

Plan Overview

A Data Management Plan created using DMPTool

DMP ID: <https://doi.org/10.48321/D11G6C>

Title: Understanding a Natural Disease Tolerance Phenomenon Using a Metabolic Modeling-Based Approach

Creator: James Borneman - **ORCID:** [0000-0003-0661-4970](https://orcid.org/0000-0003-0661-4970)

Affiliation: University of California, Riverside (ucr.edu)

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

Funding opportunity number: A1141

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

Project abstract:

Huanglongbing (HLB) is a devastating citrus disease that threatens the existence of the citrus industry in the United States and worldwide, and there currently are no effective and sustainable strategies to manage HLB. This project is to use a novel integrated experimental and computational approach to understand a natural HLB disease tolerance phenotype, and then translate that knowledge into HLB management solutions. More specifically, the Long-Term Goal of this proposed project is to create citrus trees that are resistant or tolerant to Huanglongbing (HLB) disease, and/or to create effective prophylactic or curative HLB treatments. The Objective of this Seed Proposal is to use metabolic models and multi-omics data to identify the metabolic pathways and molecules in citrus and its associated microbes that are responsible for the Survivor Tree Phenotype in Florida. Survivor Trees exhibit a very slow rate of decline even though they are in orchards where most of the trees exhibit the normal rapid HLB decline. After 4.5 years of an ongoing study, we have identified trees that have only declined by 0.5 in a 0 to 5 disease rating system. Since this proposed project is to identify the metabolic pathways and molecules in citrus and its associated microbes that are responsible for the Survivor Tree Phenotype, it specifically addresses one of the Program Area Priorities of the Pests and Beneficial Species Program, which is "Biotic and abiotic factors affecting the abundance or spread of agriculturally-important plant pests, disease vectors, or beneficial species relevant to pest management."

Start date: 11-01-2021

End date: 10-31-2023

Last modified: 05-10-2023

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

Understanding a Natural Disease Tolerance Phenomenon Using a Metabolic Modeling-Based Approach

Expected Data Type

Describe the type of data (e.g. digital, non-digital), how it will be generated, and whether the data are primary or metadata.

- **Research examples include: lab work, field work and surveys.**
- **Education examples include: number of students enrolled/participated, degrees granted, curriculum, and training products.**
- **Extension examples include: outreach materials, number of stakeholders reached, number of activities, and assessment questionnaires.**

Primary non-digital and digital data generated will come from computational work and lab work. Data will be diverse and include hand-written observations, images, videos, nucleotide and protein sequences, genome sequences, genome annotations, metabolic models, metabolomics and transcriptomics data, as well as genetic and phenotypic data. Metadata will include collaborating institution, researcher, date, experimental methods, HLB disease severity rating, Survivor or Non-Survivor Tree status, conditions, locations, and digital file names associated with individual experiments. We will also incorporate the FAIR Framework that can be found at this link – <https://www.go-fair.org/fair-principles>.

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. If the data will be managed in domain-specific workspaces or submitted to public databases, indicate that their required formats will be followed. Regardless of the format used, the data set must contain enough information to allow independent use (understand, validate and use) of the data.

Data formats will be non-proprietary, unencrypted, machine-readable, recognizable by the scientific community, and interoperable among platforms and applications (e.g., TXT, DOC, XML, PDF, CSV, TIFF, and JPEG). In the unlikely event of using a proprietary data format, clear instructions for data access and software source (i.e., software name, version, and company) will be included as a simple text file in the data directory. The format of the metabolic models follows the guidelines presented at BIGG Models (<http://bigg.ucsd.edu/>), which is the recognized standard in this field. Non-digital data will be digitized by scanning or manual input. Data submitted to public databases (e.g. NCBI), meet all format requirements. A University of California librarians specializing in data services is available to assist with annotation of research data, formatting, and metadata workflows for submission to archiving and for use by the scientific community. We will also incorporate the FAIR Framework that can be found at this link – <https://www.go-fair.org/fair-principles>.

Data Storage and Preservation

Data must be stored in a safe environment with adequate measures taken for its long-term preservation. Applicants must describe plans for storing and preserving their data during and after the project and specify the data repositories, if they exist. Databases or data repositories for long-term preservation may be the same that are used to provide Data Sharing and Public Access. Estimate how much data will be preserved and state the planned retention period. Include any strategies, tools, and contingency plans that will be used to avoid data loss, degradation, or damage.

Labs using digital notebooks and LIMS systems will be backed up on hard drives. Laboratory computers are routinely backed up on hard drives and a cloud system. Lab members' personal computers are backed up monthly on an external hard drive. Datasets for publications will be submitted in Dryad for curation and preservation. The project will also use the University of California Curation Center (UC3) Merritt Repository Service to manage, archive, and share digital content. Merritt provides public access via persistent URLs, tools for long-term data management, and permanent storage options, with built-in contingencies for disaster recovery. All data will be preserved for a minimum of five years after project completion. High-value genome-edited plants will be maintained in greenhouses and tissues sent for long-term storage in the USDA cryopreservation facility at Fort Collins, Colorado. Products including plasmids, nucleic acids, will be retained for at least three years and often longer by storage at -20C and -80C. Plasmids of potential general use (e.g. CRISPR-based genome editing vectors and citrus-specific promoters) will be deposited at Addgene. We will also incorporate the FAIR Framework that can be found at this link – <https://www.go-fair.org/fair-principles>.

Data Sharing and Public Access

Describe your data access and sharing procedures during and after the grant. Name specific repositories and catalogs as appropriate. include a statement, when applicable, of plans to protect confidentiality, personal privacy, proprietary interests, business confidential information, and intellectual property rights. Outline any restrictions such as copyright, confidentiality, patent, appropriate credit, disclaimers, or conditions for use of the data by other parties.

During the grant, data will be deposited in Merritt (as described above) which allows public sharing. Research data will also be cataloged in the Ag Data Commons as required. Final published data will be made publicly available. PDs will deposit papers published without open access in the UC “eScholarship” digital repository. All publications and presentations acknowledge USDA-NIFA support. Datasets on genomes, gene expression, or metabolomics profiling will be available through NCBI or Dryad. Progress and final reports will include a persistent identifier that provides links to the full text. All final data associated with the project will be retained for a minimum of five years after project conclusion or any project publication. If requested, data will be shared with qualified parties, as long as such a request does not compromise intellectual property interests or interfere with a publication. All members of the research team will make presentations at stakeholder events and scientific conferences. We will also incorporate the FAIR Framework that can be found at this link – <https://www.go-fair.org/fair-principles>.

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.

PD James Borneman with assistance from Co-PD Karsten Zengler (or if needed their replacements, which would likely be the other project participants), will provide oversight of all data management activities and responsibilities. No funds will be needed for data management because will be using all public domain databases, software and/or services. All members of the project's research team with access to data will receive instruction in the Responsible Conduct of Research, which includes proper maintenance of laboratory notebooks. We will also incorporate the FAIR Framework that can be found at this link – <https://www.go-fair.org/fair-principles>.

Planned Research Outputs

Dataset - "Annotated Citrus Genomes"

1. Annotated Citrus Genomes

**Model representation - "Metabolic Models will be archived at BIGG Models
<https://bigg.ucsd.edu>"**

Metabolic Models will be archived at BIGG Models <https://bigg.ucsd.edu>

Dataset - "Omics Datasets"

Omics Datasets

Text - "Putative Molecules and Pathways that explain the Survivor Tree Phenotype - Provided in Publications and at <https://escholarship.org>"

Putative Molecules and Pathways that explain the Survivor Tree Phenotype - Provided in Publications and at <https://escholarship.org>

Planned research output details

Title	Type	Anticipated release date	Initial access level	Intended repository(ies)	Anticipated file size	License	Metadata standard(s)	May contain sensitive data?	May contain PII?
Annotated Citrus Genomes	Dataset	2023-10-30	Open	NCBI	800 MB	Creative Commons Attribution Non Commercial 4.0 International	None specified	No	No
Metabolic Models will be archived at BIGG Models h ...	Model representation	2023-10-30	Open	None specified	1,000 MB	Creative Commons Attribution Non Commercial 4.0 International	None specified	No	No
Omics Datasets	Dataset	2023-10-30	Open	NCBI	50 GB	Creative Commons Attribution Non Commercial 4.0 International	None specified	No	No
Putative Molecules and Pathways that explain the S ...	Text	2023-10-30	Open	None specified	1 MB	Creative Commons Attribution Non Commercial 4.0 International	None specified	No	No