

# 1001 Mycobiomes: Mining public sequencing datasets to study plant–microbe interactions at scale

Colvin, Charles<sup>1\*</sup>, He, Sheng Yang<sup>1</sup>

<sup>1</sup>*Department of Biology, Duke University, Durham, NC 27708*

\*Presenting author: [charles.colvin@duke.edu](mailto:charles.colvin@duke.edu)

## Abstract:

Fungal communities inhabiting aboveground plant tissues influence plant health, disease dynamics, and ecosystem functioning in agricultural systems. Despite their importance, the factors that structure phyllosphere fungal communities remain poorly understood, and the contribution of host plant traits is still debated. Most prior studies have focused on bacterial communities, while fungal studies have relied largely on DNA-based amplicon sequencing. These approaches can introduce technical biases and cannot distinguish metabolically active fungi from dormant or dead cells.

Here we show that standard polyA-enriched RNA sequencing of plant leaf tissue can be repurposed to profile metabolically active phyllosphere fungal communities without targeted microbial enrichment. Using population-scale sequencing datasets from the *Arabidopsis* 1001 Genomes and Epigenomes projects, we detected millions of previously unexamined fungal transcripts. Although fungal reads comprise only a small fraction of total sequencing output, they enable retrospective, large-scale profiling of plant-associated fungal communities.

Integration of these fungal profiles with curated host genotype, gene-expression, and genome-wide methylation data revealed extensive host-associated variation in fungal abundance. Genome-wide association analyses (GWAS) identified numerous host loci associated with the abundance of individual fungal taxa, while transcriptome-wide association (TWAS) and expression quantitative trait locus (eQTL) analyses uncovered coordinated host gene-expression patterns associated with fungal presence.

Together, these results demonstrate that plant genetic and transcriptional variation are associated with the composition and abundance of active phyllosphere fungal communities. More broadly, this work highlights an underutilized opportunity to mine existing plant sequencing datasets for ecological and microbiome insights. By repurposing population-scale sequencing data, this approach provides a scalable framework for studying plant–microbe interactions and suggests that microbiome-associated traits may be genetically tractable targets for improving plant resilience.