

5-2017

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

Hernan Vazquez-Miranda

Brent Thoma

Juliet Wong

Darryl Felder

Keith A. Crandall

George Washington University

See next page for additional authors

Follow this and additional works at: https://hsrc.himmelfarb.gwu.edu/smhs_centers_cbi

 Part of the [Computational Biology Commons](#), [Research Methods in Life Sciences Commons](#), and the [Structural Biology Commons](#)

APA Citation

Vazquez-Miranda, H., Thoma, B., Wong, J., Felder, D., Crandall, K. A., & Bracken-Grissom, H. (2017). Transcriptomic effects of dispersed oil in a non-model decapod crustacean. *PeerJ Preprints*, 5 (). <http://dx.doi.org/10.7287/peerj.preprints.2977v1>

This Journal Article is brought to you for free and open access by the Institutes, Centers, and Laboratories at Health Sciences Research Commons. It has been accepted for inclusion in Computational Biology Institute by an authorized administrator of Health Sciences Research Commons. For more information, please contact hsrc@gwu.edu.

Authors

Hernan Vazquez-Miranda, Brent Thoma, Juliet Wong, Darryl Felder, Keith A. Crandall, and Heather Bracken-Grissom

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.

16 **ABSTRACT.**

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

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

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

34 **Discussion.** Our results provide evidence that exposure to chemically dispersed oil causes
35 a generalized cellular shutdown and muscular repair attempts. Our results suggest current
36 oil-spill treatment procedures could be detrimental to crustaceans and indicate additional
37 research is needed to evaluate the impact of oil spills in gene expression. Finally,

38 traditional quality metrics such as N50s have limitations to explain the nature of RNA-
39 seq compared to new methods in non-model decapod crustaceans.
40

41 INTRODUCTION

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

56 Understanding the effect of dispersed oil on organisms is critical as the most toxic
57 compounds of crude oil, polycyclic aromatic hydrocarbons (PAHs), are released in large
58 amounts after dispersion (NRC 2005). PAHs and other petroleum hydrocarbons
59 accumulated in high concentrations following the DWH: 500 parts per million (ppm) in
60 water and 200,000 ppm in sediments (Sammarco et al. 2013). While early toxicology
61 tests have suggested that dispersant formulations used in DWH are less lethal than oil
62 itself (Clark et al. 2001), few recent studies have revealed that dispersants have
63 significant toxic effects on marine life (DeLeo et al. 2015; Paul et al. 2013; Wise & Wise

64 2011) and can be more than 50 times more potent when mixed with oil than either oil or
65 dispersant alone (Rico-Martínez et al. 2013). For example, corals show more severe
66 health declines in both dispersant-only and oil-dispersant mixes, compared to oil-only
67 treatments (DeLeo et al. 2015). Dispersant presence impaired survivorship, and muscle
68 configuration and contractions showed sublethal toxicity in jellyfish (Echols et al. 2016).
69 Likewise whale fibroblast show more chromosomal aberrations (Wise et al. 2014).
70 Additionally, oil-dispersant toxicity can persist for months in the water column and
71 sediments (Sammarco et al. 2013), perpetuating detrimental effects.

72 Transcriptomics allows exploring and assessing oil spill effects for genes
73 expressed at the time of exposure. For example, if genomic references exist, microarrays
74 and targeted quantitative polymerase chain reaction (qPCR) allow for screening hundreds
75 to thousands of genes known to be affected by oil exposure (Bowen et al. 2016; Hansen
76 et al. 2016b; Olsvik et al. 2011; Olsvik et al. 2012). While genomic references are
77 lacking for most marine organisms, next generation sequencing (NGS) technologies
78 allow for using transcriptomic approaches in any study system. RNA-seq, the direct
79 massive-parallelized sequencing of all expressed genes at a given time (Wang et al. 2009),
80 can thus be applied to understanding oil spill effects in non-model organisms. In recent
81 years studies assessing immune response and other key biological processes at the gene
82 level have employed RNA-seq to examine the effects of mechanically dispersed crude oil
83 (Whitehead et al. 2012; Yednock et al. 2015; Zhu et al. 2016) but unfortunately there is
84 limited information on the consequences of chemically dispersed oil exposure (Jenny et
85 al. 2016; Liu et al. 2016; Bayha et al. 2017). Here, we examined the impact of oil and
86 dispersant exposures on a non-model decapod crustacean to test whether chemically

87 dispersed oil increases detrimental effects on gene expression beyond oil-only exposure.
88 Due to the increased availability of harmful compounds to animal tissues after oil
89 dispersion, we predict key biological processes will be severely impacted. Our objective
90 was to find changes in gene expression in a controlled setting reducing variation sources,
91 instead of aiming to replicate field conditions that are too complex to simulate *in vitro*
92 and to avoid multiple sources of environmental variation.

93 We used RNA-seq to quantify gene expression in oyster bed-associated flatback
94 mudcrabs (*Eurypanopeus depressus*) exposed to oil-only and oil-dispersant treatments
95 (Fig. 1). We employed acute non-renewal exposures (Weber 1991) to generate changes in
96 gene expression, albeit our objective was not to perform a comprehensive
97 ecotoxicological study but rather a baseline transcriptomic comparison. Mudcrabs are an
98 ideal system to test oil spill effects because they are associated to oysters and benthic
99 sediments, for which larval stages and juveniles have shown detrimental effects from oil
100 (Cucci & Epifanio 1979) and dispersal exposures (Anderson et al. 2014; Vignier et al.
101 2016). We collected crabs from sites outside of the DWH area (Fig. 1A) to avoid prior
102 oil-spill exposure as a confounding factor. After a period of *in vitro* acclimatization, we
103 exposed crabs to the following conditions, each with 3 replicates (Fig. 1C): non-aerated
104 negative control (NC), non-aerated oil-only treatment (OO), and non-aerated oil-
105 dispersant treatment (OD). Additionally, we included an aerated control (AC) to identify
106 effects of oxygen stress, continuous oxygen supply, and experimental manipulation
107 potentially present when using closed, static non-renewal exposures. As the impact of
108 aeration appeared to have little effect on gene expression associated with oil or

109 dispersant, further comparisons AC are not considered here. (See Supplementary
110 Information *SI*.)

111

112 MATERIALS AND METHODS

113

114 Animals

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

127

128 Experimental design, conditions, and treatments.

129 Treatments were mechanically enhanced water-accommodated fraction (WAF)
130 and chemically enhanced water-accommodated fraction (CEWAF) following (Singer et
131 al. 2000) using oil from the Marlin platform Dorado, an oil surrogate with no discernible
132 differences in chemical composition to the MC252 oil from the DWH Macondo Prospect

133 (Worton et al. 2015). The WAF was generated using a loading rate of 1g/l and mixed in a
134 sealed container with minimal headspace and no visible vortex for 24 hrs before being
135 allowed to settle for 30 minutes. We tried multiple loading rates of dispersant (1:10, 1:20,
136 1:30, 1:40, 1:100) to determine which concentration could produce a CEWAF with
137 visible effects on crabs. Crabs exposed to concentrations lower than 1:50 did not show
138 motility and behavioral differences compared to controls after gentle prodding with blunt
139 forceps. Thus, the selected CEWAF had a final loading rate of 0.1g/l and a COREXIT®
140 9500 concentration of 1:50 to capture the responses of crabs to acute exposures. The
141 CEWAF was mixed in a sealed container with minimal headspace and a vortex extending
142 approximately 25% of the total depth of the container for 18 hours before being allowed
143 to settle for 6 hours. After being allowed to settle, 3.5l of WAF and CEWAF were
144 transferred to separate 3.75l vessels with care taken to not disturb the oil layer. We
145 utilized a closed, static design with minimal headspace to limit loss of volatile
146 compounds; vessels were sealed with foil-lined caps (OECD. 2002). Even though
147 industry standard concentrations are about 1:20, dispersant concentration in our CEWAF
148 (1000 parts per million - ppm) that generated an acute response falls within the lower
149 bound of bioaccumulated total petroleum hydrocarbon (TPHs) concentrations found in
150 Deepwater Horizon oil spill water and sediments (Sammarco et al. 2013). We did not
151 verify exposure concentration analytically. Due to the lack of analytical data, our putative
152 experimental concentration are likely lower than reported nominal loading rates due to
153 multiple factors during exposures.

154 In addition to the WAF, CEWAF, and a sealed, non-aerated control, we included
155 an aerated control to allow us to understand the potential impact of oxygen deprivation in

156 the closed design to disentangle hypoxia genes from chemical-response genes. By having
157 these two controls we were also able to assess potential gene expression effects due to
158 experimental manipulation. These conditions are henceforth: Aerated control “AC”, non-
159 aerated negative control “NC”, Oil-Only Treatment “OO” (non-aerated WAF), and Oil-
160 Dispersant Treatment “OD” (non-aerated CEWAF). Each treatment and controls were
161 replicated in separate exposure vessels per condition, 4 vessel replicates, and 3 crabs per
162 vessel maintained at $22\pm 2^{\circ}\text{C}$ for 72-h static, non-renewal exposures (Weber 1991) for a
163 total of 48 crabs. Crabs were not fed for the duration of the exposures. We were not able
164 to include a Dispersant-Only Treatment due to limitations on read sequencing per lane
165 (see Library Preparation) and budget constrains. Therefore the effects of dispersants
166 independent from the effects of dispersed oil were not surveyed (see Discussion).

167

168 **Tissue harvesting**

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

172

173 **RNA extraction, quality control, and RNA-seq Libraries**

174 We isolated total RNA from 48 individuals. Muscle was chosen over alternative
175 tissue-types that were found to be high in RNAses (i.e. hepatopancreas) or that yielded
176 insufficient amount of high quality RNA (i.e. gill). Tissues were lysed with a hand
177 homogenizer PRO200 (BioGen, Pro Scientific). We used the column-based Nucleospin®
178 RNA kit (Macherey-Nagel), and treated with DNase (Clontech) to extract RNA

179 following manufactures' instructions. Total RNA was eluted in 50 ul of RNase-free
180 water (Sigma) and stored at -80 °C. RNA integrity and concentration was determined in a
181 Bioanalyzer 2100 (Agilent). Muscle RNA of 12 individuals with the highest
182 concentration and quality were chosen for sequencing. These individuals comprise 12
183 samples in our experimental design: three repetitions of two controls (AC, NC) and two
184 experimental conditions (OO, and OD) [Table S1].

185 Our 12 selected samples showed some RNA degradation. Thus, to ensure a high
186 quality *de novo* transcriptome assembly we collected additional crabs from Florida (Fort
187 Pierce; 27.436667°N, 80.335556°W) on 23 September 2014. These crabs were kept in 5-
188 gallon tanks prior to RNA extraction. We extracted total RNA from muscle from three
189 Florida crabs with the Trizol® Reagent (Life Technologies) method. Approximately 10-
190 12 2mm ceramic beads were used in a MiniBead Beater (BioSpec) for homogenization in
191 Trizol®, followed by chloroform-isopropanol/ethanol precipitation including the DNase
192 and elution steps mentioned previously following manufacturer's guidelines.

193

194 **Library preparation**

195 Samples containing 1ug of total RNA from the four experimental conditions were sent to
196 the Vanderbilt University Core Lab in 2013 for sample preparation using the Ribo-
197 Zero™ rRNA Removal Kit and cDNA library generation (Epicentre) using a TruSeq
198 Stranded Library kit (Illumina). Library concentrations were verified with a Qubit®
199 Fluorometer (Invitrogen) and library sizes were verified on a Bioanalyzer. The resulting
200 12 libraries were then multiplexed and sent to Beckman Coulter Genomics for RNA
201 sequencing of paired-end 100 bps reads on one Illumina HiSeq 2000 lane aiming for 20

202 million reads per experiment as recommended for *de novo* assemblies (Francis et al.
203 2013). We did not multiplex additional individuals (i.e. a Dispersant-Only treatment) in a
204 single sequencing lane as it would have reduced the number of reads per needed per
205 individual.

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

215

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

217 Open source software packages and pipelines were employed for this project
218 bioinformatics' needs. All paired-end reads from 12 experimental and 3 additional
219 libraries, 15 total, were passed through quality control (QC) first by checking for
220 presence of adapters, indexes, repetitive kmers, and low-quality sequences in FASTQC
221 (Andrews 2010) to determine trimming parameters. Adapters, indexes, and low-quality
222 sequences were removed in TRIMMOMATIC 0.33 (Bolger et al. 2014) with the
223 following parameters: 30GB of total RAM memory, -six threads, a phred scores of 33,
224 cropping the first 9 bases, LEADING:30, TRAILING:30, SLIDINGWINDOW:4:15, and

225 minimum kmer length of 36. FASTQC was run again on processed read files to verify
226 that trimming was successful.

227

228 **De novo transcriptome assembly and annotation**

229 Due to the lack of an *E. depressus* reference genome we conducted a *de novo*
230 transcriptome assembly in TRINITY 2.0.3 (Haas et al. 2013) using paired-end reads that
231 passed QC from the 15 total libraries. All libraries were included to generate a single
232 transcriptome, as it is necessary for a reference to include all individuals to detect
233 differentially expressed genes across samples (Conesa et al. 2016; Haas et al. 2013).
234 TRINITY was run on the FIU Panther Cluster in the High Performance Computer (HPC)
235 environment with 24 cores and 256 GB of RAM with the following parameters:
236 minimum kmer coverage of 4, maximum memory 252 GB, reverse single-stranded
237 libraries, 24 CPUs, Butterfly maximum heap space 10GB, Butterfly initial heap space
238 10GB, Butterfly CPUs 24, and Inchworm CPUs 24.

239 We evaluated our assembly using traditional metrics including number of
240 transcripts and N50 values, and supplemented them with the novel ExN50 (B. Haas op
241 cit; [https://github.com/trinityrnaseq/trinityrnaseq/wiki/Transcriptome-Contig-Nx-and-](https://github.com/trinityrnaseq/trinityrnaseq/wiki/Transcriptome-Contig-Nx-and-ExN50-stats)
242 [ExN50-stats](https://github.com/trinityrnaseq/trinityrnaseq/wiki/Transcriptome-Contig-Nx-and-ExN50-stats)) with TRINITY perl scripts. Recently, N50 values to assess quality have
243 been deemed unreliable as they are often biased due to: (1) filtering artifacts and
244 combinatorial isoform assembly that skew values upwards, or conversely (2) skewed
245 downwards if assemblies are efficient at finding low-abundance rare splice variants
246 particularly with deep sequencing. Thus, pondering the expression levels of each contig
247 with the ExN50 is more appropriate (<http://www.molecularecologist.com/2017/05/n50->

248 for-transcriptome-assemblies/). We further checked for completeness according to
249 conserved ortholog content in BUSCO 3.0.0 (Benchmarking Universal Single-Copy
250 Orthologs; Simão et al. 2015) using annotated Eukaryota, Metazoa, Arthropoda, and
251 Insecta databases (since there are no decapod crustacean databases) from ORTHODB v9
252 (Waterhouse et al. 2013; <http://busco.ezlab.org/datasets/>). These databases include an
253 increasing number of orthologs as the taxonomic level gets more specific, i.e. there are
254 fewer orthologs shared by all eukaryots compared to the number of orthologs shared
255 between insects.

256 Additionally, we compared our *de novo* assembly with additional decapod
257 crustacean and model arthropod transcriptomes to test how much deviated our assembly
258 is from other taxa. For this purpose we downloaded and assembled nine single-sample
259 Genbank SRA raw reads: four additional true crabs (*Eriocheir sinensis*-SRR1735536;
260 *Portunus trituberculatus*-SRR768319; *Scylla olivacea*-SRR2440122; *Callinectes*
261 *sapidus*-SRR2140752), five shrimp (*Macrobrachium nipponense*-SRR3196792; *Caridina*
262 *rubella*-SRR1248238; *Neocaridina denticulate*-SRR1185328; *Penaeus monodon*-
263 SRR1648423; *Fenneropenaeus mergiensis*-SRR1756093), one butterfly (*Bicyclus*
264 *anyana*-ERR1022646) and two model arthropods (*Daphnia pulex*-SRR2350794;
265 *Drosophila melanogaster*-SRR2930822). We repeated the exact same TRINITY
266 parameters previously mentioned after read quality control on the same HPC cluster to
267 reduce sources of variation. Lastly we also calculated traditional N50s, contig (‘gene’)
268 number, contig length with TRINITY perl scripts, and created density plots of arthropod
269 assembly contig variation in R 3.1 (R Core Development Team 2010).

270

271 The resulting transcriptome was annotated with TRINOTATE 2.0.2
272 (<http://trinotate.sourceforge.net>) and TRINITY plug-in perl scripts on the same HPC
273 node. The plug-in TRANSDECODER was used to obtain likely protein-coding regions
274 and extract the longest open reading frames. Functional annotation was achieved by
275 comparing transcriptome contigs to peptide and transcript databases with BLAST
276 (Altschul et al. 1990), protein families with PFAM (Bateman et al. 2004) and
277 HMMER(Finn et al. 2011), signal peptides with SIGNALP (Petersen et al. 2011),
278 transmembrane proteins in TMHMM (<http://www.cbs.dtu.dk/services/TMHMM/>), and
279 remnant RNA with RNAMMER (Lagesen et al. 2007). Custom databases were
280 downloaded for TRINITY from the Broad Institute's website
281 (https://data.broadinstitute.org/Trinity/Trinotate_v3_RESOURCES/) for SwissProt,
282 UniProt90, and PFAM. Annotated contigs were compiled and summarized in a SQLITE3
283 database by translating the transcriptome genes to a transcript map using TRINITY's
284 TRANSMAP utility and loading those results into the SQLITE database. All annotations
285 blasted to a record with a 1×10^{-5} identification E-value for gene ontology (GO)
286 determination. Table 1 includes a brief summary of environmental and experimental
287 conditions used to assemble the flatback mudcrab transcriptome compliant with the
288 MIGS standard.

289 **Differential gene expression statistical analyses**

290 Reads passing QC for each replicate per condition (AC, NC, OO, OD) were
291 included in the differential gene expression estimates. Libraries from Florida crabs were
292 not included to avoid geographic, genetic, and sequencing artifacts (Gleason & Burton
293 2015). Pair-end reads were merged prior to alignment to the resulting *de novo*

294 transcriptome. Pair-end reads with no overlapping regions were concatenated with a
295 custom bash shell script to merged reads. Since single stranded libraries only contain the
296 reverse read, all reads were reverse-complemented using SEQTK
297 (<https://github.com/lh3/seqtk>). We did not incorporate RNA spike-ins as controls (see
298 above), therefore our reported values are relative to normalization between biological
299 replicates.

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

311 Differential gene expression (DE) on normalized counts from the experimental
312 reads was estimated on the BIOCONDUCTOR platform (Gentleman et al. 2004) in R 3.1
313 (R Core Development Team 2010) with packages edgeR (Robinson et al. 2010) and
314 DESeq2 (Love et al. 2014). The transcript and gene normalized count matrices were used
315 to calculate pairwise differential expression between each treatment from the mean and
316 standard deviation of each biological replicate to generate Bland–Altman plots (Bland &

317 Altman 1986) measuring logarithmic fold change in gene expression (M) over mean
318 average expression on log counts (A), known in microarray and RNAseq studies as MA
319 plots (Dudoit et al. 2002). Statistical significance was assigned on a \log_2 scale values
320 passing a False Discovery Rate (FDR) of 1%. Additionally, volcano plots were generated
321 to visualize gene expression fold change by their statistical significance (Cui & Churchill
322 2003). Parameters in edgeR were left as defaults. DESeq2 calculations were run with
323 comparisons per condition, a maximum sharing mode, and a local regression fit type.

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

339

340 **RESULTS**

341 We obtained 347,082,962 pair-end reads from our libraries with an average of 23 million
342 reads per replicate. TRINITY (Haas et al. 2013) produces sequence contigs categorized
343 as two *features*: ‘*genes*’ represent unique contigs, and ‘*isoforms*’ correspond to
344 alternatively spliced transcripts in a given contig. Our TRINITY assembly yielded a
345 transcriptome of 500,008 genes (i.e. unique contigs), 660,546 isoforms (i.e. alternatively
346 spliced transcripts in a given contig), with an N50 of 421 base pairs (bps) and a mean
347 length per contig of 428.03 bps. The total number of assembled bases was 282,733,223
348 (table S1). In contrast, between 60-80% ExN50 values were more than double the
349 traditional N50 (Fig. 2A). The level of conserved single-copy orthologs ranged between
350 70.2-88.3% complete, 13.9-2.3% missing, and 15.9-9.2% fragmented BUSCOs for
351 exclusive and inclusive taxonomic hierarchies, respectively (Fig. 2B).

352 Our comparative contig density plots revealed an unprecedented level of
353 variability across single-sample Arthropoda assemblies and our *de novo* transcriptome
354 (Fig. 3). TRINITY found fewest contigs in the model arthropods (*D. pulex* and *D.*
355 *melanogaster*, Fig. 3C). In non-model arthropods (crabs, shrimp, and butterfly) the
356 number of contigs ranged between 11,534-243,857. Shrimp ranged between 11,534-
357 158,098 contigs (Fig. 3B), and crabs had the largest between 86,287-243,857 contigs
358 (Fig. 3A). For every species the contig size distribution peaked between 400-600 base
359 pairs (bps). However several N50 values were above the 1000 bps traditional benchmark
360 inversely correlated with peaks’ height, i.e. the higher the density peak, the lower the N50
361 (Fig. 3). Two crabs, three shrimp, and *D. pulex* had N50s <1000. For model arthropods
362 and the butterfly, and in particular for *D. melanogaster*, a large portion of the density

363 distribution exceeded 1000 bps. The contig size density for two shrimp (*F. mergiensis*
364 and *N. denticulata*) and two crabs (*C. sapidus* and *S. olivacea*) partially exceeded 1000
365 bps. Our *de novo* transcriptome had a similar contig size distribution to two crabs
366 (*Eriocheir sinensis* and *Portunus trituberculatus*) and three shrimp (*Macrobrachium*
367 *nipponense*, *Caridina rubella*, and *Penaeus monodon*). The number of contigs in our
368 transcriptome produced with 15 samples was approximately double compared that of
369 *Eriocheir sinensis* with a single sample, and its N50 30.5% smaller.

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

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

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

385 experimental treatments but did not blast to any GO (*SI* Fig. S1). Nevertheless, 99% of
386 differentially expressed genes and isoforms were detected in OO and OD.

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

392 When comparing gene expression across replicates within treatments, replicates
393 within NC and OD had nearly identical heatmap profiles (Figs. 5, S4); however,
394 replicates within OO shared heatmap profile characteristics with both NC and OD (Figs.
395 5, *SI* S4). In 75% of heatmaps, 66% of OO replicates clustered more closely with OD (*SI*
396 Fig. S4) indicating that OO and OD replicates have similar gene expression profiles and
397 are statistically distinguishable from NC replicates.

398 Two clear feature clusters are present in all heatmaps: one cluster containing over
399 90% of differentially expressed features, upregulated in NC and downregulated in OD,
400 and a second cluster of less than 10% with the opposite pattern of regulation
401 directionality (clusters 1 and 2 respectively on Figs. 5, *SI* Fig. S4).

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

403 Gene ontology annotations allow for the identification of features and their functions.
404 Pairwise expression measured in log₂ fold changes (logFC, *SI* DS14) between
405 experimental conditions ranged -16 to 12. Most differentially expressed features (logFC
406 > |8|) did not blast to a known protein, to an undescribed protein, or matched muscle

407 components. Immune response features logFC ranged -4 to -2. Although gene expression
408 studies typically show gene rankings per fold change limiting detailed description of the
409 top 10-30 genes (See full rankings in *SI DS14*), we report herein transcriptome-wide GOs
410 with differential gene expression instead of a handful of gen-per-gene comparisons. 70%
411 of shared features detected as differentially expressed in NC and OD (*SI Fig. S1*) did not
412 BLAST to a GO and were left as unknown (*SI Figs. S5-8*). Nevertheless, 19 features
413 upregulated in OD (and thus downregulated in NC) were annotated with GO terms
414 including: cell division, proliferation, and regulation, as well as muscle structure,
415 attachment, development, and other functional components (*SI Table S4, SI Figs. S5-S8*).
416 More specifically, upregulation of structural muscular elements included sarcomeres, I-
417 bands, Z-discs, and muscle-associated actin binding responsible for muscle formation and
418 contraction. Mechanisms involved in cell division, including chromatid cohesion and
419 chromosome condensation, were upregulated. Lastly, we found one upregulated
420 invertebrate immune response mechanism (hemocyte proliferation) involved in the
421 protection against microbial infection (Schulz & Fossett 2005).

422 The largest GO annotation set included 151 features that were downregulated in
423 OD. These features were associated with A) fundamental cellular components including
424 the nucleus, membranes, cytoskeleton, organelles, and cytoplasm (*SI Table S4, SI Figs.*
425 *S5-S8*) B) molecular functions such as energy transfer by ATP for ubiquitines, metal ion
426 binding, mRNA processing, nucleotide binding, protein phosphorylation, stabilization,
427 proteolysis, DNA synthesis, RNA processing and binding, transcription, translation and
428 C) cellular processes such as apoptosis, apoptosis regulation, cell adhesion, cell migration
429 and protein development, cellular response to hypoxia, glucose and glycogen

430 metabolism, intracellular signal transduction, splicing speckles, transmembrane transport,
431 and the mitotic cell cycle. We found several downregulated actin elements involved in
432 microfilament formation for cytoskeleton structure and cell movement, locomotion,
433 including filament binding and assembly. Downregulated components of the nervous
434 system included dendrite morphogenesis, neuron projection, neuron apoptotic process,
435 and neuronal cell bodies. Three metabolic/signaling pathways downregulated in OD
436 included: A) general cell development and immune response pathway as indicated by
437 cytokine signaling; B) cell growth, proliferation, survival, and apoptosis pathways
438 including activation of mitogen-activated protein kinases (MAPK), SAPK/ Jun amino-
439 terminal kinases (JNK), epidermal growth factor receptor (EGFR), fibroblast growth
440 factor receptor (FGFR), and transforming growth factor- β receptor (TGF- β); and C)
441 protective response of the immune system pathways including Fc receptors (FcR), the
442 I κ B kinase (IKK) and NF- κ B receptors, T cell receptors, and Toll-like receptors (TLRs).
443 Lastly, in OD we found 3 and 19 downregulated isoforms of Tumor Necrosis receptors
444 (TNF) and Cytochrome P450 (CYP), respectively (table S5). Importantly, TNFs work as
445 natural defenses against tumors by killing abnormal cell growth (Fransen et al. 1985) and
446 the CYP protein family of membrane receptors are widely used as hydrocarbon exposure
447 biomarkers (Rewitz et al. 2006). These CYP transcripts belong to three protein clans
448 (clans 3, 4, and 18). Unknown features, not matching a functional GO, constituted 30%
449 and 8% of downregulated genes and isoforms, respectively.

450 **Gene ontology terms were statistically over and under represented between controls**
451 **and dispersed oil.**

452 GO term enrichment analysis tests for categories that dominate differential gene
453 expression. Only 11 out of 24 possible pairwise enrichment tests included DE genes
454 (DEGs; Fig. 6): 8 from DESeq2 and 3 from edgeR. In most cases Biological Processes
455 (BPs) were the dominant ontology, followed by Molecular Functions (MFs), and lastly
456 Cellular Components (CCs). Those DE GO terms were consistent with our cross-
457 reference list (see above; *SI* Table 4). DE counts between NC and OO totaled 20 GO
458 terms, and between OO and OD there were 470 GO terms. Pairwise DE counts between
459 NC and OD were much higher, totaling 3578 GO terms. However, significantly enriched
460 and depleted terms were only found in NC in NC-OD DESeq2 pairwise comparisons
461 after the FDR correction, totaling 1832 and 75 respectively (Fig. 6). Statistically
462 significant overrepresented GO terms included fundamental metabolic, developmental,
463 and regulation processes, signaling, cellular differentiation, organization, and
464 communication, and morphogenesis. Underrepresented GO terms included fundamental
465 molecular activities, integration, transposition, DNA replication and recombination, in
466 ribosomes and mitochondria (File DS 15). We found the same overall pattern in edgeR
467 although none of the GO terms were significantly enriched.

468

469 **DISCUSSION**

470 **Assembly quality and transcriptome variability in crustaceans**

471 Our transcriptomic assembly yielded a relatively high number of contigs and relatively
472 low N50s. These numbers would typically suggest a poor assembly, possibly stemming
473 from the RNA degradation we reported. However, our ExN50s indicate the N50 based on
474 100% of contigs is biased due to short and lowly expressed isoforms. Further, 70-88%

475 complete BUSCOs suggest that our assembly contains mostly complete genes present in
476 all animals, and comparable to studies with seemingly no degradation and various
477 degrees of divergence to references (57-99%; Levin et al. 2016). Crustaceans have many
478 more splice variants compared to other metazoans, elevating contig count and low gene
479 ontology annotation (~32%) due to highly divergent or novel sequences (Havird &
480 Santos 2016). We were able to annotate ~50% of our transcriptome, suggesting that even
481 though some gene fragmentation was present, our transcriptome has similar
482 characteristics to other crustacean assemblies. Moreover, we found a little over ~4000
483 DEGs, similar to the average seen in previous studies (Havird & Santos 2016) indicating
484 the number of contigs did not inflate detected DEGs. Our data supports recent
485 observations that traditional metrics such as N50s are not a good proxy for RNA-seq data
486 quality ([https://github.com/trinityrnaseq/EMBOtrinityWorkshopSept2016/wiki/De-novo-](https://github.com/trinityrnaseq/EMBOtrinityWorkshopSept2016/wiki/De-novo-Assembly,-Quantitation,-and-Differential-Expression;)
487 [Assembly,-Quantitation,-and-Differential-Expression;](http://www.molecularecologist.com/2017/04/the-first-problem-with-n50/)
488 <http://www.molecularecologist.com/2017/04/the-first-problem-with-n50/>), in particular
489 for non-model organisms where known metrics are not “one-size-fits-all” (Havird &
490 Santos 2016).

491 Moreover, we showed an unprecedented amount of transcriptomic
492 assembly variation in arthropods, particularly in non-model decapod crustaceans. The
493 wide range of N50 values and contig numbers from uniform assembly and quality control
494 further exposes how traditional expectations such as N50s do not accurately describe the
495 nature of RNA-seq in many crustaceans. In all arthropods the bulk of the contig density
496 was smaller than 1000 bps, implying “large” N50s are a misleading metric that does not
497 represent contig size accurately. While many of our contigs (*genes*) were shorter, their

498 distribution fits with other crab and shrimp transcriptomes suggesting decapods could be
499 more complex than model organisms and could have more, shorter genes (Havird &
500 Santos 2016); not necessarily poorly assembled. Even if TRINITY overestimated the
501 number of genes by a factor of two in fruit flies and water fleas (probably due to having a
502 single sample per assembly), the number of contigs found in decapods is consistent with
503 genomes at least an order of magnitude larger (0.17-0.2 Gbases *versus* >1-40 Gbases;
504 <http://www.genomesize.com/>). Water fleas have high number of duplicated genes
505 (Colbourne et al. 2011), thus it is possible that gene duplication could explain larger
506 number of contigs in other crustaceans. Future research will help disentangle number of
507 genes, assembly artifacts, and gene duplication in many non-model organisms.

508 **Downregulated features dominate differential gene expression**

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

516 Because our primary focus was on differential gene expression in response to OD
517 exposure and methodological constraints, we did not include a dispersant-only treatment
518 and cannot quantify the separate effects of dispersants. Nevertheless, acute toxicity tests
519 on mysid shrimp (*Americamysis bahia*) and the inland silverside (*Menidia beryllina*)
520 revealed that dispersant-only treatments had lower toxicity effects when compared to OO

521 and OD conditions (Hemmer et al. 2011). Additionally, toxicity studies in decapod
522 shrimp showed sensitivity increase of an order of magnitude in OD over OO (Fisher &
523 Foss 1993). These studies suggest that crabs exhibit increased negative effects when
524 exposed to dispersed oil.

525 In individuals exposed to OD, we found evidence for the downregulation of
526 >4000 genes involved in fundamental cellular mechanisms, suggesting a generalized
527 shutting down of crabs' cellular processes and functions (*SI* Table S3). Some of these
528 included genes involved in actin binding, condensation, and folding in the cytoskeleton
529 (*SI* Table S4). Failures associated with these mechanisms cause multiple diseases in
530 humans (Lundin et al. 2010) and flight muscle alterations in *Drosophila* (Sparrow et al.
531 2003). For instance, downregulation of dendrite morphogenesis and neuron projection
532 could be associated with actin defects in the nervous system. Another relevant finding
533 was the downregulation of receptor genes involved in apoptosis, in particular TNFs (*SI*
534 Table S5). When apoptotic mechanisms are suppressed, abnormal cell growth leads to
535 tumors (Lowe & Lin 2000). Past studies have shown tumor growth after oil spills
536 (Suchanek 1993), including evidence for neoplastic tumors in clams following the 1971
537 Maine oil spill (Barry & Yevich 1975). Genotoxic and cytotoxic effects, mainly
538 chromosomal aberrations known to be associated with several types of cancer (Mitelman
539 et al. 2007), have also been witnessed in whale fibroblasts exposed to oil-dispersant
540 (Wise et al. 2014). The carcinogenic effects of PAHs are well documented (Phillips
541 1999), and our findings suggest there could a link between OD exposure in crabs, tumor
542 necrosis receptor/apoptosis downregulation, tumor growth, and chromosomal
543 abnormalities observed in previous studies.

544 In addition to downregulated features, we identified a limited number of
545 upregulated features related to muscular structure and development in OD. This suggests
546 that crabs may be undergoing damage-repair mechanisms, or that muscle upkeep-related
547 genes far outweigh immune system genes due to relative abundance normalization.
548 Interestingly, the same features were downregulated in NC (Fig. 5), indicating
549 upregulation is not involved in general muscle maintenance under control conditions.

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

562 **The role of cytochrome P-4501 (CYP1As) proteins as universal proxy for** 563 **hydrocarbon contamination**

564 One known link between hydrocarbon exposure and signaling pathways is the production
565 of cytochrome P-4501A (CYP1As) in *clan-1*, a protein crucial in the aryl-hydrocarbon
566 receptor (AHR) signaling pathway. CYP1A is commonly used to test for toxic

567 hydrocarbon exposure (Petrulis et al. 2001), because its expression is linked with AHR-
568 binding and metabolism of contaminants such as PAHs, polychlorinated biphenyl (PCB)
569 in vertebrate systems (Petrulis et al. 2001; Whitehead et al. 2012). Fish populations
570 naturally and experimentally exposed to DWH oil showed CYP1A upregulation (Brown-
571 Peterson et al. 2015; Whitehead et al. 2012). However, major differences exist between
572 vertebrate and invertebrate AHR pathways in that vertebrates have sophisticated aromatic
573 binding capabilities (Barron 2012). While AHR pathways exist in most animals, a past
574 review found dioxins and PAHs do not bind to invertebrate CYP1A-AHRs (Hahn 2002)
575 although there are recent evidences of CYP1A upregulation in oysters (Jenny et al. 2016)
576 and in copepods (Han et al. 2016; Hansen et al. 2016a; Hansen et al. 2016b; Olsvik et al.
577 2012). We found no evidence of CYP1A upregulation in crabs, congruent with past
578 studies using similar species (Yednock et al. 2015). The lack of AHR regulation mediated
579 by CYP1A in crabs could be due to a lower sensitivity to PAHs to those particular
580 receptors or, alternatively, CYP1A protein sequences in crabs may be too divergent to be
581 annotated. Moreover, CYP1A differential gene expression could be restricted to other
582 organs' specific responses not evident in muscle tissues. However, tissue specificity
583 artifacts seem unlikely due to a similar absence of CYP1As in blue crab hepatopancreas
584 and gills exposed to oil (Yednock et al. 2015). Future research including dispersed oil and
585 multiple tissue types will help assess crabs' response to those seemingly universal
586 toxicology markers. Interestingly, we did find differential gene regulation in other CYPs
587 (clans 3, 4 and 18) used by invertebrates to metabolize aromatic toxins (*SI* Table S5,
588 (Rewitz et al. 2006); however, these proteins were downregulated in OD presence.
589 Nonetheless, in light of our findings, the continued use of CYP1A upregulation as a

590 universal proxy for PAH, PCB, or dioxin exposure could have limited utility when used
591 in decapod crustaceans being important fisheries. We advocate for screening additional
592 CYPs as a complementary proxy for PAH exposure.

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

594 In addition to AHRs signaling pathway, we identified the suppression of several immune
595 response receptors (FcRs, IKKs, T cells, and TLRs) known to be conserved across
596 animals (Ottaviani et al. 2007). Our findings identified nine TLRs that were
597 downregulated in OD, compared to a past study that found only three in an equivalent
598 OO (Yednock et al. 2015), suggesting increased immune repression in the presence of
599 OD in crustaceans. Conversely, we found upregulation for the hemocyte proliferation
600 pathway in OD. Hemocytes are the first line of defense to infection in invertebrates
601 (Barron 2012; Hamoutene et al. 2004), suggesting crabs are combating stress related to
602 pathogens. Several studies showed that exposure to oil resulted in an increased
603 susceptibility to disease and compromised reproductive, sensory, and neurological
604 systems. Therefore, exposure to chemically dispersed oil may exacerbate effects. For
605 example, fish species exposed to the Exxon Valdez spill had a 17-fold parasite load
606 increase (Khan 1990) and tuna fish impacted by DWH had impaired heart contractions
607 (Brette et al. 2014). In invertebrates (see references in (Suchanek 1993): hydroids and
608 jellies showed teratogenic effects and neurological changes; hermatypic corals had
609 decreased ovaria and less planula per polyp, ova degeneration, and lack of gonad
610 development; kelp crabs presented chemosensory-induced bradycardia and suppressed
611 chemoreception abilities; sea urchins showed delayed embryogenesis, asynchronism,
612 non-viable larvae, and functional loss of tube foot and spine movement; multiple bivalves

613 exhibited reduced respiration and increased energy expenditures. Specifically in *E.*
614 *depressus*, occurrence of supernumerary and morphologically abnormal megalopal stages
615 was shown to be associated with crude oil (Cucci & Epifanio 1979). Our results identify
616 several immune response pathways that could be playing a role in the adverse effects
617 witnessed in other marine organisms (Peterson et al. 2003; Wise & Wise 2011) and
618 humans involved in oil spill clean-up (D'Andrea & Reddy 2013). This is of particular
619 importance because human lung cell cultures showed similar immune response impact to
620 gene expression when exposed to chemically dispersed oil (Liu et al. 2016). Therefore,
621 we suggest these receptors, such as TLRs, could serve as potential candidates as
622 complementary biomarkers to CYPs for aquatic toxicology.

623 We did not verify analytically concentrations of hydrocarbons prior and after
624 experiments. In acute non-renewal exposure it is possible than many toxic compounds
625 and oxygen are depleted due to consumption or transformation, leading to confounding
626 reasons for gene expression differences. All changes we document in this study are
627 consistent with lethal and sublethal effects of chemically dispersed oil in previous
628 toxicological (DeLeo et al. 2015; Echols et al. 2016; Wise et al. 2014) and transcriptomic
629 studies (Jenny et al. 2016; Liu et al. 2016; Bayha et al. 2017) where detailed analytical
630 data is available. Moreover, the lack of significant changes between aerated controls and
631 non-aerated treatments suggest that low oxygen is not driving gene expression. Future
632 transcriptomic analyses coupled comprehensive analytical testing and renewal exposures
633 will shed light into more precise points of concentrations and gene expression
634 correlations beyond the baseline changes we report presently.

635 The need for developing less to non-toxic dispersants is pressing. Current and
636 future oil exploration contingency in the U.S.A. (Bureau of Safety and Environmental
637 Enforcement 2015; Shell Offshore Inc 2010) relies on the same dispersant,
638 COREXIT®9500, that our results and number of other studies (e.g., Hemmer et al. 2011;
639 Paul et al. 2013; Rico-Martínez et al. 2013; Wise & Wise 2011; Bayha et al. 2017), have
640 shown to be highly detrimental to animal health particularly in systems exposed to
641 xenobiotic bioaccumulation. Although COREXIT®9500 is one of the relatively less toxic
642 formulations vetted by the United States Environmental Protection Agency (Hemmer et
643 al. 2011), future research needs to find safer solutions for spill cleanup and explore the
644 longer-term consequences of oil and dispersant impacts on marine community health.

645 **CONCLUSIONS**

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

658

659 **ACKNOWLEDGMENTS**

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

666 **ADDITIONAL INFORMATION AND DECLARATIONS**667 **Funding**

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

670 **Competing interests.**

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

673 **Additional Information**

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

677 **Author Contributions**

678 -Hernán Vázquez-Miranda performed laboratory procedures, analyzed the data, wrote the
679 paper with help of coauthors, prepared figures/tables, reviewed drafts of the paper.

680 -Brent P. Thoma and Juliet M. Wong performed experimental treatments and laboratory
681 procedures, reviewed drafts of the paper.

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

685 **Data deposition**

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

696 **Supplemental Information**

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

698

699 **REFERENCES**

700

701 Altschul SF, Gish W, Miller W, Myers EW, and Lipman DJ. 1990. Basic local alignment
702 search tool. *Journal of molecular biology* 215:403-410.

- 703 Anderson JA, Kuhl AJ, and Anderson AN. 2014. Toxicity of oil and dispersed oil on
704 juvenile mud crabs, *Rhithropanopeus harrisi*. *Bulletin of environmental*
705 *contamination and toxicology* 92:375-380.
- 706 Andrews S. 2010. FastQC: A quality control tool for high throughput sequence data.
707 *Reference Source Babraham Bioinformatics*.
- 708 Barron MG. 2012. Ecological impacts of the Deepwater Horizon oil spill: implications
709 for immunotoxicity. *Toxicologic pathology* 40:315-320.
- 710 Barry M, and Yevich P. 1975. The ecological, chemical and histopathological evaluation
711 of an oil spill site: Part III. Histopathological Studies. *Marine pollution bulletin*
712 6:171-173.
- 713 Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A,
714 Marshall M, Moxon S, and Sonnhammer EL. 2004. The Pfam protein families
715 database. *Nucleic acids research* 32:D138-D141.
- 716 Bayha KM, Ortell N, Ryan CN, Griffitt KJ, Krasnec M, Sena J, Ramaraj T, Takeshita R,
717 Mayer GD, and Schilkey F. 2017. Crude oil impairs immune function and
718 increases susceptibility to pathogenic bacteria in southern flounder. *PloS one*
719 12:e0176559.
- 720 Bland JM, and Altman D. 1986. Statistical methods for assessing agreement between two
721 methods of clinical measurement. *The lancet* 327:307-310.
- 722 Bolger AM, Lohse M, and Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina
723 sequence data. *Bioinformatics*:btu170.

- 724 Bowen L, Miles AK, Ballachey B, Waters S, and Bodkin J. 2016. Gene Transcript
725 Profiling in Sea Otters Post-Exxon Valdez Oil Spill: A Tool for Marine
726 Ecosystem Health Assessment. *Journal of Marine Science and Engineering* 4:39.
- 727 Brette F, Machado B, Cros C, Incardona JP, Scholz NL, and Block BA. 2014. Crude oil
728 impairs cardiac excitation-contraction coupling in fish. *science* 343:772-776.
- 729 Brown-Peterson NJ, Krasnec M, Takeshita R, Ryan CN, Griffitt KJ, Lay C, Mayer GD,
730 Bayha KM, Hawkins WE, and Lipton I. 2015. A multiple endpoint analysis of the
731 effects of chronic exposure to sediment contaminated with Deepwater Horizon oil
732 on juvenile Southern flounder and their associated microbiomes. *Aquatic*
733 *Toxicology* 165:197-209.
- 734 Bureau of Safety and Environmental Enforcement. 2015. Chukchi Sea oil spill response
735 plan. Chukchi Sea regional exploration program. 3.0 ed.
736 [http://www.bsee.gov/uploadedFiles/BSEE/BSEE_Newsroom/Publications_Librar](http://www.bsee.gov/uploadedFiles/BSEE/BSEE_Newsroom/Publications_Library/OSRPs/Arctic/2015_05_15_Revision%20Redacted_Shell%20Chukchi%20Sea%20OSRP%20with%20Cover%20Letters.pdf)
737 [y/OSRPs/Arctic/2015_05_15_Revision%20Redacted_Shell%20Chukchi%20Se](http://www.bsee.gov/uploadedFiles/BSEE/BSEE_Newsroom/Publications_Library/OSRPs/Arctic/2015_05_15_Revision%20Redacted_Shell%20Chukchi%20Sea%20OSRP%20with%20Cover%20Letters.pdf)
738 [a%20OSRP%20with%20Cover%20Letters.pdf](http://www.bsee.gov/uploadedFiles/BSEE/BSEE_Newsroom/Publications_Library/OSRPs/Arctic/2015_05_15_Revision%20Redacted_Shell%20Chukchi%20Sea%20OSRP%20with%20Cover%20Letters.pdf) (accessed 16.12.01)
- 739 Clark JR, Bragin GE, Febbo EJ, and Letinski DJ. 2001. Toxicity of physically and
740 chemically dispersed oils under continuous and environmentally realistic
741 exposure conditions: Applicability to dispersant use decisions in spill response
742 planning. International Oil Spill Conference: American Petroleum Institute. p
743 1249-1255.
- 744 Colbourne JK, Pfrender ME, Gilbert D, Thomas WK, Tucker A, Oakley TH, Tokishita S,
745 Aerts A, Arnold GJ, and Basu MK. 2011. The ecoresponsive genome of *Daphnia*
746 *pulex*. *Science* 331:555-561.

- 747 Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A,
748 Szcześniak MW, Gaffney DJ, Elo LL, and Zhang X. 2016. A survey of best
749 practices for RNA-seq data analysis. *Genome Biology* 17:13.
- 750 Cucci T, and Epifanio C. 1979. Long-term effects of water-soluble fractions of Kuwait
751 crude oil on the larval and juvenile development of the mud crab *Eurypanopeus*
752 depressus. *Marine Biology* 55:215-220.
- 753 Cui X, and Churchill GA. 2003. Statistical tests for differential expression in cDNA
754 microarray experiments. *Genome Biol* 4:210.
- 755 D'Andrea MA, and Reddy GK. 2013. Health consequences among subjects involved in
756 Gulf oil spill clean-up activities. *The American journal of medicine* 126:966-974.
- 757 DeLeo DM, Ruiz-Ramos DV, Baums IB, and Cordes EE. 2015. Response of deep-water
758 corals to oil and chemical dispersant exposure. *Deep Sea Research Part II:*
759 *Topical Studies in Oceanography*.
- 760 Dudoit S, Yang YH, Callow MJ, and Speed TP. 2002. Statistical methods for identifying
761 differentially expressed genes in replicated cDNA microarray experiments.
762 *Statistica sinica*:111-139.
- 763 Echols B, Smith A, Gardinali P, and Rand G. 2016. The use of ephyrae of a scyphozoan
764 jellyfish, *Aurelia aurita*, in the aquatic toxicological assessment of Macondo oils
765 from the Deepwater Horizon incident. *Chemosphere* 144:1893-1900.
- 766 Finn RD, Clements J, and Eddy SR. 2011. HMMER web server: interactive sequence
767 similarity searching. *Nucleic acids research*:gkr367.

- 768 Fisher WS, and Foss SS. 1993. A simple test for toxicity of number 2 fuel oil and oil
769 dispersants to embryos of grass shrimp, *Palaemonetes pugio*. *Marine pollution*
770 *bulletin* 26:385-391.
- 771 Francis WR, Christianson LM, Kiko R, Powers ML, Shaner NC, and Haddock SH. 2013.
772 A comparison across non-model animals suggests an optimal sequencing depth
773 for de novo transcriptome assembly. *BMC genomics* 14:167.
- 774 Fransen L, Müller R, Marmenout A, Tavernier J, Van der Heyden J, Kawashima E,
775 Chollet A, Tizard R, Van Heuverswyn H, and Van Vliet A. 1985. Molecular
776 cloning of mouse tumour necrosis factor cDNA and its eukaryotic expression.
777 *Nucleic acids research* 13:4417-4429.
- 778 Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier
779 L, Ge Y, and Gentry J. 2004. Bioconductor: open software development for
780 computational biology and bioinformatics. *Genome Biology* 5:R80.
- 781 Gleason LU, and Burton RS. 2015. RNA-seq reveals regional differences in
782 transcriptome response to heat stress in the marine snail *Chlorostoma funebris*.
783 *Molecular Ecology* 24:610-627.
- 784 Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB,
785 Eccles D, Li B, and Lieber M. 2013. De novo transcript sequence reconstruction
786 from RNA-seq using the Trinity platform for reference generation and analysis.
787 *Nature protocols* 8:1494-1512.
- 788 Hahn ME. 2002. Aryl hydrocarbon receptors: diversity and evolution. *Chemico-*
789 *biological interactions* 141:131-160.

- 790 Hamoutene D, Payne J, Rahimtula A, and Lee K. 2004. Effect of water soluble fractions
791 of diesel and an oil spill dispersant (Corexit 9527) on immune responses in
792 mussels. *Bulletin of environmental contamination and toxicology* 72:1260-1267.
- 793 Han J, Won E-J, Kang H-M, Lee M-C, Jeong C-B, Kim H-S, Hwang D-S, and Lee J-S.
794 2016. Marine copepod cytochrome P450 genes and their applications for
795 molecular ecotoxicological studies in response to oil pollution. *Marine pollution*
796 *bulletin*.
- 797 Hansen BH, Jager T, Altin D, Øverjordet IB, Olsen AJ, Salaberria I, and Nordtug T.
798 2016a. Acute toxicity of dispersed crude oil on the cold-water copepod *Calanus*
799 *finmarchicus*: Elusive implications of lipid content. *Journal of Toxicology and*
800 *Environmental Health, Part A* 79:549-557.
- 801 Hansen BH, Lie KK, Størseth TR, Nordtug T, Altin D, and Olsvik PA. 2016b. Exposure
802 of first-feeding cod larvae to dispersed crude oil results in similar transcriptional
803 and metabolic responses as food deprivation. *Journal of Toxicology and*
804 *Environmental Health, Part A* 79:558-571.
- 805 Havird JC, and Santos SR. 2016. Here we are, but where do we go? a systematic review
806 of crustacean transcriptomic studies from 2014–2015. *Integrative and*
807 *Comparative Biology*:icw061.
- 808 Hemmer MJ, Barron MG, and Greene RM. 2011. Comparative toxicity of eight oil
809 dispersants, Louisiana sweet crude oil (LSC), and chemically dispersed LSC to
810 two aquatic test species. *Environmental Toxicology and Chemistry* 30:2244-2252.
- 811 Jenny MJ, Walton WC, Payton SL, Powers JM, Findlay RH, O'Shields B, Diggins K,
812 Pinkerton M, Porter D, and Crane DM. 2016. Transcriptomic evaluation of the

- 813 American oyster, *Crassostrea virginica*, deployed during the Deepwater Horizon
814 oil spill: Evidence of an active hydrocarbon response pathway. *Marine*
815 *Environmental Research* 120:166-181.
- 816 Khan R. 1990. Parasitism in marine fish after chronic exposure to petroleum
817 hydrocarbons in the laboratory and to the Exxon Valdez oil spill. *Bulletin of*
818 *environmental contamination and toxicology* 44:759-763.
- 819 Kujawinski EB, Kido Soule MC, Valentine DL, Boysen AK, Longnecker K, and
820 Redmond MC. 2011. Fate of dispersants associated with the Deepwater Horizon
821 oil spill. *Environmental science & technology* 45:1298-1306.
- 822 Lagesen K, Hallin P, Rødland EA, Stærfeldt H-H, Rognes T, and Ussery DW. 2007.
823 RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic*
824 *acids research* 35:3100-3108.
- 825 Langmead B. 2010. Aligning short sequencing reads with Bowtie. *Current protocols in*
826 *bioinformatics*:11.17. 11-11.17. 14.
- 827 Levin M, Anavy L, Cole AG, Winter E, Mostov N, Khair S, Senderovich N, Kovalev E,
828 Silver DH, and Feder M. 2016. The mid-developmental transition and the
829 evolution of animal body plans. *Nature*.
- 830 Li B, and Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq
831 data with or without a reference genome. *BMC bioinformatics* 12:1.
- 832 Liu Y-Z, Roy-Engel AM, Baddoo MC, Flemington EK, Wang G, and Wang H. 2016.
833 The impact of oil spill to lung health—Insights from an RNA-seq study of human
834 airway epithelial cells. *Gene* 578:38-51.

- 835 Love M, Anders S, and Huber W. 2014. Differential analysis of count data—the DESeq2
836 package. *Genome Biology* 15:550.
- 837 Lowe SW, and Lin AW. 2000. Apoptosis in cancer. *Carcinogenesis* 21:485-495.
- 838 Lundin VF, Leroux MR, and Stirling PC. 2010. Quality control of cytoskeletal proteins
839 and human disease. *Trends in biochemical sciences* 35:288-297.
- 840 Mitelman F, Johansson B, and Mertens F. 2007. *Mitelman database of chromosome*
841 *aberrations in cancer*: Cancer Genome Anatomy Project.
- 842 Munro S, Lund S, Pine P, Binder H, Clevert D, Conesa A, Dopazo J, Fasold M,
843 Hochreiter S, and Hong H. 2014. Assessing technical performance in differential
844 gene expression experiments with external spike-in RNA control ratio mixtures.
845 *Nature Communications* 5:5125.
- 846 National Research Council (NRC). 2005. *Oil Spill Dispersants: Efficacy and Effects*:
847 National Academies Press.
- 848 O'Connor TP. 1998. Mussel Watch results from 1986 to 1996. *Marine pollution bulletin*
849 37:14-19.
- 850 OECD. 2002. Guidance Document on Aquatic Toxicity Testing of Difficult Substances
851 and Mixtures. OECD Series on Testing and Assessment,: Organization for
852 Economic Cooperation and Development.
- 853 Olsvik PA, Hansen BH, Nordtug T, Moren M, Holen E, and Lie KK. 2011.
854 Transcriptional evidence for low contribution of oil droplets to acute toxicity from
855 dispersed oil in first feeding Atlantic cod (*Gadus morhua*) larvae. *Comparative*
856 *Biochemistry and Physiology Part C: Toxicology & Pharmacology* 154:333-345.

- 857 Olsvik PA, Lie KK, Nordtug T, and Hansen BH. 2012. Is chemically dispersed oil more
858 toxic to Atlantic cod (*Gadus morhua*) larvae than mechanically dispersed oil? A
859 transcriptional evaluation. *BMC genomics* 13:1.
- 860 Ottaviani E, Malagoli D, and Franceschi C. 2007. Common evolutionary origin of the
861 immune and neuroendocrine systems: from morphological and functional
862 evidence to in silico approaches. *Trends in immunology* 28:497-502.
- 863 Paul JH, Hollander D, Coble P, Daly KL, Murasko S, English D, Basso J, Delaney J,
864 McDaniel L, and Kovach CW. 2013. Toxicity and mutagenicity of Gulf of
865 Mexico waters during and after the Deepwater Horizon oil spill. *Environmental*
866 *science & technology* 47:9651-9659.
- 867 Petersen TN, Brunak S, von Heijne G, and Nielsen H. 2011. SignalP 4.0: discriminating
868 signal peptides from transmembrane regions. *Nature methods* 8:785-786.
- 869 Peterson CH, Rice SD, Short JW, Esler D, Bodkin JL, Ballachey BE, and Irons DB.
870 2003. Long-term ecosystem response to the Exxon Valdez oil spill. *science*
871 302:2082-2086.
- 872 Petrusis JR, Chen G, Benn S, LaMarre J, and Bunce NJ. 2001. Application of the
873 ethoxyresorufin-O-deethylase (EROD) assay to mixtures of halogenated aromatic
874 compounds. *Environmental toxicology* 16:177-184.
- 875 Phillips DH. 1999. Polycyclic aromatic hydrocarbons in the diet. *Mutation*
876 *Research/Genetic Toxicology and Environmental Mutagenesis* 443:139-147.
- 877 R Core Development Team. 2010. The R project for statistical computing. Vienna,
878 Austria.

- 879 Rewitz KF, Styrishave B, Løbner-Olesen A, and Andersen O. 2006. Marine invertebrate
880 cytochrome P450: emerging insights from vertebrate and insect analogies.
881 *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology*
882 143:363-381.
- 883 Rico-Martínez R, Snell TW, and Shearer TL. 2013. Synergistic toxicity of Macondo
884 crude oil and dispersant Corexit 9500A® to the *Brachionus plicatilis* species
885 complex (Rotifera). *Environmental Pollution* 173:5-10.
- 886 Risso D, Ngai J, Speed TP, and Dudoit S. 2014. Normalization of RNA-seq data using
887 factor analysis of control genes or samples. *Nature biotechnology* 32:896-902.
- 888 Robinson MD, McCarthy DJ, and Smyth GK. 2010. edgeR: a Bioconductor package for
889 differential expression analysis of digital gene expression data. *Bioinformatics*
890 26:139-140.
- 891 Sammarco PW, Kolian SR, Warby RA, Bouldin JL, Subra WA, and Porter SA. 2013.
892 Distribution and concentrations of petroleum hydrocarbons associated with the
893 BP/Deepwater Horizon Oil Spill, Gulf of Mexico. *Marine pollution bulletin*
894 73:129-143.
- 895 Schulz RA, and Fossett N. 2005. Hemocyte development during *Drosophila*
896 embryogenesis. *Developmental Hematopoiesis: Methods and Protocols*:109-121.
- 897 Shell Offshore Inc. 2010. Gulf of Mexico regional oil spill response plan. 6.1 ed.
898 http://housedocs.house.gov/energycommerce/Docs_06152010/Shell.Gulf.Oil.Spill
899 [.Response.Plan.pdf](#) (accessed 16.12.01)

- 900 Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, and Zdobnov EM. 2015.
901 BUSCO: assessing genome assembly and annotation completeness with single-
902 copy orthologs. *Bioinformatics*:btv351.
- 903 Singer M, Aurand D, Bragin G, Clark J, Coelho G, Sowby M, and Tjeerdema R. 2000.
904 Standardization of the preparation and quantitation of water-accommodated
905 fractions of petroleum for toxicity testing. *Marine pollution bulletin* 40:1007-
906 1016.
- 907 Sparrow JC, Nowak KJ, Durling HJ, Beggs AH, Wallgren-Pettersson C, Romero N,
908 Nonaka I, and Laing NG. 2003. Muscle disease caused by mutations in the
909 skeletal muscle alpha-actin gene (ACTA1). *Neuromuscular Disorders* 13:519-
910 531.
- 911 Suchanek TH. 1993. Oil impacts on marine invertebrate populations and communities.
912 *American Zoologist* 33:510-523.
- 913 Trustees DWH. 2016. Deepwater Horizon oil spill: Final Programmatic Damage
914 Assessment and Restoration Plan and Final Programmatic Environmental Impact
915 Statement. In: Trustees DWH, editor: National Oceanic and Atmospheric
916 Administration. [http://www.gulfspillrestoration.noaa.gov/restoration-
917 planning/gulf-plan/](http://www.gulfspillrestoration.noaa.gov/restoration-planning/gulf-plan/) (accessed 16.12.01)
- 918 Vignier J, Soudant P, Chu F, Morris J, Carney M, Lay C, Krasnec M, Robert R, and
919 Volety A. 2016. Lethal and sub-lethal effects of Deepwater Horizon slick oil and
920 dispersant on oyster (*Crassostrea virginica*) larvae. *Marine Environmental
921 Research* 120:20-31.

- 922 Wang Z, Gerstein M, and Snyder M. 2009. RNA-Seq: a revolutionary tool for
923 transcriptomics. *Nature Reviews Genetics* 10:57-63.
- 924 Waterhouse RM, Tegenfeldt F, Li J, Zdobnov EM, and Kriventseva EV. 2013. OrthoDB:
925 a hierarchical catalog of animal, fungal and bacterial orthologs. *Nucleic acids*
926 *research* 41:D358-D365.
- 927 Weber CI. 1991. *Methods for measuring the acute toxicity of effluents and receiving*
928 *waters to freshwater and marine organisms*: Environmental Monitoring Systems
929 Laboratory, Office of Research and Development, US Environmental Protection
930 Agency.
- 931 Whitehead A, Dubansky B, Bodinier C, Garcia TI, Miles S, Pilley C, Raghunathan V,
932 Roach JL, Walker N, and Walter RB. 2012. Genomic and physiological footprint
933 of the Deepwater Horizon oil spill on resident marsh fishes. *Proceedings of the*
934 *National Academy of Sciences* 109:20298-20302.
- 935 Williams AB. 1984. *Shrimps, lobsters, and crabs of the Atlantic coast of the eastern*
936 *United States, Maine to Florida*. Washington, D. C.: Smithsonian Institution
937 Press.
- 938 Wise CF, Wise JT, Wise SS, Thompson WD, and Wise JP. 2014. Chemical dispersants
939 used in the Gulf of Mexico oil crisis are cytotoxic and genotoxic to sperm whale
940 skin cells. *Aquatic Toxicology* 152:335-340.
- 941 Wise J, and Wise JP. 2011. A review of the toxicity of chemical dispersants. *Reviews on*
942 *environmental health* 26:281-300.
- 943 Worton DR, Zhang H, Isaacman-VanWertz G, Chan AW, Wilson KR, and Goldstein AH.
944 2015. Comprehensive chemical characterization of hydrocarbons in NIST

- 945 standard reference material 2779 Gulf of Mexico crude oil. *Environmental*
946 *science & technology* 49:13130-13138.
- 947 Yednock BK, Sullivan TJ, and Neigel JE. 2015. De novo assembly of a transcriptome
948 from juvenile blue crabs (*Callinectes sapidus*) following exposure to surrogate
949 Macondo crude oil. *BMC genomics* 16:1.
- 950 Young MD, Wakefield MJ, Smyth GK, and Oshlack A. 2010. Gene ontology analysis for
951 RNA-seq: accounting for selection bias. *Genome Biology* 11:1.
- 952 Zhu L, Qu K, Xia B, Sun X, and Chen B. 2016. Transcriptomic response to water
953 accommodated fraction of crude oil exposure in the gill of Japanese flounder,
954 *Paralichthys olivaceus*. *Marine pollution bulletin* 106:283-291.
- 955

956 **Table1.** Environmental and experiment information for the flatback mudcrab

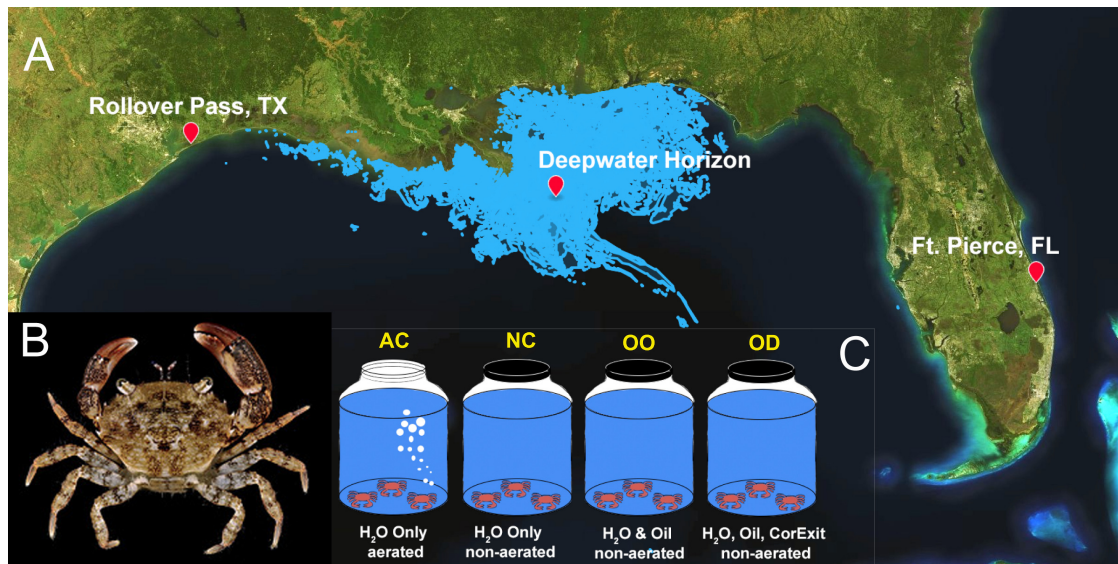
957 transcriptome compliant with the MIGS standard.

Item	Description
Investigation_type	Eukaryote
Project_name	Transcriptomic effects of dispersed oil in <i>Eurypanopeus depressus</i>
Collection_date	09-July-2013 and 23-September-2014
Lat_lon	29.521667 N 94.502100 W 27.436667 N 80.335556 W
Country	USA (Texas and Florida)
Environment	Oyster beds
Geo_loc_name	Rollover Bay, Texas (USA) Fort Pierce, Florida (USA)
env_biome	ENVO_00002030 (aquatic biome)
env_feature	ENVO_00000485 (sea shore)
env_material	ENVO_00002149 (sea water)
Ref_biomaterial	Primary transcriptomic report
biotic_relationship	free living
Rel_to_oxygen	Aerobe
Ploidy	Diploid
Sequencing_meth	Illumina HiSeq2000
Num_replicons	NA
Assembly	Trinity2.0.3

Finishing_strategy	Draft
Estimated_size	2.80E9
Sample_material	muscle
Motility	Yes
Assembly method	Trinity2.0.3
Assembly name	de Novo
Sequencing technology	Illumina HiSeq2000

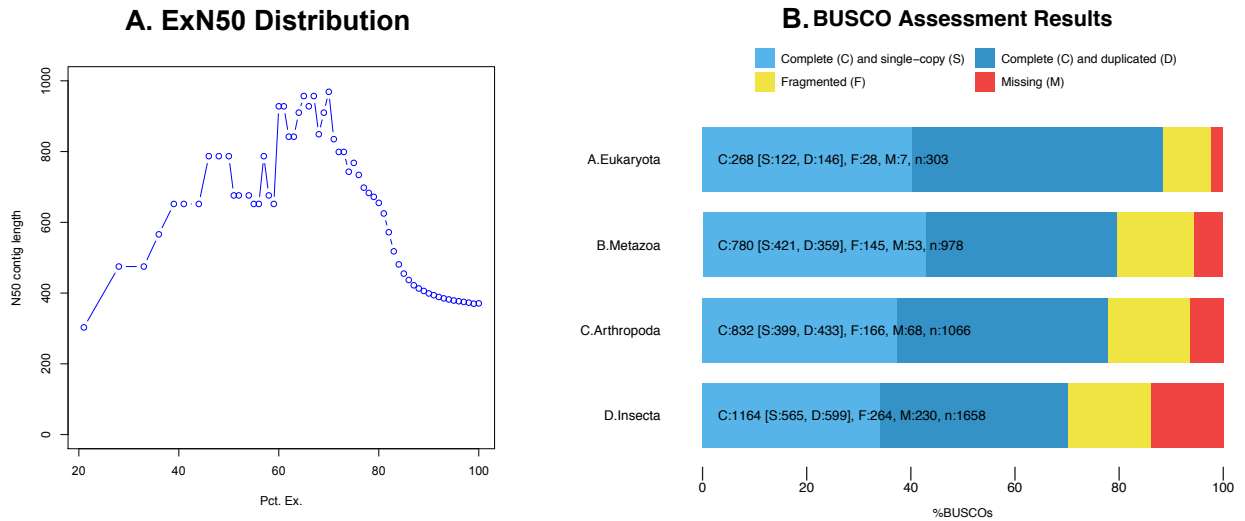
958

959



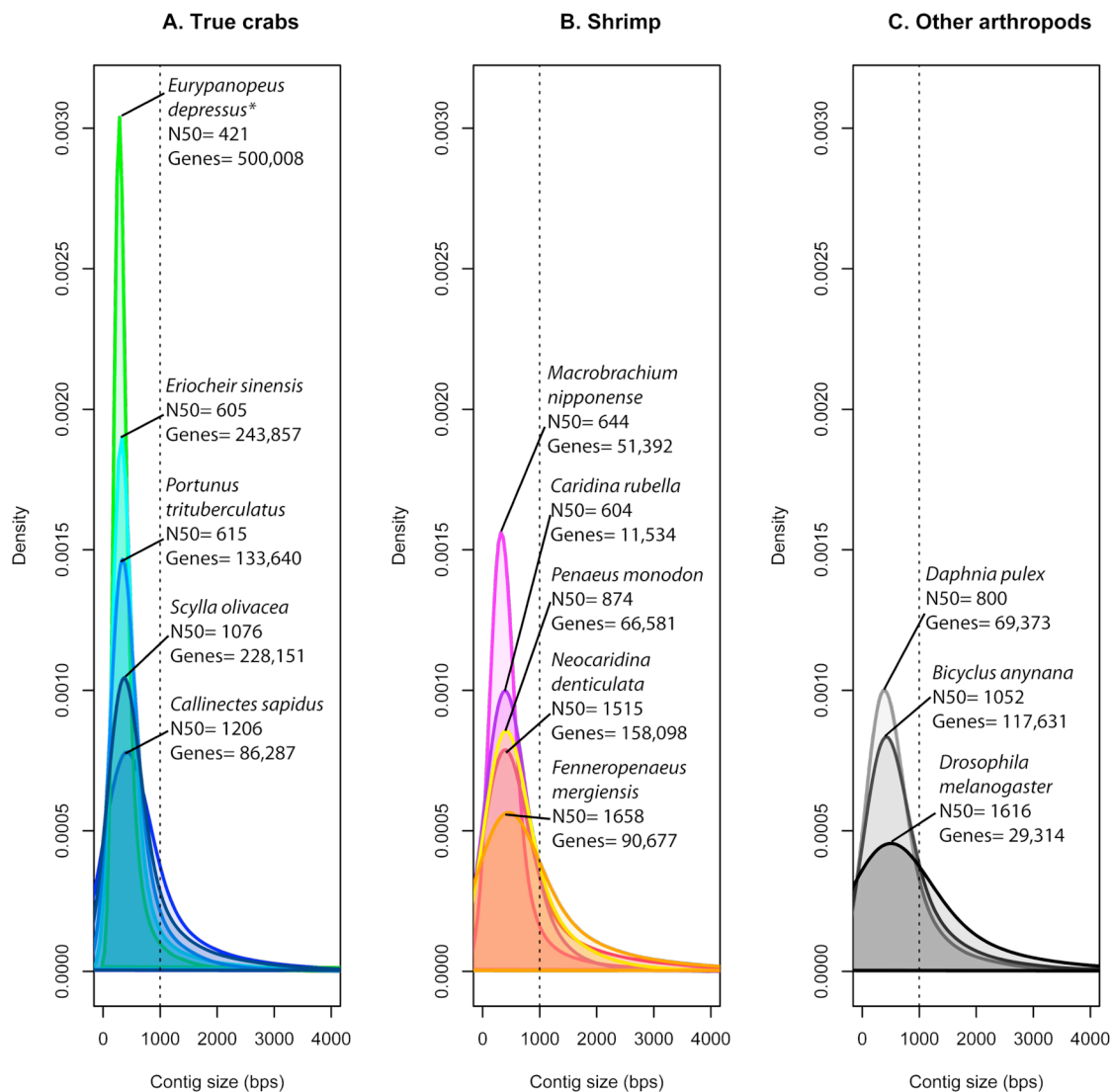
960

961 **Figure 1. Sampling sites, Deepwater Horizon Oil Spill (DWH) area, species of**
 962 **interest, and experimental design. (A)** We collected samples in two sites not affected
 963 by DWH [light blue area
 964 (http://invertebrates.si.edu/boem/maps/BOEM_NMNH_GOM_collections.kmz); map
 965 credit: created in Mapbox v2015-09-26-00-46-36 by B.P.T. (<http://www.mapbox.com>);
 966 map data: OpenStreetMap© (<http://www.openstreetmap.org/copyright>)] of the (B)
 967 flatback mudcrab *Eurypanopeus depressus* (Photo credit: B.P.T.). (C) Our experimental
 968 design (Image credit: B.P.T.) consisted of two controls: aerated (AC), negative (NC); two
 969 experimental treatments: oil-only (OO), oil-dispersant (OD).



970

971 **Figure 2. Assembly quality and transcriptomic completeness. (A)** ExN50
 972 distribution pondering N50 contig length by percentage expressed (Pct. Ex); and (B)
 973 assembly completeness based on Benchmarking Universal Single-Copy Orthologs
 974 (BUSCOs). Each bar represents ORTHODB ortholog assesment in a hierarchical
 975 taxonomic order from Eukaryota to Insecta. Legend: complete genes in blue, missing
 976 genes in red, and in yellow fragmented genes.



977

978 **Figure 3 Contig size density distribution and variability in decapod**979 **crustaceans and model arthropods. (A) Distribution of contig size in true crabs**

980 [Decapoda: Brachyura], (B) Distribution of contig size in shrimp [Decapoda:

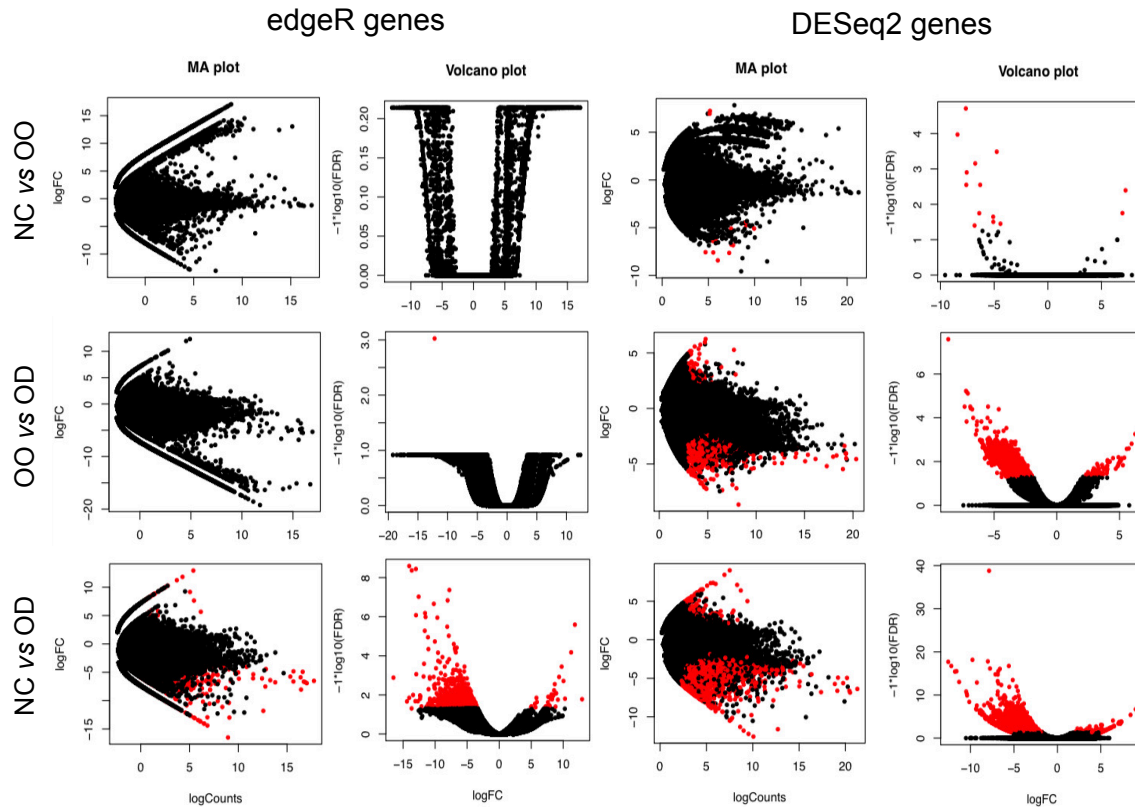
981 Dendrobranchiata & Caridea], and (C) in other arthropods including model

982 organisms fruit fly and water flea. Mudcrab assembly (*Eurypanopeus depressus*)

983 constructed with 15 samples is marked with an asterisk (*); every other assembly

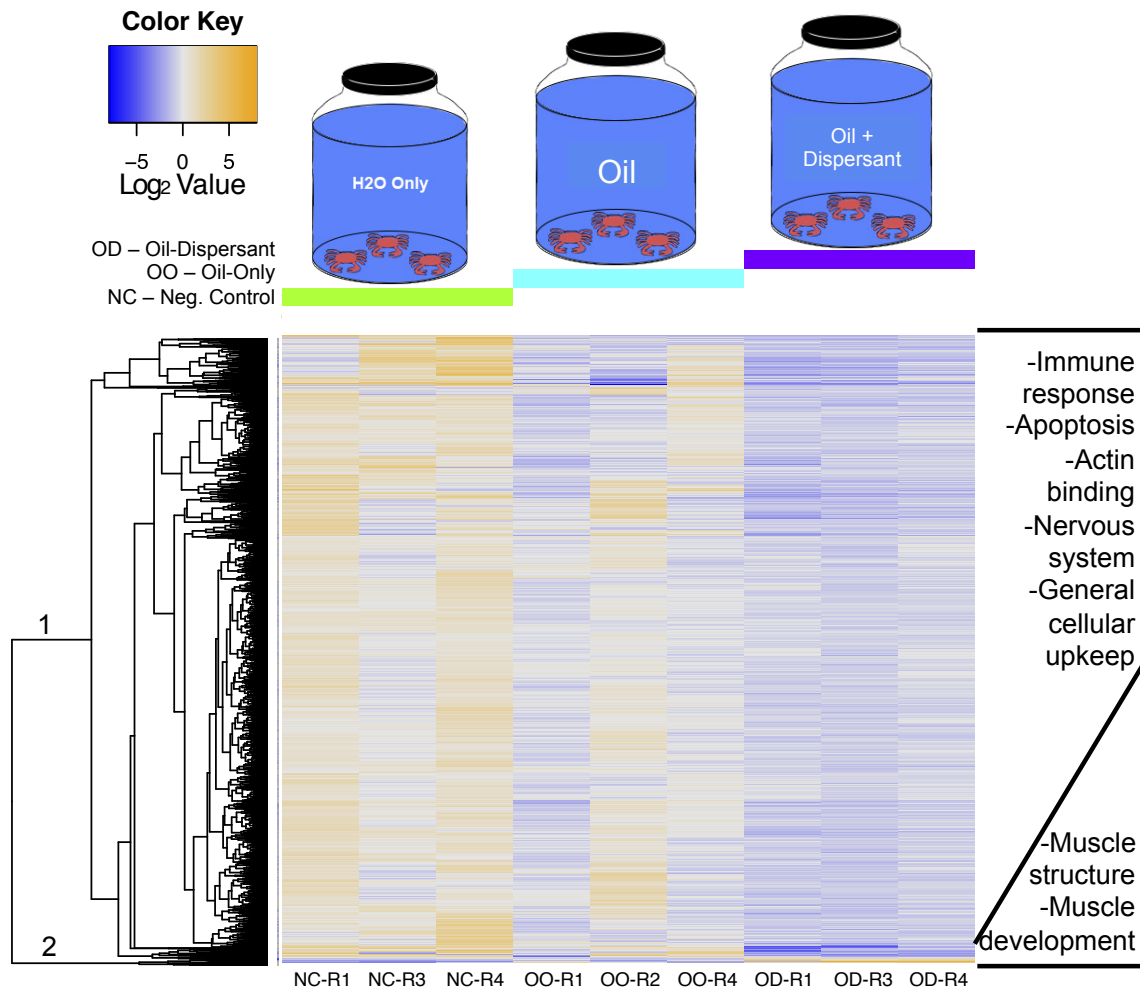
984 includes a single Genbank SRA sample. All species were assembled in TRINITY with

985 the exact same parameters in the same HPC cluster. N50 values represent at least
986 50% of contigs in an assembly are a given size in bps or larger. Dashed vertical lines
987 indicate a 1000 bps contig size threshold. Genes correspond to number of contigs
988 assembled by TRINITY.



989

990 **Figure 4. Pairwise differential gene expression MA and volcano plots between the**
 991 **negative control (NC) and experimental treatments (OO and OD).** In MA plots X-
 992 axis represents mean average expression on counts on a \log_2 scale and Y-axis represents
 993 fold-changes in gene expression on a logarithmic scale. In volcano plots X-axis
 994 represents fold-changes in gene expression on a logarithmic scale, and Y-axis represents
 995 pairwise *t*-test negative \log_{10} *P*-values Red dots represent significant comparisons passing
 996 a 1% false discovery rate (FDR).



997

998 **Figure 5. Differential expression and regulation of genes estimated in DESeq2.**

999 Columns correspond to treatments: NC (negative control) in green, OO (oil-only) in cyan,

1000 and OD (oil-dispersant) in purple. The “R” label below columns represents replicates

1001 (Table S1). Rows correspond to significant genes passing a 1% FDR. Numbers on the

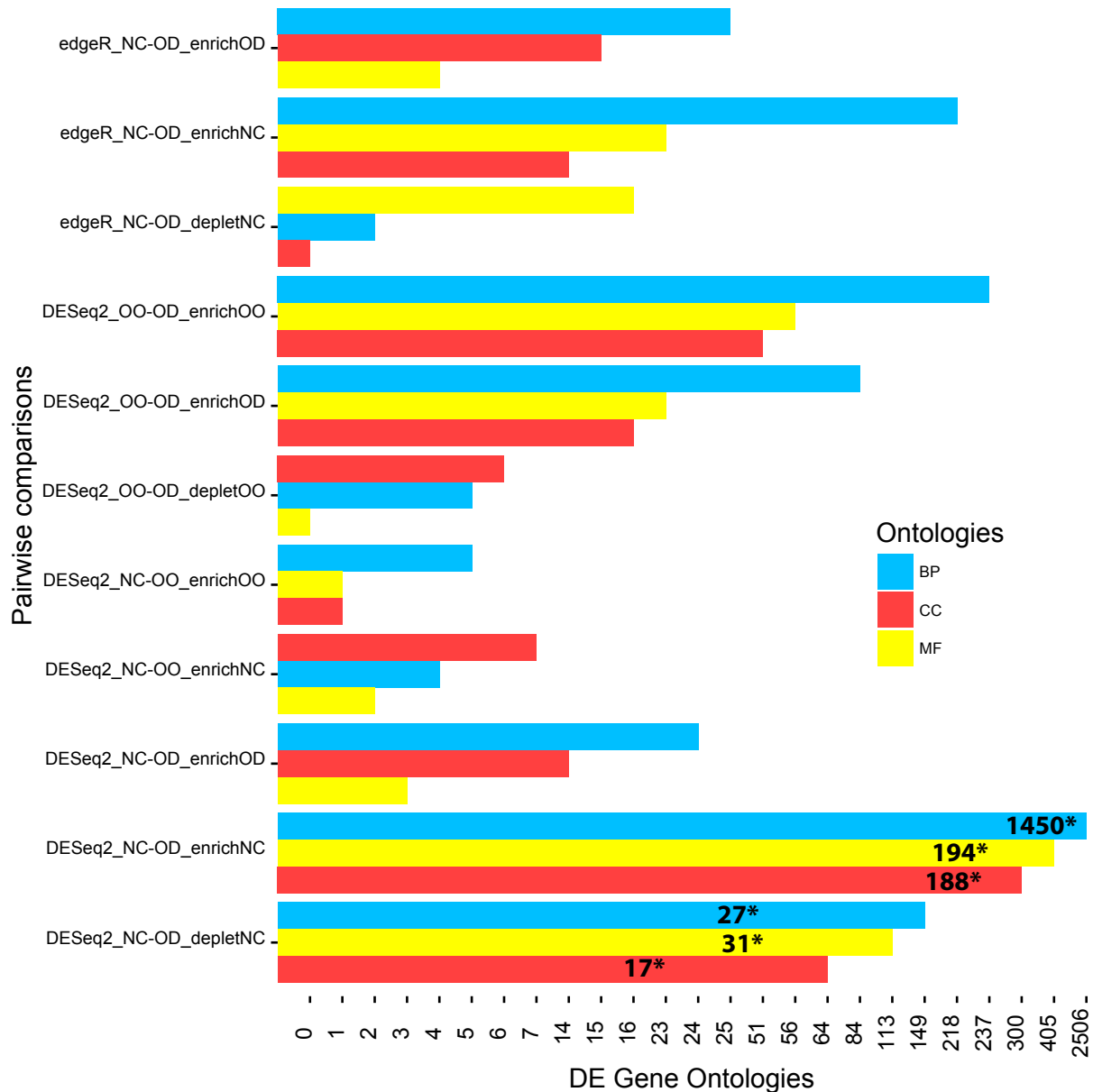
1002 left-side dendrogram are feature clusters 1 and 2. Branch lengths on dendrogram

1003 correspond to Euclidean distances. Heatmap colors represent regulation direction from

1004 normalized, log-transformed fragments per kilobase of transcript per million mapped

1005 reads (FPKM) and zero-centered fold-change values: blue cells are downregulated genes,1006 orange cells are upregulated genes, and grey values show no differential expression. For

- 1007 all comparisons see Figs. S7-8. List on right side are gene ontology (GO) terms discussed
- 1008 in text (full list: Table S4).



1009

1010 **Figure 6. Gene ontology enrichment analyses of differentially expressed (DE) genes**1011 **estimated in GOseq.** Pairwise comparisons of DE gene enrichment tests between NC

1012 (negative control), OO (oil-only), and OD (oil-dispersant) treatments listed by statistical

1013 method (edgeR and DESeq2). Only 11 out of 24 possible comparisons included DE genes

1014 (DEGs) and thus were tested for enrichment. Each test includes method, pairwise

1015 comparison, and whether an experimental condition included overrepresented (enriched –

1016 enrich) or underrepresented (depleted – deplet) gene ontology (GO) terms. Ontology

1017 categories are: Biological Process (BP), Cellular Component (CC), and Molecular
1018 Function (MF). Statistically significant enrichment counts ($P < 0.05$) passing a 1% FDR
1019 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

Hernán Vázquez-Miranda, Brent P. Thoma, Juliet M. Wong, Darryl L. Felder, Keith A. Crandall,
and Heather D. Bracken-Grissom.

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. 2, 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).

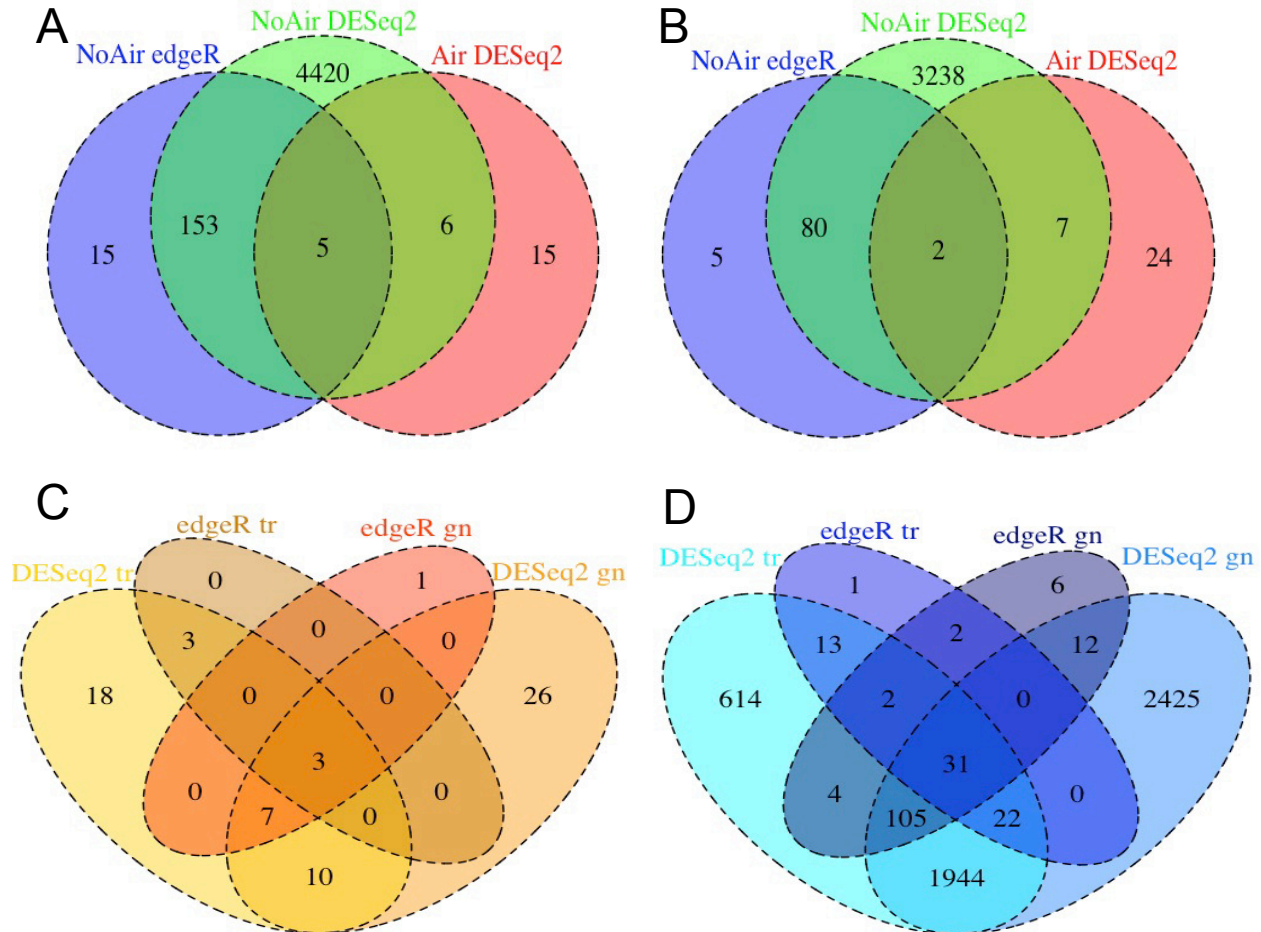


Figure S1. Venn diagrams of shared differentially expressed features between controls (AC-NC) and experimental treatments (OO-OD). (A) Trinity 'genes', (B) Trinity isoforms.

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.

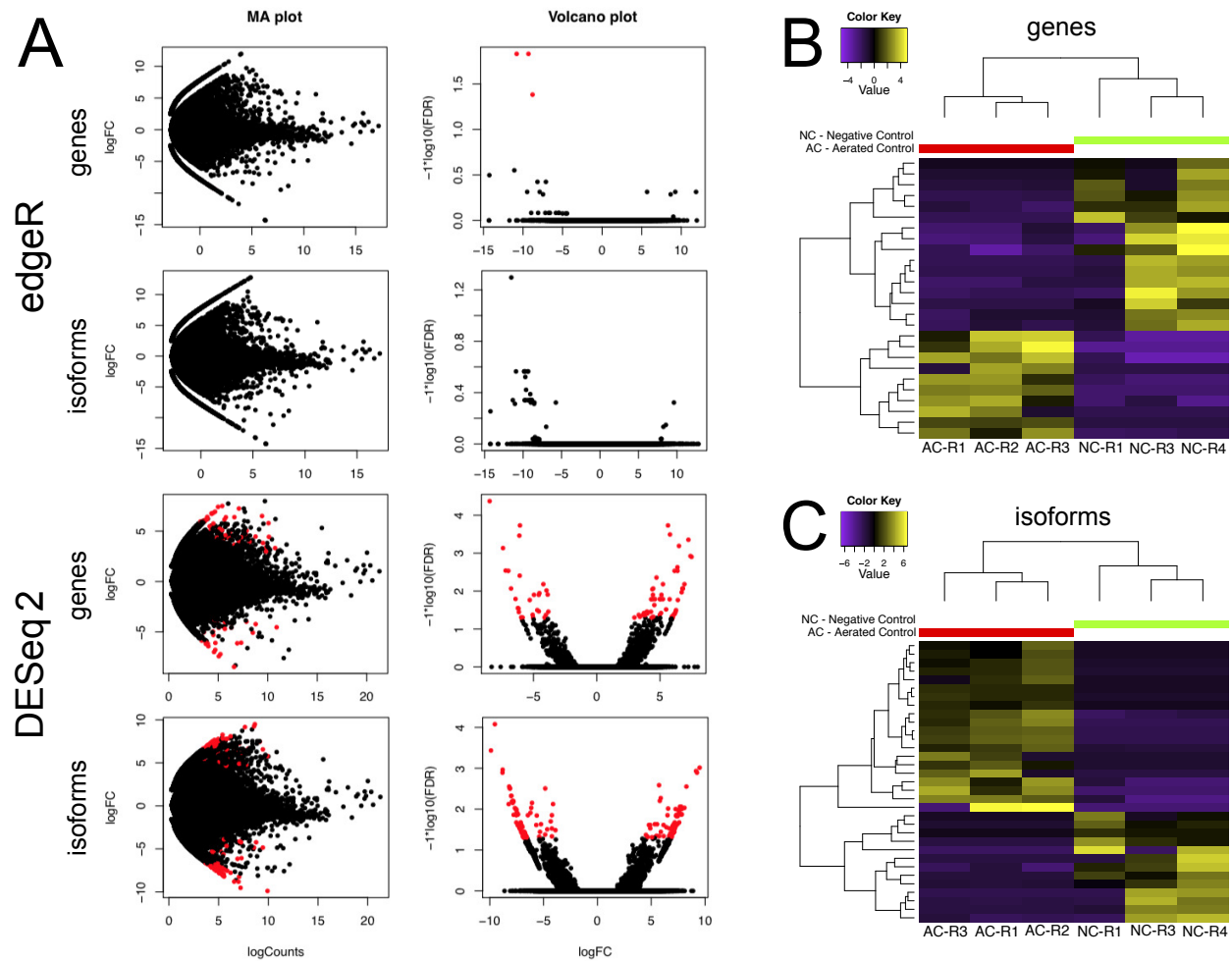


Fig. S2. Differential feature expression in low oxygen conditions between aerated (AC) and non-aerated (NC) controls. A) Left column boxes are pairwise MA plots where X-axes represent mean average expression on counts on a logarithmic scale Y-axes represents fold-changes in gene expression on a logarithmic 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.

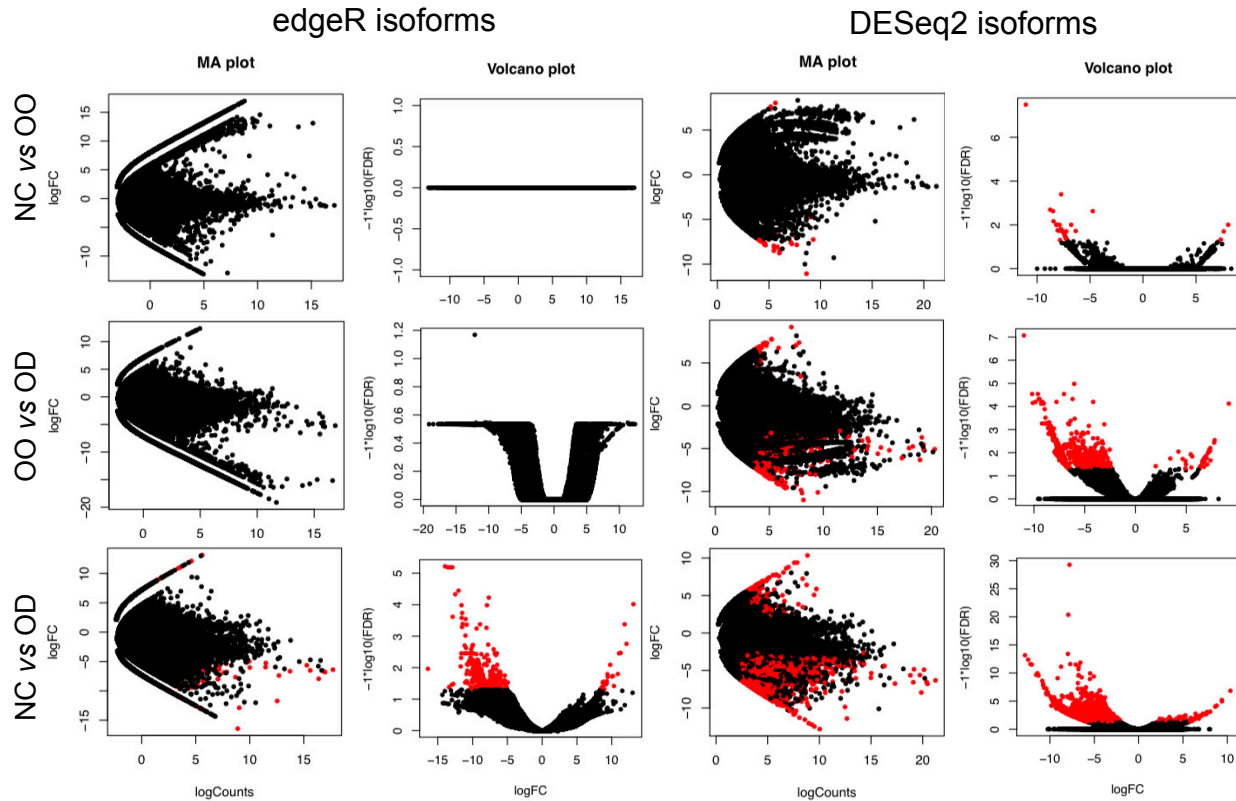


Fig. S3. Pairwise differential isoform expression MA and volcano plots between the negative control (NC) and experimental treatments (OO and OD). In MA plots X-axis represents mean average expression on counts on a \log_2 scale and Y-axis represents fold-changes in gene expression on a logarithmic scale. 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).

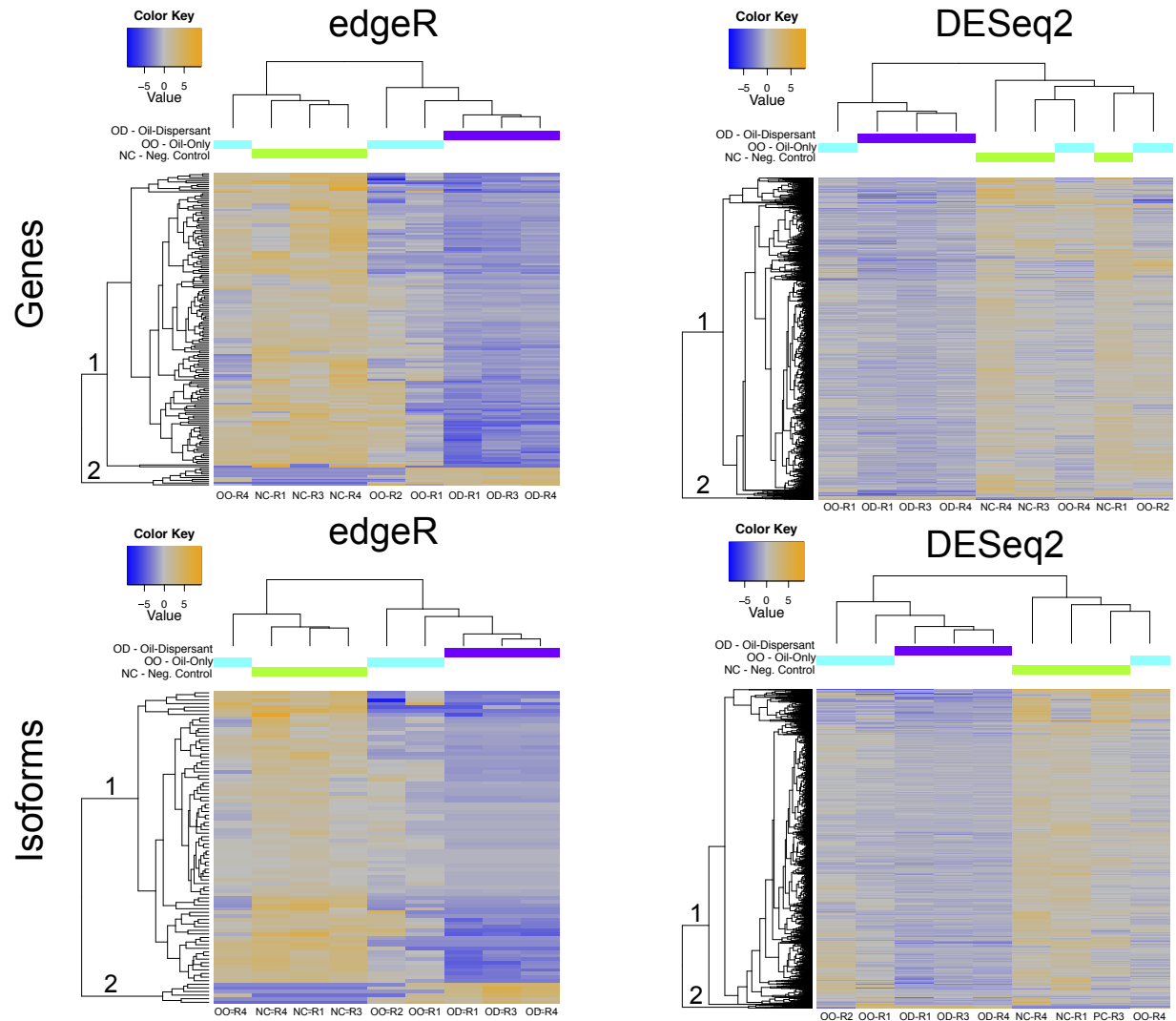


Fig. S4. Differential expression and regulation of features for all experimental treatments related to presence of oil and dispersant ordered by expression profile distances. Columns and top dendrogram correspond to treatments with three replicates each: NC (non-aerated control) in green, OO (non aerated with oil) in cyan, and OO (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

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.

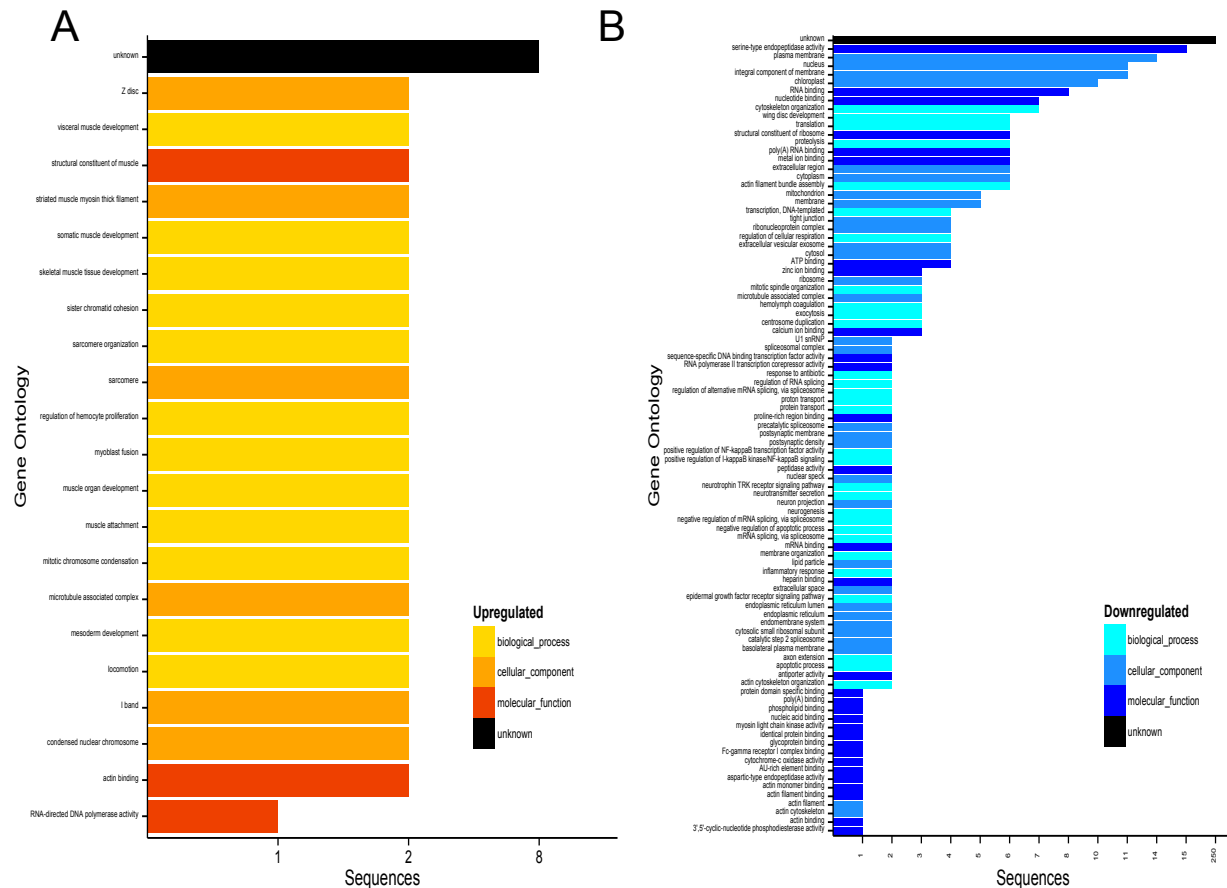


Fig. S5. Gene ontology annotations for genes from edgeR in oil-dispersant (OD). Colors in bar represent major upregulated (A; warm colors) and (B; cool colors) downregulated GO terms (see legend). X-axis indicates element count found in the database. Note last scale number is adjusted to the largest 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.

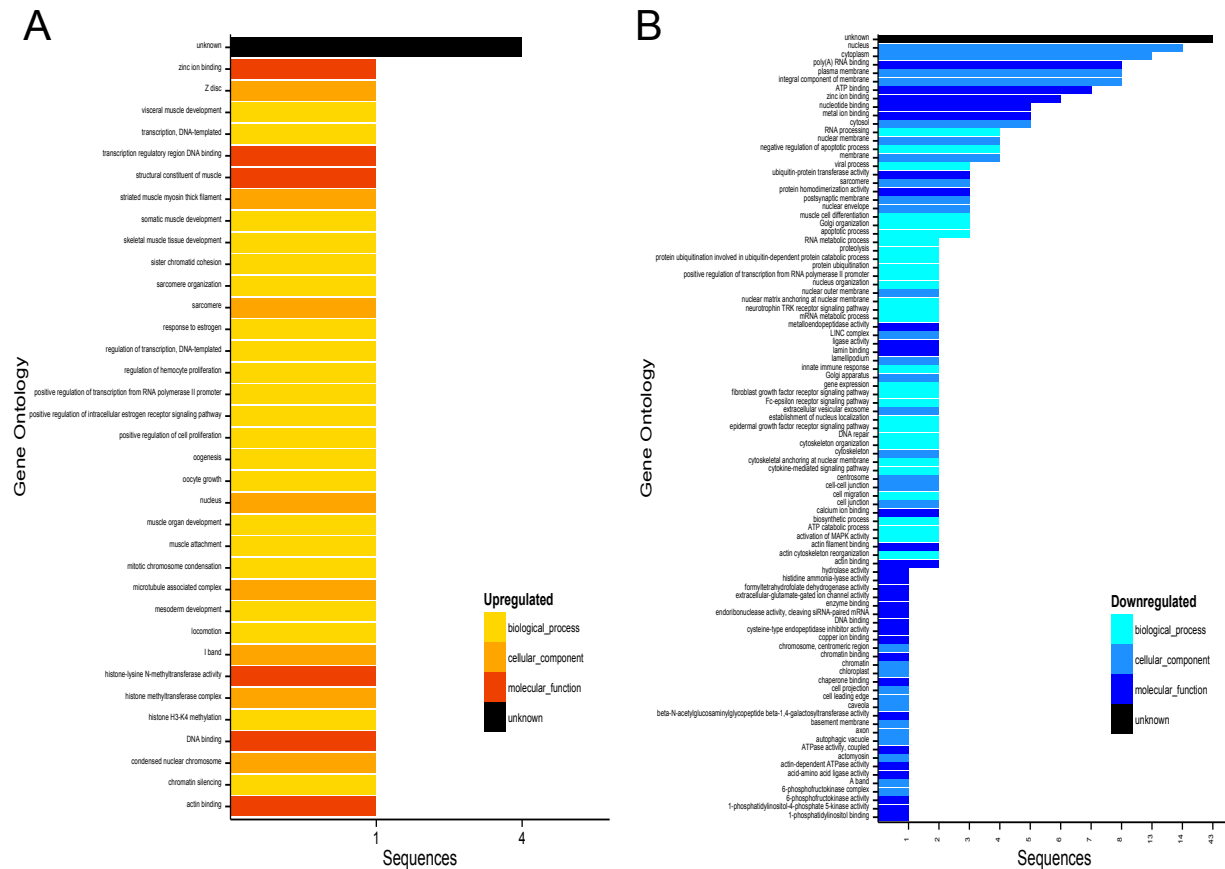


Fig. S6. Gene ontology annotations for upregulated isoforms from edgeR in oil-dispersant (OD). Colors in bar represent major upregulated (A; warm colors) and (B; cool colors) downregulated GO terms (see legend). X-axis indicates element count found in the database. Note last scale number is adjusted 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.

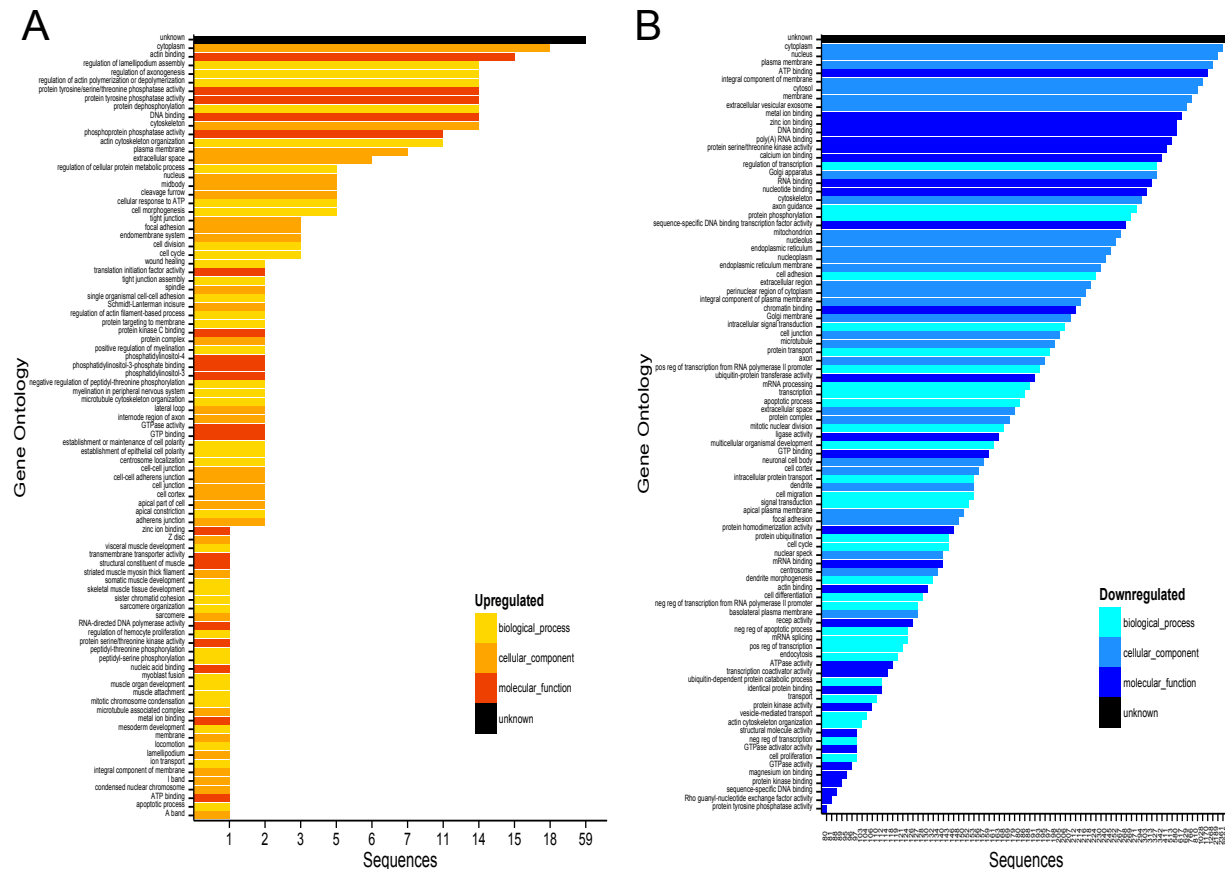


Fig. S7. Gene ontology annotations for genes from DESeq2 in oil-dispersant (OD). Colors in bar represent major upregulated (A; warm colors) and (B; cool colors) downregulated GO terms (see legend). X-axis indicates element count found in the database. Note last scale number is adjusted to the largest 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.

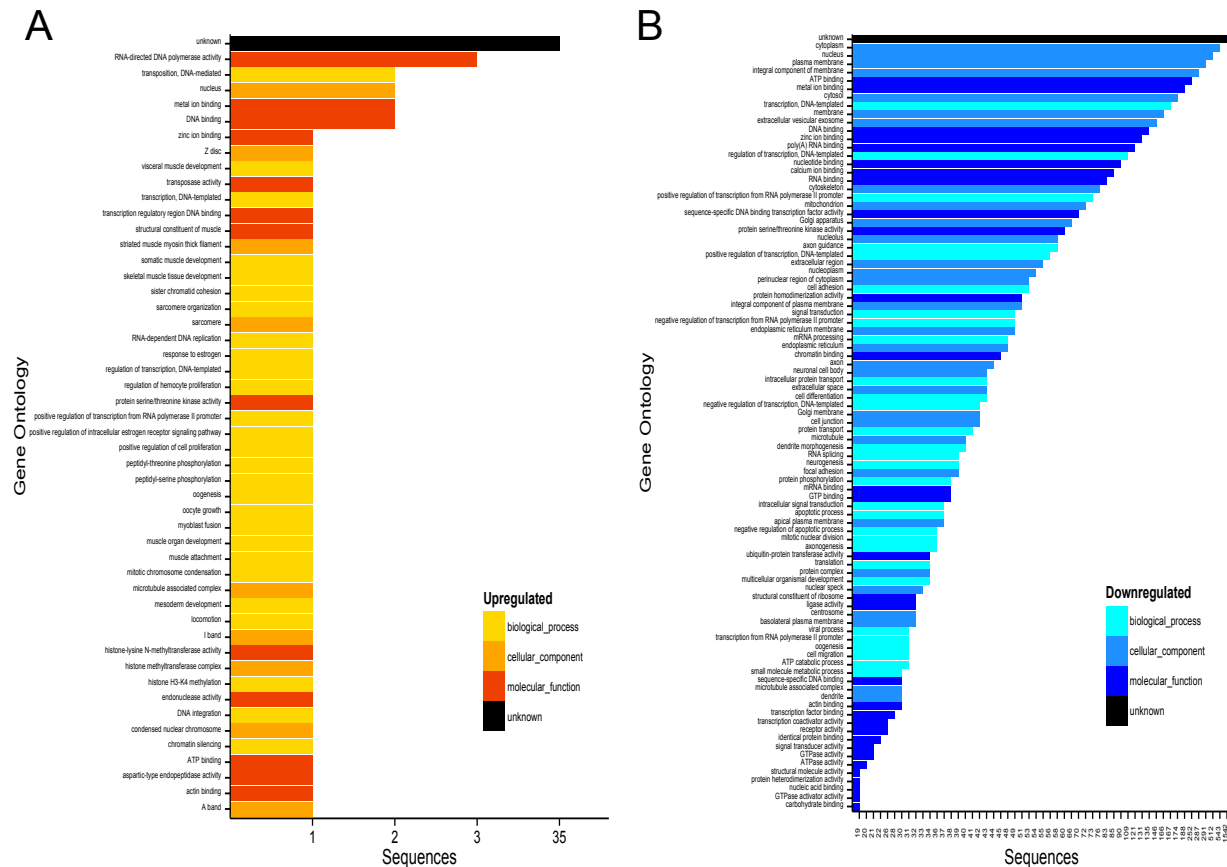


Fig. S8. Gene ontology annotations for upregulated isoforms from DESeq2 in oil-dispersant (OD). Colors in bar represent major upregulated (A; warm colors) and (B; cool colors) downregulated GO terms (see legend). X-axis indicates element count found in the database. Note last scale number is adjusted 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>). (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

- Knowles JR. 1980. Enzyme-catalyzed phosphoryl transfer reactions. *Annual review of biochemistry* 49:877-919.
- Mellors Aa, and Tappel A. 1966. The inhibition of mitochondrial peroxidation by ubiquinone and ubiquinol. *Journal of Biological chemistry* 241:4353-4356.
- Pompella A, Visvikis A, Paolicchi A, De Tata V, and Casini AF. 2003. The changing faces of glutathione, a cellular protagonist. *Biochemical pharmacology* 66:1499-1503.