Differential Expression in *Hematodinium sp.*

Aidan Coyle

2021-11-09

```
library("kableExtra")
library("tidyverse")
library("knitr")
library("magick")
```

Methods

Collection

Between October TK and TK, 2017, TK male *C. bairdi* were collected from Stephen's Passage in southeastern Alaska by the Alaska Department of Fish and Game (ADF&G). This site was selected due to its reliably high rate of *Hematodinium sp.* infection - approximately 50% (ADF&G, unpub. data). Crabs were caught using pots (TK: was this part of an ADFG survey?). Once collected, crabs were transported to the (TK: NOAA facility) within TK hours, and placed in flow-through seawater tanks (TK: assuming that's what was used) at 7.5°C - the benthic water temperature within Stephen's Passage at time of capture.

Verification of Infection

To verify crabs were infected by *Hematodinium sp.*, a TK ml aliquot of hemolymph was drawn from each crab and preserved in 800 µl 95% ethanol. 200 µl of that sample was centrifuged and pelleted. The pellet was then air-dried, and DNA was extracted with invertebrate lysis buffer. Extraction protocol followed a modified version of Ivanova et al.(2006), in which two washes with Wash Buffer were performed and eluted DNA (50 µl) was adjusted to 10 mM Tric-Cl, pH 8.0, and 0.1 mM EDTA. Once extracted, DNA was subjected to two rounds of cPCR with two primer pairs - Univ-F-15 / Hemat-R-1654 (Gruebl et al. 2002), and Hemat 18Sf / Hemat 18Sr (Bower et al. 2004). Both primer pairs are designed for the *Hematodinium spp.* small subunit (SSU) rRNA gene. After PCR, reaction aliquots were pooled and visualized on ethidium bromide-stained 2% agarose gels. If both *Hematodinium spp.* bands were visible on the gel, samples were scored as positive. If neither band was visible, samples were scored as negative. If one band amplified, samples were scored as indeterminate and (TK not selected for further analysis?)

Temperature Variation

All crabs were acclimated to 7.5°C for nine days. 60 (TK) crabs were then selected based off cPCR results for temperature treatments and apparent recovery from capture stress. 0.2 ml of hemolymph was drawn from each selected crab and preserved in 1200 µl RNAlater. Crabs were divided randomly among six replicate tanks, with 10 crabs per tank. In three of the tanks, the water temperature was gradually raised to 10°C over a two-day period. In the three other tanks, water temperature was held at 7.5°C. At the end of the two-day acclimation period, another 0.2 ml hemolymph sample was taken from each crab and preserved in RNAlater. Following the conclusion of the experiment - 17 days after the acclimation period began - 3

Table 1: Individual libraries of infected crab

Crab ID	Treatment group	Day 0 sample ID	Day 2 sample ID	Day 17 sample ID
A	Ambient	178	359	463
B	Ambient	118	349	481
C	Ambient	132	334	485
G	Elevated	173	272	NA
H	Elevated	72	294	NA
I	Elevated	127	280	NA

samples of hemolymph were taken from each surviving crab in the ambient temperature treatment group, and 6 samples were taken from each surviving crab in the elevated temperature treatment group. The increased number of per-crab hemolymph samples in the elevated temperature treatment group was due to a mass mortality event between days 2 and 17 within that treatment group.

TK: edit methods to specify examining elev vs amb

RNA Extraction and Sequencing

TK hemolymph samples were centrifuged at 14000 g for 10 minutes. RNA was then extracted from the pellet (TK: did it pellet?) with Quick DNA/RNA Microprep Plus Kit (Zymo Research) using the manufacturer's protocol. To quantify RNA, 2 µl samples were then run on Qubit 3.0 with the Qubit RNA HS Kit (Invitrogen). Based on RNA yield and infection severity as determined by qPCR, six crabs were selected - three from the elevated-temperature treatment group and three from the ambient-temperature treatment group. All samples from these six crabs (Table 1) were sent to the Northwest Genomics Center at Foege Hall at the University of Washington for library construction and sequencing. Due to the mass mortality event in the elevated-temperature treatment groups, no libraries from Day 17 were available for these crabs.

Transcriptome Assembly and Annotation

TK. Grace's chapter describes Transcriptome 3.1. Looks similar except 3.1 has a filter - should check with Sam if Transcriptome 2.0 construction method matches. Also look at Sam's notebook posts.

Differential Expression Analysis

An index of TK transcriptome was created with kallisto (TK citation), and each library was pseudoaligned to obtain counts. An abundance matrix for each pairwise comparison (Table 2) was then created using Trinity (v2.TK, TK citation?). Differential contig expression was calculated using a negative binomial GLM [TK: check if correct] using the R package DESeq2. Read counts were normalized using size factors and fit to a negative binomial distribution. The Wald test for significance of GLM terms for each comparison was used to obtain unadjusted p-values. For comparisons of crabs at different temperatures, a table of significantly differentially-expressed transcripts (Benjamini-Hochberg adjusted $p < 0.05$) was also obtained (Table 2).

TK title

Talk here about how lots of genes from the unfiltered transcriptome didn't match closely to either the host genome or parasite genes? Also did we make sure there was no overlap in transcripts between the two (i.e. none that were assigned to both the host and parasite transcriptome)?

Table 2: PairwiseComparisons

Pair 1			Pair 2			DEGs analyzed individually?
Crab IDs	Temp. when sampled	Sample day	Crab IDs	Temp. when sampled	Sample day	
A,B,C	Ambient	0	A,B,C	Ambient	2	No
A,B,C	Ambient	0	A,B,C	Ambient	17	No
A,B,C	Ambient	2	A,B,C	Ambient	17	No
A,B,C	Ambient	2	G,H,I	Elevated	2	Yes
G,H,I	Ambient	0	G,H,I	Elevated	2	Yes

Enrichment Analysis

For each comparison (Table 2), the output from DESeq2 was cross-referenced with the annotated transcriptome and the UniProt database (citation TK) to produce a table of UniProt Accession IDs and GO terms, along with a table of UniProt Accession IDs and unadjusted p-values. GO categories were then tested for significant enrichment with the R package GO-MWU (citation TK), which utilizes the Mann-Whitney U test.

Individual DEG Examination

Process TK, haven't completed this yet

Results

DESeq2

The DESeq2 package was used to examine differential expression between libraries, and to perform various pairwise comparisons between sample groups. Principal component analyses of samples taken from the elevated-temperature treatment group showed clustering by day, and thus by temperature. This was observed for libraries aligned to both the unfiltered and host-only transcriptomes. Due to low counts, a PCA could not be created for libraries aligned to the parasite-only transcriptome. No such clustering was observed for the ambient-temperature libraries, regardless of transcriptome, along this same timeframe.

[TK: CHANGE IMAGE LEGENDS FROM TEMPERATURE TO DAY. Done for hemat1.6, just rerun DESeq for other PCAs needed (after updating exp_design table)].

[TK: Should we perform GO-MWU but looking at MF or CC?]

GO-MWU

Pairwise comparisons were performed using GO-MWU to determine which biological processes were enriched. Unfiltered libraries from the elevated-temperature treatment group saw changes in expression for numerous biological processes, including TK, TK, and TK. These modules were not enriched in the ambient-temperature treatment group over the same timespan.

Time

Over the 17 days of the experiment, GO term expression changed substantially within the control group (due to the mortality event, no enrichment data is available for the experimental group). Overall, the samples

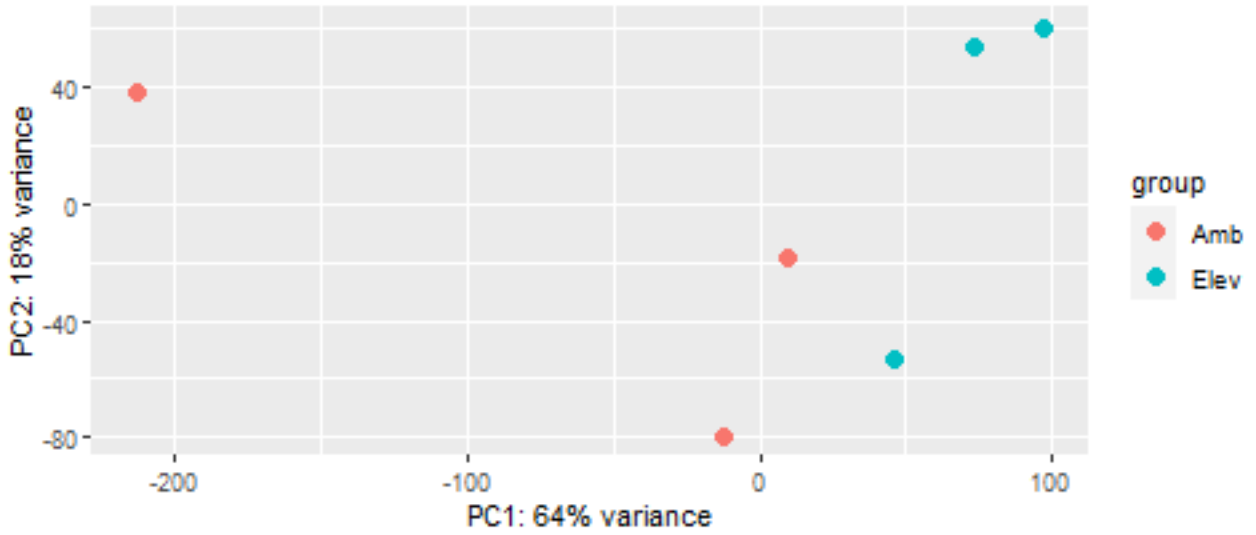

Figure 1: PCA for elevated-temperature libraries, Days 0-2 (unfiltered transcriptome)

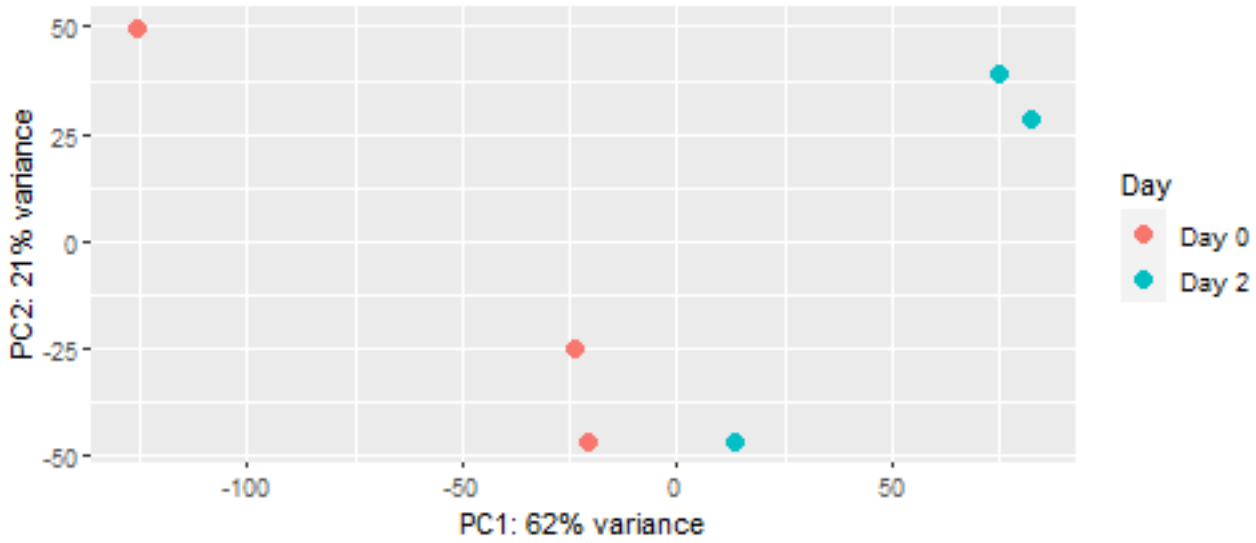

Figure 2: PCA for elevated-temperature libraries, Days 0-2 (crab transcriptome)

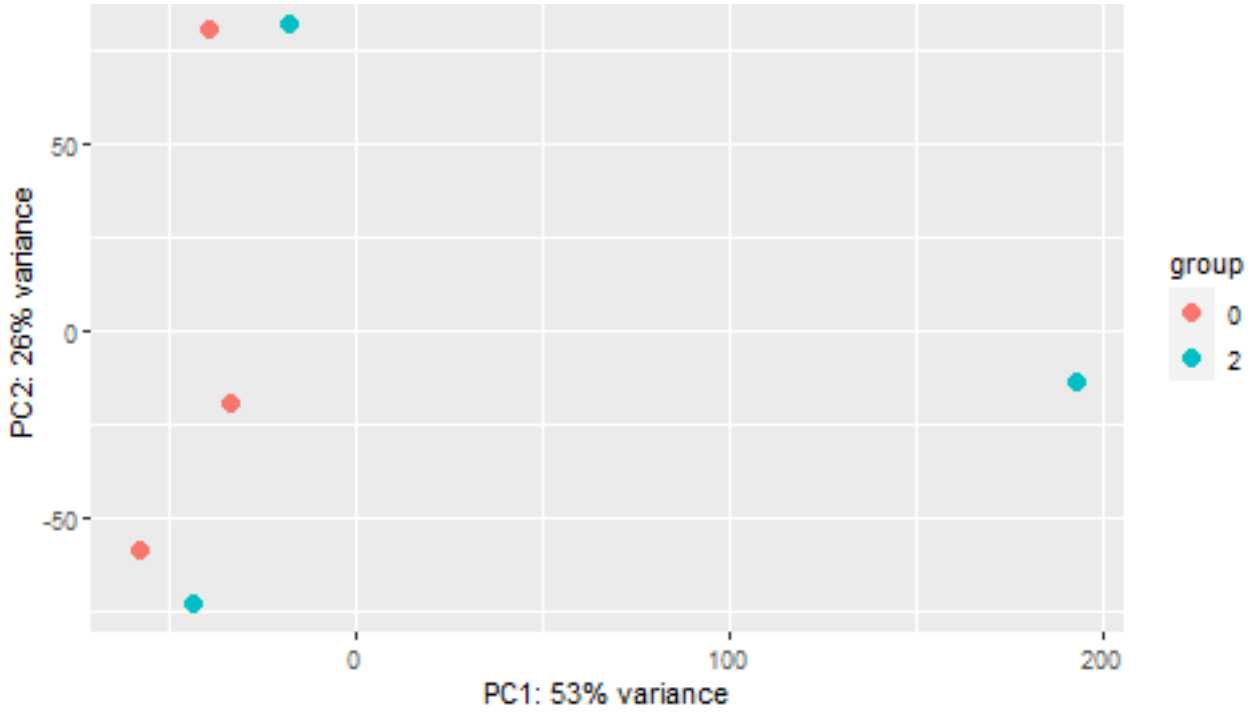

Figure 3: PCA for ambient-temperature libraries, Days 0-2 (unfiltered transcriptome)

Figure 4: PCA for ambient-temperature libraries, Days 0-2 (crab transcriptome)

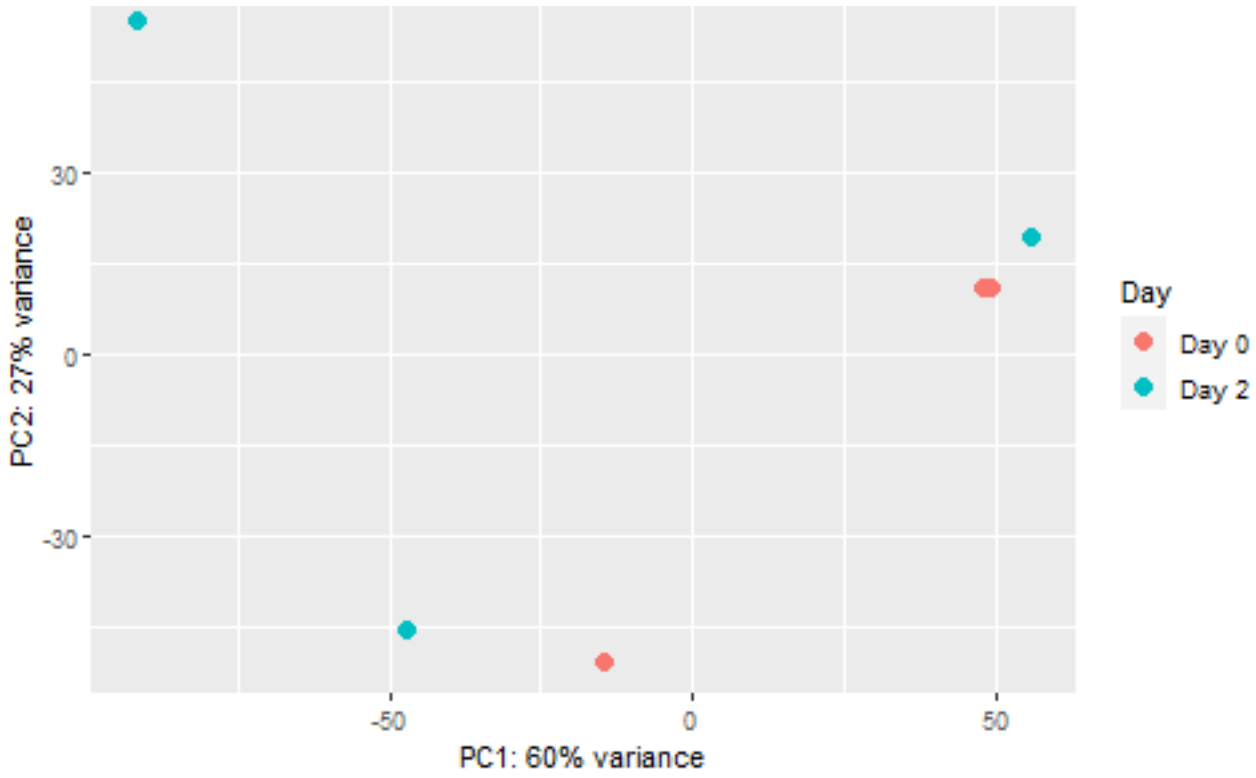

Figure 5: PCA for ambient-temperature libraries, Days 0-2 (parasite transcriptome)

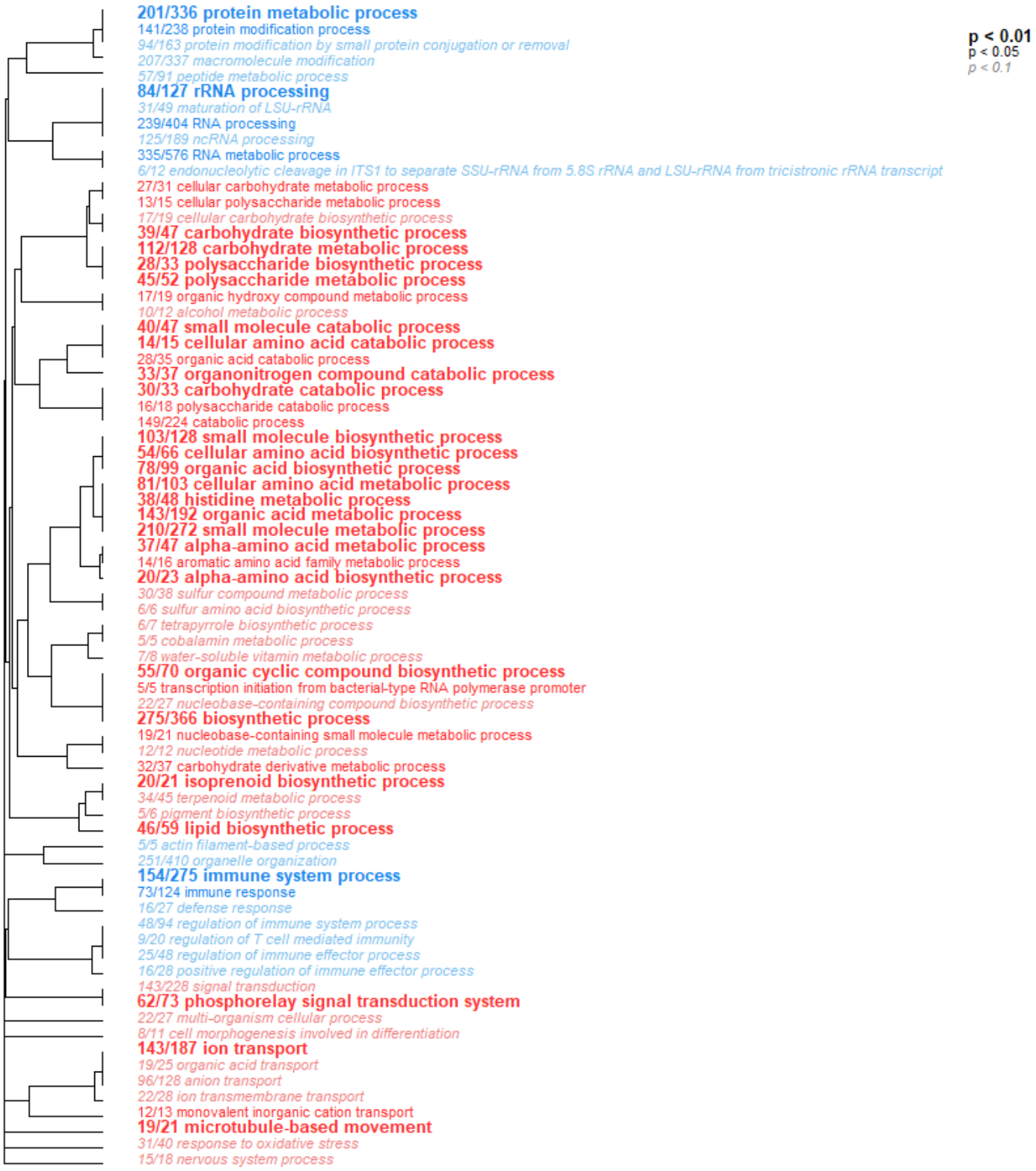

Figure 6: GO term enrichment for elevated-temperature libraries, Days 0-2 (unfiltered transcriptome)

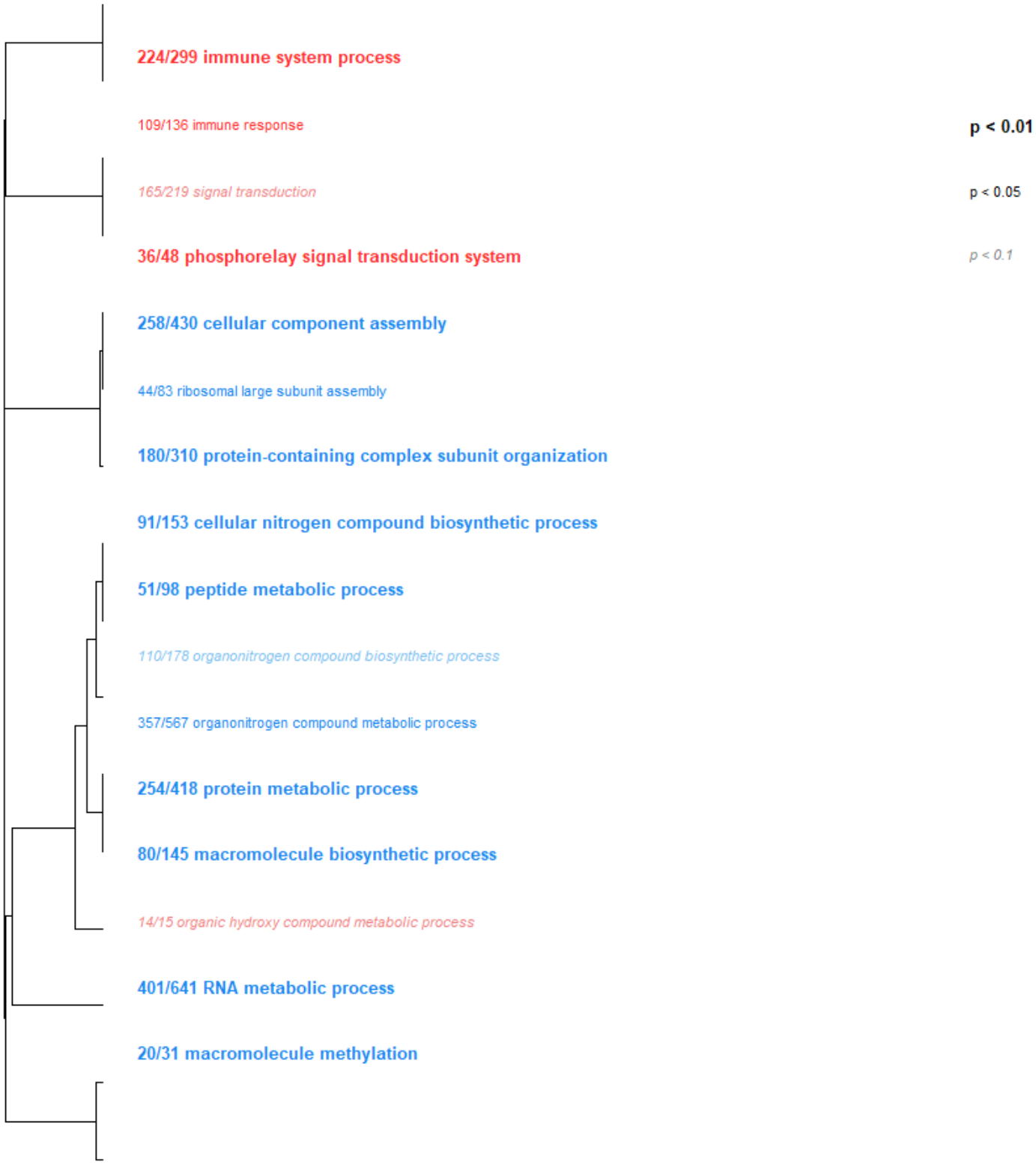

Figure 7: GO term enrichment for ambient-temperature libraries, Days 0-2 (unfiltered transcriptome)

saw changes in TK, TK, and TK. However, when examining host expression, only minor pathway expression changes were observed. This indicates TK MAYBE IMMUNOSUPPRESSIVE TALK HERE.

While no major changes were observed in the host, *Hematodinium* expression did shift substantially over the same time period. Generally, those changes were associated with TK and TK. This indicates TK DISCUSSION OF POSSIBLE CHANGES IN MORPHOLOGY

TK: TALK ABOUT HOW CRAB TRANSCRIPTION DOESN'T CHANGE MUCH BUT UNFILTERED DOES. COULD INDICATE IMMUNOSUPPRESSION, OTHER STUFF IS GETTING IN AND CHANGING. OR LIMITATION OF GO-MWU, AS SAME PROB FOR ELEV0 VS ELEV2

TK: DISCUSSION, TALK ABOUT HOW THIS COULD EITHER BE AN INDICATION OF CHANGES OVER COURSE OF INFECTION OR INDICATION OF TANK ADAPTATION. COULD LOOK AT LOWERED 0 VS 17 UNINFECTED TO GET AN IDEA, BUT ADDS TEMP AS COMPLICATING FACTOR...

Figure 8: GO term enrichment for ambient-temperature libraries, Days 0-17 (unfiltered transcriptome)

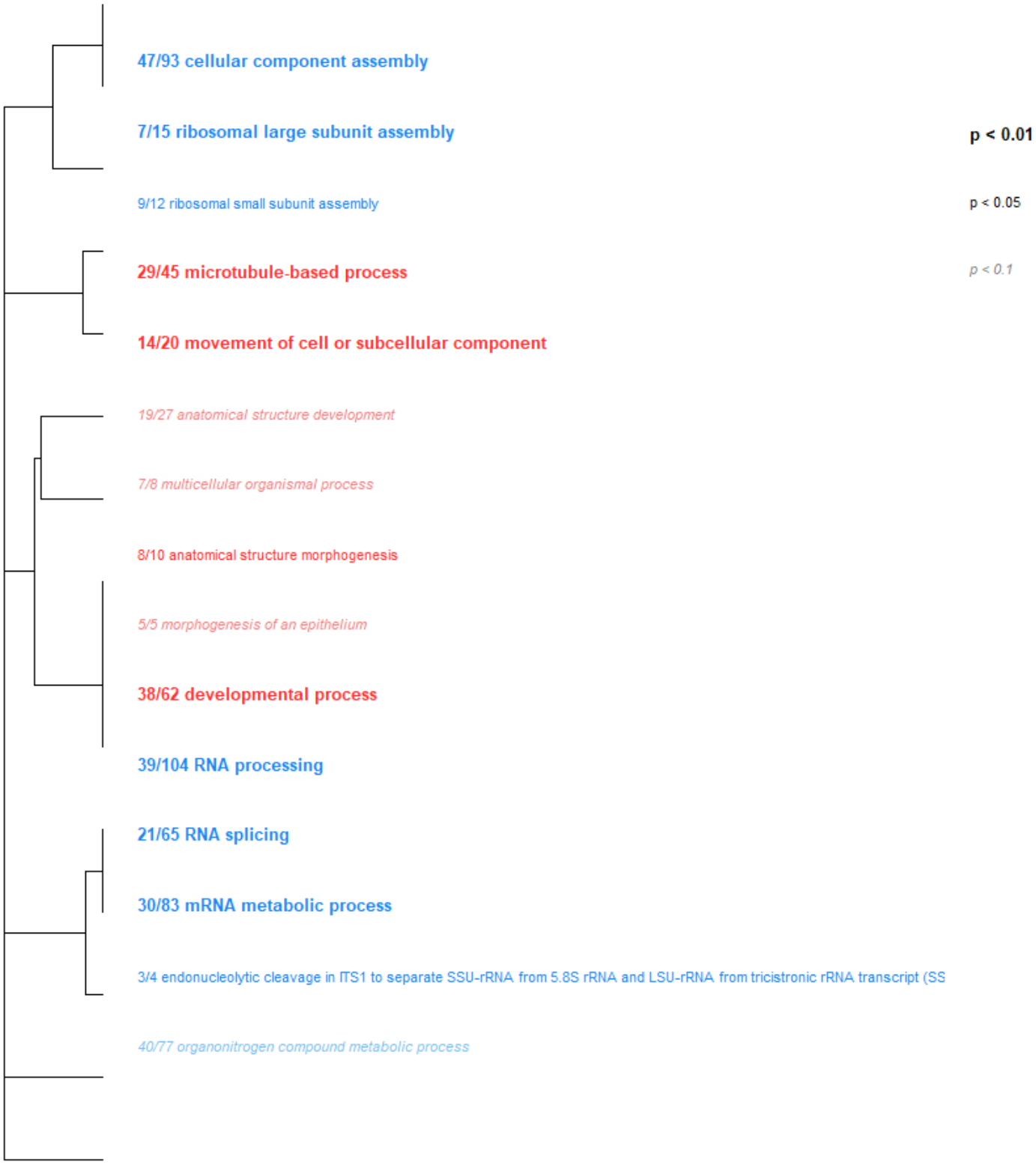

Figure 9: GO term enrichment for parasite, Days 0-17

Immune Genes

Host

Numerous genes (n = TK) within the *C. bairdi* transcriptome were associated with immune function (GO: 0006955). Many were members of the Cathepsin family, with Cathepsins C, J, L, S, U, V, and W all present. Cathepsin L was particularly broadly expressed, with seven distinct genes coding for Cathepsin and Procathepsin L [TK: does this make sense to say?]. Furthermore, Procathepsin L was differentially-expressed in the experimental group. Several types of MAPKs (mitogen-activated protein kinases) were also present within the transcriptome, including two p38 MAPKs and one one MAP4K. MAPKs are part of the IMD (immune deficiency) pathway, a notable component of the crustacean immune system. Several other genes associated with the IMD pathway were observed, including the transcription factor Relish and the kinase inhibitor IKK [TK: change 1st K to a kappa]. NFIL3, a nuclear factor which has been found to regulate Relish expression in similar systems, was also present.

Other notable immune-linked genes observed were Transcription Activator Protein-1 (TF AP-1) and Granzyme A. TF AP-1 acts as an immune system regulator within other crab species, along with a potential role as an osmoregulator. Little research on the role of Granzyme A in invertebrates has been performed, but in vertebrates it has a cytotoxic role against intracellular pathogens.

Parasite

Within the *Hematodinium sp.* transcriptome, 4 genes were linked to immune function. All four of these were cysteine proteases, which TK CP DESCRIPTION. Three of the four were cathepsins, including both Procathepsin and Cathepsin L. TK: PROBBALY TALK ABOUT ROLE OF CAT L IN PARASITES WITHIN DISCUSSION, BUT NEED A BIT MORE TO ROUND THIS SECTION OUT

TK: SOME SORT OF TABLE FOR DESCRIBING IMMUNE GENES?

Characterizing Overall Expression Patterns

Prior to filtering by taxa, samples from the lowered-temperature treatment group saw an average overall decrease in expression in 42% of transcripts, while the control group averaged a 33% decrease

Table TK: Overall expression in samples unfiltered by taxa

	Ambient	Lowered
Increase	30.8%	27.4%
Decrease	33.9%	42.9%
Neither	35.5%	29.6%

However, this same pattern was not observed when examining expression within the host or parasite specifically. Within the host, overall expression patterns were remarkably similar regardless of temperature. And within the parasite, expression increased within the lowered-temperature treatment group (TK STATISTICAL TESTS ON THESE RESULTS - CHI-SQUARE?) TK: MENTION CAVEAT OF 2 UNINFECTED IN LOWERED-TEMP GROUP, OR DO IN DISCUSSION?

Table TK: Overall host transcript expression

	Ambient	Lowered
Increase	28.9%	28.6%
Decrease	31.8%	32.2%

	Ambient	Lowered
Neither	39.3%	39.2%

Table TK: Overall parasite transcript expression

	Ambient	Lowered
Increase	32.3%	43.4%
Decrease	29.5%	29.1%
Neither	42.3%	30.1%

Characterizing Immune Gene Expression Patterns

NOTE: I really don't think our sample size of immune genes is large enough to make overall judgments on expression patterns, so this section is probably ripe for cutting. Still, I'll wait till I run a chi-square (or similar) test on this to see.

I'll also avoid writing up a more detailed analysis until I run those tests

Table TK: Immune gene expression in samples unfiltered by taxa

	Ambient	Lowered
Increase	37.1%	8.1%
Decrease	28.5%	67.2%
Neither	34.4%	24.7%

Table TK: Immune gene host transcript expression

	Ambient	Lowered
Increase	31.4%	9.7%
Decrease	28.5%	57.6%
Neither	40.0%	32.7%

Parasite expression: not available, only 5 genes total

WGCNA

A signed weighted correlation network analysis (WGCNA) was run on all libraries aligned to each transcriptome (TK citation). This clustered genes into modules according to expression pattern, and then correlated them with our variables. We took all modules that were significantly correlated, and discarded those in which the correlation to the variable appeared to be due to extremely strong correlation to a single crab. This produced the following modules (Table TK).

Transcriptome	Module	Trait and p-values
Unfiltered	black	Day (0.04)
Unfiltered	tan	Low vs. Ambient (0.05)
Unfiltered	cyan	Low vs. Ambient (0.02), Elevated vs. All (0.04)
Unfiltered	brown	Low vs. Elevated (0.02), Elevated vs. All (0.03)

Transcriptome	Module	Trait and p-values
Host	black	Low vs. Ambient (6x10^-4), Elevated (0.05)
Host	red	Low vs. Ambient (0.01), Infection Level (0.01), Carapace Width (0.04)
Host	blue	Low vs. Elevated (0.02), Elevated (0.03)
Parasite	black	Day (0.04)
Parasite	turquoise	Low vs. Ambient (0.02), Infection Level (0.01)
Parasite	blue	Infection Level (0.01)

Each of these modules was then analyzed using GO-MWU. No groups of GO terms were differentially enriched for any unfiltered or host module. However, all three modules within the parasite transcriptome saw differential enrichment. TK: Discussion of differential enrichment comes here

TK: Consider running DESeq2 on Hemat_Level H vs L

TK: Also consider running DESeq with contrasts to run three-way comparison on Amb 0 vs 2 vs 17 all in one

TK: Check whether, for the All Crabs PCAs, I should put multiple legends into the plot

TK: When we decide when/if to use WGCNA heatmaps, expand em to make em prettier

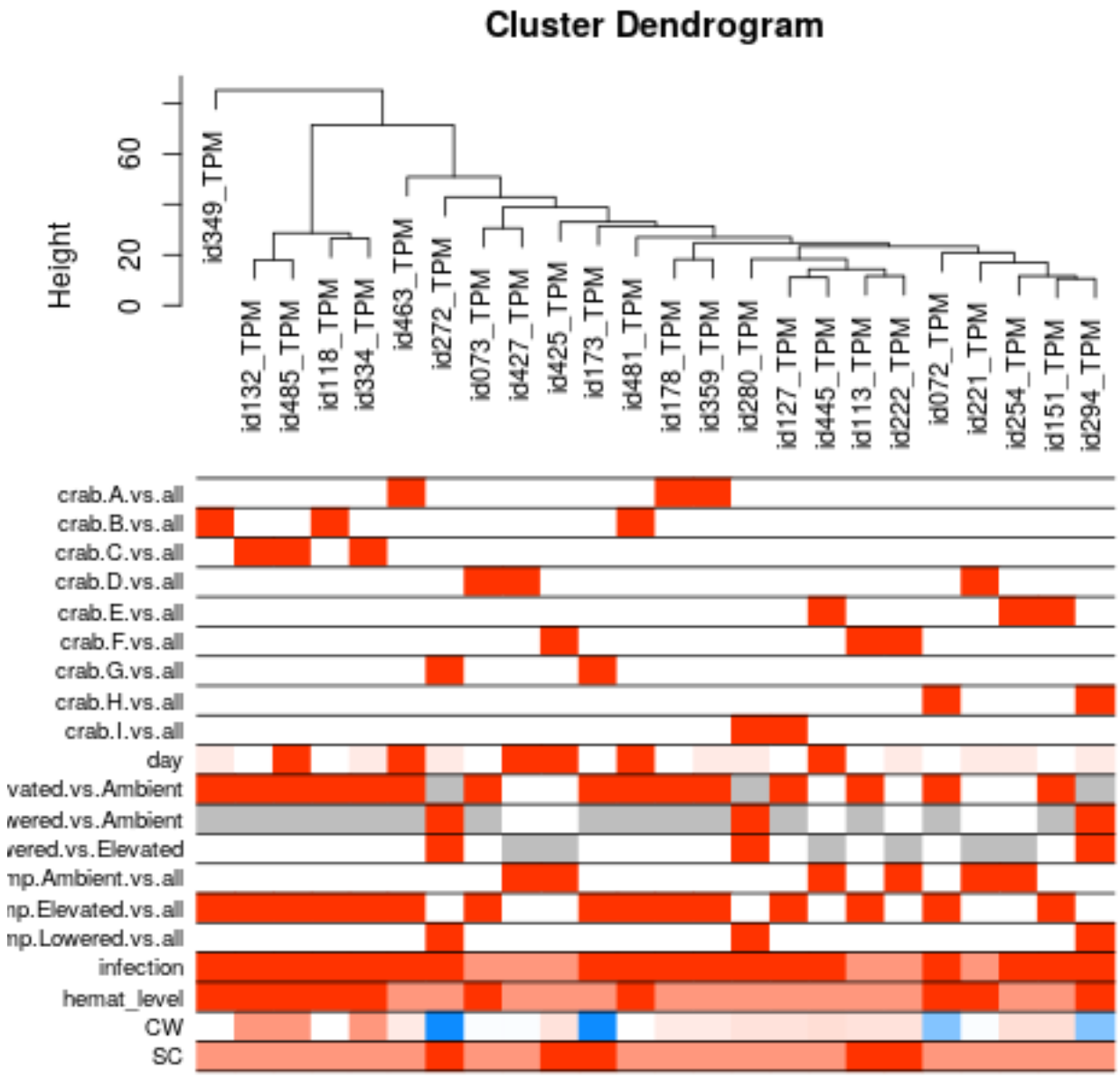

Figure 10: WGCNA Cluster Dendrogram for traits in parasite libraries

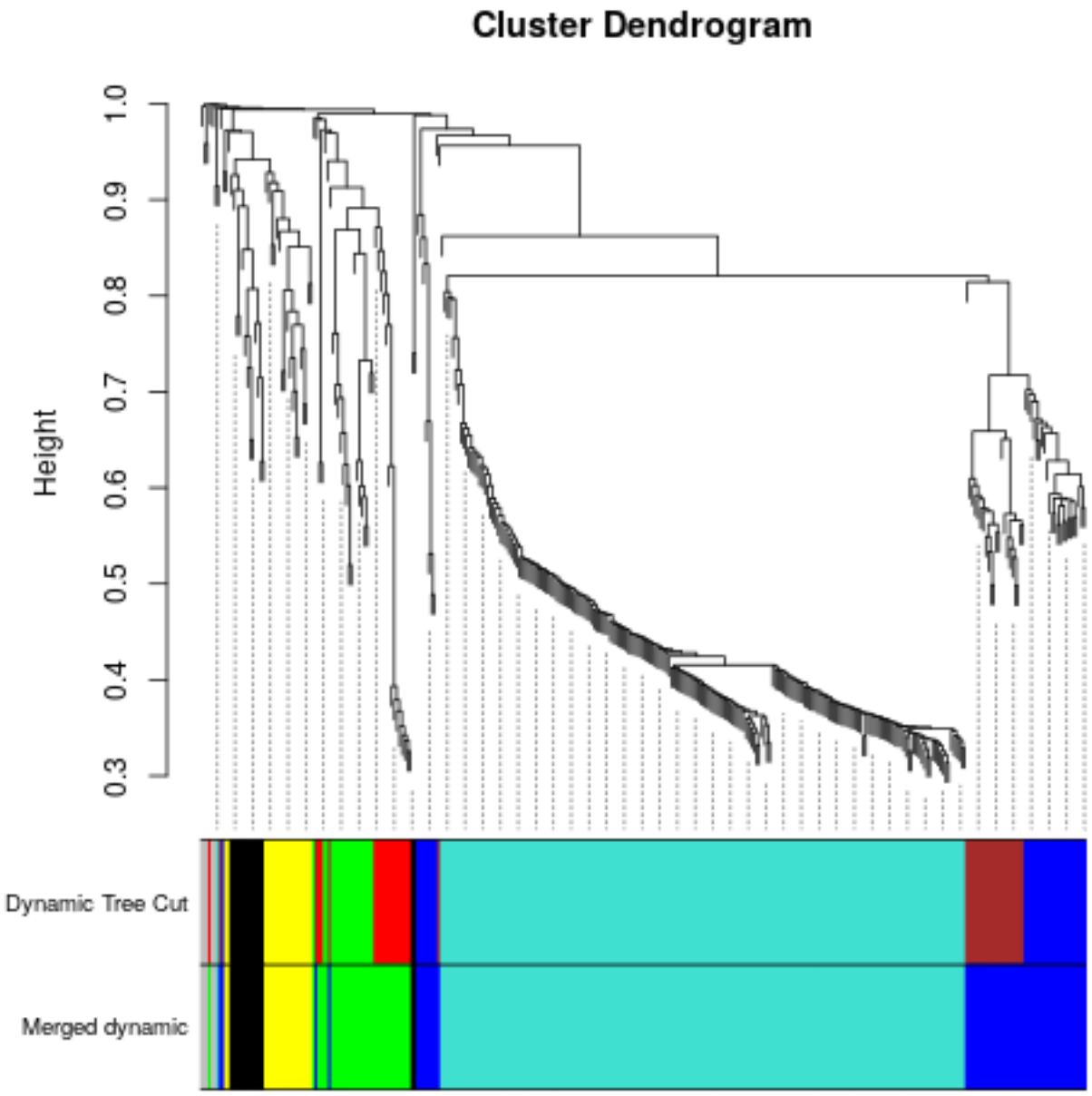

Figure 11: WGCNA Cluster Dendrogram of original and merged eigengenes for parasite libraries

Figure 12: GO term enrichment for parasite black module (linked to day)

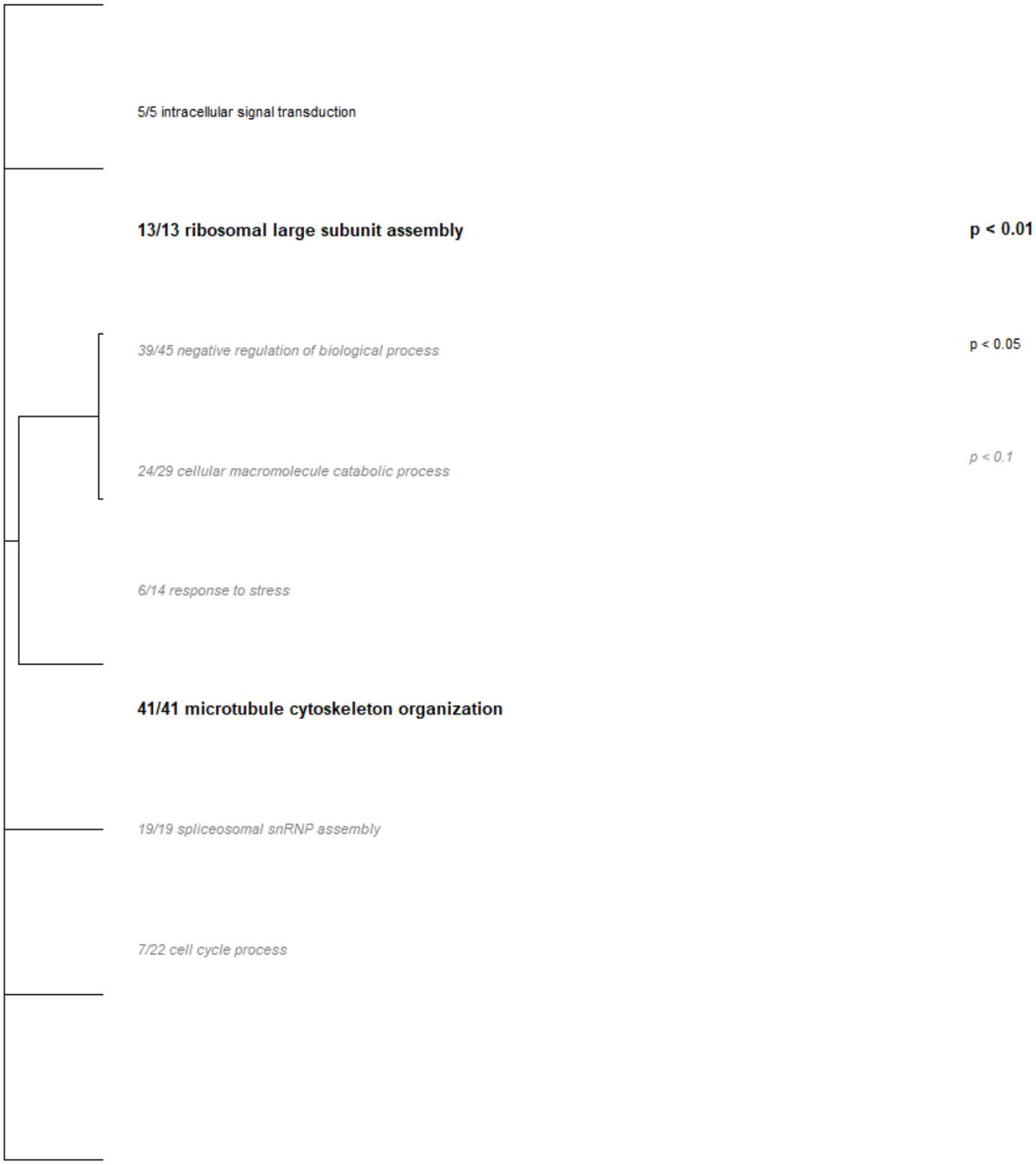

Figure 13: GO term enrichment for parasite turquoise module (linked to temperature and infection level)

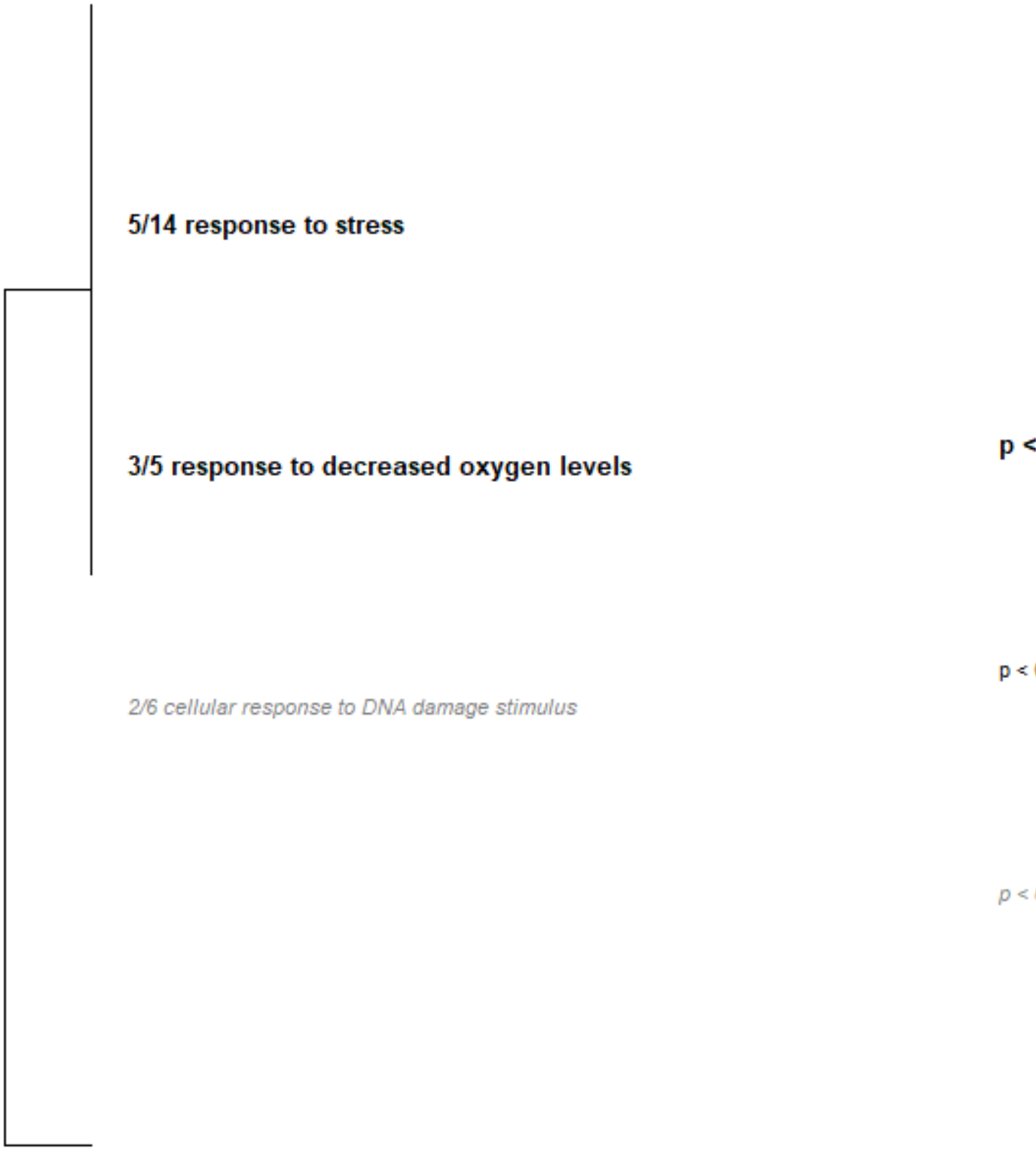

Figure 14: GO term enrichment for parasite blue module (linked to infection level)