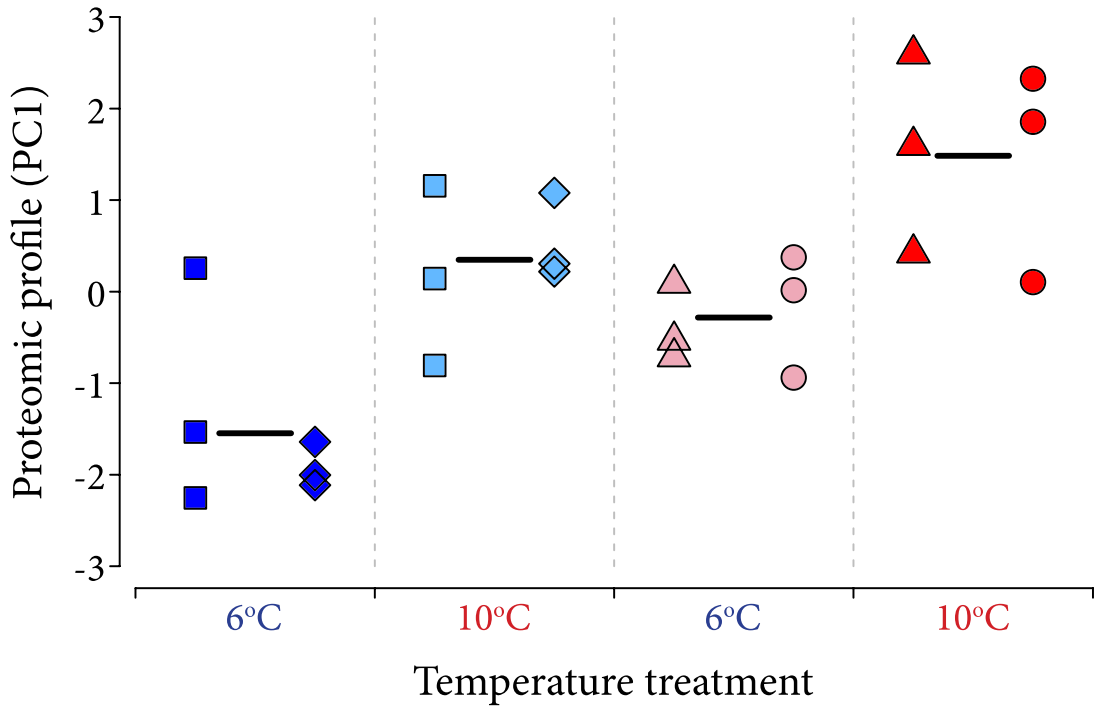
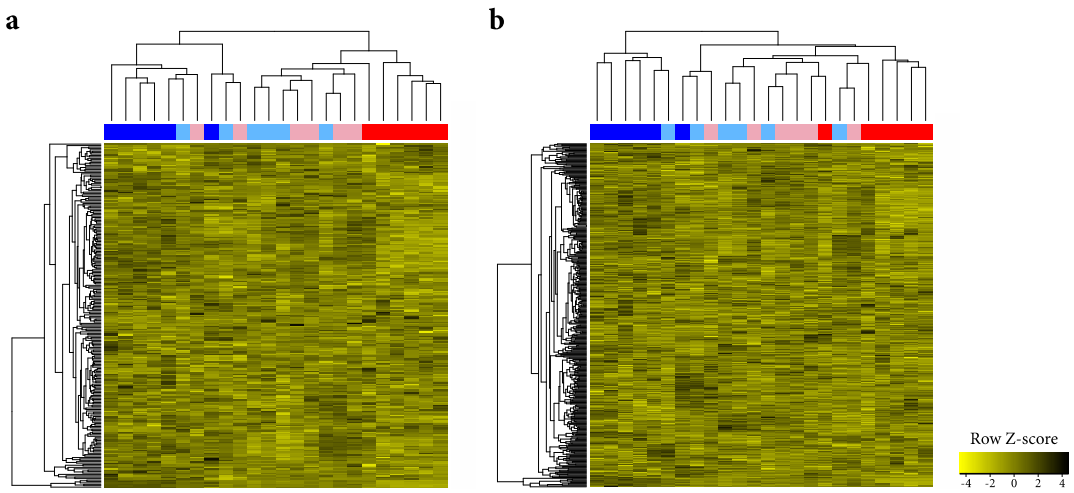
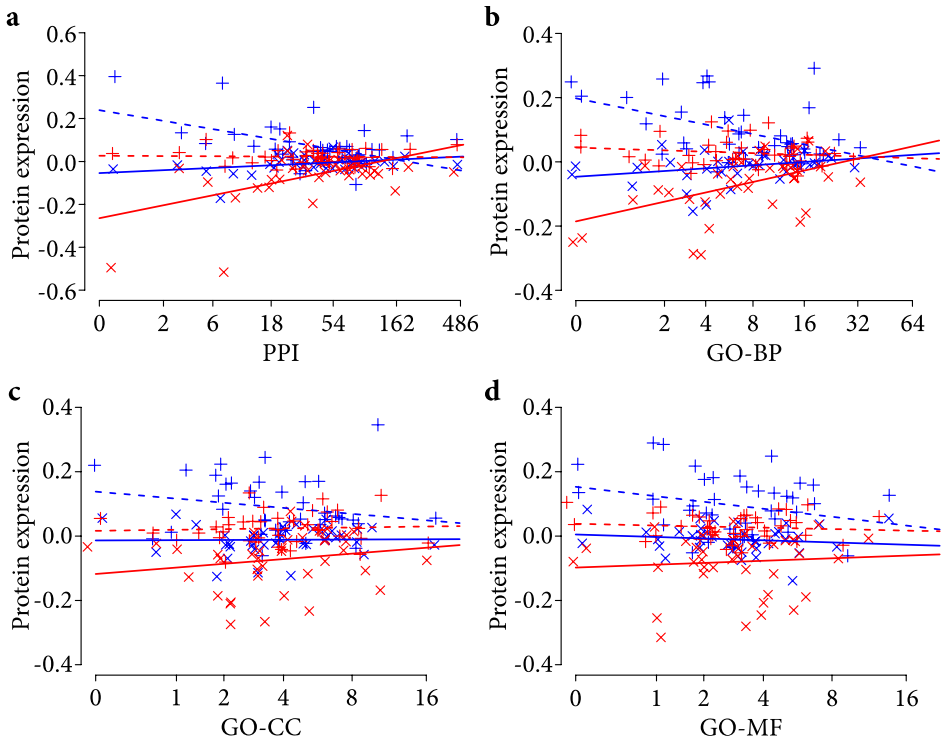
Supplementary Figures



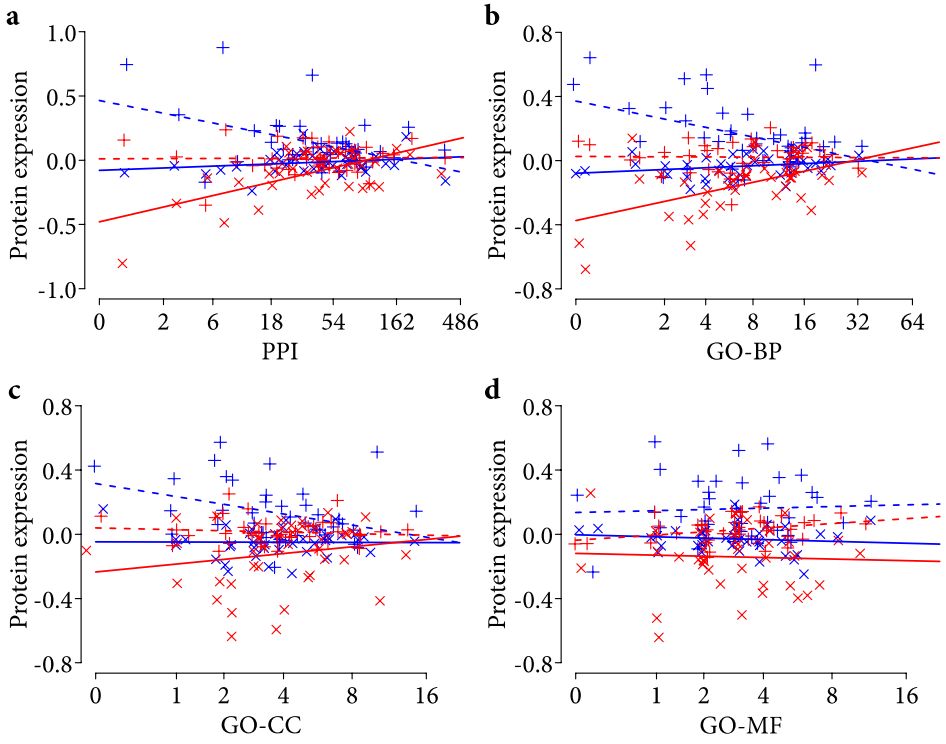
Supplementary Figure 1 | Global thermal profiles of 408 proteins with imputation for missing values in the grayling sub-populations. Overview of individual grayling protein expression profiles on the first principal component (PC1: 21.8% of the variance in expression level of the quantified proteins). The GLMM revealed a highly significant effect of temperature treatment on protein expression profiles (plastic component; GLMM, type II Wald F tests with Kenward-Roger df: *P* = 3.2E-05, F = 30.19, df = 1, df.res = 18, *n* = 24), a near-significant effect of thermal origin on protein expression profiles (evolutionary component; *P* = 0.069, F = 12.94, df = 1, df.res = 2, *n* = 24), and no interaction between them (*P* = 0.848, F = 0.04, df = 1, df.res = 18, *n* = 24). Symbols indicate different sub-populations and colours reflect thermal origin and temperature treatment (blue = cold thermal origin - 6°C, light blue = cold thermal origin - 10°C, pink = warm thermal origin - 6°C, red = warm thermal origin - 10°C), i.e. lighter colours indicate "non-local" origin-treatment combinations. Black-coloured horizontal lines represent the average over all six biological replicates of the same thermal origin within a temperature treatment.



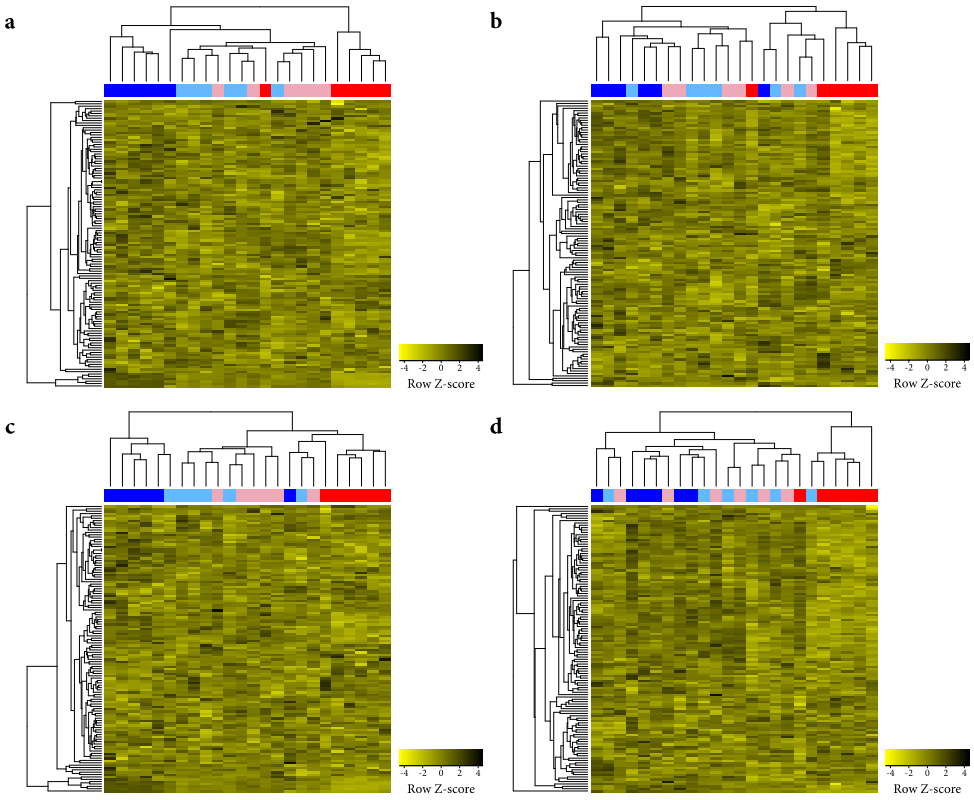
Supplementary Figure 2 | Profiles of the studied proteins visualised using heatmaps. Hierarchical clustering was performed on normalised expression data and was based on Euclidean distances for (**a**) 244 proteins without missing data and (**b**) 408 proteins with imputation for missing values. Clustering clearly distinguished cold from warm sub-populations when reared in the natal-resembling temperature. Colours reflect thermal origin and temperature treatment (blue = cold thermal origin - 6^{o}C, light blue = cold thermal origin - 10^{o}C, pink = warm thermal origin - 6^{o}C, red = warm thermal origin - 10^{o}C) i.e. lighter colours indicate "non-local" origin-treatment combinations. Raw Z-score denotes the scaled expression level of each protein across all samples in the heatmap.



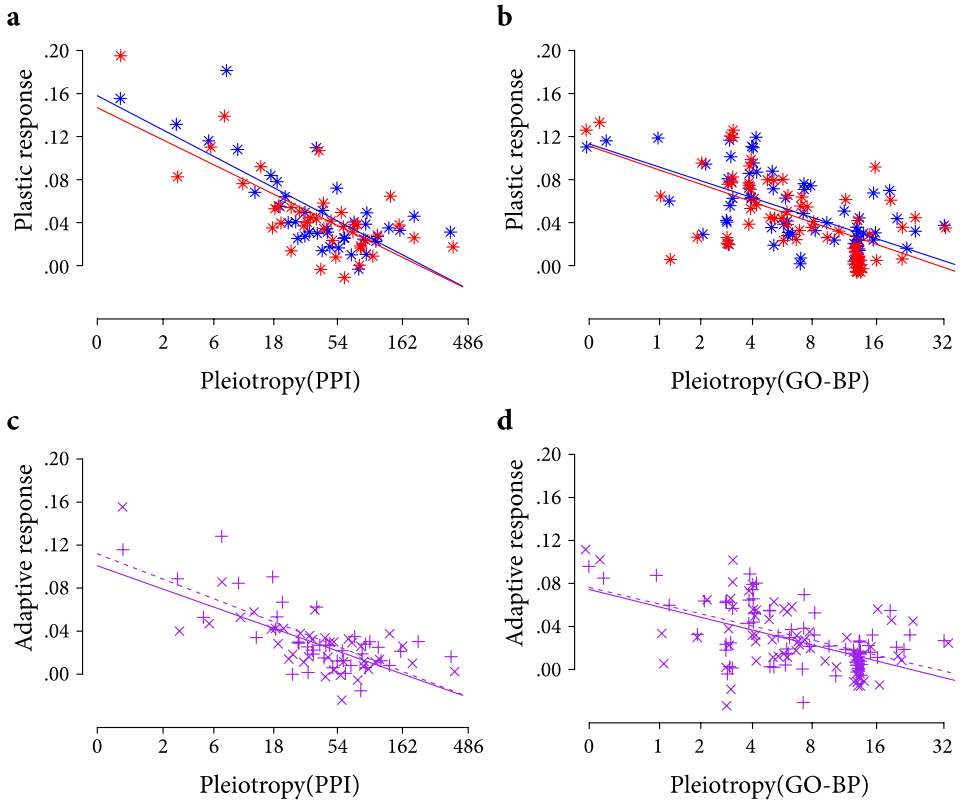
Supplementary Figure 3 | The effect of gene pleiotropy on protein expression responses using the 30% most variable proteins. Regression lines illustrate that gene pleiotropy, i.e. (**a**) PPI or (**b**) GO-BP, constrains plastic and evolutionary protein expression response, an effect weaker or absent for (**c**) GO-CC and (**d**) GO-MF. (**a**) GLMM, type II Wald F tests with Kenward-Roger df: P_{PL} = 2.67E-39, n = 960, bootstrap = 100%; P_{EV} = 6.56E-19, n = 960, bootstrap = 100%. (**b**) P_{PL} = 2.33E-10, n = 960, bootstrap = 95%; P_{EV} = 7.69E-05, n = 960, bootstrap = 83%. (**c**) P_{PL} = 1.77E-02, n = 960, bootstrap = 51%; P_{EV} = 2.51E-01, n = 960, bootstrap = 32%. (**d**) P_{PL} = 8.02E-01, n = 960, bootstrap = 27%; P_{EV} = 7.89E-01, n = 960, bootstrap = 17%. P_{PL} is the significance of the plastic response and P_{EV} is the significance of the evolutionary response in protein expression. Mild jittering of the points along the x-axis was applied to improve plot clarity. Lines are linear regression fits used for visualisation. Blue/red colour indicates cold/warm thermal origin; Continuous/dashed lines indicate 6°C/10°C temperature treatment. Plastic response is the difference in solid vs. dashed lines of the same colour and evolutionary response the difference between colours of the same type of line.



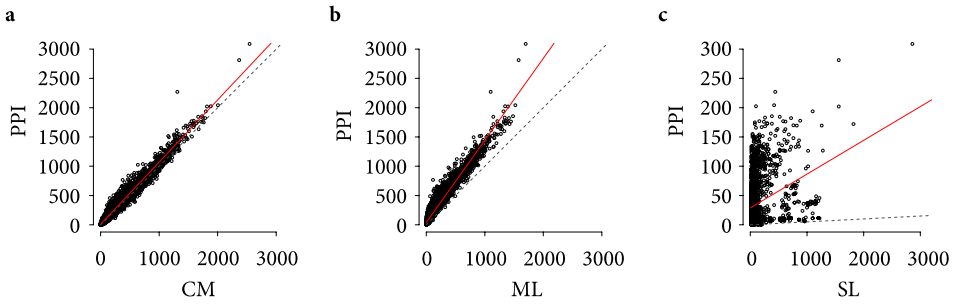
Supplementary Figure 4 | The effect of gene pleiotropy on protein expression responses using multidimensional scaling. These results are based on the 408 proteins with imputation for missing values. Regression lines illustrate that gene pleiotropy, i.e. (**a**) PPI or (**b**) GO-BP, constrains plastic and evolutionary protein expression response, an effect weaker or absent for (**c**) GO-CC and (**d**) GO-MF. (**a**) GLMM, type II Wald F tests with Kenward-Roger df: P_{PL} = 3.61E-13, n = 960, bootstrap = 94%; P_{EV} = 3.63E-09, n = 960, bootstrap = 92%. (**b**) P_{PL} = 1.24E-06, n = 960, bootstrap = 87%; P_{EV} = 3.62E-04, n = 960, bootstrap = 85%. (**c**) P_{PL} = 1.29E-02, n = 960, bootstrap = 54%; P_{EV} = 9.37E-03, n = 960, bootstrap = 49%. (**d**) P_{PL} = 3.82E-01, n = 960, bootstrap = 29%; P_{EV} = 3.70E-01, n = 960, bootstrap = 24%. P_{PL} is the significance of the plastic response and P_{EV} is the significance of the evolutionary response in protein expression. Mild jittering of the points along the x-axis was applied to improve plot clarity. Lines are linear regression fits used for visualisation. Blue/red indicates cold/warm thermal origin; Continuous/dashed indicates 6°C/10°C temperature treatment. Plastic response is the difference in solid vs. dashed lines of the same colour and evolutionary response the difference between colours of the same type of line.



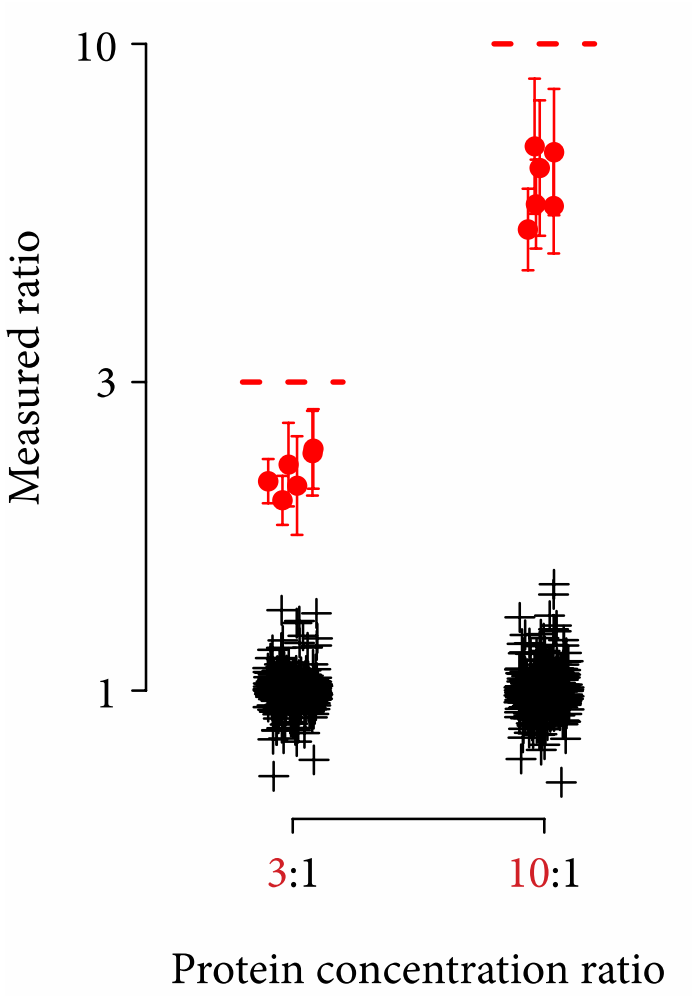
Supplementary Figure 5 | Proteomic profiles of genes with lower and higher pleiotropy visualised using heatmaps. Hierarchical clustering was performed on normalised expression data and was based on Euclidean distances. Clustering between thermal origin-temperature treatment combinations was better for the genes with low pleiotropy levels. (**a**) Lower PPI number. (**b**) Higher PPI number. (**c**) Lower GO-BP counts. (**d**) Higher GO-BP counts. These results are based on the 244 proteins with no missing values based on Euclidean distances. Colours reflect thermal origin and temperature treatment (blue = cold thermal origin - 6^{o}C, light blue = cold thermal origin - 10^{o}C, pink = warm thermal origin - 6^{o}C, red = warm thermal origin - 10^{o}C) i.e. lighter colours indicate "non-local" origin-treatment combinations. Raw Z-score denotes the scaled expression level of each protein across all samples in the heatmap.



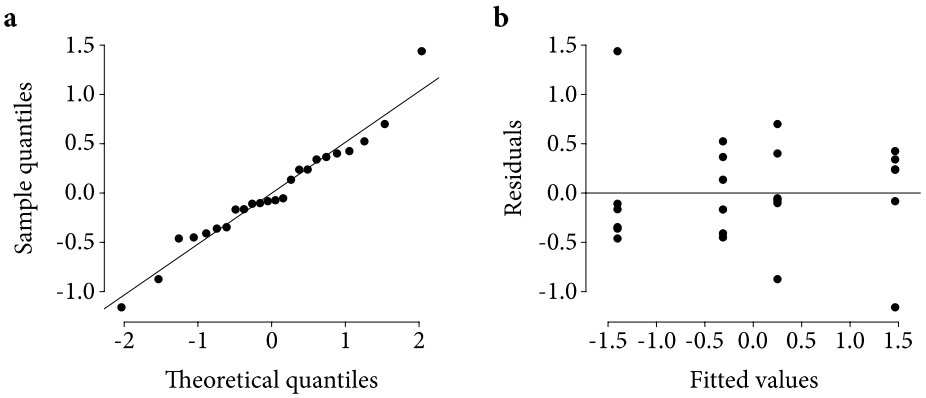
Supplementary Figure 6 | The effect of gene pleiotropy on protein expression responses after taking into account the expression level of the proteins. In all cases, the constraining effect of gene pleiotropy remained highly significant. GLMM, type II Wald F tests with Kenward-Roger df: (**a**) P = 7.03E-26, n = 960. (**b**) P = 5.08E-13, n = 960. (**c**) P = 5.99E-15, n = 960. (**d**) P = 2.21E-09, n = 960. Similar results were obtained with the rest of the approaches described. Plastic response of protein expression is represented as the difference in mean protein expression levels between 6^{o}C and 10^{o}C temperature treatments in grayling of cold (blue colour) and warm (red colour) thermal origin. Evolutionary response in protein expression is represented as the difference in mean protein expression levels between grayling of cold and warm thermal origins in the 6^{o}C (× symbol, continuous line) and 10^{o}C (+ symbol, dashed line) temperature treatment. *P*-values for the plastic response represent the interaction between gene pleiotropy and temperature treatment and for evolutionary response the interaction between gene pleiotropy and thermal origin. Lines are linear regression fits used for visualisation. Mild jittering of the points along the x-axis was applied to improve plot clarity. Lines are linear regression fits. These results derived from GLMM analysis on mean standardised protein expression levels of grouped proteins (bin = 10 proteins).



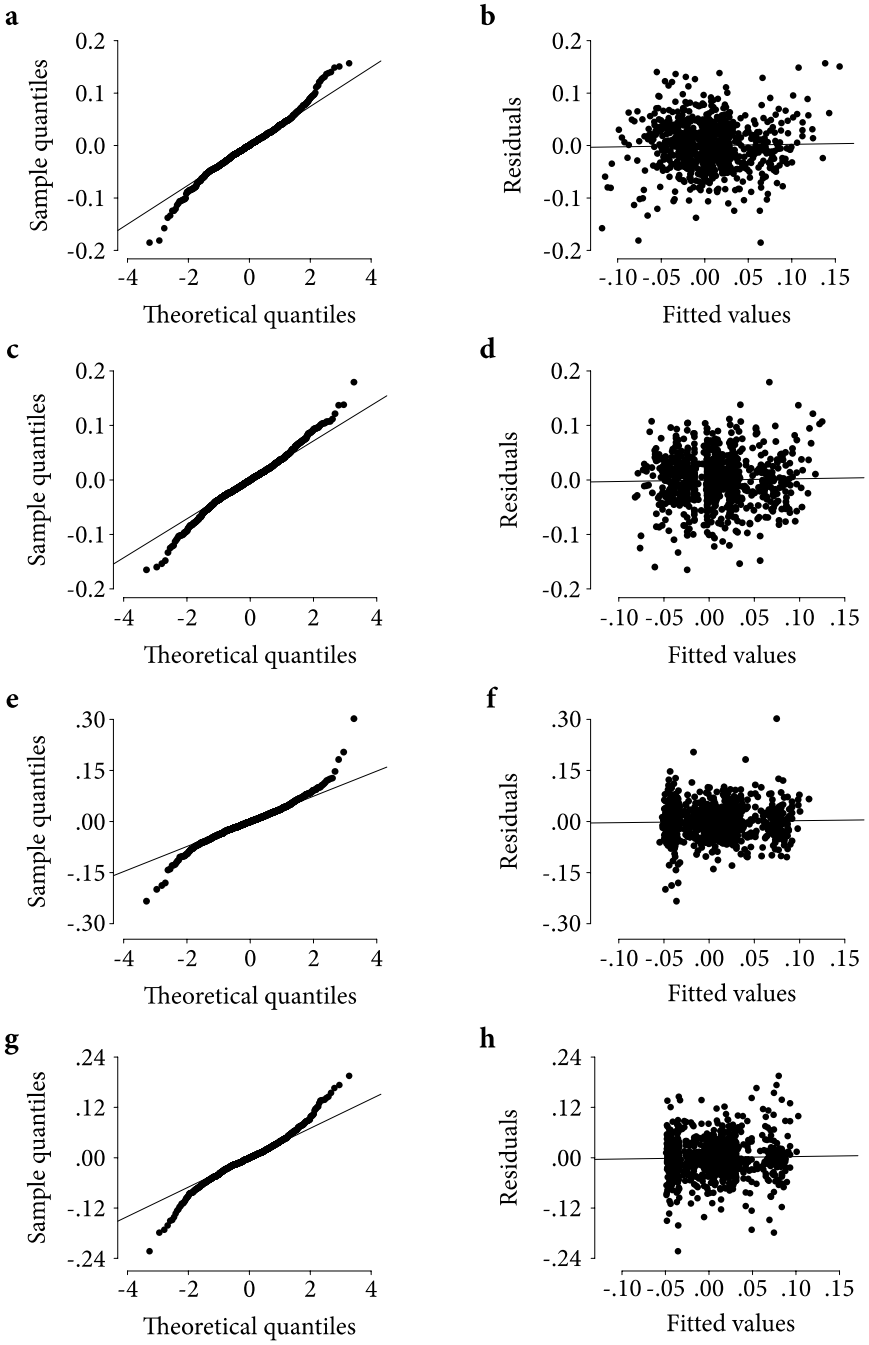
Supplementary Figure 7 | Correlation between predicted protein interactions for *Homo sapiens* and other PPI proxies. In the Funcoup 2.0 database[1], predicted protein-protein interactions (PPI) correlate strongly with (**a**) links between pairs of protein-members of the same complex (CM, Spearman's rho = 0.97, *n* = 10371) and (**b**) links between proteins from the same metabolic pathways (ML, rho = 0.94, *n* = 10371) but poorly with (**c**) links between proteins from the same signalling pathways (SL, rho = 0.42, *n* = 10371). CM and ML thus seem to be good proxies for predicted PPI to be used for *Danio rerio* genes in cases when predicted PPIs are missing. Dashed lines are identity lines (slope = 1, intercept = 0). Red lines are linear regression fits used for visualisation.



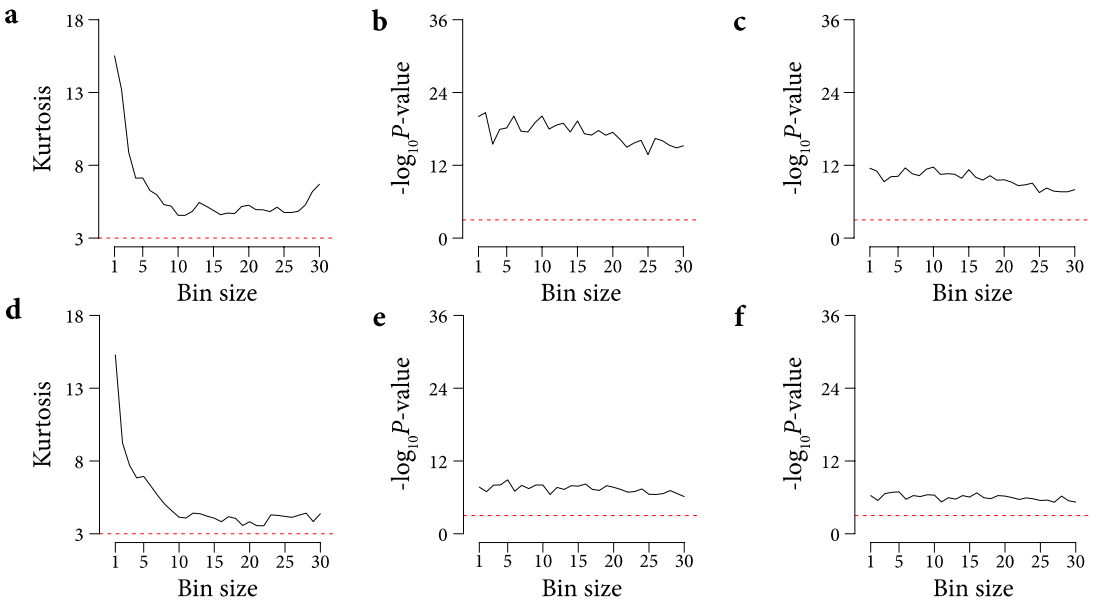
Supplementary Figure 8 | Measured expression ratios for six proteins spiked at known ratios. Spiked proteins (red colour) were close to the expected 3:1 and 10:1 ratios (dashed lines). The amount of the background proteins (black colour) was left unchanged. Errors bars indicate 95% confidence intervals.



Supplementary Figure 9 | Assessing the assumptions of GLMM used to identify plastic and evolutionary components in protein expression profiles. (**a**) We determined whether residuals roughly followed a normal distribution. Line indicates the estimated normal distribution. (**b**) We also confirmed the absence of a trend between residuals and fitted values (absence of heteroscedasticity). Line is linear regression fit.



Supplementary Figure 10 | Assessing the assumptions of GLMM used to describe the effect of gene pleiotropy on protein expression profiles. We determined whether residuals roughly followed a normal distribution (sample vs. theoretical quantiles, lines indicate the estimated normal distribution), and confirmed the absence of a trend between residuals and fitted values (absence of heteroscedasticity, lines are linear regression fits) for (**a**), and (**b**) PPI. (**c**), and (**d**) GO-BP. (**e**), and (**f**) GO-CC. (**g**), and (**h**) GO-MF. The presented plots are from the analysis with mean standardised expression of grouped proteins.



Supplementary Figure 11 | The effect of bin size on the kurtosis of GLMM residual distribution and on the significance of the effect of gene pleiotropy on plastic and evolutionary responses. GLMM analyses were performed using mean standardised protein expression per bin. (**a**) PPI, kurtosis. (**b**) PPI, plastic response. (**c**) PPI, evolutionary response. (**d**) GO-BP, kurtosis. (**e**) GO-BP, plastic response. (**f**) GO-BP, evolutionary response. Regardless of bin size, the effect of gene pleiotropy on both plastic and evolutionary protein expression responses remained highly significant. Kurtosis approximated that of normal distribution for bin sizes of 10 and above. Red dashed lines indicate P = 0.001 in (**b**), (**c**), (**e**), (**f**), and the kurtosis of a normal distribution (kurtosis = 3) in (**a**), (**c**).

Supplementary Tables

Supplementary Table 1. Thermal characteristics of the studied streams. The average number of days per month each of the studied streams had water temperature above 6°C and 9°C during the egg/juvenile incubation period. Measurements have been taken in years between 2000 and 2008. **Red**/**blue** indicate warm/cold streams.

Stream (years of measurements, *n*)	June		July	
	> 6°C	> 9°C	> 6°C	> 9°C
Steinbekken (4)	30	3	31	31
Sandbekken (4)	22	0	30	26
Hyrion (8)	5	0	30	19
Valåe (7)	7	0	30	15

Supplementary Table 2. F_{ST} and geographic distances (km) between sub-populations. Above diagonal: pairwise F_{ST} values estimated using 19 microsatellite loci. Significant F_{ST} values are indicated in **bold**. Below diagonal: pairwise geographic distances. **Red**/**blue** indicate warm/cold streams. Data taken from Junge et al[2].

	Steinbekken (n = 4)	Sandbekken (*n* = 4)	Hyrion (*n* = 8)	Valåe (*n* = 7)
Steinbekken		**0.0205**	**0.0067**	-0.0003
Sandbekken	8.5		**0.0212**	**0.0131**
Hyrion	4.81	3.66		0.0019
Valåe	6.73	1.83	1.84	

Supplementary Table 3. Details of the development stage of the studied samples. Grayling embryos sampled for protein extraction were of similar developmental stage as estimated based on number of degree-days in relation to average degree-days for 50% hatching. Eggs were sampled daily during the common garden experiment and therefore successive samples differed by 6 or 10 degree-days depending the temperature. **Red**/**blue** indicate warm/cold streams.

Name (and type) of the spawning site of origin	6°C temperature treatment			10°C temperature treatment		
	Degree-days of embryos selected for protein extraction	Degree-days to 50% hatching (SE, *n*)	Proportion of degree-days for embryos selected for protein extraction to degree days to 50% hatching	Degree-days of embryos selected for protein extraction	Degree-days to 50% hatching (SE, *n*)	Proportion of degree-days for embryos selected for protein extraction to degree days to 50% hatching
Steinbekken	168.00	201.75 (± 3.73, 276)	0.83	150.00	170.82 (± 4.06, 106)	0.88
Sandbekken	168.00	196.82 (± 3.31, 273)	0.85	150.00	162.7 (± 3.55, 109)	0.92
Hyrion	162.00	195.49 (± 3.65, 191)	0.83	150.00	176.85 (± 4.07, 111)	0.85
Valåe	162.00	187.27 (± 3.38, 189)	0.87	150.00	183.04 (± 2.57, 123)	0.82

Supplementary Table 4. Predicted upstream regulators for the protein expression profiles. Listed are the names of the regulators, calculated significance (Fisher's Exact Test: $P < 0.001$; studied genes (n) found regulated by the regulator: $n_{MYCN} = 65$, $n_{MYC} = 87$, $n_{TP53} = 58$, $n_{HSF1} = 12$, $n_{TFAP2A} = 8$, $n_{HSF2} = 6$, $n_{HNF4A} = 68$) for the overlap between the studied genes and the genes regulated by the regulator and Z-score for activation (positive values) or inhibition (negative) whenever significant [abs(Z-score) ≥ 2] in the different experimental groups. Role in the cell is as described in the IPA database. **Red**/**blue** colours visualise predicted activation/inhibition of the regulator, respectively.

Upstream regulator	Entrez gene name	Cellular role	*P*-value	Activation Z-score			
				Cold origin - 6°C	Cold origin - 10°C	Warm origin - 6°C	Warm origin - 10°C
MYCN	v-myc myelocytomatosis viral related oncogene	proliferation, apoptosis, transformation, expression in, cell death, growth, transactivation in, differentiation, transcription in, survival	1.93E-50	-2.575	7.405		6.801
MYC	v-myc myelocytomatosis viral oncogene homolog	apoptosis, proliferation, transformation, growth, cell cycle progression, differentiation, expression in, S phase, colony formation, G1 phase	1.83E-39		4.529		3.580
TP53	tumor protein p53	apoptosis, cell cycle progression, proliferation, cell death, growth, expression in, G1 phase, senescence, transformation, survival	5.62E-09				2.116
HSF1	heat shock transcription factor 1	expression in, apoptosis, cell death, activation in, proliferation, binding in, quantity, recruitment in, abnormal morphology, replication in	4.09E-05	2.140	-2.311		-2.397
TFAP2A	transcription factor AP-2 alpha	migration, growth, expression in, apoptosis, cell death, proliferation, development, invasion by, differentiation, invasion	4.87E-04		-2.121		
HSF2	heat shock transcription factor 2	quantity, production, transactivation in, morphology, mislocalization, apoptosis	3.68E-06	2.433		2.433	
HNF4A	hepatocyte nuclear factor 4, alpha	expression in, transactivation in, proliferation, transcription in, activation in, differentiation, apoptosis, growth, gluconeogenesis in, mitosis	1.99E-04			-2.214	2.768

Supplementary Table 5. Sample combinations in the iTRAQ 4-plex reactions. Samples were multiplexed to minimise technical artefacts. Reference sample contained the same amount of protein extract from the 24 samples that were analysed. Reaction represents the order that multiplex reactions entered the mass spectrometer. Two kits of iTRAQ 4-plex reagents were used (Kit A and Kit B). **Red/blue** indicates warm/cold sub-populations, respectively.

Reaction	iTRAQ kit	114 tag	115 tag		116 tag		117 tag	
1	Kit A	*Reference*	Hyrion	6 ^{o}C	Steinbekken	10 ^{o}C	Valåe	6 ^{o}C
2	Kit B	*Reference*	Steinbekken	10 ^{o}C	Valåe	6 ^{o}C	Sandbekken	10 ^{o}C
3	Kit A	*Reference*	Valåe	6 ^{o}C	Sandbekken	10 ^{o}C	Hyrion	10 ^{o}C
4	Kit B	*Reference*	Sandbekken	10 ^{o}C	Hyrion	10 ^{o}C	Steinbekken	6 ^{o}C
5	Kit A	*Reference*	Hyrion	10 ^{o}C	Steinbekken	6 ^{o}C	Valåe	10 ^{o}C
6	Kit B	*Reference*	Steinbekken	6 ^{o}C	Valåe	10 ^{o}C	Sandbekken	6 ^{o}C
7	Kit A	*Reference*	Valåe	10 ^{o}C	Sandbekken	6 ^{o}C	Hyrion	6 ^{o}C
8	Kit B	*Reference*	Sandbekken	6 ^{o}C	Hyrion	6 ^{o}C	Steinbekken	10 ^{o}C

Supplementary Table 6. The effect of genes with differing pleiotropy levels on protein expression responses. The results are of the GLMM analysis used to evaluate the relative contribution of genes with different pleiotropy levels on the plastic (effect of temperature treatment) and the evolutionary (effect of thermal origin) protein expression responses. GLMM, type II Wald F tests with Kenward-Roger df: ***P* < 0.05**, *n* = 24.

	PPI								GO-BP							
	low				high				low				high			
	F	df	df.Res	*P*	*F*	df	df.Res	*P*	*F*	df	df.Res	*P*	*F*	df	df.Res	*P*
Plastic response	65.19	1	18	2.15E-07	0.96	1	18	3.40E-01	50.53	1	18	1.26E-06	32.63	1	18	2.04E-05
Evolutionary response	30.63	1	2	3.11E-02	2.66E-07	1	2	1.00E+00	23.33	1	2	4.03E-02	14.82	1	2	6.13E-02
Interaction	0.11	1	18	7.46E-01	0.06	1	18	8.07E-01	0.13	1	18	7.21E-01	2.54E-05	1	18	9.96E-01
PC1 described variance			0.53				0.21				0.43				0.29	

Supplementary Methods

Protein extraction. For protein extraction, embryos were solubilised into 50 µL of fresh 1% SDS buffer using a Qiagen TissueLyser such that the larvae disintegrated without thawing. Samples were then incubated for 10 min at 70°C to complete the extraction and centrifuged for 10 min at 13,200 rpm (16168 × g) in an Eppendorf Centrifuge 5415D. The supernatant was then transferred to a fresh Eppendorf LoBind tube, precipitated for 4 h at -20°C using 300 μL of acetone, centrifuged for 20 min at 13,200 rpm, re-dissolved in 100 μL of 1% SDS and sonicated for 30 min. Protein concentration was then determined with a NanoDrop ND-1000 spectrophotometer.

iTRAQ labeling. For trypsin digestion and peptide-level iTRAQ labeling, samples were peptide-labeled with isobaric tags by iTRAQ reagents (4-plex, Applied Biosystems) following the manufacturer's instructions: 25 µg of each sample was suspended in 5 µL of 0.5 M triethyl ammonium bicarbonate (TEAB) and 0.25 µL of 2% sodium dodecyl sulfate (SDS). Samples were reduced by the addition of 0.5 µL of 50 mM tris-carboxyethyl phosphine hydrochloride (TCEP) and incubation at 60℃ for 1 h. Alkylation was performed by the addition of 0.25 µL of 84 mM iodoacetamide (IAA) and incubation at room temperature for 10 min. Samples were digested with 2.5 µL trypsin (1 µg/µL) (Promega, Madison, WI), and overnight incubation at 37℃. Ethanol was then added to the iTRAQ labels, and the reconstituted labels were added to their respective samples. Samples were incubated at room temperature for 1 h, combined into their respective multiplexes, and then evaporated.

Peptide fractionation. To clean and fractionate each sample we used a strong cation exchange (SCX) column and buffers supplied as part of the ICAT Cation Exchange

Buffer Pack - Cation Exchange Cartridge kit (Applied Biosystems), according to the manufacturer's protocol. Briefly, the pH of each sample was adjusted to pH = 3 with 2-3 ml of loading buffer before injection to the pre-conditioned cation exchange cartridge using 2 ml of load buffer. The cartridge column was washed with another 1 ml of loading buffer and then four peptide fractions were collected using 500 µl of four concentrations of the elution buffer (50 mM, 100 mM, 150 mM, and 350 mM KH_2PO_4). Each fraction was collected to a fresh Eppendorf LoBind tube and dried in a SpeedVac. Peptides were then brought into solution using 10 µl of 2% acetonitrile (ACN) - 0.1% trifluoroacetic acid (TFA) and desalted using C18 macrospin columns (The Nest Group, Southborough, MA) according to the protocol of Rappsilber et al[27]. Finally, samples were transferred to 30 µl of 2% formic acid (FA), sonicated for 5 min and analysed using nano-LC-coupled mass spectrometry.

Mass spectrometry. For LC-MS/MS, we loaded 64 samples (=4 fractions × 8 iTRAQ reactions × 2 technical replicates). Peptides were separated by reversed-phase chromatography using an Easy-nLC II nanoflow system connected to an Orbitrap Velos mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). A 100 µm x 3 cm trap column in-house packed with Magic C18AQ resin (200 Å, 5 µm; Michrom Bioresources, Auburn, USA) and a 75 µm x 15 cm PicoFrit analytical column (New Objective, Woburn, USA) packed with the Magic C18AQ resin were used. The mobile phases were 0.2% formic acid/2% acetonitrile (A) and 0.2% formic acid/95% acetonitrile (B). Peptides were separated at a flow rate of 300 nl/min with 102 min gradients as follows: initially 2 % B to 25% B (60 min), 40% B (90 min), and 100% B (92-102 min). The ionisation voltage was applied using a liquid junction. The Orbitrap was operated at a 200°C capillary temperature and 2.3 kV spray voltage.

Data were acquired in the data-dependent mode with up to ten MS/MS scans being recorded for each precursor ion scan. Precursor ion spectra were recorded in profile mode in the Orbitrap (m/z 300-1800, R = 30,000 at m/z 400, max injection time 100 ms, and max 1,000,000 ions), and MS/MS spectra were acquired in centroid mode in the Orbitrap (R = 7,500 at m/z 400, max injection time 200 ms, max 50,000 ions, HCD stepped normalised CE 40% and 50%, isolation width 2, activation time 0.1 ms, and the first mass fixed at m/z 100). Note, the instrument control software LTQ Tune Plus version 2.6.0.1050 was used in this study, and the HCD energy of 40 on that version would be equivalent to 35 on the updated version 2.7.0.1093. Mono-isotopic precursor selection was enabled, singly charged ions and ions with an unassigned charge state were rejected, and each fragmented ion was dynamically excluded for 90 s. The lock-mass option was enabled (m/z 445.120025).

Measuring protein expression levels. Thermo Orbitrap files (*.RAW) were converted to Mascot generic format (*.mgf) using the Proteome Discoverer software (Thermo Scientific). In the parameters we selected the Paragon method, iodoacetamide for cys alkylation, trypsin for digestion, thorough ID for search effort, biological modifications and amino acid substitutions for ID focus, and bias correction in quantification. Enabling the Paragon algorithm to tolerate for amino acid substitutions, 376 types of substitutions in particular[3], is expected to maximise the identification of peptide sequences despite the inherent variability between the grayling and the Atlantic salmon protein sequences used as a search database. To improve quantification accuracy, peaklists were preprocessed using an in-house developed algorithm[4] and technical replicates were averaged in the same analysis. With Cd-hit v.4.3[5] identical sequences were removed and only the longest entry was

retained, leaving 10033 sequences. Remaining redundancy was handled by the Pro Group algorithm of the ProteinPilot software. ProteinPilot searches were aligned using the Protein Alignment Template v.2.0 (ABSciex) and protein expression ratios using a common reference sample (Supplementary Table 5) were calculated for all samples.

To evaluate the accuracy of the quantification method, we spiked the six-protein mix provided with the iTRAQ kit in ratios 1:3:10 in samples that contained an equal amount of protein extract from a grayling embryo. Then we labeled each sample with a different isobaric tag (114:115:116, respectively). The aim was to employ the proposed methodology to quantify a subset of proteins that change across samples in a cloud of background proteins that remain unchanged. The amount of the six-protein mix compared to the background proteins was 1:1.25 μg. This is a likely an overrepresentation of the mix proteins in the lysates that regardless evaluates the accuracy and precision of the method, ensuring identification of the six proteins by mass spectrometry. Search database, identification and quantification parameters were as described in the main experiment. The UniProt accession numbers of the six proteins in the mix were P02769, P00722, P02754, P00711, P00698, and P02787 (Supplementary Data 3). Other than a compression of ratios, a known issue with iTRAQ in complex samples[6], measured ratios for the six proteins showed our employed methodology has the required precision and accuracy (Supplementary Data 3; Supplementary Fig. 8).

Additional information about the GLMM analyses. To test the effect of thermal origin (cold vs. warm sub-populations), temperature treatment (6^{o}C vs. 10^{o}C), and

their interaction on protein expression profiles, sub-population was included as a random factor nested within thermal origin and data for the individual embryos within each sub-population constituted the biological replicates for the first component (PC1) values of a principal component analysis (PCA). In this way, the analysis of the effect of the independent predictor variables can be performed in a single step and their interaction can be tested at the same time. PCA analysis on the data was performed with centering but no scaling since the difference in expression variance between proteins is assumed to have a biological meaning. For this analysis, we included 24 individuals (samples) and 244 proteins (variables) as proteins with missing data were removed. However, we also performed the PCA and subsequent GLMM analysis with all 408 quantified proteins by imputation of missing data[7, 8]. Data imputation was performed on log2-transformed data and missing data were replaced by values drawn from normal distribution (width = 0.3, shift = 1.8; Supplementary Data 4) using the program Perseus v.1.3.0.4 of the MaxQuant proteomic analysis suite[9] assuming missing data came from proteins close to the detection limit of the mass spectrometer in the experiment. For the GLMM analysis we used the *lmer* function from the *lme4* package in R[10]. Significance was tested using the *Anova* function from the *car* package[11], using Wald F-tests with Kenward-Roger approximation[12].

When testing the effect of gene pleiotropy on protein expression responses using the mean standardised expression of the 30% most variable proteins per group of 10 proteins, we used only the subset of 3 proteins that had the largest variance in expression level among the 24 individuals (30% upper quantile of variance in protein expression level) to calculate the mean expression level for each individual and repeated the GLMM analysis on mean standardised expression.

When testing the effect of gene pleiotropy on protein expression responses by performing multidimensional scaling analysis (MDS), for each set of 10 proteins along the pleiotropy gradient we performed a MDS analysis on one dimension based on Euclidean distances between protein expression levels among individuals and extracted the scores for each individual. These scores were calculated directly from the normalised protein expression ratios. Standardisation of the direction of the MDS scores within each group of 10 proteins was also performed in this case with regard to the Valåe sub-population in the 6°C temperature treatment since the signs of the values are arbitrary in MDS. The MDS scores were then used as the dependent variable in the GLMM. MDS analysis has the same sensitivity as PCA for missing values. For this reason we performed two types of analyses, one excluding the proteins with missing values and another with all proteins but with missing values replaced by imputed data.

Estimating expression levels using spectral counts. For these calculations we used the number of peptides identified with 95% confidence for each protein (Supplementary Data 4). This number roughly correlates with protein abundance as more abundant proteins will be selected for fragmentation by the mass spectrometer more often and will produce a higher abundance of spectra and thus more peptides[13]. However, the number of peptides also correlates with the length of a protein while the expected protein coverage also depends on the type of mass spectrometer and the protease that was used. To account for these factors, we normalised the peptide counts by the length of each protein and the expected coverage for each protein for Orbitrap instruments using trypsin. The latter was calculated with the tool IPEP[14]. We validated this approach using the results from the experiment in which we spiked six

proteins with known ratios but also known expression level ratios, namely 1.23/0.64/0.62/0.34/0.33/0.33 nmol respectively for each of the proteins with UniProt accessions P02754/P00711/P00698/P00722/P02769/P02787 (Supplementary Data 3). We observed a very good linear fit between calculated and known expression levels (Linear model, F test: R^2 = 0.76, P = 0.022, n = 6). We then performed a linear regression analysis within each experimental group using expression level estimates for each protein in our dataset as the independent variable and standardised expression level as the dependent variable to find that the effect of gene pleiotropy remains very significant (Supplementary Fig. 6).

Predicted PPI proxies for *Danio rerio*. The second version of the Funcoup database contains predicted interactions for *Homo sapiens* and *Danio rerio* and each interaction have been given Final Bayesian Scores (FBS) for four potential categories of coupling (protein protein interactions - PPI, involvement in a protein complex - CM, link in metabolic pathways - ML, link in signaling pathways -SL). Funcoup 2.0 database was found to be a good predictor for the experimentally observed PPI for the studied proteins (vs. Funcoup predicted human PPI: Spearman's rho = 0.54, P = 7.47E-32, n = 399). Due to a lack of available high quality data to train the model for PPI prediction in zebrafish, the Funcoup 2.0 database contains only predicted interactions based on CM, ML and SL scores for this species. In order to assess the consistency of the results we obtained with experimental PPI in human with the results obtained using Funcoup data, we first used the Funcoup human database to calculate the counts of predicted PPI for each protein of our dataset and rerun the GLMM analyses. Since no predicted PPI was available for *D. rerio*, we examined which of the three other categories correlated the most with PPI in Funcoup human database by comparing the

number of interactions predicted by each category, using all human proteins in the Funcoup database. CM and ML gave predicted counts highly correlated with PPI predicted counts while SL counts were less correlated (Supplementary Fig. 7). Following those results, we used CM and ML categories as proxies to predict interaction counts with the Funcoup *D. rerio* database and rerun our GLMM analysis with those estimates.

Supplementary References

1. Alexeyenko A, Schmitt T, Tjarnberg A, Guala D, Frings O, Sonnhammer ELL. Comparative interactomics with Funcoup 2.0. *Nucleic Acids Research* **40**, D821-D828 (2012).

2. Junge C, *et al.* Strong gene flow and lack of stable population structure in the face of rapid adaptation to local temperature in a spring-spawning salmonid, the European grayling (*Thymallus thymallus*). *Heredity* **106**, 460-471 (2011).

3. Shilov IV, *et al.* The paragon algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra. *Molecular & Cellular Proteomics* **6**, 1638-1655 (2007).

4. Rissanen J, Moulder R, Lahesmaa R, Nevalainen OS. Pre-processing of Orbitrap higher energy collisional dissociation tandem mass spectra to reduce erroneous iTRAQ ratios. *Rapid Communications in Mass Spectrometry* **26**, 2099-2104 (2012).

5. Li WZ, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658-1659 (2006).

6. Ow SY, Salim M, Noirel J, Evans C, Rehman I, Wright PC. iTRAQ underestimation in simple and complex mixtures: the good, the bad and the ugly. *Journal of Proteome Research* **8**, 5347-5355 (2009).

7. Deeb SJ, D'Souza RCJ, Cox J, Schmidt-Supprian M, Mann M. Super-SILAC allows classification of diffuse large B-cell lymphoma subtypes by their protein expression profiles. *Molecular & Cellular Proteomics* **11**, 77-89 (2012).

8. Smits AH, Jansen P, Poser I, Hyman AA, Vermeulen M. Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics. *Nucleic Acids Research* **41**, e28 (2013).

9. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology* **26**, 1367-1372 (2008).

10. Bates D, Maechler M, Bolker B. lme4: Linear mixed-effects models using S4 classes. R package version 0.999999-2, http://CRAN.R-project.org/package=lme4 (2013).

11. Fox JWSFJ. *An R companion to applied regression*. SAGE Publications (2011).

12. Kenward MG, Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* **53**, 983-997 (1997).

13. Neilson KA, *et al.* Less label, more free: approaches in label-free quantitative mass spectrometry. *Proteomics* **11**, 535-553 (2011).

14. Lu DH, *et al.* IPEP: an *in silico* tool to examine proteolytic peptides for mass spectrometry. *Bioinformatics* **24**, 2801-2802 (2008).