New transcriptome manuscript
Crab paper repo: https://github.com/RobertsLab/paper-crab

Methods

Crab collection

In the fall of 2017, approximately 400 immature, male Tanner crabs were sampled by the Alaska Department of Fish and Game (ADF&G) in Juneau, Alaska. The crabs were transported to Ted Stevens Marine Research Institute (NOAA facility, Juneau, AK) and initial hemolymph draws and hemolymph smears were performed for conventional PCR (Jensen et al. 2010) and *in situ* visualization of *Hematodinium sp.* infection. The crabs were left in the tanks for 9 days for acclimation.

Experimental design

Of the crabs that survived the acclimation period, 180 were selected, such that half were infected and half were uninfected as determined by cPCR, to continue with the experiment. Crabs were placed in three different temperature regimes; 10˚C (elevated), 7.5˚C (ambient), and 4˚C (decreased). Each temperature regime consisted of three tanks, each with 20 crabs (10 infected and 10 uninfected). Prior to the initiation of the temperature trial (day 0) hemolymph was sampled - 0.2ml for transcriptomic analysis - and stored in RNAlater. Hemolymph was sampled again after 2 days and 17 days of the temperature trial.

RNA Extraction and Sequencing

Hemolymph samples were centrifuged for 10 minutes at 14000 g with RNA extracted using Quick DNA/RNA Microprep Plus Kit (Zymo Research) according to the manufacturer's protocol. RNA-seq libraries were constructed and sequenced at the Northwest Genomics Center at Foege Hall at the University of Washington, as well as GeneWiz. Descriptions of the samples sent for RNAseq are listed in **Table 1**.

Table 1. Samples sent to UW NWGC and GeneWiz for library prep and RNAseq. For some of the pooled samples, temperature treatments were combined or infection statuses were combined. For the individual pools, crabs were sampled across time points, and the letters are listed the number of times an individual crab was sampled.

Library ID	Crab ID	Pooled or individual	Sample day	Infection status	Temperature treatment
1	*combined*	pool	17	*combined*	*combined*
2	*combined*	pool	0	infected	ambient
3	*combined*	pool	0	uninfected	ambient

4	*combined*	pool	2	infected	*combined*
5	*combined*	pool	2	uninfected	*combined*
6	*combined*	pool	17	infected	*combined*
7	*combined*	pool	17	uninfected	*combined*
8	*combined*	pool	2	infected	decreased
9	*combined*	pool	2	uninfected	decreased
10	*combined*	pool	2	infected	elevated
11	*combined*	pool	2	uninfected	elevated
12	A	individual	0	infected	ambient
13	B	individual	0	infected	ambient
14	C	individual	0	infected	ambient
15	D	individual	0	uninfected	decreased
16	E	individual	0	infected	decreased
17	F	individual	0	uninfected	decreased
18	G	individual	0	infected	elevated
19	H	individual	0	infected	elevated
20	I	individual	0	infected	elevated
21	A	individual	2	infected	ambient
22	B	individual	2	infected	ambient
23	C	individual	2	infected	ambient
24	D	individual	2	uninfected	decreased
25	E	individual	2	infected	decreased
26	F	individual	2	uninfected	decreased
27	G	individual	2	infected	elevated
28	H	individual	2	infected	elevated
29	I	individual	2	infected	elevated
30	A	individual	17	infected	ambient
31	B	individual	17	infected	ambient
32	C	individual	17	infected	ambient
33	D	individual	17	uninfected	decreased

34	E	individual	17	infected	decreased
35	F	individual	17	uninfected	decreased

Table 2. Crabs sampled throughout experiment - 3 time points (exception: G, H, I died prior to day 17)

Crab ID	Infection status	Temperature treatment
A	infected	ambient
B	infected	ambient
C	infected	ambient
D	uninfected	decreased
E	infected	decreased
F	uninfected	decreased
G	infected	elevated
H	infected	elevated
I	infected	elevated

Transcriptome Assembly and Annotation.

Raw sequence data was assessed using FastQC (v0.11.8; Andrews) and MultiQC (v1.6; Ewels) pre- and post-trimming. Data was quality trimmed using fastp (v.0.20.0; Chen et al. 2018) with the "--detect_adapter_for_pe" setting. A transcriptome was *de novo* assembled using Trinity (v2.9.0; Grabherr et al. 2011; Haas et al. 2013). Transcriptome completeness was assessed using BUSCO (v3; Simão et al. 2015; Waterhouse et al. 2018) with the metazoa_odb9 in transcriptome mode and species set as fly. The transcriptome was annotated using DIAMOND BLASTx (v0.9.29; Buchfink et al. 2015) against the UniProt/SwissProt database (downloaded 20200123). Corresponding Gene Ontology information was obtained based on the UniProt GO databases.

^ Add that libraries 1-35 were combined for transcriptome

Differential Gene Expression Analysis

Gene expression differences were assessed by the comparison of RNA-seq data from infected crabs sampled during the trial to uninfected crabs sampled during the trial. Specifically

this included libraries of individuals pooled across temperature regimes, sample at days 2 and 17 (4 libraries). Kallisto was used to obtain count data for each library and an abundance matrix was then produced using a perl script (abundance_estimates_to_matrix.pl) provided as part of Trinity (v2.8.6). Differential expression of contigs was calculated using a negative binomial GLM in the R package DESeq2. The read counts were first normalized using the size factors method and fit to a negative binomial distribution. Significantly differential contig expression (Benjamini-Hochberg adjusted $p<0.05$) between infected and uninfected crabs was determined using the Wald test for significance of GLM terms.

^ Can talk about comparisons by using Library IDs listed in table 1

Enrichment Analysis

Gene enrichment analysis was performed using DAVID v. 6.8 (Huang et al. 2009a; Huang et al. 2009b). The Uniprot Accession IDs from the crab transcriptome were set as the background, and the gene list was the Uniprot Accession IDs from the differentially expressed genes list.

Results

Survival

On day 4 of the temperature trial, a significant mortality event began in the elevated temperature tanks. By day 10 of the temperature trial, 95% of the crabs perished at the elevated temperature. In addition, one crab died in the decreased temperature treatment and three mortalities in ambient conditions.

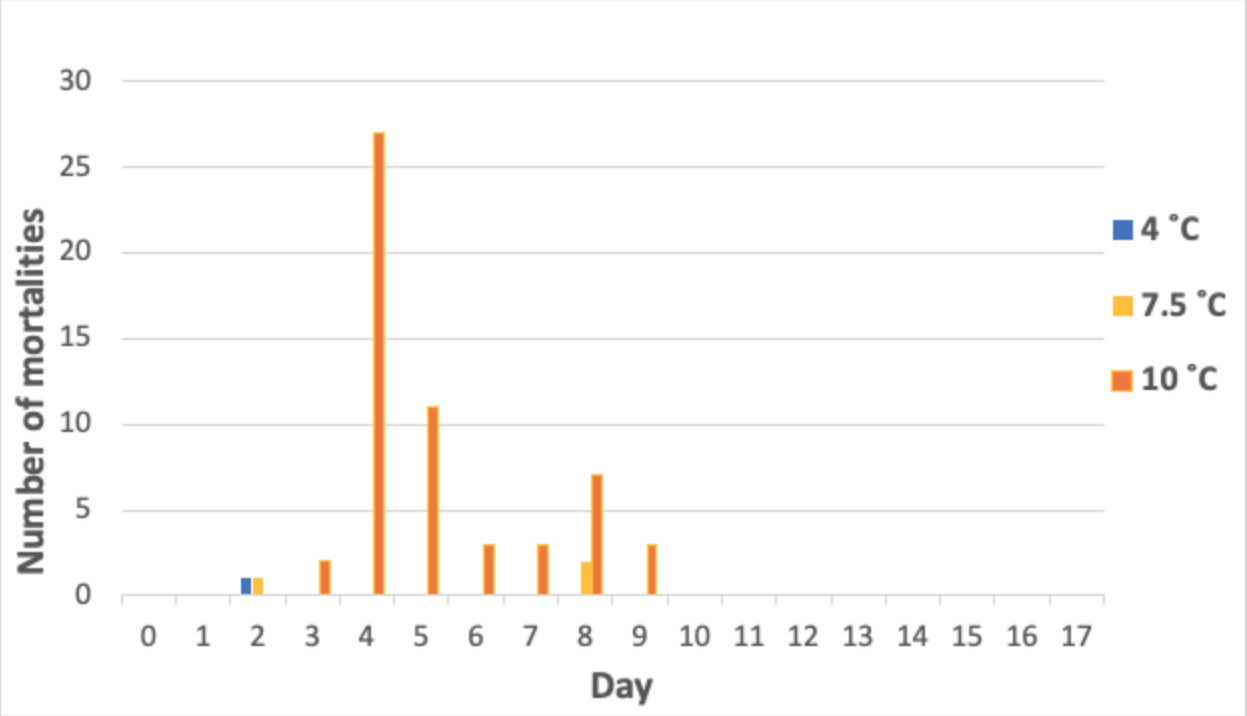

Figure 1. The number of mortalities associated with each treatment over time. The first hemolymph sampling for RNAseq occurred on Day 0. The second occurred on Day 2. The majority of crabs in the warm tanks died between Day 3 (after the second hemolymph sampling) and Day 9. The final hemolymph sampling day occurred on Day 17, and there were only 3 crabs remaining in the warm temperature treatment.

Transcriptome assembly and annotation

Assembly of 819,000,346 base pairs (bp) resulted in 783,006 contigs. The median contig length was 325 bp, with an average contig length of 579.2 bp and an N50 of 811 bp. A total of XXX potential open reading frames (ORFs) were identified. Of the 783,006 contigs, 147,454 were able to be annotated using the Swiss-Prot databases, with 70,078 having corresponding Gene Ontology information (Supplemental material: transcriptome-blast-GO.tab). Genes involved in stress response (35478) were identified and annotated with 15,545 having corresponding Gene Ontology information (Supplemental material: stress-response-genes.tab)

	C. bairdi (this study)	*L. vannamei* (Ghaffari et al. 2014)	*C. maenas* (Verbruggen et al. 2015)
Contigs	783,006	87,307	212,427

Median contig length	325 bp	429 bp	380 bp
Average contig length	579.92 bp	1137.44 bp	992 bp
N50	811 bp	2,701 bp	2,102 bp

Table 1. Comparing summary *de novo* transcriptome assembly statistics between this study's crab transcriptome, the whiteleg shrimp (*Litopenaeus vannamei*) transcriptome (Ghaffari et al. 2014), and the European shore crab (*Carcinus maenas*) transcriptome (Verbruggen et al. 2015).

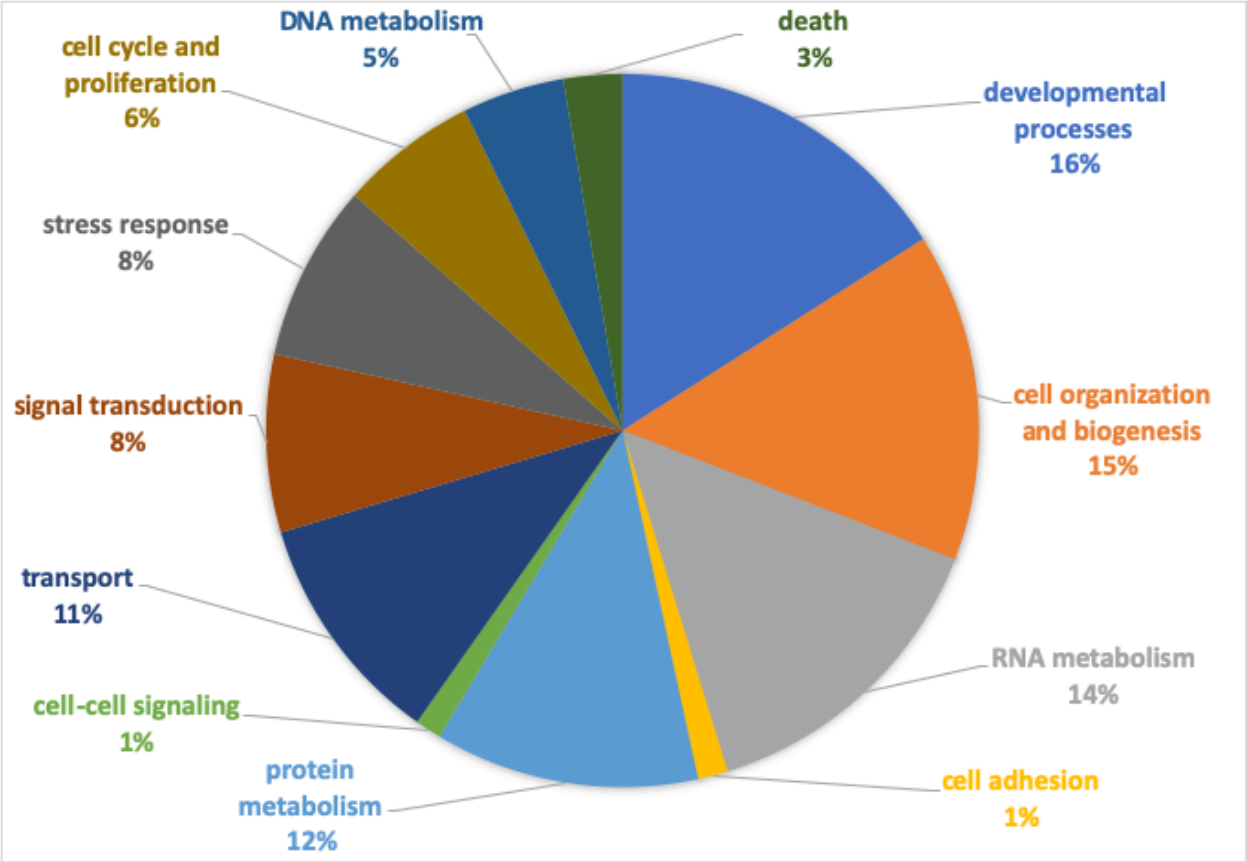

Figure 2. GOslim terms for the biological processes present in the crab transcriptome. Non-descriptive GOslim terms (e.g., other biological processes) were omitted. The percent composition of the pie represents the relative number of genes from the crab transcriptome that contribute to that GOslim term.

Differential Gene Expression

Infection status comparison

A total of 772 differentially expressed transcripts were identified, 646 of which were annotated. There were 698 expressed at an elevated level in the infected individuals and 74 expressed at a decreased level in infected individuals. Additionally, 244 were unique to the infected crabs, while 16 were unique to uninfected crabs. (Supplemental material: 2019-infection_annot-DEGlist.tab). In further analysis of the differentially expressed genes, only one significantly enriched biological process, macropinocytosis (GO:0044351), was identified in the infected crabs when compared to uninfected crabs. There are 12 genes that contribute to the macropinocytosis process (Supplemental material: macropinocytosis_gene_list.tab).

Temperature treatment comparison

A total of 423 differentially expressed transcripts were identified, 339 of which were annotated. There were XXX expressed at an elevated level in the [CONDITION] and XX expressed at a decreased level in [CONDITION] individuals. Additionally, XXX were unique to the [CONDITION], while XXX were unique to [CONDITION]. (Supplemental material: pool_temp-annnot_DEGlist.tab). In further analysis of the differentially expressed genes, significantly enriched biological process,...

DAVID output: https://raw.githubusercontent.com/RobertsLab/paper-crab/master/analyses/DAVID-enrich-temp.txt
DEG list (423) uniprot accession IDs as gene list; Uniprot accession IDs from transcriptome v. 1.5 as background

Figure 3. Gene expression levels in crab hemolymph in [CONDITION VS CONDITION2]. Each dot represents a single contig with red indicating those contigs determined to be differentially expressed (>2 fold change and adjusted p<0.05). XXX contigs expressed at an elevated level in the infected individuals (>2 Log2 Fold Change) and XXX contigs expressed at a decreased level in infected individuals (<-2 Log2 Fold Change).

Discussion

Temperature infection comparison enriched bp notes (DAVID output)

None of the enriched biological processes from the temperature comparison differentially expressed gene list were significantly enriched, however there were a few in the list of 8 processes that were interesting.

Enriched in the elevated (10˚C) temperature treatment when compared to the decreased (4˚C) temperature treatment.

GO:0055114~oxidation-reduction process
- http://www.informatics.jax.org/vocab/gene_ontology/GO:0055114
- http://www.candidagenome.org/cgi-bin/GO/go.pl?goid=55114
    - *Candida* genome → human fungal pathogen
    - A metabolic process that results in the removal or addition of one or more electrons to or from a substance, with or without the concomitant removal or addition of a proton or protons.

GO:0071333~cellular response to glucose stimulus
- http://www.informatics.jax.org/vocab/gene_ontology/GO:0071333
- https://www.yeastgenome.org/go/GO:0071333
- http://www.candidagenome.org/cgi-bin/GO/go.pl?goid=71333
    - Any process that results in a change in state or activity of a cell (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a glucose stimulus

GO:0075522~IRES-dependent viral translational initiation
- http://www.candidagenome.org/cgi-bin/GO/go.pl?goid=75522
- https://www.yeastgenome.org/go/GO:0075522
    - Process by which viral mRNA translation is initiated, where a domain in the 5' untranslated region (UTR) of the viral mRNA called an internal ribosome entry site (IRES) binds the host 43S preinitiation complex, circumventing regular cap-dependent translation initiation.

GO:0000398~mRNA splicing, via spliceosome
- https://www.yeastgenome.org/go/398
- http://www.candidagenome.org/cgi-bin/GO/go.pl?goid=398
    - The joining together of exons from one or more primary transcripts of messenger RNA (mRNA) and the excision of intron sequences, via a spliceosomal mechanism, so that mRNA consisting only of the joined exons is produced

GO:0071398~cellular response to fatty acid
- http://www.informatics.jax.org/vocab/gene_ontology/GO:0071398

GO:0043065~positive regulation of apoptotic process
- http://www.informatics.jax.org/vocab/gene_ontology/GO:0043065

GO:0000245~spliceosomal complex assembly
- https://www.yeastgenome.org/go/GO:0000245

GO:0061418~regulation of transcription from RNA polymerase II promoter in response to hypoxia

- http://www.informatics.jax.org/vocab/gene_ontology/GO:0061418
    - Any process that modulates the frequency, rate or extent of transcription from an RNA polymerase II promoter as a result of a hypoxia stimulus.
- Hypoxia → decreased oxygen
  https://www.usgs.gov/media/images/temperature-affects-dissolved-oxygen-concentrations
    - Hypoxia more of a problem in warm water, so the crabs in warm water not only are dealing with higher temp, but also less oxygen