Plan Overview

A Data Management Plan created using DMPTool-Stage

Title: Investigating the Genetic Diversity of Pantoea ananatis strains endemic to Georgia

Creator: Spencer Stumpf

Affiliation: University of Georgia (uga.edu)

Principal Investigator: Spencer Stumpf

Data Manager: Spencer Stumpf

Funder: United States Department of Agriculture (usda.gov)

Funding opportunity number: 29159

Template: USDA - NIFA: National Institute of Food and Agriculture

Last modified: 07-01-2024

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customize it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

Investigating the Genetic Diversity of Pantoea ananatis strains endemic to Georgia

Expected Data Type

Describe the type of data (e.g. digital, non-digital) and how they will be generated (lab work, field work, surveys, etc.). Are these primary or metadata?

Quantitative and qualitative data will be ascertained for the purposes of this study. Experiments will be conducted in lab settings. Strains isolated from 1997 to 2015 from symptomatic onion, onion seed, various weed species, *Frankliniella fusca* throughout southern Georgia will be analyzed. DNA sequences from genes fusA, gyrB, leuS, pyrG, rplB, and rpoB will be gathered and analyzed by extracting genomic DNA from strains, amplifying specific genes, and submitting amplicons to Eurofins for Sanger Sequencing. Data will be visualized in Geneious for quality and phylogenetic analysis will be conducted using MEGA software. The Tamura-Nei model for assessing phylogeny will be used. Assays to assess ice nucleation and copper tolerance will be performed and repeated in the lab. For the ice nucleation assay, each strain will be replicated five times with a total of two experiments. In the event of mixed results (some water droplets solidify and some do not), a threshold of >6 droplets freezing will be required for strains to be considered ice nucleation positive. Each strain will be spread plated onto 3 plates of copper amended media one plate of non amendia media. Colonies must be present on all plates to be considered copper tolerant to account for a chance mutation on one plate not previously existent in the genome. Existing data such as isolation year, host, location, and pathogenicity will be considered for the purposes of assessing diversity. All of this data (excluding phylogenetic tree) will be arranged in an excel file. Ice nucleation and copper tolerance will be demarcated as simply a '+' or a '-'.

Data Format

For scientific data to be readily accessible and usable it is critical to use an appropriate community-recognized standard and machine readable formats when they exist. The data should preferentially be stored in recognized public databases appropriate for the type of research conducted. Regardless of the format used (notebook, samples, images, spreadsheet, etc.), that data set should contain enough information to allow independent investigators to understand, validate, and use the data.

Sequences in FASTA format will be submitted to GenBank for public access upon publication. Phenotypic data will be gathered using MS Excel (.xls), Portable Document Format (.pdf), Joint Photographic Experts Group (.jpg), sequence (fasta), and Tagged Image File Format (.TIFF). Experiment conditions will be saved in a .pdf format outlining the specifc procedural methods for replication.

Data Storage and Preservation

Scientific data should be stored in a safe environment with adequate measures taken for its long-term preservation. Applicants should describe plans for storing and preserving their data during and after the project and specify the data repositories, if they exist.

They should outline strategies, tools, and contingency plans that will be used to avoid data loss, degradation, or damage.

Culture suspensions are stored in 15% glycerol solution at -80 degrees Celsius and filed in the Molecular Diagnostic Lab and available for routine culture if necessary. Genomic DNA gathered will be submitted to Genbank for public access. The data backup and preservation must be maintained by the overseer of the Molecular Diagnostic Lab for smooth transition from one graduate student to the next.

Data Sharing and Public Access

Describe your data access and sharing procedures during and after the grant. Provide any restrictions such as copyright, confidentiality, patent, appropriate credit, disclaimers, or conditions for use of the data by other parties.

There will be no permission restrictions placed on data gathered form this project. Data will be made available upon publication. No issues should arise from our intention to share data publicly and data does not intend to be witheld.

Roles and Responsibilities

Who will ensure DMP implementation? This is particularly important for multi-investigator and multi- institutional projects. Provide a contingency plan in case key personnel leave the project. Also, what resources will be needed for the DMP? If funds are needed, have they been added to the budget request and budget narrative? Projects must budget sufficient resources to develop and implement the proposed DMP.

In order to successfully share information with collaborators, information will be deposited in a secure cloused-based access such as DropBox. Once we have completed our work, our data will be accessible to the public. We will use Genbank as the standard repository for datasets. Collaborators working on Whole Genome Sequencing will submit those results to Genbank, while I will submit the phenotypic data and housekeeping gene sequences. An umbrella project number will be assigned to all data to quickly access information regarding the project.

Monitoring and Reporting

Successful projects should monitor the implementation of the DMP throughout the life of the project and after, as appropriate. Implementation of the DMP should be a component of annual and final reports to NIFA (REEport) and include progress in data sharing (publications, database, software, etc.). The final report should also describe the data that was produced during the award period and the components that will be stored and preserved (including the expected duration) after the award ends.

We acknowledge that data gathered from this project will be monitored as specified by NIFA. Principal investigators will be review and revise this data management plan.