Differential Expression in *Hematodinium sp.*

Aidan Coyle

2021-11-09

```
library("kableExtra")
library("tidyverse")
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.4     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.0.2     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter()     masks stats::filter()
## x dplyr::group_rows() masks kableExtra::group_rows()
## x dplyr::lag()        masks stats::lag()
```

```
library("knitr")
library("magick")
```

```
## Linking to ImageMagick 6.9.12.3
## Enabled features: cairo, freetype, fftw, ghostscript, heic, lcms, pango, raw, rsvg, webp
## Disabled features: fontconfig, x11
```

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

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

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 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.

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

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).

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

We used the DESeq2 package 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].

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

[TK: Perform GO-MWU on WGCNA module with p-val = 0.05]

[TK: RUN PCA ON ELEV 0v2 AND AMB 0v2 IMMUNE GENES?]

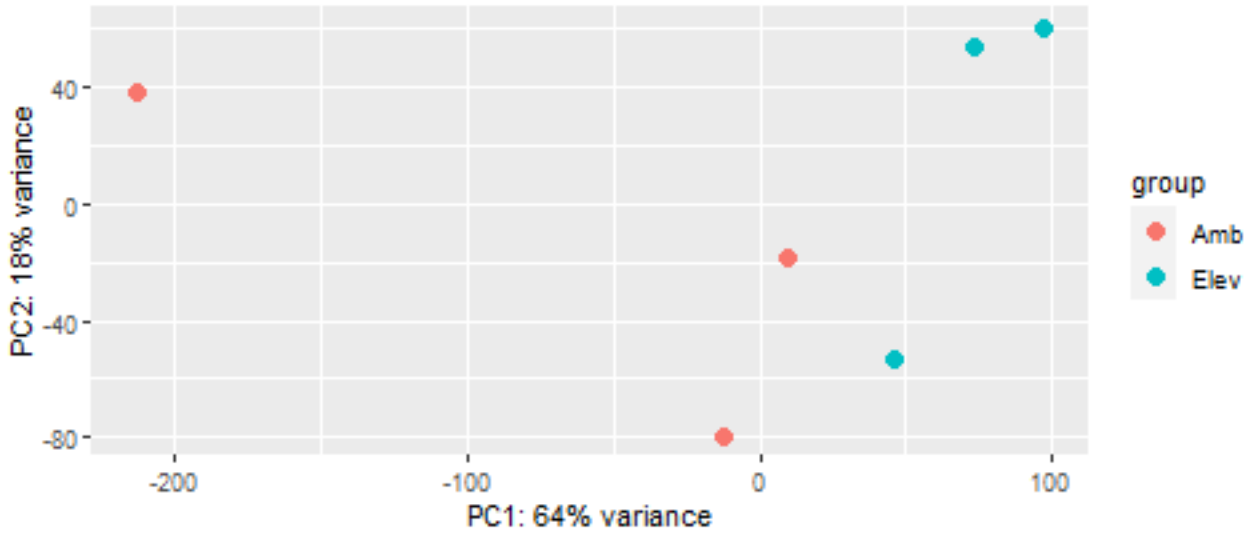

Figure 1: PCA for elevated-temperature libraries, Days 0-2 (unfiltered transcriptome)

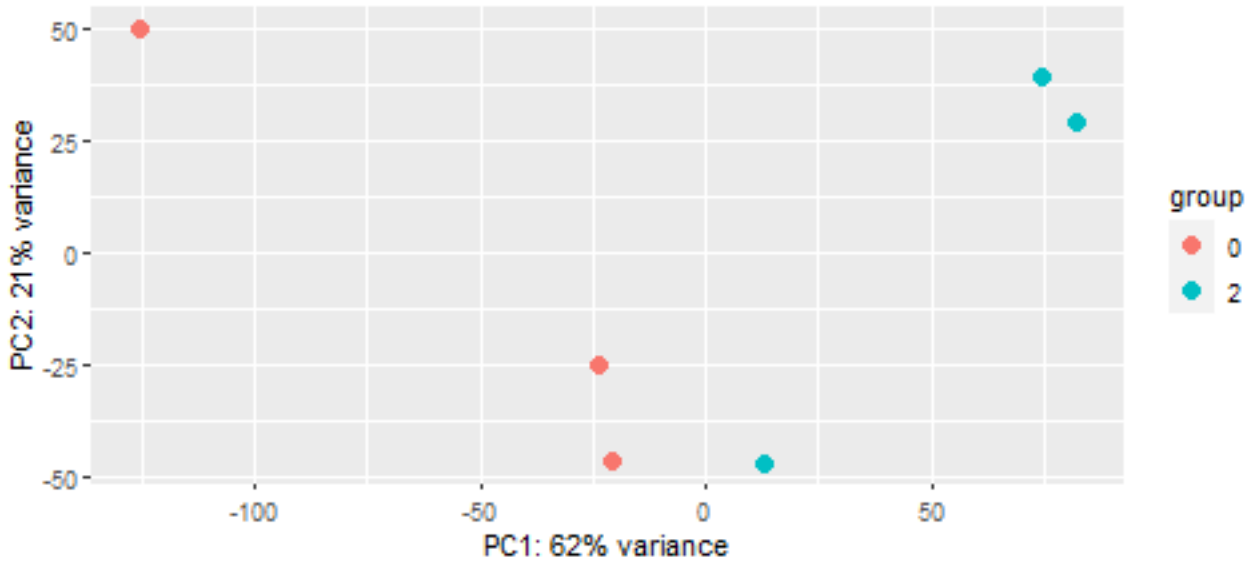

Figure 2: PCA for elevated-temperature libraries, Days 0-2 (crab transcriptome)

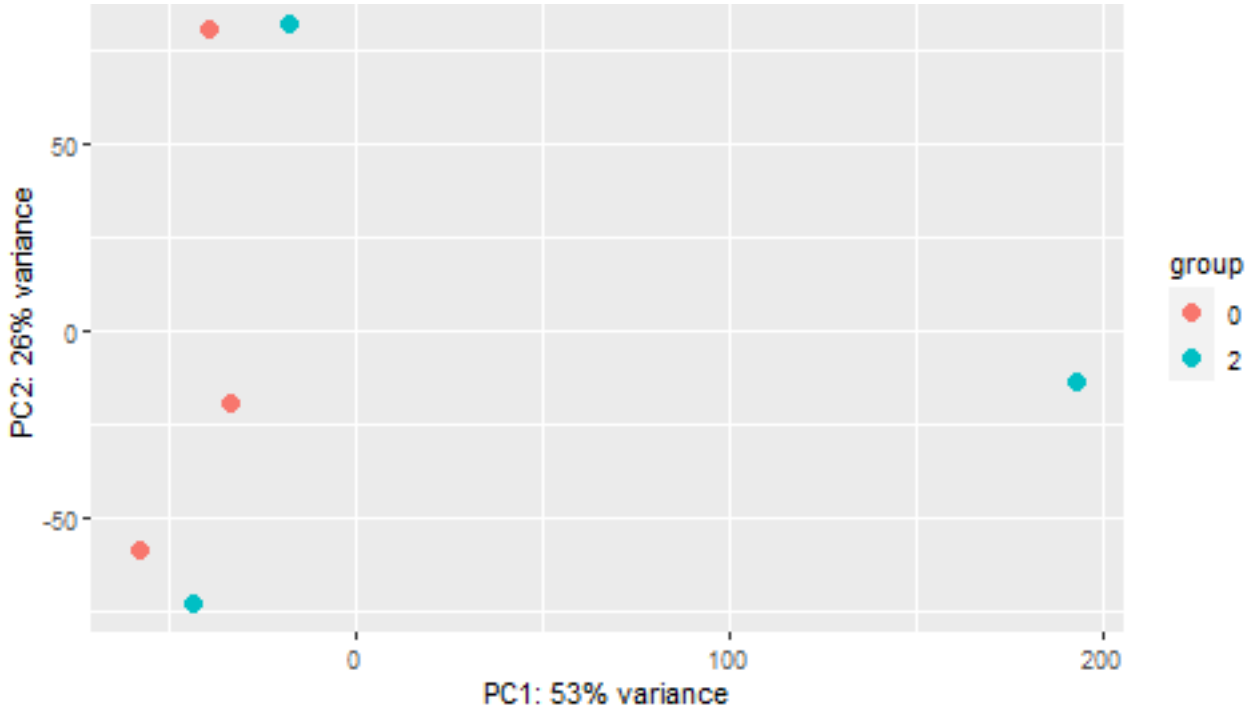

Figure 3: PCA for ambient-temperature libraries, Days 0-2 (unfiltered transcriptome)

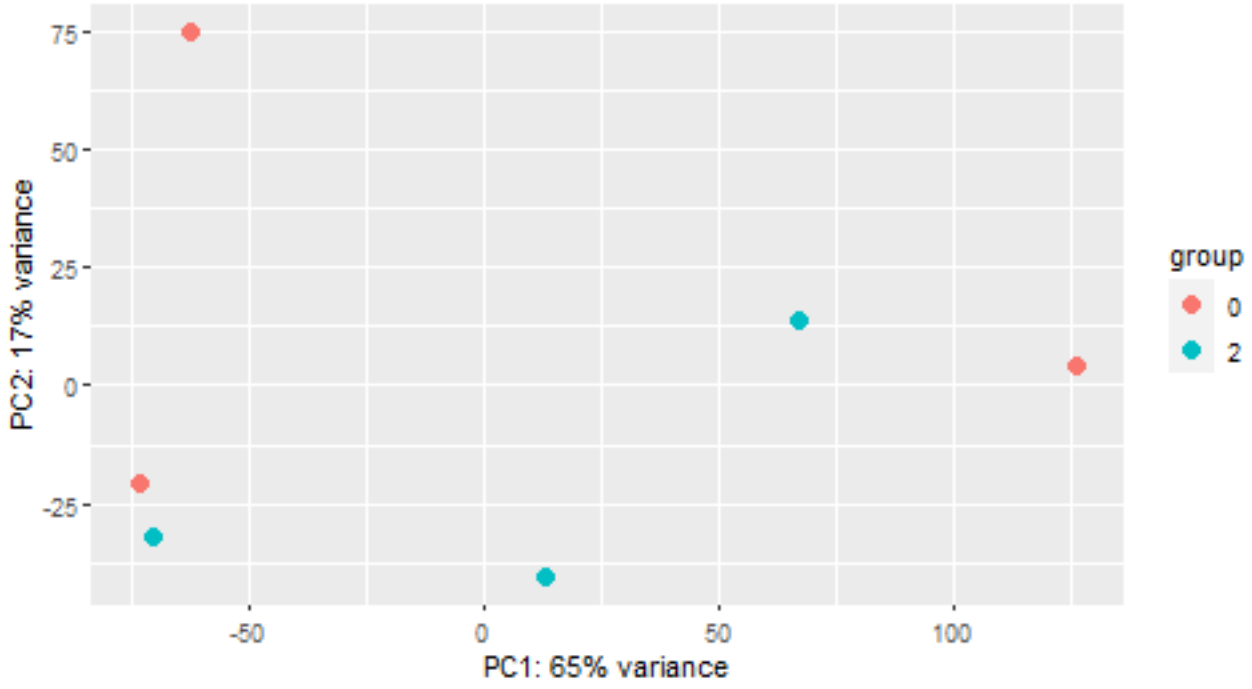

Figure 4: PCA for ambient-temperature libraries, Days 0-2 (crab transcriptome)

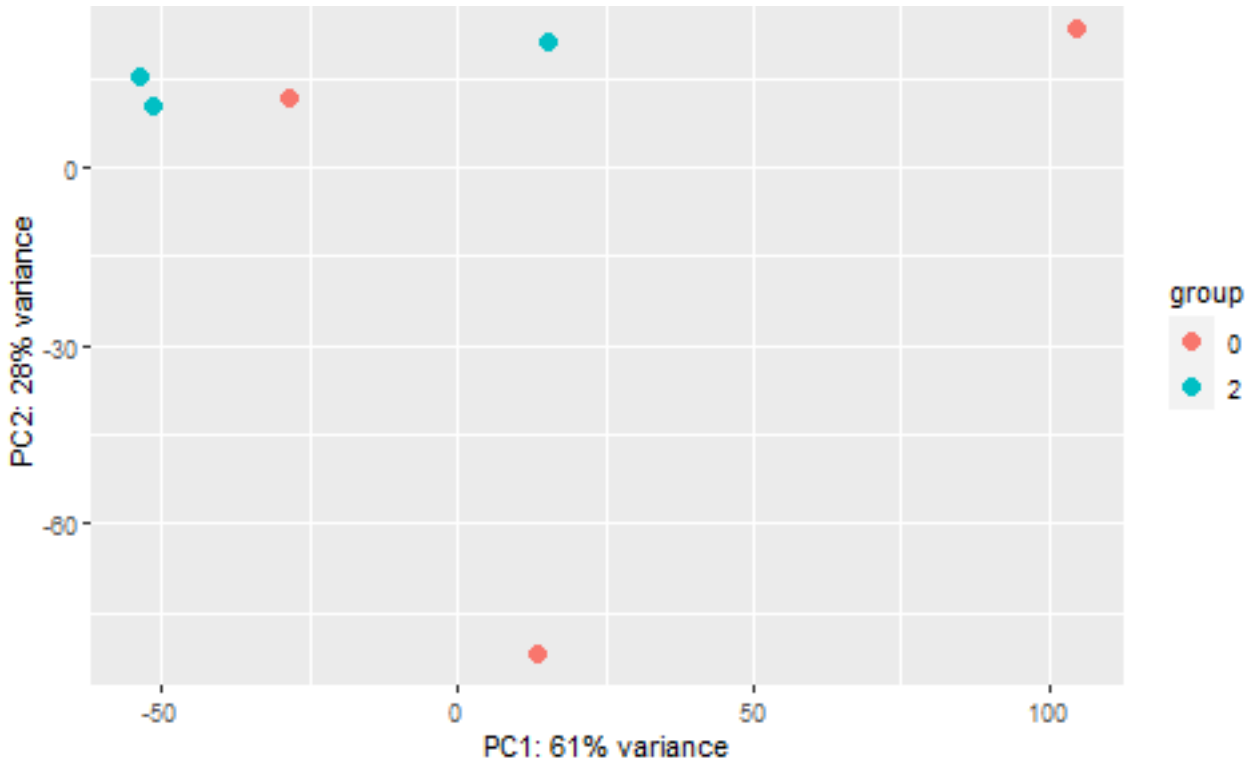

Figure 5: PCA for ambient-temperature libraries, Days 0-2 (parasite transcriptome)

[TK: Try running GO-MWU on cbaiv4.0 Elev2 vs Elev0 + Amb 0 + Amb 2? Same setup as our signif module for WGCNA. Might conflate crab/temp effects, but also might increase sample size to good point]

GO-MWU

We then performed pairwise comparisons 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

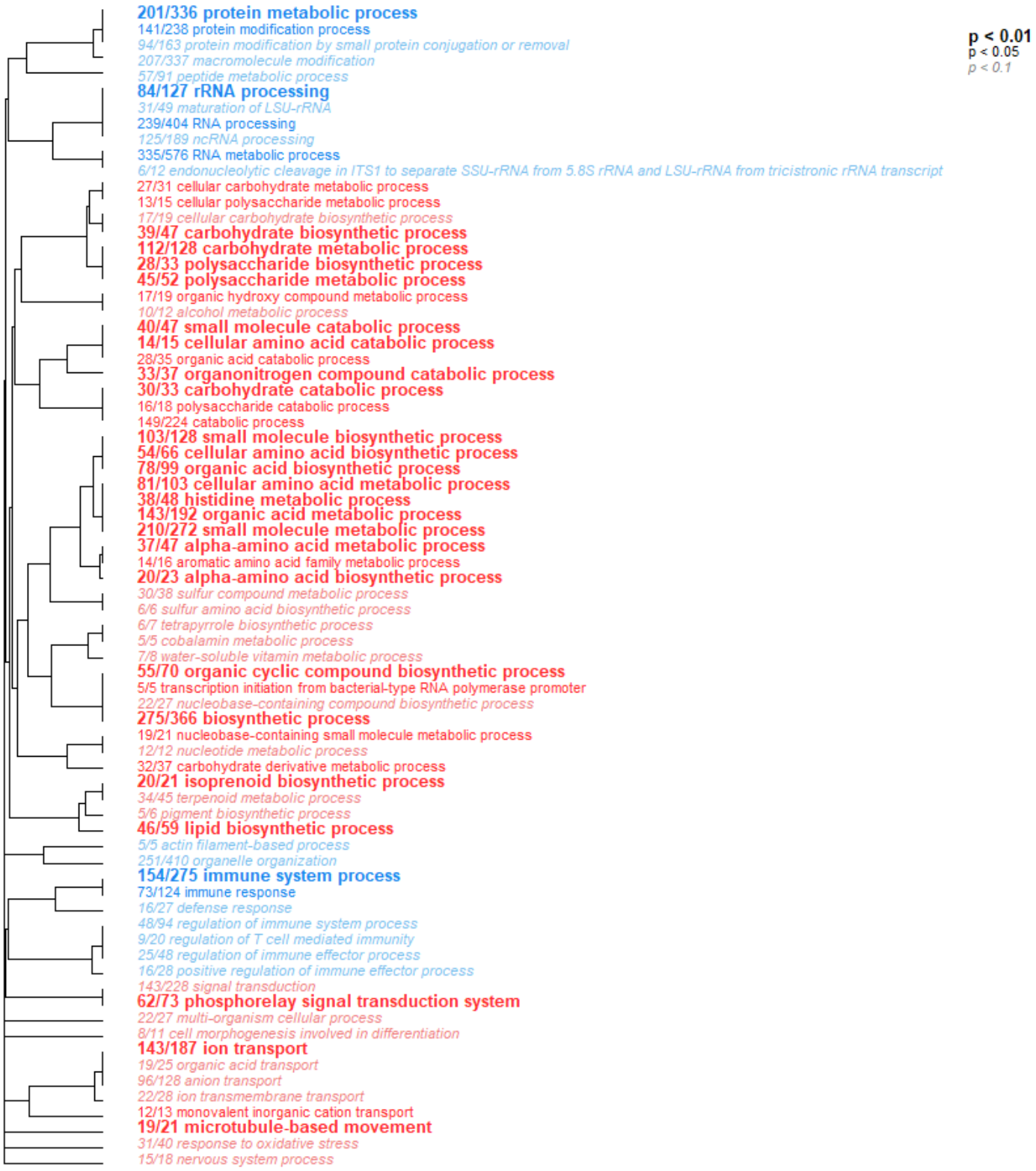

Figure 6: Module enrichment for elevated-temperature libraries, Days 0-2 (unfiltered transcriptome)

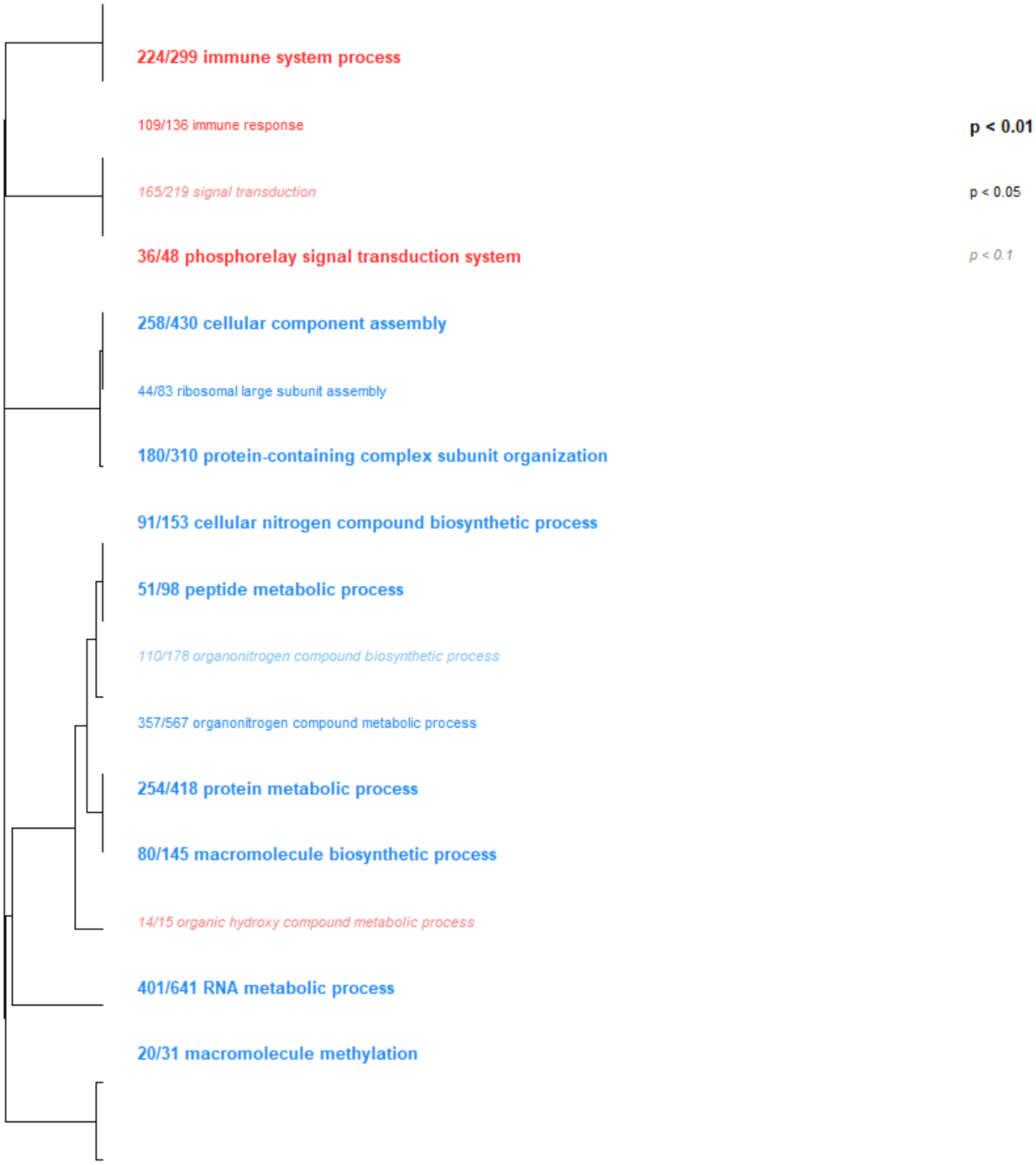

Figure 7: Module enrichment for ambient-temperature libraries, Days 0-2 (unfiltered transcriptome)