Transcriptomic effects of dispersed oil in a non-model decapod crustacean

Hernan Vázquez-Miranda
*Florida International University & University of Nebraska- Lincoln, hernan.vm@unl.edu*

Brent P. Thoma
*University of Louisiana at Lafayette & Jackson State University*

Juliet M. Wong
*Florida International University & University of California*

Darryl L. Felder
*University of Louisiana at Lafayette*

Keith A. Crandall
*George Washington University*

See next page for additional authors

Follow this and additional works at: [http://digitalcommons.unl.edu/natrespapers](http://digitalcommons.unl.edu/natrespapers)

Part of the Natural Resources and Conservation Commons, Natural Resources Management and Policy Commons, and the Other Environmental Sciences Commons

Vázquez-Miranda, Hernan; Thoma, Brent P.; Wong, Juliet M.; Felder, Darryl L.; Crandall, Keith A.; and Bracken-Grissom, Heather D., "Transcriptomic effects of dispersed oil in a non-model decapod crustacean" (2017). *Papers in Natural Resources*. 773.

[http://digitalcommons.unl.edu/natrespapers/773](http://digitalcommons.unl.edu/natrespapers/773)
Transcriptomic effects of dispersed oil in a non-model decapod crustacean

Hernan Vazquez-Miranda Corresp., 1, Brent P Thoma 2,3, Juliet M Wong 1,4, Darryl L Felder 2, Keith A Crandall 5, Heather D Bracken-Grissom 1

1 Department of Biological Sciences, Florida International University, North Miami, Florida, United States
2 Department of Biology, University of Louisiana at Lafayette, Lafayette, Louisiana, United States of America
3 Department of Biology, Jackson State University, Jackson, Mississippi, United States
4 Department of Ecology, Evolution, & Marine Biology, University of California, Santa Barbara, Santa Barbara, California, United States
5 Computational Biology Institute, George Washington University, Ashburn, Virginia, United States

Corresponding Author: Hernan Vazquez-Miranda
Email address: hernan@umn.edu

**Background.** Oil spills are major environmental disasters. Dispersants help control spills, as they emulsify oil into droplets to speed bioremediation. Although dispersant toxicity is controversial, the genetic consequences and damages of dispersed oil exposure are poorly understood. We used RNA-seq to measure gene expression of flatback mudcrabs (*Eurypanopeus depressus*, Decapoda, Brachyura, Panopeidae) exposed to dispersed oil.

**Methods.** Our experimental design included two control types, oil-only, and oil-dispersant treatments with three replicates each. We prepared 100 base pair-ended libraries from total RNA and sequenced them in one Illumina HiSeq2000 lane. We assembled a reference transcriptome with all replicates per treatment, assessed quality with novel metrics, identified transcripts, and quantified gene expression with open source software.

**Results.** Our mudcrab transcriptome included 500,008 transcripts from 347,082,962 pair-end raw reads. In oil-only treatments, we found few significant differences. However, in oil-dispersant treatments, over 4000 genes involved with cellular differentiation, primordial cellular component upkeep, apoptosis, and immune response were downregulated. A few muscle structure and development genes were upregulated.

**Discussion.** Our results provide evidence that exposure to chemically dispersed oil causes a generalized cellular shutdown and muscular repair attempts. Our results suggest current oil-spill treatment procedures could be detrimental to crustaceans and indicate additional research is needed to evaluate the impact of oil spills in gene expression. Finally, traditional quality metrics such as N50s have limitations to explain the nature of RNA-seq compared to new methods in non-model decapod crustaceans.
Title: Transcriptomic effects of dispersed oil in a non-model decapod crustacean

Authors: Hernán Vázquez-Miranda¹*, Brent P. Thoma²‡, Juliet M. Wong¹†, Darryl L. Felder², Keith A. Crandall³, and Heather D. Bracken-Grissom¹*.

Affiliations:

¹Florida International University, Department of Biological Sciences, North Miami, FL, 33181.
²University of Louisiana at Lafayette, Department of Biology, Lafayette, LA 70504.
³George Washington University, Computational Biology Institute, Ashburn, VA 20147.

*Corresponding author. Email address: hernan@umn.edu
‡Current address: Jackson State University, Department of Biology, Jackson, MS 39217.
†Current address: University of California, Santa Barbara, Department of Ecology, Evolution, & Marine Biology, CA 93106.
**ABSTRACT.**

**Background.** Oil spills are major environmental disasters. Dispersants help control spills, as they emulsify oil into droplets to speed bioremediation. Although dispersant toxicity is controversial, the genetic consequences and damages of dispersed oil exposure are poorly understood. We used RNA-seq to measure gene expression of flatback mudcrabs (*Eurypanopeus depressus*, Decapoda, Brachyura, Panopeidae) exposed to dispersed oil.

**Methods.** Our experimental design included two control types, oil-only, and oil-dispersant treatments with three replicates each. We prepared 100 base pair-ended libraries from total RNA and sequenced them in one Illumina HiSeq2000 lane. We assembled a reference transcriptome with all replicates per treatment, assessed quality with novel metrics, identified transcripts, and quantified gene expression with open source software.

**Results.** Our mudcrab transcriptome included 500,008 transcripts from 347,082,962 pair-end raw reads. In oil-only treatments, we found few significant differences. However, in oil-dispersant treatments, over 4000 genes involved with cellular differentiation, primordial cellular component upkeep, apoptosis, and immune response were downregulated. A few muscle structure and development genes were upregulated.

**Discussion.** Our results provide evidence that exposure to chemically dispersed oil causes a generalized cellular shutdown and muscular repair attempts. Our results suggest current oil-spill treatment procedures could be detrimental to crustaceans and indicate additional research is needed to evaluate the impact of oil spills in gene expression. Finally,
traditional quality metrics such as N50s have limitations to explain the nature of RNA-seq compared to new methods in non-model decapod crustaceans.
INTRODUCTION

World energy demands make offshore oil drilling a recurrent environmental threat. Marine oil-spill cleanup includes mechanical removal with booms and skimmers, and non-mechanical responses such as on-site burning and chemical dispersion that emulsifies oil into droplets to speed bioremediation. Oil has documented detrimental effects on marine life (Peterson et al. 2003; Trustees 2016), including increased mortality and health decline (Barron 2012; Brown-Peterson et al. 2015; Whitehead et al. 2012; Yednock et al. 2015). However, the effects of employing dispersants on many biological mechanisms have only begun to be understood (Paul et al. 2013), particularly when used in massive quantities. For example, during the Deepwater Horizon Oil Spill (DWH) in April 2010, eight million liters of COREXIT® dispersants were applied to control 507 million liters of oil leaked into the Gulf of Mexico (Kujawinski et al. 2011). It took 87 days to contain DWH, affecting a large portion of the United States gulf coast causing major ecological and economic loss, in what is considered the worst environmental disaster in U.S.’ history (Paul et al. 2013).

Understanding the effect of dispersed oil on organisms is critical as the most toxic compounds of crude oil, polycyclic aromatic hydrocarbons (PAHs), are released in large amounts after dispersion (NRC 2005). PAHs and other petroleum hydrocarbons accumulated in high concentrations following the DWH: 500 parts per million (ppm) in water and 200,000 ppm in sediments (Sammarco et al. 2013). While early toxicology tests have suggested that dispersant formulations used in DWH are less lethal than oil itself (Clark et al. 2001), few recent studies have revealed that dispersants have significant toxic effects on marine life (DeLeo et al. 2015; Paul et al. 2013; Wise & Wise...
and can be more than 50 times more potent when mixed with oil than either oil or dispersant alone (Rico-Martínez et al. 2013). For example, corals show more severe health declines in both dispersant-only and oil-dispersant mixes, compared to oil-only treatments (DeLeo et al. 2015). Dispersant presence impaired survivorship, and muscle configuration and contractions showed sublethal toxicity in jellyfish (Echols et al. 2016). Likewise whale fibroblast show more chromosomal aberrations (Wise et al. 2014). Additionally, oil-dispersant toxicity can persist for months in the water column and sediments (Sammarco et al. 2013), perpetuating detrimental effects.

Transcriptomics allows exploring and assessing oil spill effects for genes expressed at the time of exposure. For example, if genomic references exist, microarrays and targeted quantitative polymerase chain reaction (qPCR) allow for screening hundreds to thousands of genes known to be affected by oil exposure (Bowen et al. 2016; Hansen et al. 2016b; Olsvik et al. 2011; Olsvik et al. 2012). While genomic references are lacking for most marine organisms, next generation sequencing (NGS) technologies allow for using transcriptomic approaches in any study system. RNA-seq, the direct massive-paralleled sequencing of all expressed genes at a given time (Wang et al. 2009), can thus be applied to understanding oil spill effects in non-model organisms. In recent years studies assessing immune response and other key biological processes at the gene level have employed RNA-seq to examine the effects of mechanically dispersed crude oil (Whitehead et al. 2012; Yednock et al. 2015; Zhu et al. 2016) but unfortunately there is limited information on the consequences of chemically dispersed oil exposure (Jenny et al. 2016; Liu et al. 2016; Bayha et al. 2017). Here, we examined the impact of oil and dispersant exposures on a non-model decapod crustacean to test whether chemically
dispersed oil increases detrimental effects on gene expression beyond oil-only exposure. Due to the increased availability of harmful compounds to animal tissues after oil dispersion, we predict key biological processes will be severely impacted. Our objective was to find changes in gene expression in a controlled setting reducing variation sources, instead of aiming to replicate field conditions that are too complex to simulate in vitro and to avoid multiple sources of environmental variation.

We used RNA-seq to quantify gene expression in oyster bed-associated flatback mudcrabs (*Eurypanopeus depressus*) exposed to oil-only and oil-dispersant treatments (Fig. 1). We employed acute non-renewal exposures (Weber 1991) to generate changes in gene expression, albeit our objective was not to perform a comprehensive ecotoxicological study but rather a baseline transcriptomic comparison. Mudcrabs are an ideal system to test oil spill effects because they are associated to oysters and benthic sediments, for which larval stages and juveniles have shown detrimental effects from oil (Cucci & Epifanio 1979) and dispersal exposures (Anderson et al. 2014; Vignier et al. 2016). We collected crabs from sites outside of the DWH area (Fig. 1A) to avoid prior oil-spill exposure as a confounding factor. After a period of in vitro acclimatization, we exposed crabs to the following conditions, each with 3 replicates (Fig. 1C): non-aerated negative control (NC), non-aerated oil-only treatment (OO), and non-aerated oil-dispersant treatment (OD). Additionally, we included an aerated control (AC) to identify effects of oxygen stress, continuous oxygen supply, and experimental manipulation potentially present when using closed, static non-renewal exposures. As the impact of aeration appeared to have little effect on gene expression associated with oil or
dispersant, further comparisons AC are not considered here. (See Supplementary Information SI.)

MATERIALS AND METHODS

Animals
We selected the flatback mudcrab *Eurypanopeus depressus* (hereto referred as crabs; Crustacea: Brachyura, Panopeidae) as the model species because of its association with the economically important Eastern oyster (*Crassostrea virginica*). Bivalves bioaccumulate environmental toxins through filter feeding (O'Connor 1998) and show great susceptibility to oil spills (Vignier et al. 2016). *E. depressus* and oysters are co-distributed throughout the North and South Atlantic coasts, and in the USA from Maine to Florida into the Gulf of Mexico as far as Texas (Williams 1984). Specimens were collected from Rollover Bay near Gilchrist, Texas (29.521667°N, 94.502100°W) on 9 July 2013. In the field, crabs were removed from *C. virginica* interstices. In the lab, collected specimens were re-determined and checked for good physical condition (i.e., no missing limbs, lesions, parasites). Specimens passing said criteria were placed in a 10-gallon tank with artificial seawater and acclimated *ex situ* for 3 days prior to treatments.

Experimental design, conditions, and treatments.
Treatments were mechanically enhanced water-accommodated fraction (WAF) and chemically enhanced water-accommodated fraction (CEWAF) following (Singer et al. 2000) using oil from the Marlin platform Dorado, an oil surrogate with no discernible differences in chemical composition to the MC252 oil from the DWH Macondo Prospect
(Worton et al. 2015). The WAF was generated using a loading rate of 1g/l and mixed in a sealed container with minimal headspace and no visible vortex for 24 hrs before being allowed to settle for 30 minutes. We tried multiple loading rates of dispersant (1:10, 1:20, 1:30, 1:40, 1:100) to determine which concentration could produce a CEWAF with visible effects on crabs. Crabs exposed to concentrations lower than 1:50 did not show motility and behavioral differences compared to controls after gentle prodding with blunt forceps. Thus, the selected CEWAF had a final loading rate of 0.1g/l and a COREXIT® 9500 concentration of 1:50 to capture the responses of crabs to acute exposures. The CEWAF was mixed in a sealed container with minimal headspace and a vortex extending approximately 25% of the total depth of the container for 18 hours before being allowed to settle for 6 hours. After being allowed to settle, 3.5l of WAF and CEWAF were transferred to separate 3.75l vessels with care taken to not disturb the oil layer. We utilized a closed, static design with minimal headspace to limit loss of volatile compounds; vessels were sealed with foil-lined caps (OECD. 2002). Even though industry standard concentrations are about 1:20, dispersant concentration in our CEWAF (1000 parts per million - ppm) that generated an acute response falls within the lower bound of bioaccumulated total petroleum hydrocarbon (TPHs) concentrations found in Deepwater Horizon oil spill water and sediments (Sammarco et al. 2013). We did not verify exposure concentration analytically. Due to the lack of analytical data, our putative experimental concentration are likely lower than reported nominal loading rates due to multiple factors during exposures.

In addition to the WAF, CEWAF, and a sealed, non-aerated control, we included an aerated control to allow us to understand the potential impact of oxygen deprivation in
the closed design to disentangle hypoxia genes from chemical-response genes. By having these two controls we were also able to assess potential gene expression effects due to experimental manipulation. These conditions are henceforth: Aerated control “AC”, non-aerated negative control “NC”, Oil-Only Treatment “OO” (non-aerated WAF), and Oil-Dispersant Treatment “OD” (non-aerated CEWAF). Each treatment and controls were replicated in separate exposure vessels per condition, 4 vessel replicates, and 3 crabs per vessel maintained at 22±2°C for 72-h static, non-renewal exposures (Weber 1991) for a total of 48 crabs. Crabs were not fed for the duration of the exposures. We were not able to include a Dispersant-Only Treatment due to limitations on read sequencing per lane (see Library Preparation) and budget constrains. Therefore the effects of dispersants independent from the effects of dispersed oil were not surveyed (see Discussion).

Tissue harvesting

After exposures crabs flash frozen in liquid N. Specimens were held in liquid N before dissection in RNAlater-ICE (Life Technologies). Tissues from muscle, gills, and hepatopancreas were then frozen in RNAlater-ICE at -80°C until total RNA extraction.

RNA extraction, quality control, and RNA-seq Libraries

We isolated total RNA from 48 individuals. Muscle was chosen over alternative tissue-types that were found to be high in RNAses (i.e. hepatopancreas) or that yielded insufficient amount of high quality RNA (i.e. gill). Tissues were lysed with a hand homogenizer PRO200 (BioGen, Pro Scientific). We used the column-based Nucleospin® RNA kit (Macherey-Nagel), and treated with DNase (Clonetech) to extract RNA
following manufactures’ instructions. Total RNA was eluted in 50 ul of RNase-free water (Sigma) and stored at -80 °C. RNA integrity and concentration was determined in a Bioanalyzer 2100 (Agilent). Muscle RNA of 12 individuals with the highest concentration and quality were chosen for sequencing. These individuals comprise 12 samples in our experimental design: three repetitions of two controls (AC, NC) and two experimental conditions (OO, and OD) [Table S1].

Our 12 selected samples showed some RNA degradation. Thus, to ensure a high quality de novo transcriptome assembly we collected additional crabs from Florida (Fort Pierce; 27.436667°N, 80.335556°W) on 23 September 2014. These crabs were kept in 5-gallon tanks prior to RNA extraction. We extracted total RNA from muscle from three Florida crabs with the Trizol® Reagent (Life Technologies) method. Approximately 10-12 2mm ceramic beads were used in a MiniBead Beater (BioSpec) for homogenization in Trizol®, followed by chloroform-isopropanol/ethanol precipitation including the DNAse and elution steps mentioned previously following manufacturer’s guidelines.

**Library preparation**

Samples containing 1ug of total RNA from the four experimental conditions were sent to the Vanderbilt University Core Lab in 2013 for sample preparation using the RiboZero™ rRNA Removal Kit and cDNA library generation (Epicentre) using a TruSeq Stranded Library kit (Illumina). Library concentrations were verified with a Qubit® Fluorometer (Invitrogen) and library sizes were verified on a Bioanalyzer. The resulting 12 libraries were then multiplexed and sent to Beckman Coulter Genomics for RNA sequencing of paired-end 100 bps reads on one Illumina HiSeq 2000 lane aiming for 20
million reads per experiment as recommended for *de novo* assemblies (Francis et al. 2013). We did not multiplex additional individuals (i.e. a Dispersant-Only treatment) in a single sequencing lane as it would have reduced the number of reads per needed per individual.

In order to increase sequencing depth for the *de novo* reference additional libraries were generated from the three Florida crabs at the University of Georgia Genomic Facility (GGF) in 2014 using 1 ug of total RNA a stranded Kapa RNA Kit (Kapa technologies) and TruSeq adapters (Illumina). Quality control of these libraries was assessed as previously mentioned. Libraries were sent from GGF to the University of Texas Genomics Core Facility for sequencing on one Illumina HiSeq2000 lane. We did not add any RNA spike-in control as they have been found to be not reliable enough for normalization in Poly-A mRNA selection used by TruSeq and Kapa kits, and rather normalized by biological replicates (see below; Munro et al. 2014; Risso et al. 2014).

**Bioinformatics, raw read processing and quality control**

Open source software packages and pipelines were employed for this project bioinformatics’ needs. All paired-end reads from 12 experimental and 3 additional libraries, 15 total, were passed through quality control (QC) first by checking for presence of adapters, indexes, repetitive kmers, and low-quality sequences in FASTQC (Andrews 2010) to determine trimming parameters. Adapters, indexes, and low-quality sequences were removed in TRIMMOMATIC 0.33 (Bolger et al. 2014) with the following parameters: 30GB of total RAM memory, -six threads, a phred scores of 33, cropping the first 9 bases, LEADING:30, TRAILING:30, SLIDINGWINDOW:4:15, and
minimum kmer length of 36. FASTQC was run again on processed read files to verify
that trimming was successful.

**De novo transcriptome assembly and annotation**

Due to the lack of an *E. depressus* reference genome we conducted a de novo
transcriptome assembly in TRINITY 2.0.3 (Haas et al. 2013) using paired-end reads that
passed QC from the 15 total libraries. All libraries were included to generate a single
transcriptome, as it is necessary for a reference to include all individuals to detect
differentially expressed genes across samples (Conesa et al. 2016; Haas et al. 2013).

TRINITY was run on the FIU Panther Cluster in the High Performance Computer (HPC)
environment with 24 cores and 256 GB of RAM with the following parameters:
minimum kmer coverage of 4, maximum memory 252 GB, reverse single-stranded
libraries, 24 CPUs, Butterfly maximum heap space 10GB, Butterfly initial heap space
10GB, Butterfly CPUs 24, and Inchworm CPUs 24.

We evaluated our assembly using traditional metrics including number of
transcripts and N50 values, and supplemented them with the novel ExN50 (B. Haas op
cit; https://github.com/trinityrnaseq/trinityrnaseq/wiki/Transcriptome-Contig-Nx-and-
ExN50-stats) with TRINITY perl scripts. Recently, N50 values to assess quality have
been deemed unreliable as they are often biased due to: (1) filtering artifacts and
combinatorial isoform assembly that skew values upwards, or conversely (2) skewed
downwards if assemblies are efficient at finding low-abundance rare splice variants
particularly with deep sequencing. Thus, pondering the expression levels of each contig
with the ExN50 is more appropriate (http://www.molecularecologist.com/2017/05/n50-
for-transcriptome-assemblies/). We further checked for completeness according to
conserved ortholog content in BUSCO 3.0.0 (Benchmarking Universal Single-Copy
Orthologs; Simão et al. 2015) using annotated Eukaryota, Metazoa, Arthropoda, and
Insecta databases (since there are no decapod crustacean databases) from ORTHODB v9
(Waterhouse et al. 2013; http://busco.ezlab.org/datasets/). These databases include an
increasing number of orthologs as the taxonomic level gets more specific, i.e. there are
fewer orthologs shared by all eukaryots compared to the number of orthologs shared
between insects.

Additionally, we compared our de novo assembly with additional decapod
crustacean and model arthropod transcriptomes to test how much deviated our assembly
is from other taxa. For this purpose we downloaded and assembled nine single-sample
Genbank SRA raw reads: four additional true crabs (Eriocheir sinensis-SRR1735536;
Portunus trituberculatus-SRR768319; Scylla olivacea-SRR2440122; Callinectes
sapidus-SRR2140752), five shrimp (Macrobrachium nipponense-SRR3196792; Caridina
rubella-SRR1248238; Neocaridina denticulate-SRR1185328; Penaeus monodon-
SRR1648423; Fenneropenaeus mergiensis-SRR1756093), one butterfly (Bicyclus
anynana-ERR1022646) and two model arthropods (Daphnia pulex-SRR2350794;
Drosophila melanogaster-SRR2930822). We repeated the exact same TRINITY
parameters previously mentioned after read quality control on the same HPC cluster to
reduce sources of variation. Lastly we also calculated traditional N50s, contig (‘gene’)
number, contig length with TRINITY perl scripts, and created density plots of arthropod
assembly contig variation in R 3.1 (R Core Development Team 2010).
The resulting transcriptome was annotated with TRINOTATE 2.0.2 \(^{[1]}\) and TRINITY plug-in perl scripts on the same HPC node. The plug-in TRANSDECODER was used to obtain likely protein-coding regions and extract the longest open reading frames. Functional annotation was achieved by comparing transcriptome contigs to peptide and transcript databases with BLAST \(^{[2]}\) (Altschul et al. 1990), protein families with PFAM \(^{[3]}\) (Bateman et al. 2004) and HMMER \(^{[4]}\) (Finn et al. 2011), signal peptides with SIGNALP \(^{[5]}\) (Petersen et al. 2011), transmembrane proteins in TMHMM \(^{[6]}\) (http://www.cbs.dtu.dk/services/TMHMM/), and remnant RNA with RNAMMER \(^{[7]}\) (Lagesen et al. 2007). Custom databases were downloaded for TRINITY from the Broad Institute’s website \(^{[8]}\) (https://data.broadinstitute.org/Trinity/Trinotate_v3(Resources/)) for SwissProt, UniProt90, and PFAM. Annotated contigs were compiled and summarized in a SQLITE3 database by translating the transcriptome genes to a transcript map using TRINITY’s TRANSMAP utility and loading those results into the SQLITE database. All annotations blasted to a record with a \(1 \times 10^{-5}\) identification E-value for gene ontology (GO) determination. Table 1 includes a brief summary of environmental and experimental conditions used to assemble the flatback mudcrab transcriptome compliant with the MIGS standard.

**Differential gene expression statistical analyses**

Reads passing QC for each replicate per condition (AC, NC, OO, OD) were included in the differential gene expression estimates. Libraries from Florida crabs were not included to avoid geographic, genetic, and sequencing artifacts (Gleason & Burton 2015). Pair-end reads were merged prior to alignment to the resulting *de novo*
transcriptome. Pair-end reads with no overlapping regions were concatenated with a custom bash shell script to merged reads. Since single stranded libraries only contain the reverse read, all reads were reverse-complemented using SEQTK (https://github.com/lh3/seqtk). We did not incorporate RNA spike-ins as controls (see above), therefore our reported values are relative to normalization between biological replicates.

In order to separate effects of low oxygen and experimental manipulations from oil and oil-dispersant effects we first compared both controls to calculate abundance counts and differential expression analyses, and secondly we excluded the aerated control (AC) and re-estimated counts for NC, OO, and OD. Abundance and differential gene expression was calculated as follows. Transcript abundance was estimated for each experimental library using RSEM (Li & Dewey 2011) by preparing the transcriptome reference with a TRINITY utility perl scripts aligning experimental reads per treatment for transcripts and genes with BOWTIE (Langmead 2010), 7 cores, and 30 GB of memory. Raw count data from RSEM were normalized by determining fragments per kilobase of transcript per million mapped reads (FPKM) with similar TRINITY perl scripts.

Differential gene expression (DE) on normalized counts from the experimental reads was estimated on the BIOCONDUCTOR platform (Gentleman et al. 2004) in R 3.1 (R Core Development Team 2010) with packages edgeR (Robinson et al. 2010) and DESeq2 (Love et al. 2014). The transcript and gene normalized count matrices were used to calculate pairwise differential expression between each treatment from the mean and standard deviation of each biological replicate to generate Bland–Altman plots (Bland &
Altman 1986) measuring logarithmic fold change in gene expression (M) over mean average expression on log counts (A), known in microarray and RNAseq studies as MA plots (Dudoit et al. 2002). Statistical significance was assigned on a log₂ scale values passing a False Discovery Rate (FDR) of 1%. Additionally, volcano plots were generated to visualize gene expression fold change by their statistical significance (Cui & Churchill 2003). Parameters in edgeR were left as defaults. DESeq2 calculations were run with comparisons per condition, a maximum sharing mode, and a local regression fit type.

Gene regulatory direction was determined in edgeR and DESeq2 for genes and isoforms by comparing each experimental treatment replicate log-transform and zero-centered normalized FPKM counts. Those counts were used to estimate Euclidean distances between treatments and between genes to generate expression clusters by similarity visualized as heatmaps with TRINITY perl and custom R scripts. Finally, over and underrepresented GO terms were calculated with a statistical enrichment analysis in GOseq (Young et al. 2010). We used DE pairwise comparisons of gene count matrices between NC, OO, and OD estimated in DESeq2 and edgeR. We controlled for gene size with weighted average gene lengths across all experimental conditions, and GO annotations from the assembled transcriptome with TRINITY and TRINOTATE perl and python scripts. We considered statistical significance when \( P < 0.05 \) after a 1% FDR correction for multiple comparisons. Having two statistical methods, three possible pairwise comparisons, and two possible patterns (overrepresented or enriched, and underrepresented or depleted) per comparison pair, would produce 24 GO term enrichments if every comparison has DE terms.
RESULTS

We obtained 347,082,962 pair-end reads from our libraries with an average of 23 million reads per replicate. TRINITY (Haas et al. 2013) produces sequence contigs categorized as two features: ‘genes’ represent unique contigs, and ‘isoforms’ correspond to alternatively spliced transcripts in a given contig. Our TRINITY assembly yielded a transcriptome of 500,008 genes (i.e. unique contigs), 660,546 isoforms (i.e. alternatively spliced transcripts in a given contig), with an N50 of 421 base pairs (bps) and a mean length per contig of 428.03 bps. The total number of assembled bases was 282,733,223 (table S1). In contrast, between 60-80% ExN50 values were more than double the traditional N50 (Fig. 2A). The level of conserved single-copy orthologs ranged between 70.2-88.3% complete, 13.9-2.3% missing, and 15.9-9.2% fragmented BUSCOs for exclusive and inclusive taxonomic hierarchies, respectively (Fig. 2B).

Our comparative contig density plots revealed an unprecedented level of variability across single-sample Arthropoda assemblies and our de novo transcriptome (Fig. 3). TRINITY found fewest contigs in the model arthropods (D. pulex and D. melanogaster, Fig. 3C). In non-model arthropods (crabs, shrimp, and butterfly) the number of contigs ranged between 11,534-243,857. Shrimp ranged between 11,534-158,098 contigs (Fig. 3B), and crabs had the largest between 86,287-243,857 contigs (Fig. 3A). For every species the contig size distribution peaked between 400-600 base pairs (bps). However several N50 values were above the 1000 bps traditional benchmark inversely correlated with peaks’ height, i.e. the higher the density peak, the lower the N50 (Fig. 3). Two crabs, three shrimp, and D. pulex had N50s <1000. For model arthropods and the butterfly, and in particular for D. melanogaster, a large portion of the density
distribution exceeded 1000 bps. The contig size density for two shrimp (*F. mergiensis* and *N. denticulata*) and two crabs (*C. sapidus* and *S. olivacea*) partially exceeded 1000 bps. Our *de novo* transcriptome had a similar contig size distribution to two crabs (*Eriocheir sinensis* and *Portunus trituberculatus*) and three shrimp (*Macrobrachium nipponense*, *Caridina rubella*, and *Penaeus monodon*). The number of contigs in our transcriptome produced with 15 samples was approximately double compared to that of *Eriocheir sinensis* with a single sample, and its N50 30.5% smaller.

We compared NC and experimental treatments (OO, OD) to determine differences in gene expression between crabs exposed to oil-only and oil-dispersant mixes (*SI* Fig. S1, comparisons between controls [AC and NC, *SI* Fig. S2, *SI* Table S2] can be found in Supplementary Information).

**Differential gene expression happened between oil-only and oil-dispersant.**

We measured differential gene expression between NC, OO, and OD using edgeR (Robinson et al. 2010) and DESeq2 (Love et al. 2014). The main difference between statistical methods was the amount of significantly expressed features, with DESeq2 detecting larger amounts with relatively lower significance magnitude. We report feature numbers from DESeq2 herein (edgeR comparisons are found in the *SI*). Comparisons of controls and treatments detected hundreds of genes and thousands of isoforms that were differently expressed. Approximately 50% of these differentially expressed features in our *de novo* transcriptome (*SI* DS1) could not be annotated and were left as unknown by either low sequence similarity or by being putative proteins with missing gene ontology (GO). Five differentially expressed features were shared between the controls and
experimental treatments but did not blast to any GO (SI Fig. S1). Nevertheless, 99% of differentially expressed genes and isoforms were detected in OO and OD.

Comparisons between NC and OO-OD identified 10s to 1000s of features with significant differential expression (SI Figs. S3-S4, SI Table S3). Feature counts were as follows: NC vs. OD identified 4388 genes and 3243 isoforms (Figs. 4, SI S3-S4, SI Table S3); NC vs. OO identified 11 genes and 8 isoforms (Fig. 4, SI Table S3); OO vs. OD identified 448 genes and 215 isoforms (Fig. 4, SI Figs. S3-S4, SI Table S3).

When comparing gene expression across replicates within treatments, replicates within NC and OD had nearly identical heatmap profiles (Figs. 5, S4); however, replicates within OO shared heatmap profile characteristics with both NC and OD (Figs. 5, SI S4). In 75% of heatmaps, 66% of OO replicates clustered more closely with OD (SI Fig. S4) indicating that OO and OD replicates have similar gene expression profiles and are statistically distinguishable from NC replicates.

Two clear feature clusters are present in all heatmaps: one cluster containing over 90% of differentially expressed features, upregulated in NC and downregulated in OD, and a second cluster of less than 10% with the opposite pattern of regulation (clusters 1 and 2 respectively on Figs. 5, SI Fig. S4).

**Gene ontologies largely matched muscle components and immune response genes**

Gene ontology annotations allow for the identification of features and their functions. Pairwise expression measured in log2 fold changes (logFC, SI DS14) between experimental conditions ranged -16 to 12. Most differentially expressed features (logFC > |8|) did not blast to a known protein, to an undescribed protein, or matched muscle
components. Immune response features logFC ranged -4 to -2. Although gene expression studies typically show gene rankings per fold change limiting detailed description of the top 10-30 genes (See full rankings in SI DS14), we report herein transcriptome-wide GOs with differential gene expression instead of a handful of gen-per-gene comparisons. 70% of shared features detected as differentially expressed in NC and OD (SI Fig. S1) did not BLAST to a GO and were left as unknown (SI Figs. S5-8). Nevertheless, 19 features upregulated in OD (and thus downregulated in NC) were annotated with GO terms including: cell division, proliferation, and regulation, as well as muscle structure, attachment, development, and other functional components (SI Table S4, SI Figs. S5-S8).

More specifically, upregulation of structural muscular elements included sarcomeres, I-bands, Z-discs, and muscle-associated actin binding responsible for muscle formation and contraction. Mechanisms involved in cell division, including chromatid cohesion and chromosome condensation, were upregulated. Lastly, we found one upregulated invertebrate immune response mechanism (hemocyte proliferation) involved in the protection against microbial infection (Schulz & Fossett 2005).

The largest GO annotation set included 151 features that were downregulated in OD. These features were associated with A) fundamental cellular components including the nucleus, membranes, cytoskeleton, organelles, and cytoplasm (SI Table S4, SI Figs. S5-S8) B) molecular functions such as energy transfer by ATP for ubiquitines, metal ion binding, mRNA processing, nucleotide binding, protein phosphorylation, stabilization, proteolysis, DNA synthesis, RNA processing and binding, transcription, translation and C) cellular processes such as apoptosis, apoptosis regulation, cell adhesion, cell migration and protein development, cellular response to hypoxia, glucose and glycogen
metabolism, intracellular signal transduction, splicing speckles, transmembrane transport, and the mitotic cell cycle. We found several downregulated actin elements involved in microfilament formation for cytoskeleton structure and cell movement, locomotion, including filament binding and assembly. Downregulated components of the nervous system included dendrite morphogenesis, neuron projection, neuron apoptotic process, and neuronal cell bodies. Three metabolic/signaling pathways downregulated in OD included: A) general cell development and immune response pathway as indicated by cytokine signaling; B) cell growth, proliferation, survival, and apoptosis pathways including activation of mitogen-activated protein kinases (MAPK), SAPK/ Jun amino-terminal kinases (JNK), epidermal growth factor receptor (EGFR), fibroblast growth factor receptor (FGFR), and transforming growth factor-β receptor (TGF-β); and C) protective response of the immune system pathways including Fc receptors (FcR), the IκB kinase (IKK) and NF-κB receptors, T cell receptors, and Toll-like receptors (TLRs).

Lastly, in OD we found 3 and 19 downregulated isoforms of Tumor Necrosis receptors (TNF) and Cytochrome P450 (CYP), respectively (table S5). Importantly, TNFs work as natural defenses against tumors by killing abnormal cell growth (Fransen et al. 1985) and the CYP protein family of membrane receptors are widely used as hydrocarbon exposure biomarkers (Rewitz et al. 2006). These CYP transcripts belong to three protein clans (clans 3, 4, and 18). Unknown features, not matching a functional GO, constituted 30% and 8% of downregulated genes and isoforms, respectively. Gene ontology terms were statistically over and under represented between controls and dispersed oil.
GO term enrichment analysis tests for categories that dominate differential gene expression. Only 11 out of 24 possible pairwise enrichment tests included DE genes (DEGs; Fig. 6): 8 from DESeq2 and 3 from edgeR. In most cases Biological Processes (BPs) were the dominant ontology, followed by Molecular Functions (MFs), and lastly Cellular Components (CCs). Those DE GO terms were consistent with our cross-reference list (see above; SI Table 4). DE counts between NC and OO totaled 20 GO terms, and between OO and OD there were 470 GO terms. Pairwise DE counts between NC and OD were much higher, totaling 3578 GO terms. However, significantly enriched and depleted terms were only found in NC in NC-OD DESeq2 pairwise comparisons after the FDR correction, totaling 1832 and 75 respectively (Fig. 6). Statistically significant overrepresented GO terms included fundamental metabolic, developmental, and regulation processes, signaling, cellular differentiation, organization, and communication, and morphogenesis. Underrepresented GO terms included fundamental molecular activities, integration, transposition, DNA replication and recombination, in ribosomes and mitochondria (File DS 15). We found the same overall pattern in edgeR although none of the GO terms were significantly enriched.

**DISCUSSION**

**Assembly quality and transcriptome variability in crustaceans**

Our transcriptomic assembly yielded a relatively high number of contigs and relatively low N50s. These numbers would typically suggest a poor assembly, possibly stemming from the RNA degradation we reported. However, our ExN50s indicate the N50 based on 100% of contigs is biased due to short and lowly expressed isoforms. Further, 70-88%
complete BUSCOs suggest that our assembly contains mostly complete genes present in all animals, and comparable to studies with seemingly no degradation and various degrees of divergence to references (57-99%; Levin et al. 2016). Crustaceans have many more splice variants compared to other metazoans, elevating contig count and low gene ontology annotation (~32%) due to highly divergent or novel sequences (Havird & Santos 2016). We were able to annotate ~50% of our transcriptome, suggesting that even though some gene fragmentation was present, our transcriptome has similar characteristics to other crustacean assemblies. Moreover, we found a little over ~4000 DEGs, similar to the average seen in previous studies (Havird & Santos 2016) indicating the number of contigs did not inflate detected DEGs. Our data supports recent observations that traditional metrics such as N50s are not a good proxy for RNA-seq data quality (https://github.com/trinityrnaseq/EMBOtrinityWorkshopSept2016/wiki/De-novo-Assembly,-Quantitation,-and-Differential-Expression; http://www.molecularecologist.com/2017/04/the-first-problem-with-n50/), in particular for non-model organisms where known metrics are not “one-size-fits-all” (Havird & Santos 2016).

Moreover, we showed an unprecedented amount of transcriptomic assembly variation in arthropods, particularly in non-model decapod crustaceans. The wide range of N50 values and contig numbers from uniform assembly and quality control further exposes how traditional expectations such as N50s do not accurately describe the nature of RNA-seq in many crustaceans. In all arthropods the bulk of the contig density was smaller than 1000 bps, implying “large” N50s are a misleading metric that does not represent contig size accurately. While many of our contigs (‘genes’) were shorter, their
distribution fits with other crab and shrimp transcriptomes suggesting decapods could be more complex that model organisms and could have more, shorter genes (Havird & Santos 2016); not necessarily poorly assembled. Even if TRINITY overestimated the number of genes by a factor of two in fruit flies and water fleas (probably due to having a single sample per assembly), the number of contigs found in decapods is consistent with genomes at least an order of magnitude larger (0.17-0.2 Gbases versus >1-40 Gbases; http://www.genomesize.com/). Water fleas have high number of duplicated genes (Colbourne et al. 2011), thus it is possible that gene duplication could explain larger number of contigs in other crustaceans. Future research will help disentangle number of genes, assembly artifacts, and gene duplication in many non-model organisms.

**Downregulated features dominate differential gene expression**

We examined the effects of oil and oil-dispersant mixes on crab gene expression. Although numerous studies have examined acute oil-spill toxicity in animal systems (e.g., (Hemmer et al. 2011; Paul et al. 2013; Rico-Martínez et al. 2013; Wise & Wise 2011), there are relatively fewer studies that use genomic methods to investigate the response to oil-dispersant exposure (Bowen et al. 2016; Han et al. 2016; Hansen et al. 2016a; Hansen et al. 2016b; Jenny et al. 2016; Liu et al. 2016; Olsvik et al. 2011; Olsvik et al. 2012; Whitehead et al. 2012; Yednock et al. 2015).

Because our primary focus was on differential gene expression in response to OD exposure and methodological constrains, we did not include a dispersant-only treatment and cannot quantify the separate effects of dispersants. Nevertheless, acute toxicity tests on mysid shrimp (*Americamysis bahia*) and the inland silverside (*Menidia beryllina*) revealed that dispersant-only treatments had lower toxicity effects when compared to OO
and OD conditions (Hemmer et al. 2011). Additionally, toxicity studies in decapod shrimp showed sensitivity increase of an order of magnitude in OD over OO (Fisher & Foss 1993). These studies suggest that crabs exhibit increased negative effects when exposed to dispersed oil.

In individuals exposed to OD, we found evidence for the downregulation of >4000 genes involved in fundamental cellular mechanisms, suggesting a generalized shutting down of crabs’ cellular processes and functions (SI Table S3). Some of these included genes involved in actin binding, condensation, and folding in the cytoskeleton (SI Table S4). Failures associated with these mechanisms cause multiple diseases in humans (Lundin et al. 2010) and flight muscle alterations in Drosophila (Sparrow et al. 2003). For instance, downregulation of dendrite morphogenesis and neuron projection could be associated with actin defects in the nervous system. Another relevant finding was the downregulation of receptor genes involved in apoptosis, in particular TNFs (SI Table S5). When apoptotic mechanisms are suppressed, abnormal cell growth leads to tumors (Lowe & Lin 2000). Past studies have shown tumor growth after oil spills (Suchanek 1993), including evidence for neoplastic tumors in clams following the 1971 Maine oil spill (Barry & Yevich 1975). Genotoxic and cytotoxic effects, mainly chromosomal aberrations known to be associated with several types of cancer (Mitelman et al. 2007), have also been witnessed in whale fibroblasts exposed to oil-dispersant (Wise et al. 2014). The carcinogenic effects of PAHs are well documented (Phillips 1999), and our findings suggest there could a link between OD exposure in crabs, tumor necrosis receptor/apoptosis downregulation, tumor growth, and chromosomal abnormalities observed in previous studies.
In addition to downregulated features, we identified a limited number of upregulated features related to muscular structure and development in OD. This suggests that crabs may be undergoing damage-repair mechanisms, or that muscle upkeep-related genes far outweigh immune system genes due to relative abundance normalization. Interestingly, the same features were downregulated in NC (Fig. 5), indicating upregulation is not involved in general muscle maintenance under control conditions.

Our GO enrichment analyses revealed a significant proportion of fundamental biological processes, cellular components, and molecular functions are overrepresented between controls and dispersed oil, further suggesting crabs could be experiencing a generalized shutdown following chemical oil spill cleanup. Although enrichment was not significant between controls and oil, and oil and dispersed oil, our results point towards an increase in differential gene expression after dispersion. On the other hand, after controlling for droplet size OO/WAFs and OD/CEWAFs have similar gene expression effects in North Atlantic cod larvae in vitro (Olsvik et al. 2012). This discrepancy between OO and OD effects could be due to multiple sources of variation, including biological response between different taxa, different oil, and different analytical methods. These contrasts showcase the need for more studies as generalized oil spill conclusions based on a handful of studies are quite limited.

The role of cytochrome P-4501 (CYP1As) proteins as universal proxy for hydrocarbon contamination

One known link between hydrocarbon exposure and signaling pathways is the production of cytochrome P-4501A (CYP1As) in clan-1, a protein crucial in the aryl-hydrocarbon receptor (AHR) signaling pathway. CYP1A is commonly used to test for toxic
hydrocarbon exposure (Petrulis et al. 2001), because its expression is linked with AHR-
binding and metabolism of contaminants such as PAHs, polychlorinated biphenyl (PCB) in vertebrate systems (Petrulis et al. 2001; Whitehead et al. 2012). Fish populations naturally and experimentally exposed to DWH oil showed CYP1A upregulation (Brown-Peterson et al. 2015; Whitehead et al. 2012). However, major differences exist between vertebrate and invertebrate AHR pathways in that vertebrates have sophisticated aromatic binding capabilities (Barron 2012). While AHR pathways exist in most animals, a past review found dioxins and PAHs do not bind to invertebrate CYP1A-AHRs (Hahn 2002) although there are recent evidences of CYP1A upregulation in oysters (Jenny et al. 2016) and in copepods (Han et al. 2016; Hansen et al. 2016a; Hansen et al. 2016b; Olsvik et al. 2012). We found no evidence of CYP1A upregulation in crabs, congruent with past studies using similar species (Yednock et al. 2015). The lack of AHR regulation mediated by CYP1A in crabs could be due to a lower sensitivity to PAHs to those particular receptors or, alternatively, CYP1A protein sequences in crabs may be too divergent to be annotated. Moreover, CYP1A differential gene expression could be restricted to other organs’ specific responses not evident in muscle tissues. However, tissue specificity artifacts seem unlike due to a similar absence of CYP1As in blue crab hepatopancreas and gills exposed to oil (Yednock et al. 2015). Future research including dispersed oil and multiple tissue types will help assess crabs’ response to those seemingly universal toxicology markers. Interestingly, we did find differential gene regulation in other CYPs (clans 3, 4 and 18) used by invertebrates to metabolize aromatic toxins (SI Table S5, (Rewitz et al. 2006); however, these proteins were downregulated in OD presence. Nonetheless, in light of our findings, the continued use of CYP1A upregulation as a
universal proxy for PAH, PCB, or dioxin exposure could have limited utility when used in decapod crustaceans being important fisheries. We advocate for screening additional CYPs as a complementary proxy for PAH exposure.

**Crustacean immune response pathway downregulation and dispersed oil**

In addition to AHRs signaling pathway, we identified the suppression of several immune response receptors (FcRs, IKKs, T cells, and TLRs) known to be conserved across animals (Ottaviani et al. 2007). Our findings identified nine TLRs that were downregulated in OD, compared to a past study that found only three in an equivalent OO (Yednock et al. 2015), suggesting increased immune repression in the presence of OD in crustaceans. Conversely, we found upregulation for the hemocyte proliferation pathway in OD. Hemocytes are the first line of defense to infection in invertebrates (Barron 2012; Hamoutene et al. 2004), suggesting crabs are combating stress related to pathogens. Several studies showed that exposure to oil resulted in an increased susceptibility to disease and compromised reproductive, sensory, and neurological systems. Therefore, exposure to chemically dispersed oil may exacerbate effects. For example, fish species exposed to the Exxon Valdez spill had a 17-fold parasite load increase (Khan 1990) and tuna fish impacted by DWH had impaired heart contractions (Brette et al. 2014). In invertebrates (see references in (Suchanek 1993): hydroids and jellies showed teratogenic effects and neurological changes; hermatypic corals had decreased ovaria and less planula per polyp, ova degeneration, and lack of gonad development; kelp crabs presented chemosensory-induced bradycardia and suppressed chemoreception abilities; sea urchins showed delayed embryogenesis, asynchronism, non-viable larvae, and functional loss of tube foot and spine movement; multiple bivalves
exhibited reduced respiration and increased energy expenditures. Specifically in *E. depressus*, occurrence of supernumerary and morphologically abnormal megalopal stages was shown to be associated with crude oil (Cucci & Epifanio 1979). Our results identify several immune response pathways that could be playing a role in the adverse effects witnessed in other marine organisms (Peterson et al. 2003; Wise & Wise 2011) and humans involved in oil spill clean-up (D'Andrea & Reddy 2013). This is of particular importance because human lung cell cultures showed similar immune response impact to gene expression when exposed to chemically dispersed oil (Liu et al. 2016). Therefore, we suggest these receptors, such as TLRs, could serve as potential candidates as complementary biomarkers to CYPs for aquatic toxicology.

We did not verify analytically concentrations of hydrocarbons prior and after experiments. In acute non-renewal exposure it is possible than many toxic compounds and oxygen are depleted due to consumption or transformation, leading to confounding reasons for gene expression differences. All changes we document in this study are consistent with lethal and sublethal effects of chemically dispersed oil in previous toxicological (DeLeo et al. 2015; Echols et al. 2016; Wise et al. 2014) and transcriptomic studies (Jenny et al. 2016; Liu et al. 2016; Bayha et al. 2017) where detailed analytical data is available. Moreover, the lack of significant changes between aerated controls and non-aerated treatments suggest that low oxygen is not driving gene expression. Future transcriptomic analyses coupled comprehensive analytical testing and renewal exposures will shed light into more precise points of concentrations and gene expression correlations beyond the baseline changes we report presently.
The need for developing less to non-toxic dispersants is pressing. Current and future oil exploration contingency in the U.S.A. (Bureau of Safety and Environmental Enforcement 2015; Shell Offshore Inc 2010) relies on the same dispersant, COREXIT®9500, that our results and number of other studies (e.g., Hemmer et al. 2011; Paul et al. 2013; Rico-Martínez et al. 2013; Wise & Wise 2011; Bayha et al. 2017), have shown to be highly detrimental to animal health particularly in systems exposed to xenobiotic bioaccumulation. Although COREXIT®9500 is one of the relatively less toxic formulations vetted by the United States Environmental Protection Agency (Hemmer et al. 2011), future research needs to find safer solutions for spill cleanup and explore the longer-term consequences of oil and dispersant impacts on marine community health.

**CONCLUSIONS**

We examined changes in gene expression resulting from dispersed oil exposure using a non-model decapod crustacean. We identified a large set of vital cellular functions and components that are suppressed in the presence of chemically dispersed oil. Our results suggest that exposure to mixtures of petroleum hydrocarbons and dispersants may lead to generalized cellular shutdown. While exposure to oil-only and oil-dispersant mixes resulted in similar effects in gene expression, samples exposed to oil-dispersant mixes were generally an order or magnitude more intense. Our results indicate crabs retain little ability to maintain fundamental cellular structure and defense against pathogens that could help explain well-known health declines and mortality increases following oil spills. Finally, finding a concordant suite and distribution of genes affected by dispersed oil reported in previous studies supplies additional evidence for not relying solely on traditional metrics to evaluate RNA-seq assemblies.
ACKNOWLEDGMENTS

Computational analyses performed on the HPC Panther Cluster at Florida International University (FIU) with assistance from C. D’Cunha and M. Cheng. J. Pérez-Moreno assisted with custom scripts. A. Smith provided valuable comments to improve this manuscript. We thank J. Felder for help collecting specimens for this research. This is contribution # (pending) of the Marine Education and Research Center (FIU MERC) of the Institute for Water and the Environment at FIU.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This research was made possible in part by a grant from The Gulf of Mexico Research Initiative, and in part by Florida International University.

Competing interests.

Keith A. Crandall is an associate editor in PeerJ. The authors declare no competing interests.

Additional Information

All housing and experimental procedures were authorized and conducted following the ethics committees’ standards of the University of Louisiana at Lafayette and Florida International University. The authors declare no conflict of interests.

Author Contributions

-Hernán Vázquez-Miranda performed laboratory procedures, analyzed the data, wrote the paper with help of coauthors, prepared figures/tables, reviewed drafts of the paper.
- Brent P. Thoma and Juliet M. Wong performed experimental treatments and laboratory procedures, reviewed drafts of the paper.

- Brent P. Thoma, Darryl L. Felder, Keith A. Crandall, and Heather D. Bracken-Grissom designed experiments, secured funding, contributed reagents/materials/analysis tools, reviewed drafts of the paper.

**Data deposition**

Annotated transcriptome assembly and additional data are publicly available through the Gulf of Mexico Research Initiative Information & Data Cooperative (GRIIDC) at [https://data.gulfresearchinitiative.org/data/R4.x257.228:0013](https://data.gulfresearchinitiative.org/data/R4.x257.228:0013). (DOI: http://dx.doi.org/10.7266/N71C1TZC). The raw data used in this study is available at the NCBI website under BioProject ID: PRJNA376168 (https://www.ncbi.nlm.nih.gov/bioproject/376168), BioSamples SAMN06351232-SAMN06351246. Transcriptome Shotgun Assembly (TSA) has been deposited at DDBJ/EMBL/GenBank under the accession GFJG00000000 (version 1.0: GFJG01000000). TSA file prepared with Transvestigator at [http://doi.org/10.5281/zenodo.10471](http://doi.org/10.5281/zenodo.10471)

**Supplemental Information**

Supplemental information for this article can be found online at (DOI: to be included)

**REFERENCES**


Havird JC, and Santos SR. 2016. Here we are, but where do we go? a systematic review of crustacean transcriptomic studies from 2014–2015. *Integrative and Comparative Biology*:icw061.

Hemmer MJ, Barron MG, and Greene RM. 2011. Comparative toxicity of eight oil dispersants, Louisiana sweet crude oil (LSC), and chemically dispersed LSC to two aquatic test species. *Environmental Toxicology and Chemistry* 30:2244-2252.


R Core Development Team. 2010. The R project for statistical computing. Vienna, Austria.


Table 1. Environmental and experiment information for the flatback mudcrab transcriptome compliant with the MIGS standard.

<table>
<thead>
<tr>
<th>Item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Investigation_type</td>
<td>Eukaryote</td>
</tr>
<tr>
<td>Project_name</td>
<td>Transcriptomic effects of dispersed oil in <em>Eurypanopeus depressus</em></td>
</tr>
<tr>
<td>Collection_date</td>
<td>09-July-2013 and 23-September-2014</td>
</tr>
<tr>
<td>Lat_lon</td>
<td>29.521667 N 94.502100 W</td>
</tr>
<tr>
<td></td>
<td>27.436667 N 80.335556 W</td>
</tr>
<tr>
<td>Country</td>
<td>USA (Texas and Florida)</td>
</tr>
<tr>
<td>Environment</td>
<td>Oyster beds</td>
</tr>
<tr>
<td>Geo_loc_name</td>
<td>Rollover Bay, Texas (USA)</td>
</tr>
<tr>
<td></td>
<td>Fort Pierce, Florida (USA)</td>
</tr>
<tr>
<td>env_biome</td>
<td>ENVO_00002030 (aquatic biome)</td>
</tr>
<tr>
<td>env_feature</td>
<td>ENVO_00000485 (sea shore)</td>
</tr>
<tr>
<td>env_material</td>
<td>ENVO_00002149 (sea water)</td>
</tr>
<tr>
<td>Ref_biomaterial</td>
<td>Primary transcriptomic report</td>
</tr>
<tr>
<td>biotic_relationship</td>
<td>free living</td>
</tr>
<tr>
<td>Rel_to_oxygen</td>
<td>Aerobe</td>
</tr>
<tr>
<td>Ploidy</td>
<td>Diploid</td>
</tr>
<tr>
<td>Sequencing_meth</td>
<td>Illumina HiSeq2000</td>
</tr>
<tr>
<td>Num_replicons</td>
<td>NA</td>
</tr>
<tr>
<td>Assembly</td>
<td>Trinity2.0.3</td>
</tr>
<tr>
<td>Finishing_strategy</td>
<td>Draft</td>
</tr>
<tr>
<td>-------------------</td>
<td>-------</td>
</tr>
<tr>
<td>Estimated_size</td>
<td>2.80E9</td>
</tr>
<tr>
<td>Sample_material</td>
<td>muscle</td>
</tr>
<tr>
<td>Motility</td>
<td>Yes</td>
</tr>
<tr>
<td>Assembly method</td>
<td>Trinity2.0.3</td>
</tr>
<tr>
<td>Assembly name</td>
<td>de Novo</td>
</tr>
<tr>
<td>Sequencing technology</td>
<td>Illumina HiSeq2000</td>
</tr>
</tbody>
</table>

958
Figure 1. Sampling sites, Deepwater Horizon Oil Spill (DWH) area, species of interest, and experimental design. (A) We collected samples in two sites not affected by DWH [light blue area](http://invertebrates.si.edu/boem/maps/BOEM_NMNH_GOM_collections.kmz); map credit: created in Mapbox v2015-09-26-00-46-36 by B.P.T. ([http://www.mapbox.com](http://www.mapbox.com)); map data: OpenStreetMap© ([http://www.openstreetmap.org/copyright)] of the (B) flatback mudcrab *Eurypanopeus depressus* (Photo credit: B.P.T.). (C) Our experimental design (Image credit: B.P.T.) consisted of two controls: aerated (AC), negative (NC); two experimental treatments: oil-only (OO), oil-dispersant (OD).
Figure 2. Assembly quality and transcriptomic completeness. (A) ExN50 distribution pondering N50 contig length by percentage expressed (Pct. Ex); and (B) assembly completeness based on Benchmarking Universal Single-Copy Orthologs (BUSCOs). Each bar represents ORTHODB ortholog assessment in a hierarchical taxonomic order from Eukaryota to Insecta. Legend: complete genes in blue, missing genes in red, and in yellow fragmented genes.
Figure 3 Contig size density distribution and variability in decapod crustaceans and model arthropods. (A) Distribution of contig size in true crabs [Decapoda: Brachyura], (B) Distribution of contig size in shrimp [Decapoda: Dendrobranchiata & Caridea], and (C) in other arthropods including model organisms fruit fly and water flea. Mudcrab assembly (Eurypanopeus depressus) constructed with 15 samples is marked with an asterisk (*); every other assembly includes a single Genbank SRA sample. All species were assembled in TRINITY with
the exact same parameters in the same HPC cluster. N50 values represent at least
50% of contigs in an assembly are a given size in bps or larger. Dashed vertical lines indicate a 1000 bps contig size threshold. Genes correspond to number of contigs assembled by TRINITY.
Figure 4. Pairwise differential gene expression MA and volcano plots between the negative control (NC) and experimental treatments (OO and OD). In MA plots X-axis represents mean average expression on counts on a log2 scale and Y-axis represents fold-changes in gene expression on a logarithmic scale. In volcano plots X-axis represents fold-changes in gene expression on a logarithmic scale, and Y-axis represents pairwise t-test negative log10 P-values Red dots represent significant comparisons passing a 1% false discovery rate (FDR).
Figure 5. Differential expression and regulation of genes estimated in DESeq2.

Columns correspond to treatments: NC (negative control) in green, OO (oil-only) in cyan, and OD (oil-dispersant) in purple. The “R” label below columns represents replicates (Table S1). Rows correspond to significant genes passing a 1% FDR. Numbers on the left-side dendogram are feature clusters 1 and 2. Branch lengths on dendrogram correspond to Euclidean distances. Heatmap colors represent regulation direction from normalized, log-transformed fragments per kilobase of transcript per million mapped reads (FPKM) and zero-centered fold-change values: blue cells are downregulated genes, orange cells are upregulated genes, and grey values show no differential expression. For
all comparisons see Figs. S7-8. List on right side are gene ontology (GO) terms discussed in text (full list: Table S4).
Figure 6. Gene ontology enrichment analyses of differentially expressed (DE) genes estimated in GOseq. Pairwise comparisons of DE gene enrichment tests between NC (negative control), OO (oil-only), and OD (oil-dispersant) treatments listed by statistical method (edgeR and DESeq2). Only 11 out of 24 possible comparisons included DE genes (DEGs) and thus were tested for enrichment. Each test includes method, pairwise comparison, and whether an experimental condition included overrepresented (enriched – enrich) or underrepresented (depleted – deplet) gene ontology (GO) terms. Ontology
categories are: Biological Process (BP), Cellular Component (CC), and Molecular Function (MF). Statistically significant enrichment counts ($P < 0.05$) passing a 1% FDR are marked with an asterisk (*). Bar plots with no count numbers were not significant.
Supporting Information

Transcriptomic effects of dispersed oil in a non-model decapod crustacean

Supplementary Results

**Differential gene expression in aerated control (AC) and non-aerated negative control (NC).** We found relatively few genes differentially expressed between aerated controls (AC) and non-aerated controls (NC; Table S2) potentially representing the effects of low oxygen and experimental manipulation. Only three genes passed the 1% false discovery rate (FDR) on the MA plots in edgeR. When replicates are counted separately in gene expression heatmaps, edgeR found no significant comparisons passing the 1% FDR (Fig. S2). With DESeq2 between 26 and 33 features were significantly expressed (Table S2, Fig. S2). The upregulated features under presumed low oxygen in DESeq2 were ATP, calcium, and metal ion binding on the A-band muscle component, glutathione metabolic process and transferase activities in the cytoplasm and cytosol, and chitin binding and metabolic processes. Calcium, metal ion and ATP binding are necessary for energy transfer and protein activation (Knowles 1980), glutathione is a critical antioxidant (Pompella et al. 2003), and chitin is an integral component of arthropod exoskeletons. Approximately 76% of upregulated features did not BLAST to a GO. GOs of downregulated features matched ribosomal binding proteins, endonucleases, tyrosine and calmodulin phosphatases and zinc ion binding on the membrane, cytoplasm, nucleus, and ubiquinone reductases on mitochondrial respiratory chain complex and mitochondrial membrane. Most of these processes are involved in multiple parts of cellular upkeep, signaling, protein activation and synthesis. Ubiquinone reductases are critical factors in the respiratory electron transport chain responsible for cellular energy production and antioxidant protection (Mellors & Tappel 1966). Close to 77% of downregulated features under presumed low oxygen conditions did not blast to a GO.
Differential gene expression in non-aerated negative control (NC) and experimental treatments (OO-OD), and differences between statistical methods. We measured differential gene expression between the control, oil-only, and oil-dispersant treatments using edgeR and DESeq2. The main difference between those two statistical methods was the amount of significantly expressed features, with DESeq2 detecting larger amounts with relatively lower significance magnitude. We report numbers from DESeq2 first and numbers from edgeR second in this supplementary materials file SM (DESeq2-only numbers appear on the main text).

DESeq2.- Comparisons of controls and treatments detected hundreds of genes and thousands of isoforms that were differently expressed. Approximately 50% of these differentially expressed features in our de novo transcriptome (Database DS1) did not match any annotation and were left as unknown by either lack of database references or by being putative proteins with missing gene ontology (GO). There were only five differentially expressed features shared between controls and experimental treatments (Fig. S1). None of these shared features blasted to a GO. Most of the thousands of differentially expressed genes and isoforms were detected in the presence of oil-only and oil-dispersant treatments.

When comparing NC and experimental treatments (OO-OD), we detected a range of 11-4836 features with significantly differential expression (Figs. S3-S5, Table S3). When comparing NC to OD, we identified 4388 genes and 3243 isoforms (Figs. 2, S3-S5, Table S3). When comparing NC and OO (Fig. 2, Table S3) we found 11 genes and 8 isoforms, and 448 genes and 215 isoforms between OO and OD (Fig. 2, Figs. S3-S5, Table S3).

edgeR.- Comparisons of controls and treatments detected hundreds of genes and tens of isoforms that were differently expressed. Similar to DESeq2 counts, approximately 50% of these
differentially expressed features in our de novo transcriptome (Database DS1) did not match any GO annotation. There were only three differentially expressed features shared between controls and experimental treatments (Fig. S1). None of these shared features blasted to a GO. Most of the hundreds of differentially expressed genes and isoforms were detected in OO and OD.

Between NC and experimental treatments (OO-OD), we detected a range of 1-172 features with significantly differential expression (Figs. S3-S5, Table S3). When comparing NC to OD, we identified 172 genes and 87 isoforms (Figs. S3-S5, Table S3). When comparing NC and OO (Fig. 2, Table S2) we found zero genes and zero isoforms, and 1 gene and zero isoforms between OO and OD (Fig. 2, Figs. S3-S5, Table S3).

Although thousands of significantly expressed features were only detected by DESeq2, several hundred genes and isoforms were also found with edgeR (Fig. S6, Table S4).
Figure S1. Venn diagrams of shared differentially expressed features between controls (AC-NC) and experimental treatments (OO-OD). (A) Trinity ‘genes’, (B) Trinity isoforms. AC is the aerated, negative control, NC is the non-aerated, negative control, OO is the non-aerated oil-only treatment, and OD is the non-aerated oil-dispersant treatment. Circles represent significantly expressed features passing an FDR of 1%, detected by edgeR in blue including non-aerated control and treatments NC-OO-OD, features detected by DESeq2 including non-aerated control and treatments NC-OO-OD in green, and features detected by DESeq2 including aerated, negative controls AC and non-aerated, negative controls NC in red. There were zero features
passing the FDR of 1% including AC-NC controls in edgeR and thus were not plotted. Warm colors (C) correspond to upregulated features between NC and experimental treatments (OO-OD) across all statistical comparisons using OD as a reference, and cool colors (D) to downregulated features. Regulation direction occurs in the opposite sense (inverted colors) in features on NC (non-aerated control). The suffix “gn” corresponds to Trinity ‘genes’ and the suffix “tr” to isoforms. Color palettes match heatmaps in this paper.
Fig. S2. Differential feature expression in low oxygen conditions between aerated (AC) and non-aerated (NC) controls. A) Left column boxes are pairwise MA plots where X-axes represent mean average expression on counts on a logarithmic scale Y-axes represent fold-changes in gene expression on a logarithmic scale. Right column boxes are Volcano plots where X-axes represent fold-changes in gene expression on a logarithmic scale and Y-axes represent the probability of each pairwise t-test gene comparison between treatments. Red dots represent significant comparisons passing an FDR of 1%. B) and C) are heatmaps representing direction of gene regulation for genes that passed the 1% FDR for Trinity ‘genes’ and isoforms respectively from DESeq2. “R” labels represent individual replicates (Supplementary
Information Table S1). Warmer colors (gold) indicate upregulated features and cool colors (purple) downregulated genes for log$_2$ fold-changes. Colors in black indicate zero-centered values. There were no significantly expressed genes found in edgeR passing the 1% FDR for heatmap generation.
Fig. S3. Pairwise differential isoform expression MA and volcano plots between the negative control (NC) and experimental treatments (OO and OD). In MA plots X-axis represents mean average expression on counts on a log$_2$ scale and Y-axis represents fold-changes in gene expression on a logarithmic scale. In volcano plots X-axis represents fold-changes in gene expression on a logarithmic scale, and Y-axis represents pairwise t-test negative log$_{10}$ P-values Red dots represent significant comparisons passing a 1% false discovery rate (FDR).
Fig. S4. Differential expression and regulation of features for all experimental treatments related to presence of oil and dispersant ordered by expression profile distances. Columns and top dendrogram correspond to treatments with three replicates each: NC (non-aerated control) in green, OO (non aerated with oil) in cyan, and OD (non aerated with oil and dispersant) in purple. “R” labels represent individual replicates (Table S1). Rows and left-side dendrogram correspond to features with statistical significance passing a FDR of 1%. Numbers on the left-side dendrogram branches correspond to feature clusters 1 and 2 mentioned in the main text. Branch lengths on both dendrograms correspond to gene expression Euclidean distance.
distances. Heatmap colors represent regulation direction from normalized, log-transformed FPKM and centered fold-change values: blue cells indicate down-regulated genes, orange cells correspond to up-regulated genes, and grey values show no differential expression from a log2 count value centered at zero.
Fig. S5. Gene ontology annotations for genes from edgeR in oil-dispersant (OD). Colors in bar represent major upregulated (A; warm colors) and (B; cool colors) downregulated GO terms (see legend). X-axis indicates element count found in the database. Note last scale number is adjusted to the largest count value. GO detailed functions are arranged in reversed count number and then alphabetical order. Colors correspond to heatmaps in Figures 3, Extended Data Fig. S4. For full annotations see supplementary files DS2 DS6 and DS10.
**Fig. S6. Gene ontology annotations for upregulated isoforms from edgeR in oil-dispersant (OD).** Colors in bar represent major upregulated (A; warm colors) and (B; cool colors) downregulated GO terms (see legend). X-axis indicates element count found in the database. Note last scale number is adjusted to the largest count value. GO detailed functions are arranged in reversed count number and then alphabetical order. Colors correspond to heatmaps in Figures 3, Extended Data Fig. S4. For full annotations see supplementary files DS3, DS7 and DS11.
Fig. S7. Gene ontology annotations for genes from DESeq2 in oil-dispersant (OD). Colors in bar represent major upregulated (A; warm colors) and (B; cool colors) downregulated GO terms (see legend). X-axis indicates element count found in the database. Note last scale number is adjusted to the largest count value. GO detailed functions are arranged in reversed count number and then alphabetical order. Colors correspond to heatmaps in Figures 3, Extended Data Fig. S4. For full annotations see supplementary files DS4, DS8 and DS12.
Fig. S8. Gene ontology annotations for upregulated isoforms from DESeq2 in oil-dispersant (OD). Colors in bar represent major upregulated (A; warm colors) and (B; cool colors) downregulated GO terms (see legend). X-axis indicates element count found in the database.

Note last scale number is adjusted to the largest count value. GO detailed functions are arranged in reversed count number and then alphabetical order. Colors correspond to heatmaps in Figures 3, Extended Data Fig. S4. For full annotations see supplementary files DS5, DS9 and DS13.
**Supplementary information databases.** Files and spreadsheets available through the Gulf of Mexico Research Initiative Information & Data Cooperative (GRIIDC) at [https://data.gulfresearchinitiative.org/pelagos-symfony/dataset-submission?regid=R4.x257.228:0013](https://data.gulfresearchinitiative.org/pelagos-symfony/dataset-submission?regid=R4.x257.228:0013). (DOI:10.7266/N71C1TZC.)

DS1 – Transcriptome assembly in Trinity format and spreadsheet with complete transcriptomic annotations

DS2 – spreadsheet with significant regulated genes from edgeR

DS3 – spreadsheet with significant regulated isoforms from edgeR

DS4 – spreadsheet with significant regulated genes from DESeq2

DS5 – spreadsheet with significant regulated isoforms from DESeq2

DS6 – spreadsheet with upregulated genes’ GOs from edgeR

DS7 – spreadsheet with upregulated isoforms’ GOs from edgeR

DS8 – spreadsheet with upregulated genes’ GOs from DESeq2

DS9 – spreadsheet with upregulated isoforms’ GOs from DESeq2

DS10 – spreadsheet with downregulated genes’ GOs from edgeR

DS11 – spreadsheet with downregulated isoforms’ GOs from edgeR

DS12 – spreadsheet with downregulated genes’ GOs from DESeq2

DS13 – spreadsheet with downregulated isoforms’ GOs from DESeq2

DS14 – spreadsheet collection of pairwise comparisons of feature counts ranked by log2 fold change for DESeq2 and edgeR

DS15 – spreadsheets with GO term enrichment analyses with GOseq, based on gene count matrices from edgeR and DESeq2
References