## *Patterns of diversity in leaf endophyte communities across highly differentiated populations of* **Theobroma cacao**

**Background and Rationale:** Human activity is rapidly increasing carbon dioxide levels in the atmosphere, leading to a real threat of extreme temperature changes in tropical regions[1]. Drastic weather alterations are assumed to have severely negative impacts on agriculture, with low-input farming systems, like the chocolate tree (*Theobroma cacao*), predicted to suffer the most[1, 2]. Climate change is also expected to influence the distribution and range of infectious diseases, which poses a greater challenge to both predict and prevent[3]. However, the extent to which individuals can effectively adjust to these rapid changes remains unknown. The microbial communities found within a host presents one mechanism by which plants can respond to changing biotic or abiotic pressures[4,5,6]. **Therefore, the overarching research question I aim to address is: Can the microbiome provide a mechanism by which the host adaptability can be optimized?**

Research over the last fifteen years has shown that it is necessary to consider the plant microbiome in characterizing the adaptive potential of species. If we intend to characterize the adaptive potential of plant species in view of the rapidly changing environmental conditions, we are required to explore not only the genetic variation in plant populations, but also their microbiome composition and functional diversity. Studies have found evidence that a plant's natural microbiome plays a significant role its ability to become acclimated to a new or changing environment[4]. Additionally, there is evidence of microbial communities contributing to required immune responses and actively protecting against pathogenic invasions[5,6].

The combined evidence suggests that the microbiome provides critical, additive genetic variance, which ultimately modulates host responses to environmental and biotic factors. Yet, there is an insufficient understanding of the contribution that genetic diversity of the host plant provides in formulating their microbial composition. **I hypothesize that by better understanding the correlation between host and microbe genotypic variation among several cacao populations, there is potential to optimize the crop's overall adaptability.**

**Intellectual Merit:** As the incorporation of microbial DNA into the host's "pan-genome" is becoming more widely accepted, there is a need to understand how the host genome impacts its microbiome composition. Due to limitations in culture-dependent techniques, which omits the majority of microbial diversity in the samples[7], I will focus on genomic methods for taxonomic classification and functional characterization using next-generation sequencing (NGS) data.

Leaf samples have been collected and DNA sequenced from two germplasm collections; one located in Trinidad and the other in Costa Rica. I have access to sequence data for the genomes and metagenomes of 200 individuals across 11 distinct populations of *Theobroma cacao,* along with hybrids across populations, enabling me to characterize each individual microbiome. This unique dataset provides an opportunity to examine questions regarding microbial composition in cacao and allows me to generate an extended pan-genome reference set for *T. cacao*. Furthermore, with cacao being located in regions most susceptible to high climate variability, along with its natural tendency to be particularly susceptible to disease and declining production[8], we propose that cacao is an ideal system to explore the association between the endophyte composition and the genetic makeup of the host population.

**Aim #1: Characterize the alpha and beta diversity of leaf endophyte communities within and across *T. cacao* populations.** Alpha diversity describes the species richness within a particular region or ecosystem, while beta diversity explores the change in species composition

between ecosystems. Although there are different factors that can influence endophyte species composition, the impact of genetic differences across populations has not been evaluated.

We propose to use an efficient k-mer based classification system[9] to identify endophyte invertebrates, fungi, bacteria, archaea, and viral organisms from 200 cacao samples located at two distinct sites: Trinidad and Costa Rica. By comparing the alpha diversity of endophyte compositions across populations, I can then investigate correlations between genetic variation of host populations and community diversity. In addition, we propose to compare endophyte species composition across populations to test if there is increased similarity in microbiome composition across more related populations.

**Aim #2: Compare the phylogeny and relatedness of endophytes within and across populations.** By using the mapped reads of the NGS data, we can identify common homologous proteins across the species sets. We will then use these homologous regions to construct a comprehensive phylogeny for major dominant taxa shared across populations. I predict that there will be distinct correlations in microbial relatedness among metagenomic strains within each population for all the plant samples obtained from a common location. If there is co-evolution between the host and commensal species of the endophyte communities, we expect to find phylogenies that are consistent with the expectation of increased relatedness among microbes obtained from the same genetic group of *T. cacao* plants.

**Aim #3: Identify functional differences in endophyte microbiome across cacao populations.** Although plants typically produce a range of antimicrobial molecules for immunity, many hormonal compounds are recognized by a wide range of organisms. There are several studies suggesting that plants are selective in recruiting their specific microbiomes, often signaling microbes containing particular metabolic capabilities[10]. Given this evidence, it is likely that the functional profiles across differential cacao microbiomes varies in association with the host. In order to compose microbial proteomic catalogs, I will utilize a software that translates each possible open reading frame within individual NGS reads and aligns those results to amino acid sequences within a composed reference database. From there, the metabolic profiles of each microbial community can be compared for contrasting or relating functionalities.

**Broader Impacts:** As anthropogenic climate change continues to threaten the future of economically important crops, it is vital to determine mechanisms that maintain host adaptability to abiotic and biotic stressors. Not only can we recognize areas where the microbiome aids in environmental and biotic adaptation, but we could potentially identify physical associations between host and microbiome that are required for essential metabolic processes within cacao. With the microbiome playing a key role in plant health and immunity, this research will also broaden the scope of current knowledge on general host-microbe interactions. Support from the NSF-GRFP would allow further exploration on this project, permitting me to create additional opportunities for collaborations and learning experiences for new, motivated students. Moreover, due to the interdisciplinary nature of the project, I will be able to close the gap between biology, bioinformatics, and agricultural sciences by incorporating team members across each department. With strong connections to the agriculture, I am also able to present these novel scientific findings to organic farmers, connecting the biological sciences and agronomy.

**References: [1]** Sultan B (2016) *Frontiers in Plant Science*, 7:1262 **[2]** Schroth G (2016) *Science of The Total Environment* 556:231-241 **[3]** Brajesh S (2010) *Nature Reviews Microbiology* 8:779–790 **[4]** Zilber-Rosenberg I (2008) FEMS Microbiology Review 32(8):723-735 **[5]** Mendes R (2013) FEMS Microbiology Review 37:634-663 **[6]** Compant S (2005) AEM 71(9): 4951-4959 **[7]** Jackson C (2013) *BMC Microbiology* 13(1):274 **[8]** Lahive F (2013) *Agronomy for Sustainable Development* 39:5 **[9]** Breitwieser F (2018) *Genome Biology* 19(1):198 **[10]** Mendes R (2011) Science 332: 1097-1100