Supplementary material for

Heat stress responsive genes are not affected by ocean warming: long-term environmental monitoring and acute thermal stress experiments identify non-overlapping sets of differentially expressed genes in a marine fish

Marie V. Brasseur[1,2*], Christina Bakowski[3], Mark Christie[4,5], Nico Fuhrmann[1], Anja Locker[1], Lars Podsiadlowski[6], Joachim Sturve[7], Stefan Prost[8,9], Christoph Mayer[6], Henrik Krehenwinkel[1]

Affiliations

[1]Department of Biogeography, Trier University, Universitaetsring 15, 54296 Trier, Germany

[2]Ecotoxicology, University of Duisburg-Essen, Universitaetsstrasse 5, 45141 Essen, Germany

[3]Marine Evolutionary Ecology, GEOMAR Helmholtz Centre for Ocean Research Kiel, Wischhofstr. 1-3 24148 Kiel, Germany

[4]Department of Biological Sciences, Purdue University; 915 W. State St., West Lafayette, Indiana 47907-2054, USA

[5]Department of Forestry and Natural Resources, Purdue University; 715 W. State St., West Lafayette, Indiana 47907-2054, USA

[6] Leibniz Institute for the Analysis of Biodiversity Change, Adenauerallee 127, 53113 Bonn, Germany

[7] Department of Biological and Environmental Sciences, University of Gothenburg, Medicinaregatan 7B, 413 90 Gothenburg, Sweden

[8] Ecology and Genetics Research Unit, University of Oulu, Oulu 90014, Finland

[9] Central Research Laboratories, Natural History Museum Vienna, Vienna 1010, Austria

Section S1 - Acute thermal stress experiment

Evaluating the impact of a potential tank effect on the set of differentially expressed genes in response to acute heat stress

Due to the slightly unbalanced experimental design of the heat stress experiment, we explored if the expression of individuals that were kept in the same tank during the exposure period covaried. The ordination-based analysis suggested that the transcriptomes of individuals being kept in the same tank were not more similar than the transcriptomes of individuals kept in different tanks (Fig. S1A). This was further supported by hierarchical clustering analysis based on Manhattan distances between samples, which were driven by treatment but not tank affiliation (Fig. S1B).

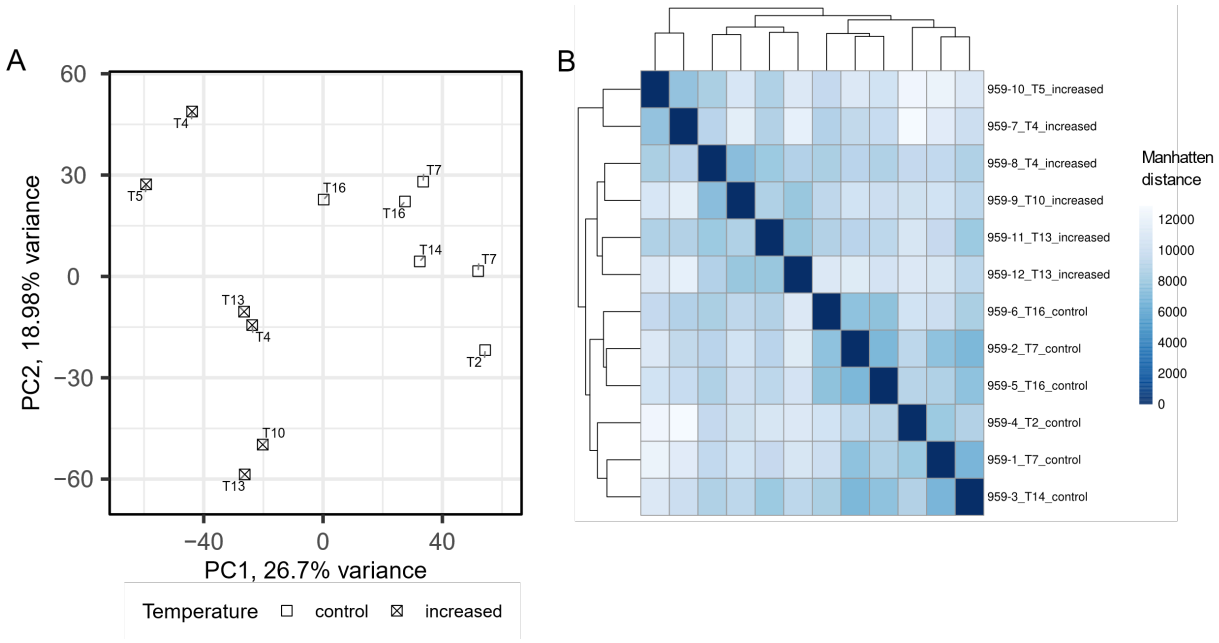

Fig S1: Multivariate analysis of the expression data set obtained from the exposure experiment to infer a potential tank effect. A: Biplot of the first two principal component axes of individuals exposed to acute heat stress. The labels indicate the tank affiliation of the sequenced individual. B: Hierarchical clustering of samples based on Manhattan distances. The sample ID ('959-X') is followed by the tank ID and the temperature treatment level (control vs. increased).

Further, we estimated a relatively low intra-tank correlation of 0.17 with limma. Including tank as random effect reduced the number of differentially expressed genes detected with limma from 2,845 to 1,919 due to less degrees of freedom and therefore a lower statistical power. The smaller set of differentially expressed genes was completely contained in the larger set (Fig. S2).

DESeq2 identified the largest number of differentially expressed genes i.e. 2,441 (Fig. S3), from which 1,439 (59%) were also detected with limma (Fig. S2). The discrepancy between the tools is related to the different modelling approaches implemented in DESeq2 and limma rather than to a potential tank effect. Since the tank affiliation appears to have only a minor impact, if any, on the variance in gene expression between samples, compared to the inter-individual variation, we argue that it is valid to treat the samples as independent biological replicates. To avoid loss in statistical power and since we aimed for an exploratory analysis, we decided to include only temperature as fixed effect in the model design and to focus on the largest set of differentially expressed genes, obtained from DESeq2, for the subsequent analyses.

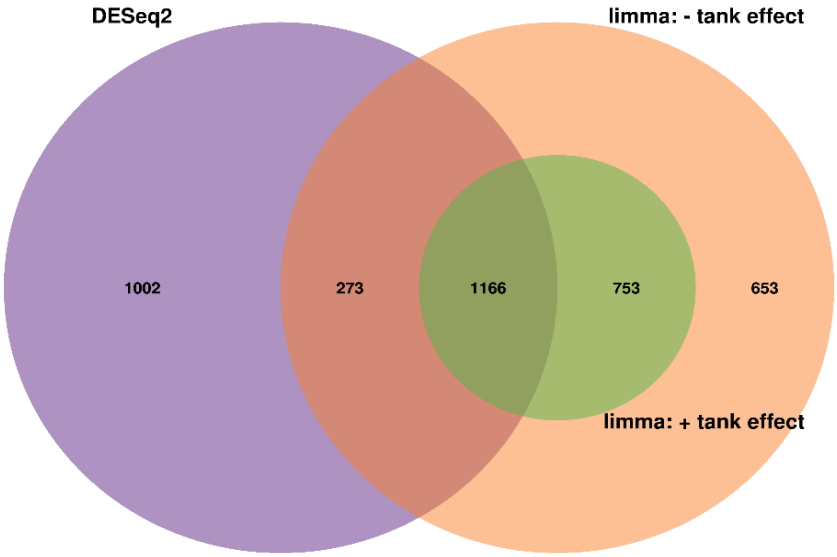

Fig. S2: Sets of differentially expressed genes detected with the GLM implemented in DESeq2 and the LM implemented in limma with (+) and without (-) the inclusion of tank as random effect.

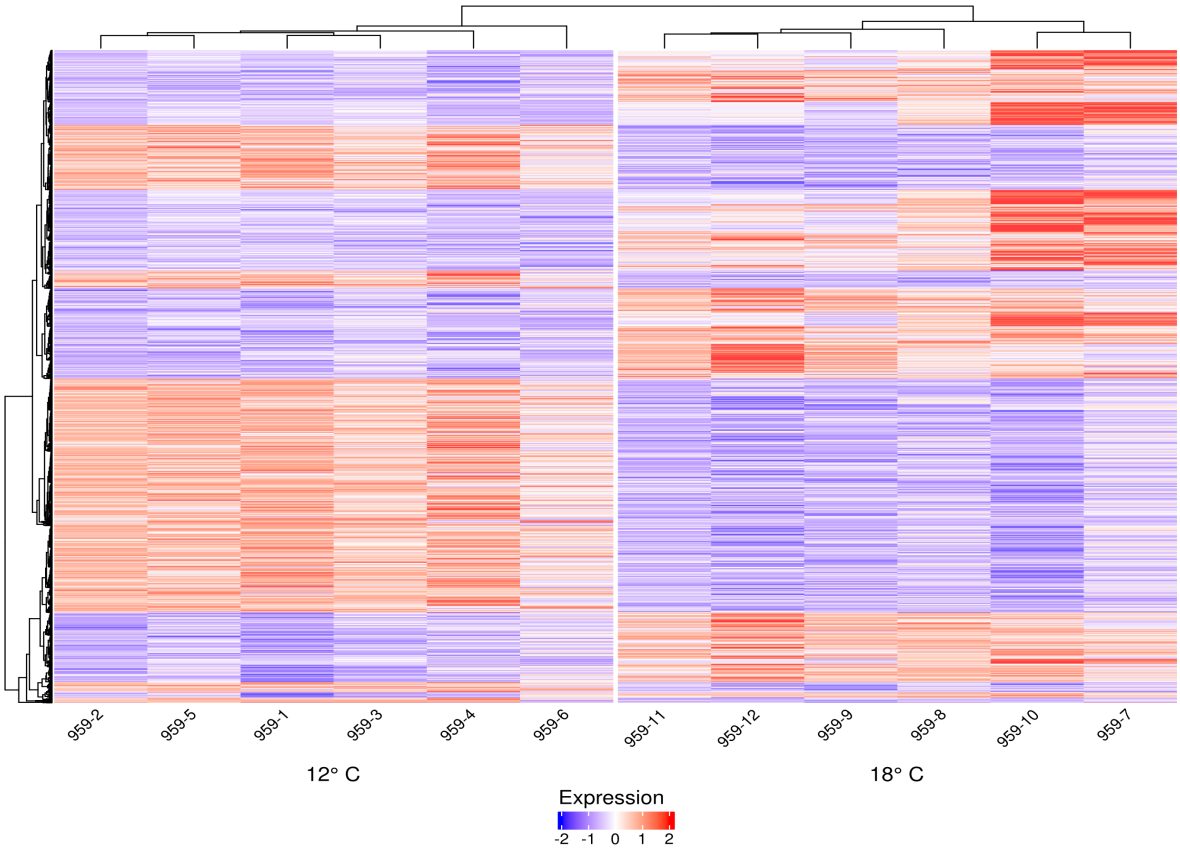

Fig. S3: Heatmap of differentially expressed genes detected with DESeq2. Expression is plotted as z-score, representing scaled and centered gene counts.

Functional enrichment of GO terms annotated to genes differentially expressed due to acute heat stress

The set of differentially expressed genes in response to acute heat stress in individuals was tested for overrepresentation of biological process (BP), molecular function (MF) and cellular compartment (CC) terms. Upregulated and downregulated genes were separately tested against all genes that were tested for differential expression in the heat stress experiment data set.

Section S2 – Environmental Specimen Bank time series expression data

Expression profiles of heat stress responsive genes over time

We detected in total 34 clusters with a minimum cluster size of 20 genes. These 34 clusters contained in total 1,500 of the 2,441 genes being differentially expressed in response to heat stress. The remaining 941 genes were not assigned to clusters containing at least 20 genes.

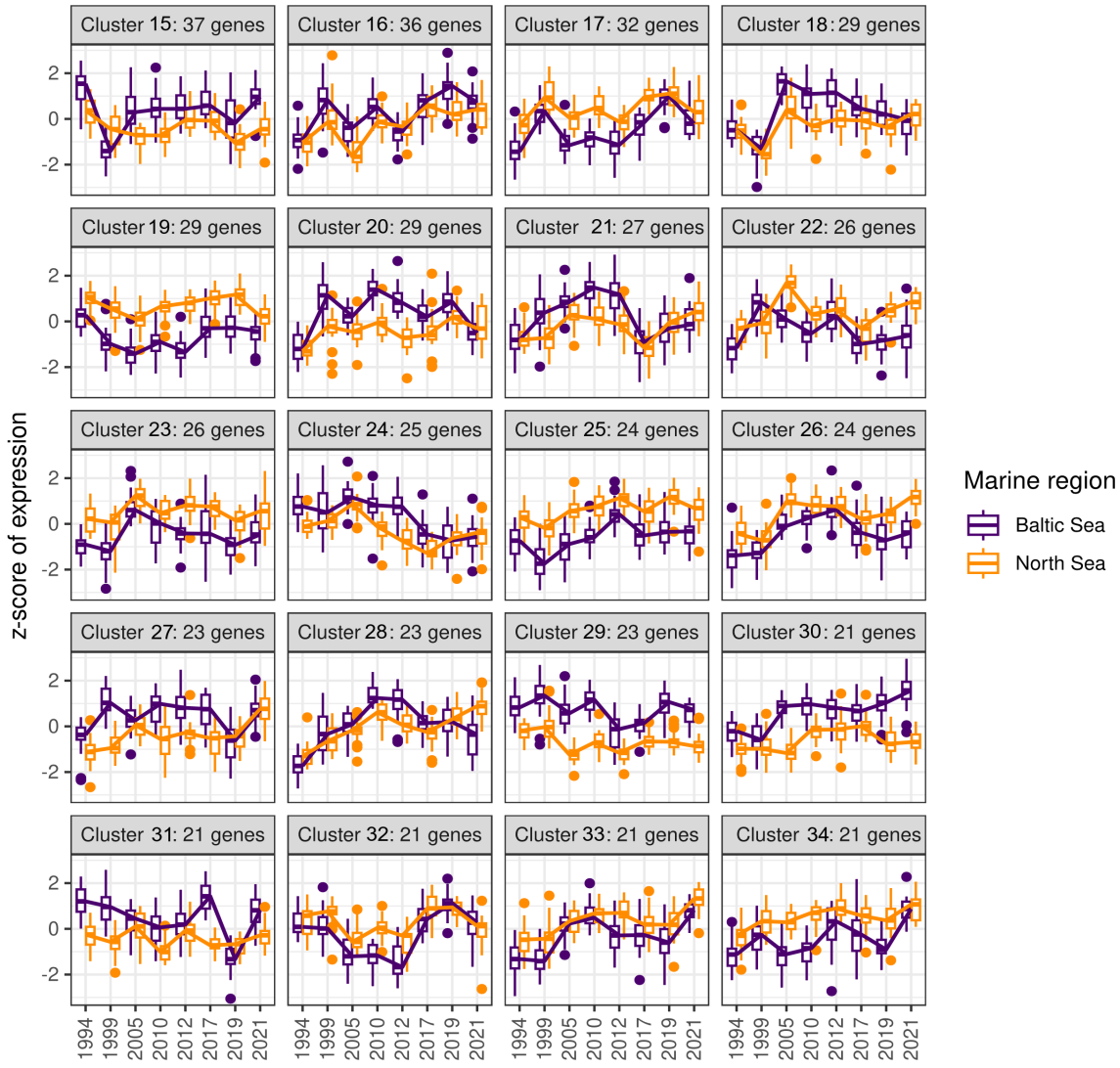

Fig. S4: Changes in expression profiles of heat-stress responsive genes over time in North Sea and Baltic Sea populations. The z-score represents the scaled and centered gene counts, i.e., positive and negative z-scores indicate gene expression above and below the average across samples, respectively. The lines connect the median expression values. Clusters with more than 20 but less than 40 genes are shown.

Sequencing batch effect in the time series data set

We observed a sequencing batch effect in the PCA based on the raw gene count data obtained from the RNA pools of the German Environmental Specimen Bank samples. The samples cluster according to their sequencing batch along the first PC, which accounts for 37% of the total variance in the data set (Fig. S5).

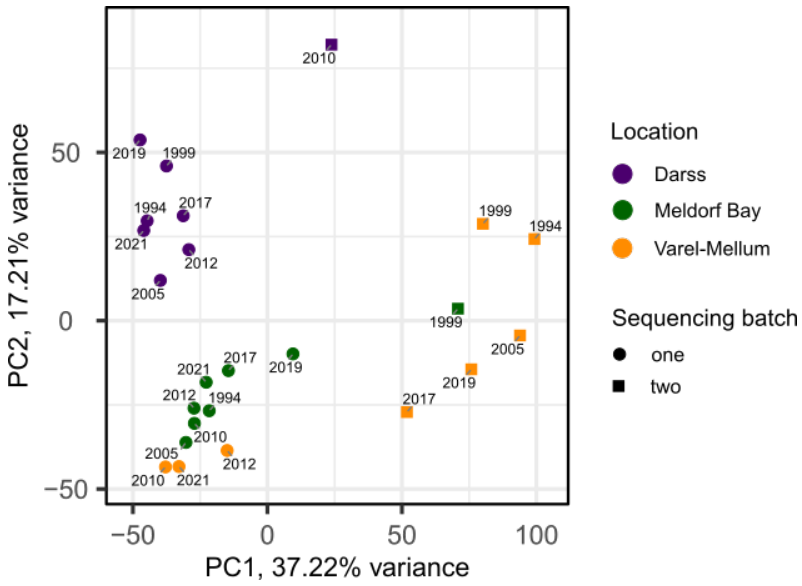

Fig. S5: Biplot of the first two principal component axes of unadjusted counts estimated from the pooled RNA-seq libraries of the German Environmental Specimen Bank time series data set.

Expression profiles of time responsive genes over time

Assuming that time might serve as a proxy for different types of environmental change, including increasing ocean temperature, we used the likelihood ratio test to identify differentially expressed genes in *Z. viviparus* populations over time (Fig. S6). Clustering of these genes according to their temporal expression profiles resulted in 10 distinct clusters which contained 339 of the time responsive genes. The remaining genes were not assigned to clusters with a minimum size of 20 genes. While some of these expression profiles revealed an approximately linear change in expression over time (i.e., clusters 4, 8), others showed strong transcriptional differences in particular years as e.g., a strong decrease in expression after 1999 (clusters 6,7).

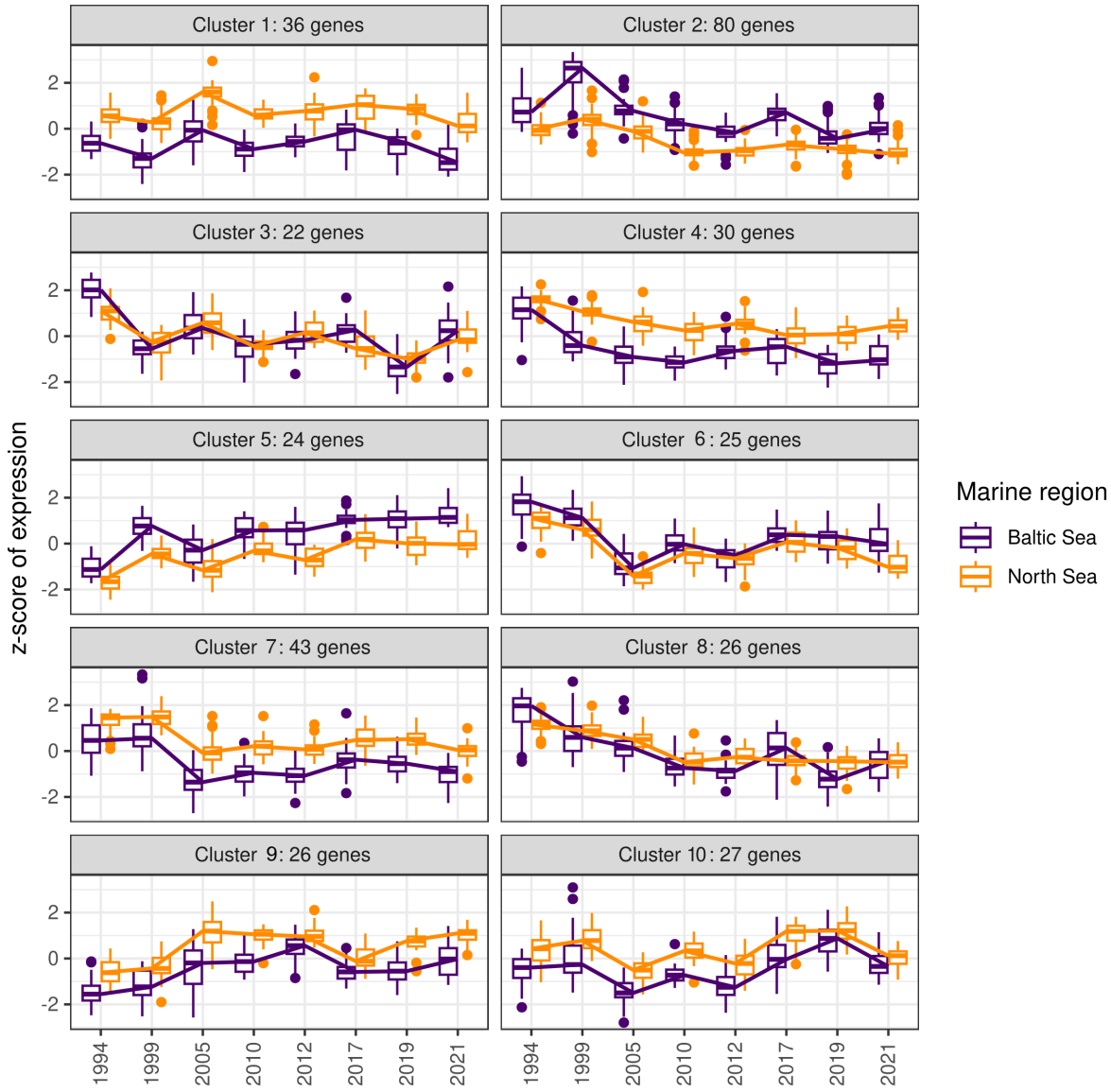

Fig. S6: Expression profiles of genes that changed their expression profile over time in North Sea and Baltic Sea populations. The z-score represents the scaled and centered gene counts, i.e., positive and negative z-scores indicate gene expression above and below the average across samples, respectively. The lines connect the average expression values. Only clusters containing at least 20 genes are shown.

Weighted gene co-expression network analysis (WGCNA)

The WGCNA modules *blue* and *turquoise* contained by far the largest number of genes of the in total 23 detected modules (Fig. S7). Their Eigengene expression profiles (Fig. S8) suggest that these gene differ in their expression between eelpout populations from the Northern and the Baltic Sea. Likewise, these populations are naturally exposed to different temperature conditions (Fig. S9) and also differ in their biometric traits (Fig. S10), which might lead to the significant correlations between the Eigengene expression profiles and the tested external traits (Fig. S11).

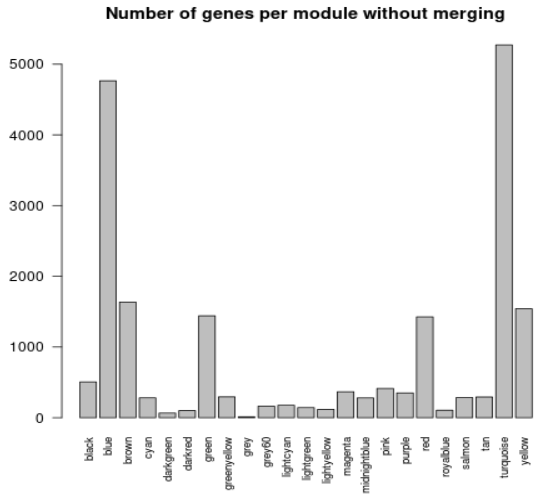

Fig. S7: Number of genes assigned to each co-expression module identified by the WGCNA.

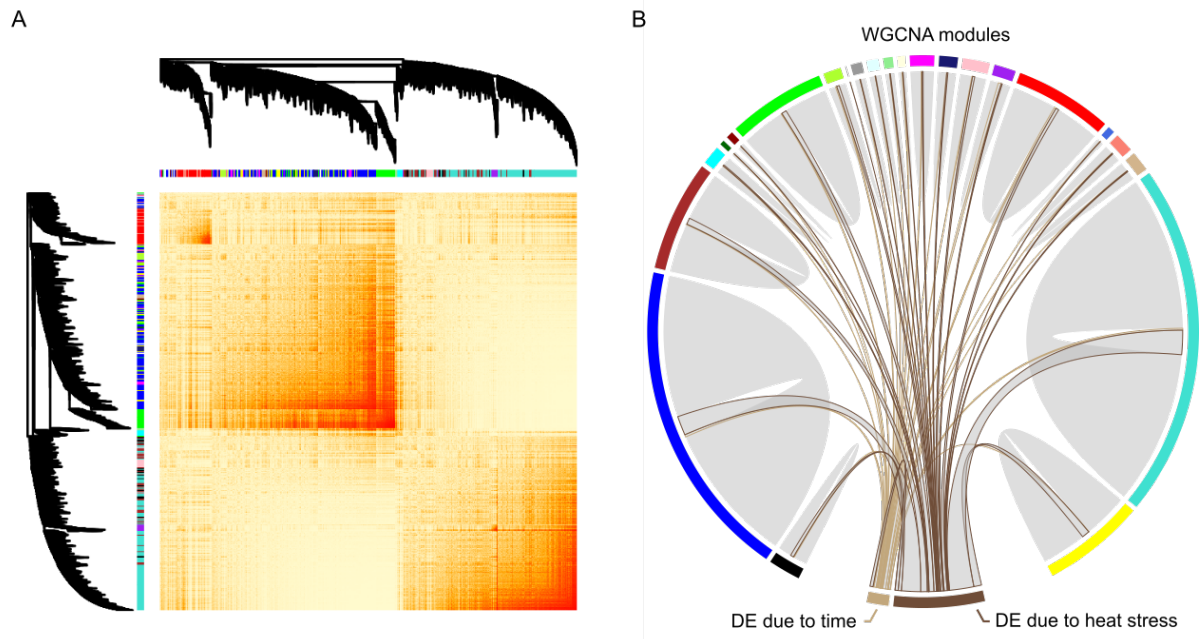

Fig. S8: (A) Co-expression network heatmap of *Z. viviparus* population time series transcriptomes. Each row and column represent a gene, the color indicate the strength of correlation in expression. (B) The identities of genes differentially expressed (DE) due to heat stress (dark brown arc) or time (light brown arc) in the network modules. All 23 modules identified during the WGCNA are shown. The arcs are colored according to the color name assigned by the WGCNA package.

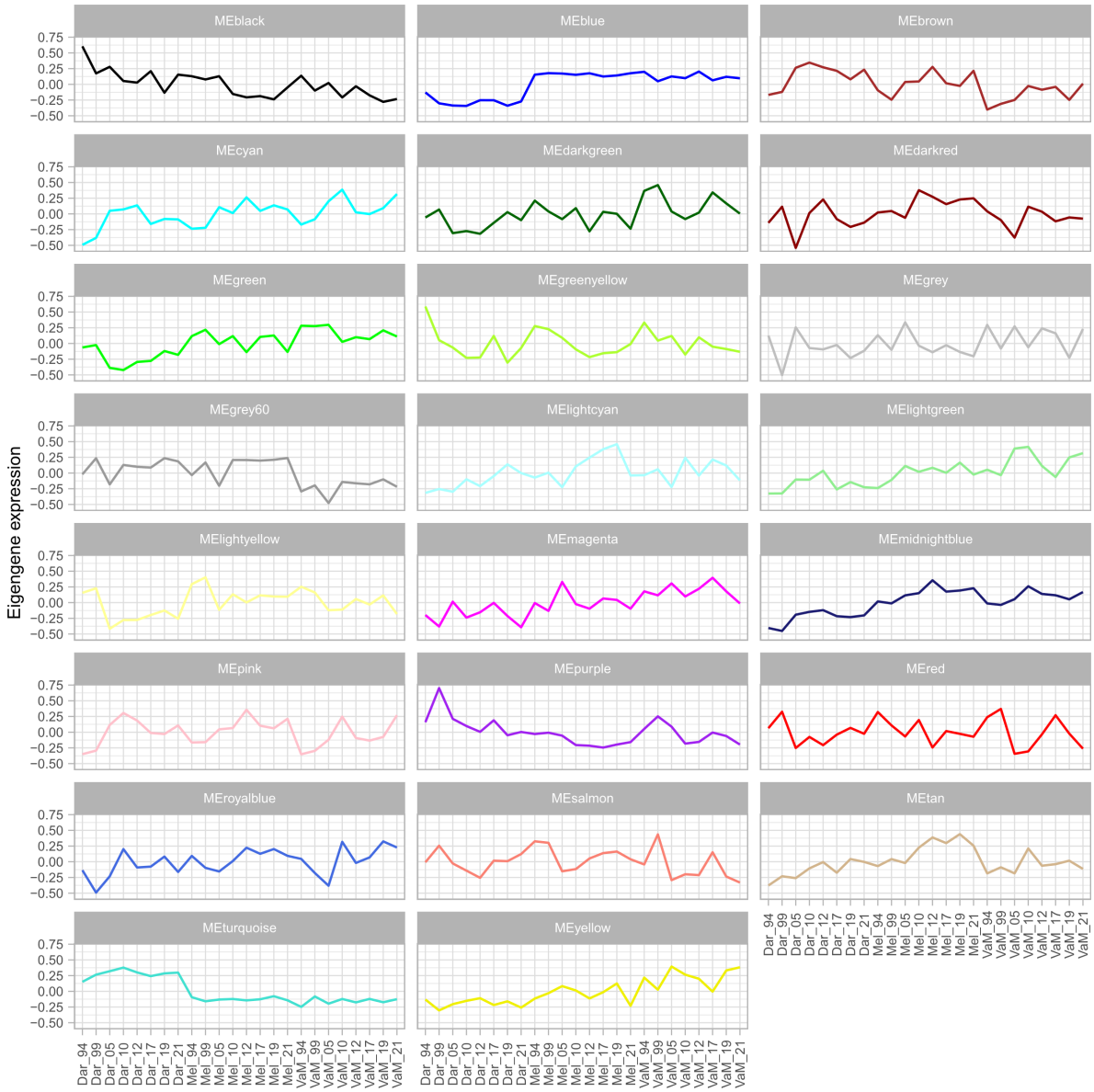

Fig. S9: Eigengene expression profiles of all identified modules. Module Eigengenes are the first principal component of the respective module and are representative for the module's expression pattern.

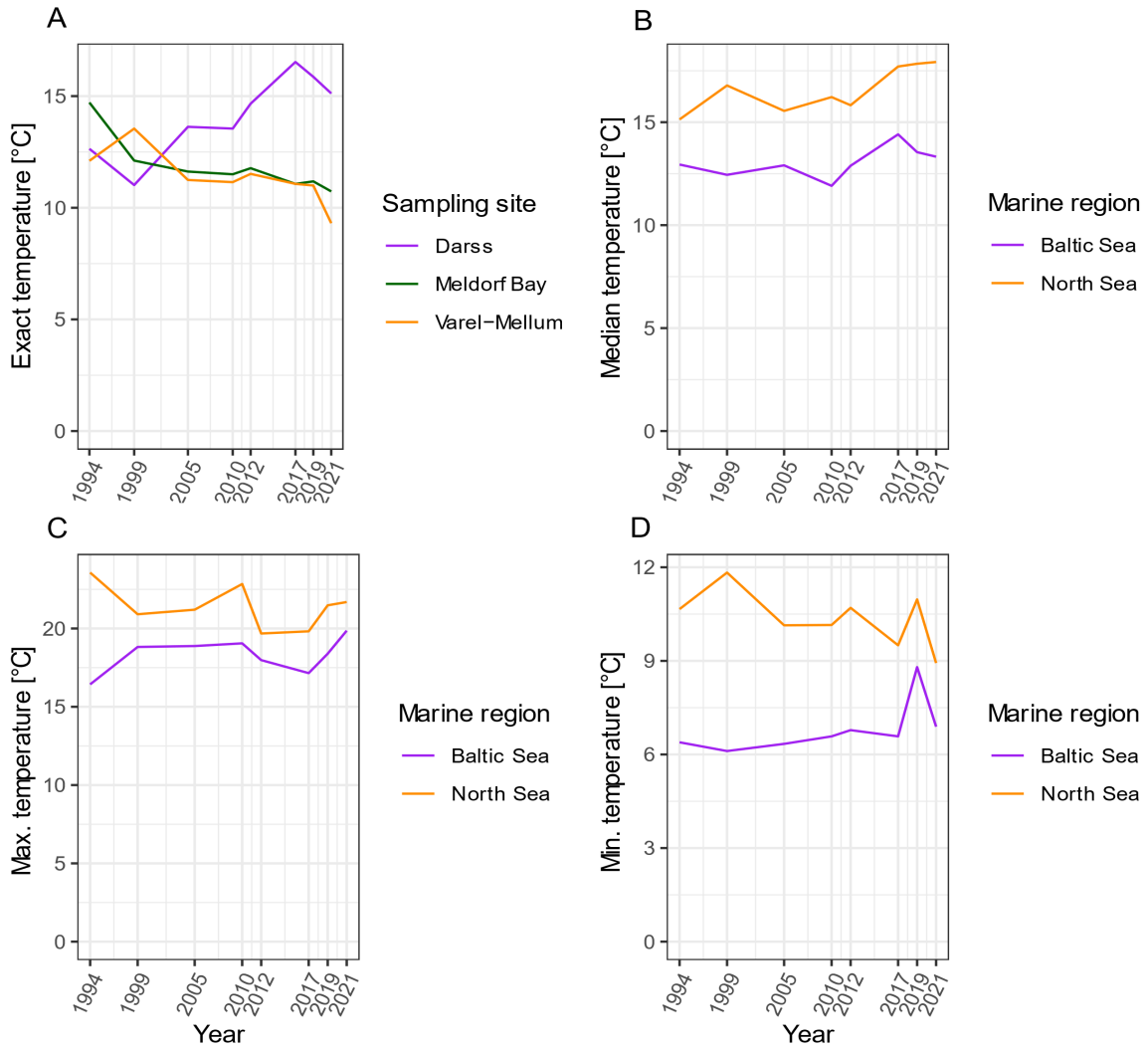

Fig. S10: Temperature profiles in the Northern Sea and the Baltic Sea across the years included in the transcriptomic time series data set. The ocean temperature was extracted from ocean physics data sets predicted by NEMO. Due to the proximity of the two North Sea sampling sites, the temperature profile for these populations was extracted from the same coordinates. Exact temperature (A) either refers to the daily average ocean temperature or, if sampling was conducted over more than one day, to the average temperature of the sampling period. Median (B), maximum (C) and minimum (D) ocean temperature refer to a fixed period i.e., from 1st of May until 31st of July, which corresponds to the eelpout sampling period of German ESB.

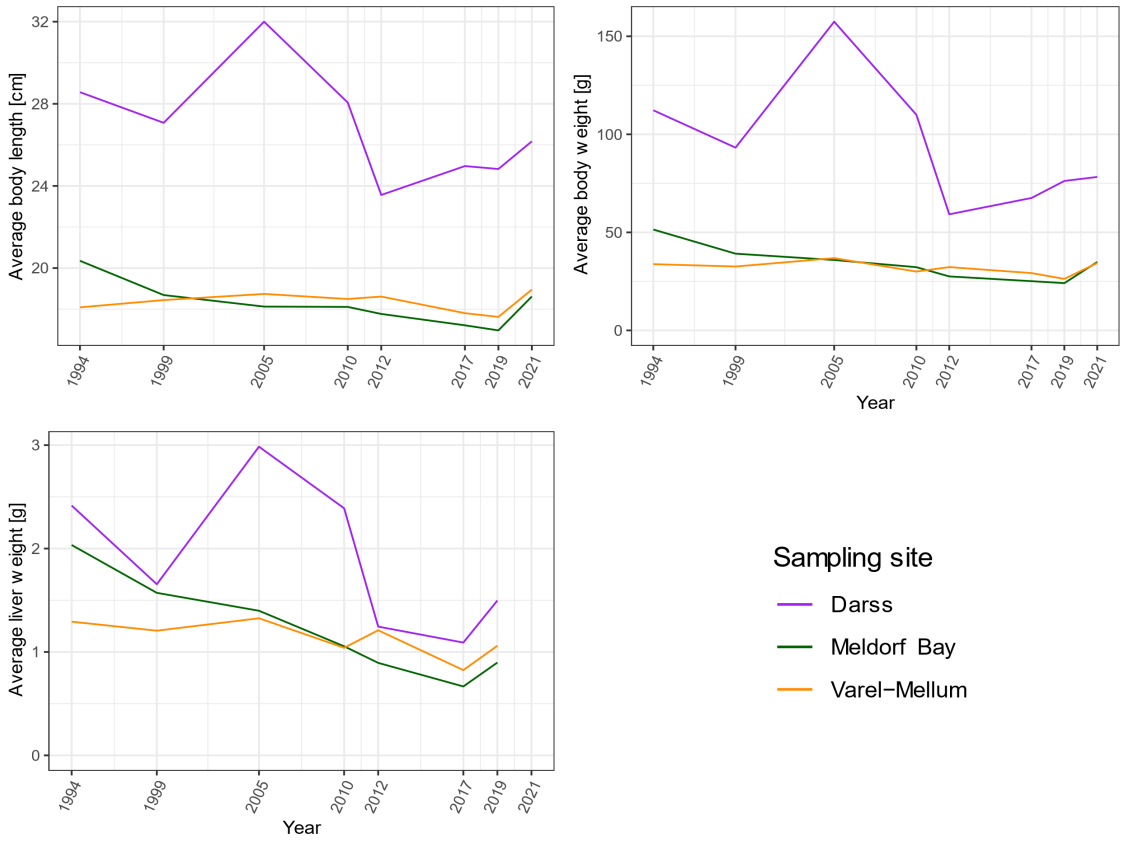

Fig. S11: Biometric trait data of eelpout populations sampled by the German Environmental Specimen Bank.

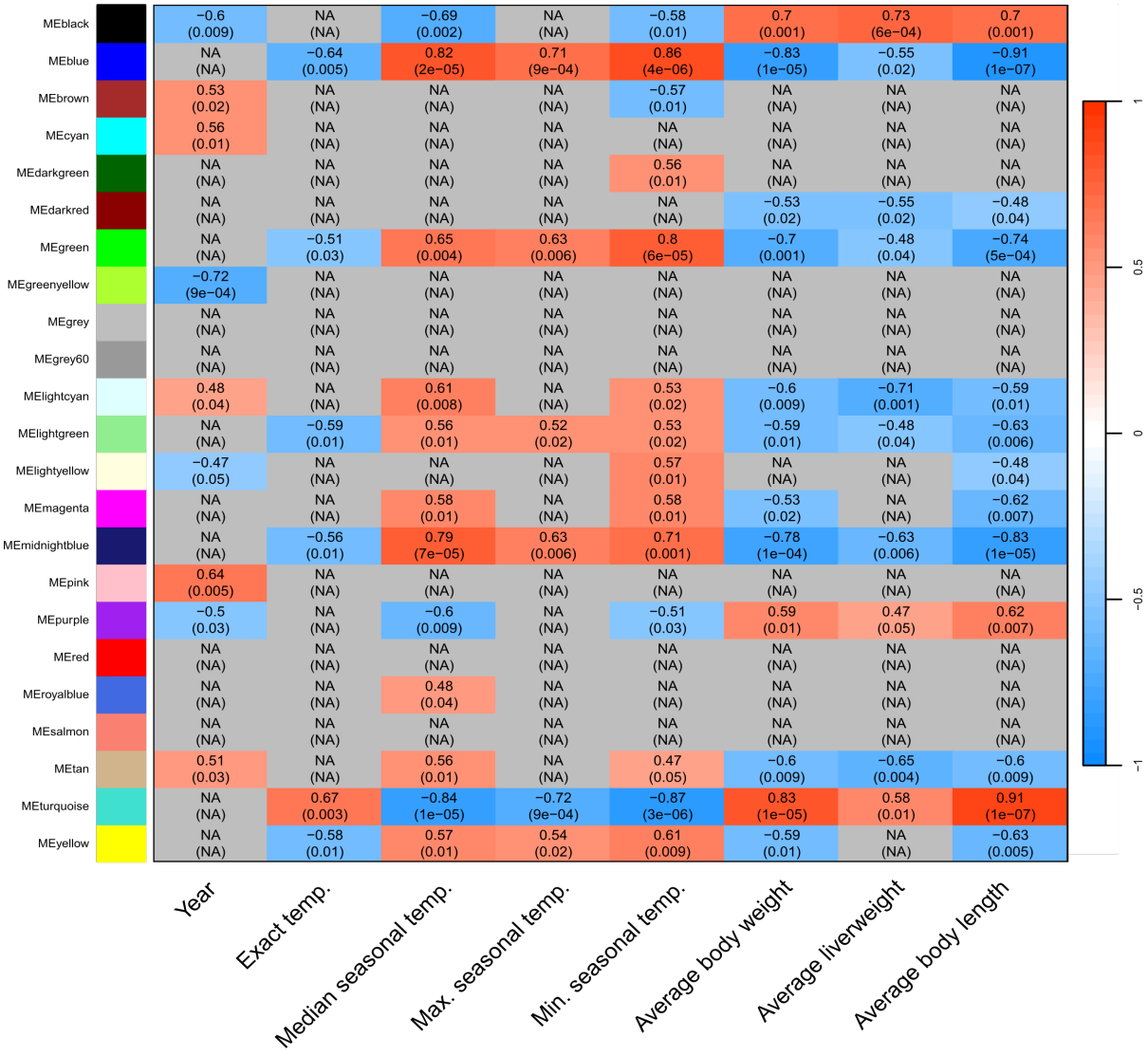

Fig. S12: Pearson correlation between Eigengene expression profiles and the external traits. Only correlation coefficients with an adjusted p-value < 0.05 are shown, otherwise NA is indicated.

Functional enrichment of GO terms annotated to time responsive genes, genes differentially expressed between North Sea and Baltic Sea populations and the main WGCNA modules 'turquoise' and *'blue'*

The set of DEGs over time in populations was tested for overrepresentation of BP and MF terms against all genes that were tested for differential expression in the time series data set.

The set of DEGs between populations of the North and Baltic Sea was tested for overrepresentation of BP and MF terms. Upregulated and downregulated genes were separately tested against all genes that were tested for differential expression in the time series data set.

Genes present in the two main WGCNA modules *turquoise* or *blue* were tested separately for overrepresentation of BP and MF terms against all genes that were included in the WGCNA.

Section S3 Visualization

The R package ggplot2 v3.4.1 [1] was used to generate the Volcano plot of the gene expression data set obtained from the heat stress experiment and to visualize the individual profiles of gene expression clusters obtained from DEGreport as well as Eigengene expression profiles and external trait data. The R package ComplexHeatmap v2.14.0 [2] was used to plot the z-score of gene expression of genes being differentially expressed due to acute heat stress across individual replicates. The Venn diagram comparing the set of DEGs detected with DESeq2 and limma (with and without including tank as random effect) was created with the R package VennDiagram v1.7.3 [3]. To determine if heat or time responsive genes correspond to a WGCNA module, a Chord diagram was created with the R package circlize v0.4.15 (Gu 2014).

Bibliography

[1] Wickham H (2016) ggplot2: elegant graphics for data analysis. Springer-Verlag, New York

[2] Gu Z, Eils R, Schlesner M (2016) Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics 32(18):2847–2849

[3] Chen, H., Boutros, P.C. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics* 12, 35 (2011). https://doi.org/10.1186/1471-2105-12-35

[4] Gu Z, Gu L, Eils R, Schlesner M, Brors B (2014) Circlize implements and enhances circular visualization in R. Bioinformatics 30(19):2811–2812