



WASHINGTON STATE
UNIVERSITY

Finding Genes and High Throughput Imagery Features Predictive of Resistance to Aphanomyces (Root Rot)

Antonio Franco, Dr. Stephen Ficklin



Overview

- Background
- Research objectives
- Results
- Future Discussion



Lentils

- Lentils are legumes
- High in nutritional value
- Produced in areas with little rainfall
- Canada produces the most lentils
- Used as rotation crops
- Fix nitrogen



Big, A. (2020). *World's top lentil producing countries*. AtlasBig. <https://www.atlasbig.com/en-us/countries-lentil-production>.

Aphanomyces euteiches

- Fungus that causes root rot
- Symptoms appear between 7 to 14 days
- We lack a lentil variety with high resistance
- Chemical control is ineffective
- Development of resistant varieties is urgently needed



Research Objective

- Four Objectives
 1. Identify genes whose expression can predict *Aphanomyces* resistance
 2. Identify imagery features predictive of *Aphanomyces* resistance
 3. Identify genes whose expression predicts imagery features
 4. Identify genes and imagery features that together are predictive of resistance
- Purpose
 - Breeders can use both genes and imagery features as markers in breeding
 - Markers can help breeders focus on better varieties



Experimental design

- Sampling
 - 9 varieties: 3 low resistance, 3 moderate, 3 higher
 - 3 replicates (repeats) for statistical power
 - 2 treatments: control, inoculated
 - Two time points: 6 hrs, 24 hrs



Experimental design

- Measurements
 - RGB data(standard imagery)
 - Hyperspectral imagery
 - Gene expression measurements (49097genes)
 - Resistance levels
 - Phenotype data (e.g., dry weight)



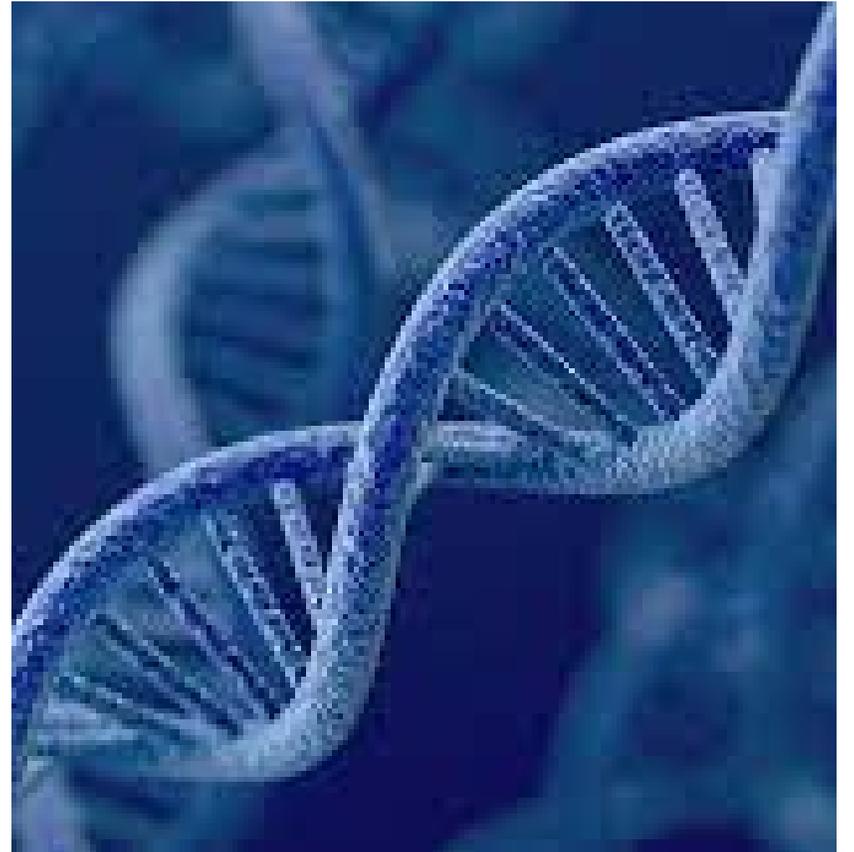
Experimental design

- Analytical approach
 - Differential gene expression
 - Supervised machine learning



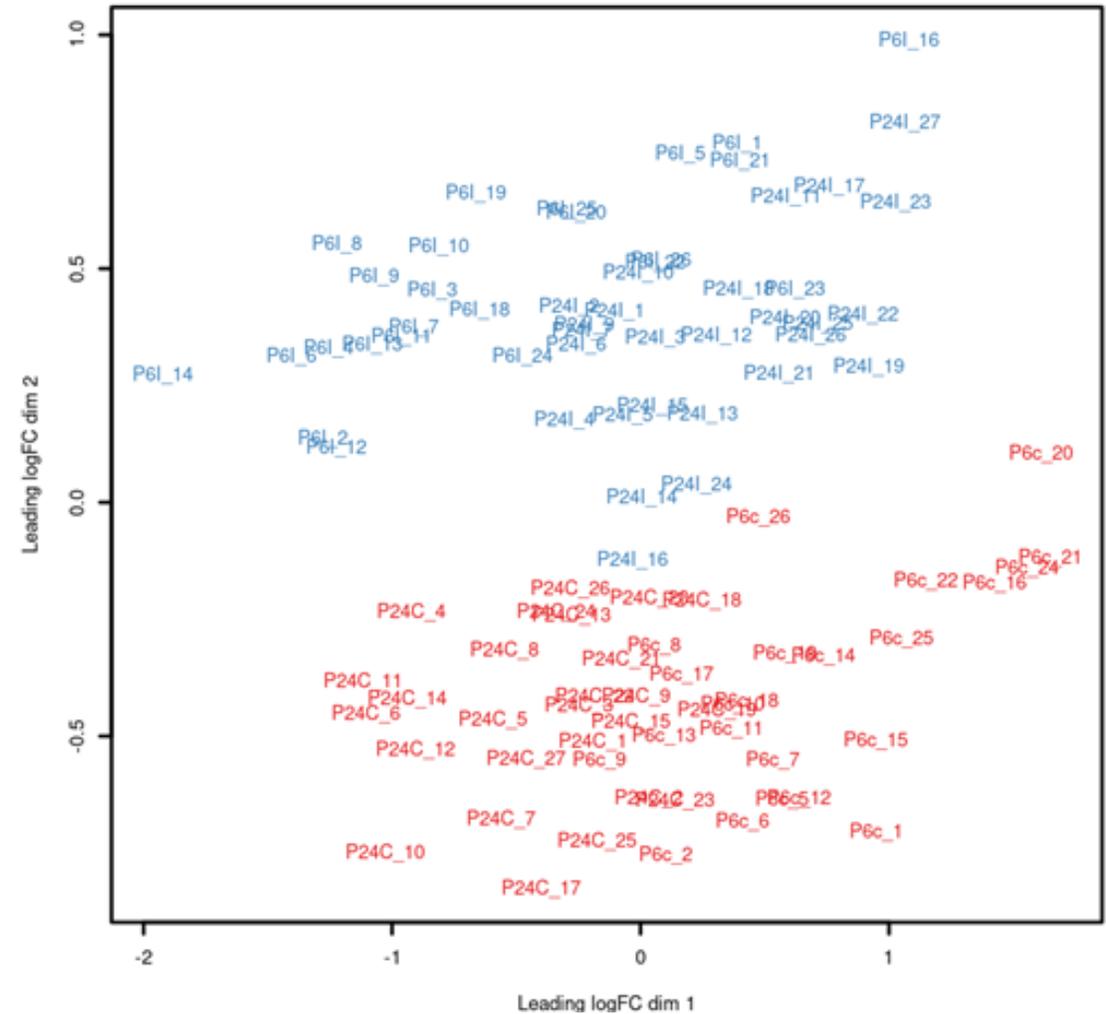
Accomplishment

- Learned Python programming (new language for me)
 - Took majority of the time
- Learned to use basic GIT (managing software and data)
- Performed objective one
 - “Identify genes whose expression can predict *Aphanomyces* resistance”



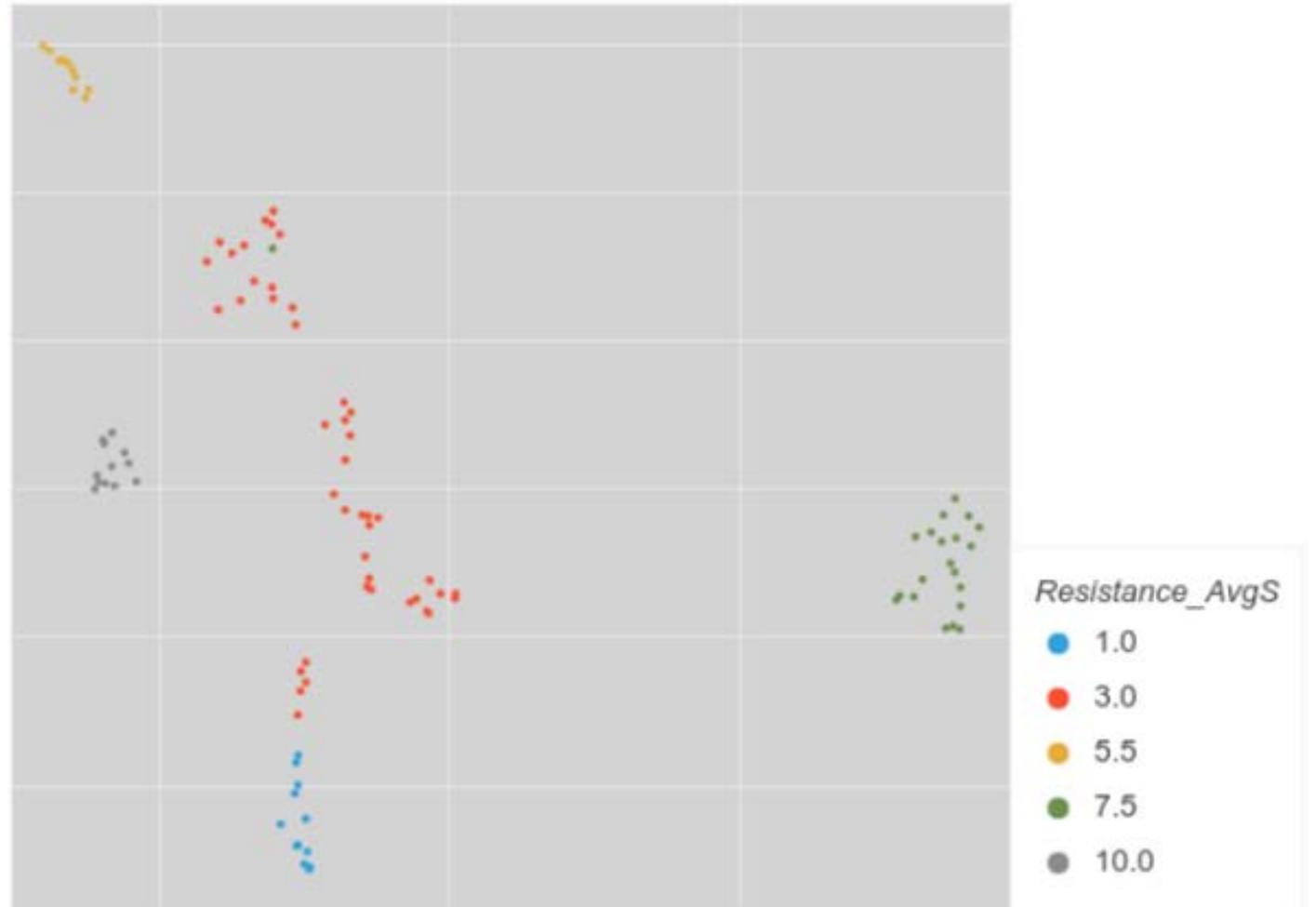
Results-Differential Gene Expression

- 1230 genes were selected
 - Different between treatments
 - Different between time points
 - Biological function is “response to stimuli”



Results- Supervised Machine Learning

- Supervised machine learning
 - Random Forest
 - Boruta
- Started with 1230 genes
- 149 predictive genes
- Predictive of resistance



Future Work

- Complete objective 2 through 4
- Other machine learning tools
- Validation of the genes





United States Department of Agriculture
National Institute of Food and Agriculture

Acknowledgements

- Mentorship
 - Dr. Stephen Ficklin
 - Dr. Tyler Biggs
- Funding
 - McNair
 - USDA REEU project: 1021788
 - CAHNRS Emerging Research Initiative Project
- Data sources
 - Gene expression and phenotype: Yu Ma, Rebecca McGee
 - Phenomics: Afef Marzougui, Sindhuja Sankaran





Summer 2021

**FACT: Research Experience for Undergraduates on Phenomics Big
Data Management**



**United States Department of Agriculture
National Institute of Food and Agriculture**

**Research and Extension Experiences for Undergraduates (REEU)
Project: 1021788**