Data Management Plan (PI: Austen Apigo)

I am committed to employing high standards of data integrity and transparency to maintain a reproducible workflow from data generation to publication. Data collection for non-sequencing based assays will be written on hard copy data sheets that will be regularly entered and digitized for long-term storage. Upon publication, final datasets, code, metadata and sequence files will be publicly shared through GitHub, Figshare and the National Center for Biotechnology Information's Sequence Read Archive (NCBI SRA). PI Apigo will be responsible for all aspects of data generation, management and release.

Description of Generated Data and Data Formats

This research will generate data related to field site environmental conditions from iButton sensors and precipitation gauges. Endophyte DNA and RNA sequence data, qPCR-based endophyte biomass data, chemical composition data from FTIR spectroscopy, endophyte enzymatic data and mass loss data will all be generated from pine litter bags. DNA and RNA sequence files will be stored as compressed FASTA or FASTQ files. I will generate R markdown (.Rmd) files and Python scripts to analyze bioinformatic and ecological data that will be accompanied with README text (.txt) files that explain the analytical pipeline for future users that wish to reproduce my work. Climate, qPCR, chemical composition, enzymatic and mass loss data will be stored as separate comma-separated values (.csv) files. Curricula for the coding workshop and climate science lesson plans for local middle schools will be stored as Rmarkdown (.Rmd) and Powerpoint (.pptx) files, respectively.

Data Storage and Protection

Hard copy field and lab data sheets will be digitized as scans, entered in .csv spreadsheets and will be stored on Google Drive, an external hard drive and UCI's BioCluster server. Climate, qPCR, chemical composition, enzymatic and mass loss data will be also stored on Google Drive, UCI's BioCluster server and an external hard drive that will be backed up weekly. DNA and RNA sequence data will be stored on an external hard drive, Illumina's cloud-based server, UCI's BioCluster server and uploaded to the NCBI's Sequence Read Archive. Rmarkdown and Python scripts for ecological and sequence analysis along with outreach curricula will be stored and publicly available on PI Apigo's GitHub.

Data Sharing and Dissemination

I am committed to open-source data practices with my sponsoring mentors (Dr. Treseder, Dr. Vázquez-Lobo and Dr. Chávez-Vergara) and the broader scientific community. Therefore, all collaborators will have access to Google Drive folders, sequencing files on Illumina's cloud-based server and coding scripts through PI Apigo's GitHub repositories. Final datasets, code, metadata and README files will be publicly available on Figshare with a project-specific DOI that will be linked on the corresponding publication.

Accountability: Roles and Responsibilities

PI Apigo will be responsible for all data generation in the field and lab, data management during the research project (digitizing spreadsheets, entering data), data storage on external hard drives and cloud-based servers (Google Drive, Illumina Basespace, UCI's BioCluster) and submitting data to publicly accessible repositories (GitHub, NCBI SRA, Figshare).