) and PnG014307 (Pn.Hsf), respectively. Remarkably, Pn.Hsf does not have a TCTAGAA site in its promoter region, while Pv.Hsf is one of the two genes that have double TCTAGAA sites with a spacer shorter than 50 nt (*SI Appendix*, Fig. S5), the other being a desiccation-induced oxidoreductase PvG023509. Together with the observation that after 24-h desiccation Pv.Hsf exhibits much stronger activation than Pn.

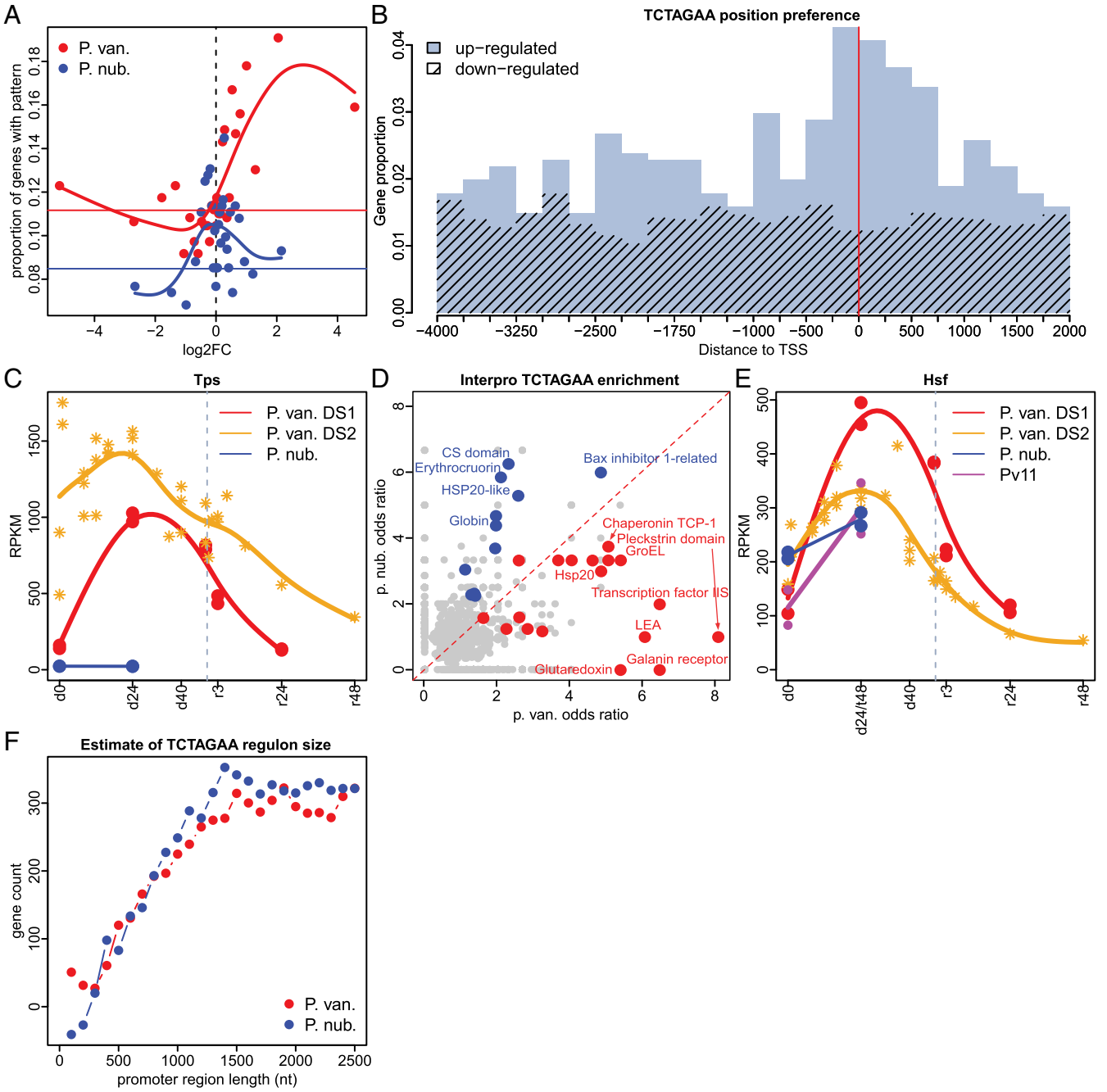

Fig. 2. Hsf activates genes during desiccation. (*A*) Dependence of the proportion of genes that have TCTAGAA site in the promoter region on average logtwofold change in *P. nubifer* (blue) and *P. vanderplanki* (red). Genes with similar logtwofold change are binned in groups of 100 genes. (*B*) Proportion of genes that have a TCTAGAA site in a given region (*x* axis, relative to the start of the first exon, nt), for genes activated after 48-h desiccation (gray), and for suppressed genes (black shading) in *P. vanderplanki*. (*C*) Expression of TPS during desiccation. (*D*) Enrichment (odds ratio) of InterPro families by genes with a TCTAGAA site in the promoter region in *P. nubifer* (*y* axis) and *P. vanderplanki* (*x* axis). Families with a *P* value below 0.01 (Fisher test) in either *P. nubifer* or *P. vanderplanki* are colored blue and red, respectively. No family has a *P* value below 0.01 in both species. (*E*) Expression of Hsf during larval desiccation or trehalose treatment of Pv11 cells. (*F*) Estimate of the TCTAGA regulon size in *P. vanderplanki* (red) and *P. nubifer* (blue). The regulon size was estimated as the number of genes that have a TCTAGAA site in the promoter region minus the number of genes that have a TCTAGAA site in the region of the same length 4,000 nt upstream of the TSS.

Hsf (3.7- and 1.3-fold, respectively), this suggests self-activation of Pv.Hsf (Fig. 2*E*).

The comparison of the (−1,000, +1,000) promoter region with the (−3,000, −1,000) region used as a negative control allowed us to estimate the number of genes likely regulated by the TCTAGAA motif (*Methods*). The estimated sizes of the TCTAGAA regulons are about the same in both species (~300 genes; Fig. 2*F*). Genes with a TCTAGAA site in *P. vanderplanki* but not in *P. nubifer* significantly more frequently demonstrate stronger activation under desiccation conditions in *P. vanderplanki* than in *P. nubifer* compared with genes that do not have TCTAGAA in both species (Fisher test, $P < 10^{-4}$; odds ratio, 1.4). At the same time, such behavior is not observed in *P. nubifer*, suggesting that the involvement of the TCTAGAA motif in the regulation of *P. vanderplanki* genes accounts for the interspecies difference in the response to desiccation.

According to previous studies, the active form of Hsf is a homotrimer that binds to a motif consisting of three AGAAA/TTTCT sequences (14). We used the PWM identified by HOMER to make a *P. vanderplanki*-specific model of the Hsf monomer binding motif (*Methods*). Then we searched for ungapped series of such sites (the weight of each site exceeds one bit; two or more sites, on any strand, total weight >12 bits). We compared such series found in the (−1,000, +1,000) promoter regions of desiccation-induced and desiccation-suppressed genes in *P. vanderplanki* and found that the former are enriched where there are two, three, or four sites arranged in strictly alternating directions (*Methods* and *SI Appendix*, Table S9).

Hsf Activates Genes During Trehalose Treatment of the Pv11 Cell Line. We used the Pv11 cell line as a model to further evaluate the role of Hsf in desiccation tolerance. We performed RNA-seq experiments on Pv11 cells (two biological replicates) under control conditions and after 48-h trehalose treatment. A fraction (14%) of genes whose expression changed in desiccating larvae were also regulated in trehalose-treated Pv11 cells. In most cases (77%), the expression changed in the same direction (Fisher test, $P < 10^{-10}$; odds ratio, 10.5; rho = 0.55; Fig. 3*A*). Hsf expression is also activated by trehalose treatment of Pv11 cells (Fig. 2*E*). Strikingly, the promoter regions of genes activated by trehalose treatment are also enriched in the TCTAGAA motif (Fisher test, $P < 0.0005$; odds ratio, 1.7; Fig. 3*B*). Of the 99 genes activated by both desiccation of larvae and trehalose treatment of Pv11 cells, 33 genes (one-third, 33%) harbor the motif within their promoter regions, while only 12% of genes that were suppressed under both conditions, or showed no significant changes or showed discordant changes, have the motif. Twelve genes (out of 99) should contain the TCTAGAA motif by chance. Therefore, we predict that about 21 of the 33 genes that are activated in both larvae and Pv11 cells and harbor the TCTAGAA motif within their promoter regions could be regulated by Hsf during both desiccation and trehalose treatment.

Hsf Knockdown Suppresses Activation of Multiple Genes During Trehalose Treatment. We used a recently established RNAi procedure for Pv11 cells (15) to validate this prediction. We excluded 12 genes that overlapped with other genes from the set of 33 genes identified above (test set 1). Additionally, we included four genes involved in trehalose metabolism, two additional LEA protein genes, and an oxidoreductase (PvG023509) that also have TCTAGAA motifs, but did not pass the differential-expression thresholds (test set 2). The final test set consisted of 29 genes including Hsf (*SI Appendix*, Table S10). We designed three siRNAs for the Hsf transcript. Two of them suppressed Hsf expression by 80% (Fig. 3*C*). The most efficient siRNA (HSF-200) was used in the analysis below. We transfected Pv11 cells with HSF-200 or with one of two random siRNA negative controls. Then we used real-time PCR to measure the expression levels of genes from the test sets before and after trehalose treatment of the HSF-200–treated cells and control cells. All experiments were performed with three replicates. Eighteen out of 21 genes from test set 1 showed significantly different expression (*t* test, BH-adjusted $P < 0.05$; Fig. 3 *D* and *E*) between cells transfected with HSF-200 and negative controls after trehalose treatment. Sixteen of them were down-regulated by the HSF-200 transfection, and this was significantly more than expected by chance (binomial test, $P < 0.002$). Thus, the Hsf knockdown appears to suppress trehalose-induced activation of genes predicted to be Hsf-regulated.

To further evaluate the significance of Hsf for Pv11 desiccation tolerance, we subjected Pv11 cells transfected with HSF-200 to dehydration. After rehydration of dry cells, we observed a fivefold reduction in survival rate of HSF-200–treated cells (Fig. 3*F*, *Left*, Mann–Whitney test, $P < 10^{-6}$), compared with controls, although the proliferation rate was not affected (Fig. 3*F*, *Right*; ANOVA, $P = 0.3$; *Methods*). The apparently reduced proliferation of HSF-200–treated cells between rehydration days 1 and 3 can be explained by the fact that most of the RNAi-treated cells are dead at day 1.

Posttranscriptional Regulation of Hsf. Up to 53% and 42% of genes encode more than one transcript in *P. vanderplanki* and *P. nubifer*, respectively, which is similar to the fraction of 58% recently observed in *D. melanogaster* (16). Differential splicing analysis showed that 510 exons from 338 genes significantly [BH correction, $q < 0.05$, difference in percent spliced in (dPSI) > 0.1; *Methods*] change their splicing pattern during the dehydration–rehydration cycle (for details, see *SI Appendix*, Notes). One example of this desiccation-induced alternative splicing was observed for the fifth exon of Hsf itself. This exon encodes a part of the transactivation domain (17, 18) and its inclusion dropped under desiccation conditions, especially in the anhydrobiotic state of *P. vanderplanki*, where the fifth exon was almost absent (Fig. 4 and *SI Appendix*, Fig. S13). Surprisingly, both introns adjacent to the fifth exon can be retained. The proportion of transcripts that retain these introns was reduced during desiccation in both species. However, these introns were retained much more frequently in *P. nubifer* than in *P. vanderplanki* under both control and dehydrated conditions (Fig. 4).

Additionally, we analyzed pre-mRNA polyadenylation and RNA editing sites. The former analysis allowed us to refine the boundaries of 833 final exons. We discovered 6,622 candidate RNA-editing sites, but we were unable to distinguish bona fide RNA editing events from DNA polymorphisms and sequencing errors (*SI Appendix*, Notes).

Discussion

The mechanisms of anhydrobiosis in *P. vanderplanki* larvae have been studied for several years and recent studies, including transcriptomics and comparative genomics, identified a large number of effector genes involved in desiccation tolerance (1, 4). However, the molecular mechanisms underlying the regulation of expression of these genes were still completely unknown before this study. Here, we performed extensive transcriptome analysis comparing developmental stages of *P. vanderplanki*, desiccation/rehydration processes of larvae, and the response of Pv11 cells to trehalose treatment and desiccation. This allowed us to identify many new genes that are differentially expressed during the desiccation and rehydration processes related to anhydrobiosis. Differential expression of histone deacetylases as well as altered expression of hundreds of long noncoding RNAs suggests a role for epigenetic factors in desiccation-related regulation of gene expression. The same patterns are common in plant molecular responses to desiccation (19). Widespread alternative splicing and promoter switching (*SI Appendix*, Table S11 and Notes) adds an extra level of complexity to the system. The major finding of this study is the identification of the heat shock transcription factor Hsf as an important regulator of anhydrobiosis-related gene expression.

Under anhydrobiosis, all metabolic processes including transcription and translation are shut down. However, previous studies have shown that, under our drying protocol, water content in *P. vanderplanki* larvae drops slightly (from ∼82 to 75%) under first 10 h of desiccation. Then it is preserved on constant level until desiccation hour 32 when it starts to quickly decrease (20). The larvae are still active during first 32 h of desiccation. During the rehydration process of dry larvae, the pharynx and heart begin to beat in 7–12 min (21), showing that metabolic activity resumes rapidly, at least for some organs. Larvae can be fully rehydrated and active within 1 h. Consequently, all of the stages of our dataset (except the dry state) are hydrated enough to allow transcriptional and translational activity. Our previous published studies on a restricted number of genes showed that,

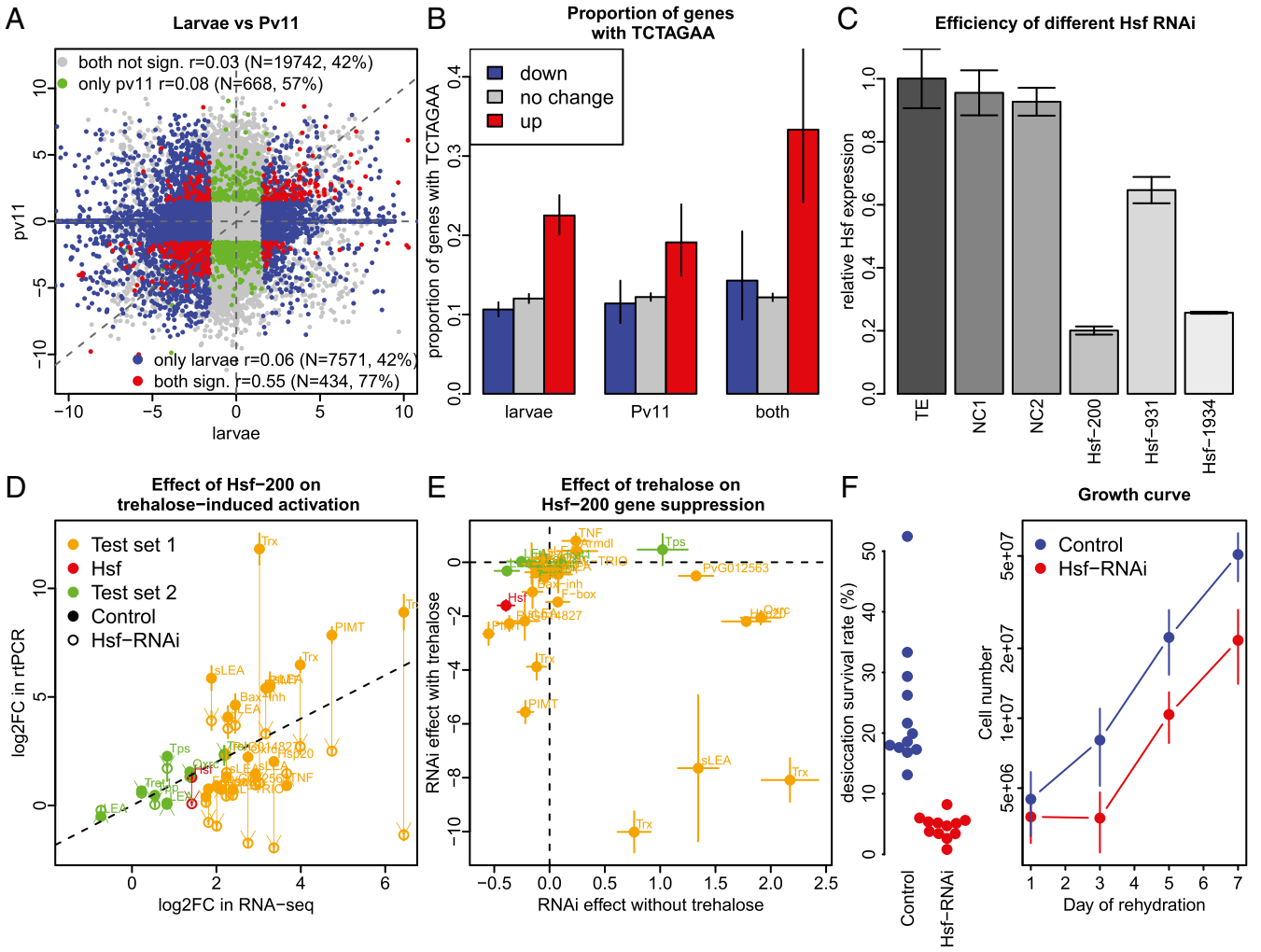

Fig. 3. Hsf activates desiccation-related genes in Pv11 cells subjected to trehalose treatment. (*A*) Comparison of logtwofold changes between desiccating larvae (48 h) and trehalose treated Pv11 cells (48 h). Genes where change in expression is not significant in both comparisons, significant only in larvae, significant only in Pv11 cells, and significant in both are shown by gray, green, blue, and red, respectively. (*B*) Proportion of genes with TCTAGAA in promoter regions among genes with constant (gray), down-regulated (blue), and up-regulated (red) expression. (*C*) Relative expression level of Hsf in Pv11 cells after mock transfection (TE), or after transfection with one of the negative control siRNAs (Nc1 and Nc2), or with one of the siRNAs against Hsf. (*D*) Logtwofold change due to trehalose treatment measured by real-time PCR (*y* axis) or RNA-seq (*x* axis); control and HSF-200 transfected cells are shown by filled and open cycles, respectively; observations relating to the same genes are linked by vertical arrows. (*E*) Logtwofold change of Hsf knockout measured in Pv11 cells under control conditions (*x* axis) or in response to trehalose treatment (*y* axis). On *D* and *E*, error bars show the 4-SD intervals. Gene set 1, gene set 2, and Hsf are shown in orange, green, and red, respectively. (*F*) Proportion of live Pv11 cells after 1 d of rehydration; cells were transfected with negative control siRNA or HSF-200 siRNA (*Left*); proliferation after rehydration of cells transfected with control siRNA or HSF-200 siRNA (*Right*); error bars show the 4-SD intervals.

during the induction of anhydrobiosis, the transcription levels were well correlated to the protein levels, with the expected normal delays due to translation and protein accumulation (22–24).

Although it has been suggested previously, no evidence of the involvement of heat shock-related transcription factors in the whole-genome control of transcription during anhydrobiosis had been obtained (4, 25). However, HSPs have been implicated in desiccation tolerance in many anhydrobiotes (26–28). Specifically, in the sleeping chironomid we previously showed that the expression of at least several HSPs is strongly linked to anhydrobiosis (29). Heat shock itself has been demonstrated to enhance desiccation tolerance in yeast (30). In *P. vanderplanki*, cross-tolerance is observed for a large variety of abiotic stresses. Salt stress and ionizing radiation have been shown to induce the expression of desiccation tolerance-related genes (20, 31). Thus, it would not be surprising to find that heat stress enhances desiccation tolerance in *P. vanderplanki*, as in other anhydrobiotes (30). Actually, our analysis suggests that Hsf activates hundreds of desiccation-related genes, including those critical for successful anhydrobiosis, that is, genes encoding LEA proteins, thioredoxins, and proteins involved in trehalose metabolism. Here, we obtained detailed data on Hsf gene structure and related expression profiles and suggest the following mechanism of action: first, Hsf is activated during dehydration, probably in response to protein denaturation caused by low water activity (32). Then Hsf binds to the upstream region of its own gene, resulting in self-activation. The increased concentration of Hsf leads to the activation of regulated downstream genes. The concentration of Hsf mRNA drops to prestress levels soon after the start of rehydration, but the underlying mechanisms for this remain to be investigated. Desiccation-induced activation of Hsf expression is accompanied by a dramatic decrease in the inclusion frequency of its fifth exon, which encodes a part of the transactivation domain. Whereas the presence of Hsf fifth exon seems linked to the hydrated state in both species, the retention of the flanking introns is much more pronounced in *P. nubifer*. These rearrangements impacting the transactivation domain

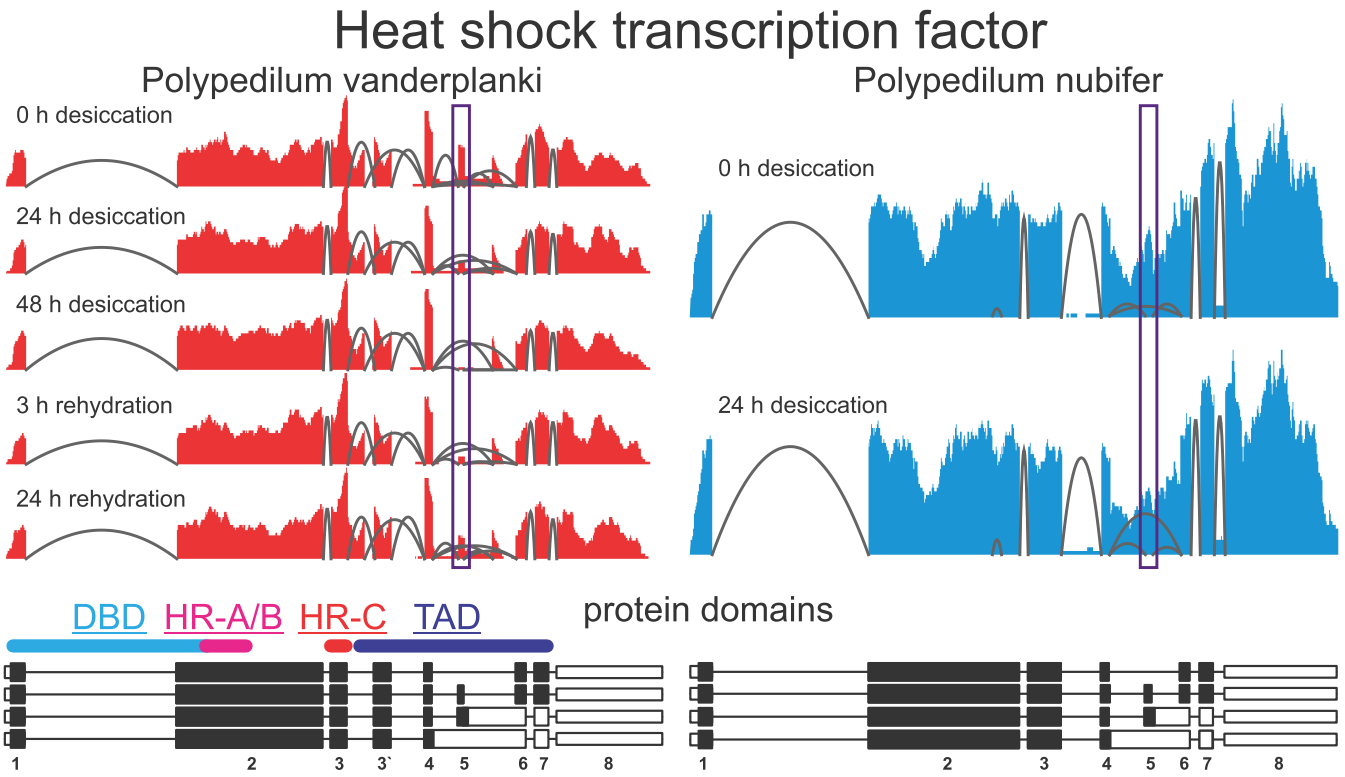

Fig. 4. Alternative splicing of *Hsf* transcripts. RNA-seq read coverage of the *Hsf* gene in *P. vanderplanki* DS1 and in *P. nubifer* is shown in the *Top*. DNA-binding domain (DBD), heptad repeat (HR) domains responsible for trimerization (HR-A/B) and trimerization inhibition (HR-C), and transactivation domain (TAD) are shown by colored lines above the exon structures (black boxes, coding regions; white boxes, noncoding regions). RNA-seq read coverage for exon 5 is highlighted by a violet rectangle.

point to a possible role for alternative splicing in Hsf function in the response to desiccation. The precise function of Hsf splicing variants in anhydrobiosis should be investigated in the future.

Remarkably, trehalose treatment is sufficient to activate Hsf in Pv11 cells. We speculate that high trehalose concentrations in the preconditioning medium might induce osmotic stress resulting in misfolding of proteins, which would lead to the activation of Hsf (33). Indeed, *P. vanderplanki* larvae massively accumulate trehalose in the process of induction of anhydrobiosis caused by dehydration (1). Further studies are need to elucidate the exact mechanistic link between the desiccation signal mimicked by osmotic stress, trehalose treatment, or salt stress (20), and its transduction by Hsf to induce the expression of anhydrobiosis-related genes.

Genes activated by Hsf during desiccation include not only its common targets such as chaperones, but also genes that encode LEA proteins and some trehalose metabolism genes. In *P. nubifer*, the Hsf gene lacks an Hsf binding site and Hsf regulatory targets are not activated by desiccation, and hence we posit that the heat shock regulatory system has been coopted to regulate, at least in part, the response to desiccation in *P. vanderplanki*. Remarkably, the larvae of the Antarctic midge *Belgica antarctica* are tolerant to temperatures as low as −15 °C and also to desiccation, inducing trehalose accumulation as in *P. vanderplanki* larvae (34), and these larvae show constitutively high expression of genes encoding HSPs, resulting in persistent high-level protection against protein aggregation (35). In this species, dehydration also enhances cold and heat stress tolerance (34). Thus, it would not be surprising if Hsf had also been coopted to regulate the response to desiccation and other stresses in *B. antarctica*. Since many anhydrobiotes show enhanced expression of HSPs during desiccation (27, 28), the involvement of Hsf in desiccation tolerance could be worth investigating more widely in other species.

In Pv11 cells treated with trehalose, Hsf knockdown suppressed expression of Hsf targets, but under control conditions we observed changes in both directions (Fig. 3*E*). Additionally, many genes that seem to be regulated by Hsf in desiccating larvae are not activated in trehalose-treated Pv11 cells. These observations point to the existence of other factors involved in the regulation of these genes, probably tissue- or stress-specific factors not present in Pv11 cells. The majority of the genes activated under desiccation in larvae or Pv11 cells do not seem to be regulated directly by Hsf itself. Whether they are regulated independently or by downstream factors activated by Hsf remain to be investigated. Nevertheless, Hsf knockdown reduces the desiccation survival rate of Pv11 cells by 80%. Thus, Hsf is clearly one of the main regulators of the desiccation-tolerance system in *P. vanderplanki*.

Conclusions

Here, we report an in-depth analysis of the transcription changes that take place during the *P. vanderplanki* desiccation–rehydration cycle and compare it to the changes in desiccation-sensitive *P. nubifer*. Our results suggest that the heat shock regulatory system was coopted in *P. vanderplanki* to regulate several genes involved in the desiccation response, a phenomenon that is not observed in *P. nubifer*. In particular, we discovered the potential for self-activation of the heat shock transcription factor Hsf and a specific alternative splicing pattern of *Hsf* mRNA in *P. vanderplanki*, but not in *P. nubifer*. We have shown that several groups of genes previously implicated in desiccation tolerance, encoding LEA proteins, thioredoxins, or trehalose metabolism genes, harbor candidate Hsf-binding sites in their promoter regions. We directly confirmed the role of Hsf in the activation of some of these genes in the Pv11 cell line. Our results provide a better mechanistic understanding of *P. vanderplanki* desiccation tolerance, represent a striking example of the evolutionary plasticity of regulatory

circuits, and open a route to understanding desiccation tolerance in other anhydrobiotic species.

Methods

Insects. Highly inbred lines of the congeneric species *P. vanderplanki* and *P. nubifer*, which differ in their ability to tolerate complete desiccation, were used for RNA extraction. The larvae were reared on a 1% agar diet containing 2% commercial milk under controlled photoperiod (13-h light:11-h dark) and temperature (27–28 °C) conditions.

Desiccation and Rehydration of *P. vanderplanki* Larvae. The desiccation was performed as described in ref. 20. Briefly, groups of eight final instar larvae were placed on filter paper imbibed with 0.44 mL of distilled water (DW) into a glass Petri dish (diameter, 65 mm; height, 20 mm). Five of these dishes were transferred into a plastic desiccator (25 × 25 × 30 cm, filled with 1.3 kg of silica gel) at <5% relative humidity, so that the larvae reached the anhydrobiotic state after 48 h. For rehydration, the dried larvae were immersed in DW.

Transcriptome Sequencing. To develop a comprehensive insight into differential gene expression during dehydration and rehydration, and to improve the coverage of transcriptome data, we performed deep RNA-seq of various RNA samples. Total RNA from hydrated, dehydrating, and rehydrated (*P. vanderplanki* only) larvae (20 individuals) was extracted using TRIzol (Thermo Fisher Scientific) and an RNeasy Mini Kit (Qiagen) according to the manufacturers' recommendations. A TruSeq RNA Sample Preparation kit, version 2 (Illumina), was used to prepare RNA-seq libraries for DS1, while a TruSeq Stranded mRNA Library Prep Kit (Illumina) was used for DS2. For *P. vanderplanki*, RNA was collected from whole larvae at different stages of dehydration (20 individuals each time point). RNA was also sampled from whole larvae at several time points after rehydration. For *P. nubifer*, RNA was only extracted at 0 and 24 h of dehydration (50 whole larvae at each time point). These samples were subjected to deep sequencing on the Illumina HiSeq 2000 platform. Further details of RNA-seq reads are given in *SI Appendix*, Tables S1 and S2. The data generated for this study have been deposited to the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under accession number SRP070984.

Cell Culture. Pv11 cells derived from *P. vanderplanki* embryos were grown in IPL-41 medium (Thermo Fisher Scientific) supplemented with 2.6 g/L tryptose-phosphate broth (TPB), 10% (vol/vol) FBS, and an antibiotic and antimycotic mixture (50 U/mL penicillin, 125 ng/mL amphotericin B, and 50 μg/mL streptomycin) (all Sigma-Aldrich), designated hereafter as complete IPL-41 medium. The cells were maintained at 25 °C and subcultured every 7 d into a 25-cm^2 cell culture flask. Dry-preservation and rehydration were carried out based on previous work (8) with the following modifications: cells were incubated in a preconditioning medium (0.6 M trehalose containing 10% complete IPL-41 medium) for 48 h at 25 °C. After 48-h incubation, cells were suspended in 400 μL of the preconditioning medium at a concentration of 1×10^8 cells/mL. Forty microliters of this cell suspension were dispensed as a droplet into a 35-mm Petri dish. Pv11 cell suspension-containing dishes were immediately transferred into a desiccator (250 × 250 × 250 mm, UD-1; AS ONE; <10% relative humidity) at 25 °C (20 dishes per desiccator) and left to dry for 10 d. For information, after 5 d of desiccation in these conditions, the water content of the dry cells mixture already dropped at 6.4 ± 0.3% of the dry mass (8). After rehydration by immersion in 1 mL of fresh complete IPL-41 medium, Pv11 cells were grown at 25 °C and monitored for 7 d after rehydration.

Proliferation Analysis. Cell numbers of Pv11 cells were counted with a disposable cell counter plate (Neubauer Improved type; Watson Bio Lab) using a fluorescence microscope (BZ-X700; KEYENCE). Viable and dead cells were evaluated with -Cellstain- Double Staining Kit (Dojindo). To compare the proliferation rates of control and Hsf-RNAi-treated cells after rehydration, we used the linear model [log(number of cells) ~ day + treatment + day:treatment] and *F* test. To exclude the effect of dead cells, data for the first day of the rehydration were eliminated.

Pv11 Transcriptomics. Total Pv11 RNA (4×10^7 cells for each dehydration or rehydration time point) was extracted with RNAiso Plus (TaKaRa Bio) and an RNeasy Mini Kit (Qiagen), according to the instruction manuals. Library preparation and sequencing for RNA-seq were conducted as described previously (4). Pv11 RNA-seq data generated for this study have been deposited to the NCBI SRA under accession number SRP070984.

Hsf Knockdown. *Hsf* gene siRNAs were designed using siDirect (36). The uniqueness of each candidate siRNA was confirmed by BLAST search of the *P. vanderplanki* genome (MidgeBase: bertone.nises-f.affrc.go.jp/midgebase) to minimize off-target effects. All designed siRNAs were synthesized and annealed by Hokkaido System Science Company. Sequences of siRNAs were as follows: HSF-200 sens: 5′-GCAAAAUUUGCAAAAGAAUTT-3′; HSF-931 sens: 5′-GAAUAUUGCAAAACAAUAUTT-3′; and HSF-1934 sens: 5′-CAUAUAUGUCAUAAUUUAATT-3′.

Simultaneously, we designed two types of siRNA with randomized sequences as negative controls. BLAST searches showed that the sequences were not present in the *P. vanderplanki* genome: NC1 sens: 5′-GCACUGCUACGAUCGUUAATT-3′; and NC2 sens: 5′-GUAGAGAGCGCGAUCUAUATT-3′.

Transfection was performed as described (15). Briefly, samples of 10^6 Pv11 cells were transfected with either siRNAs (0.5 μM final concentration) or TE (10 mM Tris·HCl, pH 8, 1 mM EDTA) buffer as a mock treatment using the 4D-Nucleofector System (Lonza) according to the manufacturer's instructions with a DC134 pulse code in SG medium. Following transfection, Pv11 cells were grown in 1 mL of complete IPL-41 medium in a single well of a 12-well plate for 48 h and subsequently incubated for an additional 48 h either in complete IPL-41 medium (control conditions) or in 0.6 M trehalose preconditioning medium (stress conditions). After incubation (total 96 h), Pv11 cells were collected by centrifugation at 1,000 × *g* for 5 min, and total RNA was extracted with the ReliaPrep RNA Tissue Miniprep System (Promega) and reverse-transcribed with a Ready-To-Go T-primed First-Strand cDNA Synthesis Kit (GE Healthcare) before real-time PCR analysis. Alternatively, after the 48-h pretreatment step in 0.6 M trehalose preconditioning medium, Pv11 cells were desiccated as described below for the desiccation survival assay.

Quantitative Real-Time PCR. The quantitative real-time PCR primers were designed using Primer3Plus software and are listed in *SI Appendix*, Table S10. The relative expression of Hsf and downstream genes was quantified using the CFX96 RT-PCR detection system (Bio-Rad) with an SYBR Premix Ex Taq kit (Tli RNase H Plus; TaKaRa Bio). PCR was performed using two-step cycling conditions: 95 °C for 60 s; 95 °C for 10 s and 60 °C for 40 s, repeated for 40 cycles; 95 °C for 10 s. Amplification was followed by melting-curve analysis with continual fluorescence data acquisition at 65 °C for 5 s and then at 95 °C for 5 s. Melt curve analysis of qPCR samples revealed that there was only one product for each gene primer reaction. Results were normalized to *PvEf1-α* (GenBank accession number AB490338) as an internal control. Relative expression data for each gene were obtained in triplicate.

Pv11 Desiccation Survival Assay. After rehydration in fresh complete IPL-41 medium, Pv11 cells were collected at 1, 3, 5, and 7 d following rehydration and stained with a Cellstain Double Staining Kit (DOJINDO), following the manufacturer's instructions. Bright-field images and calcein-AM fluorescence were visualized under a Biozero BZ-X700 microscope (Keyence), and Pv11 cells were quantified using BZ-H3C call count software (Keyence). One day after rehydration, cell viability was estimated from the number of calcein-AM–positive (live) cells, divided by the total number of cells observed by bright-field microscopy. Cell proliferation was assessed from the total number of cells counted every 2 d by bright-field microscopy. Each measurement was made with 12 replicates.

RNA-Seq Read Mapping. Genome sequences (version 0.9) and gene annotation (version 0.91) in the gtf format were downloaded from bertone.nises-f.affrc.go.jp/midgebase/. All sequencing reads from DS1 were mapped in two steps: first, reads from all samples of a given species were merged and mapped by tophat, version 2.0.13 (37), to the respective genome allowing prediction of new exon–exon junctions and using known junctions with the following parameters "–read-mismatches 4–read-edit-dist 4–read-realign-edit-dist 0–min-intron-length 10–max-intron-length 3,000,000–coverage-search–microexon-search–max-segment-intron 3,000,000." In the second step, extended gene annotations were reconstructed using these read alignments (see below) for each species. Then, RNA-seq reads from both datasets were aligned sample by sample using the exon–exon junctions from the new gene annotation without allowing prediction of new exon junctions (–no-novel-juncs option). These final alignments were used for the differential expression and splicing analyses.

Reconstruction of Exon–Intron Structures. Gene annotations for both species were reconstructed using the existing annotations and mapped RNA-seq reads by a splicing graph-based approach that allowed us to include only those exons and introns from the external annotation that did not contradict experimental data. For a detailed description of the pipeline, see *SI Appendix*, *SI Methods*.

ORF Prediction and Functional Annotation. Given the exon–intron structure of transcripts, the following procedure was used for the prediction of encoded proteins. First, analyzed transcript sequences were searched against the reference proteome set. This set consisted of two parts. The so-called "local" part was constructed by searching proteins in the NCBI database (38) for the "Nematocera" keyword. For the "global" part, the Uniref90 dataset (39) was used. The search program BLAT (40) was used in the protein-against-dnax mode, producing local similarity hits. Then the search result for each transcript was analyzed separately. All hits for a transcript were filtered with the following criteria. Each hit (*i*) should be at least 30 aa long (*ii*); should have at least 15 matched aa in the alignment; (*iii*) should have at least 30% protein identity; and (*iv*) should not contain stop codons. The hits satisfying these criteria were divided into two groups corresponding to each DNA strand. If the transcript strand was known (as for multiexon genes), the hits on the other strand were discarded. Otherwise, both strands were treated competitively. For each transcript, the best hit was selected by maximizing the number of matching amino acids in the alignment. Then the hit was extended in both directions to obtain the longest ORF. If the hit length was shorter than 30% of the indicated ORF, then this ORF was marked as unsupported by the hit. If a transcript had no hits, the longest ORF was selected (taking into account the strand, if known). ORFs supported by hits were filtered with the requirement to be at least 50 aa long. ORFs unsupported by hits should contain at least 100 aa. Pairs of orthologous genes were identified using orthoMCL (41) with default parameters. Only proteins from the two studied species were used.

Read Counting and Gene Expression Analysis. Numbers of reads that overlap constitutively spliced exons for each gene (gene read counts) were counted using SAJR (42). Only reads mapped to unique genomic locations were used. Read counts were normalized by the total length of constitutive exons and the library size (the sum of gene read counts) to obtain normalized gene expression values [fragments per kilobase of transcript per million mapped reads (FPKMs)]. Differential expression analysis was performed using the edgeR package (43). For differential expression analysis in DS1, the dispersion parameter was estimated using all data from the desiccation/rehydration cycle as the maximum between common, tagwise, and trended estimates. In DS2, the trended dispersion was used. Then each desiccation/rehydration time point was compared with the control condition. In the absence of replicates, *P. vanderplanki* developmental stages were compared using the dispersion estimates obtained from the desiccation/rehydration samples. Genes with a BH-corrected *P* value below 0.05 and fold-change above 3 were considered to be differentially expressed.

MDS Analysis. Nonmetric MDS to two dimensions was performed by the isoMDS function from the MASS package (44) of R (45). One minus the Spearman correlation coefficient between *z*-score transformed FPKM values was used as the distance measure between samples.

Differential Splicing Analysis. Differential splicing analysis was performed by our SAJR method as described (42). Briefly, each gene was divided into segments that were split between the two nearest splice sites. For each segment, the number of inclusion reads (reads that overlap the segment by at least 1 nt) and the number of exclusion reads (reads mapping to exon–exon junctions that span the segment) were calculated in each sample. For alternative first/last exons, only reads that map to the exon–exon junction connecting the exon to the second/next-to-last exon were considered as the inclusion reads. The number of exclusion reads for alternative first/last exons was calculated as the sum of inclusion reads for other alternative first/last exons. The inclusion and exclusion reads were considered as results of binomial trials and modeled by the pseudobinomial distribution using the generalized linear model (the glm function in R). The inclusion and exclusion reads were used to calculate exon usage frequency [percent spliced in (PSI)] as described (42). Segments with BH-corrected *P* value (pseudolog-likelihood test) below 0.05 and with difference in PSI (dPSI) above 0.1 were considered significant.

Hexamer-Enrichment Analysis. The 2,000-nt upstream regions of genes that were significantly up- and down-regulated after either 24- or 48-h desiccation in either species were extracted. Unassembled upstream regions were removed. Then, proportions of genes that contain each of 4,096 possible hexamers were compared between up- and down-regulated genes by the Fisher test. Hexamers with BH-corrected *P* value below 0.05 were considered significant.

Motif Structure by PWM from HOMER. We performed a differential search for motifs on the transcript strand of 2,000-nt upstream regions of *P. vanderplanki* genes that were significantly up- and down-regulated after 48-h desiccation using the HOMER software tool (13). According to previous studies (14), the Hsf-binding motif consists of several monomeric binding sites with the consensus AGAAA/TTTCT. We used motifs found by HOMER to construct the PWM for a monomeric site. The weight of nucleotide *n* in position *i* was defined as $w_{ni} = \log_2(f_{ni}/b_n)$, where f_{ni} and b_n are the frequencies of nucleotide *n* in position *i* of the motif and in the whole genome, respectively. Then we used the PWM to find ungapped sequences of matches with a weight greater than 1 bit on any strand in the same upstream regions. Only sequences with a weight greater than 12 bit were considered. Identified sequences were characterized by their strand and position. To analyze the motif structure, strand patterns were collected. We used the Fisher test (BH-corrected $P < 0.1$) to find motif structures significantly enriched in the upstream regions of up-regulated genes, compared with down-regulated genes.

Gene Ontology Enrichment Analysis. Gene ontology annotation was obtained using the InterProScan program (46). The enrichment analysis was performed using the goseq package (47) with the Wallenius method (logarithm average read count was used as the bias data); all genes were used as the background. Only terms with more than three genes were considered. All terms with a BH-corrected *P* value above 0.2 were considered significant.

Hsf Domain Structure. We mapped the domain structure of *D. melanogaster* Hsf (17, 18) to the *P. vanderplanki* and *P. nubifer* Hsf amino acid sequences using the MUSCLE (48) multiple protein alignment tool (*SI Appendix*, Fig. S12).

Data Availability. The RNA-seq data generated by this study have been deposited to the NCBI SRA under accession number SRP070984.

ACKNOWLEDGMENTS. We thank Anna Shmelkova for help with manuscript preparation. We are also grateful to Tomoe Shiratori for the maintenance of Pv11 cell cultures and to Yuki Sato-Kikuzato for her kind help with RNA-seq library preparation. Experimental aspects of this study were supported by Agriculture, Forestry and Fisheries Research Council of the Ministry of Agriculture, Forestry and Fisheries of Japan–Russian Science Foundation Joint Research Groups Grant 17-44-07002, while bioinformatic analysis was supported by Russian Science Foundation Grant 14-50-00150. Insect rearing and wet experiments were done at Institute of Agrobiological Sciences, National Institute of Agriculture and Food Research Organization and funded by the Ministry of Agriculture, Forestry and Fisheries of Japan, as well as Grants-in-Aid for Scientific Research (KAKENHI) Grants 15H05622, 23128512, 25128714, 25252060, 16K07308, and 16K15073 from the Ministry of Education, Culture, Sports, Science and Technology of Japan and/or Japan Society for the Promotion of Science.

1. Cornette R, Kikawada T (2011) The induction of anhydrobiosis in the sleeping chironomid: Current status of our knowledge. *IUBMB Life* 63:419–429.
2. Nakahara Y, et al. (2008) Effects of dehydration rate on physiological responses and survival after rehydration in larvae of the anhydrobiotic chironomid. *J Insect Physiol* 54:1220–1225.
3. Sakurai M, et al. (2008) Vitrification is essential for anhydrobiosis in an African chironomid, *Polypedilum vanderplanki*. *Proc Natl Acad Sci USA* 105:5093–5098.
4. Gusev O, et al. (2014) Comparative genome sequencing reveals genomic signature of extreme desiccation tolerance in the anhydrobiotic midge. *Nat Commun* 5:4784.
5. Hatanaka R, et al. (2015) Diversity of the expression profiles of late embryogenesis abundant (LEA) protein encoding genes in the anhydrobiotic midge *Polypedilum vanderplanki*. *Planta* 242:451–459.
6. Gusev O, et al. (2010) Anhydrobiosis-associated nuclear DNA damage and repair in the sleeping chironomid: Linkage with radioresistance. *PLoS One* 5:e14008.
7. Wang C, Grohme MA, Mali B, Schill RO, Frohme M (2014) Towards decrypting cryptobiosis—analyzing anhydrobiosis in the tardigrade *Milnesium tardigradum* using transcriptome sequencing. *PLoS One* 9:e92663.
8. Watanabe K, Imanishi S, Akiduki G, Cornette R, Okuda T (2016) Air-dried cells from the anhydrobiotic insect, *Polypedilum vanderplanki*, can survive long term preservation at room temperature and retain proliferation potential after rehydration. *Cryobiology* 73:93–98.
9. Nakahara Y, et al. (2010) Cells from an anhydrobiotic chironomid survive almost complete desiccation. *Cryobiology* 60:138–146.
10. Arbeitman MN, et al. (2002) Gene expression during the life cycle of *Drosophila melanogaster*. *Science* 297:2270–2275.
11. Bailey TL, et al. (2009) MEME SUITE: Tools for motif discovery and searching. *Nucleic Acids Res* 37:W202–W208.
12. Zhu LJ, et al. (2011) FlyFactorSurvey: A database of *Drosophila* transcription factor binding specificities determined using the bacterial one-hybrid system. *Nucleic Acids Res* 39:D111–D117.
13. Heinz S, et al. (2010) Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol Cell* 38:576–589.
14. Perisic O, Xiao H, Lis JT (1989) Stable binding of *Drosophila* heat shock factor to head-to-head and tail-to-tail repeats of a conserved 5 bp recognition unit. *Cell* 59:797–806.

15. Sogame Y, et al. (2017) Establishment of gene transfer and gene silencing methods in a desiccation-tolerant cell line, Pv11. *Extremophiles* 21:65–72.
16. Brown JB, et al. (2014) Diversity and dynamics of the *Drosophila* transcriptome. *Nature* 512:393–399.
17. Wu C (1995) Heat shock transcription factors: Structure and regulation. *Annu Rev Cell Dev Biol* 11:441–469.
18. Wisniewski J, Orosz A, Allada R, Wu C (1996) The C-terminal region of *Drosophila* heat shock factor (HSF) contains a constitutively functional transactivation domain. *Nucleic Acids Res* 24:367–374.
19. Han SK, Wagner D (2014) Role of chromatin in water stress responses in plants. *J Exp Bot* 65:2785–2799.
20. Watanabe M, Kikawada T, Okuda T (2003) Increase of internal ion concentration triggers trehalose synthesis associated with cryptobiosis in larvae of *Polypedilum vanderplanki*. *J Exp Biol* 206:2281–2286.
21. Hinton HE (1960) Cryptobiosis in the larva of *Polypedilum vanderplanki* Hint. (Chironomidae). *J Insect Physiol* 5:286–288.
22. Mitsumasu K, et al. (2010) Enzymatic control of anhydrobiosis-related accumulation of trehalose in the sleeping chironomid, *Polypedilum vanderplanki*. *FEBS J* 277:4215–4228.
23. Kikawada T, et al. (2007) Trehalose transporter 1, a facilitated and high-capacity trehalose transporter, allows exogenous trehalose uptake into cells. *Proc Natl Acad Sci USA* 104:11585–11590.
24. Kikawada T, et al. (2006) Dehydration-induced expression of LEA proteins in an anhydrobiotic chironomid. *Biochem Biophys Res Commun* 348:56–61.
25. MacRae TH (2016) Stress tolerance during diapause and quiescence of the brine shrimp, *Artemia*. *Cell Stress Chaperones* 21:9–18.
26. Reuner A, et al. (2010) Stress response in tardigrades: Differential gene expression of molecular chaperones. *Cell Stress Chaperones* 15:423–430.
27. King AM, Toxopeus J, MacRae TH (2014) Artemin, a diapause-specific chaperone, contributes to the stress tolerance of *Artemia franciscana* cysts and influences their release from females. *J Exp Biol* 217:1719–1724.
28. Erkut C, et al. (2013) Molecular strategies of the *Caenorhabditis elegans* dauer larva to survive extreme desiccation. *PLoS One* 8:e82473.
29. Gusev O, Cornette R, Kikawada T, Okuda T (2011) Expression of heat shock protein-coding genes associated with anhydrobiosis in an African chironomid *Polypedilum vanderplanki*. *Cell Stress Chaperones* 16:81–90.
30. Welch AZ, Gibney PA, Botstein D, Koshland DE (2013) TOR and RAS pathways regulate desiccation tolerance in *Saccharomyces cerevisiae*. *Mol Biol Cell* 24:115–128.
31. Ryabova A, et al. (2017) Genetic background of enhanced radioresistance in an anhydrobiotic insect: Transcriptional response to ionizing radiations and desiccation. *Extremophiles* 21:109–120.
32. Sorger PK (1991) Heat shock factor and the heat shock response. *Cell* 65:363–366.
33. Caruccio L, Bae S, Liu AY, Chen KY (1997) The heat-shock transcription factor HSF1 is rapidly activated by either hyper- or hypo-osmotic stress in mammalian cells. *Biochem J* 327:341–347.
34. Benoit JB, Lopez-Martinez G, Elnitsky MA, Lee RE, Jr, Denlinger DL (2009) Dehydration-induced cross tolerance of *Belgica antarctica* larvae to cold and heat is facilitated by trehalose accumulation. *Comp Biochem Physiol A Mol Integr Physiol* 152:518–523.
35. Rinehart JP, et al. (2006) Continuous up-regulation of heat shock proteins in larvae, but not adults, of a polar insect. *Proc Natl Acad Sci USA* 103:14223–14227.
36. Naito Y, Yoshimura J, Morishita S, Ui-Tei K (2009) siDirect 2.0: Updated software for designing functional siRNA with reduced seed-dependent off-target effect. *BMC Bioinformatics* 10:392.
37. Kim D, et al. (2013) TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 14:R36.
38. Geer LY, et al. (2010) The NCBI BioSystems database. *Nucleic Acids Res* 38:D492–D496.
39. UniProt Consortium (2015) UniProt: A hub for protein information. *Nucleic Acids Res* 43:D204–D212.
40. Kent WJ (2002) BLAT—the BLAST-like alignment tool. *Genome Res* 12:656–664.
41. Li L, Stoeckert CJ, Jr, Roos DS (2003) OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res* 13:2178–2189.
42. Mazin P, et al. (2013) Widespread splicing changes in human brain development and aging. *Mol Syst Biol* 9:633.
43. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26: 139–140.
44. Venables WN, Ripley BD (2002) *Modern Applied Statistics with S* (Springer, New York), 4th Ed.
45. R Development Core Team (2008) *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna).
46. Mitchell A, et al. (2015) The InterPro protein families database: The classification resource after 15 years. *Nucleic Acids Res* 43:D213–D221.
47. Young MD, Wakefield MJ, Smyth GK, Oshlack A (2010) Gene ontology analysis for RNA-seq: Accounting for selection bias. *Genome Biol* 11:R14.
48. Edgar RC (2004) MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113.