R. Kirsten Tyler

Since June 2013, millions of sea stars in at least 20 different species (Asteroidea) from Alaska to Mexico have perished due to a suite of symptoms called sea star wasting disease (SSWD), which involves loss of turgor pressure, formation of lesions, loss of limbs and ultimately, death (Hewson et al., 2014). As keystone species in many of marine ecosystems, mass mortality of sea stars could cause dramatic alteration and biodiversity loss of benthic communities (Lessios, Robertson, & Cubit, 1984; Paine, 1966). To avoid trophic cascades that would upset the balance of marine ecosystems and potentially cause irreversible damage, it is imperative to obtain information on both the pathology of SSWD and on the biology of the seastars. *Pisaster ocraceus* was the focal species in this study. As a keystone predator involved in the regulation of mussel populations and an Asteroid species that has exhibited severe and unprecedented population declines due to the 2013 SSWD outbreak, P. ochraceus is an appropriate candidate for this study.

The cause of SSWD is still unknown (Hewson et al., 2014). A previous study has identified the sea star associated densovirus (SSaDV) as a potential causative agent of the disease when comparing gene expression and immune response to SSWD between sick and healthy individuals (Hewson et al. 2014). However, our data show that the echinoderms displayed symptoms and became infected even in the absence of the densovirus, indicating the possibility that another pathogen (or multiple pathogens) cause SSWD. Many environmental factors have also been shown to affect the onset of the disease both directly and indirectly, rendering the task of finding its causes even more challenging. Previous sea star wasting events have been associated with warmer ocean temperatures (Bates, Hilton, & Harley, 2009; Staehli, Schaerer, Hoelzle, & Ribi, 2017) and (Eisenlord et al., 2016) exhibited in their lab experiment a faster progression of SSWD in ochre stars subjected to warmer temperatures. Recently, work was published on the effect of warmer temperatures facilitating the increase in abundance of dangerous bacteria such as *Vibrio* spp. that is another potential causative agent (Vezzulli et al., 2016). These studies indicate that we have very little knowledge as to what pathogen or pathogens cause(s) SSWD, how those respective pathogen(s) work, and how environmental factors including temperature and microbiota can facilitate its spread through altering the resistance or susceptibility of sea stars to the disease, which can consequently yield to disease outbreaks.

The purpose of this study was to determine whether 1) there is a genetic basis for resistance to SSWD in sea stars and/or 2) if the microbiomes of the sea stars are responsible for the disease’s progression. Based on previous findings by Hewson *et al.* (2014), we hypothesize that sea stars free of the disease are potentially doing so through genetic means such as having more suitable alleles of immune response genes. If our hypothesis is true, we predict to see a difference in allele frequencies between healthy and sick individuals at certain distinct loci.

The second part of our study compared the microbiomes of healthy and sick individuals over time to investigate the roles, if any, that microbiota play in susceptibility or resistance of sea stars to the pathogen. We hypothesize that an individual’s microbiome does play a role in the onset and progression of SSWD. If our hypothesis is true, we predict to see a significant difference in microbiome composition between healthy and sick individuals and a significantly more stable microbiome over time in healthy individuals than in sick individuals. To distinguish whether the change in microbiome composition is causing/exacerbating the disease or if it is simply a side-effect of the disease, we examined individuals that were initially healthy but then turned sick during the course of our study; if a drastic change in microbiome precedes individuals showing any symptom of the disease, there is a possibility that the change in microbiome is the cause of the disease or at least partly responsible for the disease affecting the sea star.

Methods

To determine whether there is a genetic basis or microbiome association for resistance to SSWD in *P. ocraceus*, we collected 24 healthy individuals from intertidal and subtidal locations of Monterey, CA. They were shipped to our lab and placed in individual tanks, remaining separate. They were reared for 15 days, while every 3 days, epidermal tissue was biopsied and health status and symptoms were recorded (Days 3, 6, 9, 12, and 15). “H” designated healthy and “S” designated sick. Individuals that stayed healthy throughout the study were labeled as “HH”. Individuals that started healthy, then eventually showed symptoms of wasting were labeled as “HS”. Individuals that showed symptoms of wasting on Day 3 (the first day of recording health status) were labeled as “SS” for analyses.

RNA and 16S was extracted from tissue samples for analysis. Using a poly-A tail, mRNA was sequenced by Illumina HiSeq 3000. To assess the quality of the reads, the program FastQC was used to give each read a Phred (quality) score. Then using Trimmomatic, the reads were cleaned for quality and remaining adapter sequences were trimmed. The quality of the reads was checked once more with FastQC.

Trinity was used to assemble a de novo transcriptome using paired and concatenated reads from four individuals. Open reading frames (ORFs) were predicted and only transcripts at least 100 amino acids long were kept. To do this, we used the program TransDecoder. The cleaned reads were mapped to the transcriptome assembly to make sequence alignment (SAM) files. From these SAM files, we extracted single nucleotide polymorphisms (SNPs) using the reads2snps program. We used variant call format (VCF) files to hold the large and complex SNP data files and used VCFtools to filter and analyze SNP data.

To determine if there was genetic divergence between groups based on health status, a discriminant analysis of principle components (DAPC) was used. DAPC is a method like PCA but whereas PCA searches arbitrarily for the largest total variance in the data, DAPC maximizes the separation between pre-defined groups while minimizing variation within each group. Thus, group differences are adequately displayed. Contributions of alleles to regions of the genome with genetic divergence can be calculated among groups, which has the potential for gene location and identification. Before DAPC was applied, the raw SNP data was subset and filtered in VCFtools. Removal of all subtidal individuals was completed to create a subset of 16 individuals found only in the intertidal (3 HH, 5 HS, 8 SS). Then the file was filtered with the following criteria: all sites were kept biallelic because mutations occurring at the same site are usually rare, therefore we assumed >2 or <2 were probably sequencing errors. Minor allele frequency was set to 0.02 to remove very rare SNPs (also probably sequencing errors). Max-missing was set to 0.8 to remove sites where fewer than 80% of samples had data because if many individuals were missing data for a particular SNP, it could bias the analysis. After filtering, all 16 individuals were kept and 5,827 out of a possible 7486938 sites were kept. The filtered VCF SNP file was loaded into R for analysis.

\*\*\*\*\*\*\*\*\*\*\*\*\*explain how both sites were analyzed here\*\*\*\*\*\*\*

The adegenet package was for all DAPC analysis. A genlight object was created from the VCF SNP file for efficient analysis (R runs the analysis on this big data faster). The metadata file, previously created, was also loaded into R to assign locality information and disease status to the genlight object. The DAPC analysis was run using disease status (HH, HS, SS) to group samples in the intertidal. In the PCA step, 8 axes were retained because 65% of the variation was explained in 8 PCA bands. Using too many PC’s runs the risk of overfitting the discriminant functions when determining group membership (Jombart & Collins 2015). The number of axes retained in the DA step was 2 because the default DA is n(group) – 1. An annotation table was created by downloading BLAST+ to our server and running BLAST to the de novo assembly on the server. The annotation table was subset to include only SNPs found to significantly differentiate the groups.

\*\*\*\*\*\*\*\*\*\*\*\*\*microbiome stuff here\*\*\*\*\*\*\*

Results

A total of 24 ochre stars were collected from intertidal and subtidal habitats in Monterey, CA. All individuals were healthy upon collection but most started showing signs of the disease in the days following arrival to the lab. We ran a DAPC analysis on the intertidal habitat between healthy and sick sea stars (16 individuals) to measure the extent to which groups possess different genes that distinguish them apart based on health status.

The resulting scatter plot (figure 1), shows separation between groups, mostly between the HH individuals and the other two defined groups (HS, SS). However, there is one HH outlier that is grouping closer to the HS and SS groups.

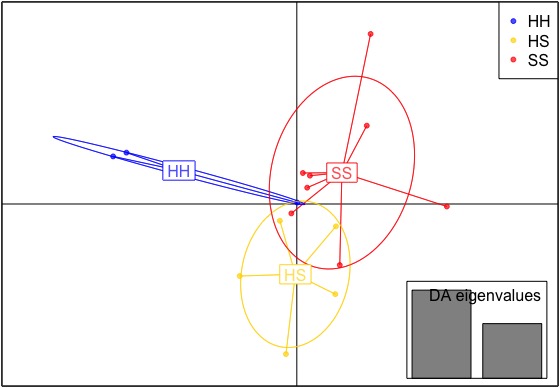


Figure : Scatter-plot of the three genetic clusters of P. ochraceus resulting from discriminant analysis of principal components (DAPC). Labels indicate health status of the ochre stars. HH means the individual stayed healthy throughout the study; HS means the individual stayed healthy until after Day 3 of rearing and remained sick until death; SS means the individual showed signs of symptoms on Day 3 of rearing and continued to stay sick until death.

To identify which sites were contributing the most to differentiating between groups, we created a loading plot (figure 2). The significant allele contributions (loadings) are noted as the spikes that rise above the threshold line, which was set to 0.02. Probability threshold “Probs = 0.999” was set to show only the top 0.1% of loci loading onto the DAPC that separate HH, HS, and SS. We found 11 alleles that significantly contributed to the differentiation between groups.

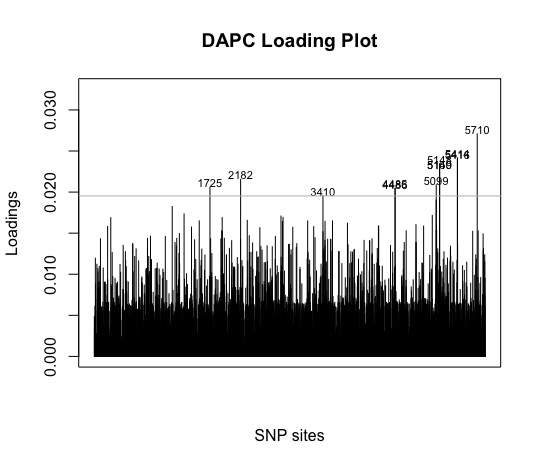


Figure : Loading plot retained from DAPC analysis which contains the loci that contribute the most to distinguishing healthy and sick individuals. In this case, 11 SNPs reflect differences between the 3 groups.

To put the results in a biological context, we used a BLASTx (\*\*\*\*\*\*\*\*\*\*\*reference\*\*\*\*\*\*\*\*\*\*\*\*) annotation table to explore functional categories of the significant genes found in our study. We found that only 5 out of the 11 genes were recognized in BLAST. Table 1 lists the genes, coded proteins, and respective GO functions.

Nucleolin (NCL, a DNA binding protein) and Endonuclease (mutS2, DNA cleaving) could be involved in maintenance processes for DNA replication. The G-protein coupled receptor is probably involved in mediating a signaling pathway. Nicotinate-nucleotide adenylyltransferase (NaMN) is an enzyme that catalyzes a reaction for biosynthesis of nicotinic acid adenine dinucleotide (NAD) (Consortium, 2017). The function of SOCS box protein 3 (SSB-3) is protein ubiquitination which may have implications for immune response to pathogens (Consortium, 2017; Coordinators, 2016).

|  |  |  |  |
| --- | --- | --- | --- |
| Protein Names | Gene Names | Organism | Gene Ontology |
| SPRY domain-containing SOCS box protein 3 (SSB-3) | SPSB3 C16orf31 SSB3 | Homo sapiens | protein ubiquitination [GO:0016567] |
| Nucleolin (Protein C23) | NCL | Gallus gallus | DNA binding [GO:0003677] |
| Endonuclease MutS2 (EC 3.1.-.-) | mutS2 Sez\_0334 | Streptococcus equi subsp. Zooepidemicus | endonuclease activity [GO:0004519] |
| Probable nicotinate-nucleotide adenylyltransferase (EC 2.7.7.18) | nadD CFF8240\_1430 | Campylobacter fetus subsp. Fetus | nicotinate-nucleotide adenylyltransferase activity [GO:0004515] |
| Smoothened homolog (SMO) (Fragment) | SMO SMOH | Gallus gallus | G-protein coupled receptor activity [GO:0004930] |

Table 1: Summary of the 5 significant genes found including names, subsequent protein names, organism sources, and GO terms

Discussion

Sea Star Wasting Disease has become increasingly prevalent in Asteroids along the Pacific coast, and with incredible mass mortality of these keystone predators, it is imperative to obtain information regarding the pathogen and host. In this study, we focused on determining whether there is a genetic basis for resistance to SSWD in sea stars and if the microbiomes of the sea stars are responsible for the disease’s progression. As expected, we observed a difference in allele frequencies between healthy and sick individuals at certain distinct loci. After running a DAPC analysis on 3 groups based on health status, 11 loci significantly differentiated the groups based on genetic divergence. We could identify 5 of the 11 genes and their GO functions, listed in table 1. Two subsequent proteins, NCL and mutS2, aided in DNA binding and regulation of mismatched strands, respectively (Consortium, 2017). The SMO protein is most likely involved in mediating signal pathways (Consortium, 2017).

Recent studies on the NAD synthesis enzyme (NaMN) have found a protective function against degeneration arising from injury or misfolded proteins in Drosophila and mammalian neurons (Ali, McCormack, Darrand, & Zhai, 2011). This gene has also been identified as a stress response protein that is required for thermotolerance and oxidative stress mitigation (Ali et al., 2011). Testing expression levels of NaMN in susceptible and resistant individuals would allow for a greater understanding of the role it plays in SSWD. Delving further into the specific functions of NaMN in sea stars is an important next step to find the role it plays in mitigating symptoms caused by SSWD. This result brings up another possible cause of SSWD symptoms because of the connection with environmental stress. Previous studies have shown that warmer ocean temperatures facilitate prevalence of disease symptoms in sea stars (Eisenlord et al., 2016) so it is possible that sea stars able to mitigate effects of environmental factors such as temperature don’t show symptoms. To test possible factors, additional environmental measurements such as temperature and pH are needed at collection sites in future studies.

Studies on SSB-3 have determined that it functions as a mediator of inducible nitric-oxide synthase (iNOS, NOS2) which is critical in host immune response to pathogenic bacterial infections (Kuang et al., 2010; Nishiya et al., 2011). The SSB-3 protein not only facilitates production of NO via iNOS, it also specifically regulates NO to prevent over-production of NO that could potentially trigger intracellular cytotoxicity within the organism. Perhaps this gene has implications for resistance to SSWD such that individuals lacking a particular allele may not be capable of the immune response necessary to survive the pathogen. Future studies should focus on the function of SSB-3 in sea stars and its relation to immune response in echinoderms. Experimental studies with symptomatic and asymptomatic individuals harboring differing alleles of the SSB-gene should be performed to understand the role SSB-3 might play in resistance to SSWD. Moreover, expression of SSB-3 should be studied in the context of SSWD. Perhaps resistant individuals can up-regulate SSB-3 upon infection with greater efficiency than susceptible individuals.

The cause of SSWD remains unknown and we have many knowledge gaps to fill regarding this deadly outbreak and the mechanisms that facilitate it. In this study, we observed differential allele frequencies in 11 specific loci between sick and healthy individuals in the intertidal zone of Monterey, CA. To confirm the association of these genes with resistance or susceptibility to SSWD, it would be helpful to replicate this study in other *P. ochraceus* populations to demonstrate a trend. If other populations exhibited the same or similar gene discrepancies, it could infer genetic resistance or immunity to the pathogen. If the same genes are found in allele frequency tests, this would suggest that they were related to (or linked to genes that were related to) pathogen resistance. If this was the case, implications for management of the pathogen would be vast, including a large-scale restoration project using individuals harboring resistant alleles. It would be useful to replicate this study in populations of unaffected *P. ochraceus* to better understand the possible roles of resistant genes to this pathogen. Moreover, larger sample sizes would better display significant results and give us a clearer picture of the roles of genes here.

* Differentially expressed (Up regulated and down regulated) genes similar to these?
* Maybe it’s not a disease per se but the effects of environmental factors or a suite of pathogens.