Differential Gene Expression Analysis And Gene Ontology In Triploid and Diploid *Pocillopora Acuta*

Deeksha Misri, Erin E. Chille, Timothy G. Stephens
Debashish Bhattacharya (Faculty Advisor)

Abstract

Corals are marine invertebrates that are facing life-threatening environmental stressors due to climate change. Polyploidy can, in such cases, be an important source of variation and adaptation in corals and other species. Polyploidy is the genomic condition wherein the cells of a normally diploid organism have more than one pair of chromosomes. Pocillopora acuta, also known as the cauliflower coral, is a brooding coral that can also reproduce asexually. It is a stress-sensitive coral, which means it shows clear physiological changes in response to environmental stressors like temperature, salinity, and pH. In this study, about 60% of the stony coral Pocillopora acuta samples collected from Kāne'ohe Bay, Oahu, HI, were triploid. The aim of this study was to identify the differences in gene expression patterns between triploid cluster 1 (T1), triploid cluster 2 (T2), and diploid samples (D) of P. acuta. Pairwise comparisons were carried out between all categories: T1 vs. D, T2 vs. D, and T1 vs. T2. While there were a large number of genes exhibiting similar expression patterns in both triploid clusters, many genes were differentially regulated in T1 when compared to T2. This result provides evidence suggesting that the two triploid lineages originated from separate triploidization events in Kāne'ohe Bay. The differentially expressed genes shared between these two triploid lineages, when compared to the diploid coral lineage, suggests changes in cellular physiology as a result of polyploidization. Functional analysis of the P. acuta genes can provide deeper insight into the specific, differentially regulated molecular functions and biological processes in triploids when compared to diploid P. acuta. Future studies involving comparative functional enrichment analysis with more triploid and diploid samples of P. acuta will provide more insight into events that caused triploidization and the coral's response to environmental stressors.

Key terms: corals, coral bleaching, polyploidy, bioinformatics

1 Introduction

Polyploidy is the genomic condition wherein an organism possesses multiple chromosome copies. This heritable condition can be a result of genome duplication within a species (autopolyploidy) or from hybridization of two different species (allopolyploidy). Polyploidy in an organism can cause a drastic change in cell organization and genome structure as well as problems in gene expression, genome stability, cell physiology, and cell cycle processes (Wertheim et al., 2021). Thus, it can result in physiological changes in the organism in response to various kinds of environmental stressors like salinity stress, thermal stress, and pCO2 stress. Polyploidy is a relatively common occurrence in plants, and until recently, was believed to be a rare occurrence in animals (Wertheim et al., 2021). However, in recent years, polyploidy has been found in all major animal phyla and occurs relatively frequently in some groups, such as fish and amphibians (Wertheim et al., 2021). In some animals, triploidy may be beneficial to the organism with respect to improved growth, pathogen resistance, and creation of more fit genotypes (Kang et al., 2004).

Coral reefs are large marine ecosystems formed by colonies of coral polyps inhabited by a diverse microbiome, including algal endosymbionts (Symbiodiniaceae), bacteria, fungi, archaea, and viruses. Due to climate change, coral reefs are subjected to acutely high temperatures that disrupt the coral-algal symbiosis leading to coral bleaching

(Williams et al., 2021). When corals bleach, they expel their symbionts, which provides most of thecarbohydrates, lipids, amino acids, and O2 that fulfill the host's energy requirements. Extended periods of bleaching, which cause compromised immunity and starvation, are a leading cause of the mass mortality of coral reefs globally (Williams et al., 2021).

Corals need effective and efficient mechanisms to adapt and ensure their survival against environmental stressors such as prolonged heat and oxidative stress. Hawai'i houses about 80 species of Scleractinia corals, and the islands are the world's most isolated archipelago, geographically isolated from other reef systems. This reef system thus provides an ideal platform for advancing coral biology and conservation using multi-omics and genetic tools. (Bhattacharya et al., 2022). To study the stress response of the stony coral Pocillopora acuta (also known as the cauliflower coral), 30 samples of the species were collected in 2018 from Kāne'ohe Bay in Hawai'i and exposed to ambient temperature and ambient CO2. From these samples, about 60% of the samples were triploid. While the origin of triploidy in this population is still unknown, there are currently three mechanistic explanations for the origin of triploidy within this population, including 1) self-fertilization of a P. acuta egg followed by fertilization by a foreign sperm, 2) diploidization of one of the two parental gametes, and 3) failure of the ovum to expel the second polar body (Stephens et al., 2021). It is also hypothesized that triploid genotypes may have spread throughout Kāne'ohe Bay by the generation of asexual brooded larvae, (Nakajima et al., 2018), and clonal propagation through colony fragmentation (Highsmith, 1982).

Because this is the first known incidence of widespread triploidy in a coral population, we aimed to gain baseline information about gene expression regulation and differential gene expression between diploid and triploid P. acuta corals in the Kāne'ohe Bay ecosystem (Stephens et al., 2021). In the study conducted by Stephens et al., an unrooted phylogenetic tree built using the SNPs identified in P. acuta RNA-Seq data showed a putative split between diploid and triploid lineages. There were two triploid lineages identified, named triploid cluster 1 (T1) and triploid cluster 2 (T2). We studied the differences in gene expression between the two triploid clades (T1 and T2) and the diploid clade (D) P. acuta. Gene expression in an organism can be measured by examining the mRNA, the protein made by the mRNA transcripts. By exploring the differences in gene expression between the diploid and triploid colonies we hope to discover if, and in what way, polyploidy provides a selective advantage. In light of the mass mortality of reef-building corals under climate change, it is essential to generate this baseline information on how triploidy affects gene expression and regulation to develop an understanding of how triploid coral populations may respond to this burgeoning stressor.

2 Methodology

Coral Sampling and Collection

Corals were collected from six reefs in Kāne'ohe Bay, a 45 km² sheltered water body in O'ahu, Hawai'i in September 2018. 30 of the 119 collected samples were cultured under ambient temperature ambient CO2 (ATAC) for a 4-month period. Molecular samples for RNA-Seq and bisulfite sequencing were obtained on days 1 and 2, and on weeks 1, 2, 4, 6, 8, 12, and 16. RNA extraction and sequencing, read trimming, and gene count estimation were conducted as described in Stephens et al. (2021). RNA was extracted from a small clipping using the Zymo Quick-DNA/RNA™ Miniprep Plus Kit (Zymo Research, Irvine, CA, USA) following the manufacturer's protocol for tissue samples. RNA library preparation and sequencing was then performed at Genewiz (South Plainfield, New Jersey, USA) following the TruSeq Stranded mRNA Sample Preparation Protocol (Illumina) with poly-A enrichment and HiSeq sequencing workflow targeting 15 million reads per sample. Quality trimming, adaptor removal, and gene count estimation were performed, and the resulting gene count matrix was used for differential expression analysis as described below.

Differential Gene Expression

Using R version 4.0.5 and the edgeR package v. 3.15 (Robinson et al., 2010), pairwise

comparisons of differential gene expression were conducted on the 27,254 genes that passed the low coverage filter of gene counts less than 10 (Robinson et al. 2010). Three pairwise gene expression comparisons were conducted: T1 vs. D, T2 vs. D and T1 vs. T2, using DESeq2 (v. 3.14) (Love et al. 2014). The detected differentially expressed genes were filtered based on adjusted *p*-values less than 0.05 and fold changes (fc) greater than 1 for upregulated genes and less than -1 for downregulated genes. The UpSet R package was used to represent the six sets of differentially expressed genes (DEGs) (upregulated and downregulated genes in T1, T2, and D samples) (Conway et al. 2017). This plot represents nonempty and overlapping genes that were differentially regulated between the pairwise comparisons.

Gene Ontology

Gene Ontology (GO) terms are a hierarchy of terms/descriptions that uses a controlled vocabulary to describe the known or inferred properties of a gene/protein. In this study, gene ontology and annotation were conducted following the GO pipeline outlined in Chille et al. (2021) to annotate genes by relevant protein accession IDs for triploid and diploid samples of *P. acuta*. Briefly, *DIAMOND* (v. 0.0.14) was used to align protein and translated DNA sequences against the RefSeq non-redundant protein sequence database (*nr* database) compiled by the NCBI (Buchfink et al. 2015). To obtain gene ontology and annotation information from multiple databases, *InterProScan* (v. 5.53-87.0) was used to search the *InterPro* database to compile additional information about *P. acuta* protein sequences (Jones et al. 2014). *Blast2GO* is a powerful bioinformatics platform for gene annotation and ontology. Using the previously generated *DIAMOND* results as input, the mapping and annotation tools on *Blast2GO* were used to map and annotate the *P. acuta* gene sequences. Out of the 38,532 gene sequences, 5073 were mapped. Then, *InterProScan* GO results were merged with the mapped and annotated sequences from *Blast2GO*. The accession IDs for the gene sequences (generated by *Blast2GO* after mapping and annotation) were then run against the UniProt database.

3 Results

RNA Sequencing, Quality Control and Clustering of Samples

For all 30 samples collected at ATAC, 38,532 genes were predicted as described in Stephens et al. (2021). Pre-filtering to retain genes with cpm (counts per million) over 3.33 in at least 2 samples resulted in 27,252 genes being selected for analysis. Out of the 30 *P. acuta* samples, 10 were diploid (D), 9 belonged to triploid cluster 1 (T1), and the remaining 11 were triploid cluster 2 (T2).

Pairwise Differential Gene Expression using R

For the first pairwise comparison (T1 vs. D), 1354 genes were upregulated in T1, 1,090 were downregulated in T1, and there was no significant difference in expression for the remaining 24,810 genes. For the second pairwise comparison (T2 vs. D), 1,157 genes were upregulated in T2, 1,330 were downregulated in T2, and there was no significant difference in expression for the remaining 24,767 genes. For the third and last pairwise comparison between the two triploid clusters (T1 vs. T2), 1,527 genes were upregulated in T1, 985 were downregulated in T1, and there was no significant difference in expression for the remaining 24,742 genes. This information is represented in Table 1.

Table 1: *Numbers of differentially expressed genes in pairwise comparisons using edgeR, with adjusted p-value < 0.05, and fold change (fc) > 1 for upregulated genes and fc <-1 for downregulated genes.*

Pairwise Comparison	Upregulated (log(fc)>1)	Downregulated (log(fc)<1)
T1 vs. D	1354	1090
T2 vs. D	1157	1330
T1 vs. T2	1527	985

An UpSet plot is a graph used to represent intersections of two or more categories that have some shared similarity (Conway et al. 2017). For this study, an UpSet plot was used to visualize the DEGs similarly regulated across pairwise comparisons.

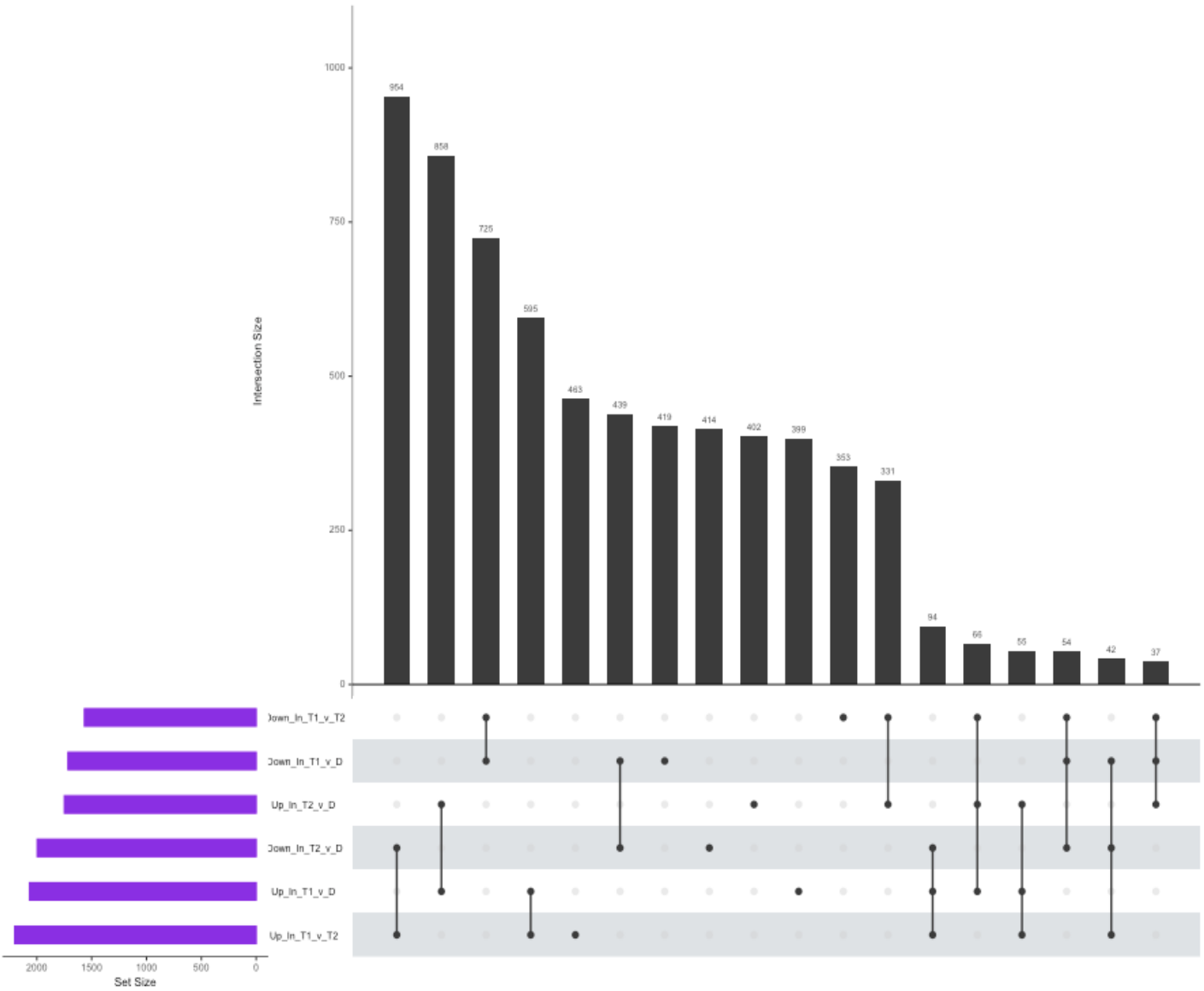

FIGURE 1: *UpSet plot of differentially expressed genes in pairwise comparisons between T1, T2, and D samples. The violet histogram on the bottom left of the figure represents the number of DEGs in each category of the pairwise comparisons. The black histogram represents the number of DEGs common between the highlighted pairwise comparisons (exact count shown on top of each black bar). For the black bars with a single dot across all sets, there is no overlap of differentially expressed genes with other categories.*

954 out of the 1,330 genes downregulated in T2 (versus D) were also upregulated in T1 (versus T2). The second largest set of intersecting genes consisted of 856 genes that were upregulated in both triploid clusters against diploid samples. In contrast, only 439 genes were downregulated in both triploid clusters against diploid samples. Other intersecting sets of genes are shown in the UpSet plot above (FIGURE 1).

4 Discussion/Conclusion

Differential Gene Expression Analysis

The aim of this study was to analyze the differential expression of genes in T1, T2, and D and to identify the physiological differences conferred by triploidy. DESeq2 was used to find the differentially expressed genes and *DIAMOND, InterProScan,* and *Blast2GO* were used for gene ontology and annota-

tion of the DEGs. While there were many genes exhibiting similar expression patterns for the two triploid clusters, there were also many differences. The number of differentially regulated genes for each triploid cluster is different when compared to each other and to diploid samples. The downregulated genes in T1 and T2 with log(fc) < -1 and adjusted *p*-value < 0.05 exhibit interesting differences in gene expression. Although both T1 and T2 clusters are triploid, the UpSet plot (Figure 1) indicates that the genes differentially expressed in both clusters show differences in gene regulation. 954 genes downregulated in T2 in comparison to diploid samples are upregulated in T1 compared to T2 and are not differentially expressed in T1 compared to D. This indicates that these 954 genes may be downregulated *only* in T2 and not T1. Similarly, there are 419 DEGs that are *only* downregulated in T1 and 414 DEGs downregulated in T2 that have no intersections with the other triploid cluster across all comparisons. However, there are 439 DEGs that are downregulated in both T1 and T2 against D. These genes depict no significant differential expression in the pairwise comparison between T1 and T2. Hence, while there are a significant number of genes that are exclusively downregulated within each triploid cluster, some are common between the two clusters as well. This indicates that although both T1 and T2 are triploid, there are a large number of differences and similarities in gene expression patterns.

This analysis is supported by the differential expression patterns for upregulated genes as well. Notably, 331 genes upregulated in T2 against D are downregulated in T1 against T2. These genes may be *only* upregulated in T2. There are 402 upregulated genes in T2 that are not differentially expressed in T1 and 399 genes upregulated in T1 but not in T2. 595 DEGs are upregulated in T1 against both T2 and D but show no differential expression in T2. However, 858 genes are upregulated in both T1 and T2 against D and show no differential expression in the pairwise comparison between T1 and T2. Again, since there are many upregulated genes showing differences and similarities in expression patterns in both T1 and T2, it is possible that after independent triploidization events, T1 and T2 adapted in somewhat similar ways to the same environmental conditions.

While both triploid clusters have many DEGs regulated in the same way, there is an equally large number of DEGs expressed in completely different ways even between the two clusters. This may indicate that there were two separate sources of triploidization in the *P. acuta* reef in Kāneʻohe Bay, but adaptation of similar forms of triploidization in the coral can cause some genes in both clusters to be differentially expressed in similar ways.

In future studies, we will compile gene ontology data obtained from *DIAMOND, InterProScan,* and *Blast2GO* and perform functional enrichment analysis will be performed on all 119 samples collected in the study. This will be done in order to identify the biological processes and molecular functions that are differentially regulated in triploid *P. acuta* across all stress conditions as described in the methodology section (ATAC, ATHC, HTAC, and HTHC). This analysis will shed more light on the physiological and biochemical differences between triploid and diploid *P. acuta*. Comparative functional enrichment analysis will provide insight into pathways impacted by differential expression of some genes in triploid lineages in *P. acuta* under environmental stress. It will also shed light on how triploidization can help the stress-sensitive *P. acuta* adapt to changing environmental conditions to ensure survival in the face of rampant climate change in the decades to come.

5 Acknowledgements

I would like to thank my mentor, Erin Chille, for her guidance and support. I would also like to thank Dr. Timothy Stephens for detecting triploidy in the samples and letting us use his findings and data as the basis for our study. Lastly, I would like to thank Dr. Debashish Bhattacharya for his continued scientific guidance.

6 References

[1] Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data.

[2] Bhattacharya, D., Stephens, T. G., Tinoco, A. I., Richmond, R. H., & Cleves, P.A. (2022). Life on the edge: Hawaiian model for coral evolution. *Limnology and Oceanography*.
[3] Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature methods*, *12*(1), 59-60.
[4] Chille, E. E., Strand, E., Neder, M., Schmidt, V., Sherman, M., Putnam, H. M., & Mass, T. (2021). Developmental series of gene expression clarifies maternal mRNA provisioning and maternal-to-zygotic transition in the reef-building coral Montipora capitata. *bioRxiv*.
[5] Conway, J. R., Lex, A., & Gehlenborg, N. (2017). UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics*.
[6] Highsmith, R. C. (1982). Reproduction by fragmentation in corals. *Marine ecology progress series. Oldendorf*, *7*(2), 207-226.
[7] Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., ... & Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics*, *30*(9), 1236-1240.
[8] Kang, H. J., & Rosenwaks, Z. (2004). Triploidy-the breakdown of monogamy between sperm and egg. *International journal of developmental biology*, *52*(5-6), 449-454.
[9] Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, *15*(12), 1-21.
[10] Nakajima, Y., Chuang, P. S., Ueda, N., & Mitarai, S. (2018). First evidence of asexual recruitment of Pocillopora acuta in Okinawa Island using genotypic identification. *PeerJ*, *6*, e5915.
[11] Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *bioinformatics*, *26*(1), 139-140.
[12] Stephens, T. G., Strand, E. L., Mohamed, A. R., Williams, A., Chiles, E. N., Su, X., & Putnam, H. M. (2021). Ploidy variation and its implications for reproduction and population dynamics in two sympatric Hawaiian coral species. *bioRxiv*.
[13] Wertheim, B., Beukeboom, L. W., & Van de Zande, L. (2013). Polyploidy in animals: effects of gene expression on sex determination, evolution and ecology. *Cytogenetic and Genome Research*, *140*(2-4), 256-269.
[14] Williams, A., Chiles, E. N., Conetta, D., Pathmanathan, J. S., Cleves, P. A., Putnam, H. M., & Bhattacharya, D. (2021). Metabolomic shifts associated with heat stress in coral holobionts. *Science advances*, *7*(1), eabd4210.
[15] Willis, B. L., van Oppen, M. J., Miller, D. J., Vollmer, S. V., & Ayre, D. J. (2006). The role of hybridization in the evolution of reef corals. *Annu. Rev. Ecol. Evol. Syst., 37,* 489-517.

Deeksha Misri is a third-year student in the Honors College at Rutgers University New Brunswick. She is majoring in Genetics with a Certificate in Computational Genetics and a minor in Philosophy. Over the past two years, she has been working as an undergraduate researcher at Dr. Debashish Bhattacharya's Lab, which she first joined as an Aresty Summer Science Fellow during her freshman year. Her research focuses on using bioinformatic and other computational tools to study coral genomics and multi-omics and the genetic and biomolecular mechanisms coral adopt to adapt to climate change. After graduating from Rutgers, Deeksha plans to continue performing research and pursue a PhD in bioinformatics. Outside of the lab, Deeksha is also involved in the women in STEM community on campus, and previously served as President of the STEM Ambassadors – Women in STEM club at Rutgers. She is also an active member of the Douglass Residential College for Women and serves as a Research Advisory Board mentor for Douglass. Previously, Deeksha has been involved with Aresty as a Summer Science participant as well as a peer reviewer for RURJ. For any questions, please contact Deeksha at deeksha.misri@rutgers.edu.