



Published in final edited form as:

Proteomics. 2019 December ; 19(24): e1900155. doi:10.1002/pmic.201900155.

Quantitative Proteomics Reveals Remodeling of Protein Repertoire Across Life Phases of *Daphnia pulex*

Leonid Peshkin,

Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA

Myriam Boukhali,

Massachusetts General Hospital Cancer Center and Department of Medicine, Harvard Medical School, Building 149, 13th Street, Charlestown, MA 02129, USA

Wilhelm Haas,

Massachusetts General Hospital Cancer Center and Department of Medicine, Harvard Medical School, Building 149, 13th Street, Charlestown, MA 02129, USA

Marc W. Kirschner,

Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA

Lev Y. Yampolsky

Department of Biological Sciences, East Tennessee State University, Johnson City, TN 37174, USA

Abstract

Although the microcrustacean *Daphnia* is becoming an organism of choice for proteomic studies, protein expression across its life cycle have not been fully characterized. Proteomes of adult females, juveniles, asexually produced embryos, and the ephippia-resting stages containing sexually produced diapausing freezing- and desiccation-resistant embryos are analyzed. Overall, proteins with known molecular functions are more likely to be detected than proteins with no detectable orthology. Similarly, proteins with stronger gene model support in two independent genome assemblies can be detected, than those without such support. This suggests that the proteomics pipeline can be applied to verify hypothesized proteins, even given questionable reference gene models. In particular, upregulation of vitellogenins and downregulation of actins and myosins in embryos of both types, relative to juveniles and adults, and overrepresentation of cell-cycle related proteins in the developing embryos, relative to diapausing embryos and adults,

Dr. L. Peshkin, pasha@hms.harvard.edu.

Author Contributions

L.P., L.Y., and M.W.K. conceptualized the study; L.P., W.H., and M.W.K. were associated with methodology; L.P. and M.B. performed the experiments; L.P., W.H., and L.Y. performed the formal analysis; L.P., W.H., and M.W.K. were acquired resources; L.P., M.B., and L.Y. were associated with data curation; L.P., L.Y., and M.W.K. wrote the study; L.P., W.H., L.Y., and M.W.K. supervised the study.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Data Availability

Proteomics data deposited into MassIVE Repository (<https://massive.ucsd.edu/>), accession number MSV000083711. Relative abundance data by protein and by peptides are available in Supporting Information 1 and 2, respectively.

Conflict of Interest

The authors declare no conflict of interest.

are observed. Upregulation of small heat-shock proteins and peroxidases, as well as overrepresentation of stress-response proteins in the ephippium relative to the asexually produced non-diapausing embryos, is found. The ephippium also shows upregulation of three trehalose-synthesis proteins and downregulation of a trehalose hydrolase, consistent with the role of trehalose in protection against freezing and desiccation.

Keywords

Daphnia; life cycle; proteomics; trehalose synthesis; vitellogenins

1. Introduction

MS-based proteomics is a major tool for understanding changes in gene expression, biochemistry, and physiology both during development and the response to environmental change. Microcrustacean *Daphnia*, an emerging model organism for biomedical research, has several useful features allowing characterization of life-history and physiological plasticity, including clonal reproduction, short lifespan, and well-characterized genome. Here, we report proteome differences between parthenogenetically produced embryos, juveniles, adults, and sexually produced resting eggs that can be used as a base-line reference for protein expression analysis across the life cycle of an organism. Previous advances in *Daphnia* proteomics identified difficulties in obtaining a large number of intact and identifiable proteins due to high protease activity in tissue extracts,^[1–3] often resulting in a relatively small number of proteins reported (often 500–1500 out of over 25 000 predicted protein coding genes; the highest number reported close to 4000^[4]).

Despite these difficulties, a variety of proteomic responses have been reported, including responses to various ecological stimuli, such as temperature,^[5,6] food availability,^[6] hypoxia,^[7] microgravity,^[8] and presence of predators^[9–11] or toxicants.^[4,12–14] Other studies focused on characterization of specific groups of proteins, such as neuropeptides and hormones.^[15] Yet, *Daphnia* proteome remains largely uncharacterized,^[3] including the lack of proteomics data across different life cycle phases.

In favorable conditions *Daphnia* reproduce asexually throughout the life cycle producing subitaneous eggs developing into offspring genetically identical to each other and the mother.^[16] These eggs undergo immediate development in the female's brood chamber and are released as neonates largely resembling adults. In response to environmental cues such as population density, temperature, and photoperiod meiosis is induced and haploid resting eggs are produced, each sealed in protective structure called "ephippium." These eggs require fertilization to survive; they arrest development in early gastrulation phase and are reactivated after a period of exposure to darkness and low temperatures.^[16]

Several a priori predictions can be made about proteins expressed during each phase of the life cycle. Embryos, both asexually and sexually produced, are expected to contain higher levels of vitellogenins (glycolipoproteins that are precursors of yolk lipo- and phosphoproteins in many animals including *Daphnia*^[17]). Conversely, ephippia embryos that have less or no muscle tissue would be expected to contain less actin and myosin.

Diapausing eggs capable of surviving freezing, heating, and desiccation are hypothesized to contain high levels of trehalose, a disaccharide commonly used in crustaceans and insects as freezing and desiccation protectant,^[18,19] thus one might expect upregulation of trehalose synthesis and downregulation of trehalose breakdown proteins in ephippia. One may also expect elevated expression in ephippia of stress-tolerance related proteins, such as small heat-shock proteins^[20,21] and proteins involved in antioxidant pathways, such as superoxide dismutases (SODs), peroxidases, and other glutathione metabolism enzymes, and thioredoxins.^[22] Additionally, desiccation resistant mosquito shows upregulation of protein repair enzymes protein-L-isoaspartate-(D-aspartate) O-methyltransferases (PIMTs) during desiccation stress.^[22] A *Daphnia magna* ortholog of mosquito PIMT has been shown to be upregulated in freshly laid ephippia.^[23] On the other hand, parthenogenetically produced embryos are not capable of surviving freezing or desiccation, so these expectations do not apply to them.

Although *Daphnia pulex* genome is relatively well characterized,^[24,25] the exact number of protein coding genes in the genome is still subject of debate, as it varies significantly between closely related genomes and/or between gene model methodologies.^[25] Unexpectedly, only about one-third of protein coding genes in either genome assembly are each other's mutual best BLAST hits,^[26,27] which raises questions about robustness of either set of gene models. We used our proteomics data to evaluate the agreement (or lack thereof) between the two genome assemblies using the set of proteins observed in this study as a positive control set likely to be enriched in true positives, i.e., genes that are actually expressed.

2. Results

2.1. Outcomes of MS Analysis and Protein Identification

Proteomic analysis revealed 34 158 peptides that mapped to unique 5099 proteins in the UNIPROT reference. Of these, 5062 proteins were matched to *D.pulex* TCO annotated proteins (version jgi060905.^[22] Out of the 5099 proteins, 3549 were identified with two or more unique peptides and 2662 proteins with three or more peptides. The likelihood of discovering a protein using our protocol was strikingly non-uniform across molecular function (MF) categories (Table 1). Specifically, proteins detected were highly enriched among enzymes, structural proteins, transporters, and RNA-binding proteins; in these categories, we detected between one-third and one-half of all annotated proteins. In contrast, we detected fewer than the expected number of transcription factors and, notably, fewer of the proteins with unknown functions, which were identified by the genome gene model analyses. Here, we were only able to identify 4% of UniProt-listed proteins. These proteins were also represented by the fewest peptides (2.89 ± 1.16), significantly less than the yield of transporters 13.73 ± 4.24), structural proteins (14.38 ± 2.23), and metabolic enzymes (6.64 ± 0.24); ANOVA on log-transformed peptide numbers per protein: $df= 8,5089$; $F= 34.4$; $p < 2.5 \times 10^{-53}$; Tukey test for the difference between unknown proteins and transporters, structural proteins, and metabolic enzymes: $p < 0.001$.

We also observed that proteins identified as mutual best BLAST hits between the TCO protein reference used in this analysis and a more recent PA42 *D. pulex* protein reference^[25]

are highly enriched among the proteins we were able to detect. Out of 11 697 such proteins, we detected 4312 (expected = 2004), while among the 18 063 proteins with ambiguous hits or no hits, we detected only 786 (expected = 3094; Fisher exact test $p < 1 \times 10^{-248}$). This enrichment remains highly significant when analyzed within each MF category.

2.2. A Priori Expected Enrichment of Molecular Functions Across Life Stages

Normalized signals mapped to individual proteins showed good concordance among biological replicates and these served to convincingly separate the samples on the plane of the first two principal components (Figure 1) with PC1 largely separating juveniles and adults from embryos and PC2, the ephippium sample from asexual individuals of any life cycle phase. Functional genomics analysis revealed that proteins downregulated in embryos relative to adults were significantly enriched in enzymes and underrepresented among transcription factors and other DNA-binding proteins, RNA-binding proteins, and proteins with “other” molecular functions (Table 2). In addition, in the ephippium sample, in the comparison to subitaneous embryos, proteins with unknown molecular function were overrepresented in the subset of upregulated proteins. Other molecular functions showed no significant enrichment among proteins differentially expressed in samples representing different stages of the life cycle.

Majority of proteins showing significant differences between adults and embryos tended to be highly expressed in adults, but showed low abundance in the embryos (Figure 2A, red dots). However, there were also a number of proteins upregulated in the embryos. Several functional groups of proteins a priori expected to show differences among life cycle classes showed, as a group, the expected behavior (Figure 2B): small HSPs, antioxidant pathway proteins (except thioredoxins), and vitellogenins were more abundant in the embryos than in adults with juveniles showing intermediate values. Unexpectedly, actins and myosins were equally abundant in the embryos juveniles and adults on average (Figure 2B), although individual actins and myosins were slightly enriched among genes with a significant downregulation in embryos (Fisher’s exact test two-tailed $p < 0.035$). A similar enrichment among proteins upregulated in embryos was observed for the antioxidant pathway proteins (Fisher’s exact test two-tailed $p < 0.031$).

The enrichment analysis in comparison between the ephippial sample and subitaneous embryos also revealed a significantly higher occurrence of HSPs and chaperons and antioxidant pathway proteins among proteins upregulated in the ephippial sample (Fisher’s exact test two-tailed $p < 0.001$ and $p < 0.0002$, respectively).

Small HSPs were enriched among proteins upregulated in the ephippium (Table 2, Figure 3C) with some showing 10–50 times lower abundance in the subitaneous embryos (Figure 3C; although one small HPS showed an inverse pattern). Although not significantly enriched (due to the low number of protein in the pathway), the trehalose metabolism genes all show differential expression in the ephippium. Three alpha-alpha-trehalose-6-phosphate synthases (that catalyze the rate-limiting step in trehalose synthesis pathway,^[28] are upregulated in the ephippium (Figure 3D; FDR < 0.01), while the trehalose hydrolase is downregulated (FDR < 0.01), relative to adult tissues. Similarly, there were numerous peroxidases and small heat shock proteins with higher relative abundance in the ephippium

than in the non-diapausing embryos (Figure 3B,C). Finally, the only PIMT ortholog detected in our samples (UniProt: E9G8Q2) had approximately three- to fivefold higher abundance than either in subitaneous embryos or maternal tissues (FDR = 0.022 and FDR = 0.055, respectively).

2.3. Biological Processes Overrepresentation Across Life Stages

Proteins with higher abundance in subitaneous (developing) embryos relative to adults showed overrepresentation of a variety of biological processes related to cell division (Table S1 and Figure S1, Supporting Information). These included DNA replication (11-fold enrichment, FDR < 10^{-7}) microtubule nucleation and polymerization (both 24-fold overrepresentation and FDR < 10^{-3}), mitotic phase transition (12-fold overrepresentation, FDR < 10^{-3}). Likewise, overrepresented were transcription and ribosome-building biological processes, indicating active gene expression, as well as histone acetylation, possibly related to ongoing differentiation.

In contrast, among the proteins with higher abundance in adults relative to embryos, the overrepresented biological processes included largely catabolic, respiratory, and oxidation-reduction processes (Table S2, Supporting Information), such as tricarboxylic acid cycle (20-fold overrepresentation, FDR < 10^{-8}), proton transmembrane transport (12-fold overrepresentation, FDR < 10^{-8}), oxidative phosphorylation (ninefold overrepresentation, FDR < 10^{-6}), pyruvate metabolism (22-fold overrepresentation, FDR < 10^{-6}), Cell adhesion was another BP category overrepresented (fivefold overrepresentation, FDR < 10^{-6}). Among the BPs underrepresented in this category of proteins there is, notably, DNA repair pathways (expected 6.54, observed 0, FDR < 0.05).

Similar results were obtained in the comparison of both types of embryos to both adults and juveniles combined (Tables S3 and S4, Supporting Information). The only exception was cell-cycle related proteins that were not overrepresented in this comparison (because there was no active cell division in the arrested embryos in the ephippia, where there was cell division in juveniles).

The overrepresentation analysis of BPs among the proteins with elevated abundance in the single ephippium sample relative to subitaneous embryos revealed, surprisingly, an elevated metabolic activity in the resting embryos than in actively developing ones (Table S5 and Figure S2, Supporting Information). Overrepresented BPs included metabolism of nucleotides, carbohydrates, lipids, pyruvate, amino sugars, and aminoglycans (3–16-fold overrepresentation). More expectedly, overrepresented were proteins involved in response to stress, including response to reactive oxygen species (sevenfold overrepresentation, FDR < 0.05), and toxicants (eightfold overrepresentation, FDR < 10^{-8}). Likewise, cell adhesion, extracellular matrix organization, and intracellular signal transition BPs were overrepresented in the ephippium sample (two- to fivefold overrepresentation).

In contrast, in the subset of proteins with low abundance in the ephippium sample relative to the subitaneous embryos, we observed overrepresentation of nucleotide-, ribonucleoside monophosphate, and DNA biosynthesis processes (11–51-fold overrepresentation),

translation, and protein folding (6–33fold overrepresentation), as one would expect for embryos with arrested development and growth (Table S6, Supporting Information).

Finally, in the comparison of adult females to juveniles no individual protein achieved $FDR < 0.01$; relaxation of this criterion to $FDR < 0.05$ results in 17 proteins with higher relative abundance in the adults, in which no BPs were significantly overrepresented and 58 proteins with higher relative abundance in juveniles, among which proteins involved in sister chromatid segregation showed a 37-fold overrepresentation ($FDR < 0.05$).

Notably the “unclassified” category is underrepresented in all these comparisons. It would be difficult to hypothesize that constitutively expressed proteins are, on average, less well characterized or less conserved across animals. Rather, we suspect that the unclassified category is more likely to include false positives or low expression/high noise proteins.

3. Discussion

3.1. Proteomics

Despite previous reports that quantitative proteomics in *Daphnia* presents significant challenges,^[2,3] we were able to obtain a deeper coverage of the proteins expressed across very different states of life cycle and conditions, using our standard proteomic sample preparation protocols and analysis procedures and identifying over 5000 proteins. The only previous study that was similarly successful in identifying individual *D. pulex* proteins^[4] utilized the label-free MS protocol. There is a well-established tradeoff between the depth of coverage and accuracy of quantitation when comparing label-free and labeled MS proteomics analysis with that label free protocol allowing deeper coverage given the same running time per sample, but has a significantly higher variance among samples.^[29] Thus, this study reports so far the most successful *Daphnia* proteomics results in terms of total number of proteins identified and in terms of balance between depth and reliability of quantification. It should be noted that both most successful *Daphnia* (ref. [4] and this one) considered a protein positively identified by just a single unique peptide. In this study, the increase of stringency of the minimal number of identifying peptides to 2 or 3 reduce the number of positively identified proteins to 3549 and 2662, respectively, which is still above the number of *Daphnia* proteins reported by most previous studies. This indicates that our dataset, along with that of ref. [4] provides a useful reference for future studies, the lack of which has been a challenge for *Daphnia* proteomics so far.

3.2. Inference for Genomic Reference

Success of protein identification by MS critically depends on the quality of protein reference used. The reference genome of the clone used in this experiment (TCO) includes a significantly larger number of gene models^[24] (31 472, as annotated in UniProt) than a more recently published reference from a closely related member of *D. pulex* group^[25] (18 440). Our results point to a significant contamination of these gene models with false positives. We were much more likely to detect proteins with known molecular functions (enzymes, structural proteins, transporters, and RNA-binding proteins in particular) than those whose molecular function is unknown (Table 1). Of course, this can be partially explained by the

fact that proteins with low expression are more likely not to be assigned any known molecular function; additionally, such proteins also include *Daphnia-specific* ones with no detectable homology in better annotated genomes. The lower number of peptides by which proteins with unknown MF have been identified may be indicative of their lower expression. Yet, it is also likely that we failed to detect at least some of these proteins because they are artifacts of false gene models. Indeed, the highly significant enrichment of mutual best BLAST hit proteins between the two independently sequenced, assembled, and annotated genomes (among proteins that we were able to detect) suggests that low detection success is enriched in proteins with less reliable independent protein or RNA confirmation.

3.3. Functional Differences Between Life Cycle Stages

We observed several a priori expected instances of differential expression between embryos and adults and between sexually and asexually produced embryos, which are, respectively, dormant and subitaneous. These signal include upregulation of vitellogenins and downregulation of the muscle proteins, actin and myosin in the embryos and upregulation of several stress resistance-related groups of proteins in the diapausing embryos, including small HSPs, peroxidases, trehalose synthesis enzymes, and the single PIMT ortholog, consistent with the previous studies on desiccation-resistant organisms.^[22] While overall expression of all peroxidases combined slightly increased along the life cycle (Figure 2B), there was evidence of numerous individual peroxidases that showed ephippium-specific expression (Figure 3B). Similarly, the overrepresentation analysis of BPs revealed overrepresentation of proteins involved in response to stress among those with higher relative abundance in the single ephippium sample. Furthermore, we observed a significant overrepresentation of proteins implicated in DNA replication-and mitosis-related BPs in subitaneous embryos relative to both the adults and the single ephippia sample (Tables S1, S5, and S6, and Figure S1, Supporting Information). This is consistent with the hypothesis that adults *Daphnia* are largely postmitotic (despite continued growth and molting throughout the lifespan) and any DNA replication may be limited to tissue undergoing endopolyploidization.^[16] Overrepresentation of catabolism and respiration processes among proteins with higher abundance in adults (Table S2, Supporting Information) is consistent with higher energetic expenditures in adults relative to embryos.

Although the lack of replication within the diapausing embryos prevented us for more in-depth analysis, it is clear that the procedure we employed is capable of detecting life cycle differences. The paucity of DNA-binding proteins among those upregulated in embryos, puzzling at the first glance, can be readily explained by lower number of nuclei in the embryonic samples than in a comparable amount of adult tissues, leading to under sampling of proteins with nuclear localization. Overall, uncharacterized proteins were significantly underrepresented in any BP gene ontology comparison. However, a significant overrepresentation of proteins with unknown MFs among those that are upregulated in the diapausing embryos was observed. It indicates that those proteins with no assignable ontology that can be identified by proteomics are likely to include diapause-specific proteins, including possibly yet unknown proteins, with functionality related to adaptations for freezing-, heat-, and desiccation tolerance.

4. Experimental Section

Sample Preparation:

Adult and juvenile *Daphnia* as well as subitaneous (asexually produced non-diapausing) embryos and resting eggs (ephippia) were sampled from a mass culture of *D. pulex* “TCO” clone. In order to roughly match the amount of protein, three adult animals, ten juvenile, and 30 embryos as well as 11 ephippia per sample were used, resulting in three adult females (over 15 days old), three juvenile (3–5 days old), three embryonic samples (within 2 days after egg-laying), and a single ephippium sample. Adults and juveniles were sampled from a mass culture; the subitaneous embryos were dissected from the same females used for the “adult” samples. The ephippia were collected from the bottom of the culture tank within 7 days after fertilization/shedding by the female. The ephippia were homogenized whole, thus containing both embryos and maternal tissues. The single ephippia sample available makes all the comparisons that involve this un-replicated sample preliminary in nature as they are severely affected by likely variation among biological specimens and preparation procedures. This is an important caveat of this part of the analysis. However, at least for the overrepresentation of GO terms analysis (see below), this was partly ameliorated by downstream statistical analysis eliminating any GO terms associated with false positives caused by lack of replication within the ephippial embryos category.

Following refs. [30] and [31], *Daphnia* were lysed in a buffer containing 75 mM NaCl, 3% SDS, 1 mM NaF, 1 mM β -glycerophosphate, 1 mM sodium orthovanadate, 10 mM sodium pyrophosphate, 1 mM PMSF, and 1 \times Roche Complete Mini EDTA free protease inhibitors in 50 mM HEPES, pH 8.5. Lyses was supported by using a pellet pestle. Lysates were then sonicated for 5 min in a sonicating water bath before cellular debris was pelleted by centrifugation at 14 000 rpm for 5 min. Proteins were then reduced with DTT and alkylated with iodoacetamide as previously described and precipitated via methanol-chloroform precipitation.^[30] Precipitated proteins were reconstituted in 300 μ L of 1 M urea in 50 mM HEPES, pH 8.5 and digested in a two-step process starting with an overnight digest at room temperature with Lys-C (Wako) followed by 6 h of digestion with trypsin (sequencing grade, Promega) at 37 °C. The digest was acidified with TFA and peptides were desalted with C₁₈ SPE (Sep-Pak, Waters) as previously described.^[30] The concentration of the desalted peptide solutions was measured with a BCA assay, and peptides were aliquoted into 50 μ L portions, which were dried under vacuum and stored at –80 °C until they were labeled with ten-plex TMT reagents (Thermo Scientific) as described previously.^[30] Pooled samples were desalted via C₁₈ SPE on Sep-Pak cartridges as described above and subjected to basic pH reversed-phase liquid chromatography (bRPLC)^[30] over a 4.6 mm \times 250 mm ZOR-BAX Extend C₁₈ column (5 μ m, 80 Å, Agilent Technologies) with concatenated fraction combining as previously described, and 12 fractions were subjected to quantitative proteomics analysis.

Liquid Chromatography-Mass Spectrometry:

All LC-MS2/MS3 experiments were conducted on an Orbitrap Fusion (Thermo Fisher Scientific) coupled to an Easy-nLC 1000 (Thermo Fisher Scientific) with chilled autosampler. Peptides were separated on an in-house pulled, in-house packed microcapillary column (inner diameter, 100 μ m; outer diameter, 360 μ m). Columns were packed to a final

length of 30 cm with GP-C₁₈ (1.8 μm , 120 \AA , Sepax Technologies). Peptides were eluted with a linear gradient from 11% to 30% ACN in 0.125% formic acid over 165 min at a flow rate of 300 nL min^{-1} while the column was heated to 60 $^{\circ}\text{C}$. ESI was achieved by applying 1500 V through a stainless steel T-junction at the inlet of the microcapillary column. The Orbitrap Fusion was operated in data-dependent mode, with a survey scan performed over an m/z range of 500–1200 in the Orbitrap with a resolution of 6×10^4 , automatic gain control (AGC) of 5×10^5 , and a maximum injection time of 100 ms. The most abundant ions detected in the survey scan were subjected to MS2 and MS3 experiments to be acquired in a 5 s experimental cycle. For MS2 analysis, doubly charged ions were selected from an m/z range of 600–1200, and triply and quadruply charged ions from an m/z range of 500–1200. The ion intensity threshold was set to 5×10^4 and the isolation window to 0.5 m/z . Peptides were isolated using the quadrupole and fragmented using CID at 30% normalized collision energy at the rapid scan rate using an AGC target of 1×10^4 and a maximum ion injection time of 35 ms. MS3 analysis was performed using synchronous precursor selection (SPS).^[32,33] Up to six MS2 precursors were simultaneously isolated and fragmented for MS3 analysis with an isolation window of 2.5 m/z and HCD fragmentation at 55% normalized collision energy. MS3 spectra were acquired at a resolution of 5×10^4 with an AGC target of 5×10^4 and a maximum ion injection time of 86 ms. The lowest m/z for the MS3 scans was set to 110.

Data Processing and Analysis:

Data were processed using an in-house developed software suite.^[34] MS2 data were annotated using the Sequest algorithm^[35] to search the set of *D. pulex* proteins present in UniProt database^[36] (sequences downloaded July 28, 2018) with the addition of known contaminants such as trypsin, and a target-decoy database strategy was applied to measure false discovery rates of peptide and protein identifications (Supporting Information 1). Searches were performed on combined data from all samples with a 50 ppm precursor mass tolerance; ten-plex TMT tags on lysine residues and peptide *n*-termini (+229.162932 Da) and carbamidomethylation of cysteines (+57.02146 Da) were set as static modifications and oxidation of methionine (+15.99492 Da) as a variable modification. Data were filtered to a peptide and protein false discovery rate of <1% using the target-decoy search strategy.^[37] MS2 assignments were filtered using linear discriminant analysis using a combined score generated from the following peptide and spectral properties: XCorr, dCn, number of missed tryptic cleavages, mass deviation, and peptide length (see Supporting Information 2).^[34] The probability of an incorrect peptide annotation was calculated using a posterior error LDA score histogram from sorting peptide annotations from the forward and the reversed database.^[34] The probability of an incorrect protein assignment was calculated by multiplying the LDA probabilities of all peptides assigned to the protein.^[34] Peptides that matched to more than one protein were assigned to that protein containing the largest number of matched redundant peptide sequences following the law of parsimony.^[34] TMT reporter ion intensities were extracted from the MS3 spectra selecting the most intense ion within a 0.003 m/z window centered at the predicted m/z value for each reporter ion and spectra were used for quantification if the sum of the S/N values of all reporter ions was >100 and the isolation specificity for the precursor ion was ≥ 0.5 . Protein intensities were calculated by summing the TMT reporter ions for all peptides assigned to a protein.

Intensities were first normalized by the average intensity across all samples relative to the median average across all proteins. In a second normalization step protein intensities measured for each sample were normalized by the average of the median protein intensities measured across the samples. Principal components analysis was conducted in R using PCA function of FactoMineR 1.40 package^[38] with data scaled to unit variance. Biological processes (BP) GO terms overrepresentation analysis was conducted by PANTHER14.1 gene ontology tool (<http://www.pantherdb.org>; ref. [39]) using *D. pulex* UniProt annotations and proteins showing differences among life stages with FDR < 0.01. FDR for Fisher's Exact Test for overrepresentation was set to 0.05.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

L.P. and M.K. were supported by NIH grant R01HD091846 and by a Paul G. Allen Family Foundation grant to MK. The authors are grateful to Neel Aluru for providing *D. pulex*TCO clone and to Yan Galimov and Oleg Gusev for unpublished data on PIMT overexpression in *Daphnia* ephippia.

References

- [1]. Fröhlich T, Arnold GJ, Fritsch R, Mayr T, Laforsch C, BMC Genomics 2009, 10, 171. [PubMed: 19383153]
- [2]. Otte KA, Proteomic analysis of stress responses in *Daphnia*. Ph.D. Thesis, LMU München, (Munich) 2016.
- [3]. Steevensz A, Gombar R, Vergilino R, Cristescu ME, Vacratsis PO, Proteomics, 18, 2018, 1700460.
- [4]. Borgatta M, Hernandez C, Decosterd LA, Chèvre N, Waridel P.J. Proteome Res, 2015, 14, 279. [PubMed: 25350372]
- [5]. Schwerin S, Zeis B, Lamkemeyer T, Paul RJ, Koch M, Madlung J, Fladerer C, Pirow R, BMC Physiol 2009, 9, 8. [PubMed: 19383147]
- [6]. Becker D, Reydelet Y, Lopez JA, Jackson C, Colbourne JK, Hawat S, Hippler M, Zeis B, Paul RJ, BMC Genomics 2018, 19, 376. [PubMed: 29783951]
- [7]. Zeis B, Lamkemeyer T, Paul RJ, Nunes F, Schwerin S, Koch M, Schutz W, Madlung J, Fladerer C, Pirow R, BMC Physiol 2009, 9, 7. [PubMed: 19383146]
- [8]. Trotter B, Otte KA, Schoppmann K, Hemmersbach R, Fröhlich T, Arnold GJ, Laforsch C, npj Microgravity 2015, 1, 15016. [PubMed: 28725717]
- [9]. Otte KA, Fröhlich T, Arnold GJ, Laforsch C, BMC Genomics 2014, 15, 306. [PubMed: 24762235]
- [10]. Otte KA, Schrank I, Fröhlich T, Arnold GJ, Laforsch C, Mol. Ecol 2015, 24, 3901. [PubMed: 26111196]
- [11]. Effertz C, Müller S, von Elert E,J. Proteome Res. 2015, 14, 888. [PubMed: 25494525]
- [12]. Rainville LC, Carolan D, Varela AC, Doyle H, Sheehan D, Analyst 2014, 139, 1678. [PubMed: 24482795]
- [13]. Rainville LC, Coelho AV, Sheehan D, Environ. Toxicol. Chem 2015, 34, 84. [PubMed: 25263122]
- [14]. Toumi H, Boumaiza M, Immel F, Sohm B, Féltén V, Ferard JF, Aquat. Toxicol 2014, 148, 40. [PubMed: 24441279]
- [15]. Dircksen H, Neupert S, Predel R, Verleyen P, Huybrechts J, Strauss J, Hauser F, Stafflinger E, Schneider M, Pauwels K, Schoofs L, Grimmelikhuijzen CJP,J. Proteome Res 2011, 10, 4478. [PubMed: 21830762]
- [16]. Smirnov N. Physiology of the Cladocera, 2nd ed Academic Press, London, UK 2017.

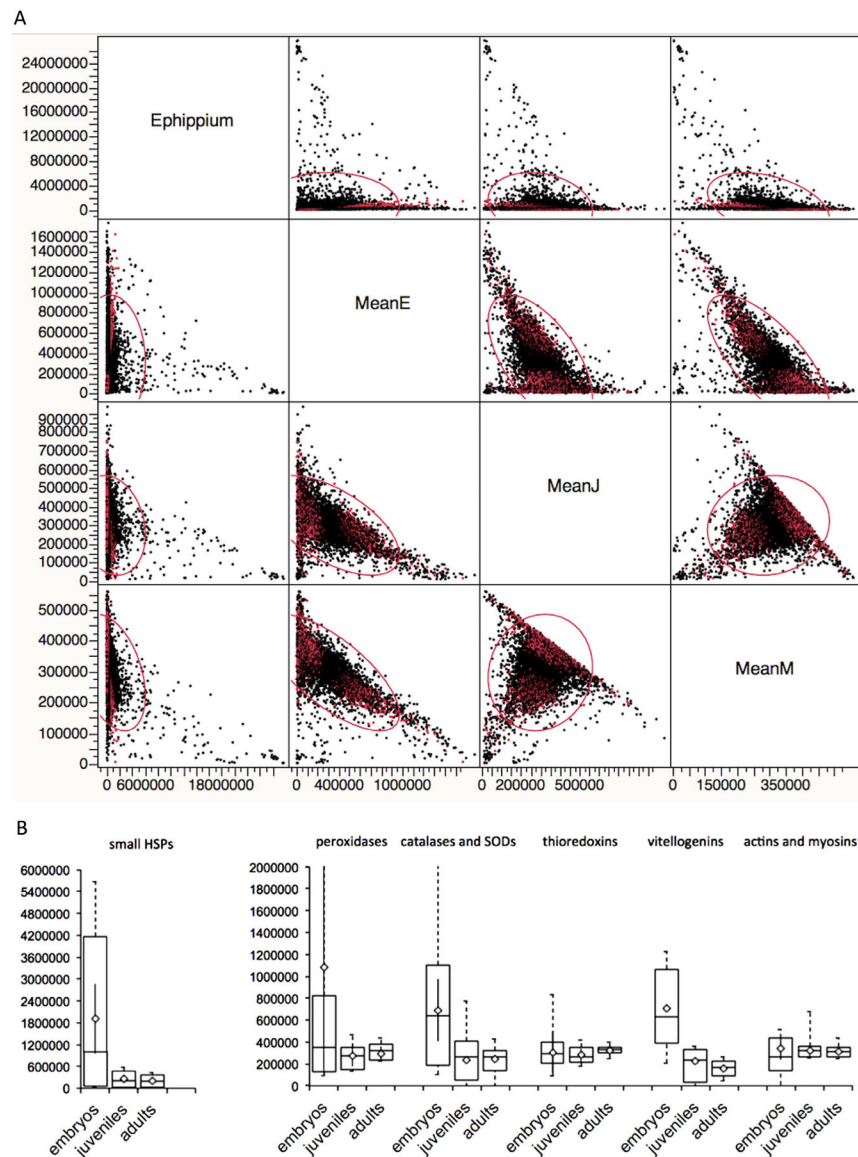
- [17]. Kato Y, Tokishita S, Ohta T, Yamagata H, Gene 2004, 334, 157. [PubMed: 15256265]
- [18]. Watanabe M, Appl. Entomol. Zool 2006, 41, 15.
- [19]. Hengherr S, Heyer AG, Bruemmer F, Schill RO, Physiol. Biochem. Zool 2011, 84, 147. [PubMed: 21460525]
- [20]. Bakthisaran R, Tangirala R, Rao CM, Biochim. Biophys. Acta 2015, 1854, 291. [PubMed: 25556000]
- [21]. Tan J, MacRae TH, PLoS One 2018, 13, e0200153. [PubMed: 29979776]
- [22]. Gusev O, Suetsugu Y, Cornette R, Kawashima T, Logacheva MD, Kondrashov AS, Penin AA, Hatanaka R, Kikuta S, Shimura S, Kanamori H, Katayose Y, Matsumoto T, Shagimardanova E, Alexeev D, Govorun V, Wisecaver J, Mikheyev A, Koyanagi R, Fujie M, Nishiyama T, Shigenobu S, Shibata TF, Golygina V, Hasebe M, Okuda T, Satoh N, Kikawada T, Nat. Commun 2014, 5, 4784. [PubMed: 25216354]
- [23]. Galimov Y and Gusev O, personal communication.
- [24]. Colbourne JK, Pfrender ME, Gilbert D, Thomas WK, Tucker A, Oakley TH, Tokishita S, Aerts A, Arnold GJ, Basu MK, Bauer DJ, Caceres CE, Carme L, Casola C, Choi J-H, Detter JC, Dong Q, Dusheyko S, Eads BD, Fröhlich T, Geiler-Samerotte KA, Gerlach D, Hatcher P, Jogdeo S, Krijgsveld J, Kriventseva EV, Kültz D, Laforsch C, Lindquist E, Lopez J, et al., Science 2011, 331, 555.. [PubMed: 21292972]
- [25]. Ye Z, Xu S, Spitze K, Asselman J, Jiang X, Ackerman MS, Lopez J, Harker B, Raborn RT, Thomas WK, Ramsdell J, Pfrender ME, Lynch M, G3 2017, 7, 1405. [PubMed: 28235826]
- [26]. Pfrender M, personal communication.
- [27]. Yampolsky L, unpublished.
- [28]. Loomis SH, Madin KAC, Crowe JH, J. Exper. Zool.: Comp. Physiol. Biochem 1980, 211, 311.
- [29]. Sonnett M, Yeung E, Wühr M, Anal. Chem 2018, 90, 5032. [PubMed: 29522331]
- [30]. Edwards A, Haas W, Methods Mol. Biol 2016, 1394, 1. [PubMed: 26700037]
- [31]. Lapek JD Jr., Greninger P, Morris R, Amzallag A, Pruteanu-Malinici I, Benes CH, Haas W, Nat. Biotechnol 2017, 35, 983. [PubMed: 28892078]
- [32]. Ting L, Rad R, Gygi SP, Haas W, Nat. Methods 2011, 8, 937. [PubMed: 21963607]
- [33]. McAlister GC, Nusinow DP, Jedrychowski MP, Wühr M, Huttlin EL, Erickson BK, Rad R, Haas W, Gygi SP, Anal. Chem 2014, 86, 7150. [PubMed: 24927332]
- [34]. Huttlin EL, Jedrychowski MP, Elias JE, Goswami T, Rad R, Beausoleil SA, Villen J, Haas W, Sowa ME, Gygi SP, Cell 2010, 143, 1174.. [PubMed: 21183079]
- [35]. Eng JK, McCormack AL, Yates JR, J. Am. Soc. Mass Spectrom 1994, 5, 976. [PubMed: 24226387]
- [36]. The UniProt Consortium, Nucleic Acids Res. 2019, 47, D506. [PubMed: 30395287]
- [37]. Elias JE, Gygi SP, Nat. Methods 2007, 4, 207. [PubMed: 17327847]
- [38]. Lê S, Josse J, Husson F, J. Stat. Softw 2008, 25, 1.
- [39]. Mi H, Muruganujan A, Thomas PD, Nucleic Acids Res. 2013, 41, D377. [PubMed: 23193289]

Significance Statement

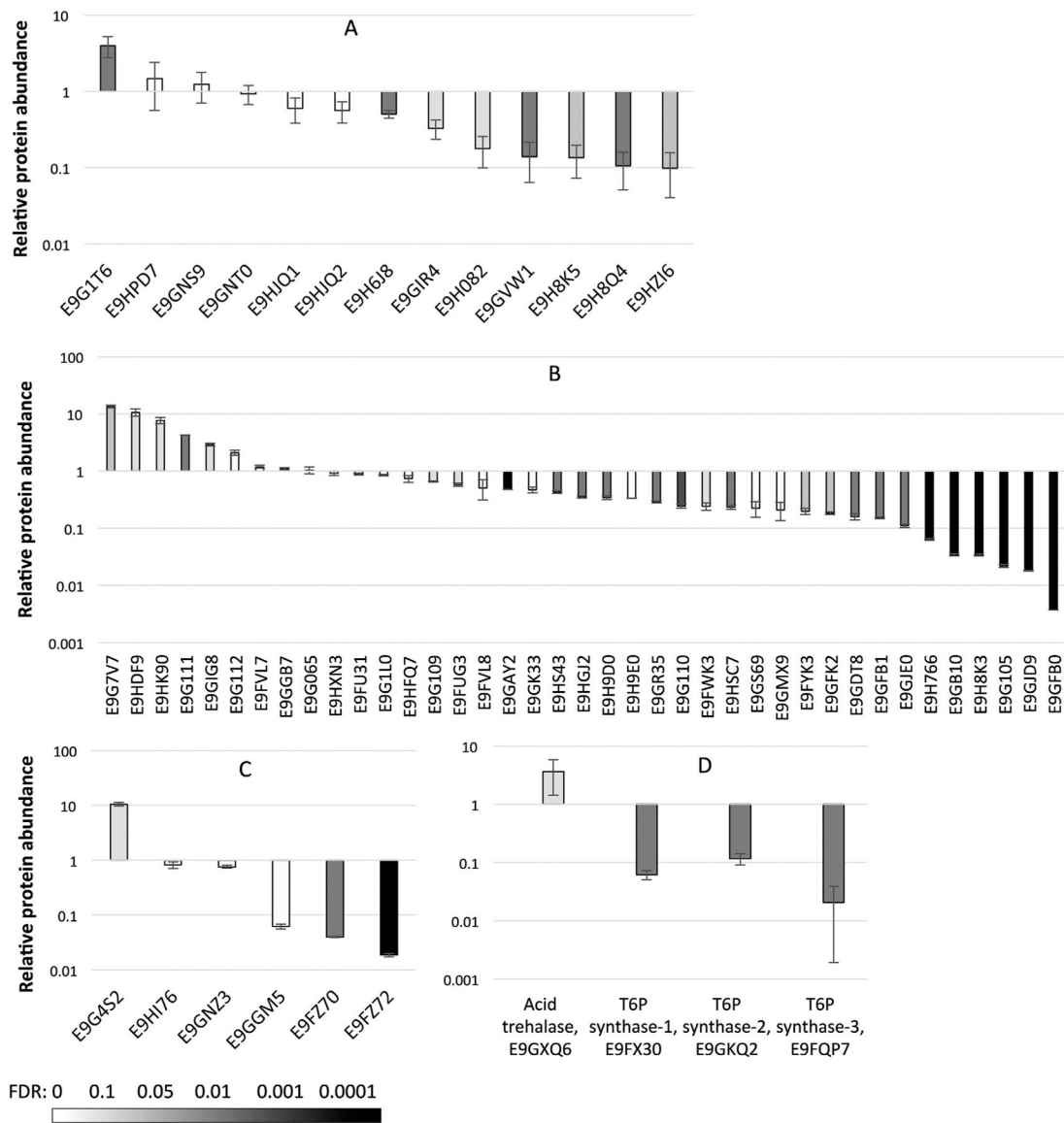
Freshwater plankton crustacean *Daphnia* is rapidly becoming a model organism of choice for ecological and developmental genomics. While there have been several advances toward establishing the protocols and reference datasets for proteomics, a detailed dataset covering several main steps of asexual and sexual phases of *Daphnia* life cycle is not yet available. Moreover, different versions of *D. pulex* genome differ in the number of protein-coding genes identified; it is unclear whether these differences are caused by differences between sequenced genotypes or between gene model methodology used. In this study we report LC-MS2/MS3 proteomes of whole-body adult females, juvenile females, asexually produced embryos, and diapausing eggs capable of surviving freezing and desiccation.



Figure 1. Ten samples analyzed on the plane of the first two principal components explaining 47.1% and 24.4% of variance in protein-specific signals.

**Figure 2.**

A) Scatterplot of relative abundances of 5098 detected proteins in the ehippium (Eph), asexually produced embryos (MeanE), juveniles (MeanJ), and adult females (MeanM). The values are median abundances averaged across replicates. Red: proteins with FDR of the t -test of all embryos combined versus adult females < 0.01 . B) Relative abundances of several functional groups of proteins a priori expected to differ among life cycle stages. Boxplots represent minimums, quartiles and maximum across all proteins within the functional category, diamonds and solid vertical lines—means and SEs across proteins. Note a different scale for small HSPs.

**Figure 3.**

Relative abundance of select proteins in adults versus embryos and/or the ephippium. A) Vitellogenins in juveniles and adults relative to embryos of both types; B) peroxidases in the subitaneous embryos relative to ephippium; C) small HSPs in the subitaneous embryos relative to ephippium; D) trehalose metabolism proteins in the subitaneous embryos relative to ephippium. A) Average in all embryos is set to 1; B-D) ephippium is set to 1. White to black shades indicate individual proteins FDR for differential expression test.

Table 1.

Enrichment of molecular function categories among proteins discovered by MS.

Molecular function	Observed	Expected	<i>p</i> -Value	Portion discovered	Number of peptides (SE)
Enzyme	2598	1443.1	****	0.29	6.64 (0.24)
Structural	243	87.7	****	0.44	14.38 (2.23)
Transporter/channel	272	185.6	****	0.24	13.73 (4.24)
Receptor	100	102.0	n.s.	0.16	4.02 (0.62)
Transcription factor	30	70.6	****	0.07	2.63 (0.65)
Other DNA binding	43	29.7	**	0.23	3.88 (0.71)
RNA binding	43	13.2	****	0.52	4.53 (0.62)
Other	1189	825.2	****	0.23	6.64 (0.24)
Unknown	580	2340.9	****	0.04	2.89 (1.16)
Total	5098			0.16	6.70 (0.38)

**
 $p < 0.01$;

 $p < 0.0001$; individual cells χ^2 tests.

Orange, enriched categories; blue, under-represented categories.

Table 2.

Enrichment analysis of proteins of seven common molecular functions (plus the “other” and “unknown” categories) among proteins up- or downregulated (FDR < 0.01) in comparisons between embryos and adults as well as between the ephippium and asexually produced embryos.

Molecular function	Embryos versus juveniles + adults			Ephippium versus adults			Asexually produced embryos versus adults			Ephippium versus asexually produced embryos			Out of	
	Observed	Expected	χ^2	Observed	Expected	χ^2	Observed	Expected	χ^2	Observed	Expected	χ^2		
Enzyme	289	288.9		420	433.6		414	426.0		602	611.5		2598	
Structural	14*	27.0		16***	40.6		21**	39.8		50	57.2		243	
Transporter/channel	25	30.25		33	45.4		39	44.6		58	64.0		272	
Receptor	7	11.1		17	16.7		11	16.4		29	23.45		100	
TF	5	3.33		7	5.00		4	4.91		2	7.06		30	
Other DNA binding	9	4.78		8	7.18		17***	7.05		3	10.1		43	
RNA binding	5	4.78		12	7.18		10	7.05		7	10.1		43	
Other	158*	132.2		230*	198.5		248***	195.0		245	279.9		1189	
Unknown	23	17.94		108	96.8		72*	95.1		204****	136.5		580	
Total	567			851			836			1200			5098	
Table-wide log-likelihood	22.7	$P < 0.0038$		38.8	$P < 5.4 \times 10^{-6}$		54.1	$P < 6.5 \times 10^{-9}$		65.2	$P < 4.4 \times 10^{-11}$		19.6	$P < 0.012$

Proteomics. Author manuscript; available in PMC 2020 March 10.

Orange, enriched categories; blue, under-represented categories. p -Values (cell's χ^2):

* $p < 0.025$;

** $p < 0.01$;

*** $p < 0.001$;

**** $p < 0.0001$ (**** and *****: survives table-wide Bonferroni adjustment with $p < 0.05$ and $p < 0.01$, respectively).

Bottom: table-wide log-likelihood contingency analysis.