**Machine learning algorithms identify key variables that dictate symbiont gene expression in infected corals**

**Mariah Noelle Cornelio**

With the increase of nonbiological stressors such as the rising of global temperatures and growth of pollution, stony corals all over the world are becoming more susceptible to contracting fatal diseases that infects their tissue and can cause death. Through the advancement of modern-day techniques, machine learning strategies have made it easier to identify and classify possible symbionts living in the coral host that have the potential to harm the host when it becomes vulnerable to the environment around it. The methods used in this study include BBSplit, which is a tool used to map reads to multiple transcriptomes inputs and then determines the transcriptomes that matches the best. In addition, principal component analysis (PCA) and independent principal component analysis (IPCA) were utilized to determine variables affecting gene expression. Results from these may show the top symbiont and genes associated with it and how that plays into the effects of coral disease and immunity. As the dominant symbionts expressions are classified, understanding of the behavior of such species will be identified and classified in relation to disease events. These classification problems will lead to better understanding of symbionts in corals, and the overall protection of coral reefs worldwide.

**>>> Final abstract below <<<**

**Final abstract**

With the increase of nonbiological stressors such as the rising of global temperatures and growth of pollution, stony corals all over the world are becoming more susceptible to contracting fatal diseases that infects their tissue and can cause death. Diseases like stony coral tissue loss disease (SCTLD) and white plague (WP) have similar phenotypes, making it difficult to identify outbreaks. One way of identifying these diseases is by recognizing possible symbionts living in the coral host, which are influenced by the same disease events. These symbionts have the potential to harm the host when it becomes vulnerable to the environment around it. Through the advancement of modern-day techniques, machine learning strategies have made it easier to identify and classify possible symbionts and their response to these biological stressors. The methods used in this study include BBSplit, which is a tool used to map reads to multiple transcriptomes inputs and then determines the transcriptomes that matches the best. In addition, principal component analysis (PCA) and independent principal component analysis (IPCA) were utilized to determine variables affecting gene expression. Results from these may show the top symbiont and genes associated with it and how that plays into the effects of coral disease and immunity. As the dominant symbionts expressions are classified, understanding of the behavior of such species will be identified and classified in relation to disease events. These classification problems will lead to better understanding of symbionts in corals, and the overall protection of coral reefs worldwide.

**Add on poster:**

PCA takes all the gene counts as a whole and plots them to analyze trends while IPCA does the same thing except considers individual gene counts rather than looking at the whole. **(methods)**

Corals play an important role on the world’s matrix. They are a source of food and medicine for many organisms and can prevent erosion of coastlines. **(conclusion)**