



OPEN

Abiotic and past climatic conditions drive protein abundance variation among natural populations of the caddisfly *Crunoecia irrorata*

Joshua Niklas Ebner¹  , Danilo Ritz² & Stefanie von Fumetti¹

Deducing impacts of environmental change on species and the populations they form in nature is an important goal in contemporary ecology. Achieving this goal is hampered by our limited understanding of the influence of naturally occurring environmental variation on the molecular systems of ecologically relevant species, as the pathways underlying fitness-affecting plastic responses have primarily been studied in model organisms and under controlled laboratory conditions. Here, to test the hypothesis that proteome variation systematically relates to variation in abiotic conditions, we establish such relationships by profiling the proteomes of 24 natural populations of the spring-dwelling caddisfly *Crunoecia irrorata*. We identified protein networks whose abundances correlated with environmental (abiotic) gradients such as in situ pH, oxygen- and nitrate concentrations but also climatic data such as past thermal minima and temperature seasonality. Our analyses suggest that variations in abiotic conditions induce discrete proteome responses such as the differential abundance of proteins associated with cytoskeletal function, heat-shock proteins and proteins related to post-translational modification. Identifying these drivers of proteome divergence characterizes molecular “noise”, and positions it as a background against which molecular signatures of species’ adaptive responses to stressful conditions can be identified.

Organisms are able to maintain fitness via physiological and phenotypic plasticity by continuously sensing and integrating environmental cues^{1–3}. These responses allow them to adjust to environmental variation, a feat that influences the outcome of evolution and the likelihood of extinction^{4–6}. Due to the reversible nature of plasticity and its contribution to fitness, it has the potential to moderate the loss of global biodiversity during environmental change^{7–10}. The advantage of species’ plasticity during such change depends on the threshold-limits of their more- or less robust molecular systems that underlie plastic responses^{11,12}. In many, if not all species, these networks of interacting macromolecules are likely tuned to the fluctuations in the environment over space and time. Accordingly, substantial variation in gene expression and protein abundance profiles within and between populations can be observed^{13–18}. Attributing this variation to environmental cues experienced by the organism can not only give insights into the molecular mechanisms underlying regulatory plastic responses^{19–21} but also inform about population health and how a population at one site responds to environmental change not encountered by another population. Often, the environmental cues and underlying molecular pathways involved are largely unknown but increasingly studied by identifying large-scale gene expression changes, providing fundamental knowledge on adaptive processes in response to environmental challenges such as global warming and pollution^{22–25}.

It has been repeatedly shown that the correlation between transcript and protein abundance is modest^{26–28}. Phenomena such as pre, co- and post-transcriptional and translational modification^{29,30}, protein turnover³¹ and the inherently stochastic nature of gene expression contribute to this pronounced difference^{32,33}. Accordingly, the molecular signatures of plasticity triggered by environmental (selective) pressure might frequently only be detectable at the protein level and against a pre-established background of environmentally induced proteome variation. Our knowledge on the concrete environmental cues that induce protein abundance variation among

¹Geocology Research Group, Department of Environmental Sciences, University of Basel, Basel, Switzerland. ²Proteomics Core Facility, University of Basel, Biozentrum Basel, Switzerland. ✉email: joshua.ebner@unibas.ch

natural populations is limited due to a lack of field-based studies that relate abundance variation to variation in environmental conditions encountered at the sites.

In this study, we proposed to tackle two fundamental questions. First, what is the standing protein abundance variation between natural populations of different geographic regions? Second, are differences in the abiotic environment linked to this variation, and, if so, do functional profiles of proteins reflect these differences? To this end, 24 populations of the caddisfly *Crunoecia irrorata* (Trichoptera: Lepidostomatidae, Curtis 1834) were used as a study system for two reasons: (1) The increasing need to monitor population health of non-model species (i.e. species without sequenced genomes) via systems-wide approaches and (2) the continuous exposure of larval individuals to surrounding aqueous conditions. Population-wide liquid chromatography-tandem mass spectrometry (LC-MS/MS) was used to examine abundance variation patterns amongst commonly identified proteins within this species. We tested the hypothesis that this variation can be systematically attributed to abiotic measurements including elevation, geographic distance and past climatic conditions. As cellular conditions change rapidly, often in seconds or minutes^{34,35}, functional physiological changes must necessarily occur before or during transcriptional and translational changes. We therefore focus on abiotic point-measurements, i.e. we measured abiotic variables at the same time as organismal sampling to obtain a “physiological snapshot” of these natural populations. Such a characterization of the “normal state” of a given system—the ranges of its variation³⁶—could aid in defining baselines of molecular responses that must be exceeded in order to be considered potentially adaptive changes or physiological responses triggered by changes in the environment.

Material and methods

Organism collection and environmental variables. Eight *C. irrorata* larval individuals of similar sizes were collected from 24 (near-) natural freshwater springs, i.e. from eight springs in three study regions across Germany: Rhoe Biosphere Reserve (R), Harz National Park (H) and Black Forest (BF) (Supplementary Table S1). Sites were distributed in a randomized design between low (275 m) and high (967 m) elevations. Individuals from one spring were pooled into one sample since pooling has been shown to match mean protein abundances of the individuals making up the pool³⁷, thereby reducing the biological variation compared with that between individuals³⁸. After identification, larvae were washed with dH₂O and transferred on-site into 5.0 ml Protein LoBind tubes (Eppendorf, Germany) containing 1 ml of RNA-later (Invitrogen). RNA-later was chosen, since snap-freezing was infeasible and, despite its limitations, proteins of various tissues are well preserved and biological information is largely kept intact^{39–41}. Samples were stored at ~4 °C in a portable fridge (Dometic TropiCool TC 33) until storage at –20 °C was possible. From each spring, additional larvae were collected, conserved in EtOH and their head capsule widths ($n = 88$) were measured using a Leica S9i stereo microscope, serving as a proxy control for possible instar stage distribution differences between sampling regions. Abiotic variables were measured at each spring using a multi-parameter portable meter (MultiLine Multi 3650 IDS). Inorganic nutrient contents were determined from 30 ml of spring water via Ion Chromatography (IC) using a 940 Professional IC Vario ONE/SeS/PP (Metrohm) and iron (Fe) content via Inductively Coupled Plasma Emission Spectrometry (ICP-OES) using a 5100 ICP-OES (Agilent Technologies). In order to assess the influence of past climatic differences between sites on protein abundance profiles, we collected macroclimate data (BioClim1–19) from the WorldClim v.2⁴² database based on spring coordinates using the raster package v.3.0.12⁴³ (resolution: 5-min of a longitude/latitude degree), presenting various indices of environmental data calculated from 30 years of average monthly data (1971–2000).

Protein extraction. Prior to protein extraction, samples were thawed on ice and centrifuged at 2,800g. RNA-later was decanted and larvae transferred from RNA-later into 1.5 ml protein LoBind tubes containing 1 ml cold 10 × phosphate-buffered saline (PBS). After vortexing for 10 s, larvae were transferred into 1.5 ml protein LoBind tubes containing 400 µl cold lysis buffer (1% sodium deoxycholate [SDC], 10 mM TCEP, 100 mM Tris, pH 8.5 [adjusted with NaOH/HCl]). Samples were in-tube homogenized with a sterile glass pestle followed by 10 min incubation at 4 °C and 3 × 1 ultrasonication (Bandelin Sonoplus HD270). Samples were spun at 2800g for 10 min at –4 °C, and 200 µl supernatant was subsequently transferred to a new tube. To desalt absorbed RNA-later, trichloroacetic acid (TCA)/acetone precipitation was performed according to the protocol by Luis Sanchez⁴⁷, resulting in a pellet that was dissolved in 200 µl lysis buffer. After 10 × 1 s sonication, samples were incubated for 10 min at 95 °C at 10 g in a Thermomixer C (PCR 96 heating block, Eppendorf). At this point, protein concentrations were determined via a Bicinchoninic acid assay (BCA; Thermo Scientific Pierce BCA Protein Assay Kit) according to manufacturer’s instructions. After letting samples cool down at room temperature, they were spun down at 2800g for 10 s. Four µl of 0.75 M chloroacetamide solution was added to each sample and incubated at 37 °C for 10 min at 28 g and again spun down at 2800g for 10 s. After checking if the pH of each sample was around 8, 1 µg trypsin (Sequencing Grade Modified Trypsin, Promega) was added to 50 µg extracted proteins per sample which then were digested overnight at 37 °C and 300 rpm. Samples were acidified with 50 µl 5% trifluoroacetic acid (TFA) and peptides were purified using PreOmics cartridges (Martinsried) according to manufacturer’s instructions. Eluted peptides were transferred to a 96-well plate and concentrated to dryness by applying vacuum for 2 h. Peptides were subsequently dissolved in 20 µl 0.1% formic acid by 10 × 1 s ultrasonication and shaking at 28 g at 25 °C for 5 min. After spinning dissolved peptides down at 1800g for 10 min, peptide concentrations were determined based on absorbance values using a SPECTROstar Nano Absorbance Plate Reader (BMG Labtech). Protein and peptide concentrations are given in Supplementary Table S2. Peptides were diluted to a concentration of 0.5 µg/µl in 0.1% formic acid. IRT peptides (Biognosys AG, Schlieren, Switzerland) were added to the wells to control for LC-MS performance, and samples were stored at –20 °C prior to LC-MS/MS analysis.

LC–MS/MS analysis. Samples were subjected to LC–MS/MS analysis using an Orbitrap Fusion Lumos Tribrid Mass Spectrometer fitted with an EASY-nLC 1200 (both Thermo Fisher Scientific) and a custom-made column heater set to 60 °C. Peptides were resolved using an RP–HPLC column (75 µm × 36 cm) packed in-house with C18 resin (ReproSil–Pur C18–AQ, 1.9 µm resin; Dr. Maisch GmbH) at a flow rate of 0.2 µl/min. The following gradient was used for peptide separation: from 5 to 12% B over 5 min to 35% B over 65 min to 50% B over 20 min to 95% B followed by 20 min at 95% B. Buffer A was 0.1% formic acid in water, and buffer B was 80% acetonitrile, 0.1% formic acid in water. The mass spectrometer was operated in Data-Dependent Acquisition (DDA) mode with a cycle time of 3 s between master scans. Each master scan was acquired in the Orbitrap at a resolution of 120,000 full width at half maximum (at 200 m/z, MS1) and a scan range from 375 to 1600 m/z followed by MS/MS (MS2) scans of the most intense precursors in the linear ion trap at “Rapid” scan rate with isolation of the quadrupole set to 1.4 m/z. Maximum ion injection time was set to 50 ms (MS1) and 35 ms (MS2) with an AGC target of 1.0E6 and 1.0E4, respectively. Monoisotopic precursor selection (MIPS) was set to peptide, and the intensity threshold was set to 5.0E3. Peptides were fragmented by HCD (higher-energy collisional dissociation) with collision energy set to 35%, and one microscan was acquired for each spectrum. The dynamic exclusion duration was set to 30 s.

Protein identification. Raw spectra (Thermo raw files) were submitted to an Andromeda⁴⁴ search in MaxQuant v1.6.10.43⁴⁵. The “match between runs” option was enabled (match time window: 0.7 min, alignment time window: 20 min). Instrument type was set to Orbitrap, precursor mass tolerance to 15 ppm and fragment ion tolerance to 0.05 Da. Enzyme specificity was set to fully tryptic, with a maximum of two missed cleavages. MS/MS spectra were searched against a previously described database consisting of translated gene-prediction sequences of Trichoptera species⁴⁶, additionally including candidate protein-coding regions of *Rhyacophila fasciata*⁴⁷; BioProject: PRJNA219600). All searches included a contaminants database (as implemented in MaxQuant, 267 sequences). For protein identification, unique and razor peptides were used. The peptide spectrum-match false discovery rate (FDR) and the protein FDR were set to 0.01 (based on the target-decoy approach). Oxidation of methionine (M) and acetylation (Protein N-term) were specified as variable and carbamidomethylation of cysteines (C) as fixed modifications. Enzyme specificity was set to “Trypsin/P”. Minimum peptide length was set to 7. The “evidence.txt” and “summary.txt” output files were used for LC–MS/MS quality control using R package artMS v.1.4.2⁴⁸. To investigate identification differences between the three regions, we analyzed samples in three independent MaxQuant runs (settings as above). Overlapping and unique Majority protein IDs (IDs of those proteins that have at least one half of the peptides that the leading protein has) were compared via a Venn diagram plotted via package VennDiagram v.1.6.20⁴⁹.

Statistical analysis. Data analysis was conducted in R v.3.6.2⁵⁰. Strongly correlating (Pearson’s $R^2 > 0.7$) abiotic and BioClim variables were identified with function cor.test and removed from further analysis (Supplementary Fig. S2). The in MaxQuant implemented label-free quantification (LFQ) option computed 54.2% missing values (Supplementary Fig. S3), making data imputation unreliable. Therefore, protein raw intensities were calculated through summation of peptide intensities⁵¹. Raw intensity values were \log_2 -transformed, quantile normalization was performed using function normalize.quantiles from package preprocessCore v.1.46.0⁵² and missing values ($n = 1512 = 5.37\%$) imputed using function impute.knn from package impute v.1.60.0⁵³. These relative abundance values were tested for association with abiotic variables: first, unsupervised non-metric multidimensional scaling (nMDS) ordination was performed to visualize clustering of protein abundances between populations and regions. A Bray–Curtis⁵⁴ distance matrix was calculated based on relative protein abundance values for each population and used as input to the metaMDS function implemented in vegan v.2.5.6⁵⁵ (Supplementary Fig. S4). Weighted gene co-expression network analysis (WGCNA) v.1.68⁵⁶ was used to identify suites of co-regulated proteins in an unsupervised way. These protein networks could identify important functional groups of proteins and candidate proteins underlying plastic responses, similar to gene expression analyses^{57,58}. A sample network was constructed to identify outlying samples with a standardized connectivity score of less than -2.5 ⁵⁹. A signed protein co-abundance network was constructed for all 24 populations independent of sampling region. The network was constructed with a soft threshold power (β) of 18 as this value was found to be appropriate by the function pickSoftThreshold to reach a scale-free topology index (R^2) of at least 0.90. We used the Dynamic Tree Cut approach to merge highly correlated modules using a height-cut of 0.20⁶⁰. The abundances of these modules can be summarized as the abundance of a single “eigengene”, calculated as the first principal component of the abundances of all proteins in a module across samples⁶⁰. These eigengenes were correlated (Pearson’s correlation) with abiotic variables to identify module–environment relationships (Supplementary Fig. S5). Analogous to the in situ abiotic variables, WGCNAs were independently performed for bioclimatic variables (WGCNA_{BC}, Supplementary Fig. S6) and for differentially abundant proteins (DAPs) between the three sampling regions (WGCNA_{DAP}, Supplementary Fig. S19). DAPs were identified using the Linear Models for Microarray Data (LIMMA) library⁶¹. A linear model was fit for each protein via the function lmFit and contrasts from the model fit, and summary statistics were computed via the functions contrasts.fit and eBayes. We considered proteins differentially abundant between regions if their \log_2 fold change (FC) was < -2 or > 2 and their adjusted p-value < 0.05 . We computed GLMs by regressing protein abundances against each continuous, abiotic response variable (i.e. abiotic gradient) using the stan_glm function in the rstanarm package v.2.17.4⁶². We termed a positive or negative abundance change of a specific protein with an abiotic gradient a “protein reaction norm”, akin to the concept of a reaction norm concept in the study of phenotypic plasticity and life-history theory^{1,63,64}. In this case, it describes the sensitivity of a protein, or of a set of proteins, to a change in some specific abiotic variable. Resulting protein reaction norms were assessed independently via scatterplots. To assess the effect of geographic distance on protein abundance divergence between populations, a Mantel test

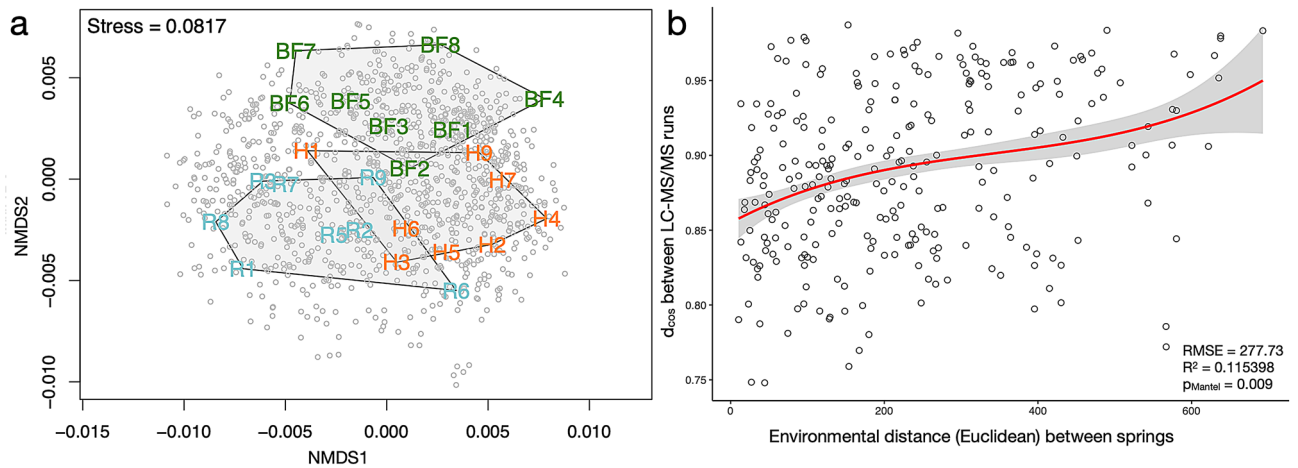


Figure 1. Ordination and spectral distance results. (a) Non-Metric Multi-Dimensional Scaling (nMDS) based on abundances of proteins identified in all populations illustrating the similarities and differences in abundance profiles of populations across three sampling regions. Relative proximity of sample-labels represents overall degree of abundance similarity between populations and grey points represent single proteins. (b) Correlation between environmental distances between springs and spectral distances between population-wide LC-MS/MS runs.

was performed between geographic- and (Euclidean) protein-distance matrices in *ade4* v.1.7.13⁶⁵. Geographic distances between springs were calculated based on longitude and latitude of springs using package *geosphere* v.1.5.10⁶⁶ and the implemented function *distm* (Vincenty great circle distances). Pairwise spectrum distances (cosine distances) between MS/MS runs (proteome-wide distances independent of peptide-identification and therefore of the homology-based protein database) were computed in R using the *DISMS2* algorithm⁶⁷.

Functional annotation of protein sequences. The first majority protein ID of each identified protein was parsed from the MaxQuant output and corresponding amino-acid sequences extracted from the database using function *ucsc-faSomeRecords* as implemented in *Bioconda*⁶⁸. All sequences were queried against the Swiss-Prot database⁶⁹ (accessed 10 January 2020) using stand-alone *blastp* v.2.2.28 with default parameters (E-value 0.001). Protein domains were determined using stand-alone *InterProScan* (IPS) v.5.39-77.0⁷⁰ with default parameters. All proteins were assigned to existing Cluster of Orthologous Groups (COGs)⁷¹ and Gene Ontology (GO) terms in *EggNOG* v5.0⁷² via *eggNOG-Mapper*⁷³ (Taxonomic scope: Insecta, Orthologs: all orthologs, GO evidence: non-electronic terms, E-value: 0.001). To test whether any GO terms were overrepresented in WGCNA modules and DAPs, we sorted proteins in selected modules by their Gene Significance (GS) for the significantly correlated abiotic variable and DAPs by their adjusted *p*-values and performed rank-based tests for each GO term assigned to these proteins by applying Kolmogorov–Smirnov tests via package *topGO* v.2.38.1⁷⁴ (“weight01” algorithm, *nodeSize*=10). Individual proteins associated with significant GO terms of interest can be easily identified using the “Supplemental R script”. To further identify and substantiate biological functions and processes associated with modules, a domain-based analysis was conducted. To this end, the IPS Protein Families (Pfam) annotations⁷⁵ of the proteins belonging to a certain module were used as entry for domain-centric GO (dcGO) enrichment (FDR *p*-value < 0.01)⁷⁶.

Results

Global drivers of protein abundance differentiation. Mass spectra matched to 4536 distinct peptide sequences (mean number of amino acids: 16, range 8–44). These peptide sequences mapped to 1173 proteins that were commonly identified in all 24 *C. irrorata* populations. Protein identifications largely overlapped (Supplementary Fig. S7b) and pairwise peptide-level correlations between samples were above 0.8. Protein abundance profiles exhibited variation between populations but showed association with sampling regions, whereby populations from any one region clustered together mostly along axis 2 of the nMDS (Fig. 1a). Therefore, proteins with significant, positive eigenvector loadings along this axis may play a dominant role in proteome differentiation over geographic range. The top proteins with positive loadings (> 0.005, *n* = 69) were enriched for GO terms *small molecule metabolic process* (GO:0044281, BP), *oxidoreductase activity* (GO:0016491, MF) and *mitochondrion* (GO:0005739, CC). More similar protein abundance profiles were found between populations occupying adjacent springs when compared to springs further apart ($p_{\text{Mantel}} = 0.006$, Supplementary Fig. S8) and pairwise spectrum distances between LC-MS/MS runs were positively correlated with environmental distances between springs ($p_{\text{Mantel}} = 0.009$), whereby spectral profiles were more similar between springs of similar abiotic environmental conditions (Fig. 1b).

Environmental distance between springs did not correlate with geographic distance ($p = 0.33$; Supplementary Fig. S9). One-hundred eighty-two DAPs were identified between sampling regions. These proteins were primarily enriched for GO terms related to *primary metabolic process* (GO:0044238; e.g. mitochondrial-processing peptidase subunit beta and probable NADH dehydrogenase [ubiquinone] iron-sulfur protein 6) and *gene expression*

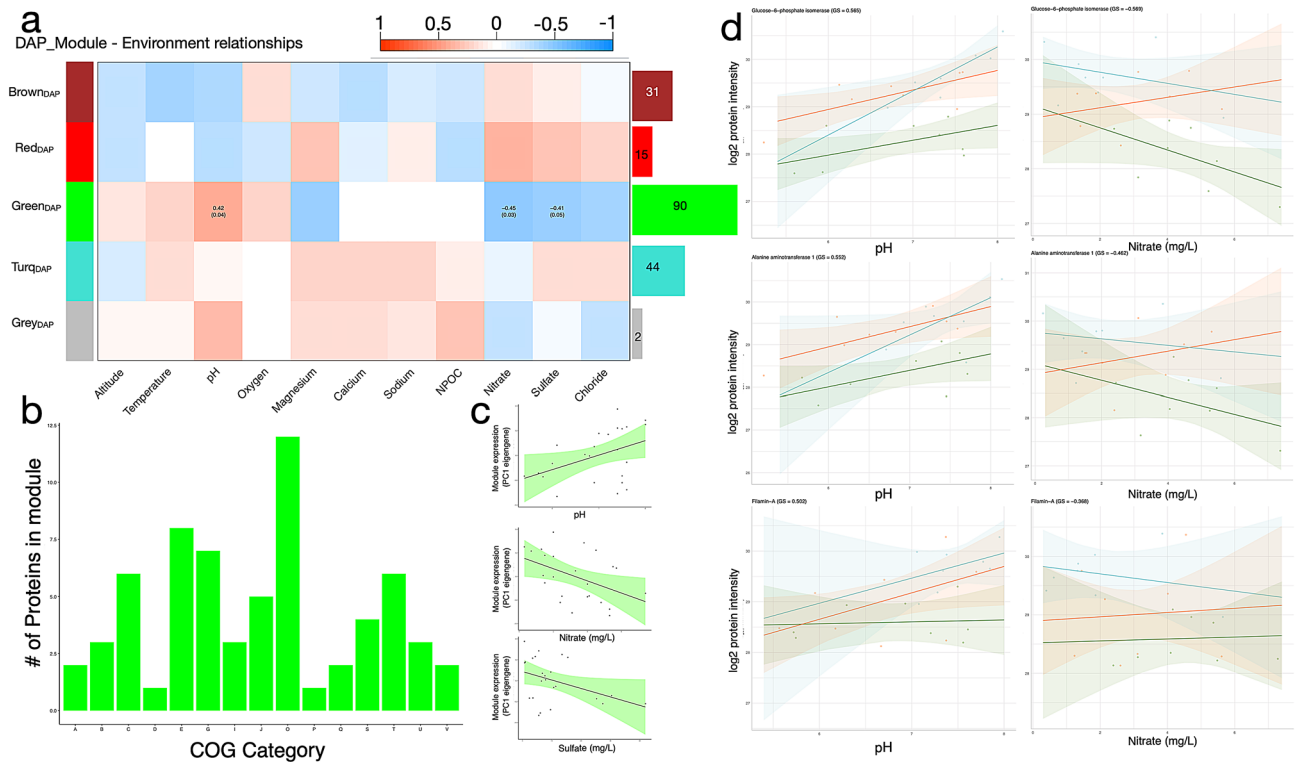


Figure 2. Association of DAPs between sampling regions ($n = 182$) and abiotic gradients. **(a)** Correlations between module eigengenes (rows) and abiotic variables (columns). The bar graph and numbers on the right indicate number of proteins belonging to each module. The strength of the correlations between abiotic factors and DAP co-abundance modules are indicated by color intensity. The numbers in the cells give Pearson's correlation coefficients between the module "eigengene" and the abiotic factor and the p-value of the correlation test (not listed for cells with $p > 0.05$). **(b)** COG distribution of "Green_{DAP}" module. One-letter abbreviations for the functional categories: A, RNA processing and modification; B, chromatin structure and dynamics; C, energy production and conversion; Y, nuclear structure; E, amino acid transport and metabolism; F, nucleotide transport and metabolism; G, carbohydrate transport and metabolism; H, coenzyme transport and metabolism; I, lipid transport and metabolism; J, translation, ribosomal structure and biogenesis; K, transcription; L, translation, ribosomal structure and biogenesis; M, cell wall/membrane/envelope biogenesis; O, post-translational modification, protein turnover, and chaperones; P, inorganic ion transport and metabolism; W, extracellular structures; Q, secondary metabolites biosynthesis, transport, and catabolism; S, unknown function; T, signal transduction mechanisms; U, intracellular trafficking, secretion, and vesicular transport; V, defense mechanisms; Z, cytoskeleton. **(c)** "GreenDAP" eigengene expression values correlated with pH, nitrate and sulfate. **(d)** Protein reaction norm examples for various DAP member proteins in relation to pH and nitrate concentrations of springs.

(GO:0010467; e.g. chromodomain-helicase-DNA-binding protein Mi-2 homolog and nuclear pore complex protein Nup93). To investigate if abundances of DAPs between regions are influenced by abiotic gradients, we related all 182 DAPs to the abiotic variables (Fig. 2a). The "green_{DAP}" module emerged as the only set of co-regulated and differentially abundant proteins that significantly correlated with pH ($r_{\text{greenDAP}} = 0.42$), nitrate ($r_{\text{greenDAP}} = -0.45$) and sulfate ($r_{\text{greenDAP}} = -0.41$) as shown in Fig. 2c. Protein families showing the highest frequencies across the "green_{DAP}" subset of DAPs were O: Post-translational modification, protein turnover, and chaperones, E: amino acid transport and metabolism and G: carbohydrate transport and metabolism (Fig. 2b). This subset of DAPs is functionally similar to the protein family distribution of all identified DAPs (Supplementary Fig. S20). Significantly changing proteins with the highest gene significance (GS) values in this module also show different, sometimes opposite, reaction norms between regions (Fig. 2d).

Protein reaction norms. We observed baseline abundance differences between geographical regions (Figs. 2d and 3). For example, the reaction norm of phosphoglycerate mutase (member of the "yellow" module), is a pH-sensitive enzyme^{77,78} and had a higher baseline abundance in populations from BF springs which had significantly higher mean pH than springs from R and H (Supplementary Fig. S10a). Multiple proteins showed pronounced abundance changes in relation to abiotic conditions (Figs. 2d, 4b–d). For example, 13 proteins showed increasing and 20 decreasing reaction norms according to spring elevation and pH variation was accompanied by 71 increasing and 5 decreasing norms of reaction. Temperature variation was accompanied by 29 increasing and 33 decreasing reaction norms whereby a high frequency of decreasing proteins belonged to the chaperonin Cpn60/TCP-1 family and a high frequency of increasing proteins to the calreticulin/calnexin family.

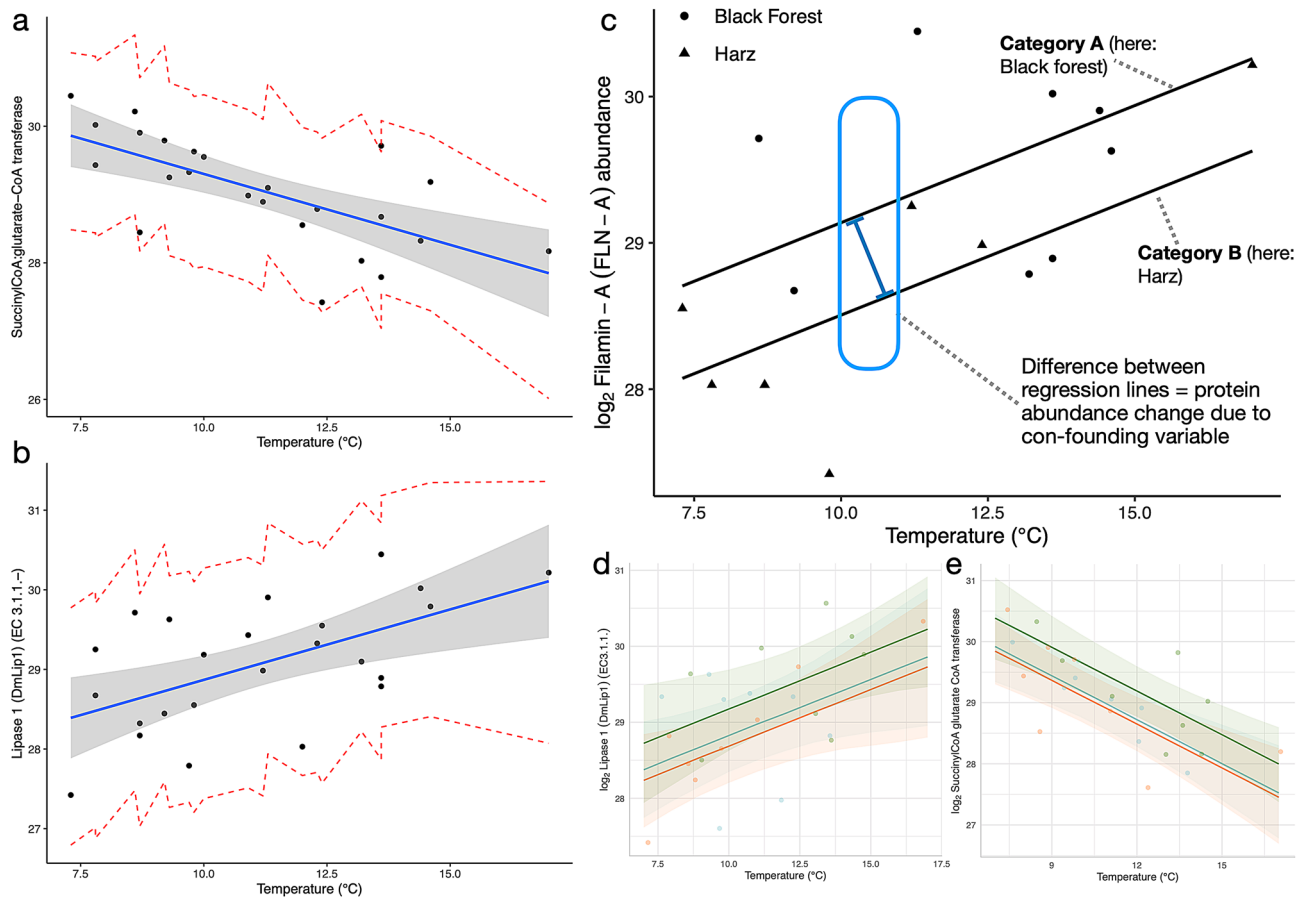


Figure 3. Illustration and concept of protein reaction norms and baseline differences. Protein reaction norms are the relative abundance changes of a single protein across a range of environments. (a) Regression model and predicted values of Succinyl-CoA-glutarate-CoA transferase in relation to spring temperature (°C). (b) Regression model and predicted values of Lipase 1 in relation to spring temperature (°C). (c) Reaction norms of filamin A in response to variation in temperature (°C). Category A: e.g. a site experiencing environmental change such as nutrient influx, temperature change or pollution; Category B: A site experiencing no such environmental change. (d) Regression as in (b) but categorized by sampling region (color-code as in Fig. 1a). (e) Regression as in (a) but categorized by sampling region.

Both families play a role in the physiological integration of temperature variation, both in cold^{79,80} and warm⁸¹ environments. Interestingly, this feature (i.e. protein's abundances change as a factor of changing abiotic conditions) can also be observed for DAPs (Fig. 2).

Global network analyses. The first WGCNA assigned all 1173 proteins to eight co-expression modules (designated randomly by colors). Five module eigengenes were significantly correlated with distinct abiotic variables (Fig. 4a): both, “brown” ($n=477$) and “red” ($n=59$) modules correlated with magnesium ($r_{\text{brown}}=-0.46$, $r_{\text{red}}=-0.42$) and sulfate ($r_{\text{brown}}=-0.43$, $r_{\text{red}}=-0.49$). “Green” ($n=81$), “yellow” ($n=133$) and “grey” ($n=73$) modules correlated positively with oxygen ($r_{\text{green}}=0.4$), pH ($r_{\text{yellow}}=0.47$) and nitrate ($r_{\text{grey}}=0.47$), respectively. Correlations between module membership and protein significance values were used to further corroborate these relationships, with correlations ranging from 0.5 for the “yellow” module to 0.66 for the “grey” module (Fig. 4f–i). Gene Ontology enrichment of the highly underrepresented “brown” module detected many significant GO terms across all GO categories (Cellular Compartment (CC), Molecular Function (MF) and BF). Gene Ontology terms within BP and MF included many terms associated with unfolded protein stimulus [e.g., *chaperone binding* (GO:0051082), *chaperone-mediated protein folding* (GO:0061077) and *cellular response to topologically incorrect protein* (GO:0035967)]. The “red” module was associated with the same abiotic variables as the “brown” module, but had no enriched GO terms. The “green” module was enriched with four GO terms in the BP category: *Intestinal stem cell homeostasis* (GO:0036335), *sleep* (GO:0030431), *response to endoplasmic reticulum stress* (GO:0034976) and *RNA interference* (GO:0016246) and one GO term in the CC category: *Intracellular membrane-bounded organelle* (GO:0043231). Lastly, the “yellow” module was enriched with only one GO term in the BP category: *Electron transport chain* (GO:0022900) but with three GO terms in the CC category: *Cytoplasmic part* (GO:0044444), *intracellular membrane bounded organelle* (GO:0043231) and *endomembrane system*

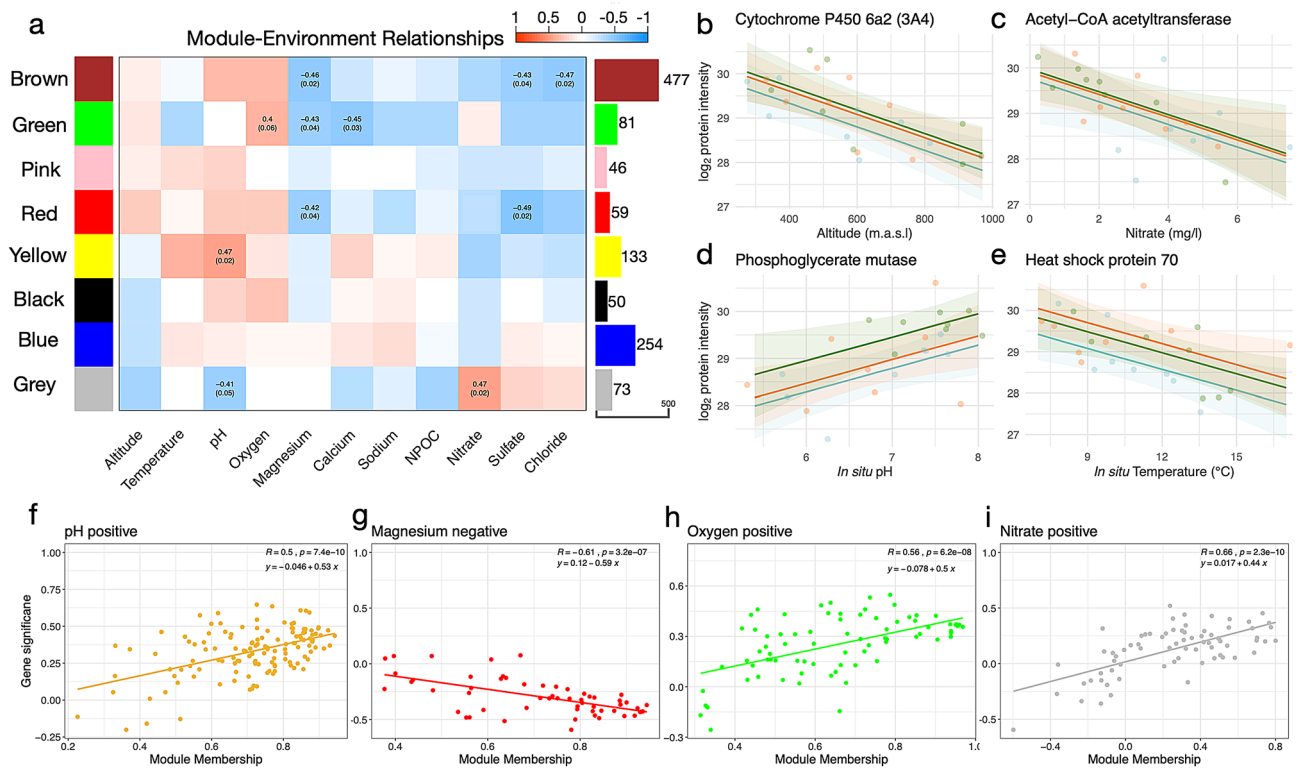


Figure 4. Network analysis of *Crumecia irrorata* protein abundances in relation to abiotic variables. **(a)** Correlations between module eigengenes (rows) and abiotic variables (columns). The bar graph and numbers on the right indicate number of proteins belonging to each module. The strength of the correlations between abiotic factors and protein co-abundance modules are indicated by color intensity. The numbers in the cells give Pearson's correlation coefficients between the module "eigengene" and the abiotic factor and the p -value of the correlation test [not listed for cells with $p > 0.05$ except "green"-oxygen ($p = 0.06$)]. **(b–e)** Scatterplots showing log₂ protein reaction norms for four proteins with high gene significance (GS) values in relation to the abiotic variable shown on the x-axis. Reaction norms are colored by sampling region to illustrate baseline differences between sampling regions according to color-scheme in Fig. 1. **(b)** Cytochrome P450 6a2 decreasing in abundance with elevation of springs. **(c)** Acetyl-CoA acetyltransferase abundance decreasing with increasing nitrate concentration of spring water. **(d)** Phosphoglycerate mutase increasing in abundance with increasing in situ pH. **(e)** Heat shock protein 70 decreasing in abundance with water temperature. **(f–i)** Scatterplots illustrating the relationship between a protein's module membership score (x-axis) and the protein's significance for the abiotic variable (y-axis). Higher correlations between these parameters indicate stronger associations of the **(f)** "yellow", **(g)** "red", **(h)** "green" and **(i)** "grey" modules with their associated abiotic variables (pH, magnesium, oxygen and nitrate, respectively).

(GO:0012505) and two GO terms in the MF category: *Hydrolase activity* (GO:0016787) and *catalytic activity acting on a protein* (GO:0140096).

Domain-centric GO enrichment of the "yellow" module (Pfam_{slim} level = 3) computed significantly enriched GO terms such as *regulation of sodium ion transport* (GO:0002028), *pH reduction* (GO:0045851) and *retina layer formation* (GO:0010842). The "green" module included significant GO terms such as *oxidoreductase activity* (GO:0016491), *oxidation-reduction process* (GO:0055114) and *response to oxidative stress* (GO:0,006,979). The "grey" module, related negatively to pH and positively to nitrate concentrations was enriched for similar GO terms as the "yellow" module such as *pH reduction* (GO:0045851), *regulation of intracellular pH* (GO:0051453) but also terms related to nitrogen such as *response to xenobiotic stimulus* (GO:0009410) and *cellular nitrogen compound metabolic process* (GO:0034641). The "yellow" module had higher frequencies of proteins associated with C: *energy production and conversion*, Z: *cytoskeleton* and O: *post-translational modification/protein turnover, and chaperones*. The "green" module included 39.5% proteins associated with O: *post-translational modification/protein turnover, and chaperones* and the "grey" module showed COG underrepresentation and many proteins with S: *unknown function* (Fig. 2e,f).

The WGCNA_{BC} assigned all proteins to six co-expression modules (designated randomly by colors; Fig. 5a). An overview of COG frequency distributions and eigengene correlations with BioClim variables is given in Fig. S17. The strongest relationship in the WGCNA_{BC} emerged between the "green_{BC}" module ($n = 214$) and bio6 ($r = 0.72$), representing the minimum temperatures (°C) of the coldest month experienced by *C. irrorata* populations in the corresponding sites over 30 years (Fig. 5c,d). This module contained a high frequency of proteins related to the cytoskeleton COG family (Fig. 5b) and was enriched for GO terms related to the regulation of cytoskeleton and tissue integrity, including terms such as *regulation of microtubule-based process* (GO:0032886),

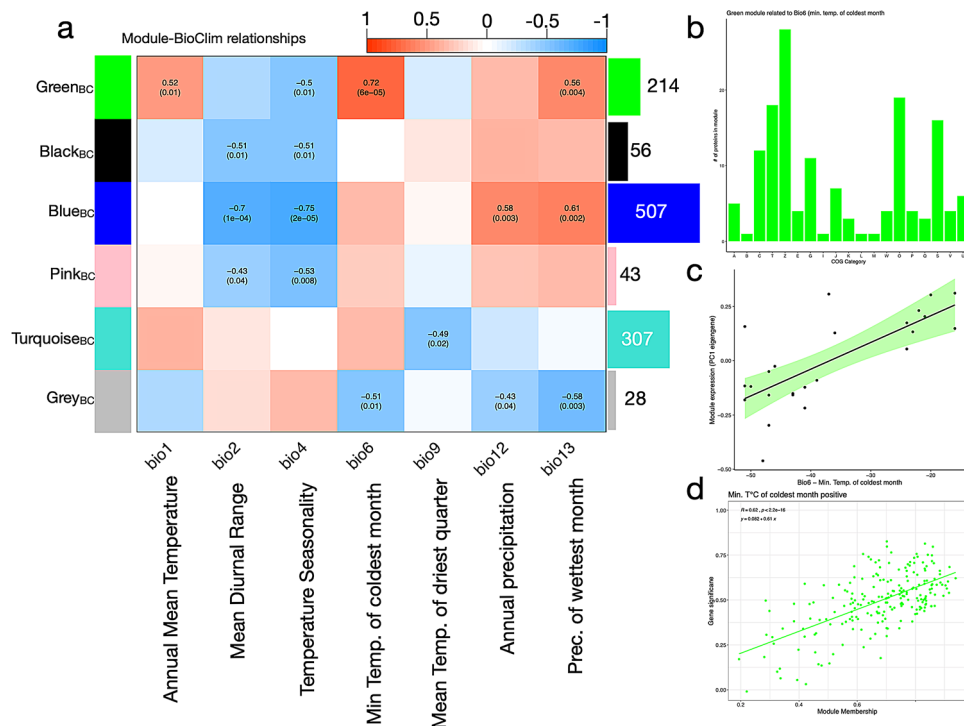


Figure 5. Network analysis of *Crunoecia irrorata* protein abundances in relation to bioclimatic variables (WGCNA_{BC}). **(a)** Correlations between module eigengenes (rows) and bioclimatic variables (columns). Colors and numbers convey information identical to Fig. 4a. **(b)** COG family distribution of proteins belonging to the “green_{BC}” module. One-letter abbreviations are identical to Fig. 2. **(c)** “Green_{BC}” module eigengene expression positively associated with BioClim Bio6. **(d)** Scatterplot illustrating the relationship between protein’s “green_{BC}” membership score (x-axis) and the protein’s significance for Bio6 (min. temperature of coldest month) (y-axis).

actin filament bundle assembly (GO:0051017), *actin-mediated cell contraction* (GO:0070252), *homotypic cell–cell adhesion* (GO:0034109), *positive regulation of wound healing* (GO:0090303), *actin filament polymerization* (GO:0030041) and *polarized epithelial cell differentiation* (GO:0030859). In the Cellular Compartment (CC) category, proteins were associated with the *outer and inner dense plaque of desmosome* (GO:0090636), *terminal web* (GO:1990357) and the *myosin II complex* (GO:0016460). The module with the second highest frequency of proteins related to the cytoskeleton COG family was the “pink_{BC}” module (n = 43; Fig. S22). Its eigengene expression correlated negatively (p < 0.05) with bio2 (mean diurnal range; r = -0.43) and bio4 (temperature seasonality; r = -0.53), indicating that temperature fluctuations (monthly and seasonal) directly influence species’ molecular responses, additionally to determining their distributional range⁸². Interestingly, the “black_{BC}” module consisted to 43% of proteins associated with the COG family O: post-translational modification, protein turnover, and chaperones (Fig. S22). Like the “pink_{BC}” module, it also correlated negatively (p < 0.05) with bio2 (r = -0.51) and bio4 (r = -0.51), hinting at the importance of PTMs and molecular chaperones in adaptive mechanisms to temperature fluctuations.

Discussion

Relating the most variable proteins that drive proteome differentiation among populations to their associated protein families and functions can provide broad insight into which physiological systems respond dynamically to abiotic variation. As a result, any environmental change that is to be deduced from molecular profiles that are associated with any of these physiological systems are potentially confounded. In this study, the differentiation among population proteomes was associated with environmental heterogeneity between springs, geographic distance between populations and past climatic conditions. These findings indicate that the presence or absence of spectra associated with certain peptides/proteins may be an important indicator of environmental differences and that relative protein abundance differences are influenced by the geographic distance between sites. Even though no relationship between protein abundances and environmental distance emerged, protein reaction norms and association of protein modules with abiotic variation show that protein abundances of *C. irrorata* populations are regulated by the species’ integration of variation in abiotic conditions commonly experienced in its habitat. The absence of such a relationship (i.e. the majority of identified proteins did not exhibit a relation or a norm of reaction) was to be expected since cells maintain their own internal environment, e.g. via ion-transport⁸³. Proteins associated with energy acquisition and storage (e.g. hexamerin-1.1⁸⁴ and solute carrier family 25 member 44), cytoskeletal functioning (e.g. calcineurin subunit B type 2⁸⁵ and filamin-A⁸⁶) and post-translational modification (e.g. FK506-binding protein 2⁸⁷) separated sampling regions along axis 2 of the nMDS (Supplementary

Fig. S21), indicating, for example, that nutrient availability differs between sites and affects populational protein abundances. Further, protein families showing high frequencies among all DAPs likely contribute substantially to proteome-divergence. High frequencies of DAPs belong to the O: *Post-translational modification, protein turnover, and chaperones*, T: *signal transduction mechanisms* and C: *energy production and conversion* COG families. These findings point to differences in cellular environmental sensing⁸⁸, and post-translational modification (PTM), regulatory modifications that may change or suppress protein functions⁸⁹, between sampling regions. Our findings therefore substantiate the importance of PTMs in integrating environmental cues and regulating physiological processes, suggesting that PTM profile differences between natural populations are substantial and are agents of plasticity. As a detailed discussion of all significant module-abiotic relationships and annotation profiles is beyond the scope of this research article, we focused the discussion on certain protein reaction norms and protein networks that corresponded with differences in abiotic conditions of springs. We first establish functional connections between modules and abiotic variables as this information is of great relevance to studies investigating molecular abundance and activity differentiation between populations^{90,91}.

Functional relationships between modules and abiotic variation. We highlight functional relationships between modules and abiotic variables with the examples of the “yellow” module, related to pH, and the “green” module, related to oxygen concentration. Proteins in the “yellow” module show functional profiles related to cell-internal and external pH conditions: Member proteins such as succinate dehydrogenase and peptidyl-prolyl cis–trans isomerase are activated by changes in pH and modulate intracellular pH homeostasis, respectively^{92,93}. Enrichment for *electron transport chain* (GO:0022900) of member proteins such as electron transfer flavoprotein [subunits alpha/beta (C- and N-terminal)] indicate a relationship between aqueous pH and the electron transport system (ETS). Fifteen proteins in this module are found in this oligomeric enzyme complex within the inner mitochondrial membrane and of these, seven show increasing reaction norms with increasing pH (Supplementary Information 1 Figs. S11 and S12). The ETS of freshwater invertebrates is dependent on external pH, whereby its activity is reduced at lower pH⁹⁴. The relationship between this module and pH variation is further corroborated when analyzing its Pfam and COG family distributions (Supplementary Fig. S12b, Fig. 2d). Higher frequencies of proteins with a zinc finger (LIM type) domain, filamin/ABP280 repeat structure and a gelsolin-like domain such as galectin-8 and a high frequency of proteins related to the Z: *cytoskeleton* family ($n = 22$) likely reflect the pH-dependent regulation of the structure and contractility of the actin-based cytoskeleton, as proteins with these domains and within this family are involved in the control of cytoskeletal function^{95–98}. These actin-binding proteins and their ability to cross-link actin filaments is pH-dependent^{99–101} and their regulation of the cytoskeleton represents an integral response of organisms to various environmental contexts including changes in pH^{102,103}. Proteins in the “green” module, associated with variation in oxygen concentration, were characterized by high frequencies of heat shock protein 70 family and insect cuticle domains (Supplementary Fig. S13a) and were predominantly associated with the COG family O: *Post-translational modification, protein turnover, and chaperones* (Fig. 2e). Translated, unfolded proteins require oxygen to form disulfide bonds¹⁰⁴ which may explain why changes in oxygen levels correlated with changes in abundance of proteins associated with unfolded protein metabolic activity. The relationship between the insect cuticle and oxygen concentration is less understood but our data points to key proteins involved in regulating cuticle characteristics such as permeability in response to varying oxygenation of freshwater, potentially to release reactive oxygen species (ROS)¹⁰⁵ or to facilitate penetration of dissolved oxygen¹⁰⁶. Multiple enriched GO terms such as *oxidoreductase activity* (GO:0016491), *oxidoreductase activity, acting on CH₂–OH group of donors* (GO:0016614) and *oxidation–reduction process* (GO:0055114) and GO terms related to externally-induced oxidative stress such as *response to oxidative stress* (GO:0006979) and *response to abiotic stimulus* (GO:0009628) were enriched for this module, providing further evidence for this relationship. The relation between GO enrichment of *intestinal stem cell homeostasis* (GO:0036335) and *regulation of stem cell differentiation* (GO:2000736) in this module and varying oxygen levels of springs suggests that *C. irrorata* larvae actively regulate oxygen levels in the low oxygen/ROS niches in which insect stem cells reside, potentially to avoid the effects of ROS on stem cells such as DNA damage and senescence^{107,108}.

Overall, the outlined functional relationships indicate the coordinated regulation of protein abundances in response to abiotic variation, validating the present approach as a way to connect protein dynamics with changes in abiotic conditions. Given this functional relationship between pH variation and the “yellow” module, testable hypotheses about adaptation and phenotypically plastic responses to pH variation can be formulated. For example, the enriched GO term *retina layer formation* (GO:0010842) in the “yellow” module may reflect a previous observation that retinal gene expression changes in response to a drop in local pH¹⁰⁹. The module contained e.g. paxillin (PF00412) and obscurin (Unc-89), a signaling (COG: T) protein containing a Immunoglobulin I-set domain (PF07679), whose deletion affects eye development in zebrafish¹¹⁰. This finding hints at the pH-induced regulation of stemmata physiology in the Trichoptera larval ocelli, a potential adaptation to live underwater.

Variation of biomarker proteins. A consequence of the relation between adaptive physiology and environment is that baselines and activity levels of protein biomarkers may shift according to abiotic variation of the studied system. Eco-toxicological studies showed that enzymatic variability across time and/or space is often not fully explained by stress factors such as pollution^{111,112}. Additionally, abundance measures of protein biomarkers are regularly applied in eco-toxicological research including freshwater invertebrates¹¹³. Members include e.g. the cytochrome P450 mixed function oxidases and enzymes such as catalase and filamin-A^{114–117}. Catalase, filamin-A, ATP-citrate(pro-5)-lyase and the two cytochromes Cyp62a and Cyp4c1 exhibited significant reaction norms in relation to geographic elevation (Supplementary Fig. S14). Both, *Cyp6a2* and *Cyp4c1* are responsible for the metabolism of numerous xenobiotics and endogenous compounds, including organophosphorus insecticides.

ticides such as dichlorodiphenyltrichloroethane (DDT)^{118–120}. Reaction norms of these two cytochromes over an elevation range hint at a potential presence of organophosphorus insecticides in groundwater-aquifers in Central Europe. Moreover, it might indicate that the contemporary application of organophosphorus insecticides in agricultural practice results in groundwater contamination as anthropogenic input of xenobiotics should increase with decreased elevation and corresponding increased settlement density. The described biomarker norms of reaction exemplify that sources and nature of variability in organism-environment relationships need to be understood and made explicit; i.e. used as information rather than dismissed as noise in eco-toxicological analyses of natural populations.

Adjusting for different baseline abundances. In comparative field-studies that investigate the systems-wide regulation of proteins in response to environmental change (e.g. comparing populations experiencing drought vs. no drought or a shift in mean temperature vs. no shift), the influence of that change on organismal physiology is commonly identified through the log₂ fold-change (FC) values of pairwise comparisons between abundances of identified proteins^{121,122}. Our study indicates that these compared abundances are directly influenced by abiotic variation and, therefore, by environmental differences between (sampling) sites (e.g. Fig. 2). Importantly, our data show that these DAPs may precisely be differentially abundant because of differences in abiotic conditions between individual sites. Results of pairwise comparisons may therefore over-, or even underestimate the differences between compared groups if only one studied factor between regions is considered in the sampling design (e.g. polluted vs. non-polluted). Including information on baseline abundance shifts due to confounding variables could, for example, be achieved by reducing the set of DAPs by those that also show abundance increases or decreases with abiotic conditions. One could account for the influence of confounding abiotic variables by subtracting the coefficients of the regression model from the original abundances used as input for e.g. LIMMA differential analysis, followed by a new differential analysis on adjusted abundances. This value should then be representative of the protein abundances of populations independent from the influence of pH differences between regions. Since we did not investigate regional differences such as pollution specifically, and, therefore, do not expect abundance differences to be associated with other than abiotic differences, future studies applying such an adjustment are warranted.

Temperature tolerance. Despite being a key environmental variable in determining the physiology of ectothermic organisms, in situ water temperature was not a key abiotic factor regulating a protein-network. In contrast, proteins regulating the cytoskeleton appear to be related to past extreme climatic conditions. “Green_{BC}” eigengene expression increased with higher minimum extreme temperatures and annual mean temperature (Fig. 5). Cytoskeletal integrity is perturbed by a variety of cell stress responses and plays a crucial role in maintenance of cell homeostasis^{123,124}. This relationship echoes previous findings that show cytoskeletal regulation in response to temperature change via e.g. associated cryoprotective dehydration^{102–106}. Population-wide abundance changes of cytoskeletal proteins in response to previously experienced thermal extremes indicates heritable protein abundance variation between populations attributable to heritable gene expression variation^{125,126}, supporting the notion that extreme environmental temperatures exert a strong selective pressure on ectothermic species¹²⁷. In support of this finding, the “green_{BC}” and “pink_{BC}” modules, both with high numbers of proteins associated with cytoskeletal functioning (Supplementary Fig. S17), show decreasing eigengene expression with increasing temperature seasonality, i.e. the standard deviation of the mean monthly temperature. These results indicate that cytoskeleton regulation is integral to organismal functioning when organisms experience less-variable temperature regimes (i.e. longer stretches of warmer or colder temperatures).

Integral parts of the cellular stress response include the evolutionary conserved heat-shock- and cold-shock-response (HSR and CSR), characterized by the expression of heat-shock proteins (Hsps). Putatively cold-adapted species have been shown to constitutively express Hsps to facilitate protein folding at low temperatures^{128,129}, a phenomenon also observed in *C. irrorata*⁴⁶. Here, we observed constitutive and increasing but also *decreasing* Hsp reaction norms with increasing in situ temperatures (Fig. 4e; Supplementary Fig. S15). Chaperones with decreasing norms of reaction such as Hsp83 and Hsp90 might be integral for alleviating cold-induced protein denaturing and increased protein folding time. These patterns indicate evolutionarily sub-divided roles within the Hsp gene family in putatively cold-adapted species, with certain Hsps functioning during HSRs and others during CSRs. In light of the present data, chaperonins belonging to the Chaperonin Cpn60/TCP-1 protein family such as Hsp68 could be indicative of cold-adaptation of *C. irrorata*, especially since chaperonins govern cell growth at cold temperatures and springs show constant cold temperatures^{79,130,131}. Species from such stable and cold environments might be ill-adapted for climate-induced warming of aquatic ecosystems^{132,133}. Somewhat counterintuitively, recent cellular evidence indicates that, for certain aquatic species, the stable cold environment might be a conditioning factor such that they evolved to be able to protect themselves from temperature-induced cellular damage^{129,134,135}.

Limitations and implications for the study of natural populations. This work assumes that if two populations have a different quantity of protein it means that the proportion of individuals which express that protein differs between populations. Inter-individual differences are not only pronounced but directly influence the population-level outcome of environmental change^{136,137}. Consequently, even in the absence of a significant population-level response to the environment, some individuals may still respond plastically to changing conditions. Due to this strong inter-individual variation, it may be worthwhile to apply omics-wide assessments on pooled samples directly, if the goal is to obtain a “screening” of the physiological status of populations. Overall, this approach can help keeping sample sizes smaller than inter-individual studies, and, as shown in this study, does not apparently mask abundance patterns in response to environmental variation including established

biomarker proteins. The systematic characterization of such relationships in a non-targeted approach, using a non-model organism, has the potential to give unprecedented insights into the molecular functioning of a variety of ecologically relevant species. We further recognize that there are many other sources of confounding factors that complicate the interpretation of the observed molecular responses. Our study design did not account for biotic influences on molecular profiles such as predator presence¹³⁸, parasitism status¹³⁹ and microbial differences between sites and individuals or populations^{140,141}. Further, since we used whole-animal lysates, information on GO term enrichment at certain life stages or tissues were likely missed. Lastly, information on other abiotic and biotic factors such as prevailing thermal conditions or co-abundant species are additionally needed to further refine baseline levels of protein abundances and to understand how natural environmental conditions might mask and alter the molecular responses of natural populations.

Conclusion

The primary findings of this study are that the variation in proteome profiles of pooled population samples are attributable to abiotic but also geographic and past climatic differences between sites. Variables such as pH, oxygen concentration, past thermal minima and seasonality “explained” a part of the variation and need to be considered when comparing sites experiencing differential environmental change. Additionally, protein families related to energy allocation, cellular information processing and PTM regulation contribute to proteome divergence overall and likely play a vital role in physiological adjustment to varying abiotic conditions. More fundamentally, protein norms of reaction are ubiquitous and often reflect the functional roles these proteins have within multicellular organisms. Additionally, signatures of local adaptation (e.g. warming temperatures or differences in oxygenation) can be gleaned through such comparative analyses, whereby abundance increases of cytoskeleton-related proteins in populations experiencing warmer temperatures at a historical timescale are just one example. This considerable variability in protein abundance patterns likely distorts abundance differences in comparative field studies if abiotic differences are not accounted for, potentially leading to over- or under-estimation of environmental change experienced by organisms. Integrating this natural variation appears to be an important step in systems-wide biomarker development and for disentangling the differences between populations experiencing differential environmental change.

Data availability

We provided all raw, transformed and filtered data required to interpret, replicate, and build upon our findings in the Supplementary Information files 1 and 2 (combining quantitative and annotation information). An R script, detailing all commands used in data analysis can be found in the “Supplementary Information”. Mass spectrometry raw data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository¹⁴² with the data set identifier PXD017959.

Received: 14 April 2020; Accepted: 2 September 2020

Published online: 23 September 2020

References

- West-Eberhard, M. J. *Developmental Plasticity and Evolution* (Oxford University Press, Oxford, 2003).
- Beldade, P., Mateus, A. R. A. & Keller, R. A. Evolution and molecular mechanisms of adaptive developmental plasticity. *Mol. Ecol.* **20**, 1347–1363 (2011).
- Dall, S. R. X., McNamara, J. M. & Leimar, O. Genes as cues: Phenotypic integration of genetic and epigenetic information from a Darwinian perspective. *Trends Ecol. Evol.* **30**, 327–333 (2015).
- Scheiner, S. M. Genetics and evolution of phenotypic plasticity. *Annu. Rev. Ecol. Syst.* **24**, 35–68 (1993).
- Ovaskainen, O. & Meerson, B. Stochastic models of population extinction. *Trends Ecol. Evol.* **25**, 643–652 (2010).
- Lawson, C. R., Vindenes, Y., Bailey, L. & van de Pol, M. Environmental variation and population responses to global change. *Ecol. Lett.* **18**, 724–736 (2015).
- Mayr, E. The growth of biological thought: Diversity, evolution, and inheritance. *Am. Biol. Teach.* **46**, 462–463 (1984).
- Seebacher, F., White, C. R. & Franklin, C. E. Physiological plasticity increases resilience of ectothermic animals to climate change. *Nat. Clim. Change* **5**, 61–66 (2015).
- Somero, G. N. The physiology of climate change: how potentials for acclimatization and genetic adaptation will determine ‘winners’ and ‘losers’. *J. Exp. Biol.* **213**, 912–920 (2010).
- Riddell, E. A., Odom, J. P., Damm, J. D. & Sears, M. W. Plasticity reveals hidden resistance to extinction under climate change in the global hotspot of salamander diversity. *Sci. Adv.* **4**, eaar5471 (2018).
- Noble, D. Claude Bernard, the first systems biologist, and the future of physiology. *Exp. Physiol.* **93**, 16–26 (2008).
- Ideker, T., Galitski, T. & Hood, L. A new approach to decoding life: Systems biology. *Annu. Rev. Genomics Hum. Genet.* **2**, 343–372 (2001).
- Whitehead, A. & Crawford, D. L. Variation within and among species in gene expression: Raw material for evolution. *Mol. Ecol.* **15**, 1197–1211 (2006).
- Wittkopp, P. J., Haerum, B. K. & Clark, A. G. Regulatory changes underlying expression differences within and between *Drosophila* species. *Nat. Genet.* **40**, 346–350 (2008).
- Gonzalez, E. G. *et al.* Population proteomics of the European hake (*Merluccius merluccius*). *J. Proteome Res.* **9**, 6392–6404 (2010).
- Papakostas, S. *et al.* A proteomics approach reveals divergent molecular responses to salinity in populations of European whitefish (*Coregonus lavaretus*). *Mol. Ecol.* **21**, 3516–3530 (2012).
- Chevalier, F. *et al.* Proteomic investigation of natural variation between *Arabidopsis* ecotypes. *Proteomics* **4**, 1372–1381 (2004).
- Mueller, R. S. *et al.* Ecological distribution and population physiology defined by proteomics in a natural microbial community. *Mol. Syst. Biol.* **6**, 374 (2010).
- Snijder, B. & Pelkmans, L. Origins of regulated cell-to-cell variability. *Nat. Rev. Mol. Cell Biol.* **12**, 119–125 (2011).
- Bradshaw, A. D. Evolutionary significance of phenotypic plasticity in plants. In *Advances in Genetics* Vol. 13 (eds Caspari, E. W. & Thoday, J. M.) 115–155 (Academic Press, New York, 1965).
- Colman-Lerner, A. *et al.* Regulated cell-to-cell variation in a cell-fate decision system. *Nature* **437**, 699–706 (2005).

22. Fisher, M. A. & Oleksiak, M. F. Convergence and divergence in gene expression among natural populations exposed to pollution. *BMC Genomics* **8**, 108 (2007).
23. Pujolar, J. M. *et al.* Surviving in a toxic world: transcriptomics and gene expression profiling in response to environmental pollution in the critically endangered European eel. *BMC Genomics* **13**, 507 (2012).
24. Davies, S. W., Marchetti, A., Ries, J. B. & Castillo, K. D. Thermal and pCO₂ stress elicit divergent transcriptomic responses in a resilient coral. *Front. Mar. Sci.* **3**, 112 (2016).
25. Palumbi, S. R., Barshis, D. J., Traylor-Knowles, N. & Bay, R. A. Mechanisms of reef coral resistance to future climate change. *Science* **344**, 895–898 (2014).
26. Gygi, S. P., Rochon, Y., Franza, B. R. & Aebersold, R. Correlation between protein and mRNA abundance in yeast. *Mol. Cell Biol.* **19**, 1720–1730 (1999).
27. Liu, Y., Beyer, A. & Aebersold, R. On the dependency of cellular protein levels on mRNA abundance. *Cell* **165**, 535–550 (2016).
28. Bludau, I. & Aebersold, R. Proteomic and interactomic insights into the molecular basis of cell functional diversity. *Nat. Rev. Mol. Cell Biol.* <https://doi.org/10.1038/s41580-020-0231-2> (2020).
29. Lodish, H. *et al.* *Molecular Cell Biology* (W. H. Freeman, New York, 2000).
30. Watson, J. D. *Molecular Biology of the Gene* (Pearson Education, London, 2004).
31. Giardi, M. T., Masojidek, J. & Godde, D. Effects of abiotic stresses on the turnover of the D1 reaction centre II protein. *Physiol. Plant.* **101**, 635–642 (1997).
32. Raj, A. & van Oudenaarden, A. Nature, nurture, or chance: Stochastic gene expression and its consequences. *Cell* **135**, 216–226 (2008).
33. Kaern, M., Elston, T. C., Blake, W. J. & Collins, J. J. Stochasticity in gene expression: From theories to phenotypes. *Nat. Rev. Genet.* **6**, 451–464 (2005).
34. Nikinmaa, M., Tiihonen, K. & Paajaste, M. Adrenergic control of red cell pH in salmonid fish: Roles of the sodium/proton exchange, Jacobs-Stewart cycle and membrane potential. *J. Exp. Biol.* **154**, 257–271 (1990).
35. Pavlov, M. Y. & Ehrenberg, M. Optimal control of gene expression for fast proteome adaptation to environmental change. *Proc. Natl. Acad. Sci. USA* **110**, 20527–20532 (2013).
36. Adams, S. M., Giesy, J. P., Tremblay, L. A. & Eason, C. T. The use of biomarkers in ecological risk assessment: recommendations from the Christchurch conference on Biomarkers in Ecotoxicology. *Biomarkers* **6**, 1–6 (2001).
37. Diz, A. P., Truebano, M. & Skibinski, D. O. F. The consequences of sample pooling in proteomics: An empirical study. *Electrophoresis* **30**, 2967–2975 (2009).
38. Karp, N. A. & Lilley, K. S. Investigating sample pooling strategies for DIGE experiments to address biological variability. *Proteomics* **9**, 388–397 (2009).
39. Bennike, T. B. *et al.* Comparing the proteome of snap frozen, RNAlater preserved, and formalin-fixed paraffin-embedded human tissue samples. *EuPA Open Proteomics* **10**, 9–18 (2016).
40. Johnsen, I. K. *et al.* Evaluation of a standardized protocol for processing adrenal tumor samples: Preparation for a European adrenal tumor bank. *Horm. Metab. Res.* **42**, 93–101 (2010).
41. Kruse, C. P. S., Basu, P., Luesse, D. R. & Wyatt, S. E. Transcriptome and proteome responses in RNAlater preserved tissue of *Arabidopsis thaliana*. *PLoS ONE* **12**, e0175943–e0175943 (2017).
42. Fick, S. E. & Hijmans, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **37**, 4302–4315 (2017).
43. Hijmans, R. J. *raster: Geographic Data Analysis and Modeling. R package version 3.3-13.* <https://CRAN.R-project.org/package=raster>. (2020).
44. Cox, J. *et al.* Andromeda: A peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10**, 1794–1805 (2011).
45. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
46. Ebner, J. N., Ritz, D. & von Fumetti, S. Comparative proteomics of stenotopic caddisfly *Crunoecia irrorata* identifies acclimation strategies to warming. *Mol. Ecol.* **28**, 4453–4469 (2019).
47. Misof, B. *et al.* Phylogenomics resolves the timing and pattern of insect evolution. *Science* **346**, 763–767 (2014).
48. Jimenez-Morales, D., Campos, A.R. & Von Dollen, J. *artMS: Analytical R tools for Mass Spectrometry. R package version 1.5.3.* <https://github.com/bioadavidjm/artMS>. (2020).
49. Chen, H. *VennDiagram: Generate High-Resolution Venn and Euler Plots. R package version 1.6.20.* <https://CRAN.R-project.org/package=VennDiagram>. (2018).
50. R Core Team. *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.* <https://www.R-project.org/>. (2019).
51. Carrillo, B., Yanofsky, C., Laboissiere, S., Nadon, R. & Kearney, R. E. Methods for combining peptide intensities to estimate relative protein abundance. *Bioinform. Oxf. Engl.* **26**, 98–103 (2010).
52. Bolstad, B. *preprocessCore: A collection of pre-processing functions. R package version 1.48.0.* <https://github.com/bmbolstad/preprocessCore>. (2019).
53. Hastie, T., Tibshirani, R., Balasubramanian, N. & Chu, G. *impute: Imputation for microarray data. R package version 1.60.0.* (2019).
54. Bray, J. R. & Curtis, J. T. An ordination of the upland forest communities of Southern Wisconsin. *Ecol. Monogr.* **27**, 325–349 (1957).
55. Oksanen, J. *et al.* *vegan: Community Ecology Package. R package version 2.5-6.* <https://CRAN.R-project.org/package=vegan>. (2019).
56. Langfelder, P. & Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **9**, 559–559 (2008).
57. Campbell-Staton, S. C. *et al.* Winter storms drive rapid phenotypic, regulatory, and genomic shifts in the green anole lizard. *Science* **357**, 495–498 (2017).
58. Campbell-Staton, S. C., Bare, A., Losos, J. B., Edwards, S. V. & Cheviron, Z. A. Physiological and regulatory underpinnings of geographic variation in reptilian cold tolerance across a latitudinal cline. *Mol. Ecol.* **27**, 2243–2255 (2018).
59. Horvath, S. *Weighted network analysis: Applications in genomics and systems biology* (Springer, New York, 2011).
60. Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* **4**, Article17 (2005).
61. Smyth, G. K. limma: Linear models for microarray data. In *Bioinformatics and Computational Biology Solutions Using R and Bioconductor* (eds Gentleman, R. *et al.*) 397–420 (Springer, New York, 2005). https://doi.org/10.1007/0-387-29362-0_23.
62. Goodrich, B., Gabry, J., Ali, I. & Brilleman, S. *rstanarm: Bayesian applied regression modeling via Stan. R package version 2.17.4.* (Comprehensive R Archive Network (CRAN), 2018).
63. Stearns, S. C. & Koella, J. C. The evolution of phenotypic plasticity in life-history traits: Predictions of reaction norms for age and size at maturity. *Evolution* **40**, 893–913 (1986).
64. Schmalhausen, I. I. *Factors of Evolution: The Theory of Stabilizing Selection* (Blakiston, Philadelphia, 1949).
65. Dray, S. & Dufour, A.-B. The ade4 package: Implementing the duality diagram for ecologists. *J. Stat. Softw.* **22**, 1–20 (2007).

66. Hijmans, R. J. *geosphere: Spherical Trigonometry. R package version 1.5-10*. <https://CRAN.R-project.org/package=geosphere>. (2019).
67. Rieder, V. *et al.* DISMS2: A flexible algorithm for direct proteome-wide distance calculation of LC-MS/MS runs. *BMC Bioinform.* **18**, 148 (2017).
68. Grüning, B. *et al.* Bioconda: Sustainable and comprehensive software distribution for the life sciences. *Nat. Methods* **15**, 475–476 (2018).
69. Bairoch, A. & Apweiler, R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* **28**, 45–48 (2000).
70. Jones, P. *et al.* InterProScan 5: Genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
71. Tatusov, R. L., Galperin, M. Y., Natale, D. A. & Koonin, E. V. The COG database: A tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* **28**, 33–36 (2000).
72. Huerta-Cepas, J. *et al.* EggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **47**, D309–D314 (2019).
73. Huerta-Cepas, J. *et al.* Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper. *Mol. Biol. Evol.* **34**, 2115–2122 (2017).
74. Alexa, A. & Rahnenfuhrer, J. *topGO: Enrichment Analysis for Gene Ontology. R package version 2.38.1*. (2019).
75. El-Gebali, S. *et al.* The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**, D427–D432 (2019).
76. Fang, H. & Gough, J. dcGO: Database of domain-centric ontologies on functions, phenotypes, diseases and more. *Nucleic Acids Res.* **41**, D536–D544 (2013).
77. Kuhn, N. J., Setlow, B. & Setlow, P. Manganese(II) activation of 3-phosphoglycerate mutase of *Bacillus megaterium*: pH-Sensitive interconversion of active and inactive forms. *Arch. Biochem. Biophys.* **306**, 342–349 (1993).
78. Chander, M., Setlow, B. & Setlow, P. The enzymatic activity of phosphoglycerate mutase from gram-positive endospore-forming bacteria requires Mn²⁺ and is pH sensitive. *Can. J. Microbiol.* **44**, 759–767 (1998).
79. Ferrer, M., Chernikova, T. N., Yakimov, M. M., Golyshin, P. N. & Timmis, K. N. Chaperonins govern growth of *Escherichia coli* at low temperatures. *Nat. Biotechnol.* **21**, 1266–1267 (2003).
80. Strocchi, M., Ferrer, M., Timmis, K. N. & Golyshin, P. N. Low temperature-induced systems failure in *Escherichia coli*: Insights from rescue by cold-adapted chaperones. *Proteomics* **6**, 193–206 (2006).
81. Visudtiphoke, V., Watthanasurorot, A., Klinbunga, S., Menasveta, P. & Kirtikara, K. Molecular characterization of Calreticulin: A biomarker for temperature stress responses of the giant tiger shrimp *Penaeus monodon*. *Aquaculture* **308**, S100–S108 (2010).
82. Wehrly, K. E., Wang, L. & Mitro, M. Field-based estimates of thermal tolerance limits for trout: Incorporating exposure time and temperature fluctuation. *Trans. Am. Fish. Soc.* **136**, 365–374 (2007).
83. Alberts, B. *et al.* *Molecular Biology of the Cell* (Garland Science, New York, 2002).
84. Hagner-Holler, S., Pick, C., Girgenrath, S., Marden, J. H. & Burmester, T. Diversity of stonefly hexamerins and implication for the evolution of insect storage proteins. *Insect Biochem. Mol. Biol.* **37**, 1064–1074 (2007).
85. Descazeaud, V., Mestre, E., Marquet, P. & Essig, M. Calcineurin regulation of cytoskeleton organization: A new paradigm to analyse the effects of calcineurin inhibitors on the kidney. *J. Cell. Mol. Med.* **16**, 218–227 (2012).
86. Urra, H. *et al.* IRE1a governs cytoskeleton remodelling and cell migration through a direct interaction with filamin A. *Nat. Cell Biol.* **20**, 942–953 (2018).
87. Tong, M. & Jiang, Y. FK506-binding proteins and their diverse functions. *Curr. Mol. Pharmacol.* **9**, 48–65 (2015).
88. Miranti, C. K. & Brugge, J. S. Sensing the environment: A historical perspective on integrin signal transduction. *Nat. Cell Biol.* **4**, E83–E90 (2002).
89. Walsh, C. T., Garneau-Tsodikova, S. & Gatto, G. J. Protein posttranslational modifications: The chemistry of proteome diversifications. *Angew. Chem. Int. Ed.* **44**, 7342–7372 (2005).
90. Snape, J. R., Maund, S. J., Pickford, D. B. & Hutchinson, T. H. Ecotoxicogenomics: The challenge of integrating genomics into aquatic and terrestrial ecotoxicology. *Aquat. Toxicol.* **67**, 143–154 (2004).
91. Nikinmaa, M. & Rytönen, K. T. Functional genomics in aquatic toxicology—Do not forget the function. *Aquat. Toxicol.* **105**, 16–24 (2011).
92. Kearney, E. B., Ackrell, B. A. C., Mayr, M. & Singer, T. P. Activation of succinate dehydrogenase by anions and pH. *J. Biol. Chem.* **249**, 2016–2020 (1974).
93. Bissoli, G. *et al.* Peptidyl-prolyl cis-trans isomerase ROF2 modulates intracellular pH homeostasis in Arabidopsis. *Plant J. Cell Mol. Biol.* **70**, 704–716 (2012).
94. Simčič, T. & Brancelj, A. Effects of pH on electron transport system (ETS) activity and oxygen consumption in *Gammarus fossarum*, *Asellus aquaticus* and *Niphargus sphagnicolus*. *Freshw. Biol.* **51**, 686–694 (2006).
95. Kadrmas, J. L. & Beckerle, M. C. The LIM domain: From the cytoskeleton to the nucleus. *Nat. Rev. Mol. Cell Biol.* **5**, 920–931 (2004).
96. van der Flier, A. & Sonnenberg, A. Structural and functional aspects of filamins. *Biochim. Biophys. Acta BBA Mol. Cell Res.* **1538**, 99–117 (2001).
97. Sun, H. Q., Yamamoto, M., Mejillano, M. & Yin, H. L. Gelsolin, a multifunctional actin regulatory protein. *J. Biol. Chem.* **274**, 33179–33182 (1999).
98. Diskin, S. *et al.* Galectin-8 promotes cytoskeletal rearrangement in trabecular meshwork cells through activation of rho signaling. *PLoS ONE* **7**, e44400 (2012).
99. Motizuki, M., Yokota, S. & Tsurugi, K. Effect of low pH on organization of the actin cytoskeleton in *Saccharomyces cerevisiae*. *Biochim. Biophys. Acta* **1780**, 179–184 (2008).
100. Wang, F., Sampogna, R. V. & Ware, B. R. pH dependence of actin self-assembly. *Biophys. J.* **55**, 293–298 (1989).
101. Sperelakis, N. *Cell Physiology Source book* (Academic Press, Amsterdam, 2012).
102. Tomanek, L. Proteomics to study adaptations in marine organisms to environmental stress. *J. Proteomics* **105**, 92–106 (2014).
103. Tomanek, L., Zuzow, M. J., Ivanina, A. V., Beniash, E. & Sokolova, I. M. Proteomic response to elevated PCO₂ level in eastern oysters, *Crassostrea virginica*: Evidence for oxidative stress. *J. Exp. Biol.* **214**, 1836–1844 (2011).
104. Koritzinsky, M. *et al.* Two phases of disulfide bond formation have differing requirements for oxygen. *J. Cell Biol.* **203**, 615–627 (2013).
105. L'Haridon, F. *et al.* A permeable cuticle is associated with the release of reactive oxygen species and induction of innate immunity. *PLoS Pathog.* **7**, e1002148 (2011).
106. Richards, A. G. Studies on arthropod cuticle—XIII: The penetration of dissolved oxygen and electrolytes in relation to the multiple barriers of the epicuticle. *J. Insect Physiol.* **1**, 23–39 (1957).
107. Wang, K. *et al.* Redox homeostasis: The linchpin in stem cell self-renewal and differentiation. *Cell Death Dis.* **4**, e537–e537 (2013).
108. Szablowska-Gadomska, I., Zayat, V. & Buzanska, L. Influence of low oxygen tensions on expression of pluripotency genes in stem cells. *Acta Neurobiol. Exp. (Warsz.)* **71**, 86–93 (2011).
109. Dreffs, A., Henderson, D., Dmitriev, A. V., Antonetti, D. A. & Linsenmeier, R. A. Retinal pH and acid regulation during metabolic acidosis. *Curr. Eye Res.* **43**, 902–912 (2018).
110. Raeker, M. Ö *et al.* Targeted deletion of the zebrafish obscurin A RhoGEF domain affects heart, skeletal muscle and brain development. *Dev. Biol.* **337**, 432 (2010).

111. Serafim, A. *et al.* Application of an integrated biomarker response index (IBR) to assess temporal variation of environmental quality in two Portuguese aquatic systems. *Ecol. Indic.* **19**, 215–225 (2012).
112. Berra, E., Forcella, M., Giacchini, R., Rossaro, B. & Parenti, P. Biomarkers in Caddisfly Larvae of the Species *Hydropsyche pellucidula* (Curtis, 1834) (Trichoptera: Hydropsychidae) measured in natural populations and after short term exposure to fenitrothion. *Bull. Environ. Contam. Toxicol.* **76**, 863–870 (2006).
113. Wernersson, A.-S. *et al.* The European technical report on aquatic effect-based monitoring tools under the water framework directive. *Environ. Sci. Eur.* **27**, 7 (2015).
114. Ryan, J. A. & Hightower, L. E. Stress proteins as molecular biomarkers for environmental toxicology. In *Stress-Inducible Cellular Responses* (eds Feige, U. *et al.*) (Birkhäuser, Basel, 1996). https://doi.org/10.1007/978-3-0348-9088-5_28.
115. Sanders, B. M. Stress proteins in aquatic organisms: An environmental perspective. *Crit. Rev. Toxicol.* **23**, 49–75 (1993).
116. Barata, C. *et al.* Combined use of biomarkers and in situ bioassays in *Daphnia magna* to monitor environmental hazards of pesticides in the field. *Environ. Toxicol. Chem.* **26**, 370–379 (2007).
117. Dorts, J. *et al.* Ecotoxicoproteomics in gills of the sentinel fish species, *Cottus gobio*, exposed to perfluorooctane sulfonate (PFOS). *Aquat. Toxicol.* **103**, 1–8 (2011).
118. Daborn, P. J. *et al.* A single P450 allele associated with insecticide resistance in *Drosophila*. *Science* **297**, 2253–2256 (2002).
119. Amichot, M. *et al.* Point mutations associated with insecticide resistance in the *Drosophila* cytochrome P450 Cyp6a2 enable DDT metabolism. *Eur. J. Biochem.* **271**, 1250–1257 (2004).
120. Dunkov, B. C. *et al.* The *Drosophila* cytochrome P450 gene Cyp6a2: Structure, localization, heterologous expression, and induction by phenobarbital. *DNA Cell Biol.* **16**, 1345–1356 (1997).
121. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Nat. Preced.* <https://doi.org/10.1038/npre.2010.4282.2> (2010).
122. Soneson, C. & Delorenzi, M. A comparison of methods for differential expression analysis of RNA-seq data. *BMC Bioinform.* **14**, 91 (2013).
123. Kim, S. & Coulombe, P. A. Emerging role for the cytoskeleton as an organizer and regulator of translation. *Nat. Rev. Mol. Cell Biol.* **11**, 75–81 (2010).
124. Parker, A. L., Kavallaris, M. & McCarroll, J. A. Microtubules and their role in cellular stress in cancer. *Front. Oncol.* **4**, 153 (2014).
125. Skelly, D. A., Ronald, J. & Akey, J. M. Inherited variation in gene expression. *Annu. Rev. Genomics Hum. Genet.* **10**, 313–332 (2009).
126. Brem, R. B., Yvert, G., Clinton, R. & Kruglyak, L. Genetic dissection of transcriptional regulation in budding yeast. *Science* **296**, 752–755 (2002).
127. Arias, M. B., Poupin, M. J. & Lardies, M. A. Plasticity of life-cycle, physiological thermal traits and Hsp70 gene expression in an insect along the ontogeny: Effect of temperature variability. *J. Therm. Biol.* **36**, 355–362 (2011).
128. Place, S. P. & Hofmann, G. E. Constitutive expression of a stress-inducible heat shock protein gene, hsp70, in phylogenetically distant Antarctic fish. *Polar Biol.* **28**, 261–267 (2005).
129. Hotaling, S. *et al.* Mountain stoneflies may tolerate warming streams: evidence from organismal physiology and gene expression. *bioRxiv* 2019.12.16.878926 (2019). <https://doi.org/10.1101/2019.12.16.878926>.
130. Cuellar, J. *et al.* Assisted protein folding at low temperature: Evolutionary adaptation of the Antarctic fish chaperonin CCT and its client proteins. *Biol. Open* **3**, 261–270 (2014).
131. Cantonati, M., Füreder, L., Gerecke, R., Jüttner, I. & Cox, E. J. Crenic habitats, hotspots for freshwater biodiversity conservation: Toward an understanding of their ecology. *Freshw. Sci.* **31**, 463–480 (2012).
132. Hofmann, G. E. & Todgham, A. E. Living in the now: Physiological mechanisms to tolerate a rapidly changing environment. *Annu. Rev. Physiol.* **72**, 127–145 (2010).
133. Pörtner, H. O., Peck, L. & Somero, G. Thermal limits and adaptation in marine Antarctic ectotherms: An integrative view. *Philos. Trans. R. Soc. B Biol. Sci.* **362**, 2233–2258 (2007).
134. Shah, A. A. *et al.* Climate variability predicts thermal limits of aquatic insects across elevation and latitude. *Funct. Ecol.* <https://doi.org/10.1111/1365-2435.12906> (2018).
135. Treanor, H. B., Giersch, J. J., Kappenman, K. M., Muhlfeld, C. C. & Webb, M. A. H. Thermal tolerance of meltwater stonefly *Lednia tumana* nymphs from an alpine stream in Waterton-Glacier International Peace Park, Montana, USA. *Freshw. Sci.* **32**, 597–605 (2013).
136. Forsman, A. & Wennersten, L. Inter-individual variation promotes ecological success of populations and species: Evidence from experimental and comparative studies. *Ecography* **39**, 630–648 (2016).
137. Cogne, Y. *et al.* Comparative proteomics in the wild: Accounting for intrapopulation variability improves describing proteome response in a *Gammarus pulex* field population exposed to cadmium. *Aquat. Toxicol.* **214**, 105244 (2019).
138. Gotelli, N. J., Ellison, A. M. & Ballif, B. A. Environmental proteomics, biodiversity statistics and food-web structure. *Trends Ecol. Evol.* **27**, 436–442 (2012).
139. Liu, F. *et al.* New Perspectives on host-parasite interplay by comparative transcriptomic and proteomic analyses of *Schistosoma japonicum*. *PLOS Pathog.* **2**, e29 (2006).
140. Nold, S. C. & Zwart, G. Patterns and governing forces in aquatic microbial communities. *Aquat. Ecol.* **32**, 17–35 (1998).
141. Pass, D. A. *et al.* The effect of anthropogenic arsenic contamination on the earthworm microbiome. *Environ. Microbiol.* **17**, 1884–1896 (2015).
142. Perez-Riverol, Y. *et al.* The PRIDE database and related tools and resources in 2019: Improving support for quantification data. *Nucleic Acids Res.* **47**, D442–D450 (2019).

Acknowledgements

We thank Stefan and Christian Zaenker and Haris Mujagic for indispensable help during field-work. We thank the Harz National Park authorities for granting sampling permission and area access. Part of the calculations was performed at sciCORE (<https://scicore.unibas.ch/>) scientific computing center at the University of Basel.

Author contributions

This study is part of the doctoral research of J.N.E. under the supervision of S.F. J.N.E. conceived the idea. S.F. and J.N.E. conceived of study design and conducted field-work. D.R. and J.N.E. generated the mass spectrometry data. J.N.E. analyzed data and wrote the manuscript.

Funding

This study was funded by Schweizerische Nationalfonds zur Förderung der Wissenschaftlichen Forschung (31003A_176234).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-72569-4>.

Correspondence and requests for materials should be addressed to J.N.E.

Reprints and permissions information is available at www.nature.com/reprints.

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



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

© The Author(s) 2020