

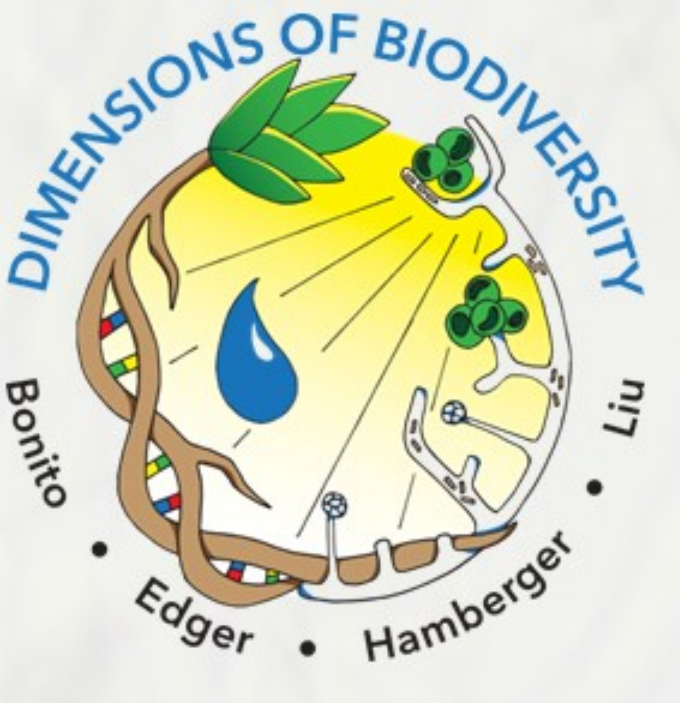
# Use of Hyphal Image Analysis and Machine Learning to Classify Mucoromycota Soil Fungal Isolates

Julian Liber (liberjul@msu.edu)<sup>1</sup>, Elizabeth Gall<sup>1</sup>, Natalie Vande Pol<sup>1</sup>, Devin Silvia<sup>2</sup>, and Gregory Bonito<sup>1</sup>

<sup>1</sup>Department of Plant, Soil, and Microbial Sciences, Michigan State University <sup>2</sup>Department of Computational Mathematics, Science, and Engineering, Michigan State University

## Introduction

- Fungi in the group Mucoromycota are valuable
  - Plant growth promotion (Osorio and Habte 2001)
  - Industrial production of lipids (Sakuradani et al. 2013)
  - Study system of bacterial endosymbiosis (Desirò et al. 2018)
- Isolates obtained from culturing of soil samples
- Computational methods including Fast Fourier Transform (FFT) and machine learning are useful classification tools (Mennitt, Sherrill, Frstrup, 2014; Orlov et al., 2008)

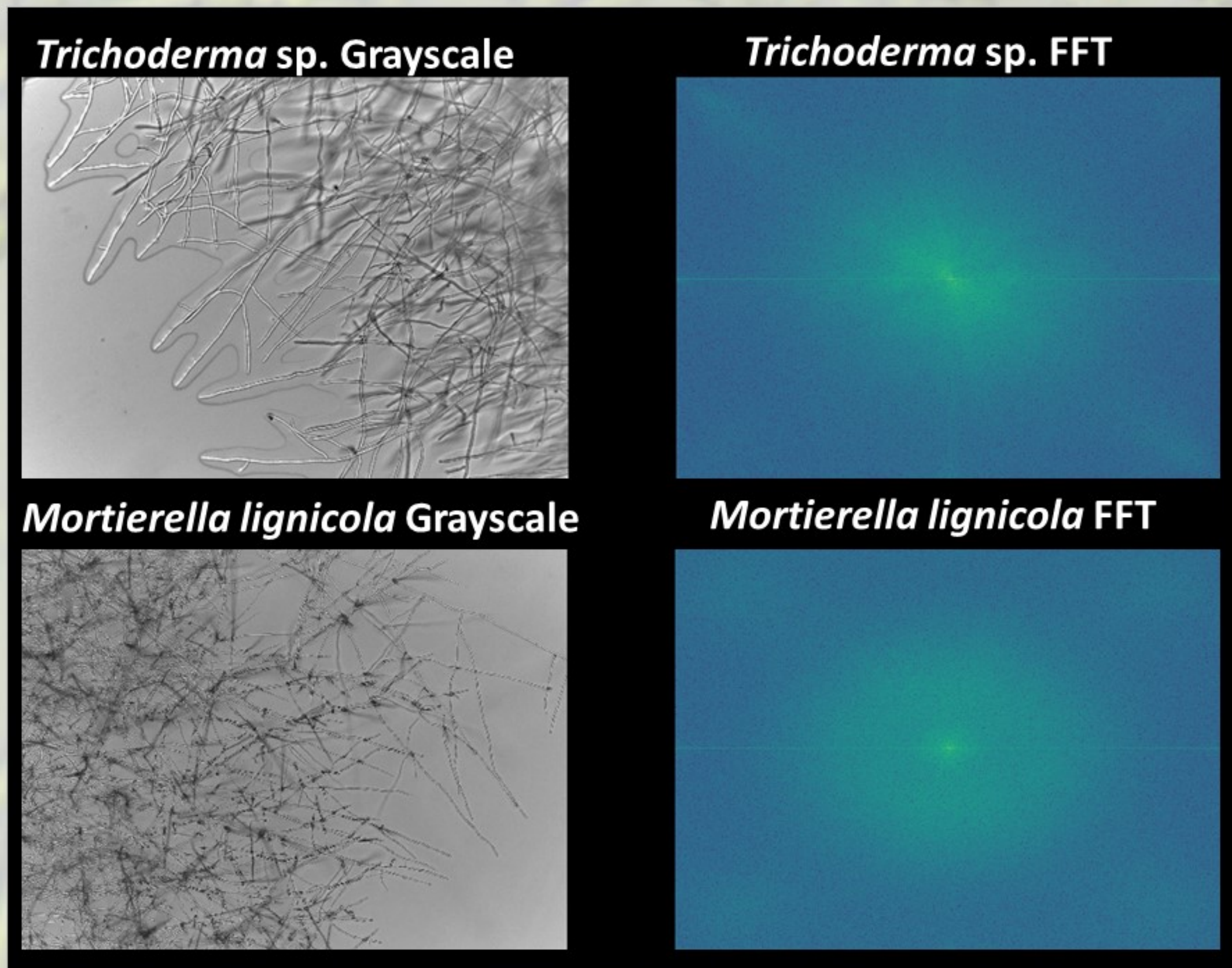


## Problem

- Identification of fungal isolates is slow, expensive
- Fungal morphological identification often relies on features absent in culture

## Design Goal

- Hyphae have few identifiable features, but have patterns which may be classifiable
- Create an image classifier using a database of sequence verified images to separate Mucoromycota from other fungi
- Obtain a usefully high precision and recall
- Implement a script capable of rapid classification using only hyphae

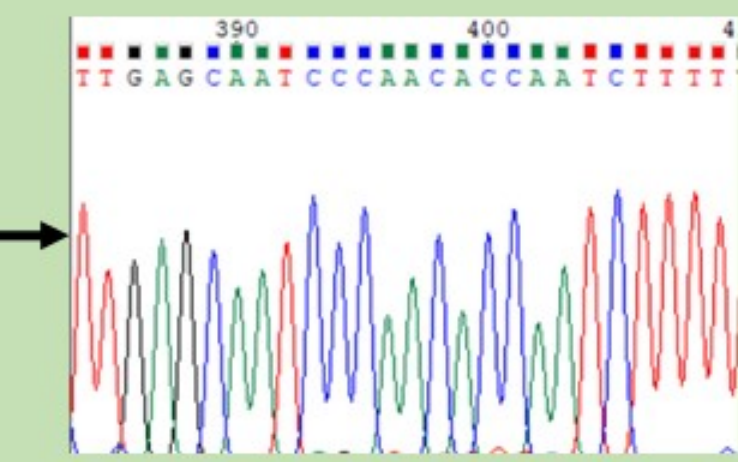
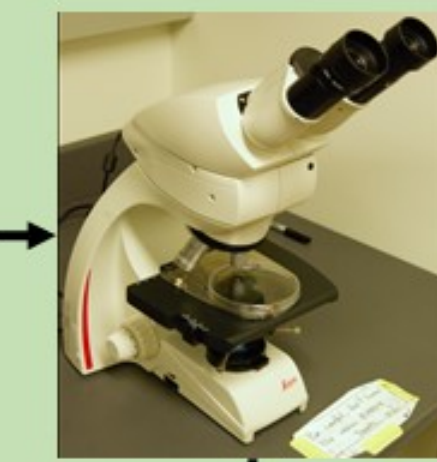
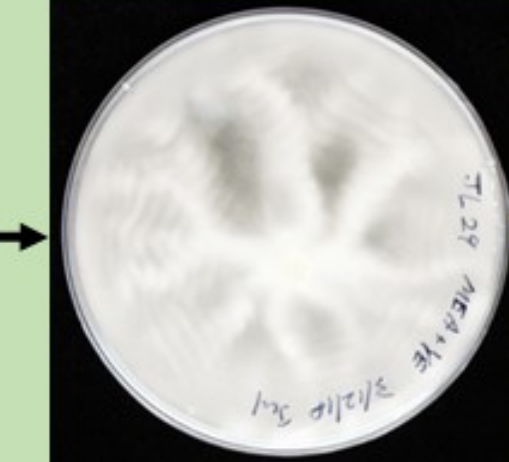


**Figure 1. Grayscale and Fast Fourier Transforms (FFTs) of Brightfield Hyphal Images.** FFT captures frequencies in the image, with low frequency (more structural) features in the center and high frequency (more detail) near the perimeter.

## Methods

### 1) Sample Preparation

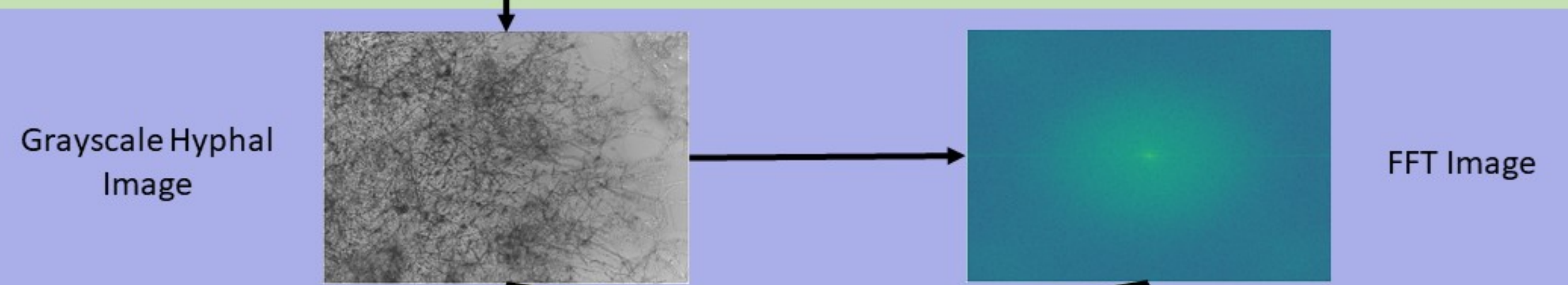
Isolates are obtained from using soil-planting and shrimp-baiting, grown on malt extract agar with yeast extract. Brightfield micrographs are obtained of isolates in Petri dishes at 100X.



**Identity Confirmation**  
DNA is extracted from isolates and the ITS1 region is amplified. Sanger sequencing and BLAST (Altschul et al. 1990) are used to confirm genus/species identity.

### 2) Data Processing

Images are processed with Python 3.6 using Numpy, Pandas, Matplotlib, and SKLearn. First, they are converted to grayscale and Fast Fourier transformed (FFT) then linearized. The images are randomly split 1:1 into training and testing data.



### 3) Feature Selection

Using Principle Component Analysis (PCA) or Random Forest Feature Selection (RFFS) important components or pixels are identified.



### 4) Classification

Processed data sets, with a reduced number of features, are classified with Random Forest (RF), Artificial Neural Network (ANN), or Support Vector Machine (SVM) classifiers.

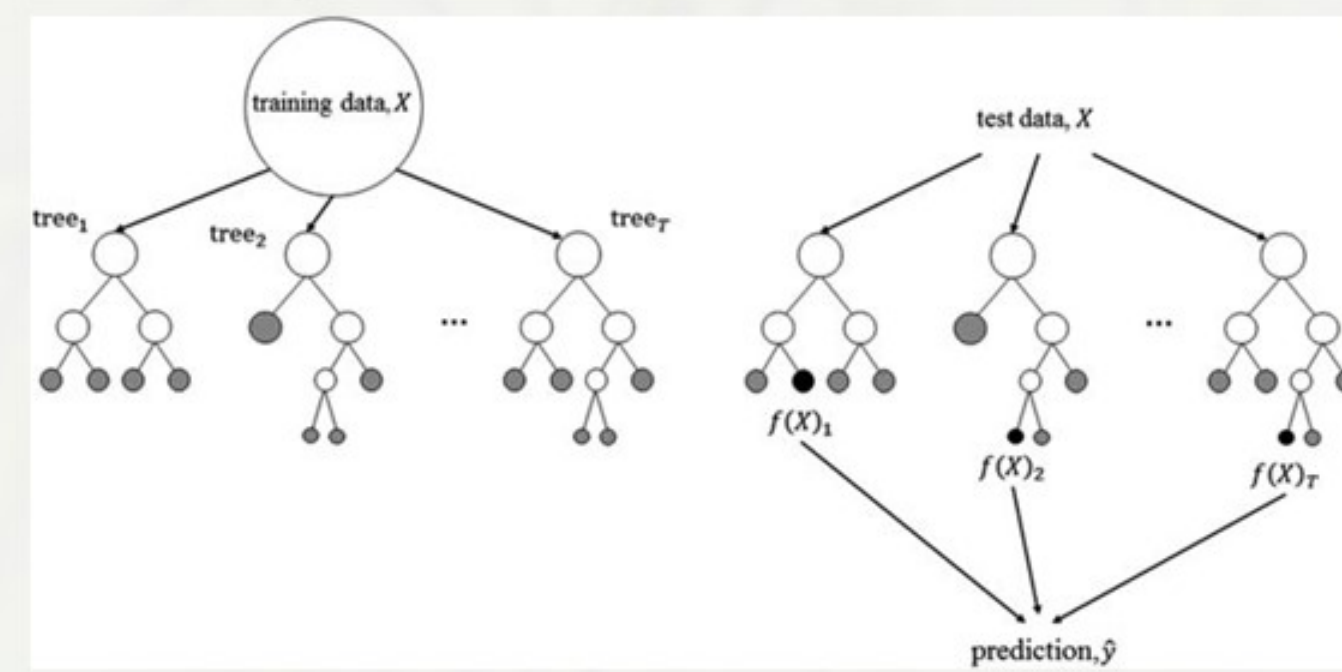


### 5) Evaluation

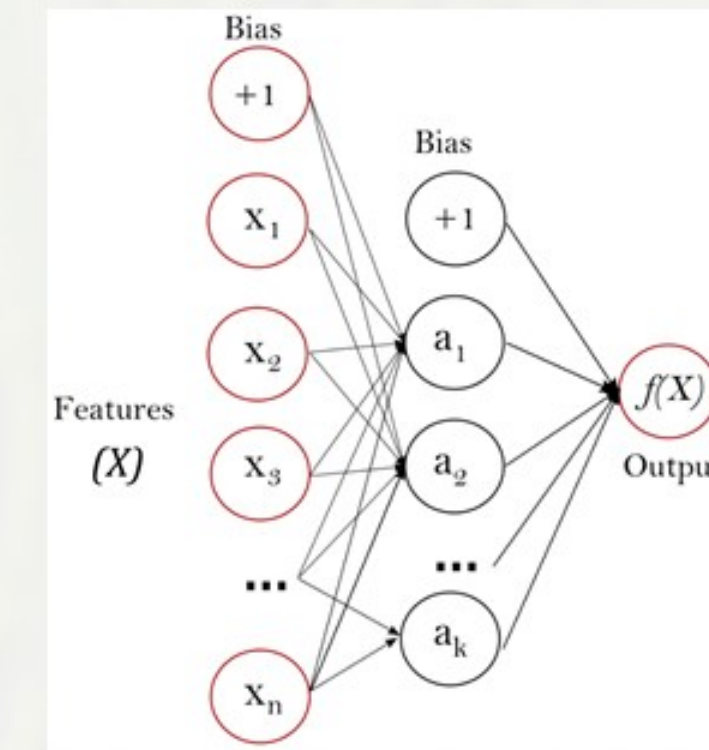
Classifiers developed on the training dataset are evaluated on the testing set. Weighted F-score is the primary metric, to ensure that incorrect classification of rare groups is penalized.



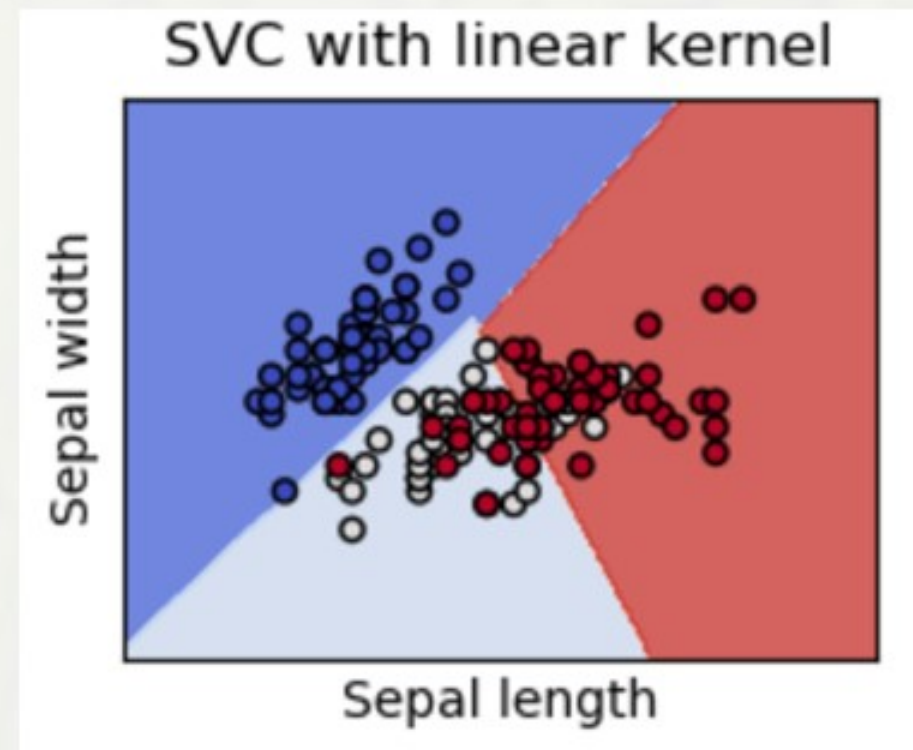
### A) Random Forest (RF)



### B) Artificial Neural Network (ANN)



### C) Support Vector Machine (SVM)



**Figure 2. Classification Methods.** A) Random Forest classifiers operate by constructing a series of decision trees and compiling results into a class prediction. B) Artificial Neural Networks use a neuron-like method of activation based on input data to produce an output. C) Support Vector Machine attempts to linearly separate the data into groups. Figures retrieved from Mennitt, Sherrill, Frstrup, 2014, and SKLearn documentation.

## Results

- 568 images included, 398 Mucoromycota
- Classification methods varied in success, measured by weighted F-score, compared using ANOVA
  - Mucoromycota vs. other fungi ( $F = 29.42$ ,  $p = 3.25e^{-12}$ )
  - Mucoromycota vs. Mortierellomycota ( $F = 326$ ,  $p < 2e^{-16}$ )
- Best method combines FFT, PCA, and RF

## Conclusion

- Hyphal characteristics are sufficient for high-level classification of fungal isolates
- Greater than 90% F-score is sufficient to inform collection process
- Creation of useable classifiers and pipeline for creating classifiers
- Future Work
  - Improve phylogenetic resolution
  - Apply to detecting effects of environment on fungus as method of rapid phenotyping
  - Compare endobacteria hosting and non-hosting strains

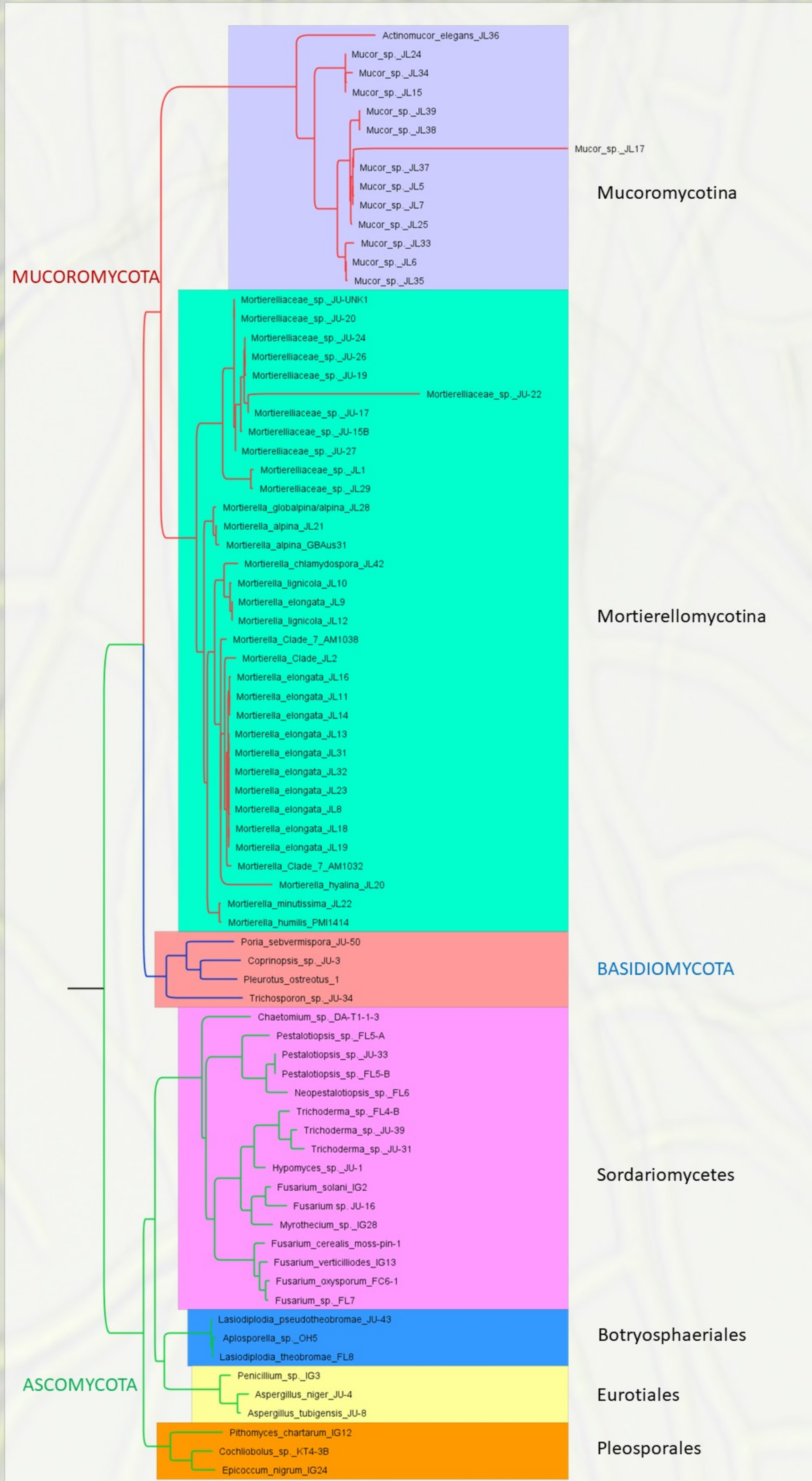
## Acknowledgements

- Alessandro Desirò – mentorship and training
- Christina Azodi – machine learning help
- MSU Honors College Professorial Assistantship Program
- Research in the Bonito lab is supported through NSF - DEB 1737898 and MSU - AgBioResearch NIFA project MICL02416

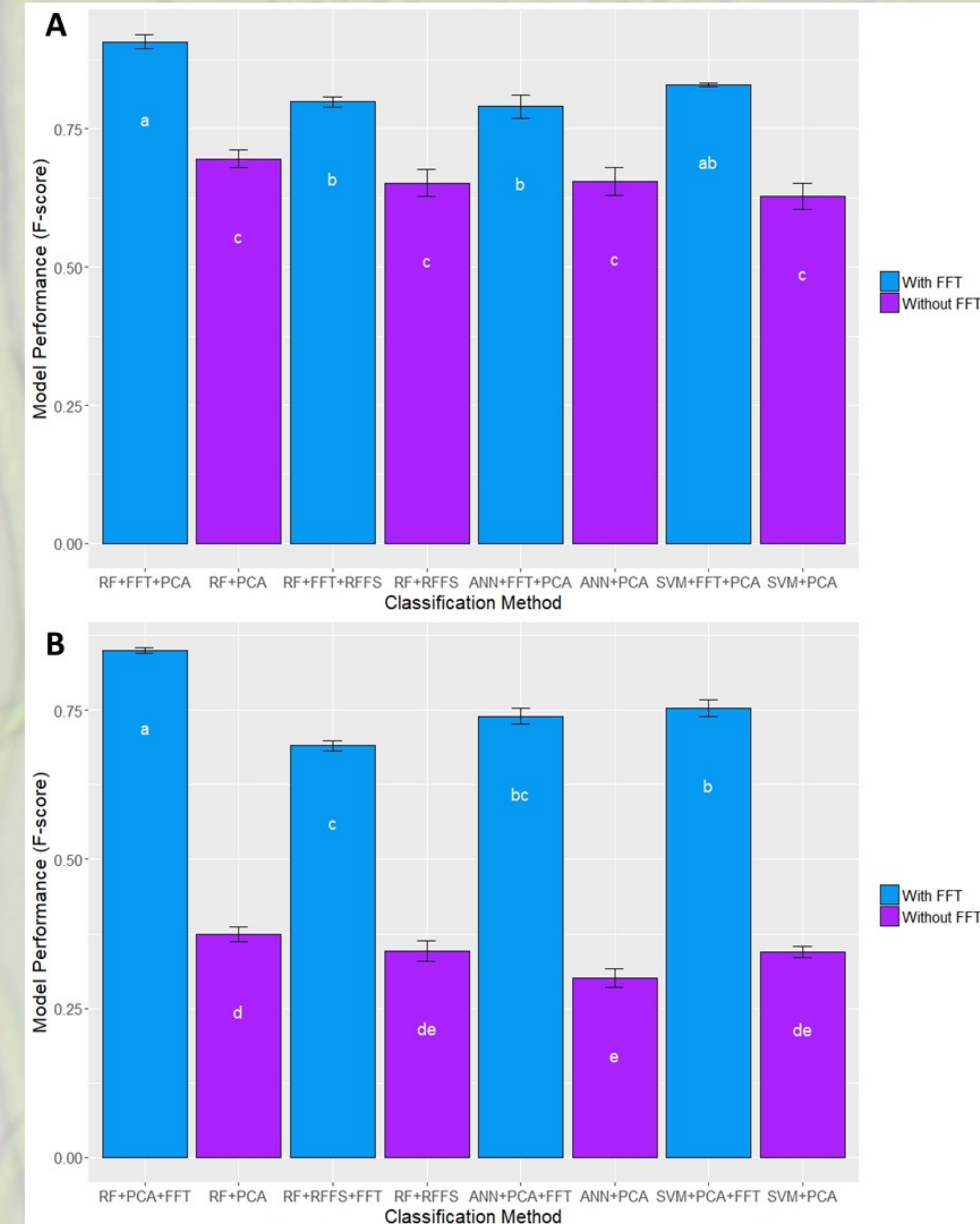
## References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *Journal of molecular biology.* 1990 Oct 5;215(3):403-10.
- Desirò A, Hao Z, Liber JA, Benucci GM, Lowry D, Roberson R, Bonito G. Mycoplasma-related endobacteria within Mortierellomycota fungi: diversity, distribution and functional insights into their lifestyle. *The ISME journal.* 2018 Feb 23:1.
- Mennitt D, Sherrill K, Frstrup K. A geospatial model of ambient sound pressure levels in the contiguous United States. *The Journal of the Acoustical Society of America.* 2014 May;135(5):2746-64.
- Orlov N, Shamir L, Macura T, Johnston J, Eckley DM, Goldberg IG. WND-CHARM: Multi-purpose image classification using compound image transforms. *Pattern recognition letters.* 2008 Aug 1;29(11):1684-93.
- Osorio NW, Habte M. Synergistic influence of an arbuscular mycorrhizal fungus and a P solubilizing fungus on growth and P uptake of *Leucaena leucocephala* in an Oxisol. *Arid Land Research and Management.* 2001 Jan 1;15(3):263-74.
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J. Scikit-learn: Machine learning in Python. *Journal of machine learning research.* 2011;12(Oct):2825-30.
- Sakuradani E, Ando A, Shimizu S, Ogawa J. Metabolic engineering for the production of polyunsaturated fatty acids by oleaginous fungus *Mortierella alpina* 1S-4. *Journal of bioscience and bioengineering.* 2013 Oct 1;116(4):417-22.

Code available at: [https://github.com/liberjul/Hyphal\\_Image\\_Classification](https://github.com/liberjul/Hyphal_Image_Classification)



**Figure 3. Maximum Likelihood Phylogeny of Fungal Strains.** Most micrographs were obtained from Mucoromycota isolates, but a diverse range of isolates were used to ensure validity of the classifier.



**Figure 4. Performance of Classification Methods.** Performance significantly varies between methods, and letter annotations are different at  $p < 0.05$  with Tukey's HSD test. Error bars indicate standard error of classifications with n trials. The combination of Random Forest, Fast Fourier Transform, and Principal Component Analysis provides the best classification. A) Classification between Mucoromycota and other fungi (n=5). B) Classification between Mucoromycota vs. Mortierellomycota fungi (n=10).