8-5-2020

Linking Crude Oil and Dispersant Effects To Gene Expression of Marine Sponge Cinachyrella

Yvain Desplat
Nova Southeastern University

Thesis of
Yvain Desplat

Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Science
Biological Sciences

Nova Southeastern University
Halmos College of Arts and Sciences

August 2020

Approved:
Thesis Committee

Committee Chair: Jose V. Lopez, Ph.D.

Committee Member: Jacob F. Warner, Ph.D.

Committee Member: Travis Craddock, Ph.D.

NOVA SOUTHEASTERN UNIVERSITY

HALMOS COLLEGE OF ARTS AND SCIENCES

Linking Crude Oil and Dispersant Effects to Gene Expression of Marine Sponge Cinachyrella

By

Yvain Desplat

Submitted to the Faculty of
Halmos College of Arts and Sciences
in partial fulfillment of the requirements for
the degree of Master of Sciences with a specialty in:

Biological Sciences

Nova Southeastern University

August 2020

Table of Contents

ACKNOWLEDGEMENTS

I would like to thank my major advisor, Dr. Jose Lopez, for giving me the opportunity to work on this project as my master's thesis as well as a research assistant in his laboratory. Thank you for showing me the way around the lab and taking the time to answer the countless questions I had. Thank you for providing the help, the input, and sharing your knowledge on this project that led to its successful completion. Thank you for providing the opportunities to travel to scientific conferences and introducing me to so many peers along the way. Special thanks to my two other committee members, Dr. Jacob Warner (University of North Carolina at Wilmington) and Dr. Tavis Craddock (Nova Southeastern University). Thank you for your lab support, technical advices, helpful comments and guidance you provided throughout this entire project. As well, sincere thanks to M.S. Emily Blake for all the help you provided on lab procedure when I encountered issues. I also thank Nova Southeastern University, its president Dr. George L. Hanbury II, and his committee members, for awarding Dr. Jose Lopez the Professional Faculty Research and Development Grant worth $15K without which this project would not have been possible. Further acknowledgement to the initial funding from the Florida Institute of Oceanography (FIO) to get this project initially started in 2010.

I would like to thank all my lab mates, Dr. Cole Easson PhD., Dr. Lauren Krausfeldt, M.S. Chase Donelly, M.S. Kyle Roebuck, Victoria Pecci, Denise Swack, Deepesh Tourani, Philip Fortman, Yasmin Khan, Elizaveta Shmakova, Shelby Cain, and Alexis Berger. Cole, a very sincere and special thank you for teaching me all the bioinformatics tools required for this project and taking the patience to help me solve the numbers of issues I faced. The completion of this project would have been compromised without your knowledge and help. Chase, thank you so much for all the help, insight and support you provided me with in the lab, helping me out with any issues I had and always keeping the good vibe around the lab; you made the frustrating days better and always found the good jokes to brighten the lab. Kyle, thank you for showing me around the lab and helping me out with the procedure at the beginning, teaching me techniques and protocols without which the completion of this project would have been impossible. To all the others, thank you for making the lab a fun but serious environment, and a place where work can be done in a good mood with great people.

I would also like to give a huge thank you to Dr. Jacob Warner (Ph.D, University of North Carolina Wilmington) who helped me greatly with the bioinformatics part of this project. Without his knowledge in command line-based programming and bioinformatics experience, the completion of this project would have been much more difficult, and I will forever be grateful for your help when this world of the unknown was challenging me. I owe you one.

A very special thank you to my girlfriend, Alexandra Friedlich, for all your love, constant support, and encouragement during the low times. The late-night discussions, your listening ear, making me laugh and giving me back confidence when I felt low, are a few things that lead to the success of this thesis. I am forever grateful for you.

Last but certainly not least, I would like to thank my family. A warm thank you to my grandmother, you supported me financially and morally throughout this project. You taught me the importance of education and the importance of doing a job you love, and I certainly am right now. Thanks to my sister who taught me patience, staying calm in difficult situations, and believed I was capable of great things.

Finally, to my mother. I am and will be forever grateful for the opportunities you gave me so I could follow my dreams. I will never be able to thank you enough for allowing me the opportunity of studying abroad, and by supporting me in every way possible. You helped me believe in myself and knew I was capable of pushing beyond my limits. Thank you for teaching me to take things day by day, that behind every low there is a high hiding, and that better days are always ahead. I dedicate this thesis to you. I promise to always make you proud.

List of Abbreviations

ACS: American Cancer Society

BAM: Binary Alignment Map file. Compressed version of a SAM file.

BLAST: Basic Local Alignment Search Tool

CCR: Center for Collaborative Research

cDNA: complementary DNA

CE-WAF: Chemically Enhanced Water Accommodated Fraction

D: Dispersant treatment

DEEPEND: Deep Pelagic Nekton Dynamics of the Gulf of Mexico

DNA: Deoxyribose Nucleic Acid

DWHOS: Deep Water Horizon Oil Spill

EBP: Earth Biome Project

FIO: Florida Institute of Oceanography

FIU: Florida International University

GIGA: Global Invertebrate Genomic Alliance

HSP: Heat Shock Protein

LN2: Liquid Nitrogen

mRNA: messenger RNA

NCBI: National Center for Biotechnology Information

NGS: Next Generation Sequencing

NSU: Nova Southeastern University

NSUOC: Nova Southeastern University Oceanographic Center

O: Oil treatment

OD: Oil:Dispersant treatment

PAH: Polycyclic Aromatic Hydrocarbon

PCA: Principal Component Analysis

PCB: Poly-Chlorinated Biphenyl

PorToL: Porifera Tree of Life

QC: Quality Control

RNA: Ribo-Nucleic Acid

RPM: Revolution Per Minute

SAM: Sequence Alignment Map

SEM: Scanning Electron Microscopy

TEM: Transmission Electron Microscopy

TSG: Tumor Suppressor Gene

WAF: Water Accommodated Fraction

List of Figures

List of Tables

ABSTRACT

Because of their filter-feeding lifestyle, sponges (Phylum Porifera) have shown to be bio-accumulators of heavy metals, and bio-monitors for polychlorobiphenyl (PCB) contaminants. Furthermore, marine sponges have been shown to be regulators of reef ecosystems by fulfilling many ecological functions. However, very little is known about their behavior in the face of environmental changes. Consequently, our lab has developed the reef sponge Cinachyrella spp. as novel experimental model. We have designed an experiment to study the effects of WAF (Water Accommodated Fraction), Corexit 9500 dispersant, and CE-WAF (Chemically Enhanced Water Accommodated Fraction)mixtures, in an effort to mimic the conditions of the Deepwater Horizon oil spill of 2010. Three replicate of dosing experiments (labeled as X1, X2, X3) were performed on Cinachyrella based on standard CROSERF protocols. Over 80 different sponges were exposed to sublethal amounts (0.5 ppm) of oil, oil mixed with 10% Corexit dispersant (OD), and dispersant for 1 and 24 hours.

This thesis took up the characterization by using RNA-Seq data to determine the differential expression genetic response on a Florida reef native sponge from the Cinachyrella genus. Overall, from 24 Cinachyrella individuals, 31,571 total transcripts were eligible for genetic profiling. Overall, 12,913 transcripts have shown significant differential expression, among which 7,863 were upregulated, and 5,058 were down regulated. These differentially expressed transcripts included transcripts from the sponge holotranscriptome, coding for protein structure and integrity, cancer related proteins, cell survival proteins, apoptosis along with other essential protein for the organism's survival. A large number of "orphan" uncharacterized sponge genes (and putative protein products) with no previously known function were identified, providing a scope for future work. Overall down-regulation of genes was dominant over upregulation of genes. Major genetic responses to oil had started mostly after 1 hour of exposure and higher response was seen after 24 hours of exposure for dispersant and oil:dispersant mixtures. CE-WAF Oil:dispersant mixtures appeared most harmful to the sponge after longer exposure This study confirmed Cinachyrella as a suitable research model organism from Florida reefs.

Key words: Bioindicator, Cinachyrella, Dispersant, Deep Water Horizon, Gene Expression, Oil, Sponges, Transcriptomics

INTRODUCTION

Marine sponges, one of the earliest branching living taxa on earth, are invertebrate eukaryotic animals that are found worldwide, and at a variety of depths (Pisani et al, 2015; Feuda et al 2017). They are capable of filtering thousands of liters of water per day. Some species can filter up to 50,000 times their body volume in one day (Tree of Life Project; www.tolweb.org/porifera/). Such filtering abilities are important for these organisms as it is their primary mean of feeding mechanism. Sponges are filter feeders, with complex anatomy. Water flows through their tissues thanks to the flagella's vibration of the amoeboid cells. The vibration creates a current that pulls the water through the sponge pores and nutrients are retained in the collar of the choanocyte cells. The water then exits the sponge through the osculum back in the water column (Fig. 1).

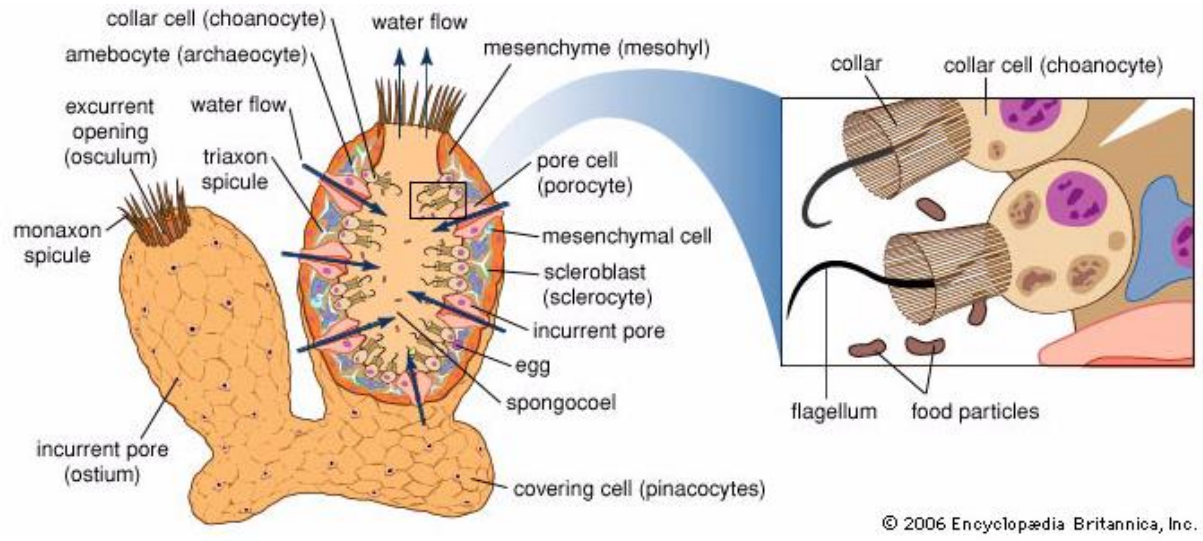


Figure 1: Basic anatomy of a marine sponge. Michele Sara "Sponges" Britannica. https://www.britannica.com/animal/sponge-animal

In addition, sponges harbor a wide array of microorganisms within their tissues (Hentschel et al 2012; Lopez 2019). The relationship between the host and those microorganisms is described as symbiosis, a complex interaction in which both parties benefit from the other. Water pumping by the host will provide the symbionts with nutrients and other life necessary compounds. On the other hand, commensal microbes may help the host by degrading some toxic substance for the host, or producing some substances needed for the survival of the host.

By pumping these important volumes of water, sponges are exposed to any suspended or dissolved matter in the water column, making them prime candidates for water quality studies. Sponges have already been shown to be potential bio-accumulators of heavy metals (Batista et al., 2014; Rao et al., 2006; Rao et al., 2008; Srikanth, K., & Rao, J. V., 2016; Webster et al. 2001) as well as bio-monitors for polychlorobiphenyl (PCB) contaminants (Pawar, 2013; Perez et al., 2003). Sponges have also shown to contain many natural product compounds capable of having anti-tumor or cancer treating properties of interest in nowadays medicine (Garson 1994; Mehbub et al, 2014; Perina et al., 2011).

While research has been done on PCBs and heavy metals in poriferan, very few studies have looked at the effect of polycyclic aromatic hydrocarbons (PAHs) on such organisms. Given the pumping abilities of these organisms, sponges may also be excellent indicators to understand the ecotoxicological impacts of such PAHs from both natural and anthropogenic sources. PAHs are more commonly known as oil, one of the most used resources of our era. With the exponential rate of increase in world population and the constant need for more oil related products, a race to extraction has been happening over the past two decade, with the challenge of finding more and more oil slicks in a worldwide diminishing stock. Oil companies are consequently forced to look in unexplored areas, sometimes very deep in the earth for this black gold. While some areas are known to have potential land reservoirs such as in the middle east (Saudi Arabia, United Arab Emirates; Lakhani et al., 2016), little is known about offshore shales. Unfortunately, with this rush of finding new slicks to drill, proper background research is sometimes overlooked, leading to environmental catastrophes with very serious impacts on ecosystems

On April 20th of 2010, the BP-operated Macondo Prospect extracting station, located in the Gulf of Mexico (28°44'17.30"N, 88°21'57.40"W), exploded and discharged a total of 780,000 m^3 (210 million US gallons) of crude oil, in what is considered the largest marine oil spill in the history of the petroleum industry (Fig. 2). Killing eleven of the platform operators, the Deepwater Horizon oil spill (DWHOS) also affected the entire marine ecosystem from pelagic to benthic organisms, from open ocean to coastlines, impacting 180,000 km^2 (69,500 sqm) (Amos, 2010) the shores of Florida, Louisiana, Mississippi, and Alabama (Fig. 3). To remove the spilled oil, 7000 m^3 (1.84 million US gallons) of COREXIT oil dispersant (COREXIT EC9500A and COREXIT AC9527A) were poured directly at the well head and across the oil spill (United States Coast Guard & Nation

Response Team, 2011). Authorities estimated the oil spill had covered a total area in the ocean comparable to the size of Oklahoma.

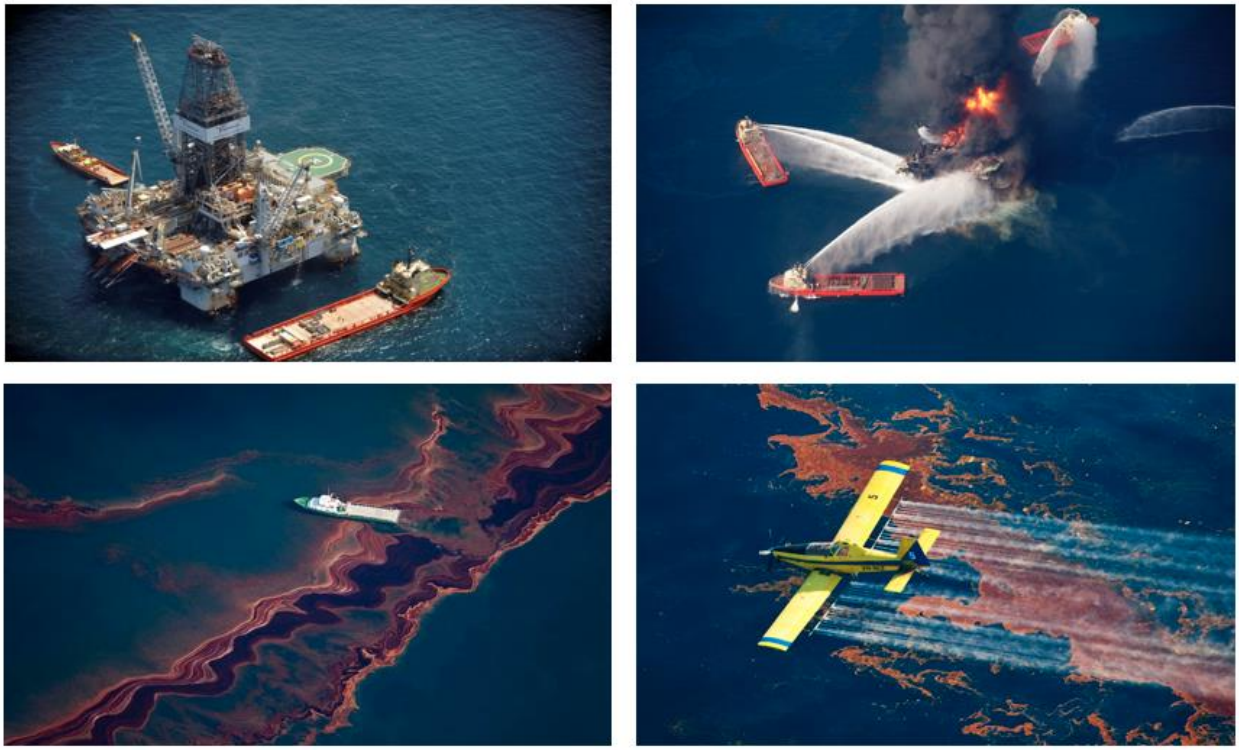


Figure 2: Snapshots of the Deepwater Horizon BP offshore rig in operation (top left), Deepwater Horizon after the explosion (top right),of the spill(bottom left) and dispersant spread over the spill bottom right. Andy Rowell. "New Evidence BP's Spill Dispersant Caused Harm to Humans and Wildlife" Oil Change International. http://priceofoil.org/2015/04/07/new-evidence-bps-spill-dispersant-caused-harm-humans-wildlife/

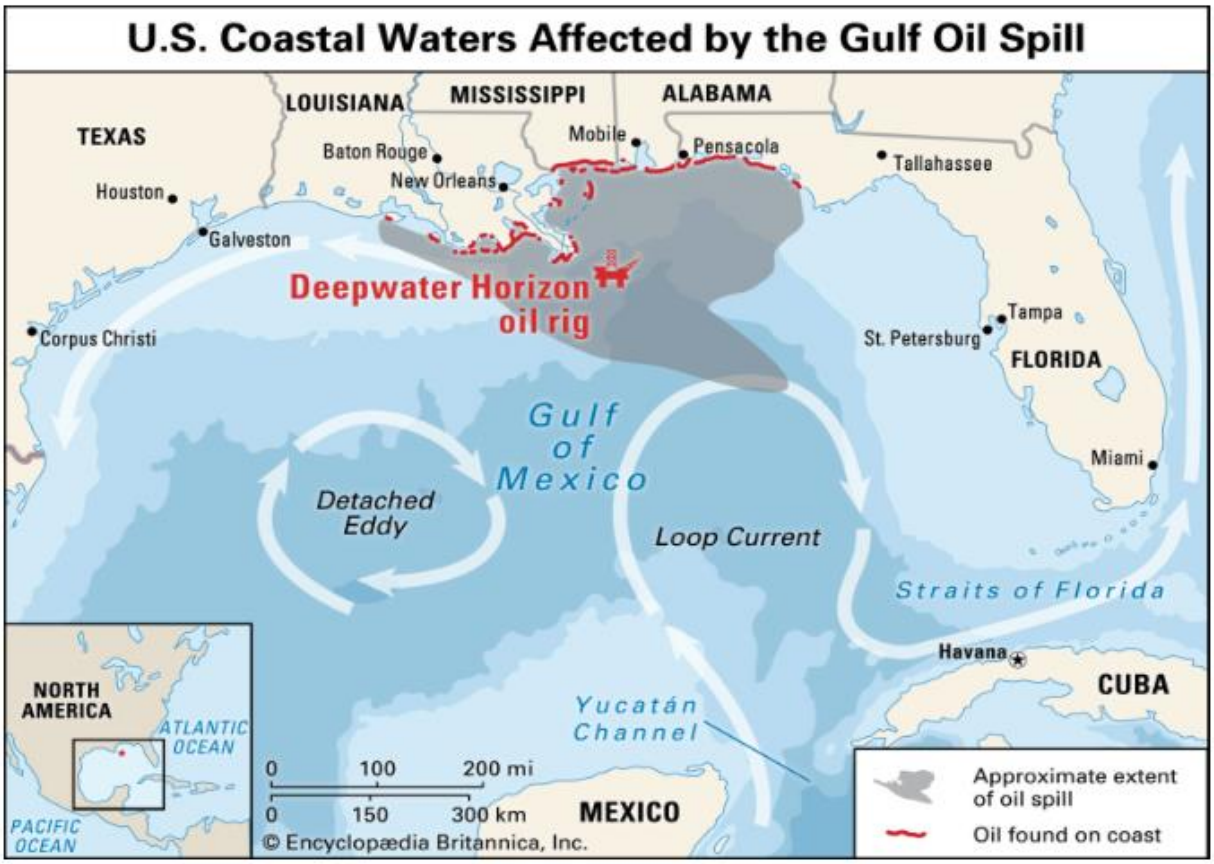


Figure 3: Estimated size of the Deepwater Horizon oil spill. A total of 210 million US gallons of oil were spilled, covering a. surface area of the size of Oklahoma (180,000 km2). Richard Pallardy. "Deepwater Horizon Oil Spill", Britannica https://www.britannica.com/event/Deepwater-Horizon-oil-spill

Environmental impacts of such oil spills can be seen throughout the entire water column and have negative impact on all levels of marine and land ecosystems. Studies have shown impacts on small organisms such as copepods (Almeda et al., 2014), smaller species of fish (Dubansky et al., 2014 ; Holth et al., 2008), microbial communities (Lamendella et al., 2014; Mason et al., 2014; Rodriguez et al., 2015), and marine megafauna (Rooker et al., 2013; Incardona et al., 2014). Most research efforts have focused on well-known organisms, umbrella species, or species of commercial interest (Ruiz-Ramos et al., 2017 ; Studivan et al., 2015;White et al., 2012; Putman et al., 2015; Venn-Watson et al., 2015; Xu et al., 2017). Not only having an impact on the marine ecosystem, the oil spill has disrupted land ecosystem as well with damages seen from marshland (Whitehead et al., 2011; Deis et al., 2017; Dubansky et al., 2017), to land animals such as birds and freshwater organisms. Finally, the oil spill also affected humans, especially the people who worked to get rid of the oil on the shorelines and coastal living areas (Liu et at., 2016; Liu et al.,

2017). Very few studies have focused on the impact of a spill from a genetic standpoint. Most studies have focused on visually assessable impacts of the spill without paying a whole lot of attention to the core and starting point of what is seen at the phenotypic scale, the genotype. On the same note, many studies have shown the harmful impact oil, dispersant, and oil:dispersant mixtures from an embryology and developmental perspective, but very few have looked at the effect of these chemicals on fully grown organisms (Milligan et al., 2019)

Consequently, as an effort to understand the effect of oil spills and dispersant on marine life, as well as on humans, more and more studies should focus on the impact of such chemical effects on the genetic scale. The genetic changes are the very first witnesses of any changes occurring in an organism and can be detected early on after exposure. Looking for the impact of these chemical at the molecular level not only provides insight on the exact changes the organisms is going through, but can also provide continuous and accurate monitoring of these changes over shorter periods of time. This allows for testing hypotheses testing that would not be feasible to the sole naked eye, as phenotypic changes could take much more time to be noticeable.

Here in this study, the effect of crude oil and dispersant was investigated, using a relatively novel model marine sponge, from the Cinachyrella genus (Fig. 4). Commonly known as the global "golf ball" sponge, Cinachyrella sponges are found worldwide in tropical waters. In this study we are focusing on Cinachyrella spp. found in the Caribbean islands, as well as of the coast of Brazil and off the coast of Florida. Our initial hypotheses included monitoring effects on the sponge microbiome, but the results did not appear conclusive (Lopez unpublished; Cuvelier et al 2014).

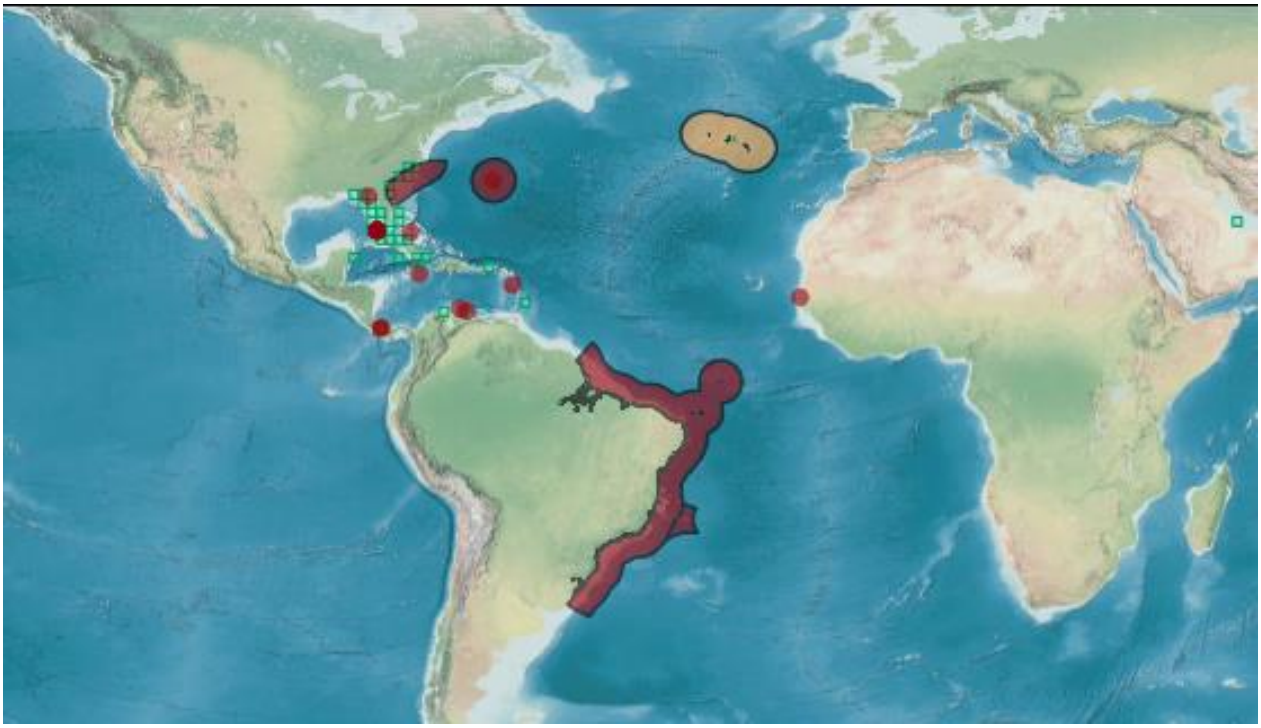


Figure 4: Geographical distribution of C. alloclada , C. Apion, and C. kuekenthali in the Western Atlantic Ocean and the Caribbean. Rob Van Soest. *"Cinachyrella alloclada"* World Porifera Database.
http://www.marinespecies.org/porifera/porifera.php?p=taxdetails&id=171291#distributions

Sponge genomics and transcriptomics in marine biology

Genomics is a multidisciplinary field of biology that focuses on the evolution, mapping, structure, function, and editing of genomes (Hilgartner, S. 2017; Koonin, E. 2009; Woese, C.R. 2004). Unlike genetics that focuses on specific gene function one at a time, genomics aims at the collective characterization and quantification of all of an organism's genes, their interrelations and influence on the organism as a whole. Genomics also involves the sequencing and analysis of genomes through uses of high throughput DNA sequencing and bioinformatics to assemble and analyze the function and structure of entire genomes. Advances in genomics have triggered a revolution in discovery-based research and systems biology to facilitate understanding of even the most complex biological systems. The genomics field has benefited many other fields such as conservation, medicine, and biological engineering. Genomics is divided in several categories including metagenomics, epigenomics, or transcriptomics to cite a few. Here we focused on the transcriptomic field.

Transcriptomics is field of study in which the set of all RNA transcripts, including coding and non-coding, of a whole organism is being studied. The term can also sometimes be used to refer to all RNAs, or just mRNA, depending on the particular experiment. The advent of high-throughput technology led to faster and more efficient ways of obtaining data about the transcriptome. mRNA-seq has become the preferred method and has been the dominant technique to study transcriptomics since the 2010s (Mortazavi et al., 2008). Data obtained from sequencing the transcriptome is used to study many different aspects of a living organism. Cellular differentiation, carcinogenesis, transcription regulation and biomarker discovery are some of the processes among others that this field can provide an insight into (Godoy et al., 2018; Szabo 2014). This type of data also finds applications in invitro fertilization or assigning phylogenetic relationships from a process of evolution perspective (Zhao et al., 2019).

The evolution of the genomics field has helped scientist understand better how organisms live in their environments and how they react to changes in their living environment. It also led to the creation of several projects, that aim to sequence, assemble, and annotate genomes of various species to better understand the natural world as a whole such as Genome 10K, GIGA, or EBP. Genome 10K is a consortium aiming to sequence the genome of at least one individual from each vertebrate genus, which represent approximately 10,000 vertebrates (Koepfli et al., 2015). GIGA (Global Invertebrate Genomics Alliance) aims to promote standards that will facilitate comparative approaches to invertebrate genomics and collaborations across the international scientific community (GIGA, 2014). The EBP (Earth Biogenome Project) aims to sequence, catalog, and characterize the genomes of all Earth's eukaryotic biodiversity over a period of ten years (Lewin et al., 2018).

Sponges are believed to be very important on corals reefs, due to the fulfilment of several ecological functions. It is believed that sponges play a big role in nutrient cycling on reefs systems. They might be important factors responsible for both good and bad shifts in water quality. Scientist have been looking at sponges respiration and the amount of released nitrogen, bacteria collection resulting from water filtration, as well as the role these bacteria play in a coral reef ecosystem. It has been hypothesized that these bacteria are able to transform the dissolved nitrogen gas in the water into usable nitrogen for the sponge host. Interest has also been put on the capability of these bacteria to transform the ammonium release by the breathing of the sponge into nitrogen gas that is then released in the atmosphere. Through these hypotheses it has been speculated that these

processes would lower the nitrogen excess in coral reef environments, which would prevent harmful ecosystem changes, as well as enhancing the survival of other species in the same area (PorToL).

However, only a small amount of studies has been realized on poriferan and their genetic heritage (Riesgo et al., 2012, Riesgo et al., 2014, Risego et al., 2019) and their development (Qiu et al., 2015). While most studies have focused on microbial communities of sponges, as well as on phylogenic assignment, a small amount of interest has been given to the pure genetic aspect of a single organism. As of today, three marine sponge genomes have been sequenced, Amphimedon queenslandica, a Demospongiae found off the coast of Australia (Fernandez-Valverde et al., 2015), Aplysina aerophoba (assembly and annotation are available on NCBI but no paper has been published), and Tethya wilhelma (Francis et al., 2017), two Demospongiae from the Mediterranean Sea. However, among these three, A. queenslandica is the most well annotated genome that can be used for genomics studies. The two other genomes are still partially annotated and render their use more challenging in genomics studies. Other genomic data such as transcriptome or mitochondrial genomes is available for other sponges (Ephydatia muelleri, Sycon ciliatum, Suberites domuncula, Oscarella carmela, Crella elegans). However, most of this data is used to focus on taxonomic on phylogeny assignment (Leis, 2017; Riesgo et al., 2019, Qiu et al., 2015). This project hence aimed to provide more resources for porifera research, as well as potentially establishing Cinachyrella as a new bio-indicator in the face of environmental events, as well as a new laboratory research model organism. In order to do so, transcriptome changes of sponges exposed to a sub-lethal concentration of Macondo crude oil (WAF, O), COREXIT CE9500A oil dispersant (D), and a mixture of Macondo crude oil and COREXIT CE9500A oil dispersant (CE-WAF, OD) were analyzed. Quantification of the transcriptome, under different conditions, will give an insight on the effect of these chemicals on gene expression (up-regulation, down-regulation, or no change).

As well, the full transcriptome of the sponge was sequenced, assembled, and annotated to provide additional genomics resources for further research projects. This follows a preliminary study done in 2013 by Emily Smith, now a Data Analyst for the Genomics Core Facility at Nova Southeastern University's Center for Collaborative Research (CCR). The first results identified 483 contigs (contiguous sequence of DNA created by assembling overlapping sequenced fragments. A group of clones representing overlapping regions of the genome) that have been

matched to existing genes. Those genes have been linked to different biological pathways and fell into different categories. Such categories included metabolism pathways, cellular processing pathways, or environmental information processing pathways (E. Smith, 2013, unpublished).

However, only partial results have been produced. Only the first set of samples have been analyzed, but in order for the results to be relevant, the experiments needed to be replicated with the analyzes of more samples. In 2012, Knight and his colleagues have highlighted the importance of replicated experimental design (Knight et al., 2012). Indeed, producing only few results based on few samples can be helpful in some cases, but most of the time it is required to replicate the experiments and analyses because the first analyses only scratch the surface and do not provide in depth results.

It is thus highly recommended to replicate these experiments in order to have statistically relevant results. And for all the reasons stated above, this oil and dispersant study is no exception. Consequently, as a follow up to Emily Smith's study, more replicates of the experiments previously performed will be executed. These replicates will allow a more statistically significant quantification of the effect of oil and dispersant on this selected marine sponge. A particular attention will be given to the human disease pathways, which are under the scope of many studies worldwide.

HYPOTHESES and OBJECTIVES

It is known that environmental disturbances can cause much damage to the ecosystem as a whole (Barange et al. 2014; Chueng et al., 2012; Hautier et al. 2015; Hofmann et al., 2014). In the case of oil spills, certain organisms may be the first to be affected. Whether we look at the lower parts of the food chain (phytoplankton, zooplankton), the megafauna like pelagic fishes, mammals, birds, or the vegetation, all face the consequences of the catastrophe, and the effects can be seen throughout the entire water column, as well as on the affected shores(Bayha et al., 2017; Deis et al., 2017; Dubansky et al., 2014; Goodbody-Gringley et al., 2013; Incardona et al. 2014; Khanna et al., 2013; Powers et al., 2013; Rodriguez et al., 2015; Venn-Watson et al., 2015). Thus using the marine sponge Cinachyrella as a bio-indicator species, along with genomics analyses, it should be possible to witness shifts in environmental conditions, and correlate them with specific events. In this study, we expect that genetic expression in the marine sponge Cinachyrella will differ, depending on the environmental conditions it has been exposed

to: only oil (WAF), only dispersant, and oil:dispersant mixtures (CE-WAF), as stated by the following hypotheses:

1- There is a difference in genetic expression of stress response related pathways, cell survival, and tissues cohesiveness when sponges are exposed to oil (WAF)
2- There is a difference in genetic expression of stress response related pathways, cell survival, and tissues cohesiveness when sponges are exposed to dispersant
3- There is a difference in genetic expression of stress response related pathways, cell survival, and tissues cohesiveness when sponges are exposed to oil:dispersant mixtures (CE-WAF)

MATERIALS and METHODS

Sponges Collection and Culturing

Nova Southeastern University Oceanographic Center (NSUOC) SCUBA collection team sampled 29 sponges from the Hollywood's first reef, Florida, USA. All sponges were collected under a standard Florida fishing license (issued by the FL Fish and Wild Commission – myfwc.com)

Ambient seawater samples were collected at the same sponge samples sites. Twenty-four sponges were brought back to the lab and were then acclimated individually in temperature-controlled culture tanks at Florida International University. After acclimation, the sponges were exposed to different environmental conditions for different periods of time. After each selected time mark, sponges were taken out of the water, sacrificed and prepared for RNA extraction.

The remaining five specimens were immediately flash frozen using LN2 and conserved in -80°C freezer as environmental samples for whole transcriptome applications. Only three samples were however considered. The three samples were denominated Cin-Env-1, Cin-Env-2, and Cin-Env-3 were selected, RNA was extracted (extraction protocol can be found in the genetic expression profiling extraction method section) and sent to GENEWIZ LLC. (South Plainfield, NJ, USA). Frozen specimen were stored in a -80°C freezer in Dr. Lopez's lab and are identified as follow: Cin-Env-1 corresponds to specimen labeled 17-8-2011-1 Environmental

Cinachyrella Field sp., Cin-Env-2 is labeled 17-8-2011-2 Environmental Cinachyrella Field sp., and ., Cin-Env-3 is labeled 17-8-2011-3 Environmental Cinachyrella Field sp.

The fully detailed sampling and culturing method has been described in Smith E., 2013. The taxonomy of the sponges was determined by spicule preparations and ultimately via the presence of a Group I mtDNA intron following the guidelines of Schuster et al 2017

Whole Transcriptome

cDNA Libraries Preparation and Sequencing

RNA library preparations and sequencing reactions were conducted at GENEWIZ, LLC. (South Plainfield, NJ, USA). RNA samples were quantified using Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA) and RNA integrity was checked using Agilent TapeStation 4200 (Agilent Technologies, Palo Alto, CA, USA).

RNA sequencing libraries were prepared using the NEBNext Ultra RNA Library Prep Kit for Illumina using manufacturer's instructions (NEB, Ipswich, MA, USA). Briefly, mRNAs were initially enriched with Oligod(T) beads. Enriched mRNAs were fragmented for 15 minutes at 94 °C. First strand and second strand cDNA were subsequently synthesized. cDNA fragments were end repaired and adenylated at 3'ends, and universal adapters were ligated to cDNA fragments, followed by index addition and library enrichment by PCR with limited cycles. The quality of the sequencing library was validated on the Agilent TapeStation (Agilent Technologies, Palo Alto, CA, USA), and quantified by using Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA) as well as by quantitative PCR (KAPA Biosystems, Wilmington, MA, USA).

The sequencing libraries were clustered on a single lane of a flowcell. After clustering, the flowcell was loaded on the Illumina HiSeq instrument (4000 or equivalent) according to manufacturer's instructions. The samples were sequenced using a 2x150bp Paired End (PE) configuration. Image analysis and base calling were conducted by the HiSeq Control Software (HCS). Raw sequence data (.bcl files) generated from Illumina HiSeq was converted into FASTQ files and de-multiplexed using Illumina's bcl2fastq 2.17 software. One mismatch was allowed for index sequence identification.

Whole Transcriptome Assembly and Annotation

After reviewing the sample QC and data QC, the data of the sample Cin-Env-1 was chosen by GENEWIZ as the best data set, and was input to Trinity v2.5, de-novo assembler, for transcriptome assembly (Grabherr et al., 2011). Trinity is a three steps program. In the first step input reads are broken down into k-mers (nucleotide of a certain length). The second step consists of clustering and assemble these k-mers together using deBruijn graphs. Finally the last step processes these deBruijn assembled nucleotides sequences into transcripts.

One de-novo assembled transcriptome was created with a minimum contig length of 200 bp. Transrate v1.0.3 was used to generate statistics for the de-novo assembled transcriptome. EMBOSS tools GETORF was then used to find the open reading frames within the de-novo assembled transcriptome. The de-novo transcriptome assembly was then annotated using Diamond BLASTx setting a e-value of $1e^{-04}$ as the threshold for a hit to be considered valid within the NR database of the NCBI repository (Buchfink et al., 2015).

Genetic Profiling

WAF, CE-WAF, and Dispersant Solution Preparation

Crude oil was obtained directly from BP (SOB-20100622-084; SOB-20100624-00) and Corexit 9500 from Nalco Holding Company. Approximately 100 L of ambient seawater that had been collected with the sponges, were used for preparing three primary treatments: i) water accommodated fractions (WAF) of oil, ii) chemically enhanced (Corexit 9500) WAFs or CE-WAFs, and iii) Corexit dispersant only.

Water accommodated fractions - WAF (oil), and chemically enhanced WAFs - CE-WAF (oil:dispersant), and dispersant (D) only solutions were prepared using standard protocols outlined by CRO-SERF (Aurand and Coelho 2005). This method was conservative approach given both the small amount of primary material available, and the little knowledge of toxicity level of these chemicals. The WAF solution was prepared using 4.07 grams of weighed Macondo crude oil, which was then mixed in 20 liters glass bottle filled with sea water collected at sample collection site. The CE-WAF solution was made of 4.07 grams weighed crude oil and 0.42 grams

weighed of COREXIT 9500 and mixed to another 20 liters glass bottle filled with ambient sea water collected at Dania Beach sampling sites. Finally, the dispersant only solution was prepared by mixing 0.42 grams of weighed COREXIT 9500 in 20 liters of sea water from sampling site. Ultimately, solutions were prepared with sublethal doses to ensure survival of specimens throughout the experiment. The complete mixture preparation protocol is highlighted in Smith E., 2013.

Three replicate oil dosing experiments (labeled as X1 – X3) were performed on a total of 24 sponges based on standard CROSERF protocols (Table. 1)

Table 1: Table representing the experimental design followed in this study with sample naming

		Experiment		
time stamp	Sample	X1	X2	X3
T=1hr	C	X1C1	X2C1	X3C1
	O	X1O1	X2O1	X3O1
	D	X1D1	X2D1	X3D1
	OD	X1OD1	X2OD1	X3OD1
T=24hrs	C	X1C24	X2C24	X3C24
	O	X1O24	X2O24	X3O24
	D	X1D24	X2D24	X3D24
	OD	X1OD24	X2OD24	X3OD24

C=	Control
O=	Oil
D=	Dispersant
OD=	Oil+Dispersant

RNA Extraction, Purification, Integrity Scoring, and Quantitation

The overall experimental design for O, D, and O/D dosing are shown in Table 1. Sponges from chosen time stamps were selected to demonstrate short-term genetic response and long-term genetic response. In that vein, one control individual (not exposed to any chemicals) was selected from each experiment at each chosen time point, for a total of 6 controls samples.

The short-term response representatives were collected after 1 hour of exposure to the chemicals and were labeled X1C1, X1O1, …, X3OD1. Long-term response will be represented by specimens collected after 24 hours of exposure and were labeled X1C24, X1O24, …, X3OD24. Samples labeling and their meaning is summarized in Table 1. A total of three experiments was performed to allow for enough replication and confidence in the data.

The sponge RNAs were extracted using a standard phenol/chloroform method as outlined in the Porifera Tree of Life (PorToL) project (Hill et al., 2013). This specific method uses TRIzol, a strong phenol reagent, that extracts high quality RNAs in high quantities. TRIzol has been the golden standard for RNA extractions because of its chemical properties and its high yield. After chloroform precipitation, this chemical allows the separation of the three biological components an extraction protocol aims for: DNA, RNA, and proteins. The extraction results in three phases: an organic phase in which the DNA is precipitated (red phase), a foamy phase (white color) in which the proteins are precipitated, and a clear aqueous phase in which the RNA is precipitated. The protocol goes as follows:

Ahead of the extraction, prepare an ice bucket as some steps are required to be executed on ice.

It is also important to have a cold centrifuge as you do not want your RNA to denature before extracting or during the extraction. If you do not have a cold centrifuge, a simple trick can be to use a normal centrifuge in a cold room at 4°C. Make sure the centrifuge is capable of spinning at 12,000xg or more.

Pro tip: when centrifuging, place the hinge of centrifuge tubes towards the outside of the centrifuge. This will indicate where the RNA pellet has precipitated.

The least you pipet your RNA the better the yield will be. RNA is very labile, and unstable, pipetting it too much or too hard will denature it.

1. Thaw out samples in TRIzol at room temperature for 5 minutes (3ml of TRIzol per gram of tissue)
2. Add 0.2 ml of chloroform per 1 mL of initial TRIzol. Cap samples securely, shake vigorously for 15 seconds, and incubate at room temperature for 2-3 minutes.

3. Centrifuge at 12,000xg for 15 minutes at 4°C. The mixture should separate into three phases, red at the bottom (with the DNA), aqueous phase on top (with the RNA) and a foamy layer in between (with the proteins). Transfer the aqueous phase into a new tube without taking any of the interphase or the bottom phase. It is better to leave some of the aqueous phase in rather than try to take it all, the resulting RNA will be cleaner.
4. During the above centrifugation step, prepare tubes with 0.5 mL of isopropanol per mL of initial TRIzol. Mix by inverting tubes a couple times and incubate at room temperature for 10 minutes.
5. Centrifuge at 12,000xg for 10 minutes at 4°C. A pellet has formed at the bottom of the centrifuge tube. Pour the supernatant out (be careful to not discard the pellet).
6. Wash the pellet by adding 0.3 ml 75% RNase free ethanol.
7. Centrifuge at 7,500xg for 5 minutes at 4°C. If the pellet does not look clean, steps 6&7 can be repeated one more time.
8. Pipette off and discard the aqueous phase. BE CAREFUL! in ethanol your pellet will not stick to the tube. Do not pipette the pellet. Air dry the pellet at room temperature for 10-15 minutes. Some protocols call for air drying the pellet on ice and under the hood. If going with the air drying on ice method, air dry the pellet for 20-25 minutes.
9. Resuspend pellet in 200-500 µL of RNase free water depending on the size of the pellet. The bigger the pellet the more water will be needed. The amount of water needed will ultimately affect the final RNA concentration of your sample.
10. RNA precipitation and clean up: add RNase free 4-7.5M LiCl to ½ the volume of your RNA. (can add the same volume as RNA, this will increase the final yield). Store at -20°C for at least an hour but the best results are obtained when precipitation occurs overnight. This step will precipitate RNA and leave remaining contaminating DNA behind, as well as unwanted phenol and salts contaminants.
11. Centrifuge on highest speed for 20-30 minutes at 4°C (can go longer if the pellet was small).
12. Take out supernatant. Add 75% RNase free ethanol to match the volume in step 10. ($V_{total} = V_{RNA} + V_{LiCl}$)
13. Centrifuge highest speed for 5 minutes at 4°C.
14. Take out supernatant. Let air dry for 10-15 minutes.

15. Resuspend in RNase free H2O (DEPC water, V≤100 μL). If the pellet doesn't dissolve easily, dissolve the pellet using a water bath heated between 50-65°C. Usually the pellet will dissolve easily with a couple pipetting mix.
16. Measure absorbance ratios using the nanodrop. If absorbance ratios do not meet passing standards, go to 18, otherwise proceed to 17.
17. Store RNAs at -80°C until downstream use.
18. If absorbance ratios in step 16 are not within the recommended range, add equal volume of 5M ammonium acetate (NH4OAc) to your RNase free water suspended RNA. Mix gently.
19. Add 2.5 volumes of 100% ice cold EtOH and mix gently.
20. Incubate 1 hour at -20°C. After 1h transfer to -80°C and incubate overnight.
21. Centrifuge at 12000xg at 4°C for 20 minutes. Discard supernatant.
22. Wash the pellet two times with 75% RNase free ice cold EtOH, and centrifuge at 12000xg at 4°C for 5 minutes.
23. Resuspend pellet in RNase free water.
24. Re-assess absorbance ratios on the NanoDrop to confirm clean up success.
25. After absorbance estimation, quantify samples using the Qubit. It is important to quantify both RNA and DNA concentrations.
26. DNase treat each extracted sample, using Turbo DNase (ThermoFisher Scientific)
27. Re-quantify nucleic acid using the Qubit to make sure DNA digestion has been effective and that RNA concentration haven't been affected.

Once extracted, RNA quality was assessed using different methods. First, RNA integrity was checked by assignment of a RIN (RNA Integrity Number) scores as shown in appendix 1. This was performed using an Agilent TapeStation 2200 and the RNA ScreenTape analysis kit from Agilent. This method assesses the integrity of RNA using the ribosomal RNA scale through the 18S and the 28S bands. RNAs with a RIN score above 6 were considered appropriate for downstream applications. RNA with RIN numbers below 6 were re-extracted. According to the literature, a RIN score of 7 is considered the passing standard. However, the Lexogen cDNA library prep kit specifically indicated that RNAs of lower quality would be possible. As well, the

whole transcriptome sequencing aspect of this study required RINs > 6. Consequently, to keep the homogeneity between the two aspect of this project, passing standards were kept consistent.

In addition to this metric, the extracted RNAs purity was checked using spectrophotometry, via a NanoDrop 2000. Spectrophotometry gives the absorbance of the samples at different wave lengths: 230 nm, 260 nm, and 280 nm. From these absorbances, ratios of A260/A280 and A260/A230 were calculated and purity of the sample can be derived from the values. Most spectrophotometer uses big volumes to determine the absorbance (>200 μL). The NanoDrop technology uses very small amounts of RNA (1-2 μL), which minimize wastes (Desjardins, P and Conklin, D, 2010). Pure RNA has a A260/A280 value of 2.1, but most protocols agree on RNA purity when this ratio is within the 1.8-2.2 range. Values under 1.8 indicates DNA or proteins contamination. On the other hand, the A260/A230 value gives another purity metric, but this time about non-nucleic contaminants in the samples. For RNA, the ideal A260/A230 ratio is 1.5 or above. Lower values indicate possible contamination of salts or phenol. Salts and phenols are commonly used reagents in RNA extraction protocols that can accidentally precipitate in the final volume of RNA (Wieczorek et al., PROMEGA). In the event the absorbance ratios standards are not met, the RNA can be resuspended and cleaned by mixing an equal volume of 5M RNase free ammonium acetate (NH4OAc). After centrifugation and ethanol wash, absorbance ratios should be measured again.

Finally, samples concentration was determined using the Qubit 2.0 from ThermoFisher Scientific. This instrument allows highly accurate concentration of RNA in the samples using minimum input. Quantification was performed using the Qubit RNA HS (High Sensitivity) assay kit and following the manufacturer's protocol.

After quantification, the samples were DNase treated using TURBO DNase from ThermoFisher Scientific and then re-quantified on the Qubit, using both dsDNA HS Assay kit and RNA HS assay kit to ensure that contaminating DNA digestion was successful.

cDNA Library Preparation

Samples passing the quality and purity standards were then prepared for sequencing using the relatively novel cDNA library prep kit from Lexogen: QuantSeq 3' mRNA-Seq Library Prep Kit FWD for Illumina. This kit contains the Illumina Read 1 linker sequence in the second strand

synthesis primer, which targets the poly A tail of mRNA, and creates the cDNA library from this end for sequencing. This allows to specifically select mRNA (eukaryotic mRNA can be selected thanks to their polyA tail) and get rid of any other type of unwanted RNA present in the cells (tRNA, siRNA, rRNA, etc.). This method called polyA selection precisely reflects the transcriptome of the organism at a given time. Another benefit of using this library preparation kit is its ability to work with low quality samples as well as low RNA quantity inputs. The kit is said to be efficient with low RIN RNAs (RNA with RIN scores as low as 4), and can create libraries with RNA input as low as 100 pg. It also provides a rapid turnover. The cDNA libraries can be prepared in a little more than 4.5 hours of time with only 2 hours of hands on manipulations (Fig. 5) The most useful property of this kit is that only one fragment is produced per transcript. This implies that no length normalization is required, thus allowing more accurate determination of gene expression. This also provides an alternative to microarrays.

As well, according to the manufacturer, the QuantSeq 3' mRNA-Seq Library Prep Kit FWD yields high strand specificity. 99.9% of the reads are correctly map to the genome, which helps in the quantification and discovery of overlapping genes as well as antisense transcripts. Finally, the kit can be cost saving thanks to its ability to multiplex. Indeed, the kit can be coupled with specific indices (i5 or i7 universal illumine indices) which allows sequencing of up to 384 samples at the same time on one Illumina flowcell lane. However, for this project a 24 reactions kit was only necessary.

The protocol works in several steps. For the cDNA library generation to be homogeneous across the board, each sample input was normalized to 132 ng total RNA, in the first step of the protocol. First, the mRNA was polyA selected by putting the extracted RNA into a reagent called FS1. During the step, the mRNA was selected by binding an oligodT primer and the first strand was then synthesized. The next step consisted of depleting the mRNA strand by adding the RNA removal reagent (RS) and denaturing it at 95°C. After denaturation, the second strand synthesis occurred by adding SS reagents and enzymes. During this step, random priming occurs on the 3'-5' strand creating a strand in the 5'-3' sense. This results in a double strand cDNA library with an cDNA insert, representative of the mRNA strand. i7 Illumina index were then added to the cDNA libraries, and a PCR amplification was executed. Here, given the 132 ng of RNA input, the optimal number of cycles for the endpoint PCR was determined to be 14. The 14 cycles were established using a qPCR approach as recommended by manufacturer's protocol.

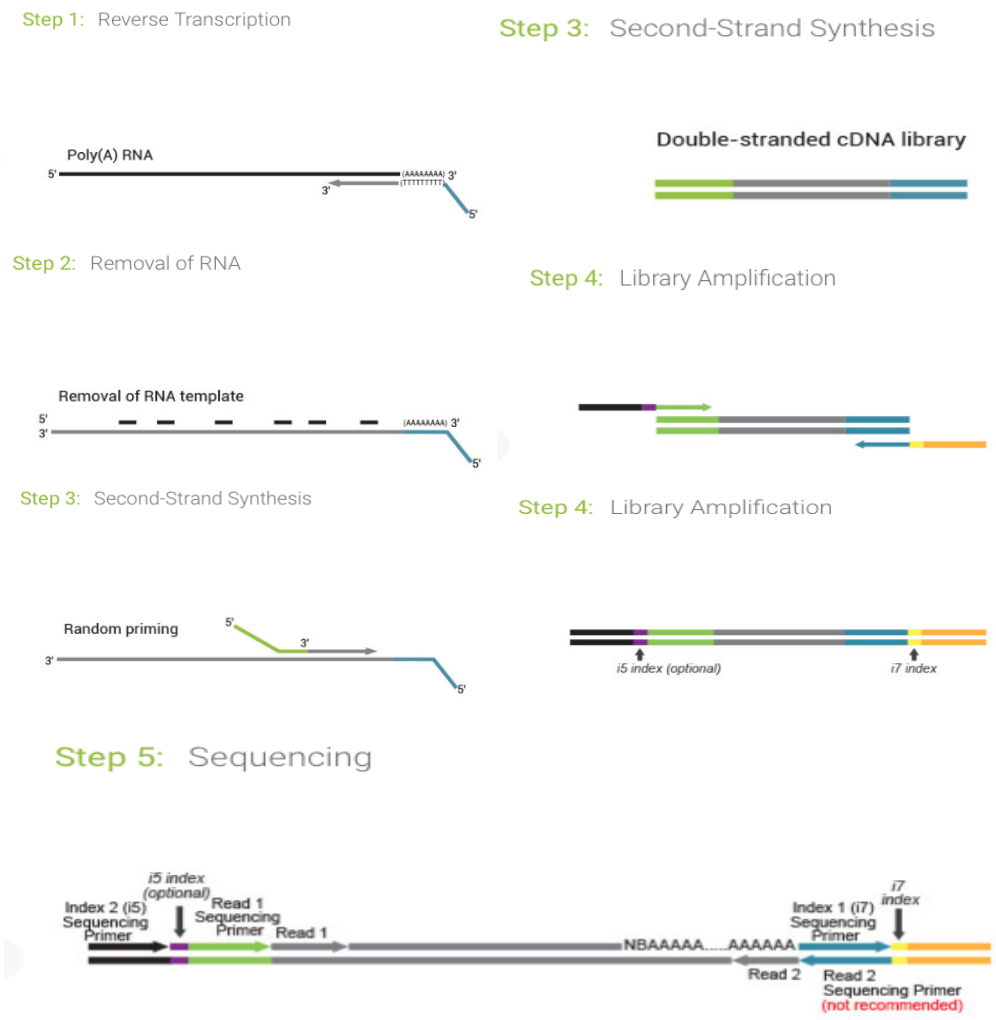


Figure 5: Lexogen's cDNA synthesis protocol. It functions in a very similar ways to other cDNA protocol but has the ability to work with lower quality RNAs. The protocol can be completed within a 4.5 hour time frame and requires only 2 hours of hands-on manipulation. Lexogen Inc "QuantSeq 3' mRNA Seq Library Prep" Lexogen. https://www.lexogen.com/quantseq-3mrna-sequencing/

Genetic Profiling Sequencing

The Lexogen cDNA libraries were sent to the NSU Genomics Core for sequencing. Final cDNA libraries generated using the Lexogen QuantSeq 3'mRNA Seq FWD Library Prep Kit were qPCR-quantified using KAPA Biosystem's Library Quantification Kit optimized for the Roche LightCycler 480 Instrument II.

A total of 24 RNA sequencing libraries were pooled and normalized to 2 nM and denatured according to Illumina's NextSeq System Denature and Dilute Libraries Guide. Final pooled libraries were spiked with 2% PhiX as an internal control and loaded at a final concentration of 1.6 pM onto the Illumina NextSeq 500 platform. Libraries were sequenced on a 1x150 bp single end run using the Illumina NextSeq 500 Mid Output v2.5 Kit (150 cycles, 130 million read flow cell). This outputted four different file for each sample corresponding to each of the 4 lanes of the flowcell, and demultiplexing was performed by the sequencing core using the Illumina's bcl2fastq 2.17 software.

Genetic Profiling Analysis (RNASeq data)

The four lanes of the NextSeq 550 created 4 cDNA sequenced data files for each sample. These four data files were first merged together in order to obtain one sequence file per sample. Merging FASTQ files together can be done using bioinformatics programs but can be done easily using a command line prompt and a small line of code. The merging code can be found in appendix 3. The merged sequence files were then uploaded to a six steps analysis pipeline, which was divided in three major steps.

The first step consisted in checking the quality of the sequences and cleaning the sequence files. To do so, the FastQC software was used (Andrews, 2005). Each sample sequence file was uploaded to Galaxy: an open, web-based, and secure platform for performing accessible, reproducible, and transparent genomic science. This platform is very user friendly and contains many tutorials for newly initiated scientists in the field of genomics. It provides a wide catalog of genomic programs used in countless genomic projects. Two servers are available, the American server can be accessed through https://usegalaxy.org and the European server can be accessed through https://usegalaxy.eu. Each sample was passed through FastQC and cleaning options were deducted from the output. Cleaning and trimming of the sequences was performed using Cutadapt (Marcel, 2011). Lexogen recommends trimming the first 6 nucleotide of each reads, and in addition polyA removal as well as adapter contamination removal was performed for each sample. Output minimum quality score was set to 25 and reads shorter than 20 bp long were discarded. Following the trimming step, the FastQC step was executed again to ensure quality of the trimmed reads. It is important to note that even though these first three steps were executed

on the Galaxy platform, they can also be performed in a Linux environment using the code provided in the appendix 3. The trimmed sequence files were then downloaded from the Galaxy platform to a local machine, to be uploaded to a cloud computing instance, as the next two steps are computationally heavy and require resources unavailable on a local computer.

After trimming and quality controls, the second step consisted in aligning and extracting raw gene counts. Trimmed read files were uploaded into Jetstream, a Linux based cloud computing server, available through XSEDE (Extreme Science and Engineering Discovery Environment). The instance was setup with 60 GB of RAM, 240 GB of storage and a total of 24 cores. Reads were then aligned to the reference Cinachyrella transcriptome using Bowtie2, an ultrafast and memory-efficient tool for aligning sequencing reads to long reference sequences (Langmead et al., 2009). Bowtie2 is a non-gaped aligner which allows to align reads to both reference genomes and transcriptome without requiring a GTF/GFF3 file. Bowtie2 is a two-step process in which the first one requires creating an index using the reference transcriptome, and the second step consists of aligning the reads to the previously created reference. The reference creating step produced six .bt2 files that allows the software to align the trimmed reads to the reference transcriptome. The alignment step produces a .SAM file that can then be used to determine genetic expression by extracting raw gene counts out of the file, and a log file in which alignment statistics are displayed and can be used for troubleshooting. Linux code to run Bowtie2 (reference and alignment step) can be found in the appendix 3. Not that Bowtie2 can also produce alignment files in .BAM format, but BAM files are unreadable to the eye as they are written in binary. SAM files can be opened and are readable, however they are much heavier in size. BAM files are favored in storage is limited.

With the SAM alignment file from Bowtie2, raw gene counts can be calculated using RSEM, a software package for estimating gene and isoform expression levels from RNA-Seq data (Li and Dewey, 2011). Just like bowtie2, RSEM is a two-step process in which a reference needs to be created using the reference transcriptome, with the possibility of creating the reference with the different isoforms of each genes. It outputs six files needed by the software to extract raw genes counts from the .SAM input file. RSEM running code can be found in the appendix 3.

The third and last step consisted in using DESeq2 (Love et al., 2014), a software program that uses raw counts table to determine differentially expressed transcripts. DESeq2 is freely

available on Galaxy but can also be executed in R as a Bioconductor package using the code present in the appendix 3. The R script provides more options than the version available on Galaxy, consequently the R version is to be favored.

Prior to running the data through DESeq2, raw counts were filtered in order to discard low expressing features and make the model stronger. This was performed by removing all transcripts that showed less than 5 counts in total over less than 25% of the sample pool. For this project this meant that a minimum of 5 transcripts in total were needed to be found across at least 6 samples. If a minimum counts of 5 was seen but only across 3 samples this transcript was discarded. DESeq2 is the golden standard program to use when assessing differential expression. It is designed to account for variability between samples and does so by taking in account the replicates of each treatment. DESeq2 works by using a negative binomial dispersion estimate for quantification, normalize counts through a variance stabilization step and can fit different type of models (local or mean) as well as different type of significant tests (Wald or Likelihood Ratio Test (LRT)). While using this program, differential expression can be assessed at both the gene and the transcript level. However, for this project, the differential expression was assessed at the transcripts level to be consistent with the TRINITY assembler output.

RESULTS

Whole Transcriptome

RNA Extraction and Quality

GENEWIZ submission guidelines required submitted samples to pass the following standards:
Sample Purity (OD260/OD280): 1.8-2.2,
Recommended RIN: ≥6.0,
Recommended Quantity: >2 µg, >50 ng/µl,
Minimum Quantity: 500 ng,
Resuspension Buffer: Nuclease-free water.

After integrity, purity and quantity assessments, all three samples passed the required standards. A summary of quality metrics can be found in Table 2

Table 2: Qualitative and quantitative metrics of extracted RNAs from environmental samples of Cinachyrella spp. Before submission to GENEWIZ

Quality control				
Sample Name	RIN	A260/A280	Conc (ng/ul) Nanodrop	Conc (ng/ul) Qubit
Cin. Env #1	8.2	1.98	90.8	120
Cin. Env. #2	7.7	2.06	165.2	387
Cin. Env. #3	7.5	1.94	70.4	206

Upon receival of the samples, GENEWIZ performed a second round of quality checks that returned concentration and RIN scores for each sample as shown in Table 3.

Table 3: Qualitative and quantitative metrics of extracted RNAs from environmental samples of Cinachyrella spp. after receival by GENEWIZ.

Sample ID	RIN	Conc. (ng/ul)
Cin-Env-1	7.8	29.2
Cin-Env-2	5.9	86.1
Cin-Env-3	6.1	27.2

The difference in concentration and RIN scores could be explained by the traveling of the samples and the possibility of samples degrading by getting thawed over the trip.

Whole Transcriptome Sequencing

Sequencing of the whole transcriptome was performed on one lane of Illumina HiSeq 4000 resulted in over 140 million paired-end reads for each submitted samples, and the yield for each sample was 44,000 million base pairs. All sequencing statistics can be found in Table 4 below.

Table 4: Whole transcriptome sequencing results of the three submitted environmental samples sent to GENEWIZ

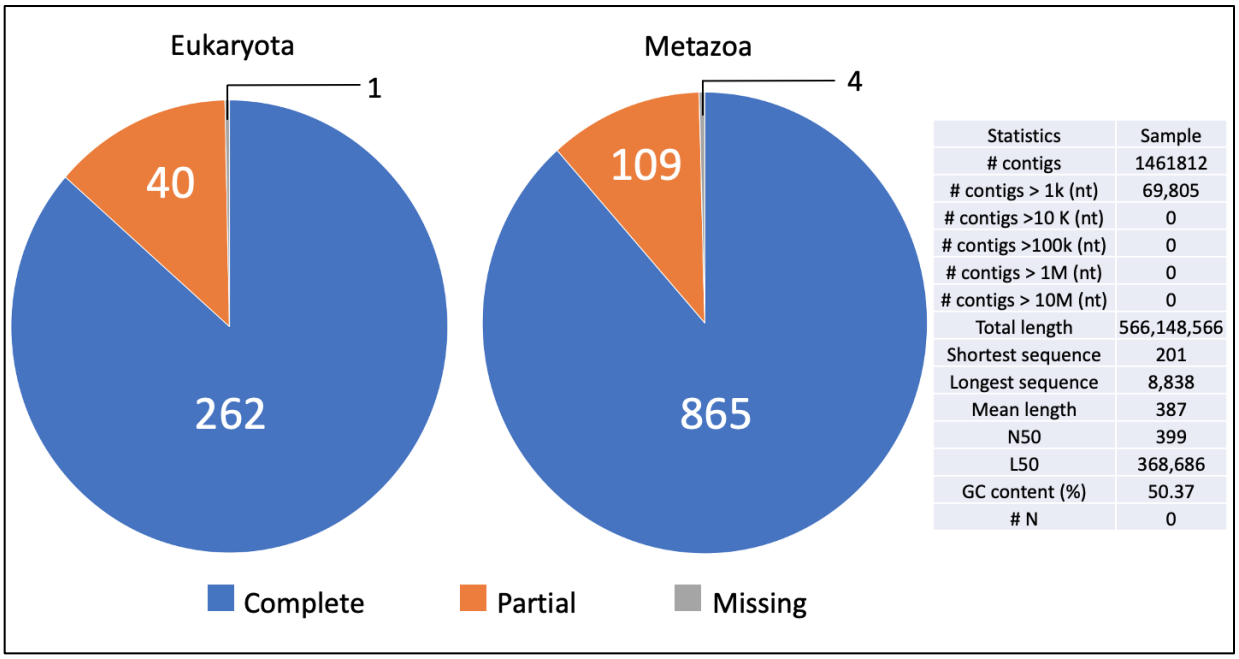
Sample ID	Barcode Sequence	# Reads	Yield (Mbases)	Mean Quality score	% Bases >=30
Cin-Env-1 Cinachyrella spp.	TCCGCGAA + GAGAGGTT	149,428,060	44,828	35.93	93.94
Cin-Env-2 Cinachyrella spp.	TCCGCGAA + ACCTGGTT	149,384,419	44,815	36.02	94.40
Cin-Env-3 Cinachyrella spp.	TCCGCGAA + AAGCGGAA	147,874,130	44,362	35.84	93.53

De-novo Cinachyrella Transcriptome Assembly and Annotation

After quality assessment and quality check of sequencing data, sample Cin-Env-1 was selected for transcriptome assembly and annotation. Because the assembly appeared to include all transcripts from the host and its symbiotic communities, the assembly can be referred as the holo-transcriptome of the sponge. The assembly consisted of 1,461,812 contigs with 69,805 contigs 1000 base pairs or longer. The largest contig was estimated to be 8,838 nucleotides long. Average contig length was found to be 387 base pair long and the N50 metric was 399 (Fig. 6). The transcriptome revealed to have a GC content of 50.361%, which is common in eukaryotes. The GETORF tool identified a total of 6,487,125 ORFs within the de-novo assembled transcriptome.

The annotation of the transcriptome using Diamond BLASTx resulted in a total of 588,048 genes with a hit in the NR NCBI database, out of which 296,810 genes where mono-coding when duplicates and isoforms were eliminated. Within these 296,810 identified genes, 82,831 genes had a gene description, leaving 213,979 genes as uncharacterized proteins, hypothetical proteins, or proteins of unknown function.

Completeness analysis using BUSCO v3 (Waterhouse et al., 2017) against the Eukaryota and Metazoa databases (Fig. 6) revealed a completeness of 99.67% (C:86.5%[S:25.1%,D:61.4%],F:13.2%,M:0.3%,n:303) and 99.59% respectively (C:88.4%[S:33.4%,D:55.0%],F:11.1%,M:0.5%,n:978). It is worth noting that within this assembly and in both databases, most core complete genes are duplicated, which indicates the sequencing of the holo-transcriptome.



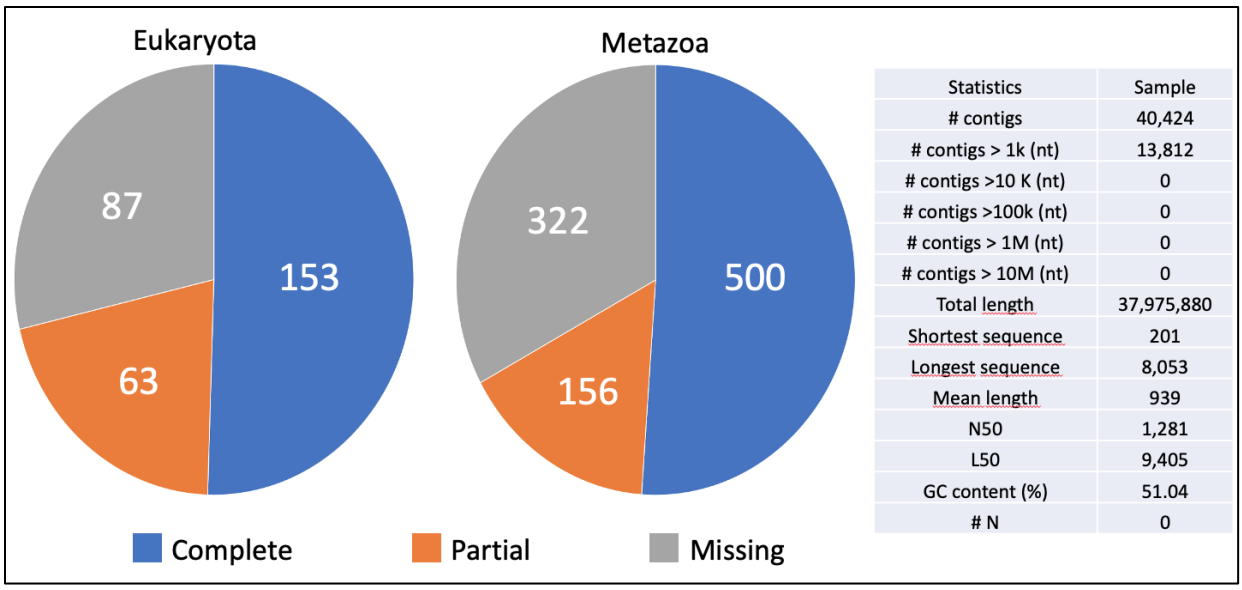
Statistics	Sample
# contigs	1461812
# contigs > 1k (nt)	69,805
# contigs >10 K (nt)	0
# contigs >100k (nt)	0
# contigs > 1M (nt)	0
# contigs > 10M (nt)	0
Total length	566,148,566
Shortest sequence	201
Longest sequence	8,838
Mean length	387
N50	399
L50	368,686
GC content (%)	50.37
# N	0

Figure 6: Pie charts diagrams representing the completeness of the TRINITY assembled holo-transcriptome as measured against two databases. A summary of the holo-transcriptome metrics is compiled in the table on the far right. Completeness of an assembly is determined based on a number of core genes commonly found across a wide array of representative organism within that taxa.

Because this is a poriferan transcriptome, it was interesting to see the number of transcripts related to a known sponge species. A total of 40,424 transcripts were identified as sponge related. The large majority of transcripts (36,459) was identified as similar to the sponge A. queenslandica, with 8,660 genes flagged as uncharacterized proteins. These 40,424 transcripts coded for 8,496 unique proteins. Thus a large difference exists between the number of hits matching a sponge and the total number of genes with a hit in the annotation. Perhaps this can be explained by the fact that sponges are highly colonized with symbionts, with some species being

made of up to 35% of symbionts or bacterial biomass (Santavy et al, 1990; Webster and Thomas, 2016; Lopez 2019). This would explain why so many genes returned a hit different from a porifera representative.

When only the 40,424 sponge matches were considered as the assembly, the transcriptome ended up being 37,975,880 nucleotide long, with the shortest and longest sequences being 201 and 8053 bases respectively. The average sequence length was found to be 939 bp. N50 sequence length metric revealed a value of 1281 bp, while the L50 sequence count was determined to be 9405. Finally, this assembly was comprised of 13,812 sequences longer than 1000 bp, which represented 34.2% of the assembly. No sequences with lengths over 10K, 100K, 1M, or 10M bp were reported. Overall, the sponge hits based assembly resulted in a GC content of 51.04%, which is within the range of GC bases in a eukaryotic transcriptome (Fig. 7). The analysis of the assembly with only the contigs bouncing off as Porifera contigs, showed a completeness of 71.29 % against the Eukaryota database (C:50.5%[S:20.8%,D:29.7%],F:20.8%,M:28.7%,n:303; Fig. 7). Against the Metazoa database, completeness analysis suggested a completeness of 67.08% (C:51.1%[S:22.1%,D:29.0%],F:16.0%,M:32.9%,n:978; Fig. 7).



Statistics	Sample
# contigs	40,424
# contigs > 1k (nt)	13,812
# contigs >10 K (nt)	0
# contigs >100k (nt)	0
# contigs > 1M (nt)	0
# contigs > 10M (nt)	0
Total length	37,975,880
Shortest sequence	201
Longest sequence	8,053
Mean length	939
N50	1,281
L50	9,405
GC content (%)	51.04
# N	0

Figure 7: Pie charts representing the completeness of the sponge related transcripts assembly as measured against two databases. A summary of the assembly metrics is compiled in the table on the far right. Completeness of an assembly is determined based on a number of core genes commonly found across a wide array of representative organism within that taxa.

Differential Gene Expression (DGE) Analysis

RNA Integrity, Purity, and Quantitation

Extracted RNA integrity using the Agilent TapeStation 2200 revealed that all samples had a RIN value superior to 6. Sample X2C24 had the lowest RIN of the sample pool with a value of 6.5. Consequently, no samples needed to be re-extracted as all of them were above passing standards. A maximum RIN value of 8.9 was observed for samples X2OD1 and X2OD24. Overall, the RIN distribution was homogeneous across the pool of sample, confirming the efficacity of the PorToL extraction protocol.

Absorbance ratio measurements revealed that all 24 samples where within the literature recommended range for the A260/A280 with the lowest ratio being 1.84 for the X1C24 and X3C24 samples and the highest being 2.33 for the X2OD1 sample. A260/A230 ratios revealed most samples were equal or above passing standard with X1C24 being the purest sample with a ratio of 2.49, confirming again the efficacity of the PorToL extraction protocol. X1C1, X1O1, X1O24, X2O1, X3O1, X3OD1, and X3OD24 samples all had ratio values below the recommended 1.5 standard. However, given the very close numbers, these samples were still considered good for downstream analysis.

Finally, after DNase treatment, each sample concentration was calculated, and a maximum of 88,7 ng/µl was found for the X2C24 sample and a minimum concentration of 19.5 ng/µl for the X3O1 sample. Lexogen's cDNA preparation kit allows RNA input as low as 100 pg and up to 2 µg, consequently putting all samples within the manufacturer's recommended range. The quality and quantitation results of the extracted RNA are summarized in Table 5.

Table 5: Qualitative and quantitative metrics of dosed samples selected for differential gene expression assessment

	Species	RNA Concentration (ng/ul)	RIN	A260/A280	A260/A230	i7 index sequence
X1C1	*Cinachyrella spp.*	48.4	6.8	2.03	1.41	CAGCGT
X1O1	*Cinachyrella spp.*	30.7	6.8	1.96	1.45	GATCAC
X1D1	*C. alloclada*	41.6	7.4	1.98	1.91	ACCAGT
X1OD1	*C. alloclada*	30.7	8.8	2.01	2.08	TGCACG
X1O24	*C. alloclada*	37.5	7.2	2.10	1.42	ACATTA
X1D24	*Cinachyrella spp.*	48.5	8.0	2.08	2.32	GTGTAG
X1OD24	*Cinachyrella spp.*	46.3	6.9	2.11	1.56	CTAGTC
X1C24	*Cinachyrella spp.*	83	8.4	1.84	2.49	TGTGCA
X2C1	*C. alloclada*	26.6	8.3	1.90	1.55	TCAGGA
X2O1	*Cinachyrella spp.*	29.5	7.8	1.94	1.43	CGGTTA
X2D1	*C. alloclada*	52.2	7.2	2.03	2.14	TTAACT
X2OD1	*Cinachyrella spp.*	26.8	8.9	2.33	2.39	ATGAAC
X2O24	*Cinachyrella spp.*	70	7.8	2.07	2.34	CCTAAG
X2D24	*Cinachyrella spp.*	44	7.8	2.07	1.50	AATCCG
X2OD24	*Cinachyrella spp.*	29.2	8.9	1.98	2.00	GGCTGC
X2C24	*Cinachyrella spp.*	88.3	6.5	1.9	2.41	TACCTT
X3C1	*C. alloclada*	53	7.5	2.03	1.65	TCTTAA
X3O1	*C. alloclada*	19.5	6.6	2.09	1.45	GTCAGG
X3D1	*C. alloclada*	37	7.3	1.94	1.96	ATACTG
X3OD1	*C. alloclada*	57	7.3	2.09	1.44	TATGTC
X3O24	*Cinachyrella spp.*	42.6	6.9	2.03	1.59	GAGTCC
X3D24	*Cinachyrella spp.*	45.4	7.5	1.92	1.67	GGAGGT
X3OD24	*Cinachyrella spp.*	29	7.0	1.92	1.47	CACACT
X3C24	*Cinachyrella spp.*	65.7	7.9	1.84	2.30	CCGCAA

It is interesting to note that based on the study from Schuster et al (2017) who used the same samples as those for this study, only 9 of our 24 samples were identified as Cinachyrella alloclada. All other samples either had no intron found or had an intron but of too bad quality to assess taxonomy. Consequently, all further downstream conclusions will apply to the Cinachyrella genus and not to the species exactly. It is thus important to keep in mind that potential variation might be due to differences in sponge species, which were difficult to discriminate in the field.

Electronic and light microscopy analysis

Some tissues from each sponge were saved for light and scanning electronic microscopy (SEM). SEM images of sponges fresh from field and after 10 days of culturing in tanks showed no visual significant difference, giving confidence in the ability of sponges to survive in tanks for some period of time without affecting their health. The tissues kept their integrity without releasing spicules or inner tissues.

SEM imagery of control, oil, dispersant, and oil:dispersant at the 24 hours time mark showed a detrimental effect to the sponge tissue at the microscopic scale. Indeed, while control samples looked similar than fresh and 10 days tank culturing samples (Fig. 8), the tissues exposed to the chemical mixtures reacted differently.

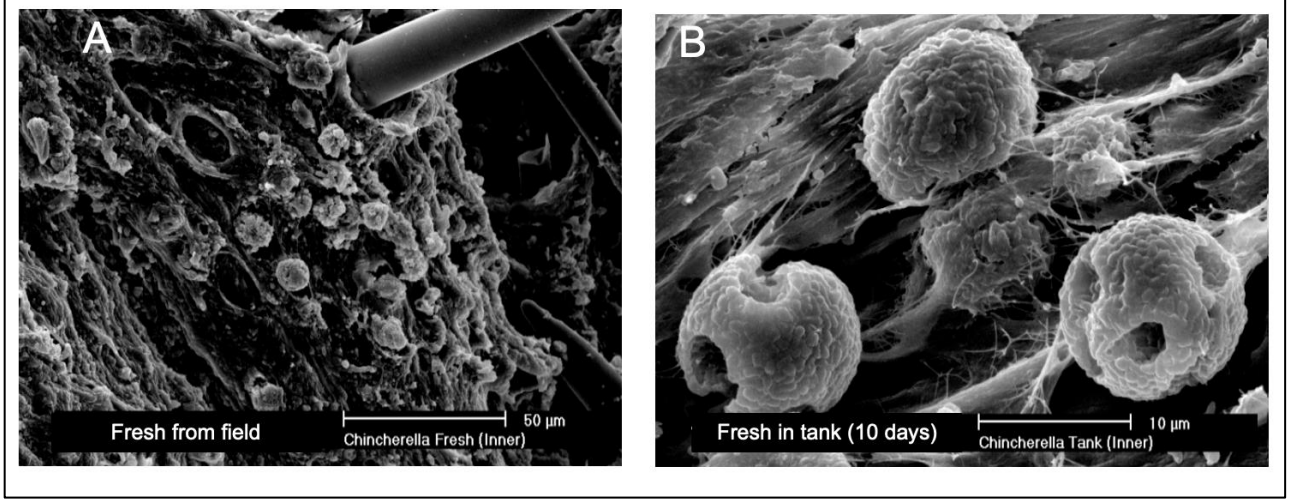


Figure 8: Micrographs of environmental (A) and 10 days tank cultured (B) sponges as seen under scanning electronic microscopy. Both tissues look healthy and compact, indicating that sponges were doing well under controlled culturing conditions with no treatments. SEM micrographs courtesy of Dr. Patricia Blackwelder

After 1 hour of exposure, across all treatments, all sponges seemed to be healthy. Tissues were compact and bulked together, didn't look like the degrading process had started, and even though some spicules could be seen, which can be explained by the sectioning of the tissue to see it under microscopy, thus exposing some areas that would not normally be exposed to the naked eye if a whole animals was used, overall, 1 hour treated samples looked really close and similar to control and uncultured samples. At the phenotypic scale it looked like the chemical exposure didn't really have any impact (Fig. 9).

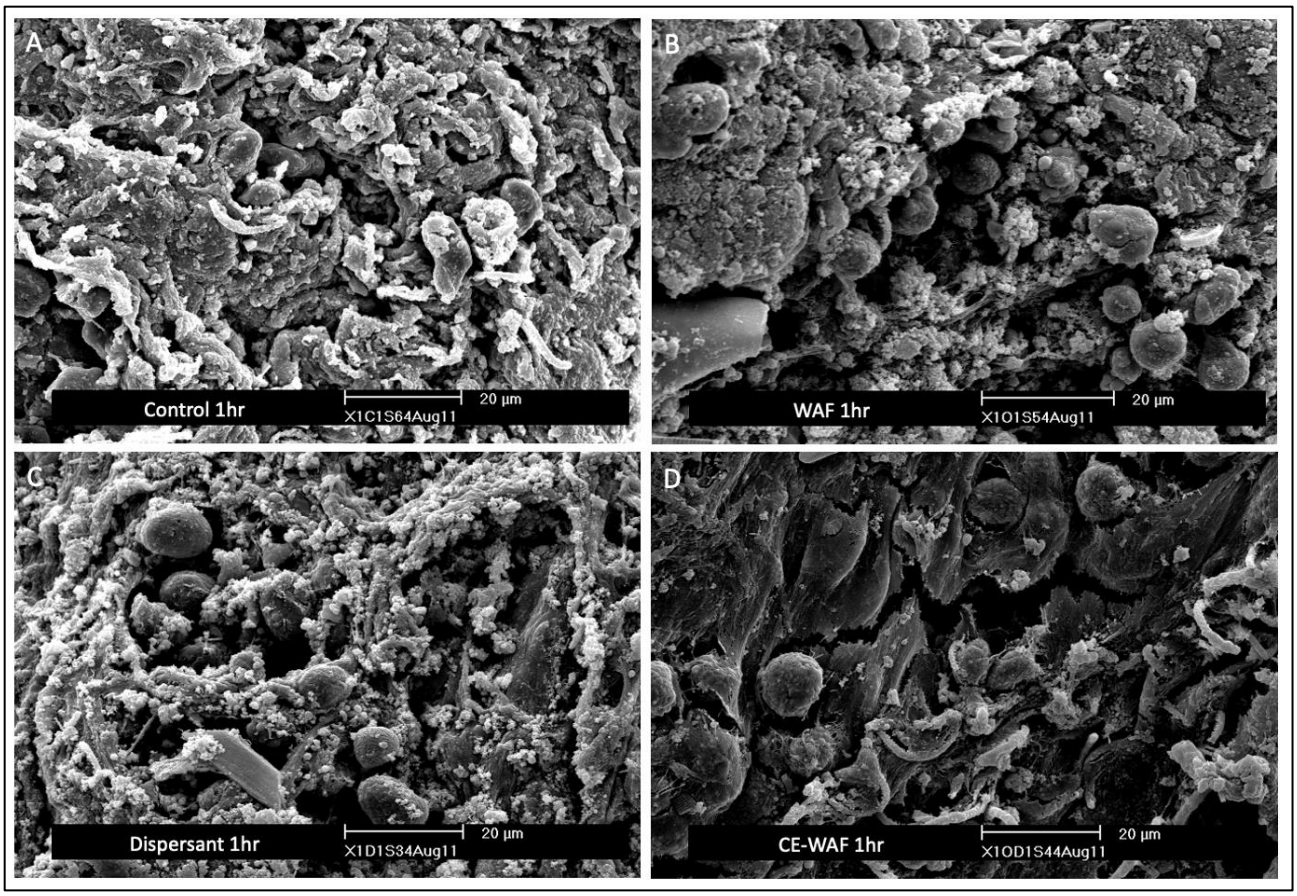


Figure 9: micrographs of 1 hour control (A), oil treated (WAF; B), dispersant (C), oil:dispersant (CE-WAF; D) sponges as seen under scanning electronic microscopy. Treated tissues looked similar to control tissues: healthy and compact, indicating that sponges were doing well phenotypically after 1 hour of exposure to the chemicals. (Electron micrographs courtesy of Dr. Patricia Blackwelder).

After 24 hours of exposure, in all three treatments tissues are degrading and starting to retract, spicules are being expulsed, and the exoskeleton of the sponge started to fall apart. These images were the first witness of an impact of the chemicals at the phenotypic scale as a result of a change at the genotypic scale (Fig. 10).

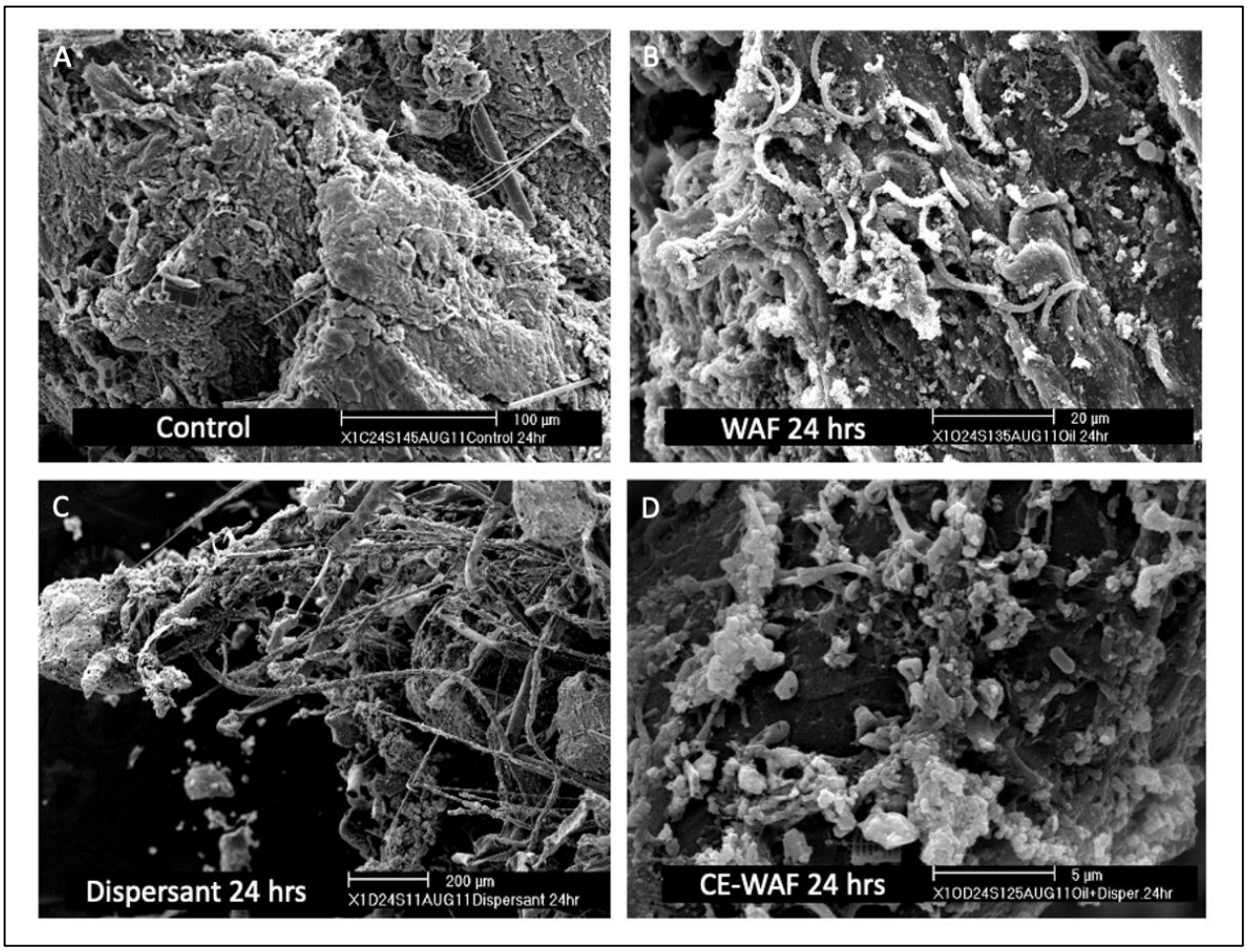


Figure 10: Micrographs of 24 hours control (A), oil treated (WAF; B), dispersant (C), oil:dispersant (CE-WAF; D) sponges as seen under scanning electronic microscopy. In comparison to the control, treated tissues looked unhealthy and less compact. Spicules started to be apparent and cohesiveness of tissues looked to be compromised. This was the phenotypic proof of a genotypic change. (Electron micrographs courtesy of Dr. Patricia Blackwelder).

Furthermore, light microscopy of sponge tissues revealed the presence of oil droplets obstructing the pumping canals of the sponge in oil and oil:dispersant samples (Fig. 11). The presence of such compounds in the tissues can impact the sponge in several ways. First by obstructing the pumping canals, this reduces the volume of water the sponge can filter per day, and ultimately in the long run if these canals stay clogged, this will have an impact on feeding mechanisms and thus the survival of the sponge. Secondly, by getting stuck in the tissues, the oil impact on the genetic expression will be increased. Indeed, if the oil just passes through the sponge tissue it only affects the organism temporarily. On the other hand, if the chemical is stuck

inside the tissues for a prolonged period, this presence can impact specific gene expression continuously.

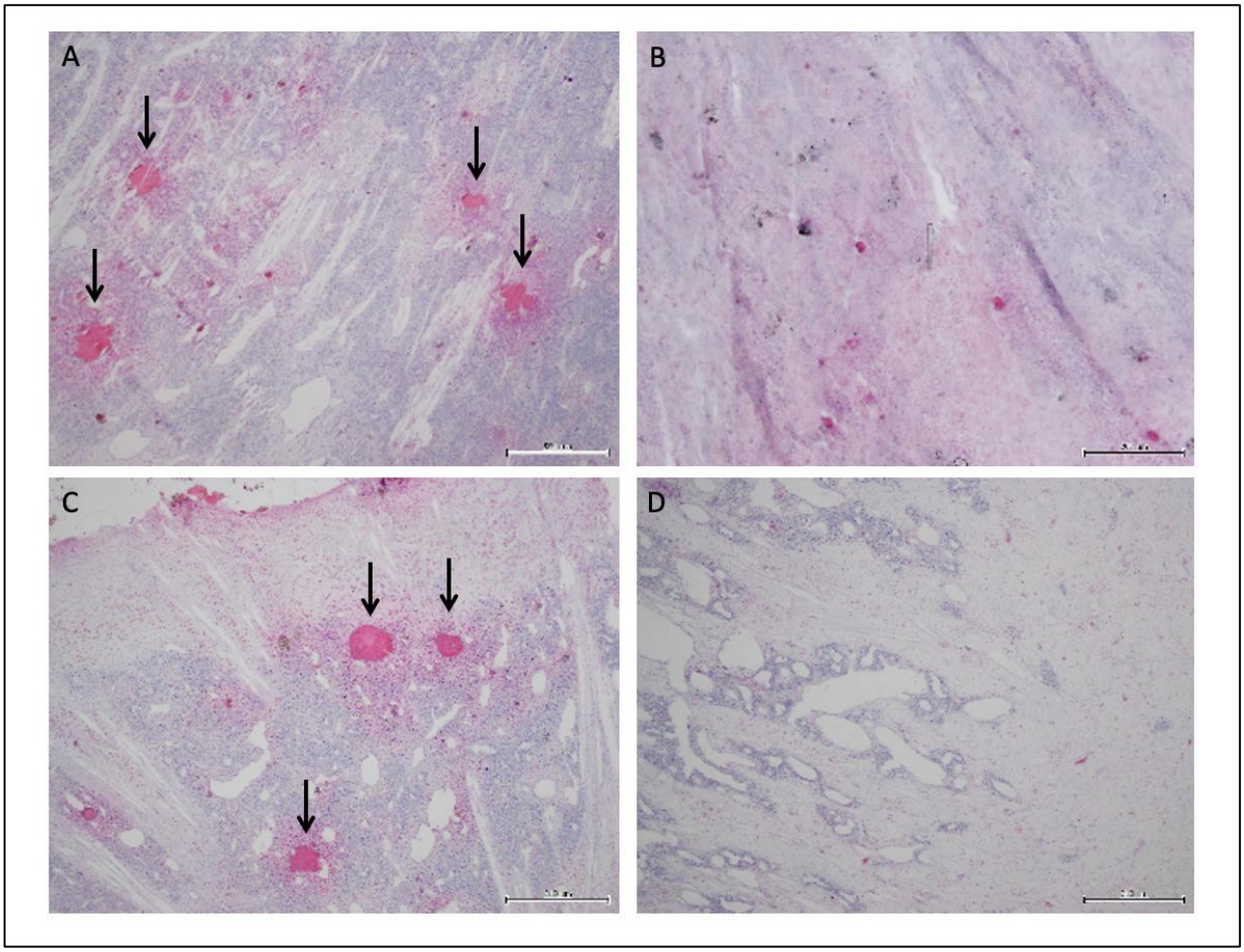


Figure 11: Micrographs of oil droplets in sponge tissues (top and bottom left, black arrows point to oil droplets trapped in the sponge tissues, scale bar = 500 μm) and broken down aquiferous systems (top and bottom right, scale bar = 200 μm). (A) and (B) represent 24 hours oil (WAF) treated samples (C) and (D) represent 24 hours oil:dispersant (CE-WAF) treated samples. In both treatments, oil droplets were clearly visible within the sponge tissue, potentially indicating the non-sufficient dispersant properties of the chemical dispersant used to disperse the oil. Both treatments resulted in broken down aquiferous systems of the sponge resulting in a reduce pumping ability by the organism. Light micrographs courtesy of Nidhi Vijayan.

cDNA Library Preparation

Twenty four cDNA libraries were generated using the Lexogen QuantSeq 3' mRNA FWD kit. For all libraries, RNA input was standardized to 132 ng, which resulted in 14 cycles of PCR amplification. All resulting libraries averaged 290 bp in size with and average insert size of

170 bp. Each library was indexed with a different universal Illumina I7 index (Table 5). Indexing, also known as barcoding, allows for multiplexing on the sequencer (simultaneous sequencing of all samples at once). An index is a unique sequence (here 6 base pair long) that allows us to identify which sample the read belongs to and map it back to the corresponding file when demultiplexing the raw sequencing data.

Genetic profiling sequencing and analysis

Sequencing of the 24 single-end libraries from dosed individuals on the four lanes of a NextSeq 550 resulted in an average of 7.8 million reads per sample, for a total coverage of 55X per sample. Table 6 summarizes the number of reads for each sample.

Table 6: Number of reads obtain by sequencing of each sample of interest with their associated index. Each index is a unique 6 base pairs identifier allowing for multiplexing during sequencing

Sample ID	Index #	Index Sequence	# Reads
X1C1	7001	CAGCGT	7,236,956
X1O1	7002	GATCAC	7,661,541
X1D1	7003	ACCAGT	7,448,452
X1OD1	7004	TGCACG	7,628,907
X1O24	7005	ACATTA	7,086,773
X1D24	7006	GTGTAG	7,517,078
X1OD24	7007	CTAGTC	7,502,756
X1C24	7008	TGTGCA	6,893,190
X2C1	7009	TCAGGA	8,480,306
X2O1	7010	CGGTTA	7,435,339
X2D1	7011	TTAACT	8,105,621
X2OD1	7012	ATGAAC	8,094,343
X2O24	7013	CCTAAG	8,984,283
X2D24	7014	AATCCG	9,106,184
X2OD24	7015	GGCTGC	8,073,798
X2C24	7016	TACCTT	7,520,429
X3C1	7017	TCTTAA	8,400,491
X3O1	7018	GTCAGG	8,301,308
X3D1	7019	ATACTG	7,519,615
X3OD1	7020	TATGTC	7,790,465
X3O24	7021	GAGTCC	6,762,163
X3D24	7022	GGAGGT	7,959,505
X3OD24	7023	CACACT	8,305,751
X3C24	7024	CCGCAA	9,110,585

Average	7,871,909.96

Initially, sequenced samples were to be analyzed using a pipeline set up and provided by Lexogen on Bluebee.com, a cloud based genomics platform with a user friendly web interface. In concept, the pipeline consisted in similar steps as the pipeline described in the material and methods to the exception of programs used to process the data. Read quality was assessed using FastQC and trimming was performed using BBduk from the BBmap suite. Reads where then aligned to a reference genome using STAR, a splice aware aligner, with ENCODE settings, and from which an alignment BAM file is outputted. This alignment step was performed using the Amphimedon queenslandica genome (Fernandez-Valverde et al., 2015) the best poriferan genome available up to date. That BAM file was put through RSEQC, to report the distribution of the reads on the annotation feature. Finally, HTSeq-counts used the BAM file to extract gene counts based on the type of library preparation kit used (FWD or REV, here in this case FWD).

However, although this pipeline is most commonly used in the literature, and recommended by Lexogen, it was intended to work mostly on model organisms and with well annotated and deeply sequenced genomes, which is not the case for A. queenslandica and sponges in general. This resulted in poor alignment rates, with very few reads aligned only one time, most of the reads too short to be aligned, and a huge fraction of "tags" were located up to 1kb upstream of the annotated TSS (Transcript Starting Site). This means that the splice aware aligner software considered the reads to be fragments located before the beginning of the ORF. For example, the sample X1C1 had an overall alignment of 33.2%. Out of this 17.72% was uniquely mapped reads and 15.48% of multi-mapped reads. But 66.8% of the reads were too short to be mapped, leaving the big majority of the reads behind and resulting in inaccurate DGE assessment. This trend was seen throughout the sample pool when using this pipeline, which resulted in this pipeline being discarded and the use of another pipeline was favored.
This pipeline, described in the materials and methods section above, was more adapted for non-model organisms made of programs able to deal with limited resources.

Sequence Alignment

Alignment with Bowtie2 resulted in a much higher alignment rate that previously found with STAR. Alignment rates are summarized in Table 7 below.

Table 7: Summary of Bowtie2 alignment rate for each sample of interest

Sample	# reads	No alignment	single mapped reads	multi-mapped reads	Overall alignment (%)
X1C1	7,118,759	2,781,613	366,730	3,970,416	60,92
X1O1	7,545,460	2,475,202	721,861	4,348,397	67,19
X1D1	7,367,066	2,876,666	366,209	4,124,191	60,95
X1OD1	7,570,763	4,906,358	298,558	2,365,847	35,19
X1C24	6,840,068	2,480,884	799,517	3,559,667	63,73
X1O24	6,986,159	2,082,889	641,922	4,261,348	70,18
X1D24	7,446,891	2,957,338	1,037,689	3,451,864	60,28
X1OD24	7,439,405	2,229,092	652,516	4,557,797	70,03
X2C1	8,333,676	2,877,741	399,649	5,056,286	65,46
X2O1	7,372,975	2,463,519	791,490	4,117,966	66,58
X2D1	8,002,365	3,252,275	450,348	4,299,742	59,35
X2OD1	7,976,089	3,087,711	1,174,466	3,713,912	61,28
X2C24	7,463,060	3,129,010	783,005	3,551,045	58,07
X2O24	8,798,536	3,888,007	1,461,176	3,449,353	55,81
X2D24	8,786,660	4,215,119	1,392,847	3,178,694	52,02
X2OD24	8,019,662	2,455,582	875,482	4,688,598	69,38
X3C1	8,228,097	5,103,265	445,403	2,679,429	37,97
X3O1	8,217,219	4,811,704	495,271	2,910,244	41,44
X3D1	7,455,339	3,632,441	390,323	3,432,575	51,27
X3OD1	7,690,291	4,528,448	421,718	2,740,125	41,11
X3C24	9,038,879	2,944,951	926,963	5,166,965	67,41
X3O24	6,706,057	2,254,746	631,248	3,820,063	66,37
X3D24	7,895,435	1,771,544	1,706,089	4,417,802	77,56
X3OD24	8,116,488	1,748,496	1,319,419	5,048,573	78,45

All samples to the exception of X1OD1, X3C1, X3O1, and X3OD1 had alignment rates greater than 50%. The maximum alignment rate was found to be X3D24 with an overall alignment rate of 77.32%.

In all samples, there were a lot of reads mapped multiple times. The high abundance of multi-mapped reads is normal when dealing with transcriptome-based alignments. A transcriptome is composed of genes that have different isoforms (different versions of a genes that codes for the same protein) and reads might align to different isoforms. These different isoforms need to be accounted for as belonging to the same gene for further downstream analysis but programs like RSEM take in account this parameter.

Read Counts

This last step is achieved by running the alignment files in .SAM format through RSEM, a transcript quantification software. Just like Bowtie2, RSEM was a two-step process in which a reference needed to be created using the reference transcriptome and with the possibility of creating the reference with the different isoforms of each genes, and outputs 6 files needed to get obtain gene counts. The quantification step outputs four files: 1 statistics files of the counting, one .BAM file for the alignment, and 2 .results file. These two .results files were the files needed to extract raw counts. One of these files was a gene.results which considered each transcript as a gene and thus did not account for isoforms and was less suitable for differential expression analysis. The other .results file was a isoforms.results file. It was from this file that raw expression was extracted, as it took in account the different isoforms of each genes and allowed us to get more accurate expression counts for each transcript and its different versions.

Linux code to run these programs can be found in appendix 3. After raw counts have been extracted, all transcripts were identified as TRINITY_DNXXX_cX_gX_iX in which ‘TRINITY_DNXXX’ was the transcript identifier, cX the cluster number, gX the gene number, and iX the isoform number. This identifier allowed us to map back these transcripts to the annotation file provided by GENEWIZ, in which the gene description and accession number were associated to each transcript.

Differential Expression Assessment

With these annotated transcripts, merging of duplicates based on the accession number was performed. Further filtering using the threshold mentioned in the Materials and Methods section, resulted in 31,751 transcripts to be considered for differential expression in DESeq2.

Differential expression analysis was performed using DESeq2 a Bioconductor package available in R, with extensive support available from the community and the developer. The R script to run DESeq2 is available in the appendix 3. DESeq2 requires the input of a metadata file in order to keep track of the samples and the replicates. In the case of this study a simple metadata file was needed with only the type of sample, the treatment, and the replicate number, as the experimental design is not a true time series. The used metadata file can be found in the appendix 3.

In order to run DESeq2, a few settings needed to be adjusted from the tutorial available online (http://bioconductor.org/packages/release/bioc/vignettes/DESeq2/inst/doc/DESeq2.html). In the DESeq2 step, the "local" fitType was preferred as it was fitting the model better with the gene dispersion estimates. Furthermore, the test method was set to "Wald" over the LRT method. A total of 6 comparisons were generated: Control 1 hour samples against treated samples of oil 1 hour, dispersant 1 hour, and oil:dispersant 1 hour. The same comparisons were executed for the 24 hours time stamp with comparing 24 hours Control samples against treated samples of oil 24 hours, dispersant 24 hours, and oil:dispersant 24 hours.

The threshold for significant differential expression was set to a difference in log2 fold change of 2 and an adjusted p-value (better known as q-value) of 0.05. Out of the 31,751 genes considered for differential expression Overall, 12,913 transcripts have shown significant differential expression, among which 7,863 were upregulated, and 5,058 were down regulated. Table 8 (A) below summarizes the number of differentially expressed transcripts across all treatments using the full assembly, and Table 8 (B) shows the top 5 up-regulated and down-regulated transcripts for each treatment at each time point.

Table 8: (A) Differentially expressed transcripts across all treatments. (B) Top 5 up-regulated and down regulated genes for each treatment at each time point.

(A)	Pairwise Comparison	# Differentially expressed transcripts	up-regulated	down-regulated
	C1_O1	8,052	7,561	491
	C1_D1	6	6	0
	C1_OD1	31	31	0
	C24_O24	268	48	220
	C24_D24	308	58	250
	C24_OD24	4,248	159	4,089
	O1_O24	203	92	111
	D1_D24	15,041	13,326	1,715
	OD1_OD24	6,116	5,240	876
	C1_C24	9,682	8,711	971

(B)	Treatment		
Top 5 regulated transcripts	Oil (WAF)	Dispersant	Oil:Dispersant (CE-WAF)
1 Hour of Exposure			
Upregulated	TRINITY_DN392278_c5_g4	TRINITY_DN395491_c4_g4	YP_009326829 : ribosomal protein
	XP_015458757 : FAD dependent oxidoreductase	XP_015458757 : FAD dependent oxidoreductase	TRINITY_DN372845_c5_g5
	OGG55450 : hypothetical protein	TRINITY_DN392278_c5_g4	TRINITY_DN392278_c5_g4
	TRINITY_DN395491_c4_g4	TRINITY_DN389181_c0_g2	TRINITY_DN368615_c12_g1
	TRINITY_DN372845_c5_g5	TRINITY_DN382666_c0_g5	XP_022100877 : uncharacterized protein
Down-regulated	TRINITY_DN386287_c3_g1	N/A	N/A
	TRINITY_DN398670_c0_g2	N/A	N/A
	TRINITY_DN381735_c0_g1	N/A	N/A
	TRINITY_DN308228_c0_g1	N/A	N/A
	TRINITY_DN381372_c5_g4	N/A	N/A
24 Hours of Exposure			
Upregulated	TRINITY_DN396576_c0_g3	TRINITY_DN384023_c12_g1	TRINITY_DN396576_c0_g3
	TRINITY_DN399762_c1_g4	XP_019020703 : hypothetical protein	TRINITY_DN384023_c12_g1
	TRINITY_DN240164_c0_g1	TRINITY_DN396576_c0_g3	TRINITY_DN75966_c0_g1
	TRINITY_DN382002_c0_g1	SBT58227: hypothetical protein	TRINITY_DN382002_c0_g1
	CDS17303 : ribosomal protein	TRINITY_DN75966_c0_g1	TRINITY_DN240164_c0_g1
Down-regulated	AMX21890 : photosystem II, protein D1	TRINITY_DN371367_c0_g1	OGG55450: hypothetical protein
	TRINITY_DN546245_c0_g1	TRINITY_DN392278_c5_g4	WP_062105845 : sulfatase hydrolase/transferase
	TRINITY_DN392184_c1_g3	TRINITY_DN361181_c0_g1	TRINITY_DN379795_c8_g1
	ABX13097 : hypothetical protein	TRINITY_DN380481_c2_g2	TRINITY_DN385861_c4_g2
	WP_041003236 : SDR family oxidoreductase	TRINITY_DN374043_c8_g2	OFW22852 : hypothetical protein

Dosage Sample Grouping

Before Pairwise comparisons, sample grouping and sample distances were assessed (Fig. 12). From the PCA plot (Fig. 12, A) we were able to determine that the samples divided into three main clusters. The first cluster on the left grouped all replicates from the 1 hour time point to the exception of the Oil 1 hour samples. All seemed relatively close to each other suggesting that the difference in expression between these samples should be minimal. Two other clusters are visible on the right side of the graph. A smaller cluster on the top of the graph including the three oil:dispersant 24 hours replicates with replicate 1 of oil 24 hours and replicate 3 of

dispersant 24 hours. The last cluster seemed to be made of replicate 1 and 2 of dispersant 24 hours, all three replicates of control 24 hours, replicates 2 and 3 of oil 24 hours, and all three replicates of oil 1 hours.

When plotting the sample to sample distance plot, clustering of samples was a bit different. Both the PCA and the distance plot agreed on 3 overall clusters. However, the second and third cluster mentioned above were divided into two sub clusters. Replicates 1 and 2 of oil 1 hours were separated from cluster 2 and replicates 1 and 2 of dispersant 24 hours were separated from cluster 3 (Fig. 12, B)

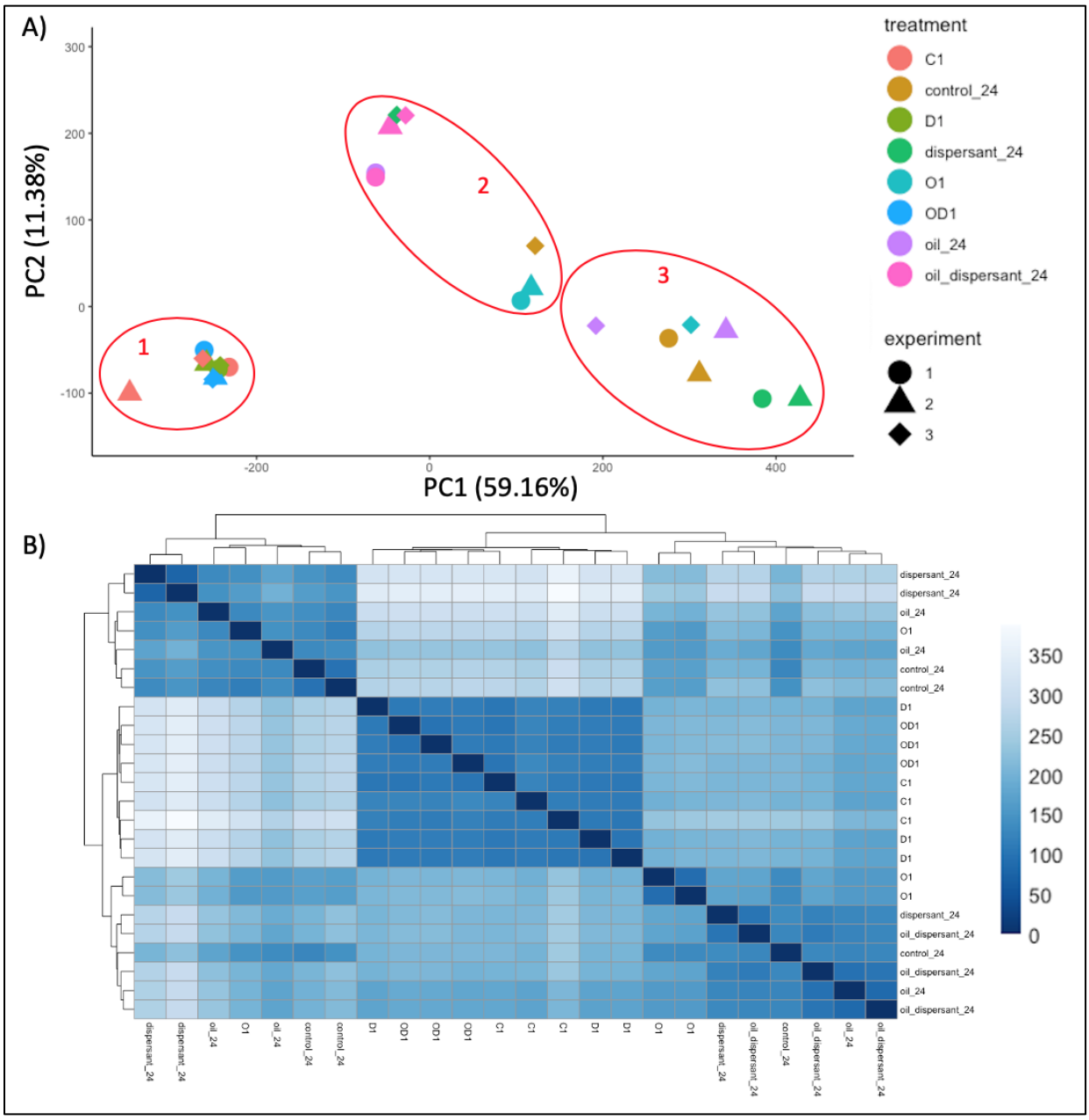


Figure 12: Principal component analysis (PCA, A), and sample to sample distance plot (B) analysis to determine sample grouping by similarity. The PCA and sample to sample plots both agreed to three clusters, with cluster 1 including all 1 hour samples to the exception of 1 hour oil treated samples, and all the remaining samples spread out between cluster 2 and 3.

Effects of Oil (WAF)

Comparison between 1 hour control samples and samples exposed to oil for 1 hour (C1 VS O1) revealed a total of 4. Out of these, 1,140 transcripts had a match with A. queenslandica.

A total of 1,904 transcripts were identified as uncharacterized protein function, meaning those transcripts have a known ORF, but no known functional annotation, and 2,803 transcript identified as TRINITY_DNXXXX_cX_gX. All transcripts identified with their TRINITY ID are transcript built by the TRINITY assembler having a potential open reading frame (ORF) but no known annotation about the function of this transcript. These will later be referenced as TRINITY transcripts. Maximum downregulation was reached by a TRINITY transcript called TRINITY_DN386287_c3_g1 with a log2 fold change of -22.
TRINITY transcript TRINITY_DN392278_c5_g4 was the most up-regulated transcript with a log2 fold change of 21.

A total of 3 heat shock proteins (HSPs) were found upregulated and linked to sponge host. Two of them were Hsp70 and one Hsp60. Other HSPs like Hsp20 or Hsp90 were found to be expressed but were attributed to bacteria, dinoflagellates, or even diatoms, for a total of 17 transcripts. Up-regulation of HSPs can reflect that the organism is facing stress and needs some defense mechanisms to prevent and protect its biological functions from any further harm (Salamanca et al.,2014)

Overall, I found 30 E3 ubiquitin genes associated with the sponge host which were differentially expressed. Two proteins (ubiquitin-like protein FUBI and E3 ubiquitin-protein ligase Siah1-like) were down regulated, and the remaining 28 (for example E3 ubiquitin-protein ligase HERC2, E3 ubiquitin-protein ligase MYCBP2, or E3 ubiquitin-protein ligase NEDD4) were upregulated. Over expression of E3 ubiquitin proteins usually inhibits apoptosis and thus favors the proliferation of cells (Sun, Y. 2006). Both apoptosis and cell proliferation are processes that happen during the cell cycle. A good balance between these two processes insures a proper turnover of the cell number in the organism. Consequently, and imbalance of one of the two will inversely enhance the activity of the other as apoptosis can occurs at any time during the cell cycle (Alenzi, 2004). Furthermore, five initiation factors have shown an up-regulation (for example initiation factor eIF-2B, LFC=5.4, and initiation factor 4E, LFC=2.1). Differential regulation of initiation factors has shown to have some control on cell proliferation and apoptosis. Over expression of such initiation factors favor the cell proliferation (Caraglia et al., 2001). Consequently, a combination of up-regulation in both E3 ubiquitin ligases and initiation factors results in an inhibition of apoptosis and an enhancement of cell proliferation.

Cell death and apoptosis related genes added up to 7 in total. Cell death related genes represented 4 genes and 3 genes identified to apoptotic related proteins. One gene called ‘‘apoptotic protease-activating factor 1-like’’ was downregulated but all others were upregulated. Up regulation of these genes induces a higher rate of cell death, indicating the need to control cell proliferation (Lee et al., 2005).

Among all genes, 87 oncogenes/ tumor suppressor genes (TSGs) have been found differentially expressed. Among them are the well-known Src, the first TSG to be discovered, Rab/Ras related protein, and MYC proteins. These genes have been found previously in sponges (Cetkovic et al, 2004) All these genes have been shown to be present in various cancers, putting the spotlight on sponges to be potential lab models for cancer research. As well this shows the harmful impact of oil exposure, and the carcinogenic consequences linked to exposure (American Cancer Society).

In conjunction to cancer related genes, galectin was found to be up-regulated by almost a 6 times fold. Galectins are a large family of 15 proteins with relatively broad specificity. Thus, they have a broad variety of functions including mediation of cell–cell interactions, cell–matrix adhesion and transmembrane signaling, and apoptosis (Drickamer and Taylor 2011). Galectins have also shown to be very implied in diseases such as cancers and critically influence tumor progression by modulating tumor cell migration, invasiveness, angiogenesis and antitumor immune responses (Sunbald et al., 2013). However, the differentially expressed transcript here only has a partial sequence available in sponges, and consequently only general conclusion can be drawn without the knowledge of which specific galectin is affected (Rabelo et al., 2012). One protein of interest, especially when dealing with oil treated samples, is cytochrome P450. This protein is known to be one of the main metabolic activator of PAHs. After 1 hour of exposure to oil, cytochrome P450 was found in 7 occurrences, always up-regulated and present in both the host and the symbiotic and microbial community. Differential expression of 1 hour oil treated samples is plotted in Fig 13, A using a volcano plot.

After 24 hours of exposure to oil, a total of 268 transcripts were differentially expressed among which only 1 transcript was identified as A. queenslandica, 66 transcripts were singled out as TRINITY, and 143 transcripts were found to be transcripts with uncharacterized protein function. 220 transcripts were downregulated with a maximum log2 fold change of -20 for the

transcript with accession number AMX21890 (psbA gene), and 48 were up regulated all the way to a log2 fold change of 7 for the transcript with the ID TRINITY_DN396576_c0_g3.

While the biggest majority of differentially expressed transcripts in this comparison are hypothetical proteins, the organisms from which it originated was still identifiable. Most of these hypothetical proteins originated from the sponges microbial and symbiotic community.
This suggests that after 24 hours of exposure, mostly the microbial and symbiotic community could be affected by the exposure to the chemical. However, for the reasons explained before about the 24 hours control issues, these conclusions must be verified.
Differential expression of 24 hours oil treated samples is plotted in Fig 13, B using a volcano plot.

Looking at the overlap between the two treatments, 7931 transcripts were uniquely found in the 1 hour exposure treatment, 147 transcripts were uniquely found in the 24 hours exposure treatment, and there was 121 transcripts that were differentially expressed in both time points, as shown in Fig 13, C. Out of the 121 transcripts, 29 were TRINITY transcripts. The majority of the remaining 92 transcripts were associated with the microbial and symbiotic community of the sponge. Unfortunately, most of the 92 transcript were identified as hypothetical or uncharacterized protein.

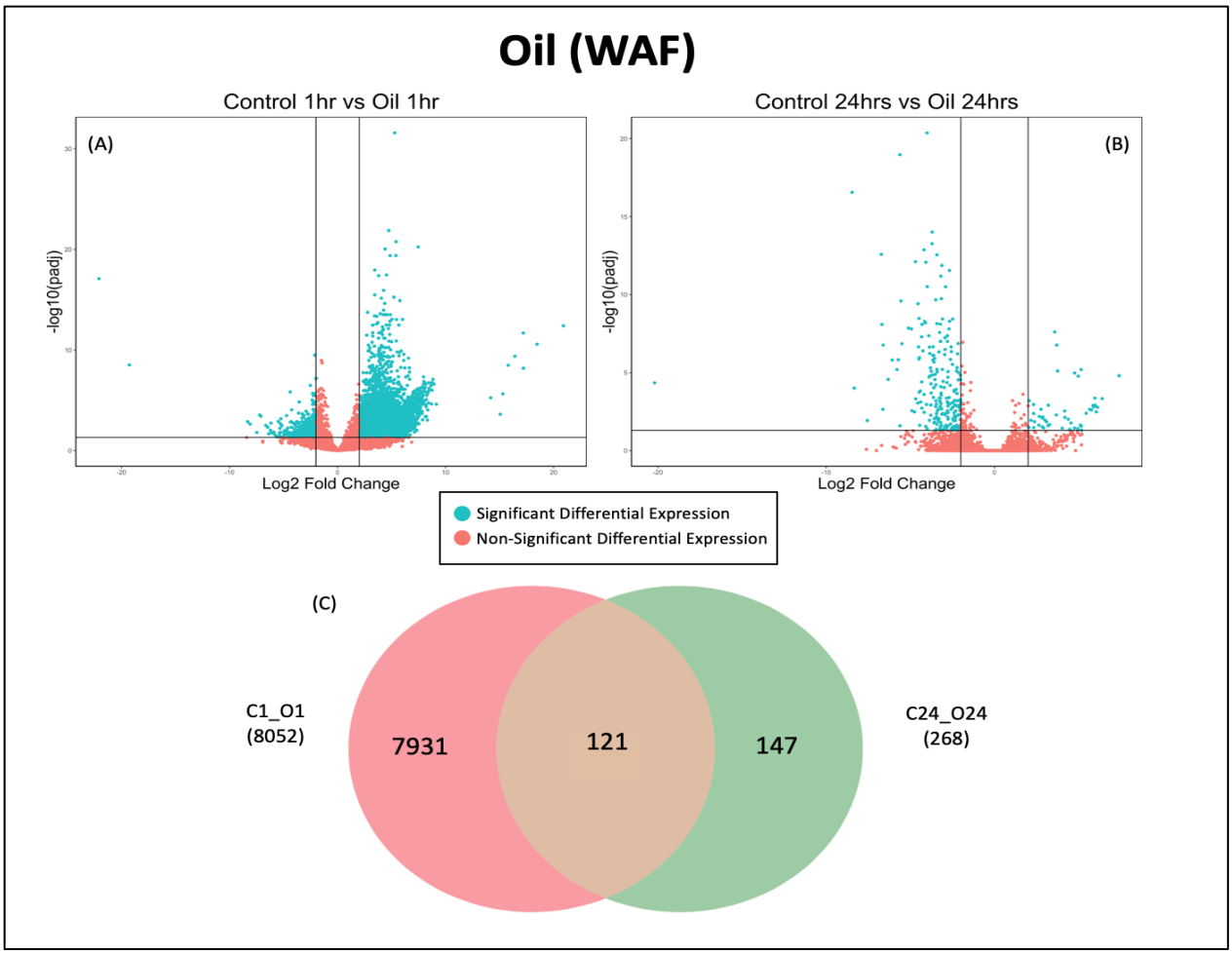


Figure 13: Volcano plots representing the significant differential expression between control samples and oil treated (WAF) samples at each time point (1 hour, A; and 24 hour, B). Each dot on the plot represent a single transcripts. Blue dots represent significant differential expression, and red dots represent non-significant differential expression (threshold was set to |log2foldchange|>2 and q-value<=0,05. Vertical lines represent log2 fold change of 2, and horizontal line represent q-value=0.05. Scale is independent on each graph.(C) Venn diagram of the number of significantly differentially expressed transcripts for each time point. In red, transcripts uniquely found in 1 hour exposure treatment (7,931 transcripts), in green, transcripts uniquely found in the 24 hours exposure treatment (147 transcripts), and in orange commonly found transcripts between the two exposure times (121 transcripts).

The comparison between 1 hour oil treated samples versus 24 hours oil treated samples was also analyzed. Differential expression analysis between 1 hour oil treated and 24 hours oil treated samples revealed few differentially expressed transcripts: 203 transcripts in total (111 downregulated transcripts and 92 upregulated transcripts; Fig. 14). This number of transcripts follows the PCA results given the similarity of the two oil treated samples.

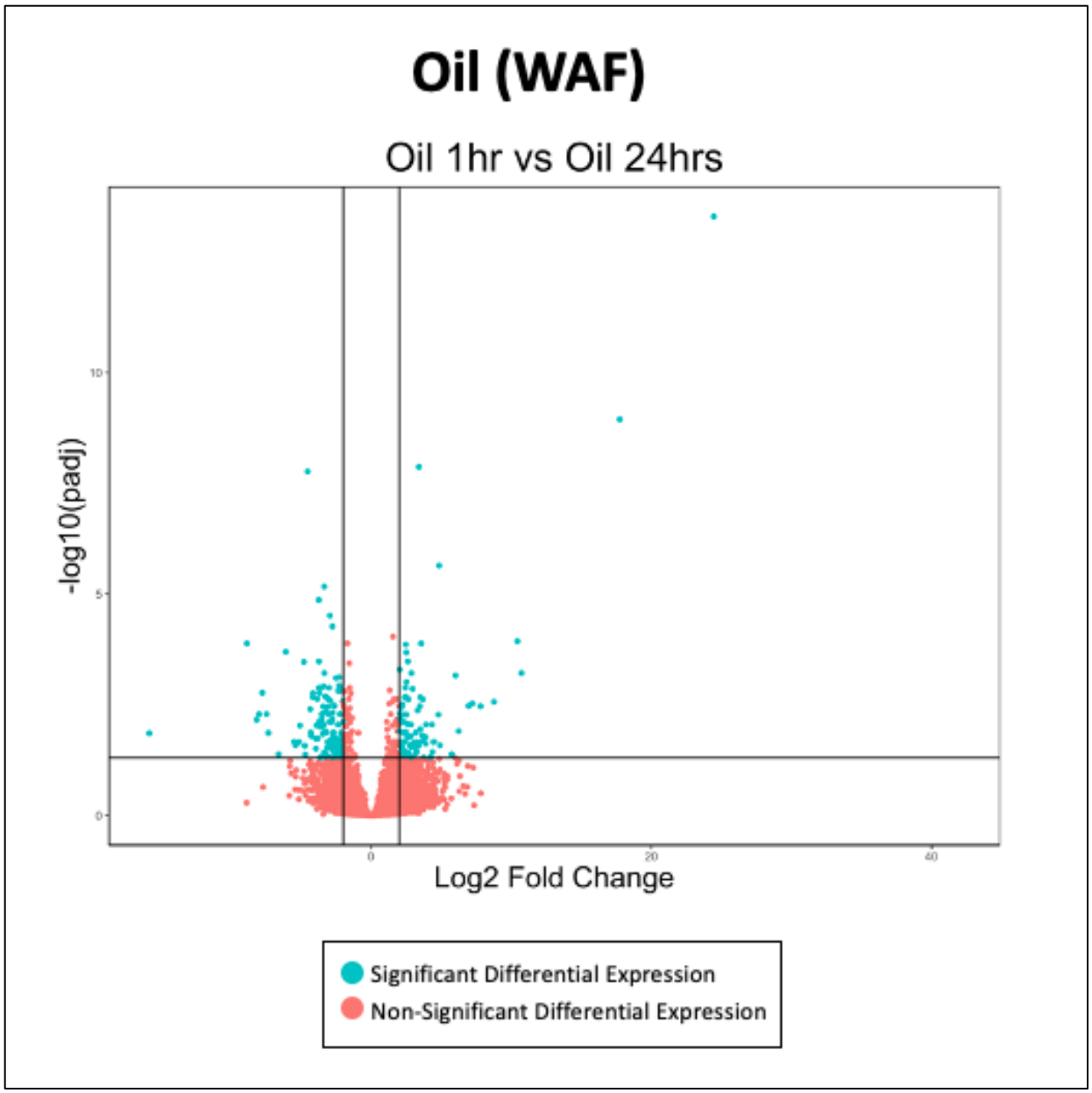


Figure 14: Volcano plot representing the significant differential expression between and 1 hour vs 24 hours oil treated (WAF) samples. Each dot on the plot represent a single transcripts. Blue dots represent significant differential expression, and red dots represent non-significant differential expression (threshold was set to |log2foldchange|>2 and q-value<=0,05). Vertical lines represent log2 fold change of 2, and horizontal line represent q-value=0.05. Scale is independent on each graph.

Effects of Dispersant

Differential expression between C1 and D1 samples revealed only 6 differentially expressed genes, all upregulated, with 4 TRINITY transcripts, 1 hypothetical protein, and one annotated gene identified as peroxisomal sarcosine oxidase. None of the genes belonged to the host sponge. The hypothetical protein belonged to an uncultured bacterium, and the peroxisomal sarcosine oxidase belonged to Astyanax mexicanus, a blind freshwater fish. None of the transcripts were found to be part of the gene categories cited before (Fig. 15, A).

About 24 hours of exposure to Dispersant resulted in differential expression of 308 transcripts, with 16 sponge transcripts (13 A. queenslandica 1 Ephydatia muelleri, 1 Suberites domuncula, 1 Lubomirskia baicalensis), 108 TRINITY transcripts, and 90 transcripts with uncharacterized protein function. Maximum down-regulation was observed for the transcript with the ID TRINITY_DN371367_c0_g1 with a log2 fold change of -41, while maximum up-regulation was observed for transcript with ID TRINITY_DN384023_c12_g1 and a log2 fold change of 7. About 58 transcripts in total were upregulated and 250 were down regulated.

Cathepsin L was found to be downregulated by a fold change of -6. Cathepsin is a known lysosomal endopeptidase enzyme involved in the initiation of protein degradation. A down regulation of this protein would imply a potential decrease in processing of unwanted proteins and thus overall a negative impact on the organism (Dickinson DP., 2002) . Another explanation is that cathepsin L has been found to resemble the central protein component of some sponge spicules or silicatein (Shimizu et al 1998), and thus the decrease in silicatein expression would be consistent with overall sponge tissue degradation.

Filamin B like was downregulated as well (LFC = -2). Filamin is a protein that connects cell membrane constituents to the actin cytoskeleton. Downregulation of this protein could imply a degradation of tissues overall ultimately resulting in the animal's death due to lack of cell cohesion. Downregulation of filamin B has also been shown to favor tumor growth and metastasis (Bandaru et al., 2014).

Protocadherin Fat 4-like expression was significantly lower (LFC = -3.5). FAT4 is a known tumor suppressor gene, protocadherin-FAT4 is part of the cadherins family which are calcium-dependent cell adhesion proteins. With a downregulation of such proteins cell adhesion would be compromised which ultimately affects the integrity of the animal. Downregulation of

Fat4 may influence the cell cycle, ubiquitin hydrolysis, mitogen-activated protein kinase, p53, and apoptosis (Zhang et al., 2016)

Septin 11-like expression was four times lower than in control samples (LFC = -4.7). Septin 11 is a filament forming cytoskeletal GTPase, and part of the septin group which are evolutionary and structurally related RAS oncogenes and comprises 13 septin proteins. It is thought to play a role into cytokinesis. Cytokinesis is the physical process through which the mother cell divides into two daughter cells. Downregulation of septin 11 could results, along of downregulation of other septins, in global disruption of the septin cytoskeleton, and a disruption of the cell cycle leading to compromised survival of the animal in the long run (Hanai et al., 2004).

DNA/RNA-binding protein KIN17-like is a protein involved in DNA replication and the cellular response to DNA damage. It also may be playing a role in the regulation of the genetic expression of an organism. A downregulation of this protein (LFC = -2.4) could imply that the exposure to the dispersant chemical is having a damaging effect on the organism's ability to replicate its DNA. As well, with the potential role played in gene expression regulation, a negative fold change of this gene could have some implications in overall gene expression of the organism, with uncontrolled expression of many life threatening genes. (Miccoli et al., 2003; Biard et al., 2003; Le Maire et al., 2006).

Rap guanine nucleotide exchange factor 1-like isoform X1 (RAPGEF1), also known as C3G or GRF2 is a guanine nucleotide-releasing protein that binds to SH3 domain of CRK and GRB2/ASH. It activates RAS by transducing signals from CRK. RAPGEF1 is involved in cell branching and adhesion mediated by BCAR1-CRK-RAPGEF1 signaling and activation of RAP1 and plays a role in the establishment of basal endothelial barrier function. (Knudsen et al., 1994; Sakakiraba et al., 2002; Hisata et al., 2007; Pannekoek et al., 2011). It has been shown to have a strong regulatory role in cell proliferation, differentiation, and apoptosis. Consequently, an up-regulation of this gene (LFC = 2.6) could imply an increase in cell mortality and thus have an impact in the organism survival in the long run. (Samuelsson et al., 2011).Visual representation of 24 hours dispersant treated samples differential expression is depicted in Fig 15, B.

When looking at the overlap of transcripts differentially expressed between the two dispersant exposure times, 5 transcripts were uniquely expressed in the 1 hour exposure treatment, 307 transcripts were uniquely found differentially expressed in the 24 hours exposure

treatment, and only 1 differentially expressed transcript was found in common between both exposure times. This transcript was a TRINITY transcript with ID TRINITY_DN392278_c5_g4 (Fig 15, C).

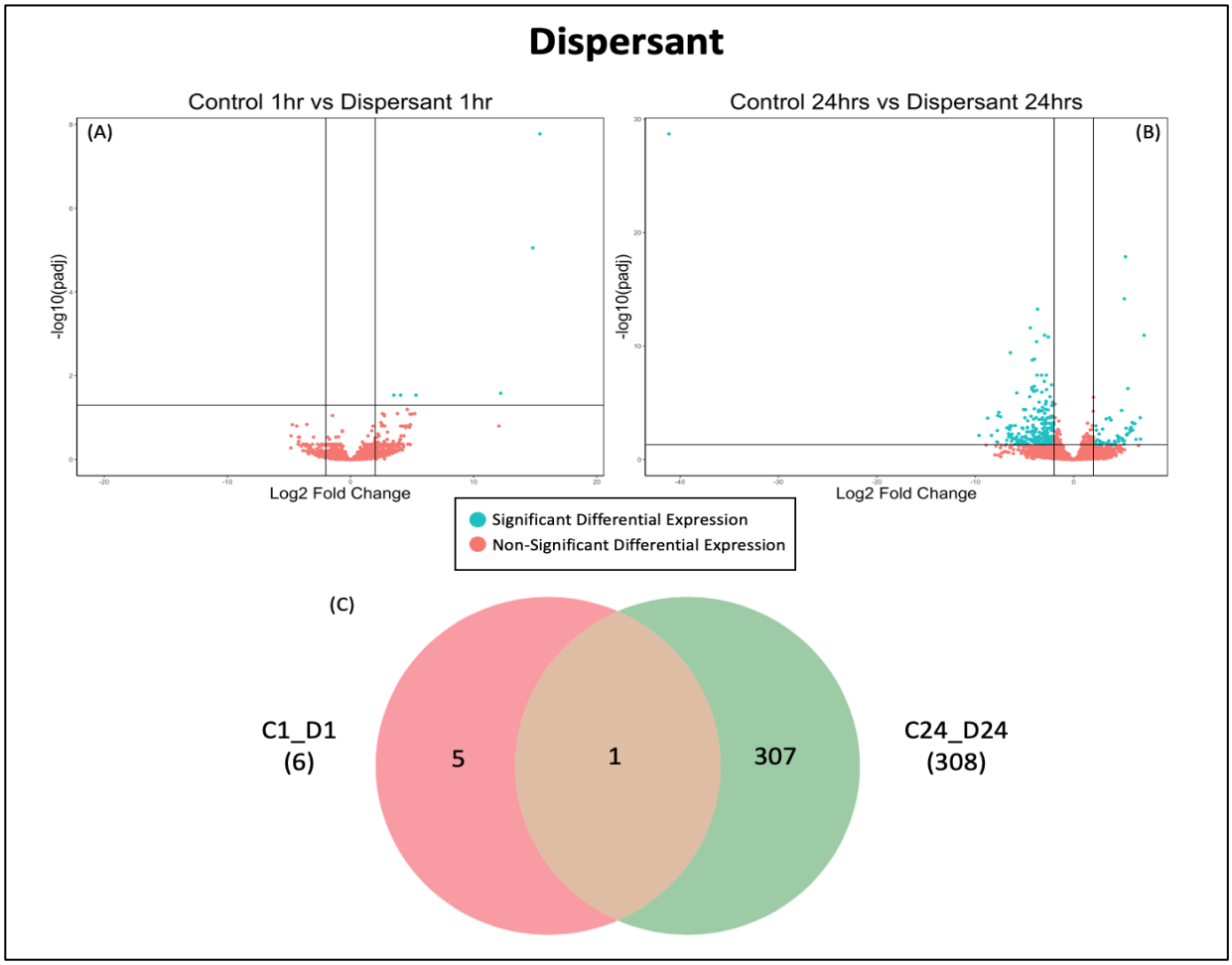


Figure 15: Volcano plots representing the significant differential expression between control samples and dispersant treated samples at each time point (1 hour and 24 hours, A and B). Each dot on the plot represent a single transcripts. Blue dots represent significant differential expression, and red dots represent non-significant differential expression (threshold was set to |log2foldchange|>2 and q-value<=0,05. Vertical lines represent log2 fold change of 2, and horizontal line represent q-value=0.05. Scale is independent on each graph. (C) Venn diagram of the number of significantly differentially expressed transcripts for each time point. In red, differentially expressed transcripts uniquely found in the 1 hour dispersant exposure (5 transcripts), in green differentially expressed transcripts uniquely found in the 24 hours dispersant exposure (307 transcripts), and in orange the only commonly found transcript between the two times of exposure.

The comparison between 1 hour dispersant treated samples versus 24 hours dispersant treated samples were also looked at.. Differential expression analysis between 1 hour dispersant treated and 24 hours dispersant treated samples revealed many differentially expressed transcripts: 15,041 transcripts in total (1,715 downregulated transcripts and 13,326 upregulated transcripts; Fig. 16). The high number of differentially expressed transcripts, is following the PCA results with both dispersant treated samples being far apart and different.

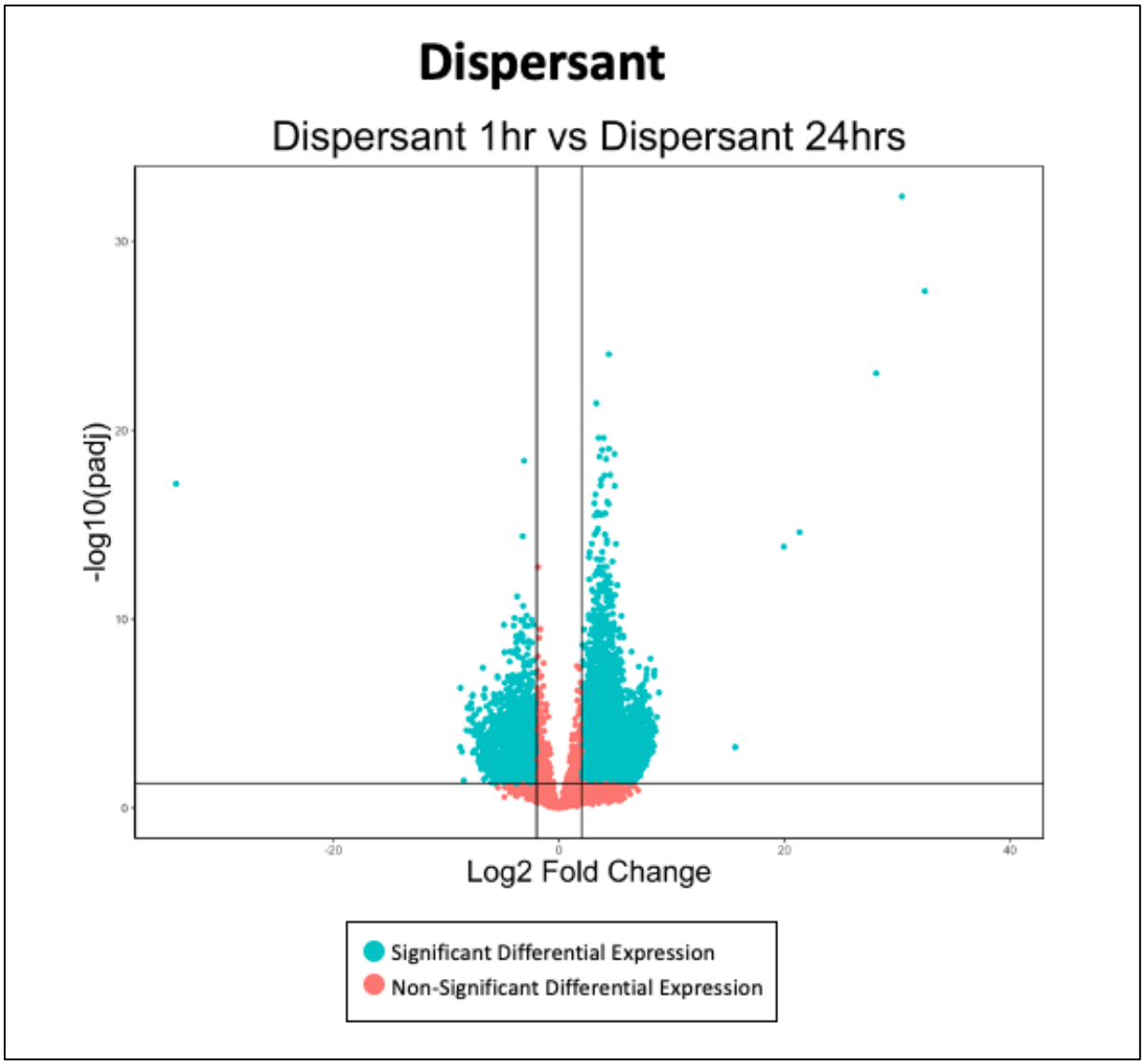


Figure 16: Volcano plot representing the significant differential expression between 1 hour vs 24 hours dispersant treated samples. Each dot on the plot represent a single transcript. Blue dots represent significant differential expression, and red dots represent non-significant differential expression (threshold was set to |log2foldchange|>2 and q-value<=0,05). Vertical lines represent log2 fold change of 2, and horizontal line represent q-value=0.05. Scale is independent on each graph.

Effects of Oil:Dispersant (CE-WAF) mixtures

When comparing CE-WAF treated samples after 1 hour of exposure, a total of 31 transcripts were found differentially expressed. 4
No representative genes of the 4 categories of interest were found expressed after 1 hour exposure.

However, a couple of genes were worth mentioning. Catalase was upregulated by a fold change of 5.4. Catalase is a common protein found in almost all living organisms exposed to oxygen. The protein transform hydrogen peroxide, a by-product of many metabolic processes, into water molecule and oxygen. It is really important to protect the cell from oxidative damage. The upregulation of this gene could potentially mean that the sponge is under great oxidative stress and needs some counter measures to keep surviving (Chelikani et al., 2004).

Tubulin was also seen over-expressed (LFC = 3.9). Tubulin is a microtubule structural protein, one of the major component of the cytoskeleton (Gunning et al., 2015). An up-regulation of this gene could imply an effect of the CE-WAF chemical on the cell integrity and structure and thus the need to produce more tubulin in order to maintain cellular activity and integrity, as well as organism cohesiveness. Differential expression of 1 hour CE-WAF exposed sample is depicted in Fig 17, A.

Exposure to CE-WAF mixture for 24 hours had an impact on 4248 transcripts. In total, 29 transcripts were identified as similar to A. queenslandica, TRINITY transcripts totaled up to 802, and 1516 transcripts had no protein function. Transcript with accession number OGG55450, which correspond to a hypothetical protein from Candidatus Handelsmanbacteria bacterium, was the most down regulated transcript with a log2 fold change of -33. The most upregulation was reached by the transcript with transcript ID TRINITY_DN396576_c0_g3 with a log2 fold change of 7. Overall. 159 transcripts were negatively impacted, and 4089 transcripts saw their expression significantly increased. Most of transcripts with a functional annotation belonged to either the microbial or the symbiotic community of the sponge host. However, several genes belonging to the host and their expression are worth noting.

First of all, three representative of the cytochrome P450 were differentially expressed. Two transcripts belonged to a bacteria and one to the sponge host.

Protocadherin FAT4 who was downregulated in the dispersant 24 hours treated sample was also downregulated in CE-WAF samples (LFC = -2.8).

Galectin, a protein that was also over expressed in 1 hour oil treated samples, was also up-regulated in CE-WAF samples (LFC = 4.2).

Continued expression of tubulin was witnessed between 1 hour CE-WAF and 24 hours CE-WAF samples, with a similar up-regulation ratio (LFC = 3.4).

Additionally, some new transcripts ended being differentially expressed. Calcineurin binding protein cabin 1 was up-regulated by a fold change of 2.6. Calcineurin has shown to potentially serve as negative regulator of T-cell receptors and may be required for replication independent chromatin assembly. But more importantly, it acts as a negative regulator of p53/TP53, a well-known TSG, by keeping p53 in an inactive state on chromatin at promoters of a subset of its target genes (Sun et al., 1998; Tagami et al., 2004).

TGF-beta receptor type-1 (TGFBR1) was up-regulated by a 3 times fold. TGFBR1 is a transmembrane serine/threonine kinase forming with the TGF-beta type II serine/threonine kinase receptor, TGFBR2, the non-promiscuous receptor for the TGF-beta cytokines TGFB1, TGFB2 and TGFB3. It transduces the TGFB1, TGFB2 and TGFB3 signal from the cell surface to the cytoplasm and is thus regulating a plethora of physiological and pathological processes including cell cycle arrest in epithelial and hematopoietic cells, control of mesenchymal cell proliferation and differentiation, wound healing, extracellular matrix production, immunosuppression and carcinogenesis (Weiser et al., 1995; Macias-Silva et al., 1995; Ozdamar et al., 2005). Differential expression of 24 hours CE-WAF exposed sample is depicted in Fig 17, B.

Overall, 22 transcripts were uniquely found differentially expressed in the 1 hour exposure treatment, 4,239 transcripts were uniquely found differentially expressed in the 24 hours exposure treatment, and 9 transcripts were found to be differentially expressed in both 1 hour and 24 hours CE-WAF treated samples. Out of these, 5 were TRINITY transcripts, and in the remaining 4, 1 transcript was associated to a bacterium and 3 to three different marine invertebrates (Fig. 17, C).

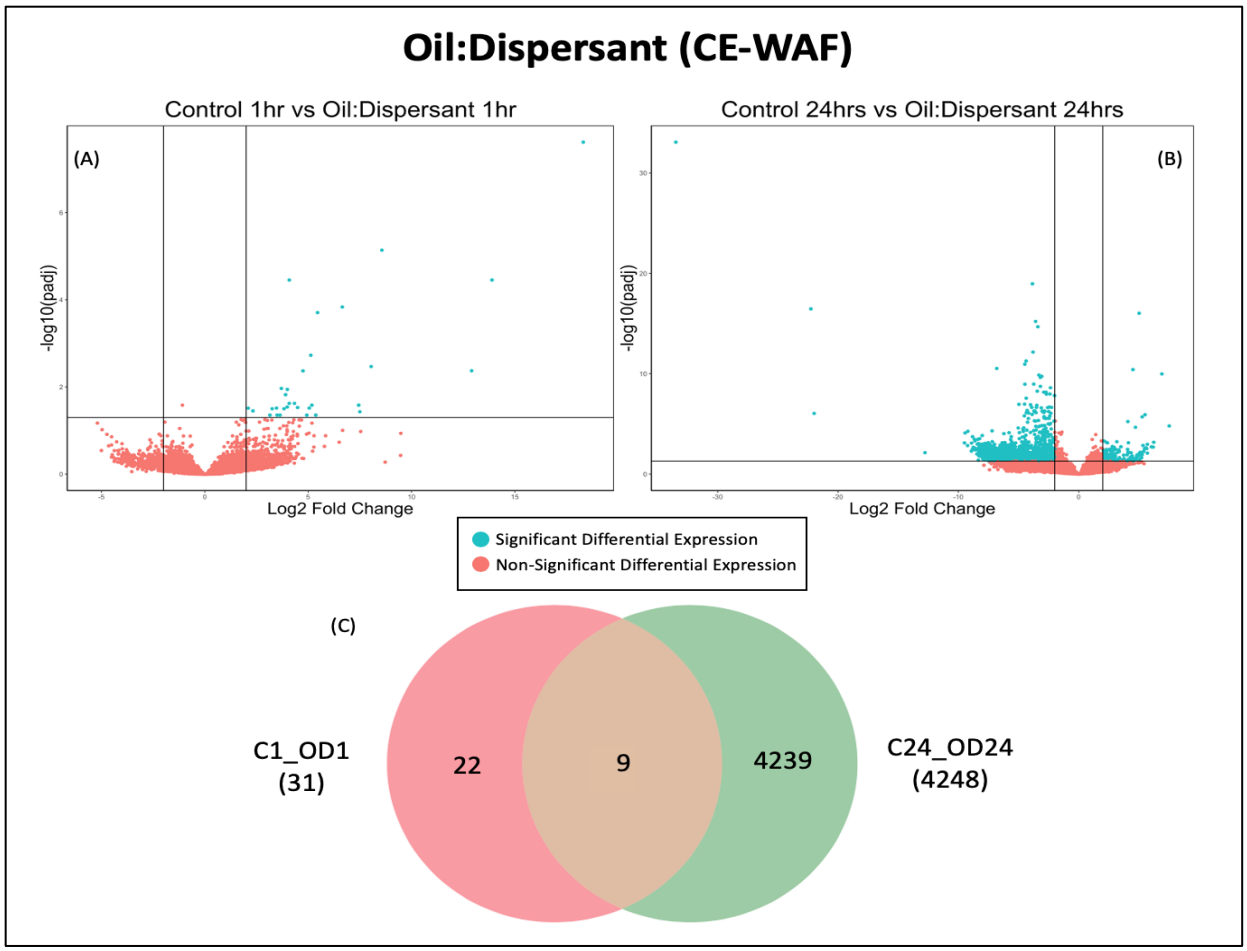


Figure 17: Volcano plots representing the significant differential expression between control samples and oil :dispersant treated (CE-WAF) samples at each time point (1 hour and 24 hours, A and B). Each dot on the plot represent a single transcript. Blue dots represent significant differential expression, and red dots represent non-significant differential expression (threshold was set to |log2foldchange|>2 and q-value<=0,05. Vertical lines represent log2 fold change of 2, and horizontal line represent q-value=0.05. Scale is independent on each graph. (C) Venn diagram of the number of significantly differentially expressed transcripts for each time point. In red, differentially expressed transcripts uniquely found in the 1 hour oil:dispersant exposure (22 transcripts), in green differentially expressed transcripts uniquely found in the 24 hours dispersant exposure (4,239 transcripts), and in orange the9 commonly found transcript between the two times of exposure.

The comparison between 1 hour oil:dispersant treated samples versus 24 hours oil:dispersant treated samples were also looked at. Differential expression analysis between 1 hour oil:dispersant treated and 24 hours oil:dispersant treated samples revealed many differentially expressed transcripts, with again a dominating trend of up-regulation: 6,116 transcripts in total (876 downregulated transcripts and 5,240 upregulated transcripts; Fig. 18, B).

The difference between the two oil:dispersant treated samples followed the PCA results as they appeared different and very distant from each other.

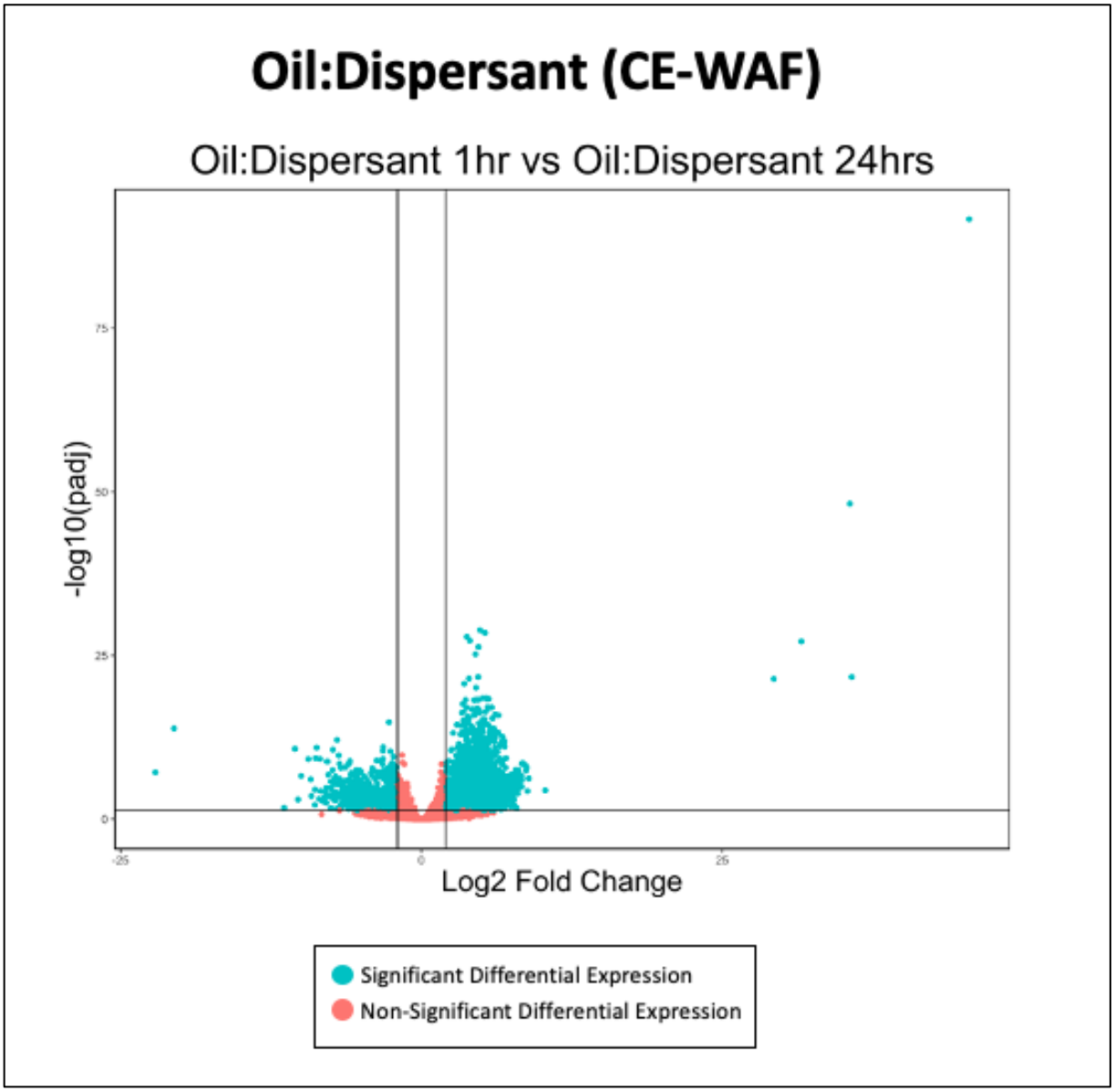


Figure 18: Volcano plots representing the significant differential expression between 1 hour vs 24 hours oil:dispersant treated (CE-WAF) samples. Each dot on the plot represent a single transcript. Blue dots represent significant differential expression, and red dots represent non-significant differential expression (threshold was set to |log2foldchange|>2 and q-value<=0,05).Vertical lines represent log2 fold change of 2, and horizontal line represent q-value=0.05. Scale is independent on each graph.

Control Samples Comparison

Comparing differential expression between the two control time point was performed. This comparison provided information regarding the effect of the culturing condition on the gene expression changes. When comparing the gene expression levels of Control 24 hours samples versus Control 1 hour samples, a total of 9,682 transcripts were significantly differentially expressed (8,711 upregulated transcript and 971 down-regulated transcripts). This means that after a 24 hours period, a lot more gene expression is witness in the organisms as shown in Figure 19.

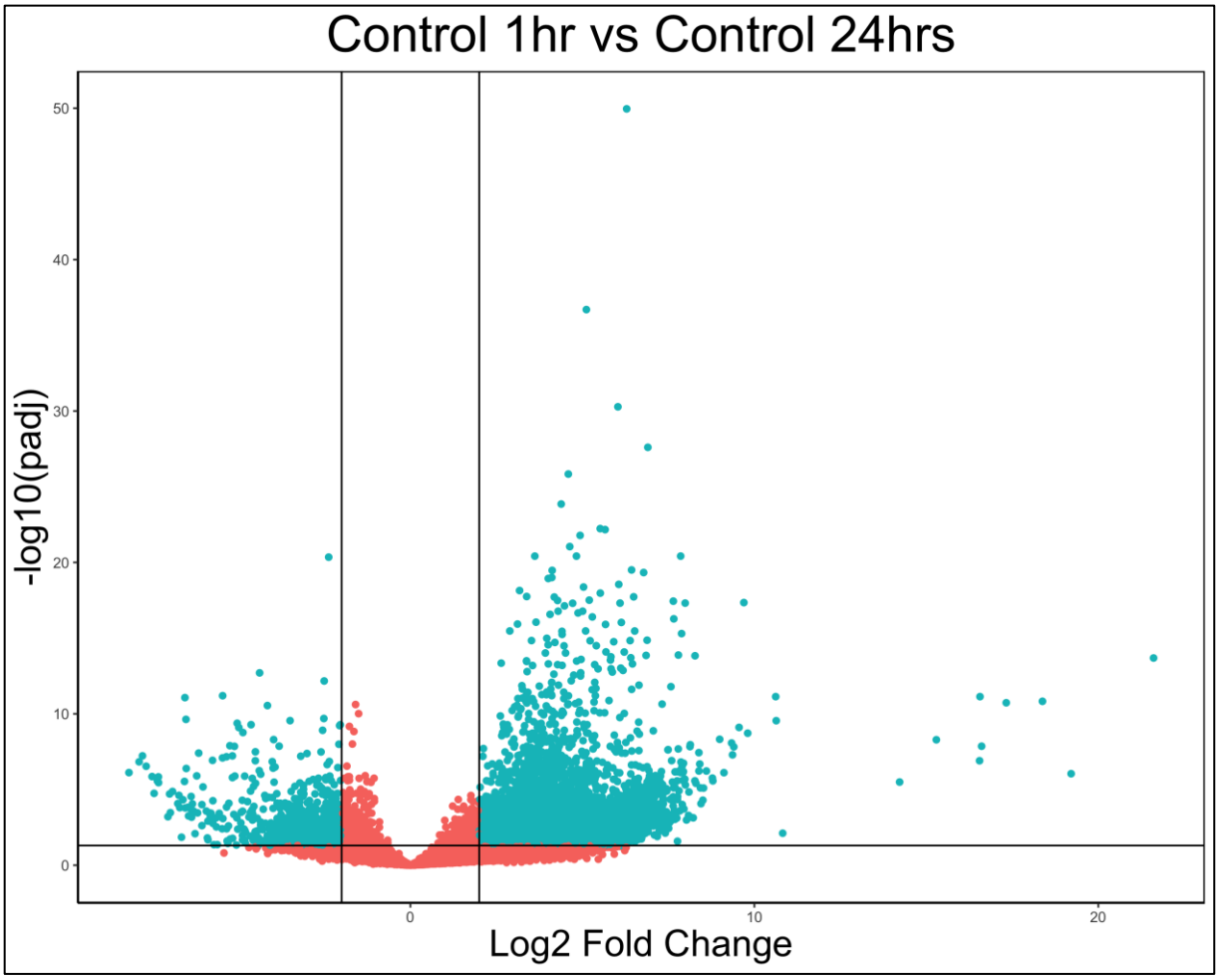


Figure 19: Volcano plots representing the significant differential expression between 1 hour control versus 24 hours control samples. Each dot on the plot represent a single transcript. Blue dots represent significant differential expression, and red dots represent non-significant differential expression (threshold was set to |log2foldchange|>2 and q-value<=0,05).Vertical lines represent log2 fold change of 2, and horizontal line represent q-value=0.05.

Shared Transcripts Between All Treatments

Commonly found genes that were differentially up or down regulated between treatments at a given time point were investigated based on the holo-transcriptome. For example, across all 1 hour comparisons, it was found that O1 and D1 had 5 common differentially expressed genes. Four transcripts had no annotation, and 1 transcript was identified as peroxisomal sarcosine oxidase (Protein abbreviation is PIPOX)

Nine transcripts were differentially expressed in both O1 and OD1 samples, 3 of which were TRINITY transcripts and 6 with an annotation. Among these annotated transcripts, 2 were hypothetical proteins. The remaining four coded for ribosomal protein (2), 1 was bacteria related, and the last one was coding for catalase.

And only 1 transcript was found to be present in both D1 and OD1 samples. This transcript had no annotation and was identified only as TRINITY_DN392278_c5_g4. Interestingly that same transcript is the only transcript differentially expressed in all 1 hour treatment (Fig. 22).

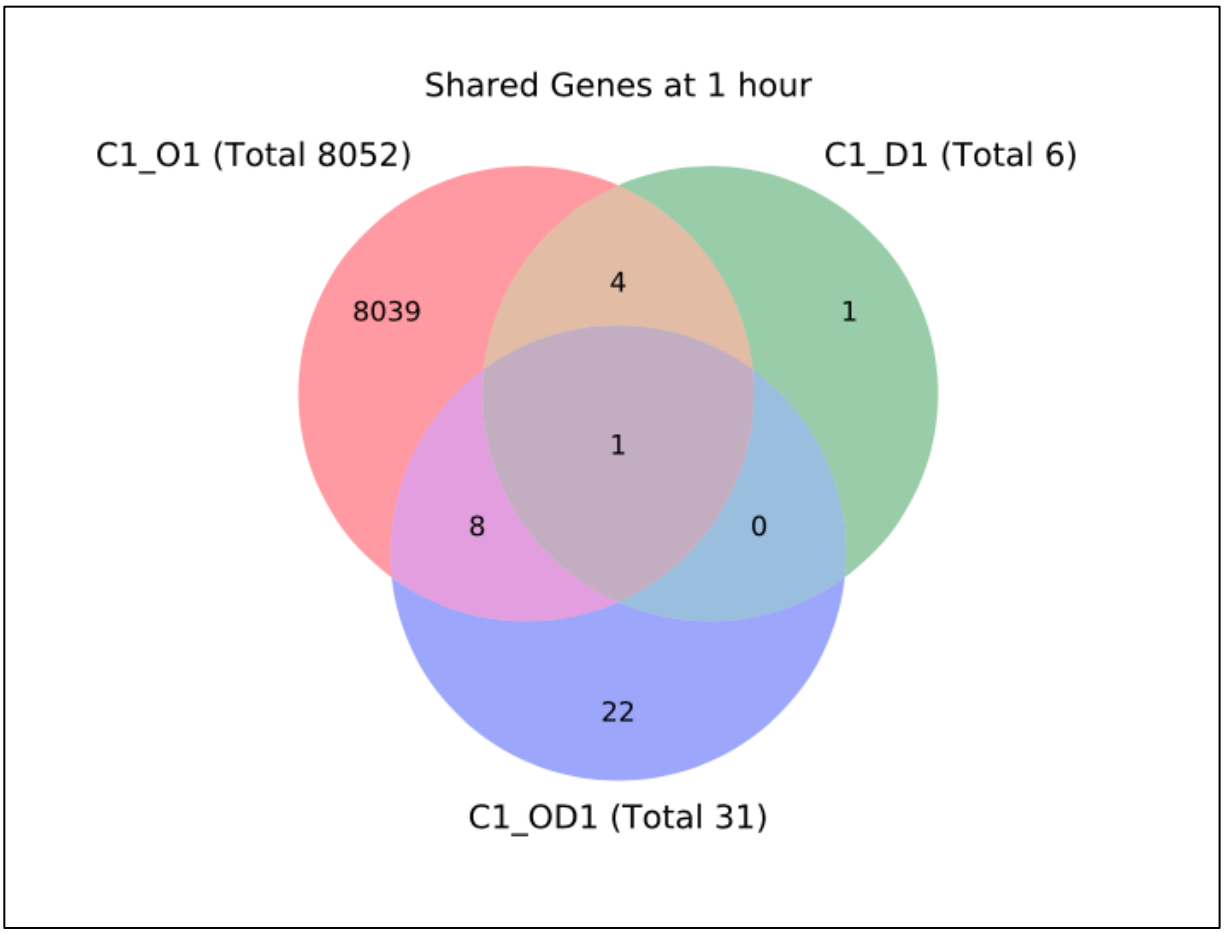


Figure 20: Venn diagram representing the number of differentially expressed genes across all 1 hour treated samples compared to 1 hour control samples, and the number of commonly found transcripts between all treated samples

A total of 116 transcripts were present across all treatments for the 24 hours of exposure samples. O24 and D24 samples ended up sharing 126 transcripts of which 91 were annotated and 35 were identified as TRINITY transcripts. Between O24 and OD24 samples, 210 common transcripts, of which 56 were TRINITY transcripts, and 154 were annotated. Finally, 144 transcripts were present in both D24 and OD24 samples, with TRINITY transcripts accounting for 48 of them, and the remaining having an annotation (FIG. 23).

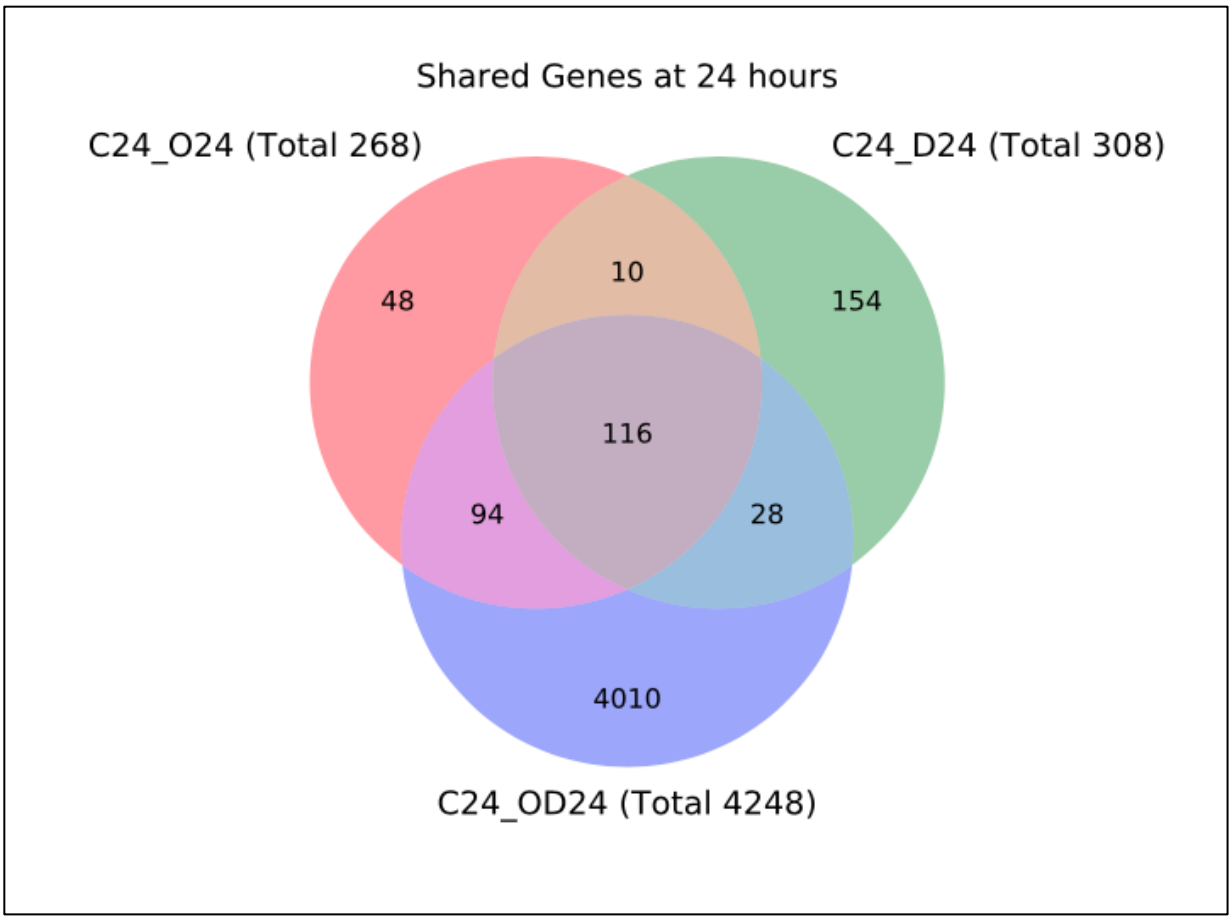


Figure 21: Venn diagram representing the number of differentially expressed genes across all 24 hours treated samples compared to 24 hour control samples, and the number of commonly found transcripts between all treated samples.

A list of genes commonly expressed between treatments, at each time point can be found in the appendix 1.

Sponge Assembly

The same analyses were run with using only the transcripts bouncing off as sponges identified transcripts from the main assembly.

Sample clustering using a PCA showed that the samples clustered in two clusters. Cluster 1 was made of the three replicates for the treatments Dispersant 1hr, Oil:Dispersant 1hr, and Control 1 hr. Cluster 2 embodied all three replicates of the remaining samples (Oil 1 hour, Control 24 hours, Oil 24hrs, Dispersant 24 hours, and Oil:Dispersant 24hrs) as seen in Fig. 24, A. This clustering was confirmed by the sample to sample distance plot (Fig. 24, B).

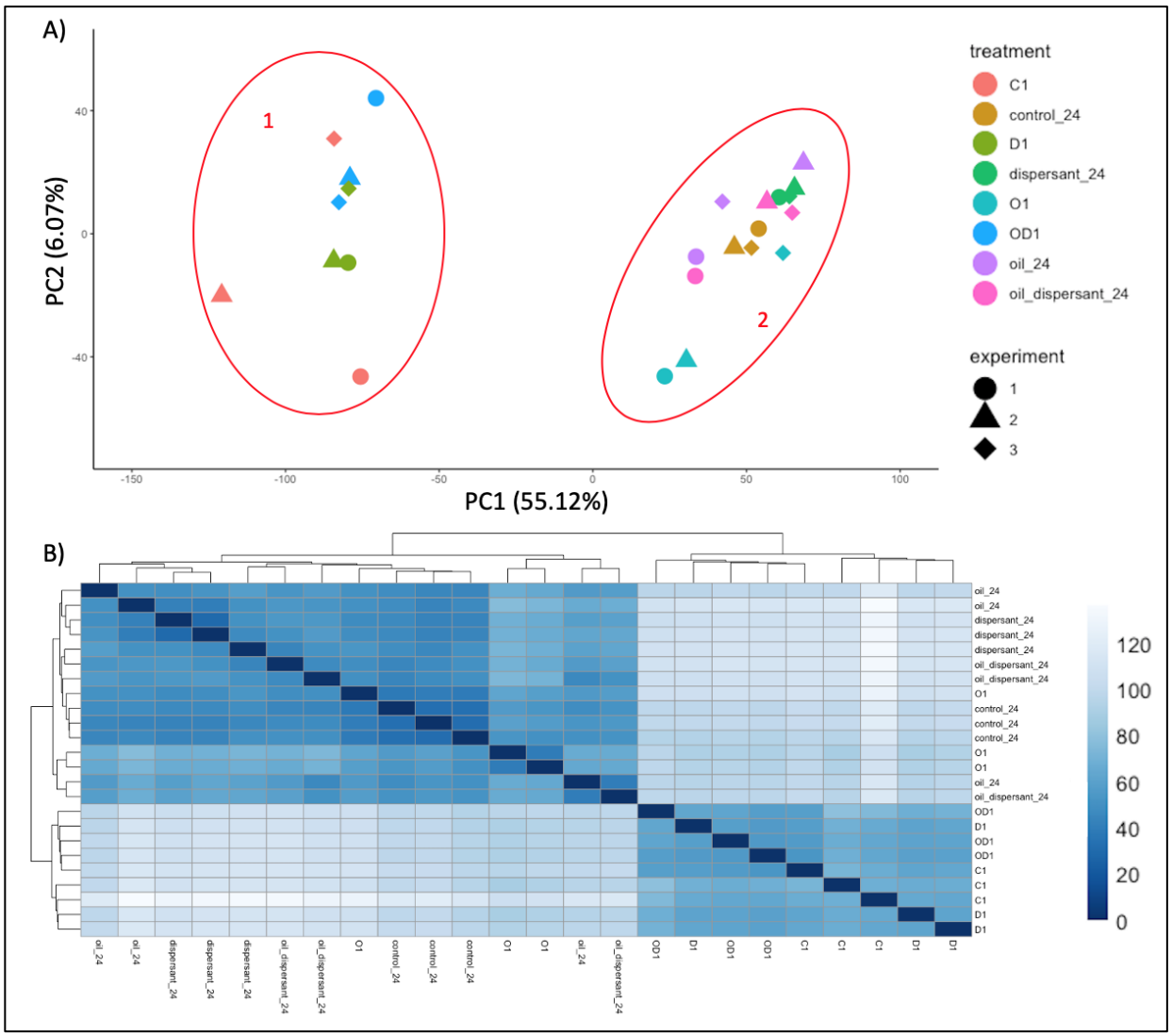


Figure 22: (A) Principal component analysis (PCA), and (B) sample to sample distance plot analysis to determine sample grouping by similarity. The PCA and sample to sample plots both agreed to two clusters, with cluster 1 including all 1 hour samples to the exception of 1 hour oil treated samples, and cluster 2 including all the remaining.

Filtering of sponges related transcripts resulted in a pool of 3968 transcripts considered for differential expression. Overall, 1391 transcripts ended being up-regulated (LFC >0), 519 were down-regulated (LFC < 0), 0 were considered low counts and 3 transcripts were considered as outliers. As seen in Table 9 and Figure 21 below, besides the effect of oil at 1 hour, very few genes were seen differentially expressed. It is however interesting to note that despite the low number of differentially expressed gene in this analysis, similar pattern as the ones with the full

assembly were followed. Oil impact seemed to reduce over time of exposure while longer exposure to dispersant and oil:dispersant mixture seemed to be more detrimental over time with an emphasis on down-regulation.

Table 9: Number of differentially expressed genes in the sponge hits assembly

Pairwise Comparison	# Differentially Expressed genes	Up-regulated	Down-regulated
C1_O1	1281	1154	1273
C1_D1	0	0	0
C1_OD1	0	0	0
C24_O24	2	0	2
C24_D24	16	3	13
C24_OD24	8	2	6
O1_O24	115	25	90
D1_D24	1696	1334	362
OD1_OD24	1487	1243	244
C1_C24	1411	1191	220

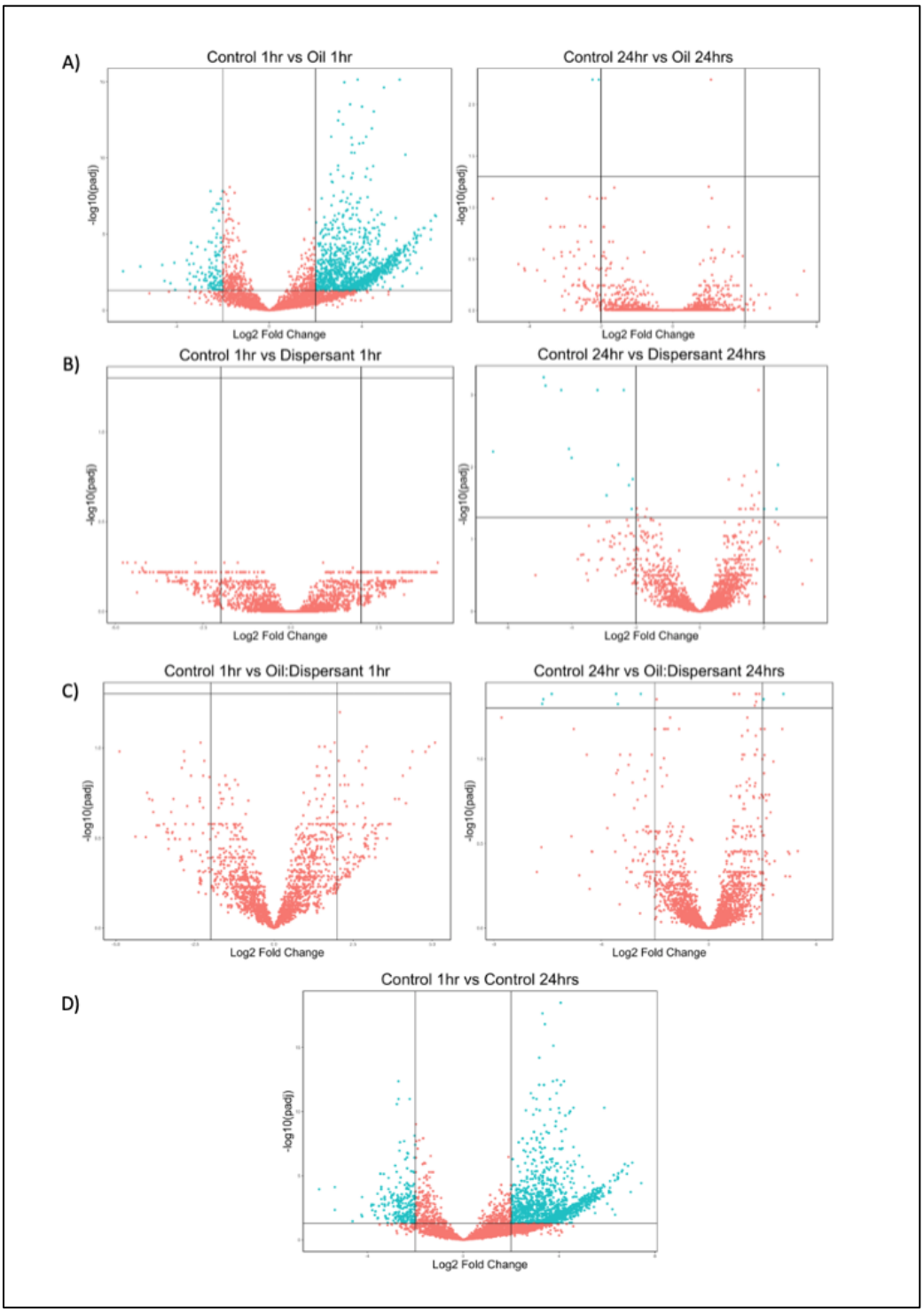


Figure 23: Volcano plots representing the significant differential expression between control samples and treated samples at each time point (oil (WAF) in (A), dispersant in (B), and oil:dispersant (CE-WAF) in (C)). (D) represents the comparison between the two time points of the controls. Each dot on the plot represent a single transcripts. Blue dots represent significant differential expression, and red dots represent non-significant differential expression (threshold was set to |log2foldchange|>2 and q-value<=0.05. Scale is independent on each graph.

It was also interesting to note that when comparing the differentially expressed genes between the general assembly and the assembly comprising only sponge hits a total of 1068 transcript were shared for the O1 treatment, no transcript were shared for the O24 treatment, 9 transcripts were found in both analysis for D24 treatment, and 5 common transcripts appeared for the OD24 treatments.

The same numbers of regulated transcripts should be expected when comparing the full assembly analysis and the assembly made of only sponge transcripts. However, because the size of the pool of transcripts was so different, even if the same parameters were kept during the analysis, the fitted model used by DESeq was slightly different between the two types of analysis. For that reason, the results were slightly different from each other.

DISCUSSION

Environmental impacts of oil spill in a marine environment can be seen throughout the water column. It affects all levels of the food chain from filter feeder to macro-fauna organism. Understanding the effect of oil and chemical dispersant at the genetic scale provides insight on ecosystem answers in the face of environmental stress and its capability of bouncing back. Sponges being very important filter feeders are prime candidates in such studies and can be used as bio-indicators in the face of environmental disturbances.

Physiological Effects of Oil and Dispersant

Like many crude oils, the Macondo crude oil is made of many different organic compounds (Reddy et al., 2012). Most of these compounds are 2 or 3 ringed PAHs that are very volatile. The effects of oil and dispersant exposure is generally witnessed in situations of extreme oiling, in which exposure is way past the sub-lethal level (Peiffer and Cohen, 2015; Gardiner et al., 2013). Such impacts are mainly seen on birds or mammals, but can also be seen in shellfish, plants and microorganisms. Because of its high volatility, the effect of crude oil on the respiratory system of birds and mammals is devastating, causing severe inflammation, incapability of breathing, immune system deficiency, and reproduction. In fur bearing animals, crude oil impairs their water repellency capabilities, making them very vulnerable to hypothermia, leading to their death by drowning. Corals, shellfish, fishes, and other micro-

organisms (microbes, phyto/zoo-plankton) all see effect on their development stages, reduced growth, and even reproductive impairment. It has also been shown that when sublethal doses occur, zoo-plankton ingest oil droplets, which ultimately results in oil introduction in the food chain.

The primary role of dispersant is to break down floating oil particles so the oil can be transferred from to the surface to the depth and be degraded more easily. After many studies, the toxicity level of dispersant appears to be species dependent and no general trend can be drawn as every organism seems to physiologically respond differently to its exposure (Wise and Wise, 2011, Cameron et al., 2018). It has however been shown that longer exposure to dispersant results in higher negative impacts. Chemical dispersant exposure has shown to impact the larvae development in many species (Adayemo et al., 2015; Beirao et al., 2019; Xu et al., 2016; Mu et al., 2014), disrupt microbial and phyto/zooplankton communities (Hook et al., 2012; Lamendella et al., 2014; Mason et al., 2014), and reproduction abilities of some species (Zhang et al., 2012; Lane et al., 2015; Han et al., 2014).

While the physiological effect of chemical dispersant seems to be species dependent, chemically dispersed oil (CE-WAF) mixtures studies have all shown an increase in toxicity compared to oil and dispersant only exposures. Mixtures of oil and chemical dispersant have shown to potentiate physiological responses in each organisms exposed to the mixtures (Hook et al., 2012; Zhang et al., 2012; Adeyemo et al., 2015; Ruiz-Ramos et al., 2017). The break-down of oil into smaller particles by the chemical dispersant makes it more harmful to the environment as it change the crude oil composition (Couillard et al., 2005) and can be ingested by some species which results in the introduction of both dispersant and oil in the food chain.

However, while extreme oiling impacts are easy to detect, with obvious effects, studying observing and quantifying of sublethal doses are more challenging. In this case study, after all sponges looked visually healthy to the naked eye, and it was only noticeable that some impact had occurred when tissues were observed under powerful microscopy (SEM). Even at this resolution, tissue effects were not obvious and might not tell the full story. In our controlled experiments, we could observe some gross changes in the mesohyl of the sponge (Fig. 10). Consequently, in the case of sublethal exposure, the genetic approach should be favored in order to understand what is happening to the organism. For example, in multiple occurrences in this study, apoptosis (cell death) related genes were up-regulated indicating that the cells were not

viable anymore and that they needed to be discarded. Other important cell structure proteins such as actin, and myosin were down-regulated, implying that the cells capabilities to keep their conformational structure was compromised.

The Cinachyrella Holo-Transcriptome

Whole transcriptome sequencing and assembly revealed how challenging it is to deal with genomic data originating from the phylum Porifera. It should not be forgotten that an extant marine sponge is composed of not just one organism, but multiple species of bacterial, fungal, and protozoan besides animals cells because of symbiosis (Webster and Thomas, 2016; Lopez et al., 2019). Indeed, the transcriptome assembly identified many bacterial, fungal, and protozoan related genes, reminding that sponges are very complex animals. However, when sorting through the genomic data for poriferan related transcripts, the resulting assembly comprised a number of transcripts agreeing with previous findings and estimations. Here, a total of 40,024 poriferan related transcripts have been found in the assembly, coding for 8,496 unique proteins. Sponges have shown to have between 17,000 and 41,000 genes in their genomes packed into 23 micro-chromosomes (Renard et al., 2018).These high numbers of genes found in sponge genomes was achieved by deep developmental sequencing of these organisms. Here we have sequenced the holo-transcriptome of fully developed organisms, which explains why 8,496 unique genes have been identified. Porifera is one of the earliest emerging taxa on earth and their genetic complexity is a witness of their survivability through the different eras. Their genomic patrimony is a good witness of the C-value paradox (a term to initially describe the observation that genome size does not correlate with the organism complexity), nowadays referred more frequently as the "C-value enigma". The paradox aspect tends to imply a one-dimensional solution, to what is in reality a multi-faced puzzle, given the knowledge of non-coding DNA we have today.

Electronic Microscopy (SEM)

The electronic microscopy of exposed specimens revealed that even though the sponges looked healthy to the naked eye, when put under an electronic microscope tissues looked to be affected even after 1 hour of exposure. Without being dead yet, after 1 hour of exposure, tissues

seemed to be starting to retract, with spicules being apparent. However, after 24 hours of exposure, it is clear in all treatments that tissues have started to degrade and that the animal is facing major stress. It is thus possible to think that the longer these sponges are exposed to these chemical, the more detrimental it is to these organisms and their tissues. Eventually this could lead to organism's death. Unpublished data (Vijayan, N., unpublished) also showed the presence of oil droplets present in pumping channels of the sponge. This could potentially provide another explanation to the tissue degradation of the sponge. Ultimately the negative impact of the chemicals would both be seen at the genetic level since the exposure of stuck droplet within the organism would impact more than just particles passing through the animal, but as well, this would impact the ability of the animal to feed and thus would impair its survival abilities.

Dosage Sample Grouping

One aspect revealed by our PCA analysis worth noting was the clustering of the 24 hours controls. These samples were expected to cluster within cluster 1 closer to 1 hour control samples. It is believed that even under the culturing methods described in the materials and methods, 1 hour and 24 hours control samples should not have been so dissimilar. This is confirmed by the comparisons between the 1 hour control samples and the 24 hours treated samples, as well as the comparisons between the hour treated samples and 24 hours treated samples. In contrast 24 hours control samples behaved like 24 hours oil treated samples, which could indicate cross-contamination between treatments. However, even if kept in the same room, and high volatility of ≤3 ringed PAHs contained in oil products, this was least likely. Those volatiles compounds are able to vaporize in the air but are not able to solubilize back into the water, most likely excluding this cross contamination theory.

Another hypothesis would imply the culturing methods. Even though as many potential confounding variables were accounted for: light cycle, temperature, water flow, and oxygenation, acclimation time could have been an issue. The experiment was performed after 24 hours of acclimation of the sponges in tanks after collection in the wild. This 24 hour acclimation period could have been too short after collection and would have consequently needed to be longer prior to starting the dosage experiment.

The third hypothesis would be concerning a potential batch effect. While batch effects can happen in many fields, they are most commonly discussed within the field of genomics. Batch effects happen when non-biological factors in an experiment cause changes in the data produced by the experiment. In the case of batch effects, some effective algorithm as R packages have been developed to correct these artifacts. The most well-known package to do so is a package called “sva” in which the ComBat function achieves that correction.

Genetic Profiling

Gene expression profiling of sponges from the Cinachyrella spp. exposed to Macondo crude oil (WAF), dispersant, and oil:dispersant (CE-WAF) mixture from the Deepwater Horizon catastrophe of 2010 in the Gulf of Mexico revealed a total of 12,913 transcripts significantly differentially expressed across all treatments (O1=8,052, D1=6, OD1=31, O24=268, D24=308, OD24=4,248). After exposure to these chemicals for up to 24 hours, a large number of differentially expressed genes was observed after 1 hour of exposure to oil which could indicate that oil and dispersant can elicit large genetic and cellular responses fairly quickly.

Here we have focused on 4 categories of proteins: Heat Shock Proteins (HSPs), E3 ubiquitin ligases, cell death and apoptosis related proteins, and oncogenes/tumor suppressor genes (TSGs). These protein categories were selected due to their identification in preliminary work started by Emily Smith in 2013.

Heat shock proteins (HSP), are often proteins produced by many organisms when under a stressful situation. Many proteins part of this group function as chaperone, meaning they act as controllers ensuring the correct folding of other proteins or the refolding of damaged proteins due to the stress the cell has been under. HSPs have various functions in an organism ranging from simple management function of proper protein conformation under non-stressful conditions, to some potential implication in cancer cell deaths apoptosis (Salamanca et al., 2014).They have been found to be part of the chemical defensome of several organisms (Goldstone 2008 ; Goldstone et al., 2006 ; Shinzato et al., 2012) and are found virtually in all living organisms from bacteria to humans. HSPs are named after their molecular weight in kilodaltons. Three HSPs have been mostly studied: Hsp60, Hsp70, and Hsp90, with respective weights of 60 kDa, 70 kDa, and 90kDa.HSPs have been found to be upregulated under stress

conditions, which is also the case in this study. The observed HSPs in this dosage experiment study have been shown to be up-regulated after exposure. This proves that no matter which treatment the sponge is exposed to, it is put under stressful conditions. Consequently it is trying to protect its physiological functions by producing the proteins that help achieving that goal.

Another protein category important to the organism survival are the ubiquitin proteins. Three types of ubiquitin exist: E1s, also called ubiquitin activating enzymes, E2s, which are conjugating enzymes, and E3s that are ubiquitin ligases. These proteins are essential in many biological processes which include: endocytic trafficking, inflammation, translation, DNA repair, or apoptosis (Miranda et al., 2007; Teixeira and Reed, 2013) Over 600 ubiquitin proteins are found in the human genome, and they are present in all living organisms. In this study mostly E3 ubiquitin have shown differential expression with the big majority being up-regulated. Furthermore, several initiation factors have also been seen up-regulated. The up-regulation of these two types of genes are favoring the cell proliferation by inhibiting apoptosis. This shows that exposure to the chemical has the capacity to change specific steps of the cell cycle and ultimately affect the organism survival.

The third category of proteins of interest were those coding for cell death, and apoptosis. These proteins are produced as marker to indicate the non-viable state of cell, and consequently the need to get rid of a non-usable component of the body. Up-regulation of these types of protein indicate an increase in cell damage and thus a decrease in cell viability. Down-regulation of such proteins, on the other hand indicate an uncontrolled cell proliferation, which for example helps tumors to evade the cell death and promotes drug resistance, a common issue seen in many cancers (Berger and Pu, 2018; lee et al., 2005; Prasad et al., 1997). In this experiment the biggest majority of cell death and apoptosis related proteins were up -regulated. This is expected given the upregulation of the initiation factors and ubiquitin stated in the previous paragraph. The sponge is trying to keep its cell proliferation at bay in order to keep the balance in cell counts and avoid overproduction of cells. Even though ubiquitins stated earlier inhibit apoptosis, apoptosis and cell death protein can be produce through a variety of different pathways, which explain why both aspects are seen after exposure.

The last group of proteins of interest represent oncogenes and tumor suppressor genes (TSGs) family. These two types of genes play major roles in cancer. Oncogenes are the mutated version of proto-oncogenes. Proto-oncogenes under normal conditions help the cells grow and

develop. However, when mutated and turned into oncogenes (too many copies or permanently turned on), cells grow out of control which leads to cancer. Such genes include BCL2, MYCL1, TFG and many others.

On the other hand, TSGs slow down cell division, repair DNA, or even induce apoptosis. By mutation of these, cell division and proliferation can get out of control, leading to cancer. Some TSGs include TP53, BRCA1/2, NOTCH1 and others. An important difference between oncogenes and TSGs is that oncogenes result from turning on proto-oncogenes (activation), but TSGs cause cancer when they are turned off (inactivated). A good balance between these two types of genes is necessary for the organism survival. (American Cancer Society, cancer.org). Here again, both types of gene were found upregulated, showing that the organisms is trying to keep a balanced state in cell proliferation. Moreover, the finding of these specific genes, which are also found in humans and other species, could hint towards the development of sponges as new lab model for cancer research.

Oil (WAF) Effect

Major genetic responses were seen after 1 hour of exposure to oil but seemed to diminish slowly over the course of 24 hours. The changes were not paralleled in the tissue images at 1 hour, showing no changes (Fig 8).

Differentially expressed transcripts in 1 hour oil treated samples included many genes from the 4 categories of interest. The presence of 20 differentially expressed HSPs (4 HSPs related to the sponge host and 16 to its symbiotic community) revealed the potential negative effect of oil on the conformation of proteins. The 30 E3 ubiquitin ligases, 87 oncogenes and TSGs, and 7 apoptosis related proteins differentially expressed are the witness that exposure to oil is affecting the survival of the sponge as whole, even after 1 hour of exposure.

Furthermore, a well-known protein found in other oil related studies (Han et al, 2014; Jung et al., 2017) Cytochrome P450, was seen differentially expressed confirming the negative impact of oil on the sponge tissues and gene expression.

The low number of differentially expressed transcripts within the 24 hours exposed samples was expected given the similarity between 24 hours control and 24 hours oil treated samples.

Unfortunately, no transcripts related to the gene of interest previously cited were found differentially expressed, but other transcripts linked in some way to these were observed. Moreover, the low number of differentially expressed transcripts in 24 hours treated samples can be partially explained by DESeq's way of assessing differential expression. DESeq2 will not account for transcripts if the variability between the replicates of the same treatments is too far off the fitted models of gene dispersion (Appendix 1). This resulted in a large portion of transcripts being discarded.

Overall it seemed that the oil would have an important impact on the gene expression of both the sponge host and its microbial and symbiotic community after 1 hour of exposure. After 24 hours of exposure, the effect of oil exposure seemed to be greatly decreased. This could be explained by the tank and the mixture reaching an equilibrium of concentrations in PAHs between the air layer and the water column. Also, after 24 hours of exposure, the non-miscible fraction of the oil ended up floating on the surface, not affecting the sponge as sponges are benthic organisms. Morever, given the sublethal dosing, after reaching equilibrium, the content of chemicals dissolved in the water seemed to not be enough to trigger a genetic response.

Dispersant Effect

Detrimental effects of dispersant seemed to be increasing with longer exposure, as observed by microscopy. However, there was very few transcripts differentially expressed in 1 hour treated sample. This was mainly explained by the way DESeq2 assess significant expression. More than half of the transcripts in the transcript pool were discarded due to too much variability between replicates. A big number of 24 hours treated samples were also discarded for the same reason, however to a lesser extent. General trend after 24 hours showed more downregulation. While no genes from the previously cited categories of interest were found differentially expressed, other genes such as filamin B, cathepsin L, protocadherin FAT4, or RGPGEF1, which have an impact on the well-being of the sponge or linked to the genes of interest were found differentially expressed.

Comparisons between 1 hour dispersant treated and 24 hours dispersant treated samples showed major differential expression, with a rather balance between up-regulated and down-regulated transcripts. This would suggest the longer sponges are exposed to dispersant, the more its gene expression is affected.

Oil:Dispersant (CE-WAF) effect

Differential expression assessment by DESeq in the 1 hour CE-WAF treated samples faced the same issues as 1 hour dispersant treated samples cited above. A big portion of transcripts were discarded by the variability factor DESeq accounts for. This could potentially also imply that chemical dispersant worked properly on the oil and limited its effect on the sponge genetic expression after short exposure. Although no transcripts related to genes of interest were found, within the 31 differentially expressed transcripts, catalase and tubulin were found up regulated.

CE-WAF treated samples followed a similar trend as dispersant treated samples with a more detrimental effect after longer exposure, with a trend of more transcripts being downregulated after 24 hours of exposure. No transcripts coding for proteins of the categories of interest were found. However, 3 P450 proteins, 1 galectin, tubulin, or TGFBR1 were found differentially expressed. All these genes are somewhat connected to the genes of interest previously cited, which could imply that even though the dispersant fulfilled its role during the first hour of exposure, after 24 hours the dispersant effect might be fading away progressively, letting the oil effect take back over.

Comparisons between 1 hour exposed and 24 hours exposed CE-WAF samples also showed many differentially expressed transcripts with a trend leaning towards a slight majority of up-regulated transcripts. This would suggest that the longer sponges are exposed to CE-WAF the more its gene expression is affected. This is also following the trend of dispersant treated samples. Such expression profile confirms the efficacity of dispersant to break down oil into smaller particles to allow mixing in the water column, and thus affect benthic organisms that would not be affected in the absence of chemical dispersant.

Overall Conclusions

The results of this study indicated that along with the host genes, several bacterial, fungal, and protists genes were differentially regulated by the experimental dosing which confirmed the findings of Kleindienst et al., 2015. Although some of the original hypotheses of the original FIO grant posited a change in microbial symbionts after oil exposure, Cuvelier at al. 2014 showed no significant changes in bacterial communities after exposure. However, bacterial taxonomy is beyond the scope of this thesis.

Besides oil treated samples which showed a reduced effect after longer exposure, the rest of the experiments showed an increase in differential expression over time. Dispersant and CE-WAF treated samples both revealed more transcripts significantly differentially expressed after 24 hours of exposure compared to 1 hour exposed samples., with a domination of down-regulation after 24 hours of exposure. After 1 hour of exposure, only 6 transcripts and 31 transcript were differentially over-expressed for the dispersant and oil:dispersant treatment respectively. Twenty four hours of exposure revealed an increase in differential expression with 308 significantly differentially expressed transcripts for the dispersant treatment and 4,248 significantly differentially expressed transcripts for the oil:dispersant treatment. While after 1 hour of exposure no transcript was found downregulated, after 24 hours most significantly differentially expressed transcripts were found down regulated (D24=258 and OD24=4089).

Overall, even if in most cases, specific genes of interest were not directly found, other transcripts linked to these proteins of interests were still observed. As well, among the differentially expressed transcripts, galectin, Rab/Ras related proteins, HSPs, and other transcripts in common with Smith (2013) were found confirming the previous findings, this time with more replicate data. In addition, the results reveal a large number of "orphan" uncharacterized sponge genes (and putative protein products) with no previously known function (7455 in total; O1= 4707, O24=209, D1=5, D24=198, OD1=18, OD24=2318). This study provides a starting point for new functional analyses.

It seemed that within 24 hours of exposure, exposure to only dispersant does not generate a big genetic response on the sponge and that more exposure time would cause an even more quantifiable response in gene expression. Additionally, the dispersant role in the CE-WAF

treated samples seems to be effective and reduce the effect of oil during the early stages of exposure. However, after longer exposure times, chemically dispersed oil, seems to be taking over and gene expression seemed to be affected again in a similar manner as when exposed to oil only after 1 hour, even though very few transcripts were found in common between both these treatments. This could imply that dispersant effect might be a good short term counter measure to oil spills but new methods would be better for long term responses.

Overall, the results presented here concerning the effect of oil spill chemicals on the gene expression of this Cinachyrella marine sponge followed previous findings from the literature. Sponges exposed to oil only displayed important differential gene expression after 1 hour of exposure which can be explained by the fact that the miscible part of the oil is still very much present in the water column. However after 24 hours, the miscible portion of oil is not present in big enough quantities to trigger as much of a genetic response.

Dispersant exposure revealed very limited effect on short term exposure with an increase in harmfulness over time.

Chemically dispersed oil seemed to be more harmful to the sponges after longer exposure. Furthermore, chemically dispersed oil seemed to be more harmful to the sponge than oil and dispersant alone after 24 hours.

Issues Faced During This Project

However successful, this project encountered a certain number of issues. One of the first issue this project had to face, is the taxonomic assignments of the sponges. Three species of Cinachyrella are mainly found on Florida reefs, in the Caribbean islands, and off the coast of Brazil: Cinachyrella alloclada, Cinachyrella apion, and Cinachyrella kuekenthali. However, these three species are phenotypically impossible to distinguish, and distinction must be made at the genotypic level. New studies have shown the potential for new markers to help taxonomic assignment of phenotypically undistinguishable sponges (Schuster et al., 2017; Belinky et al., 2012; Park and Min, 2007;Yang et al., 2017; Rua et al., 2011), but unfortunately for this project limited funding didn't allow the taxonomic differentiation. It is important to denote that the study by (Schuster et al., 2017) used samples present in this study and have revealed that 9 of the 24 samples analyzed in this study are in fact C. alloclada. Without being able to conclude about the

effect of oil (WAF), dispersant, and oil:dispersant (CE-WAF) mixtures at the species level for Cinachyrella, we are however able to conclude about the effect of such chemical on poriferan of the Caribbean area for this genus.

The main issue with this project was mostly related to the availability of poriferan genomes in the literature and databases. Indeed, only one poriferan genome is released up to date. Amphimedon queenslandica, a sponge found on Australian reefs, had its genome sequenced by the Degnan Lab in Australia, and the latest release of May 2015 (Aqu2.1) consists of a scaffold assembly with 13,397 scaffolds, and the gene annotation only accounts 9,468 scaffolds of the assembly. The gene annotation has identified 43,615 mono-transcript protein-coding genes, of which only 17,857 have an annotated 3'UTR.

For 3' end sequencing it is vital to have a well annotated genome to align to, and however the immense work that has been done by the Degnan lab on that species, the genome is not fully ready for 3' sequencing applications. The low number of 3' UTR annotated transcripts limits the successful alignment rate of 3' end generated libraries which ultimately results in inaccurate gene quantification. This is the reason why the transcriptome of Cinachyrella was sequenced, assembled, and annotated for this project.

However, it is very important to remember that A. queenslandica genome was the very first sponge genome to ever be sequenced within the Porifera taxa, and even to this day, is still the best genome available for poriferan genomics studies. With this aspect in mind, annotation of the Cinachyrella transcriptome is still a challenge, given the limited basal resources available to go off of.

Some issues also arose through the analysis pipelines used to analyze the data. Choosing the right programs and the right settings for these was crucial to allow the data to speak. When a well annotated genome is available, pipeline made of splice aware alignment programs and quantification software based on splice aware alignments can be used. But in the case the available resources are not up to par, other options need to be considered. This is why for this project a different pipeline was used rather than the one provided by Lexogen and Bluebee.

Even though a different pipeline was used to than the one offered by Lexogen, our pipeline had some limitations as well. In our pipeline, RSEM was used to quantify transcripts because of its ability to account for different isoforms. However, in order for RSEM to perform properly, it requires very stringent alignments with a very low error rate. This means that when

generating alignment file with bowtie2, some transcript will not be accounted for. With another counting software, less stringent alignment can be used but isoforms are not accounted for and thus quantification might be a little less accurate. This is all about finding the right balance.

Another important point worth noticing is how DESeq2 assess differential expression. In order to consider gene for accurate expression DESeq2 requires a minimum presence in comparison to absence of counts across the replicates. If there are counts present in all replicates the transcript is considered for differential expression analysis. If however, in some replicates the number of transcript is equal to 0, DESeq will consider the transcript as a false positive and will return a non-significant p-value. This will also be affected if the number of transcript counts are too low as DESeq will consider this as a low count representative. This is why in some treatment such as D1 or OD1 a big majority of the transcript had a log2 fold change passing the threshold, but the adjusted p-value was not meeting the passing standards. For example, in the D1 treatment, the transcript with accession ID AAA29119, which code for alpha-collagen had a log2 fold change of 2.46 but an adjusted p-value of 0.45. The non-significant p-value is the result of this transcript missing counts in one or more replicates and having low counts in general.

Added to the DESeq2 differential expression assessment issue, it has been shown by Cuvelier et al, 2014 that sponge seem to have different microbial communities depending on the time of the year. This would add some variability between the different replicates of the same sample as they were collected at three different times of the year. That could explain why a lot of the transcripts were discarded across all treatments due to the too high variability.

This project, however, put the spotlight on a bigger issue that is common in the genomics field. Some taxa have very limited data available and projects like this one highlights the gap of knowledge there is on given eukaryote groups, but more generally on species of lesser interest. With the decreasing cost of next generation sequencing and the rise of qualified people to analyze the tremendous amount of data produced, one can only hope these gaps fill quickly.

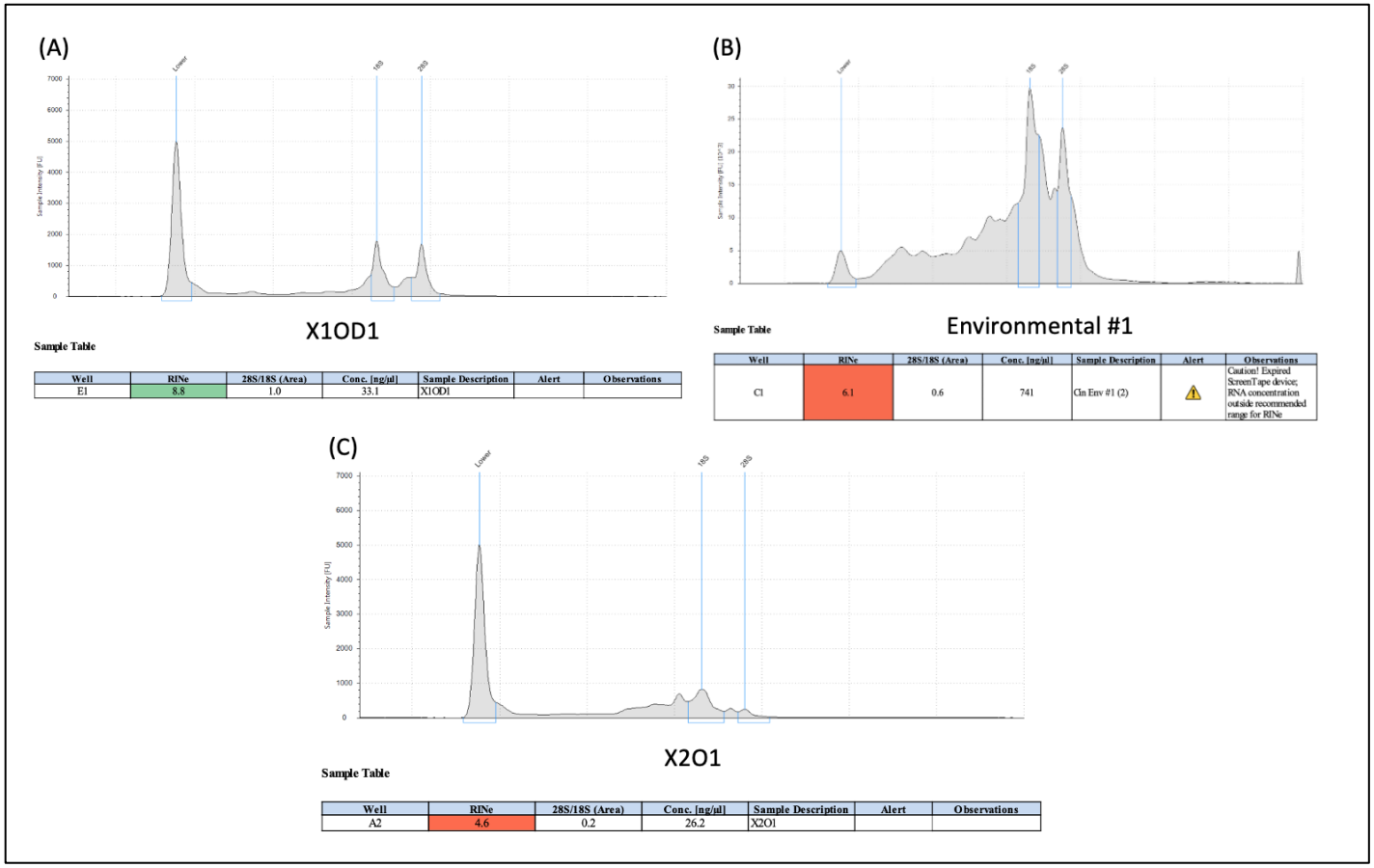
With the newly assembled transcriptome from this study, more in depth analysis of the data should be performed. As well, with the availability of this data, new opportunities will arise such as helping to sequence and further annotate the genome of other poriferans, provide more robust annotation of known transcripts, help identify the function of novel transcripts, both in toxin exposure studies and the discovery of new transcripts in general, which all have shown to be very challenging within the Porifera taxa.

Finally, one of the main goals of this project was to potentially establish the sponge model as a water quality and bio-indicator species for Florida reefs. As a result, sponges have shown great potential to be used for such purposes and can surely be considered as bio-indicators for water quality on south Florida reefs, as changes can be witness both at the genotypic and phenotypic scale.

Overall, this project has enhanced the knowledge of genomics within the Porifera taxa. These results will be the leading point of other studies such as sequencing of the genome from other Cinachyrella sponges or other poriferan by helping with the annotation and the assembly. As well, the present transcriptome has not been fully analyzed. This leaves the opportunity for future work on this aspect of the project.

APPENDICES

APPENDIX 1: Miscellaneous figures & tables



Well	RINe	28S/18S (Area)	Conc. [ng/µl]	Sample Description	Alert	Observations
E1	8.8	1.0	33.1	X1OD1		

Well	RINe	28S/18S (Area)	Conc. [ng/µl]	Sample Description	Alert	Observations
C1	6.1	0.6	741	Cln Env #1 (2)	⚠	Caution! Expired ScreenTape device; RNA concentration outside recommended range for RINe

Well	RINe	28S/18S (Area)	Conc. [ng/µl]	Sample Description	Alert	Observations
A2	4.6	0.2	26.2	X2O1		

Figure 24: (A) Example of a good quality RNA after extraction (RIN=8.8). Both the 18S and 28S peaks are clearly visible and between the lower peak and the two peaks very few to almost no peaks are seen. (B) example of a medium quality RNA after extraction (RIN=6.1). Both the 18S and 28S peak are visible but the lower marker peak is small and there is some contamination between the lower peak and the 18S peak. This would indicate potential phenol/salt contamination from the extraction protocol. In this case, the sample was re-purified using ammonium acetate and the quality was checked again. (C) Example of a bad quality RNA (RIN=4.6).In this case the lower peak is present but both the 18S and 28S peaks are unnoticeable. In this case the RNA from this sample was re-extracted.

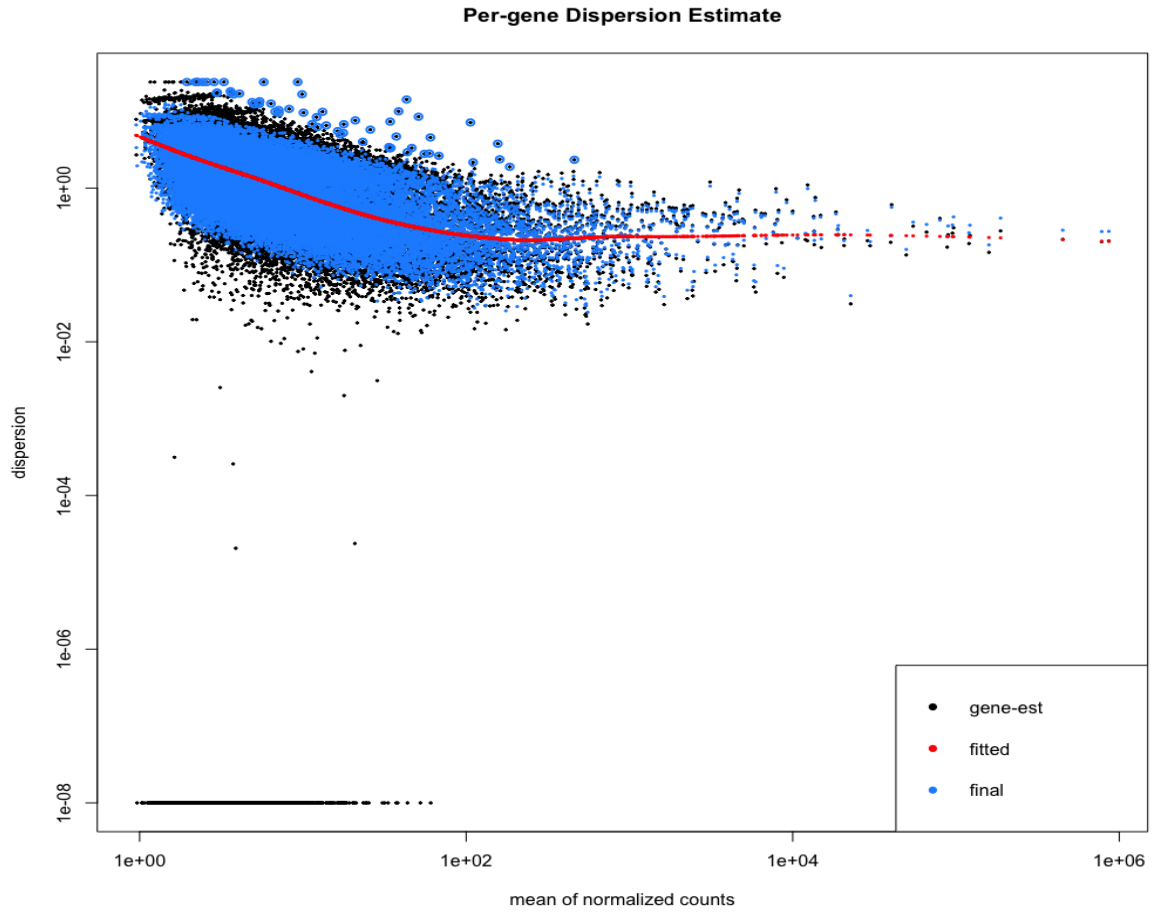


Figure 25: Model of gene dispersion estimate fitted by DESeq2 after applying the « local » fitType setting and the « Wald » test setting. Each black dot represents a single transcript of the 31,571 transcripts pool with their dispersion. Each blue dot represents a transcript after being shrunk to best fit the model. The red line represents the fitted model. All black dots at the bottom left corner of the graph are transcripts considered as outliers and are thus not shrunk towards the fitted value.

Table 10: Common transcripts found differentially expressed between all 1 hour treatments. Highlighted in yellow is the only transcript found in all three treatments

O1_D1	O1_OD1	D1_OD1
XP_015458757	ABW23216	TRINITY_DN392278_c5_g4
TRINITY_DN382666_c0_g5	AGO88073	
TRINITY_DN389181_c0_g2	XP_003389936	
TRINITY_DN392278_c5_g4	XP_009065705	
TRINITY_DN395491_c4_g4	XP_011408643	
	YP_009326829	
	TRINITY_DN372845_c5_g5	
	TRINITY_DN392278_c5_g4	
	TRINITY_DN398117_c2_g4	

Table 11: Transcripts found differentially expressed between the different treatments after 24 hours of exposure

O24_D24			O24_OD24					D24_OD24			
AAN44146	EHN95545	WP_068784382	AAA28321	CDY45547	EPR08335	OUI31666	TRINITY_DN383165_c2_g6				
ABI99685	EHY31642	XP_001617809	AAB66300	CDZ90521	EQE91113	OUJ35337	TRINITY_DN383173_c5_g1				
ABM53544	ELC06673	XP_001622475	AAN44146	CEJ83743	ETI86646	OUJ35732	TRINITY_DN384023_c12_g1				
ADI17560	ELY20072	XP_003088200	ABI99685	CEM31478	ETJ02548	OUP63279	TRINITY_DN384036_c2_g1	AAN44146	EGW50004	OSH35232	TRINITY_DN383119_c4_g1

ADI18655	EMF82881	XP_012694912	ABX13097	CNU32488	EXH74083	OUQ08062	TRINITY_DN384168_c4_g1	AAU94673	EGX75147	OUG43419	TRINITY_DN383165_c2_g6
ADI19635	ENL04231	XP_014662098	ABZ84886	CRY95954	EZX15970	OWF36832	TRINITY_DN384359_c10_g1	ABI99685	EHM94433	OUI31666	TRINITY_DN383488_c0_g13
AGT99284	EOB10733	XP_021349078	ABZ84905	CSD41531	GAD81568	OYS46957	TRINITY_DN386145_c1_g1	ABW23216	EHN95545	OUJ35337	TRINITY_DN384023_c12_g1
ALA57789	EQE91113	TRINITY_DN240164_c0_g1	ACD87749	CUN25727	GAN11837	OZG59192	TRINITY_DN386405_c23_g1	ADI17560	EHY31642	OUJ35732	TRINITY_DN384168_c4_g1
AOE14041	ETJ14928	TRINITY_DN256346_c0_g2	ADI17495	CUQ13639	GAN85277	PAV27562	TRINITY_DN386528_c0_g4	ADI18655	ELC06673	OWF36832	TRINITY_DN385912_c5_g1
BAO86772	EXH74083	TRINITY_DN370801_c0_g1	ADI17560	CUW39321	GAQ68251	SBT22684	TRINITY_DN386718_c7_g2	ADI19635	ELU02242	OZG59192	TRINITY_DN386528_c0_g4
CAJ55202	EZX15970	TRINITY_DN376216_c8_g1	ADI17561	CUW99006	GAW32655	SCI55789	TRINITY_DN386718_c8_g6	AGT99284	ELY20070	SCH95916	TRINITY_DN386718_c7_g2
CCG06582	GAD81568	TRINITY_DN377001_c0_g4	ADI17934	CYD71224	KFJ04251	SHT70305	TRINITY_DN387257_c11_g1	ALA57789	ELY20072	SHT70305	TRINITY_DN386718_c8_g6
CCH69100	GAN11837	TRINITY_DN379320_c2_g1	ADI18104	EDM97734	KFZ42566	WP_027877613	TRINITY_DN387257_c13_g2	AOE10326	EMF82881	WP_027877613	TRINITY_DN387140_c5_g4
CCH85655	GAN85277	TRINITY_DN380082_c7_g1	ADI18106	EDN76677	KMS64973	WP_041003236	TRINITY_DN387257_c16_g1	AOE14041	ENL04231	WP_041003236	TRINITY_DN387257_c11_g1
CCI74358	GAQ68251	TRINITY_DN381628_c1_g5	ADI18655	EDN76706	KMS93274	WP_048070573	TRINITY_DN387669_c0_g4	BAO86772	EOB10733	WP_048070573	TRINITY_DN387407_c0_g2
CDB39267	GAW32655	TRINITY_DN383165_c2_g6	ADI19586	EDO56336	KMV77917	WP_053216981	TRINITY_DN388992_c5_g1	CAJ55202	EQE91113	WP_053216981	TRINITY_DN387669_c0_g4
CDN41084	KFJ04251	TRINITY_DN384023_c12_g1	ADI19635	EDS76179	KMW71018	WP_053955528	TRINITY_DN389601_c1_g3	CCH69100	EXH74083	WP_068784382	TRINITY_DN387809_c1_g1

CDS17303	KFZ42566	TRINITY_DN384168_c4_g1	AGO87862	EEF58946	KMW71019	WP_068784382	TRINITY_DN389602_c0_g4	CCH85655	EZX15970	WP_086437560	TRINITY_DN388992_c5_g1
CDY19671	KMS64973	TRINITY_DN386528_c0_g4	AGO87878	EEG43412	KNA06142	WP_092499977	TRINITY_DN389838_c4_g1	CCI74358	GAD81568	XP_001617809	TRINITY_DN389602_c0_g4
CDY45547	KMS93274	TRINITY_DN386718_c7_g2	AGO88073	EEP52383	KPY16424	XP_001617809	TRINITY_DN390473_c0_g1	CDB39267	GAN11837	XP_003088200	TRINITY_DN389838_c4_g1
CDZ90521	KMV77917	TRINITY_DN386718_c8_g6	AGT99284	EES47791	KPY31326	XP_002739625	TRINITY_DN391038_c10_g2	CDN41084	GAN85277	XP_009065705	TRINITY_DN390289_c2_g1
CEJ83743	KMW71018	TRINITY_DN387257_c11_g1	AID23620	EEX24764	KRX85891	XP_003088200	TRINITY_DN391739_c0_g2	CDS17303	GAQ68251	XP_011409706	TRINITY_DN390473_c0_g1
CEM31478	KPY31326	TRINITY_DN387669_c0_g4	ALA57789	EFB40215	KTC66578	XP_004997448	TRINITY_DN392137_c2_g1	CDY45547	GAW32655	XP_012694912	TRINITY_DN391038_c10_g2
CNU32488	KTC95109	TRINITY_DN388992_c5_g1	ALG05280	EFE67883	KTC95109	XP_005845999	TRINITY_DN392184_c1_g3	CDZ90521	KCX19713	XP_014662098	TRINITY_DN391739_c0_g2
CRH31484	KUK19638	TRINITY_DN389602_c0_g4	AOE06246	EFF65279	KTD49695	XP_012694912	TRINITY_DN394578_c0_g6	CEJ83743	KFJ04251	XP_019862850	TRINITY_DN392184_c1_g3
CSD41531	KWW25567	TRINITY_DN389838_c4_g1	AOE11098	EFG00477	KUJ95333	XP_014662098	TRINITY_DN394631_c3_g5	CEM31478	KFZ42566	XP_021349078	TRINITY_DN394631_c3_g5
CYD71224	OCA56495	TRINITY_DN390473_c0_g1	AOE14041	EFG20046	KUK19638	XP_018009702	TRINITY_DN396107_c3_g2	CNU32488	KMS64973	XP_022100877	TRINITY_DN396107_c3_g2
EDM97734	ODQ68767	TRINITY_DN391038_c10_g2	ATD07737	EFL15838	KWW25567	XP_021349078	TRINITY_DN396544_c1_g3	CSD41531	KMS93274	TRINITY_DN240164_c0_g1	TRINITY_DN396544_c1_g3
EEF58946	OLC82878	TRINITY_DN391739_c0_g2	BAG46934	EFV00405	KZN08616	TRINITY_DN240164_c0_g1	TRINITY_DN396576_c0_g3	CYD71224	KMV77917	TRINITY_DN256346_c0_g2	TRINITY_DN396576_c0_g3
EEG43412	OSH35232	TRINITY_DN392184_c1_g3	BAH73833	EFV78732	OCA56495	TRINITY_DN256346_c0_g2	TRINITY_DN396576_c0_g4	EDM97734	KMW71018	TRINITY_DN368615_c12_g1	TRINITY_DN396576_c0_g4

EEP52383	OUG43419	TRINITY_DN394631_c3_g5	BAO86772	EGW50004	ODQ68767	TRINITY_DN340513_c0_g1	TRINITY_DN398358_c0_g2	EEF58946	KPY31326	TRINITY_DN376216_c8_g1	TRINITY_DN397081_c0_g1
EEX24764	OUI31666	TRINITY_DN396107_c3_g2	CAJ55202	EGX75147	OEU06062	TRINITY_DN376216_c8_g1	TRINITY_DN399353_c1_g1	EEG43412	KTC95109	TRINITY_DN377001_c0_g4	TRINITY_DN397224_c3_g5
EEY31938	OUJ35337	TRINITY_DN396544_c1_g3	CAR86202	EHM94433	OGL67088	TRINITY_DN376957_c13_g5	TRINITY_DN399762_c1_g4	EEP52383	KUK19638	TRINITY_DN378801_c7_g1	TRINITY_DN398358_c0_g2
EFB40215	OUJ35732	TRINITY_DN396576_c0_g3	CBX28568	EHN95545	OKY54895	TRINITY_DN377001_c0_g4	TRINITY_DN399871_c5_g7	EEX24764	KWW25567	TRINITY_DN379320_c2_g1	TRINITY_DN399762_c1_g4
EFE67883	OWF36832	TRINITY_DN396576_c0_g4	CCG06616	EHS56317	OLB14314	TRINITY_DN379249_c7_g2	TRINITY_DN400209_c11_g1	EFB40215	OCA56495	TRINITY_DN380082_c7_g1	TRINITY_DN434984_c0_g1
EFG00477	OZG59192	TRINITY_DN398358_c0_g2	CCH69100	EHY31642	OLC82878	TRINITY_DN379320_c2_g1	TRINITY_DN401269_c56_g1	EFE67883	ODQ68767	TRINITY_DN380481_c2_g2	TRINITY_DN546245_c0_g1
EFV00405	SHT70305	TRINITY_DN399762_c1_g4	CCH85655	EJK58442	ORC94772	TRINITY_DN380077_c8_g1	TRINITY_DN434984_c0_g1	EFG00477	OLC62184	TRINITY_DN381628_c1_g5	TRINITY_DN557693_c0_g1
EFV78732	WP_027877613	TRINITY_DN434984_c0_g1	CCI74358	ELC06673	ORD05465	TRINITY_DN380082_c7_g1	TRINITY_DN546245_c0_g1	EFV00405	OLC82878	TRINITY_DN382750_c7_g3	TRINITY_DN688382_c0_g1
EGW50004	WP_041003236	TRINITY_DN546245_c0_g1	CDB39267	ELY20072	OSH35232	TRINITY_DN380476_c5_g1	TRINITY_DN557693_c0_g1	EFV78732	OLD62857	TRINITY_DN382867_c6_g7	TRINITY_DN75966_c0_g1
EGW65230	WP_048070573	TRINITY_DN557693_c0_g1	CDN41084	EMF82881	OSX56030	TRINITY_DN380850_c6_g1	TRINITY_DN688382_c0_g1				
EGX75147	WP_048110542	TRINITY_DN688382_c0_g1	CDN41086	ENL04231	OUD48660	TRINITY_DN381628_c1_g5	TRINITY_DN75966_c0_g1				
EHM94433	WP_053216981	TRINITY_DN75966_c0_g1	CDS17303	EOB10733	OUG43419	TRINITY_DN382002_c0_g1	TRINITY_DN943_c0_g1				

Table 12: Transcripts found differentially expressed in all 24 hours exposed samples

Common Genes at 24 hours								
AAN44146	CCI74358	EEG43412	ELC06673	KFJ04251	OSH35232	XP_001617809	TRINITY_DN384023_c12_g1	TRINITY_DN392184_c1_g3
ABI99685	CDB39267	EEP52383	ELY20072	KFZ42566	OUG43419	XP_003088200	TRINITY_DN384168_c4_g1	TRINITY_DN394631_c3_g5
ADI17560	CDN41084	EEX24764	EMF82881	KMS64973	OUI31666	XP_012694912	TRINITY_DN386528_c0_g4	TRINITY_DN396107_c3_g2
ADI18655	CDS17303	EFB40215	ENL04231	KMS93274	OUJ35337	XP_014662098	TRINITY_DN386718_c7_g2	TRINITY_DN396544_c1_g3
ADI19635	CDY45547	EFE67883	EOB10733	KMV77917	OUJ35732	XP_021349078	TRINITY_DN386718_c8_g6	TRINITY_DN396576_c0_g3
AGT99284	CDZ90521	EFG00477	EQE91113	KMW71018	OWF36832	TRINITY_DN240164_c0_g1	TRINITY_DN387257_c11_g1	TRINITY_DN396576_c0_g4
ALA57789	CEJ83743	EFV00405	EXH74083	KPY31326	OZG59192	TRINITY_DN256346_c0_g2	TRINITY_DN387669_c0_g4	TRINITY_DN398358_c0_g2
AOE14041	CEM31478	EFV78732	EZX15970	KTC95109	SHT70305	TRINITY_DN376216_c8_g1	TRINITY_DN388992_c5_g1	TRINITY_DN399762_c1_g4
BAO86772	CNU32488	EGW50004	GAD81568	KUK19638	WP_027877613	TRINITY_DN377001_c0_g4	TRINITY_DN389602_c0_g4	TRINITY_DN434984_c0_g1
CAJ55202	CSD41531	EGX75147	GAN11837	KWW25567	WP_041003236	TRINITY_DN379320_c2_g1	TRINITY_DN389838_c4_g1	TRINITY_DN546245_c0_g1
CCH69100	CYD71224	EHM94433	GAN85277	OCA56495	WP_048070573	TRINITY_DN380082_c7_g1	TRINITY_DN390473_c0_g1	TRINITY_DN557693_c0_g1
CCH85655	EDM97734	EHN95545	GAQ68251	ODQ68767	WP_053216981	TRINITY_DN381628_c1_g5	TRINITY_DN391038_c10_g2	TRINITY_DN688382_c0_g1
	EEF58946	EHY31642	GAW32655	OLC82878	WP_068784382	TRINITY_DN383165_c2_g6	TRINITY_DN391739_c0_g2	TRINITY_DN75966_c0_g1

APPENDIX 2: MS DRAFT

Physiological and Genetic Effects of Deepwater Horizon Oil and Dispersant on Marine Sponge Cinachyrella sp.

Yvain Desplat*[1], Jacob Warner[2], Nidhi Vijayan[1], Emily Smith[1], Patricia Blackwelder[1], Jose V. Lopez[1]

[1] Nova Southeastern University, Florida

[2] University of North Carolina at Wilmington; warnerj@uncw.edu

* yvain.desplat@gmail.com

ABSTRACT

Marine sponges have been shown to be regulators of reef ecosystems by fulfilling many ecological functions. However, little is known about sponge behavior in the face of sudden environmental changes. The Deepwater Horizon oil spill in 2010 represents the largest environmental accident in US waters, Consequently, we developed the Caribbean reef sponge Cinachyrella spp. as a novel experimental indicator.

The experimental design exposed (N=24) individual sponges to sublethal amounts (0.5 ppm) of oil (water accommodated fractions-WAF), oil mixed with 10% Corexit 9500 dispersant (OD; or chemically enhanced-CE WAF), and dispersant (D) from 1 to 24 hours. Three replicate oil dosing experiments (labeled as X1 – X3) were performed to characterize the physiological and genetic effects of Cinachyrella spp. Microscopy revealed oil droplets in the mesohyl and observable changes only after 24 hours. cDNA library construction and differential gene expression analyses indicate 31571 verified sponge transcripts eligible for genetic profiling. A total of 12,913 transcripts have shown significant differential expression, among which 7,863 were upregulated and 5,058 were down regulated. After OD exposures, P450 proteins, galectin, tubulin, and TGFBR1 were found differentially expressed. With longer oil or dispersant exposures, down-regulation was dominant over upregulation. Strength of impact on gene expression followed this order : Oil:Dispersant > Dispersant > Oil. This study supports development of Cinachyrella as a suitable model organism from Florida reefs.

1. Introduction

The April 20, 2010 Deepwater Horizon oil platform explosion and oil spill (DWHOS) well rupture in the Gulf of Mexico represents one of the most lethal, human-mediated environmental accidents in US history. After ten years, the effects of the DWHOS have been well documented (REFS; Fisher et al, 2016; Joye et al, 2016; Milligan et al, 2019; Johnston et al, 2019). Previous work has shown that hydrocarbons, other components of crude oil, and their degradation products act as physiological stress inducers and toxicants to marine animals and microbial populations (Harwell and Gentile, 2006; Yeats et al, 2008; more recent REFS). Using the advances of high throughput molecular methods such as next generation sequencing, researchers can identify and catalogue a greater number of physiological and genetic signatures that mark marine organismal "stress."

The central aim of this study was to develop and characterize "sentinel" sponge species and their associated microbiota along with advanced molecular and genomic tools to assess the impact of oil contamination on Western Florida shelf reefs. Marine sponges and their associated microbes are excellent environmental sentinels. Using high throughput transcriptomics and metagenomics we can simultaneously trace the direct impact of crude oil (and byproducts) and dispersants on both sponge physiology (a general marker of reef health) and microbial community dynamics (a general marker of regional seawater quality). Evaluating shifts in the functional differential gene expression (DGE) of marine organisms will allow us to better understand…. the overall effects of hydrocarbon loading in the water column that have resulted from the Deepwater Horizon oil spill.

2. Methods

Sponges Collection and Culturing

Nova Southeastern University Oceanographic Center (NSUOC) SCUBA collection team sampled 29 sponges from the Hollywood's first reef, Florida, USA (coordinates: 26.051425 N; 80.112141W). All sponges were collected under a standard Florida fishing license (issued by the FL Fish and Wild Commission – myfwc.com). Ambient seawater samples were also collected at the same sponge samples sites. Live sponges were brought back to the lab and acclimated in temperature-controlled integrated closed circulating aquaria culture tanks fabricated by

AquaLogic, Inc. (California). at Florida International University. The taxonomy of the sponges was determined by spicule preparations and ultimately via the presence of a Group I mtDNA intron following the guidelines of Schuster et al (2017).

Dosing solution preparation and sponge exposure

Working solution for dosing experiments were prepared following CROSERF guidelines ("Chemical Response to Oil Spills: Ecological Effects Research Forum"), a conclusive report to standardize analytical laboratory procedures in testing the toxicity and environmental effects of dispersants and dispersed oil in oil spill response. It states that dispersed oil solutions were tested using an oil:dispersant ratio of 10:1 (Aurand & Coelho, 2005); this guideline was followed in preparation of the CE-WAF, and the representative dispersant volume was kept constant in preparation of the dispersant-only solution.

We obtained crude oil directly from BP (SOB-20100622-084; SOB-20100624-00) and Corexit 9500 from Nalco Holding Company. Three replicate oil dosing experiments (labeled as X1 – X3) were performed on a total of 24 sponges based on standard CROSERF protocols. (Figure *** of experimental design)

Approximately 100 L of ambient seawater that had been collected with the sponges, were used for preparing three primary treatments: i) water accommodated fractions (WAF) of oil, ii) chemically enhanced (Corexit 9500) WAFs or CE-WAFs, and iii) Corexit only.

Sponges were removed one at a time for processing and immediately sacrificed as 0 hr, initial time point sample, and later time points t=1 hr, and t=24 hr and processed as described above for each sponge.. Processing included quartering each sponge with a flame-sterilized knife. Three pieces were flash frozen in liquid nitrogen (LN_2) and stored at -80°C for DNA and RNA work and the latter piece saved archived voucher sample. The remaining piece was halved for separate histological processes, transmission and scanning electron microscopy (TEM/SEM) and stored at 4°C in 2% glutaraldehyde and sodium cacodylate buffer

Sponge Histology

Sponge explant samples was embedded in paraffin wax using an automatic apparatus. The apparatus was used in order to immerse the cassettes containing the sponge in two 80% ethanol

solutions, two 95% ethanol solutions, three 100% ethanol solution, three 100% xylene solutions and three paraffin wax solutions for a period of 30 minutes each. The gradual changes of percentages were required to prevent an extreme change in hydrophobicity that would damage the cells. Once the sponge is dehydrated and processed in paraffin, the cassettes were placed in a melted paraffin bath. Samples were taken out of the cassettes and cut at desired locations and placed with the cut side down into a mould, which was filled with melted paraffin.

The paraffin-embedded sponge blocks was sliced into sections with an Accu-Edge low profile microtome blades (Sakura Finetek). Sponge sections for day 2, day 8, day 5 and day 14 were cut to a width of 10µM and placed on a warm water bath where the sections float on the top in order to smooth out the sections and to make it easier to mount. Sections were then floated on top of a glass slide. The slides were placed in an incubator for 12 hours at 37^{O} C. The slides were then dewaxed by placing in xylene and ethanol solutions and air-dried.

Hematoxylin and Eosin Staining

The sections were deparaffinized by washing with xylene three times followed with 100% ethanol. Sections were hydrated with decreasing concentration of ethanol (95%, 80%) and distilled water and then stained with Harris hematoxylin for 2mins and washed with water. Destaining was performed with 0.5% acid alcohol and washed with water. Slides were washed with 0.25% ammonium hydroxide as mordant and then washed and dehydrated with 70% ethanol, followed by staining with eosin for 30 seconds (Avwioro 2011). Slides were then destained with 95% ethanol and dehydrated with 100% ethanol followed by three washed with xylene.

Electron Microscopy

cDNA libraries preparation and sequencing

A total of 24 cDNA libraries were generated using the QuantSeq mRNA 3’ FWD Library Prep kit from Lexogen, following manufacturers protocol. The total RNA input was standardized across samples to 132 ng, which resulted in a total of 14 PCR cycles for amplification of the libraries. The Lexogen cDNA libraries were then sent to the NSU Genomics Core for sequencing. Final

cDNA libraries were qPCR-quantified using KAPA Biosystem's Library Quantification Kit optimized for the Roche LightCycler 480 Instrument II. (put citation)

The 24 RNA sequencing libraries were pooled and normalized to 2 nM and denatured according to Illumina's NextSeq System Denature and Dilute Libraries Guide. Final pooled libraries were spiked with 2% PhiX as an internal control and loaded at a final concentration of 1.6 pM onto the Illumina NextSeq 500 platform. Libraries were sequenced on a 1x150 bp single end run using the Illumina NextSeq 500 Mid Output v2.5 Kit (150 cycles, 130 million read flow cell).

Differential Gene Expression (DGE) Assessment

Generated FASTQ file were put through a custom analysis pipeline on a Jetstream instance. The pipeline consisted of quality reads check using FASTQC then trimmed using cutadapt for any sequencing adapter and polyA tails removal (min. QC = 25, min. length=20 bp). Trimmed reads were then aligned to the reference transcriptome using Bowtie2 with very sensitive settings, count files were generated using RSEM, and DGE was assessed using DESeq2 in R a modified script.

Results

Dosing Solution Prep

Oil is mostly immiscible, but dissolution does occur for a small semi-soluble fraction as a function of the surface area of the oil-water interface. This study did not separate oil into their specific components. Turner and Renegar (2017) provided a comprehensive review of the effects of crude oil toxicants on marine organisms, specifically corals. Solution preparation following CORSERF guidelines revealed to be successful and usable for dosing experiment for sublethal exposures of sponges to oil spill chemicals.

Sponge Histology

In order to study the effects, the treated sponges were stained with hematoxylin-eosin to analyze the choanocyte chambers in the aquiferous system introduced separately with oil, dispersant, and oil plus dispersant. Abnormally stained parts were observed in the sponge treated with Oil (Fig X A&B) when comparing the Control sponge (X1 C0, Fig X) to sponge treated with Oil (X1 O24; Fig X A&B). The stain was ruled out to be cellular as it encompasses different types of cells in the area. The abnormal areas appear to be surrounded by a concentration of

amoebocytes/archaeocytes. This could be due to increase stress and an immune response of the sponge. Also, the inner regions of the oil treated sponge appears to be broken down compared to its upper layers with intact choanocyte chambers.

Sponges treated with only dispersant did not have the abnormally stained structures (Fig X; SI). The choanocyte chambers appear structurally normal in sponges treated with dispersant at 8 hour time point (X2 D8 and X3 D8). Sponges exposed to dispersant only for 24 hours did not have abnormally stained structures either. However, the cellular components of the sponges at 48 hour time point appears to be broken down (X2 D48; Fig X, SI). The sponges in experiment X1 treated with only dispersant were difficult to section due to the high concentration of spicules and the position of the section in the paraffin.

Sponge treated with a combination of oil and dispersant after 24 hour time point (X1 OD24; Fig X, D&E) was not significantly different from sponge treated with oil at 24 hour time point (X1 O24). Abnormally stained regions appear scattered in the aquiferous system (Fig 4). Inner regions appear to be disintegrated with the absence of choanocyte chambers. However, sponge treated with oil and dispersant at 48 hour time point (X2 OD48) did not reveal abnormally stained structured (Fig 5). This could be due to the action of the dispersant on the oil, which allows the breakdown of the oil droplets.

Electronic microscopy of sponge tissues revealed no noticeable differences between control and 1 hour exposed sponges. After short exposure, mesohyl integrity was not compromised and tissues did not seem to have retracted (Fig X, SI). On the contrary after 24 hours of exposure, significant difference was observer between control and treated samples. Mesohyl integrity was compromised, possibly hinting to a retraction of the tissues with potential degradations of the cytoskeleton of the sponge.

Dosage sample grouping

Final count table generated for DESeq2 analysis was put into primer to create a nMDS plot (Fig X). The nMDS plot analysis revealed a very low stress value (r=0.03) which confirmed high confidence with the results in further downstream analyses.

Sequencing and Differential Gene Expression (DGE) Assessment

Illumina NextSeq 500 sequencing resulted in 24 libraries with an average of 7.8 M reads/sample. Reads counts generated by RSEM assesses transcripts count at the isoforms levels, and filtering before DGE assessment resulted in a pool of 31,751 transcripts eligible for genetic profiling. Differential expression was significant when |LFC|>2 & padj<0.05. Overall, 12,913 transcripts have shown significant differential expression, among which 7,863 were upregulated, and 5,058 were down regulated. Table 8 (A, SI) summarizes the number of differentially expressed transcripts across all treatments using the full assembly, and Table 8 (B, SI) shows the top 5 up-regulated and down-regulated transcripts for each treatment at each time point.

One Hour Chemical Exposure

DGE analysis of 1 hour exposed sponges revealed 8,052, 6, and 31 differentially expressed transcripts for oil, dispersant only, and oil:dispersant mixtures treated samples respectively.

After 1 hour of exposure to oil, sponges seemed to have an important genetic reaction with 7,561 up-regulated transcripts and 91 down-regulate transcripts. Maximum up-regulation was reached by an unidentified transcript called TRINITY_DN392278_c5_g4 with a log2foldchange (LFC) of 21. The transcripts called TRINITY_DN386287_c3_g1 had the maximum down-regulation with a LFC of 22. Overall, 1,140 transcripts had a match with A. queenslandica. A total of 1,904 transcripts were identified as uncharacterized protein function, meaning those transcripts have a known ORF, but no known functional annotation, and 2,803 transcript identified as TRINITY_DNXXXX_cX_gX, later referred as TRINITY transcripts. Among the differentially expressed transcripts, several 3 Heat Shock Proteins were detected (HSP60, HSP70), along with 30 E3 ubiquitin ligases, 5 initiation factors, 7 cell death and apoptosis related protein, and 87 oncogenes and suppressor genes (including Rab/Ras, CMYC, and Src). Along with these specific proteins, P450, a protein known to be one of the main metabolic activator of PAHs, was observed in 7 occurrences, always up-regulated. These results indicated a strong gene after short term exposure to oil.

After 1 hour of dispersant exposure, very few transcripts were differentially expressed. Only 6 transcripts were significantly differentially expressed, with all of them being up-regulated. Transcripts identified as 4 TRINITY transcripts, 1 hypothetical protein, and one annotated gene identified as peroxisomal sarcosine oxidase. Dispersant only treatment didn’t seem to have a big impact on the gene expression on the host and its communities after short exposure time.

One hour of exposure to oil:dispersant also triggered a minimal gene response, similar to the dispersant only treatment. A total of 31 transcripts were significantly differentially expressed, 3 transcripts identified as A. queenslandica, 12 were TRINITY transcripts, and 6 as transcripts with uncharacterized protein function. All 31 transcripts were upregulated with a maximum log2 fold change of 18. Catalase (hydrogen peroxide processing) and tubulin (cytoskeleton structural protein) were observed differentially expressed in this treatment. These results indicated that short term exposure to oil:dispersant mixture was mostly not harmful to the sponge host and its communities.

Volcano plots representing the significant differential expression profile for each condition are shown in Figure X, with a Venn diagram summarizing the number of differentially expressed transcripts and the overlap between each treatment after 1 hour of exposure.

Figure composite C1x O1x D1x OD1 volcano plots

Twenty Four Hours Chemical Exposure

DGE analysis of 24 hours exposed samples 268, 308, and 4,248 transcripts differentially expressed when exposed to oil, dispersant only, and oil:dispersant mixtures respectively.

After 24 hours of exposure to oil, sponge gene expression had greatly decreased: 268 transcripts were differentially expressed (48 up-regulated, max LFC=7; 220 down-regulated, max LFC=-20), among which only 1 transcript was identified as A. queenslandica, 66 transcripts were singled out as TRINITY, and 143 transcripts were found to be transcripts with uncharacterized protein function coming from the sponge communities. Interestingly, P450 was not observed at that time point. These results showed that an equilibrium had been reached in the tank between the water accommodated fraction (miscible part of the oil) and the air in the tank after 24 hours of exposure. Moreover, after 24 hours, the non-miscible fraction of the oil had floated back on top of the tank, thus not affecting a benthic organism. Consequently, active PAHs were not present in enough quantities to trigger a similar response to 1 hour treated samples.

Exposure to dispersant for 24 hours resulted in an increase of gene expression. A total of 308 transcripts were differentially expressed (58 up-regulated, and 258 down-regulated) made of 16 sponge related transcripts, 108 TRINITY transcripts, and 90 transcripts with uncharacterized protein function. Maximum down-regulation was observed for the transcript with the ID

TRINITY_DN371367_c0_g1(LFC=-41), while maximum up-regulation was observed for transcript with ID TRINITY_DN384023_c12_g1 (LFC=7). Cathespin L (protein degradation), filamin B (cell membrane and actin cytoskeleton connection), protocadherin FAT4 (tumor suppressor gene), septin (cytokinesis related protein), and RAPGEF1 are some of the transcripts differentially regulated after 24 hours of dispersant exposure. These results highlighted an increase in gene expression over longer exposure, and thus an increase in harmfulness over time.

Exposure to CE-WAF mixture for 24 hours had an impact on 4248 transcripts (159 up-regulated, 4089 down-regulated). In total, 29 transcripts were identified as similar to A. queenslandica, TRINITY transcripts totaled up to 802, and 1516 transcripts had no protein function. Transcript with accession number OGG55450 (hypothetical protein) was the most down regulated transcript (LFC=-33). The most upregulation was reached by the TRINITY transcript TRINITY_DN396576_c0_g3 (LFC=7). P450, tubulin, galectin (anti-tumor agent), calcineurin binding protein cabin 1 (TP53 negative regulator), and TGFBR1 were some of the gene differentially expressed in this treatment. These results confirmed the increased toxicity of oil after being chemically dispersed and its effects on benthic organisms that would be affected under undispersed oil.

Volcano plots representing the significant differential expression profile for each conditions are shown in Figure X, with a Venn diagram summarizing the number of differentially expressed transcripts and the overlap between each treatment after 24 hour of exposure.

Figure composite C24x O24x D24xOD24 volcano plots + venn diagram

3. Discussion

After 1-24 hour treatments of sponges in aquaculture with sublethal amounts of oil and dispersant, all sponges appeared to survive, based on visual observation of active pumping through open oscules. However, we anticipated finer scale changes in gene expression and sponge physiology of sponge appeared, which is why microscopy and molecular genetics methods were applied.

This study looked at the effect of crude oil exposure on the gene expression rather than into the effect of specific components of crude oil. Extensive work has been by Turner and Renegar (2017), who provided a comprehensive review of the effects of crude oil toxicants on marine organisms, specifically corals.

Overall, exposure to oil spill chemicals revealed a strong response by sponge to oil (WAF) exposure after 1 hour of exposure with a dominance of over expressed transcripts (8,052 differentially expressed transcripts), and a decrease in gene expression was seen after 24 hour with a dominance of down-regulated transcripts (268 transcripts differentially expressed), which was most likely due to a chemical concentration equilibrium being reached between the water and the air layer in the tanks. These results would suggest two points. The first point is that crude oil if not dispersed would have very small impacts on benthic organisms. The second conclusion that can be made from this dosing experiment is that in the long run, weathered crude oil will eventually reach a concentration low enough to not be harmful to benthic organisms.

Dispersant exposure revealed that while short term exposure might be safe for benthic organisms, it is a lot more miscible that oil and would eventually mix with the entire water column resulting in a harmful effect to organisms living on the benthos in the long term. Here, 1 hour of dispersant exposure revealed 6 differentially expressed transcripts. After 24 hours of exposure, gene expression was impacted to a higher extent with a total of 308 transcripts differentially expressed and a dominance of down-regulation.

Finally, oil:dispersant (CE-WAF) exposure revealed just like dispersant an increases toxicity over time. After 1 hour of exposure, 36 transcripts were differentially expressed. Twenty four hours after the beginning of the exposure 4,248 transcripts were significantly affected with a dominance of down regulation. Overall, longer exposure time resulted in higher genetic response from the sponge. These results highlighted the working process of chemical dispersant. Chemical dispersant breaks down crude oil into smaller particles to allow them to sink down the water column. With such chemicals in action, benthic organisms would end up being expose and for longer periods of time than surface or mid water column organisms.

Among differentially expressed transcripts, galectin, Rab/Ras related proteins, HSPs, and other transcripts in common with Smith (2013) were found confirming the previous preliminary findings, although this time with more replicate data. In addition, the results reveal a large number of “orphan” uncharacterized sponge genes (and putative protein products) with no previously known function (7455 in total; O1= 4707, O24=209, D1=5, D24=198, OD1=18, OD24=2318). This study provides a starting point for new functional analyses.

Heat shock proteins (HSP), are often proteins produced by many organisms when under a stressful situation. Many proteins part of this group function as chaperone, meaning they act as controllers ensuring the correct folding of other proteins or the refolding of damaged proteins due to the stress the cell has been under. HSPs have various functions in an organism ranging from simple management function of proper protein conformation under non-stressful conditions, to some potential implication in cancer cell deaths apoptosis (Salamanca et al., 2014).They have been found to be part of the chemical defensome of several organisms (Goldstone 2008 ; Goldstone et al., 2006 ; Shinzato et al., 2012) and are found virtually in all living organisms from bacteria to humans. HSPs are named after their molecular weight in kilodaltons. Three HSPs have been mostly studied: Hsp60, Hsp70, and Hsp90, with respective weights of 60 kDa, 70 kDa, and 90kDa.HSPs have been found to be upregulated under stress conditions, which is also the case in this study. The observed HSPs in this dosage experiment study have been shown to be up-regulated after exposure. This proves that no matter which treatment the sponge is exposed to, it is put under stressful conditions. Consequently it is trying to protect its physiological functions by producing the proteins that help achieving that goal.

Another protein category important to the organism survival are the ubiquitin proteins. Three types of ubiquitin exist: E1s, also called ubiquitin activating enzymes, E2s, which are conjugating enzymes, and E3s that are ubiquitin ligases. These proteins are essential in many biological processes which include: endocytic trafficking, inflammation, translation, DNA repair, or apoptosis (Miranda et al., 2007 ; Teixeira and Reed, 2013) Over 600 ubiquitin proteins are found in the human genome, and they are present in all living organisms. In this study mostly E3 ubiquitin have shown differential expression with the big majority being up-regulated. Furthermore, several initiation factors have also been seen up-regulated. The up-regulation of these two types of genes are favoring the cell proliferation by inhibiting apoptosis. This shows that exposure to the chemical has the capacity to change specific steps of the cell cycle and ultimately affect the organism survival

The third category of proteins of interest were those coding for cell death, and apoptosis. These proteins are produced as marker to indicate the non-viable state of cell, and consequently the need to get rid of a non-usable component of the body. Up-regulation of these types of protein indicate an increase in cell damage and thus a decrease in cell viability. Down-regulation of such proteins, on the other hand indicate an uncontrolled cell proliferation, which for example

helps tumors to evade the cell death and promotes drug resistance, a common issue seen in many cancers (Berger and Pu, 2018; lee et al., 2005; Prasad et al., 1997). In this experiment the biggest majority of cell death and apoptosis related proteins were up -regulated. This is expected given the upregulation of the initiation factors and ubiquitin stated in the previous paragraph. The sponge is trying to keep its cell proliferation at bay in order to keep the balance in cell counts and avoid overproduction of cells. Even though ubiquitins stated earlier inhibit apoptosis, apoptosis and cell death protein can be produce through a variety of different pathways, which explain why both aspects are seen after exposure.

The last group of proteins of interest represent oncogenes and tumor suppressor genes (TSGs) family. These two types of genes play major roles in cancer. Oncogenes are the mutated version of proto-oncogenes. Proto-oncogenes under normal conditions help the cells grow and develop. However, when mutated and turned into oncogenes (too many copies or permanently turned on), cells grow out of control which leads to cancer. Such genes include BCL2, MYCL1, TFG and many others.

On the other hand, TSGs slow down cell division, repair DNA, or even induce apoptosis. By mutation of these, cell division and proliferation can get out of control, leading to cancer. Some TSGs include TP53, BRCA1/2, NOTCH1 and others. An important difference between oncogenes and TSGs is that oncogenes result from turning on proto-oncogenes (activation), but TSGs cause cancer when they are turned off (inactivated). A good balance between these two types of genes is necessary for the organism survival. (American Cancer Society, cancer.org). Here again, both types of gene were found upregulated, showing that the organisms is trying to keep a balanced state in cell proliferation. Moreover, the finding of these specific genes, which are also found in humans and other species, could hint towards the development of sponges as new lab model for cancer research.

A major issue with this project was related to gene annotation based on the availability of poriferan genomes in the literature and databases. Indeed, only one poriferan genome was released at the time of analyses: Amphimedon queenslandica, a sponge found on Australian reefs, had its genome sequenced by the Degnan Lab in Australia, and the latest release of May 2015 (Aqu2.1) consists of a scaffold assembly with 13,397 scaffolds, and the gene annotation only accounts 9,468 scaffolds of the assembly. The gene annotation has identified 43,615 mono-transcript protein-coding genes, of which only 17,857 have an annotated 3’UTR.

For 3’ end sequencing it is vital to have a well annotated genome to align to, and however the immense work that has been done by the Degnan lab on that species, the genome is not fully ready for 3’ sequencing applications. The low number of 3’ UTR annotated transcripts limits the successful alignment rate of 3’ end generated libraries which ultimately results in inaccurate gene quantification. This is the reason why the transcriptome of Cinachyrella was sequenced, assembled, and annotated for this project.

Furthermore, whole transcriptome sequencing of the Cinachyrella transcriptome revealed how complex organisms sponges are. Sponges are highly symbiotic animals and this was shown in the sequence data with the presence of many bacterial, fungal, and protists related transcripts. Consequently, rather than sequencing the transcriptome of the sponge, in this study the holo-transcriptome was sequenced.

Since the completion of this study, another poriferan genome has been sequence by Kenny and its colleagues (Kenny et al, 2020). The chromosome level genome assembly of Ephydatia muelleri provides additional resources for future poriferan genomic project and will certainly help in the annotation of future genomes and transcriptome assemblies.

Consequently the results presented here highlighted that sponges, even if benthic organisms follow the same sensitivity patterns as other organisms in the case of oil spills scenario. While short term crude oil exposure triggers a strong genetic response, long term exposure for benthic organisms, in the absence of chemical dispersant would eventually result in a slowly decreasing stress response from the organisms. Dispersant only exposures also agree with the previous literature as for longer exposure being more detrimental to the health of the organisms exposed to, potentially leading to death in sublethal doses exposure conditions. However, in order to conclude after very long term effects, more experimentation with longer exposure times and different concentration is needed. Furthermore, oil:dispersant exposure agreed with previous findings. Longer exposure is more harmful to the sponge as chemical dispersant allows crude oil to reach these benthic organisms.

Overall, toxicity from crude oil, dispersant, and oil:dispersant mixtures can be summarized as follows: i) at 1 hour, dispersant is the least toxic, then come the oil:dispersant mixture and finally comes crude oil only (O>OD>D); ii) at 24 hours, oil:dispersant mixtures are the most toxic, followed by dispersant only, and finally crude oil only (OD>D>O).

Supplemental Data

The holo-transcriptome raw data will be accessible as a data note published at a later date on NCBI through accession number (put SRA number). Assembly and annotation file will also be available under accession number (put TSA number) on the NCBI Database as well.

Raw sequence file from the genetic profiling section of this study are available in the NCBI database under the accession number (put SRA number)

Figures

- Fig 9 - SEM controls - Micrographs of 1 hour control (A), oil treated (WAF ; B), dispersant (C), oil:dispersant (CE-WAF ; D) sponges as seen under scanning electronic microscopy. Treated tissues looked similar to control tissues: healthy and compact, indicating that sponges were doing well phenotypically after 1 hour of exposure to the chemicals.

- Fig 10 - SEM Micrographs of 24 hours control (A), oil treated (WAF ; B), dispersant (C), oil:dispersant (CE-WAF ; D) sponges as seen under scanning electronic microscopy. In comparison to the control, treated tissues looked unhealthy and less compact. Spicules started to be apparent and cohesiveness of tissues looked to be compromised. This was the phenotypic proof of a genotypic change.

- Fig 11 - Micrographs of oil droplets in sponge tissues (top and bottom left, black arrows point to oil droplets trapped in the sponge tissues, scale bar = 500 mm) and broken down aquiferous systems (top and bottom right, scale bar = 200 mm). (A) and (B) represent 24 hours oil (WAF) treated samples (C) and (D) represent 24 hours oil:dispersant (CE-WAF) treated samples. In both treatments, oil droplets were clearly visible within the sponge tissue, potentially indicating the non-sufficient dispersant properties of the chemical dispersant used to disperse the oil. Both treatments resulted in broken down aquiferous systems of the sponge resulting in a reduce pumping ability by the organism.

- Fig 12 -composite C1x O1x D1x OD1 volcano plots, the rest go into supplemental

- Fig composite C24x O24x D24x OD24?

- Fig - composite Venn diagram C1x O1x D1x OD1, and same with 24 hours

Tables

Table 1: Table representing the experimental design followed in this study with sample naming

APPENDIX 3: Useful Codes

This section regroups a non-exhaustive collection of command line and bash codes necessary to run the different programs required for data analysis. As well, other useful codes are presented to help with formatting file for data analysis from the command line. All the followings codes and commands work in a LINUX/UNIX environment as well as bash. Bash is a command line language to speak directly with the computer. As well, it is a good idea to be familiar with python coding and language as some of the commands are extracting from python scripts. Please note that all programs used in the analysis have to be installed before their usage. All installing procedure are available on each website of the program (link provided).
Each code is presented with what it does and how to write it.
In the instance a code doesn't work because of potential updates or that the code has been changed by the developers, resources are available to the link provided or can be accessed through the command line with the following command: man software_you_need_help_with.

To merge FASTQ files together

Cat *.fastq >> merged.fastq

Be careful, if all your files are in the same directory (folder), this command will merge all of them together and you will not have separate sample representative. It is highly recommended to either create different directories for each sample and repeat the command in each different directory. In the case all files are in the same directory, you can also run the command by modifying the '*.fastq' part to match the names of your files. For example, if you have four files for a sample named:
Sample1_lane1.fastq, Sample1_lane2.fastq, Sample1_lane3.fastq, Sample1_lane4.fastq
You can run the previous command by specifying 'Sample1*' to tell the command prompt to grab only files starting with the name 'Sample1'. The results will be a single file with all file combined together.

FastQC

FastQC is a program to check the quality of your sequenced reads. Two versions are available. One version very user friendly with a user interface can be downloaded at https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. There is also a command line version of this program that works with the following code:

```
fastqc --extract file.fastq -o ~/directory/of/output/
```

This command however works well for one file at a time. If you have many files you might want to loop this command so that all your files are run one after the other, and you only have to run the command once. To perform this, follow the next command:

```
For file in *.fastq; do
fastqc --extract "$file" -o ~/directory/of/output/
done
```

Notes: 1- fastqc output is in HTML format. If you decide to use their interactive graphical application, make sure that the latest java release is installed on your machine.
2- be careful with the " symbol. It is a specific symbol in the command line and most of the time copying and pasting directly will change this symbol and the command might not work. It is better to type directly the loop in the command line.

Cutadapt (https://cutadapt.readthedocs.io/en/stable/guide.html)

Cutadapt is used to remove any adapter contamination from your reads as well as removing any polyA tail in the case of RNA-Seq data. The following code can also be looped:

```
cutadapt -j 1 -a polyA=A{10} -b
illumina=AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC -u 6 –output=out1.fq.gz –
error-rate=0.1 –times=1 –overlap=3 –minimum-length=20 –maximum-length-151 –quality-
cutoff=25 -o output_file.fastq input_file.fastq
```

If you wish to loop this command do as follows:

```
for file in *fastq; do
cutadapt -j 1 -a polyA=A{10} -b
illumina=AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC -u 6 –output=out1.fq.gz –
error-rate=0.1 –times=1 –overlap=3 –minimum-length=20 –maximum-length-151 –quality-
cutoff=25 -o $file_trimmed.fastq $file.fastq
done
```

Bowtie2 http://bowtie-bio.sourceforge.net/bowtie2/manual.shtml

Bowtie2 is the aligner program used to find matches between our sequenced reads and our reference transcriptome. It works in two steps. The first step is to create a reference for the program to work with. This reference is created by running the following command:

```
bowtie2-build -f reference.fasta reference_name
```

Note: you only need to create the reference once. The same reference will be used for all analyses. Of course if you have different type of samples that need to be aligned to different reference genomes/transcriptomes, you will have to create different references.

This command can also be looped:

```
for file in *.fasta; do
bowtie2-build -f $file.fasta $file
done
```

Create alignment

Once the reference is created you can create the alignment with your sequenced reads. Bowtie2 takes two type of input files: .fasta and .fastq. For each type of file the command is slightly different as shown below:

<u>For .fasta input file</u>

```
bowtie2 -f --end-to-end --very-sensitive --dpad 0 --gbar 99999999 --mp 1,1 --np 1 --score-min L,0,-0.1 --threads 4 --time -x bowtie2_reference_name -U input_file.fasta -S output_name.sam >> output.bowtie.log 2>&1
```

for .fastq input file

```
bowtie2 --end-to-end --very-sensitive --dpad 0 --gbar 99999999 --mp 1,1 --np 1 --score-min L,0,-0.1 --threads 4 --time -x bowtie2_reference_name -U input_file.fastq -S output_name.sam >> output.bowtie.log 2>&1
```

Bowtie2 produces two output file in that case, a log file that helps troubleshoot in case of issues and will report the alignment rate in case of success, and a file in .sam format with the reads alignments.

Again, this code can be looped if you have several files to align with the following code (only shown for .fasta files, for .fastq files simply match code above):

```
for file in *.fasta; do
bowtie2 -f --end-to-end --very-sensitive \
  --dpad 0 --gbar 99999999 --mp 1,1 --np 1 --score-min L,0,-0.1 \
  --threads 4 --time -x \
  bowtie2_reference_name \
  -U "$file" -S "$file.sam" \
  >> "$file.bowtie.log" 2>&1;
done
```

RSEM (https://deweylab.github.io/RSEM/)

RSEM is the counting program used to extract raw counts out of the alignment file. It works in two steps. The first step is to create a reference for the program to work with. This reference is created by running the following command:

To build the reference

```
rsem-prepare-reference reference_transcriptome.fasta reference_name
```

To get gene counts

rsem-calculate-expression --seed-length 15 --alignments /path/to/alignment/file /path/to/reference/directory/reference_name sample_name

Looping Bowtie2 and RSEM

From a time saving point of view, it is recommended to loop the bowtie2 and RSEM steps so they can run all your samples with one single command. Keep in mind that the code might need to be adapted or changed if needed, for .fastq files or for paired-end reads. Here it is presented for single-end reads in .fasta format:

```
for file in *.fasta;
  do bowtie2 -f --end-to-end --very-sensitive \
  --dpad 0 --gbar 99999999 --mp 1,1 --np 1 --score-min L,0,-0.1 \
  --threads 4 --time -x \
  bowtie2_reference_name \
  -U "$file" -S "$file.sam" \
  >> "$file.bowtie.log" 2>&1;
done &&
for i in *.sam;
  do rsem-calculate-expression -p 4 \
  --alignments "/path/to/alignment/$i" \
  path/to/reference/directory/reference_name \
"$i"_counts
  >> "$i.rsem.log" 2>&1;
done
```

Notes: 1- the .bowtie.log and .rsem.log were added to the code to facilitate troubleshooting and get extra information on alignment rate for bowtie and counting information for RSEM.
2- the && after the bowtie2 step is to ensure that the bowtie2 step occurs properly before going into the RSEM step. With that “&&” argument, if the bowtie2 step doesn’t finish properly the RSEM step for that specific file will not be processed.

DESEQ2 code

The R script for DESeq2 contains many lines. Consequently the full script is not presented there but can be found at the following link: https://liverootnova-my.sharepoint.com/:u:/r/personal/yd215_mynsu_nova_edu/Documents/Yvain%20Desplat%20Thesis/polished_DESeq2_script.R?csf=1&web=1&e=GCefkL. The metadata file can be found at the same link.
If access is needed to this file, you can contact the author of the thesis (Yvain Desplat, yd215@mynsu.nova.edu ; or the supervisor (Dr. Lopez, joslo@nova.edu)

Data manipulation helpful codes

These are all bash commands that can be ran through a Unix/Linux-based environment command line prompt.
Basic command line tools can be found here (https://ubuntu.com/tutorials/command-line-for-beginners#1-overview) and here (https://stackify.com/top-command-line-tools/). It is a good idea to get familiar with these before getting into further, and more difficult manipulations.
It is also a good idea to learn python coding or at least the basics of python. Tutorial can be found here (https://www.tutorialspoint.com/python/python_basic_syntax.htm)

Identifying number of protein coding genes

grep -c ‘identifier’ file.txt ; this command grabs (grep) all entries containing the identifier stated and count them (-c) in the file of interest.
These entries can then be poured into a new file if required by running the same command and adding an output extension as such

grep -c ‘identifier’ file.txt > new_file.txt

Counting the number of unique protein coding genes

cut -d ‘ ’ -f 1 file.txt | sort | uniq | wc

this command cuts all delimiters (here in this case a simple space, but can be a tab delimiter which is represented in code language by four spaces) in the specified file (-f). Using the pipe (|) it then sorts the entries so that all similar entries are next consecutive to each other. The next pipe grabs only unique entries and if two consecutives entries are the same, counts it as one. Finally the 'wc' command returns the number of sorted unique entries in the file. The number '1' in this example tells the specify the column you want the command to be run though. You might have several columns in a file and want to sort unique entries in a specific column.

Extracting specific columns of a file

Your assembly file might contain many unwanted columns and for the ease of reading through the file you might want to only keep specific columns. For that, you can use the function awk. Note that awk is part of the python language and thus its grammar is a little different than regular bash language. Awk is very powerful and almost everything can be done with awk if you know how to use it right.

Awk '{print $1, $2, $n, …}' file.txt > new_file.txt ; this commands prints the different columns you selected ($colum_number) from the selected file and pours it into a newly created file. Columns number do not have to be in increasing order it can be rearranged in which ever order you need. For example if your original file has 16 columns but you only need 4 of them starting with columns 14 you can write awk '{print $14, $1, $10, $16}' file.txt > out_file.txt

awk 'BEGIN { OFS = "\t"; ORS = "\n" } { print $1, $2, $n… }' file.txt > file_out.txt ; this command will print out the specific columns of a file separated by a tabulation. This is required when you wish to deal with only certain columns and not all the columns of one file.

The awk command can be looped but the syntax is a little tricky. Here's how you would loop an awk command to print some columns with a tabulation separator. **Be careful, a tabulation in UNIX/LINUX is represented by <u>four spaces</u> not a regular tabulation**.

for file in $(ls *.txt)

```
do
   awk 'BEGIN { OFS = "\t"; ORS = "\n" } { print $1, $2 }' ${file} > ${file}_raw.txt
done
```

Note: there are for space before the beginning of the awk command. As this is a python command, it is required that the awk command (or any other command) in a loop is preceded by 4 space to be able to be read. Otherwise the command will not run.

Awk can also be used to transform .fastq files in .fasta files:

```
awk '{if(NR%4==1) {printf(">%s\n",substr($0,2));} else if(NR%4==2) print;}' file.fastq > file.fasta
```

if needed loop it as follows:

```
for file in $(ls *.fastq)
do
   awk '{if(NR%4==1) {printf(">%s\n",substr($0,2));} else if(NR%4==2) print;}' ${file} > ${file}.fasta
done
```

Putting files together

In the even you have several file with the same first columns and want to build a table with the first common column of all files and the rest of the columns of all the files, you can use the “paste” command. That command will simply grab every file and put it next to each other in columns:

```
paste file1.txt file2.txt
```

or if you want to paste all the .txt files in your directory:

```
paste *.txt
```

Combining the paste command with awk is very useful to extract specific columns:

Paste *.txt | awk 'BEGIN { OFS = "\t"; ORS = "\n" } { print $1, $2, $4, $6, … }' > new_file.txt

Adding or removing characters in a file

Finally, a very powerful command is “sed”. Tutorial are available here) https://www.tutorialspoint.com/sed/index.htm) and here (https://www.grymoire.com/Unix/Sed.html)
Sed stands for "stream editor" and it can perform lots of function on file like, searching, find and replace, insertion or deletion. Though most common use of SED command in UNIX is for substitution or for find and replace. By using SED you can edit files even without opening it, which is much quicker way to find and replace something in file. An example of sed:

Sed ‘s/^/#/’ file.txt ; this command adds a “#” at the beginning of each line in the file called “file.txt”.

However, if this command is ran, this will only show the file in the command prompt and not actually print it in the file. To do so you will have to redirect the output to a new file:

Sed ‘s/^/#/’ file.txt > file_2.txt

References

Adeyemo, O. K., Kroll, K. J., & Denslow, N. D. (2015). Developmental abnormalities and differential expression of genes induced in oil and dispersant exposed Menidia beryllina embryos. Aquatic Toxicology, 168, 60–71. doi: 10.1016/j.aquatox.2015.09.012

Ainsworth, C. H., Paris, C. B., Perlin, N., Dornberger, L. N., Patterson, W. F., Chancellor, E., … Perryman, H. (2018). Impacts of the Deepwater Horizon oil spill evaluated using an end-to-end ecosystem model. Plos One, 13(1). doi: 10.1371/journal.pone.0190840

Alenzi, F.q.b. "Links between Apoptosis, Proliferation and the Cell Cycle." British Journal of Biomedical Science, vol. 61, no. 2, 2004, pp. 99–102., doi:10.1080/09674845.2004.11732652.

Almeda, R., Baca, S., Hyatt, C., & Buskey, E. J. (2014). Ingestion and sublethal effects of physically and chemically dispersed crude oil on marine planktonic copepods. Ecotoxicology, 23(6), 988–1003. doi: 10.1007/s10646-014-1242-6

Almeda, R., Wambaugh, Z., Chai, C., Wang, Z., Liu, Z., & Buskey, E. J. (2013). Effects of Crude Oil Exposure on Bioaccumulation of Polycyclic Aromatic Hydrocarbons and Survival of Adult and Larval Stages of Gelatinous Zooplankton. PLoS ONE, 8(10). doi: 10.1371/journal.pone.0074476

Amos, J. (2010, July 27). BP / Gulf Oil Spill – 68,000 Square Miles of Direct Impact – SkyTruth. Retrieved from https://www.skytruth.org/2010/07/bp-gulf-oil-spill-68000-square-miles-of/

Andrews, S. (2010). FastQC A Quality Control tool for High Throughput Sequence Data. Available online at : http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Aurand, D, and G Coelho. 2005. "Cooperative Aquatic Toxicity Testing of Dispersed Oil and the Chemical Response to Oil Spills: Ecological Effects Research Forum (CROSERF)."

Barange, M., Merino, G., Blanchard, J. L., Scholtens, J., Harle, J., Allison, E. H., … Jennings, S. (2014). Impacts of climate change on marine ecosystem production in societies dependent on fisheries. Nature Climate Change, 4(3), 211-216. doi:10.1038/nclimate2119

Bandaru, S., Zhou, A., Rouhi, P. et al. Targeting filamin B induces tumor growth and metastasis via enhanced activity of matrix metalloproteinase-9 and secretion of VEGF-A. Oncogenesis 3, e119 (2014). https://doi.org/10.1038/oncsis.2014.33

Batista, D., Muricy, G., Rocha, R. C., & Miekeley, N. F. (2014). Marine sponges with contrasting life histories can be complementary biomonitors of heavy metal pollution in coastal ecosystems. Environmental Science and Pollution Research, 21(9), 5785-5794. doi:10.1007/s11356-014-2530-7

Bayha, K. M., Ortell, N., Ryan, C. N., Griffitt, K. J., Krasnec, M., Sena, J., … Griffitt, R. J. (2017). Crude oil impairs immune function and increases susceptibility to pathogenic bacteria in southern flounder. PLOS ONE, 12(5), e0176559. doi:10.1371/journal.pone.0176559

Beirão, J., Baillon, L., Litt, M. A., Langlois, V. S., & Purchase, C. F. (2019). Impact of crude oil and the dispersant Corexit™ EC9500A on capelin (Mallotus villosus) embryo

development. Marine Environmental Research, 147, 90-100. doi:10.1016/j.marenvres.2019.04.004

Belinky, F., Szitenberg, A., Goldfarb, I., Feldstein, T., Wörheide, G., Ilan, M., & Huchon, D. (2012). ALG11 – A new variable DNA marker for sponge phylogeny: Comparison of phylogenetic performances with the 18S rDNA and the COI gene. Molecular Phylogenetics and Evolution, 63(3), 702-713. doi:10.1016/j.ympev.2012.02.008

Berger, K. N., & Pu, J. J. (2018). PD-1 pathway and its clinical application: A 20 year journey after discovery of the complete human PD - 1 gene. Gene, 638, 20-25. doi:10.1016/j.gene.2017.09.050

Biard DS, Miccoli L, Despras E, et al. Participation of kin17 protein in replication factories and in other DNA transactions mediated by high molecular weight nuclear complexes. Mol Cancer Res. 2003;1(7):519-531.

Bretherton, L., Kamalanathan, M., Genzer, J., Hillhouse, J., Setta, S., Liang, Y., . . . Quigg, A. (2019). Response of natural phytoplankton communities exposed to crude oil and chemical dispersants during a mesocosm experiment. Aquatic Toxicology, 206, 43-53. doi:10.1016/j.aquatox.2018.11.004

Brewton, R. A., Fulford, R., & Griffitt, R. J. (2013). Gene Expression and Growth as Indicators of Effects of the BPDeepwater HorizonOil Spill on Spotted Seatrout (Cynoscion nebulosus). Journal of Toxicology and Environmental Health, Part A, 76(21), 1198–1209. doi: 10.1080/15287394.2013.848394

Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015;12(1):59-60. doi:10.1038/nmeth.3176

Cameron et al., (2018). Impact of the Deepwater Horizon oil spill evaluated using an end to end ecosystem model. PlosOne. 13(1): e0190840, doi: 10.1371/journal.pone.0190840

Caraglia, M., Marra, M., Giuberti, G., D'alessandro, A. M., Budillon, A., Del Prete, S., ... & Abbruzzese, A. (2001). The role of eukaryotic initiation factor 5A in the control of cell proliferation and apoptosis. Amino acids, 20(2), 91-104.

Cave, E. J., & Kajiura, S. M. (2018). Effect of Deepwater Horizon Crude Oil Water Accommodated Fraction on Olfactory Function in the Atlantic Stingray, Hypanus sabinus. Scientific Reports, 8(1). doi: 10.1038/s41598-018-34140-0

Chakraborty, R., Borglin, S. E., Dubinsky, E. A., Andersen, G. L., & Hazen, T. C. (2012). Microbial Response to the MC-252 Oil and Corexit 9500 in the Gulf of Mexico. Frontiers in Microbiology, 3. doi: 10.3389/fmicb.2012.00357

Chelikani P, Fita I, Loewen PC (January 2004). "Diversity of structures and properties among catalases". Cellular and Molecular Life Sciences. 61 (2): 192–208. doi:10.1007/s00018-003-3206-5. hdl:10261/111097. PMID 14745498.

Chen, X., Zhou, Y., Yang, D., Zhao, H., Wang, L., & Yuan, X. (2012). CYP4 mRNA expression in marine polychaete Perinereis aibuhitensis in response to petroleum hydrocarbon and deltamethrin. Marine Pollution Bulletin, 64(9), 1782–1788. doi: 10.1016/j.marpolbul.2012.05.035

Cheung, W. W., Sarmiento, J. L., Dunne, J., Frölicher, T. L., Lam, V. W., Deng Palomares, M. L., … Pauly, D. (2012). Shrinking of fishes exacerbates impacts of global ocean changes on marine ecosystems. Nature Climate Change, 3(3), 254-258. doi:10.1038/nclimate1691

Conaco, C., Neveu, P., Zhou, H., Arcila, M., Degnan, S. M., Degnan, B. M., & Kosik, K. S. (2012). Transcriptome profiling of the demosponge Amphimedon queenslandica reveals genome-wide events that accompany major life cycle transitions. BMC Genomics, 13(1), 209. doi: 10.1186/1471-2164-13-209

Conaco, C., Tsoulfas, P., Sakarya, O., Dolan, A., Werren, J., & Kosik, K. S. (2016). Detection of Prokaryotic Genes in the Amphimedon queenslandica Genome. Plos One, 11(3). doi: 10.1371/journal.pone.0151092

Couillard, C. M., Lee, K., Légaré, B., & King, T. L. (2005). Effect Of Dispersant On The Composition Of The Water-Accommodated Fraction Of Crude Oil And Its Toxicity To Larval Marine Fish. Environmental Toxicology and Chemistry, 24(6), 1496. doi:10.1897/04-267r.1

Cuvelier, M. L., Blake, E., Mulheron, R., Mccarthy, P. J., Blackwelder, P., Thurber, R. L. V., & Lopez, J. V. (2014). Two distinct microbial communities revealed in the sponge Cinachyrella. Frontiers in Microbiology, 5. doi: 10.3389/fmicb.2014.00581

Dasgupta, S., Choyke, S., Ferguson, P. L., & Mcelroy, A. E. (2018). Antioxidant responses and oxidative stress in sheepshead minnow larvae exposed to Corexit 9500® or its component surfactant, DOSS. Aquatic Toxicology,194, 10-17. doi:10.1016/j.aquatox.2017.10.010

Deis, D. R., Fleeger, J. W., Bourgoin, S. M., Mendelssohn, I. A., Lin, Q., & Hou, A. (2017). Shoreline oiling effects and recovery of salt marsh macroinvertebrates from the Deepwater Horizon Oil Spill. PeerJ, 5, e3680. doi:10.7717/peerj.3680

Deleo, D. M., Herrera, S., Lengyel, S. D., Quattrini, A. M., Kulathinal, R. J., & Cordes, E. E. (2018). Gene expression profiling reveals deep-sea coral response to the Deepwater Horizon oil spill. Molecular Ecology, 27(20), 4066–4077. doi: 10.1111/mec.14847

Deleo, D. M., Ruiz-Ramos, D. V., Baums, I. B., & Cordes, E. E. (2016). Response of deep-water corals to oil and chemical dispersant exposure. Deep Sea Research Part II: Topical Studies in Oceanography, 129, 137-147. doi:10.1016/j.dsr2.2015.02.028

Demopoulos, A., Bourque, J., Cordes, E., & Stamler, K. (2016). Impacts of the Deepwater Horizon oil spill on deep-sea coral-associated sediment communities. Marine Ecology Progress Series, 561, 51-68. doi:10.3354/meps11905

Desjardins, P., & Conklin, D. (2010). NanoDrop Microvolume Quantitation of Nucleic Acids. Journal of Visualized Experiments, (-1). doi: 10.3791/2565

Dickinson DP. Cysteine Peptidases of Mammals: Their Biological Roles and Potential Effects in the Oral Cavity and Other Tissues in Health and Disease, CROBM 2002 13: 238.

Drickamer, K.; Taylor, M. (2011). "Chapter 9: Carbohydrate recognition in cell adhesion and signaling". Introduction to Glycobiology (3rd ed.). Oxford University Press. ISBN 978-0-19-956911-3.

Dubansky, B., Rice, C. D., Barrois, L. F., & Galvez, F. (2017). Biomarkers of Aryl-hydrocarbon Receptor Activity in Gulf Killifish (Fundulus grandis) From Northern Gulf of Mexico Marshes Following the Deepwater Horizon Oil Spill. Archives of Environmental Contamination and Toxicology, 73(1), 63–75. doi: 10.1007/s00244-017-0417-6

Dubansky, B., Whitehead, A., Rice, C. D., & Galvez, F. (2014). Response to Comment on "Multi-Tissue Molecular, Genomic, and Developmental Effects of the Deepwater Horizon Oil Spill on Resident Gulf Killifish (Fundulus grandis)". Environmental Science & Technology, 48(13), 7679-7680. doi:10.1021/es501185a

Dunn, Casey W, and Joseph F Ryan. "The Evolution of Animal Genomes." Current Opinion in Genetics & Development, vol. 35, 2015, pp. 25–32., doi:10.1016/j.gde.2015.08.006.

Edmunds, R. C., Gill, J. A., Baldwin, D. H., Linbo, T. L., French, B. L., Brown, T. L., … Incardona, J. P. (2015). Corresponding morphological and molecular indicators of crude oil toxicity to the developing hearts of mahi mahi. Scientific Reports, 5(1). doi: 10.1038/srep17326

Engelberts, J. P., Robbins, S. J., Goeij, J. M., Aranda, M., Bell, S. C., & Webster, N. S. (2020). Characterization of a sponge microbiome using an integrative genome-centric approach. The ISME Journal, 14(5), 1100-1110. doi:10.1038/s41396-020-0591-9

Ensembl Genomes. (2009). Retrieved from http://ensemblgenomes.org

Etnoyer, P. J., Wickes, L. N., Silva, M., Dubick, J. D., Balthis, L., Salgado, E., & Macdonald, I. R. (2015). Decline in condition of gorgonian octocorals on mesophotic reefs in the northern Gulf of Mexico: Before and after the Deepwater Horizon oil spill. Coral Reefs, 35(1), 77-90. doi:10.1007/s00338-015-1363-2

Fernández-Carrera, A., Rogers, K. L., Weber, S. C., Chanton, J. P., & Montoya, J. P. (2016). Deep Water Horizon oil and methane carbon entered the food web in the Gulf of Mexico. Limnology and Oceanography, 61(S1). doi: 10.1002/lno.10440

Fernandez-Valverde, S. L., Calcino, A. D., & Degnan, B. M. (2015). Deep developmental transcriptome sequencing uncovers numerous new genes and enhances gene annotation in the sponge Amphimedon queenslandica. BMC Genomics, 16(1). doi:10.1186/s12864-015-1588-z

Feuda, R., Dohrmann, M., Pett, W., Philippe, H., Rota-Stabelli, O., Lartillot, N., ... & Pisani, D. (2017). Improved modeling of compositional heterogeneity supports sponges as sister to all other animals. Current Biology, 27(24), 3864-3870.

Fisher, C. R., Hsing, P.-Y., Kaiser, C. L., Yoerger, D. R., Roberts, H. H., Shedd, W. W., … Brooks, J. M. (2014). Footprint of Deepwater Horizon blowout impact to deep-water coral communities. Proceedings of the National Academy of Sciences, 111(32), 11744–11749. doi: 10.1073/pnas.1403492111

Francis, W. R., Eitel, M., Vargas R., S., Adamski, M., Haddock, S. H., Krebs, S., … Wörheide, G. (2017). The Genome Of The Contractile Demosponge Tethya wilhelma And The Evolution Of Metazoan Neural Signalling Pathways. doi:10.1101/120998

Freeman, C. J., & Thacker, R. W. (2011). Complex interactions between marine sponges and their symbiotic microbial communities. Limnology and Oceanography, 56(5), 1577–1586. doi: 10.4319/lo.2011.56.5.1577

Frometa, J., Delorenzo, M. E., Pisarski, E. C., & Etnoyer, P. J. (2017). Toxicity of oil and dispersant on the deep water gorgonian octocoral Swiftia exserta, with implications for the effects of the Deepwater Horizon oil spill. Marine Pollution Bulletin, 122(1-2), 91-99. doi:10.1016/j.marpolbul.2017.06.009

Garcia, Tzintzuni I, et al. "RNA-Seq Reveals Complex Genetic Response to Deepwater Horizon Oil Release in Fundulus Grandis." BMC Genomics, vol. 13, no. 1, 2012, p. 474., doi:10.1186/1471-2164-13-474.

Gardiner, W. W., Word, J. Q., Word J .D., Perkins R. A., McFarlin K. M., Hester, B. W., Word L. S., and Ray C. M. (2013). The acute toxicity of chemically and physically dispersed crude oil to key artic species under arctic conditions during the open water season. Environmental Toxicology and Chemistry/Setac, 32(10):2284-2300. doi:10.1002/etc.2307

Garson, M. J. (1994). The biosynthesis of sponge secondary metabolites: why is it important. sponges in time and space, 427-440.

Girard, F., & Fisher, C. R. (2018). Long-term impact of the Deepwater Horizon oil spill on deep-sea corals detected after seven years of monitoring. Biological Conservation, 225, 117-127. doi:10.1016/j.biocon.2018.06.028

GIGA Community of Scientists (Lopez JV as lead author). 2014. The Global Invertebrate Genome Alliance (GIGA): Developing Community Resources to Study Diverse Invertebrates. Journal of Heredity. 105:1-18. doi: 10.1093/jhered/est084.

Godoy, Patricio; Schmidt-Heck, Wolfgang; Hellwig, Birte; Nell, Patrick; Feuerborn, David; Rahnenführer, Jörg; Kattler, Kathrin; Walter, Jörn; Blüthgen, Nils; G. Hengstler, Jan (5 July 2018). "Assessment of stem cell differentiation based on genome-wide expression profiles". Philosophical Transactions of the Royal Society B. 373 (1750): 20170221. doi:10.1098/rstb.2017.0221. PMC 5974444. PMID 29786556

Goldstone, J. V. (2008). Environmental sensing and response genes in cnidaria: The chemical defensome in the sea anemone Nematostella vectensis. Cell Biology and Toxicology, 24(6), 483-502. doi:10.1007/s10565-008-9107-5

Goldstone, J., Hamdoun, A., Cole, B., Howard-Ashby, M., Nebert, D., Scally, M., . . . Stegeman, J. (2006). The chemical defensome: Environmental sensing and response genes in the Strongylocentrotus purpuratus genome. Developmental Biology, 300(1), 366-384. doi:10.1016/j.ydbio.2006.08.066

Goodbody-Gringley, G., Wetzel, D. L., Gillon, D., Pulster, E., Miller, A., & Ritchie, K. B. (2013). Toxicity of Deepwater Horizon Source Oil and the Chemical Dispersant, Corexit® 9500, to Coral Larvae. PLoS ONE, 8(1), e45574. doi:10.1371/journal.pone.0045574

Goud, B., Liu, S., & Storrie, B. (2018). Rab proteins as major determinants of the Golgi complex structure. Small GTPases, 9(1-2), 66-75. doi:10.1080/21541248.2017.1384087

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., … Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology, 29(7), 644–652. doi: 10.1038/nbt.1883

Greer, J. B., Pasparakis, C., Stieglitz, J. D., Benetti, D., Grosell, M., & Schlenk, D. (2019). Effects of corexit 9500A and Corexit-crude oil mixtures on transcriptomic pathways and developmental toxicity in early life stage mahi-mahi (Coryphaena hippurus). Aquatic Toxicology, 212, 233-240. doi:10.1016/j.aquatox.2019.05.014

Grey, E. K., Chiasson, S. C., Williams, H. G., Troeger, V. J., & Taylor, C. M. (2015). Evaluation of Blue Crab, Callinectes sapidus, Megalopal Settlement and Condition during the Deepwater Horizon Oil Spill. Plos One, 10(8). doi: 10.1371/journal.pone.0135791

Griffith, S. M. (2013). Isolating microsatellite markers in the marine sponge Cinachyrella alloclada for use in community and population genetics studies. Retrieved from https://www.research.manchester.ac.uk/portal/files/54549200/FULL_TEXT.PDF

Gunning PW, Ghoshdastider U, Whitaker S, Popp D, Robinson RC (June 2015). "The evolution of compositionally and functionally distinct actin filaments". Journal of Cell Science. 128 (11): 2009–19. doi:10.1242/jcs.165563. PMID 25788699.

Guzman, C., & Conaco, C. (2016). Comparative transcriptome analysis reveals insights into the streamlined genomes of haplosclerid demosponges. Scientific Reports, 6(1). doi: 10.1038/srep18774

Han, J., Won, E., Hwang, D., Shin, K., Lee, Y. S., Leung, K. M., . . . Lee, J. (2014). Crude oil exposure results in oxidative stress-mediated dysfunctional development and reproduction in the copepod Tigriopus japonicus and modulates expression of cytochrome P450 (CYP) genes. Aquatic Toxicology, 152, 308-317. doi:10.1016/j.aquatox.2014.04.027

Han, L., Vickers, K. C., Samuels, D. C., & Guo, Y. (2014). Alternative applications for distinct RNA sequencing strategies. Briefings in Bioinformatics, 16(4), 629-639. doi:10.1093/bib/bbu032

Hanai N, Nagata K, Kawajiri A, et al. Biochemical and cell biological characterization of a mammalian septin, Sept11. FEBS Lett. 2004;568(1-3):83-88. doi:10.1016/j.febslet.2004.05.030

Hansen, B. R. H., Nordtug, T., Altin, D., Booth, A., Hessen, K. M., & Olsen, A. J. (2009). Gene Expression of GST and CYP330A1 in Lipid-Rich and Lipid-Poor FemaleCalanus finmarchicus(Copepoda: Crustacea) Exposed to Dispersed Oil. Journal of Toxicology and Environmental Health, Part A, 72(3-4), 131–139. doi: 10.1080/15287390802537313

Harcet, M., Roller, M., Cetkovic, H., Perina, D., Wiens, M., Muller, W. E. G., & Vlahovicek, K. (2010). Demosponge EST Sequencing Reveals a Complex Genetic Toolkit of the Simplest Metazoans. Molecular Biology and Evolution, 27(12), 2747–2756. doi: 10.1093/molbev/msq174

Hartmann, A., Sandin, S., Chamberland, V., Marhaver, K., Goeij, J. D., & Vermeij, M. (2015). Crude oil contamination interrupts settlement of coral larvae after direct exposure ends. Marine Ecology Progress Series, 536, 163-173. doi:10.3354/meps11437

Hartwick, E., Wu, R., & Parker, D. (1982). Effects of a crude oil and an oil dispersant (Corexit 9527) on populations of the littleneck clam (protothaca staminea). Marine Environmental Research, 6(4), 291–306. doi: 10.1016/0141-1136(82)90043-5

Hautier, Y., Tilman, D., Isbell, F., Seabloom, E. W., Borer, E. T., & Reich, P. B. (2015). Anthropogenic environmental changes affect ecosystem stability via biodiversity. Science, 348(6232), 336-340. doi:10.1126/science.aaa1788

Hentschel, U., Piel, J., Degnan, S. M., & Taylor, M. W. (2012). Genomic insights into the marine sponge microbiome. Nature Reviews Microbiology, 10(9), 641–654. doi: 10.1038/nrmicro2839

Hilgartner, Stephen. (2017) Reordering Life: Knowledge and Control in the Genomics Revolution. The MIT Press. ISBN-10: 0262035863.

Hill, M., Hill, A., Lopez, N., & Harriott, O. (2005). Sponge-specific bacterial symbionts in the Caribbean sponge, Chondrilla nucula (Demospongiae, Chondrosida). Marine Biology, 148(6), 1221–1230. doi: 10.1007/s00227-005-0164-5

Hill, M., Hill, A., Lopez, J., Pomponi, S., J.V., Reed, J., Tarver, Redmond, N., Diaz, M.C., Boury-Esnault, N., Cardenas, P., Reiswig, H., Dohrmann, M., Morrow, C. Rubin, E., Chaves-Fonnegra, A., Patel, J., Formica D, Hajdu E, Lobo-Hajdu G, Klontz S, Morrow CC, Patel J, Picton B, Pohlmann, D., Richie, S., di Stefano, M., Russell , Z., Peterson, K. J., Pisani, D., Thacker, R., Collins, A. 2013 Nuclear housekeeping genes help resolve demosponge relationships and taxonomy. PLoS ONE 8(1): e50437. doi:10.1371/journal.pone.0050437

Hisata, S., Sakisaka, T., Baba, T., Yamada, T., Aoki, K., Matsuda, M., & Takai, Y. (2007). Rap1-PDZ-GEF1 interacts with a neurotrophin receptor at late endosomes, leading to sustained activation of Rap1 and ERK and neurite outgrowth. Journal of Cell Biology, 178(5), 843-860. doi:10.1083/jcb.200610073

Hofmann, G. E., Evans, T. G., Kelly, M. W., Padilla-Gamiño, J. L., Blanchette, C. A., Washburn, L., … Dutton, J. M. (2014). Exploring local adaptation and the ocean acidification seascape – studies in the California Current Large Marine Ecosystem. Biogeosciences, 11(4), 1053-1064. doi:10.5194/bg-11-1053-2014

Holth, T., Nourizadeh-Lillabadi, R., Blaesbjerg, M., Grung, M., Holbech, H., Petersen, G., … Hylland, K. (2008). Differential gene expression and biomarkers in zebrafish (Danio rerio) following exposure to produced water components. Aquatic Toxicology, 90(4), 277–291. doi: 10.1016/j.aquatox.2008.08.020

Hook, S. E., & Osborn, H. L. (2012). Comparison of toxicity and transcriptomic profiles in a diatom exposed to oil, dispersants, dispersed oil. Aquatic Toxicology, 124-125, 139–151. doi: 10.1016/j.aquatox.2012.08.005

Hsing, P.-Y., Fu, B., Larcom, E. A., Berlet, S. P., Shank, T. M., Govindarajan, A. F., ... Fisher, C. R. (2013). Evidence of lasting impact of the Deepwater Horizon oil spill on a deep Gulf of Mexico coral community. Elementa: Science of the Anthropocene, 1, 000012. doi: 10.12952/journal.elementa.000012

Incardona, J. P., Gardner, L. D., Linbo, T. L., Brown, T. L., Esbaugh, A. J., Mager, E. M., … Scholz, N. L. (2014). PNAS Plus: From the Cover: Deepwater Horizon crude oil impacts

the developing hearts of large predatory pelagic fish. Proceedings of the National Academy of Sciences, 111(15), E1510-E1518. doi:10.1073/pnas.1320950111

Jeffery, N. W., Jardine, C. B., & Gregory, T. R. (2013). A first exploration of genome size diversity in sponges. Genome, 56(8), 451–456. doi: 10.1139/gen-2012-0122

Jenny, M. J., Walton, W. C., Payton, S. L., Powers, J. M., Findlay, R. H., Oshields, B., … Cunningham, C. (2016). Transcriptomic evaluation of the American oyster, Crassostrea virginica, deployed during the Deepwater Horizon oil spill: Evidence of an active hydrocarbon response pathway. Marine Environmental Research, 120, 166–181. doi: 10.1016/j.marenvres.2016.08.006

Johansen, J. L., Allan, B. J. M., Rummer, J. L., & Esbaugh, A. J. (2017). Oil exposure disrupts early life-history stages of coral reef fishes via behavioural impairments. Nature Ecology & Evolution, 1(8), 1146–1152. doi: 10.1038/s41559-017-0232-5

Jones, E. R., Martyniuk, C. J., Morris, J. M., Krasnec, M. O., & Griffitt, R. J. (2017). Exposure to Deepwater Horizon oil and Corexit 9500 at low concentrations induces transcriptional changes and alters immune transcriptional pathways in sheepshead minnows. Comparative Biochemistry and Physiology Part D: Genomics and Proteomics,23, 8-16. doi:10.1016/j.cbd.2017.05.001

Judson, R. S., Martin, M. T., Reif, D. M., Houck, K. A., Knudsen, T. B., Rotroff, D. M., . . . Dix, D. J. (2010). Analysis of Eight Oil Spill Dispersants Using Rapid, In Vitro Tests for Endocrine and Other Biological Activity. Environmental Science & Technology, 44(15), 5979-5985. doi:10.1021/es102150z

Jung, J. H., Ko, J., Lee, E. H., Choi, K. M., Kim, M., Yim, U. H., ... & Shim, W. J. (2017). RNA seq-and DEG-based comparison of developmental toxicity in fish embryos of two species exposed to Iranian heavy crude oil. Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology, 196, 1-10.

Khanna, S., Santos, M. J., Ustin, S. L., Koltunov, A., Kokaly, R. F., & Roberts, D. A. (2013). Detection of Salt Marsh Vegetation Stress and Recovery after the Deepwater Horizon Oil Spill in Barataria Bay, Gulf of Mexico Using AVIRIS Data. PLoS ONE, 8(11), e78989. doi:10.1371/journal.pone.0078989

Kim, J. N., Kim, B.-S., Kim, S.-J., & Cerniglia, C. E. (2012). Effects of Crude Oil, Dispersant, and Oil-Dispersant Mixtures on Human Fecal Microbiota in an In Vitro Culture System. MBio, 3(5). doi: 10.1128/mbio.00376-12

Kim, R.-O., Kim, B.-M., Hwang, D.-S., Au, D. W., Jung, J.-H., Shim, W. J., … Lee, J.-S. (2013). Evaluation of biomarker potential of cytochrome P450 1A (CYP1A) gene in the marine medaka, Oryzias melastigma exposed to water-accommodated fractions (WAFs) of Iranian crude oil. Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology, 157(2), 172–182. doi: 10.1016/j.cbpc.2012.11.003

Kleindienst, S., Seidel, M., Ziervogel, K., Grim, S., Loftis, K., Harrison, S., . . . Joye, S. B. (2015). Chemical dispersants can suppress the activity of natural oil-degrading microorganisms. Proceedings of the National Academy of Sciences, 112(48), 14900-14905. doi:10.1073/pnas.1507380112

Knight, R., Jansson, J., Field, D., Fierer, N., Desai, N., Fuhrman, J. A., … Gilbert, J. A. (2012). Unlocking the potential of metagenomics through replicated experimental design. Nature Biotechnology, 30(6), 513-520. doi:10.1038/nbt.2235

Knudsen BS, Feller SM, Hanafusa H. Four proline-rich sequences of the guanine-nucleotide exchange factor C3G bind with unique specificity to the first Src homology 3 domain of Crk. J Biol Chem. 1994;269(52):32781-32787.

Koepfli, K.-P., Paten, B., & O'Brien, S. J. (2015). The Genome 10K Project: A Way Forward. Annual Review of Animal Biosciences, 3(1), 57–111. doi: 10.1146/annurev-animal-090414-014900

Koonin, E. V. "Darwinian Evolution in the Light of Genomics." Nucleic Acids Research, vol. 37, no. 4, 2008, pp. 1011–1034., doi:10.1093/nar/gkp089.

Lakhani, Shyam, et al. "Oil and Gas: Field Data." GOV.UK, Aug. 2016, www.gov.uk/guidance/oil-and-gas-uk-field-data#uk-oil-and-gas-reserves-and-resources.

Lamendella, R., Strutt, S., Borglin, S., Chakraborty, R., Tas, N., Mason, O. U., … Jansson, J. K. (2014). Assessment of the Deepwater Horizon oil spill impact on Gulf coast microbial communities. Frontiers in Microbiology, 5. doi: 10.3389/fmicb.2014.00130

Lane, S. M., Smith, C. R., Mitchell, J., Balmer, B. C., Barry, K. P., Mcdonald, T., … Schwacke, L. H. (2015). Reproductive outcome and survival of common bottlenose dolphins sampled in Barataria Bay, Louisiana, USA, following the Deepwater Horizon oil spill. Proceedings of the Royal Society B: Biological Sciences, 282(1818), 20151944. doi: 10.1098/rspb.2015.1944

Langmead, Ben and Trapnell, Cole and Pop, Mihai and Salzberg, Steven L (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. In Genome Biology, 10 (3), pp. R25. doi:10.1186/gb-2009-10-3-r25

Lee, J. H., Rho, S. B., & Chun, T. (2005). Programmed Cell Death 6 (PDCD6) Protein Interacts with Death-Associated Protein Kinase 1 (DAPk1): Additive Effect on Apoptosis via Caspase-3 Dependent Pathway. Biotechnology Letters, 27(14), 1011-1015. doi:10.1007/s10529-005-7869-x

Lewin, H. A., Robinson, G. E., Kress, W. J., Baker, W. J., Coddington, J., Crandall, K. A., ... & Goldstein, M. M. (2018). Earth BioGenome Project: Sequencing life for the future of life. Proceedings of the National Academy of Sciences, 115(17), 4325-4333.

Lewis, J. P., Tarnecki, J. H., Garner, S. B., Chagaris, D. D., & Patterson, W. F. (2020). Changes in Reef Fish Community Structure Following the Deepwater Horizon Oil Spill. Scientific Reports, 10(1). doi: 10.1038/s41598-020-62574-y

Leys, S. (2017, November 24). Ephydatia muelleri Trinity transcriptome. Retrieved from https://era.library.ualberta.ca/files/b47429936z

Li, B., & Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics, 12(1). doi: 10.1186/1471-2105-12-323

Liu, Y.-Z., Roy-Engel, A. M., Baddoo, M. C., Flemington, E. K., Wang, G., & Wang, H. (2016). The impact of oil spill to lung health—Insights from an RNA-seq study of human airway epithelial cells. Gene, 578(1), 38–51. doi: 10.1016/j.gene.2015.12.016

Liu, Y.-Z., Zhang, L., Roy-Engel, A. M., Saito, S., Lasky, J. A., Wang, G., & Wang, H. (2017). Carcinogenic effects of oil dispersants: A KEGG pathway-based RNA-seq study of human airway epithelial cells. Gene, 602, 16–23. doi: 10.1016/j.gene.2016.11.028

Lönning, S., & Hagström, B. E. (1976). Deleterious effects of corexit 9527 on fertilization and development. Marine Pollution Bulletin, 7(7), 124–127. doi: 10.1016/0025-326x(76)90197-1

Love, Michael I and Huber, Wolfgang and Anders, Simon (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. In Genome Biology, 15 (12). doi:10.1186/s13059-014-0550-8

Lopez JV. 2019. After the Taxonomic Identification Phase: Addressing the Functions of Symbiotic Communities within Marine Invertebrates. In Symbionts of Marine Sponges and Corals. Z. Li (ed.), Springer- Verlag. Dordrecht, The Netherlands. Pp 105-144. https://doi.org/10.1007/978-94-024-1612-1_8

Lowe, S. R., Kwok, R. K., Payne, J., Engel, L. S., Galea, S., & Sandler, D. P. (2015). Mental health service use by cleanup workers in the aftermath of the Deepwater Horizon oil spill. Social Science & Medicine, 130, 125–134. doi: 10.1016/j.socscimed.2015.02.009

Luter, H. M., Whalan, S., Andreakis, N., Wahab, M. A., Botté, E. S., Negri, A. P., & Webster, N. S. (2019). The Effects of Crude Oil and Dispersant on the Larval Sponge Holobiont. MSystems, 4(6). doi: 10.1128/msystems.00743-19

Macías-Silva, M., Abdollah, S., Hoodless, P. A., Pirone, R., Attisano, L., & Wrana, J. L. (1996). MADR2 Is a Substrate of the TGFβ Receptor and Its Phosphorylation Is Required for Nuclear Accumulation and Signaling. Cell, 87(7), 1215-1224. doi:10.1016/s0092-8674(00)81817-6

Maire, A. L., Schiltz, M., Stura, E. A., Pinon-Lataillade, G., Couprie, J., Moutiez, M., . . . Zinn-Justin, S. (2006). A Tandem of SH3-like Domains Participates in RNA Binding in KIN17, a Human Protein Activated in Response to Genotoxics. Journal of Molecular Biology, 364(4), 764-776. doi:10.1016/j.jmb.2006.09.033

Marcel, Martin (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. In EMBnet.journal, 17 (1)

Mason, O. U., Scott, N. M., Gonzalez, A., Robbins-Pianka, A., Bælum, J., Kimbrel, J., … Jansson, J. K. (2014). Metagenomics reveals sediment microbial community response to Deepwater Horizon oil spill. The ISME Journal, 8(7), 1464–1475. doi: 10.1038/ismej.2013.254

Matsuo, A. Y. O., Woodin, B. R., Reddy, C. M., Val, A. L., & Stegeman, J. J. (2006). Humic Substances and Crude Oil Induce Cytochrome P450 1A Expression in the Amazonian Fish SpeciesColossoma macropomum(Tambaqui). Environmental Science & Technology, 40(8), 2851–2858. doi: 10.1021/es052437i

Mattos, J. J., Siebert, M. N., Luchmann, K. H., Granucci, N., Dorrington, T., Stoco, P. H., … Bainy, A. C. (2010). Differential gene expression in Poecilia vivipara exposed to diesel oil water accommodated fraction. Marine Environmental Research, 69. doi: 10.1016/j.marenvres.2009.11.002

May, L. A., Burnett, A. R., Miller, C. V., Pisarski, E., Webster, L. F., Moffitt, Z. J., . . . Woodley, C. M. (2020). Effect of Louisiana sweet crude oil on a Pacific coral, Pocillopora damicornis. Aquatic Toxicology, 222, 105454. doi:10.1016/j.aquatox.2020.105454

Mccall, B. D., & Pennings, S. C. (2012). Disturbance and Recovery of Salt Marsh Arthropod Communities following BP Deepwater Horizon Oil Spill. PLoS ONE, 7(3). doi: 10.1371/journal.pone.0032735

Mcclain, C. R., Nunnally, C., & Benfield, M. C. (2019). Persistent and substantial impacts of the Deepwater Horizon oil spill on deep-sea megafauna. Royal Society Open Science, 6(8), 191164. doi: 10.1098/rsos.191164

Medeiros, D. S., Medeiros, T. L., Ribeiro, J. K., Monteiro, N. K., Migliolo, L., Uchoa, A. F., … Santos, E. A. (2010). A lactose specific lectin from the sponge Cinachyrella apion: Purification, characterization, N-terminal sequences alignment and agglutinating activity on Leishmania promastigotes. Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology, 155(3), 211–216. doi: 10.1016/j.cbpb.2009.10.016

Mehbub, M. F., Lei, J., Franco, C., & Zhang, W. (2014). Marine sponge derived natural products between 2001 and 2010: trends and opportunities for discovery of bioactives. Marine drugs, 12(8), 4539-4577.

Miccoli, L. (2003). Selective interactions of human kin17 and RPA proteins with chromatin and the nuclear matrix in a DNA damage- and cell cycle-regulated manner. Nucleic Acids Research, 31(14), 4162-4175. doi:10.1093/nar/gkg459

Michaelsen, S., Schaefer, J., & Peterson, M. S. (2015). Fluctuating Asymmetry in Menidia beryllina before and after the 2010 Deepwater Horizon Oil Spill. Plos One, 10(2). doi: 10.1371/journal.pone.0118742

Milligan, RJ, A.M. Bernard, K.M. Boswell, H.D. Bracken-Grissom, M.A. D'Elia, S. DeRada, C.G. Easson, D. English, R.I. Eytan, C. Hu, C. Lembke, J.V. Lopez, B. Penta, T. Richards, I.C. Romero, M. Shivji, L. Timm, J.D. Warren, M. Weber, R.J.D. Wells, T.T. Sutton. 2019. The application of novel research technologies by the Deep Pelagic Nekton Dynamics of the Gulf of Mexico (DEEPEND) Consortium. Marine Technology. 52 (6), 81-86.

Miranda, M., & Sorkin, A. (2007). Regulation of Receptors and Transporters by Ubiquitination: New Insights into Surprisingly Similar Mechanisms. Molecular Interventions, 7(3), 157-167. doi:10.1124/mi.7.3.7

Montagna, P. A., Baguley, J. G., Cooksey, C., Hartwell, I., Hyde, L. J., Hyland, J. L., … Rhodes, A. C. E. (2013). Deep-Sea Benthic Footprint of the Deepwater Horizon Blowout. PLoS ONE, 8(8). doi: 10.1371/journal.pone.0070540

Mortazavi, Ali, et al. "Mapping and Quantifying Mammalian Transcriptomes by RNA-Seq." Nature Methods, vol. 5, no. 7, 2008, pp. 621–628., doi:10.1038/nmeth.1226.

Mu, J., Jin, F., Ma, X., Lin, Z., & Wang, J. (2014). Comparative effects of biological and chemical dispersants on the bioavailability and toxicity of crude oil to early life stages of marine medaka (Oryzias melastigma). Environmental toxicology and chemistry, 33(11), 2576-2583.

Negri, A. P., Luter, H. M., Fisher, R., Brinkman, D. L., & Irving, P. (2018). Comparative toxicity of five dispersants to coral larvae. Scientific Reports, 8(1). doi:10.1038/s41598-018-20709-2

Nordborg, F. M., Flores, F., Brinkman, D. L., Agustí, S., & Negri, A. P. (2018). Phototoxic effects of two common marine fuels on the settlement success of the coral Acropora tenuis. Scientific Reports, 8(1). doi: 10.1038/s41598-018-26972-7

Nurhayati, A. P. D., Pratiwi, R., Wahyuono, S., Istriyati, I., & Voogt, N. J. D. (2015). The Anticancer Activity of Marine Sponge Cinachyrella sp. (Family Tetillidae). IPTEK The Journal for Technology and Science, 25(3). doi: 10.12962/j20882033.v25i3.530

Olsen, A. J., Nordtug, T., Altin, D., Lervik, M., & Hansen, B. H. (2013). Effects of dispersed oil on reproduction in the cold water copepodCalanus finmarchicus(Gunnerus). Environmental Toxicology and Chemistry, 32(9), 2045–2055. doi: 10.1002/etc.2273

Olsvik, P. A., Lie, K. K., Nordtug, T., & Hansen, B. H. (2012). Is chemically dispersed oil more toxic to Atlantic cod (Gadus morhua) larvae than mechanically dispersed oil? A transcriptional evaluation. BMC Genomics, 13(1), 702. doi:10.1186/1471-2164-13-702

Oncogenes and Tumor Suppressor Genes. (n.d.). Retrieved May 22, 2020, from http://www.cancer.org/

Overmans, S., Nordborg, M., Díaz-Rúa, R., Brinkman, D. L., Negri, A. P., & Agustí, S. (2018). Phototoxic effects of PAH and UVA exposure on molecular responses and developmental success in coral larvae. Aquatic Toxicology,198, 165-174. doi:10.1016/j.aquatox.2018.03.008

Ozdamar B, Bose R, Barrios-Rodiles M, Wang HR, Zhang Y, Wrana JL. Regulation of the polarity protein Par6 by TGFbeta receptors controls epithelial cell plasticity. Science. 2005;307(5715):1603-1609. doi:10.1126/science.1105718

Pannekoek, W., Dijk, J. J., Chan, O. Y., Huveneers, S., Linnemann, J. R., Spanjaard, E., . . . Bos, J. L. (2011). Epac1 and PDZ-GEF cooperate in Rap1 mediated endothelial junction control. Cellular Signalling, 23(12), 2056-2064. doi:10.1016/j.cellsig.2011.07.022

Park, M., Sim, C., Baek, J., & Min, G. (2007). Identification of gene suitable for DNA barcoding of morphologically indistinguishable Korean Halichondriidae sponges. Molecules and Cells, 23(2), 220-227.

Paul, J. H., Hollander, D., Coble, P., Daly, K. L., Murasko, S., English, D., … Kovach, C. W. (2013). Toxicity and Mutagenicity of Gulf of Mexico Waters During and After the Deepwater Horizon Oil Spill. Environmental Science & Technology, 47(17), 9651–9659. doi: 10.1021/es401761h

Pawar, P. R. (2013). Monitoring of impact of anthropogenic inputs on water quality of mangrove ecosystem of Uran, Navi Mumbai, west coast of India. Marine Pollution Bulletin, 75(1-2), 291–300. doi: 10.1016/j.marpolbul.2013.06.045

Peiffer, R. F., Cohen, J. H. (2015). Lethal and sublethal effects of oil, chemical dispersant, and dispersed oil on the ctenophore Mnemiopsis leidyi. Aquatic Biology, Vol. 23: 237-250. Doi:10.3354/ab00625.

Perez-Porro, A. (2014). Transcriptomics along a sponge life cycle. Retrieved from https://figshare.com/articles/Transcriptomics_along_a_sponge_life_cycle/1254912/1

Pérez-Porro, A. R., Navarro-Gómez, D., Uriz, M. J., & Giribet, G. (2013). A NGS approach to the encrusting Mediterranean sponge Crella elegans(Porifera, Demospongiae, Poecilosclerida): transcriptome sequencing, characterization and overview of the gene expression along three life cycle stages. Molecular Ecology Resources, 13(3), 494-509. doi:10.1111/1755-0998.12085

Perez, T., Wafo, E., Fourt, M., & Vacelet, J. (2003). Marine Sponges as Biomonitor of Polychlorobiphenyl Contamination: Concentration and Fate of 24 Congeners. Environmental Science & Technology, 37(10), 2152-2158. doi:10.1021/es026234v

Perina, D., Bosnar, M. H., Bago, R., Mikoč, A., Harcet, M., Deželjin, M., & Ćetković, H. (2011). Sponge non-metastatic Group I Nme gene/protein - structure and function is conserved from sponges to humans. BMC Evolutionary Biology, 11(1). doi: 10.1186/1471-2148-11-87

Pie, H. V., Schott, E. J., & Mitchelmore, C. L. (2015). Investigating physiological, cellular and molecular effects in juvenile blue crab, Callinectus sapidus, exposed to field-collected sediments contaminated by oil from the Deepwater Horizon Incident. Science of The Total Environment, 532, 528–539. doi: 10.1016/j.scitotenv.2015.06.022

Piel, J., Hui, D., Wen, G., Butzke, D., Platzer, M., Fusetani, N., & Matsunaga, S. (2004). Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge Theonella swinhoei. Proceedings of the National Academy of Sciences, 101(46), 16222–16227. doi: 10.1073/pnas.0405976101

Pilcher, W., Miles, S., Tang, S., Mayer, G., & Whitehead, A. (2014). Genomic and Genotoxic Responses to Controlled Weathered-Oil Exposures Confirm and Extend Field Studies on Impacts of the Deepwater Horizon Oil Spill on Native Killifish. PLoS ONE, 9(9). doi: 10.1371/journal.pone.0106351

Pilcher, W.Miles, S., Tang, S., Mayer, G., & Whitehead, A. (2014) Genomic and Genotoxic Responses to Controlled Weathered-Oil Exposures Confirm and Extend Field Studies on Impacts of the Deepwater Horizon Oil Spill on Native Killifish. PLoS ONE, vol. 9, no. 9, doi:10.1371/journal.pone.0106351.

Pisani, D., Pett, W., Dohrmann, M., Feuda, R., Rota-Stabelli, O., Philippe, H., ... & Wörheide, G. (2015). Genomic data do not support comb jellies as the sister group to all other animals. Proceedings of the National Academy of Sciences, 112(50), 15402-15407

Polli, J. R., Zhang, Y., & Pan, X. (2014). Dispersed crude oil amplifies germ cell apoptosis in Caenorhabditis elegans, followed a CEP-1-dependent pathway. Archives of Toxicology. doi:10.1007/s00204-014-1198-6

Powers, S. P., Hernandez, F. J., Condon, R. H., Drymon, J. M., & Free, C. M. (2013). Novel Pathways for Injury from Offshore Oil Spills: Direct, Sublethal and Indirect Effects of the

Deepwater Horizon Oil Spill on Pelagic Sargassum Communities. PLoS ONE, 8(9), e74802. doi:10.1371/journal.pone.0074802

Prasad, K. V., Ao, Z., Yoon, Y., Wu, M. X., Rizk, M., Jacquot, S., & Schlossman, S. F. (1997). CD27, a member of the tumor necrosis factor receptor family, induces apoptosis and binds to Siva, a proapoptotic protein. Proceedings of the National Academy of Sciences, 94(12), 6346-6351. doi:10.1073/pnas.94.12.6346

Prince, R. C., & Butler, J. D. (2013). A protocol for assessing the effectiveness of oil spill dispersants in stimulating the biodegradation of oil. Environmental Science and Pollution Research, 21(16), 9506–9510. doi: 10.1007/s11356-013-2053-7

Putman, N. F., Abreu-Grobois, F. A., Iturbe-Darkistade, I., Putman, E. M., Richards, P. M., & Verley, P. (2015). Deepwater Horizon oil spill impacts on sea turtles could span the Atlantic. Biology Letters, 11(12), 20150596. doi: 10.1098/rsbl.2015.0596

Qiu, F., Ding, S., Ou, H., Wang, D., Chen, J., & Miyamoto, M. (2015). Transcriptome Changes during the Life Cycle of the Red Sponge, Mycale phyllophila (Porifera, Demospongiae, Poecilosclerida). Genes, 6(4), 1023–1052. doi: 10.3390/genes6041023

Rabelo, L., Monteiro, N., Serquiz, R., Santos, P., Oliveira, R., Oliveira, A., … Santos, E. (2012). A Lactose-Binding Lectin from the Marine Sponge Cinachyrella Apion (Cal) Induces Cell Death in Human Cervical Adenocarcinoma Cells. Marine Drugs, 10(12), 727–743. doi: 10.3390/md10040727

Rao, J. V., Kavitha, P., Reddy, N. C., & Rao, T. G. (2006). Petrosia testudinaria as a biomarker for metal contamination at Gulf of Mannar, southeast coast of India. Chemosphere, 65(4), 634-638. doi:10.1016/j.chemosphere.2006.01.072

Reddy, C. M., Arey, J. S., Seewald, J. S., Sylva, S. P., Lemkau, K. L., Nelson, R. K., ... & Van Mooy, B. A. (2012). Composition and fate of gas and oil released to the water column during the Deepwater Horizon oil spill. Proceedings of the National Academy of Sciences, 109(50), 20229-20234

Renard, E., Leys, S. P., Wörheide, G., & Borchiellini, C. (2018). Understanding Animal Evolution: The Added Value of Sponge Transcriptomics and Genomics. BioEssays, 40(9), 1700237. doi:10.1002/bies.201700237

Reuscher, M. G., Baguley, J. G., Conrad-Forrest, N., Cooksey, C., Hyland, J. L., Lewis, C., … Washburn, T. (2017). Temporal patterns of Deepwater Horizon impacts on the benthic infauna of the northern Gulf of Mexico continental slope. Plos One, 12(6). doi: 10.1371/journal.pone.0179923

Rhee, J.-S., Kim, B.-M., Choi, B.-S., Choi, I.-Y., Wu, R. S. S., Nelson, D. R., & Lee, J.-S. (2013). Whole Spectrum of Cytochrome P450 Genes and Molecular Responses to Water-Accommodated Fractions Exposure in the Marine Medaka. Environmental Science & Technology, 47(9), 4804–4812. doi: 10.1021/es400186r

Rhodes, N. (2014, July 10). A. queenslandica Transcriptome. Retrieved from http://amphimedon.qcloud.qcif.edu.au/index.html

Rico-Martínez, R., Snell, T. W., & Shearer, T. L. (2013). Synergistic toxicity of Macondo crude oil and dispersant Corexit 9500A® to the Brachionus plicatilis species complex (Rotifera). Environmental Pollution, 173, 5–10. doi: 10.1016/j.envpol.2012.09.024

Riesgo, A., Andrade, S. C., Sharma, P. P., Novo, M., Pérez-Porro, A. R., Vahtera, V., … Giribet, G. (2012). Comparative description of ten transcriptomes of newly sequenced invertebrates and efficiency estimation of genomic sampling in non-model taxa. Frontiers in Zoology, 9(1), 33. doi:10.1186/1742-9994-9-33

Riesgo, A., Farrar, N., Windsor, P. J., Giribet, G., & Leys, S. P. (2014). The Analysis of Eight Transcriptomes from All Poriferan Classes Reveals Surprising Genetic Complexity in Sponges. Molecular Biology and Evolution, 31(5), 1102-1120. doi:10.1093/molbev/msu057

Riesgo, A., Peterson, K., Richardson, C., Heist, T., Strehlow, B., Mccauley, M., … Hill, A. (2014). Transcriptomic analysis of differential host gene expression upon uptake of symbionts: a case study with Symbiodinium and the major bioeroding sponge Cliona varians. BMC Genomics, 15(1), 376. doi: 10.1186/1471-2164-15-376

Riesgo, A., Taboada, S., Pérez-Portela, R., Melis, P., Xavier, J. R., Blasco, G., & López-Legentil, S. (2019). Genetic diversity, connectivity and gene flow along the distribution of the emblematic Atlanto-Mediterranean sponge Petrosia ficiformis (Haplosclerida, Demospongiae). BMC Evolutionary Biology, 19(1). doi: 10.1186/s12862-018-1343-6

Rivers, A. R., Sharma, S., Tringe, S. G., Martin, J., Joye, S. B., & Moran, M. A. (2013). Transcriptional response of bathypelagic marine bacterioplankton to the Deepwater Horizon oil spill. The ISME Journal, 7(12), 2315–2329. doi: 10.1038/ismej.2013.129

Rodriguez-R, L. M., Overholt, W. A., Hagan, C., Huettel, M., Kostka, J. E., & Konstantinidis, K. T. (2015). Microbial community successional patterns in beach sands impacted by the Deepwater Horizon oil spill. The ISME Journal, 9(9), 1928-1940. doi:10.1038/ismej.2015.5

Rooker, J. R., Kitchens, L. L., Dance, M. A., Wells, R. J. D., Falterman, B., & Cornic, M. (2013). Spatial, Temporal, and Habitat-Related Variation in Abundance of Pelagic Fishes in the Gulf of Mexico: Potential Implications of the Deepwater Horizon Oil Spill. PLoS ONE, 8(10). doi: 10.1371/journal.pone.0076080

Rua, C. P., Zilberberg, C., & Solé-Cava, A. M. (2011). New polymorphic mitochondrial markers for sponge phylogeography. Journal of the Marine Biological Association of the United Kingdom, 91(5), 1015-1022. doi:10.1017/s0025315410002122

Rubin-Blum, M., Antony, C. P., Borowski, C., Sayavedra, L., Pape, T., Sahling, H., … Dubilier, N. (2017). Short-chain alkanes fuel mussel and sponge Cycloclasticus symbionts from deep-sea gas and oil seeps. Nature Microbiology, 2(8). doi: 10.1038/nmicrobiol.2017.93

Ruiz-Ramos, Dannise V., et al. “Stress Response of the Black Coral Leiopathes Glaberrima When Exposed to Sub-Lethal Amounts of Crude Oil and Dispersant.” Elem Sci Anth, vol. 5, 2017, p. 77., doi:10.1525/elementa.261.

Sakakibara, A. (2002). Novel function of Chat in controlling cell adhesion via Cas-Crk-C3G-pathway-mediated Rap1 activation. Journal of Cell Science, 115(24), 4915-4924. doi:10.1242/jcs.00207

Salamanca, H. H., Antonyak, M. A., Cerione, R. A., Shi, H., & Lis, J. T. (2014). Inhibiting Heat Shock Factor 1 in Human Cancer Cells with a Potent RNA Aptamer. PLoS ONE, 9(5). doi:10.1371/journal.pone.0096330

Samuelsson J, Alonso S, Ruiz-Larroya T, Cheung TH, Wong YF, Perucho M. Frequent somatic demethylation of RAPGEF1/C3G intronic sequences in gastrointestinal and gynecological cancer. Int J Oncol. 2011;38(6):1575-1577. doi:10.3892/ijo.2011.972

Santavy, D. L., Willenz, P., & Colwell, R. R. (1990). Phenotypic study of bacteria associated with the caribbean sclerosponge, Ceratoporella nicholsoni. Applied and Environmental Microbiology, 56(6), 1750-1762

Schmitt, S., Tsai, P., Bell, J., Fromont, J., Ilan, M., Lindquist, N., … Taylor, M. W. (2011). Assessing the complex sponge microbiota: core, variable and species-specific bacterial communities in marine sponges. The ISME Journal, 6(3), 564–576. doi: 10.1038/ismej.2011.116

Schöttner, S., Hoffmann, F., Cárdenas, P., Rapp, H. T., Boetius, A., & Ramette, A. (2013). Relationships between Host Phylogeny, Host Type and Bacterial Community Diversity in Cold-Water Coral Reef Sponges. PLoS ONE, 8(2). doi: 10.1371/journal.pone.0055505

Schuster, A., Lopez, J. V., Becking, L. E., Kelly, M., Pomponi, S. A., Wörheide, G., … Cárdenas, P. (2017). Evolution of group I introns in Porifera: new evidence for intron mobility and implications for DNA barcoding. BMC Evolutionary Biology, 17(1). doi: 10.1186/s12862-017-0928-9

Shinzato, C., Hamada, M., Shoguchi, E., Kawashima, T., & Satoh, N. (2012). The Repertoire of Chemical Defense Genes in the Coral Acropora digitifera Genome. Zoological Science, 29(8), 510. doi:10.2108/zsj.29.510

Smith, E. (2013). De novo Transcriptome Analysis of the Marine Sponge Cinachyrella spp: A Potential Model Organism for Oil and Dispersant Ecotoxicology. Master's thesis. Nova Southeastern University. Retrieved from NSUWorks, Oceanographic Center. (141) http://nsuworks.nova.edu/occ_stuetd/141.

Snyder, S. M., Pulster, E. L., & Murawski, S. A. (2019). Associations Between Chronic Exposure to Polycyclic Aromatic Hydrocarbons and Health Indices in Gulf of Mexico Tilefish (Lopholatilus chamaeleonticeps) Post Deepwater Horizon. Environmental Toxicology and Chemistry, 38(12), 2659-2671. doi:10.1002/etc.4583

Soysa, T. Y. D., Ulrich, A., Friedrich, T., Pite, D., Compton, S. L., Ok, D., … Barresi, M. J. (2012). Macondo crude oil from the Deepwater Horizon oil spill disrupts specific developmental processes during zebrafish embryogenesis. BMC Biology, 10(1), 40. doi: 10.1186/1741-7007-10-40

Srikanth, K., & Rao, J. V. (2016). Sponges as Biomonitors of Metal Toxicity in the Aquatic Systems. Marine Sponges: Chemicobiological and Biomedical Applications, 105-114. doi:10.1007/978-81-322-2794-6_8

Srivastava, M., Simakov, O., Chapman, J., Fahey, B., Gauthier, M. E. a., Mitros, T., Richards, G.S., Conaco, C., Dacre, M., Hellsten, U., Larroux, C., Putnam, N. H., Stanke, M., Adamska, M., Darling, A., Degnan, S. M., Oakley, T.H., Plachetzki, D. C., Zhai, Y., Adamski, M., Calcino, A., Cummins, S.

F., Goodstein, D. M., Harris, C., Jackson, D. J., Leys, S. P., Shu, S., Woodcroft, B. J., Vervoort, M., Kosik, K. S., Manning, G., Degnan, B. M., and Rokhsar, D.S. (2010).The Amphimedon queenslandica genome and the evolution of animal complexity. Nature, 466(7307):720–6

Stagg, R. M., Rusin, J., Mcphail, M. E., Mcintosh, A. D., Moffat, C. F., & Craft, J. A. (2000). Effects Of Polycyclic Aromatic Hydrocarbons On Expression Of Cyp1A In Salmon (Salmo Salar) Following Experimental Exposure And After The Braer Oil Spill. Environmental Toxicology and Chemistry, 19(11), 2797. doi: 10.1897/1551-5028(2000)019<2797:eopaho>2.0.co;2

Stroski, K. M., Tomy, G., & Palace, V. (2019). The current state of knowledge for toxicity of corexit EC9500A dispersant: a review. Critical Reviews in Environmental Science and Technology, 49(2), 81–103. doi: 10.1080/10643389.2018.1532256

Studivan, M. S., Hatch, W. I., & Mitchelmore, C. L. (2015). Responses of the soft coral Xenia elongata following acute exposure to a chemical dispersant. SpringerPlus, 4(1). doi: 10.1186/s40064-015-0844-7

Sun, L., Youn, H., Loh, C., Stolow, M., He, W., & Liu, J. O. (1998). Cabin 1, A Negative Regulator for Calcineurin Signaling in T Lymphocytes. Immunity, 8(6), 703-711. doi:10.1016/s1074-7613(00)80575-0

Sun, Y. (2006). E3 Ubiquitin Ligases as Cancer Targets and Biomarkers. Neoplasia, 8(8), 645-654. doi:10.1593/neo.06376

Sunbald, V., Mathieu, V., Kiss, R., & Rabinovich, G. A. (n.d.). Chapter 31 - Galectins: Key Players in the Tumor Microenvironment. In Cancer Immunotherpay. Academic Press.

Szabo, David (2014). "Transcriptomic biomarkers in safety and risk assessment of chemicals". Transcriptomic biomarkers in safety and risk assessment of chemicals. In Ramesh Gupta, editors:Gupta - Biomarkers in Toxicology, Oxford:Academic Press. pp. 1033–1038. doi:10.1016/B978-0-12-404630-6.00062-2. ISBN 978-0-12-404630-6.

Szitenberg, A., Rot, C., Ilan, M., & Huchon, D. (2010). Diversity of sponge mitochondrial introns revealed by cox 1 sequences of Tetillidae. BMC Evolutionary Biology, 10(1), 288. doi: 10.1186/1471-2148-10-288

Tagami, H., Ray-Gallet, D., Almouzni, G., & Nakatani, Y. (2004). Histone H3.1 and H3.3 Complexes Mediate Nucleosome Assembly Pathways Dependent or Independent of DNA Synthesis. Cell, 116(1), 51-61. doi:10.1016/s0092-8674(03)01064-x

Tarrant, A. M., Payton, S. L., Reitzel, A. M., Porter, D. T., & Jenny, M. J. (2018). Ultraviolet radiation significantly enhances the molecular response to dispersant and sweet crude oil exposure in Nematostella vectensis. Marine Environmental Research, 134, 96-108. doi:10.1016/j.marenvres.2018.01.002

Techtmann, S. M., Zhuang, M., Campo, P., Holder, E., Elk, M., Hazen, T. C., … Domingo, J. W. S. (2017). Corexit 9500 Enhances Oil Biodegradation and Changes Active Bacterial Community Structure of Oil-Enriched Microcosms. Applied and Environmental Microbiology, 83(10). doi: 10.1128/aem.03462-16

Teixeira, L. K., & Reed, S. I. (2013). Ubiquitin Ligases and Cell Cycle Control. Annual Review of Biochemistry, 82(1), 387-414. doi:10.1146/annurev-biochem-060410-105307

Tethya wilhelma genome. (n.d.). Retrieved from https://bitbucket.org/wrf/tethya-wilhelma-genome

The Global Invertebrate Genomics Alliance (GIGA): Developing Community Resources to Study Diverse Invertebrate Genomes. (2013). Journal of Heredity, 105(1), 1-18. doi:10.1093/jhered/est084

Vargas, S., Schuster, A., Sacher, K., Büttner, G., Schätzle, S., Läuchli, B., … Wörheide, G. (2012). Barcoding Sponges: An Overview Based on Comprehensive Sampling. PLoS ONE, 7(7). doi: 10.1371/journal.pone.0039345

Vazquez-Miranda, H., Thoma, B. P., Wong, J. M., Felder, D. L., Crandall, K. A., & Bracken-Grissom, H. D. (2017). Transcriptomic effects of dispersed oil in a non-model decapod crustacean. doi:10.7287/peerj.preprints.2977

Venkateswara Rao, J., Srikanth, K., Pallela, R., & Gnaneshwar Rao, T. (2008). The use of marine sponge, Haliclona tenuiramosa as bioindicator to monitor heavy metal pollution in the coasts of Gulf of Mannar, India. Environmental Monitoring and Assessment, 156(1-4), 451-459. doi:10.1007/s10661-008-0497-x

Venn-Watson, S., Colegrove, K. M., Litz, J., Kinsel, M., Terio, K., Saliki, J., … Rowles, T. (2015). Adrenal Gland and Lung Lesions in Gulf of Mexico Common Bottlenose Dolphins (Tursiops truncatus) Found Dead following the Deepwater Horizon Oil Spill. PLOS ONE, 10(5), e0126538. doi:10.1371/journal.pone.0126538

Vijayan, N. Unpublsihed. Oil histology of Cinachyrella sponges in face of the Deepwater Horizon oil spill.

Waterhouse, Robert M, et al. “BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics.” Molecular Biology and Evolution, vol. 35, no. 3, 2017, pp. 543–548., doi:10.1093/molbev/msx319.

Webster, N. S., & Taylor, M. W. (2011). Marine sponges and their microbial symbionts: love and other relationships. Environmental Microbiology, 14(2), 335–346. doi: 10.1111/j.1462-2920.2011.02460.x

Webster, N. S., & Thomas, T. (2016). The Sponge Hologenome. MBio, 7(2). doi: 10.1128/mbio.00135-16

Webster, N. S., Taylor, M. W., Behnam, F., Lücker, S., Rattei, T., Whalan, S., … Wagner, M. (2009). Deep sequencing reveals exceptional diversity and modes of transmission for bacterial sponge symbionts. Environmental Microbiology. doi: 10.1111/j.1462-2920.2009.02065.x

Webster, N. S., Webb, R. I., Ridd, M. J., Hill, R. T., & Negri, A. P. (2001). The effects of copper on the microbial community of a coral reef sponge. Environmental Microbiology, 3(1), 19-31. doi:10.1046/j.1462-2920.2001.00155.x

Wenger, Y., & Galliot, B. (2013). RNAseq versus genome-predicted transcriptomes: a large population of novel transcripts identified in an Illumina-454 Hydra transcriptome. BMC Genomics, 14(1), 204. doi:10.1186/1471-2164-14-204

White, H. K., Hsing, P.-Y., Cho, W., Shank, T. M., Cordes, E. E., Quattrini, A. M., … Fisher, C. R. (2012). Impact of the Deepwater Horizon oil spill on a deep-water coral community in the Gulf of Mexico. Proceedings of the National Academy of Sciences of the United States of America. doi: 10.1073/pnas.1118029109

Whitehead, A., Dubansky, B., Bodinier, C., Garcia, T. I., Miles, S., Pilley, C., … Galvez, F. (2011). Genomic and physiological footprint of the Deepwater Horizon oil spill on resident marsh fishes. Proceedings of the National Academy of Sciences, 109(50), 20298–20302. doi: 10.1073/pnas.1109545108

Wieser, R., Wrana, J., & Massagué, J. (1995). GS domain mutations that constitutively activate T beta R-I, the downstream signaling component in the TGF-beta receptor complex. The EMBO Journal, 14(10), 2199-2208. doi:10.1002/j.1460-2075.1995.tb07214.x

Wise, J., & Wise, J. P. (2011). A review of the toxicity of chemical dispersants. Reviews on Environmental Health,26(4). doi:10.1515/reveh.2011.035

Woese, Carl R. "A New Biology for a New Century." Microbiology and Molecular Biology Reviews, vol. 68, no. 2, 2004, pp. 173–186., doi:10.1128/mmbr.68.2.173-186.2004.

Xu, E. G., Khursigara, A. J., Magnuson, J., Hazard, E. S., Hardiman, G., Esbaugh, A. J., … Schlenk, D. (2017). Larval Red Drum (Sciaenops ocellatus) Sublethal Exposure to Weathered Deepwater Horizon Crude Oil: Developmental and Transcriptomic Consequences. Environmental Science & Technology, 51(17), 10162–10172. doi: 10.1021/acs.est.7b02037

Xu, E. G., Mager, E. M., Grosell, M. S., Pasparakis, C., Schlenker, L. S., Stieglitz, J. D., … Schlenk, D. (2016). Time- and Oil-Dependent Transcriptomic and Physiological Responses to Deepwater Horizon Oil in Mahi-Mahi (Coryphaena hippurus) Embryos and Larvae. ACSPublications. doi: 10.1021/acs.est.6b02205

Xu, E. G., Mager, E. M., Grosell, M., Hazard, E. S., Hardiman, G., & Schlenk, D. (2017). Novel transcriptome assembly and comparative toxicity pathway analysis in mahi-mahi (Coryphaena hippurus) embryos and larvae exposed to Deepwater Horizon oil. Scientific Reports, 7(1). doi: 10.1038/srep44546

Yang, Q., Franco, C. M. M., Sorokin, S. J., & Zhang, W. (2017). Development of a multilocus-based approach for sponge (phylum Porifera) identification: refinement and limitations. Scientific Reports, 7(1). doi: 10.1038/srep41422

Yednock, B. K., Sullivan, T. J., & Neigel, J. E. (2015). De novo assembly of a transcriptome from juvenile blue crabs (Callinectes sapidus) following exposure to surrogate Macondo crude oil. BMC Genomics, 16(1). doi:10.1186/s12864-015-1739-2.

Zhang, X., Liu, J., Liang, X., Chen, J., Hong, J., Li, L., . . . Cai, X. (2016). History and progression of Fat cadherins in health and disease. OncoTargets and Therapy, Volume 9, 7337-7343. doi:10.2147/ott.s111176

Zhang, Y., Chen, D., Ennis, A. C., Polli, J. R., Xiao, P., Zhang, B., . . . Pan, X. (2012). Chemical dispersant potentiates crude oil impacts on growth, reproduction, and gene expression in Caenorhabditis elegans. Archives of Toxicology, 87(2), 371-382. doi:10.1007/s00204-012-0936-x

Zhao, L; Zheng, X; Liu, J; Zheng, R; Yang, R; Wang, Y; Sun, L (1 July 2019). "The placental transcriptome of the first-trimester placenta is affected by in vitro fertilization and embryo transfer". Reproductive Biology and Endocrinology. 17 (1): 50. doi:10.1186/s12958-019-0494-7. PMC 6604150. PMID 31262321.

Zheng, M., Ahuja, M., Bhattacharya, D., Clement, T. P., Hayworth, J. S., & Dhanasekaran, M. (2014). Evaluation of differential cytotoxic effects of the oil spill dispersant Corexit 9500. Life Sciences, 95(2), 108–117. doi: 10.1016/j.lfs.2013.12.010