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

Julian Liber[1], Elizabeth Gall[1], Natalie Vande Pol[1], Devin Silvia[2], and Gregory Bonito[1]

[1]Department of Plant, Soil, and Microbial Sciences; [2]Department of Computational Mathematics, Science, and Engineering

Introduction

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

Figure 1. Phylogeny of the fungal kingdom, adapted from Choi and Kim, 2017. Mucoromycota (within Zygomycota) is a small portion of the fungal kingdom but its members are sought for their economic and scientific value.

Problem

- Identification of fungal isolates is slow, expensive
- Fungal morphological identification often relies on features absent in culture
- Hyphae have few identifiable features, but have patterns which may be classifiable

Design Goal

- 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
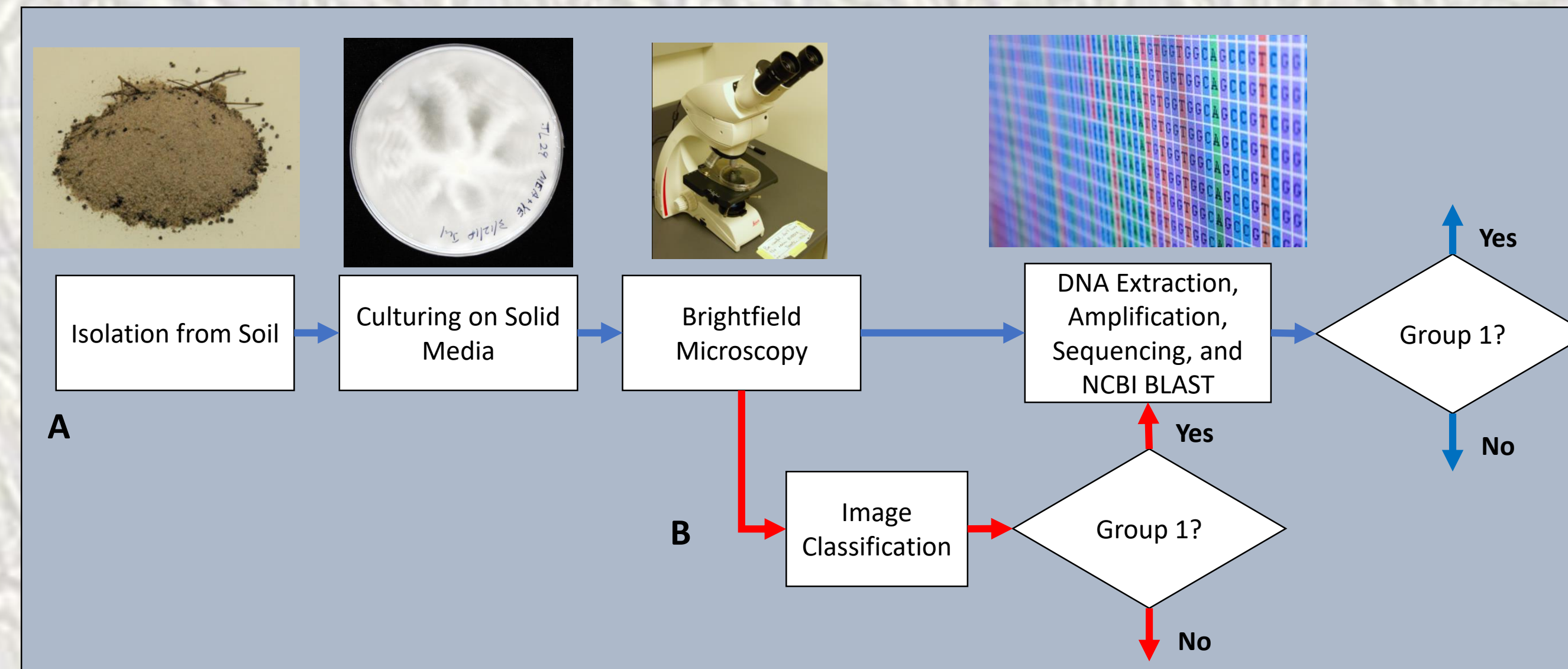

Figure 2. Conventional and proposed method of isolating and identifying soil fungi. Conventional method indicated by blue arrows, proposed addition in red. Mucoromycota is considered Group 1 for this figure. A) Once pure isolates are obtained, repeated plating on solid media and microscopy followed by DNA extraction, PCR amplification, Sanger sequencing, and NCBI BLAST are used to identify isolates. Isolates are either included for further analysis or excluded from further study. B) An additional gatekeeping step of image classification aims to reduce expense and improve speed of identification.

Methods

- Fungal isolated obtained with soil-plating and shrimp-baiting, grown on malt extract agar supplemented with yeast extract
- Photographs of hyphae in Petri dish obtained at 100X with brightfield microscopy
- Data manipulation and machine learning in Python 3.6
  - Packages included Numpy, SciKit-Learn, and Matplotlib
- Fast Fourier Transform (FFT) performed to capture pattern
- Feature Selection by Principal Component Analysis (PCA) or Random Forest (RF) feature selection (RFFS)
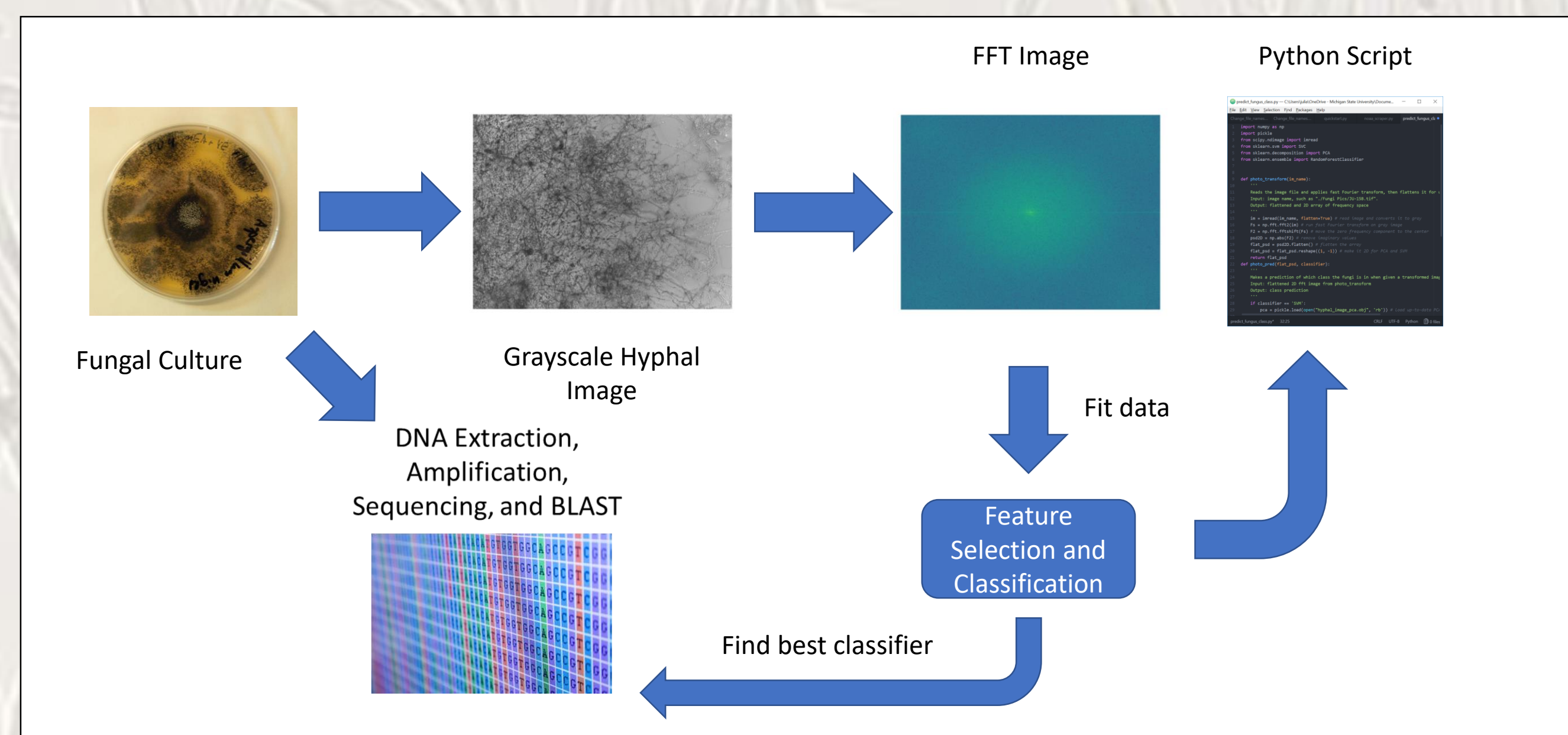- Classification with Support Vector Machine (SVM), RF, or Artificial Neural Network (ANN)

Figure 3. Construction of image classifier. Images obtained with the microscope are converted to grayscale and processed with FFT. Multiple classification algorithms were tested for ability to fit and predict the data. The best performing algorithms were constructed into a classification script.

Results

- 568 images included, 398 Mucoromycota
- Classification methods varied in success, measured by weighted F-score (ANOVA, $F = 29.42$, $p = 3.25e^{-12}$)
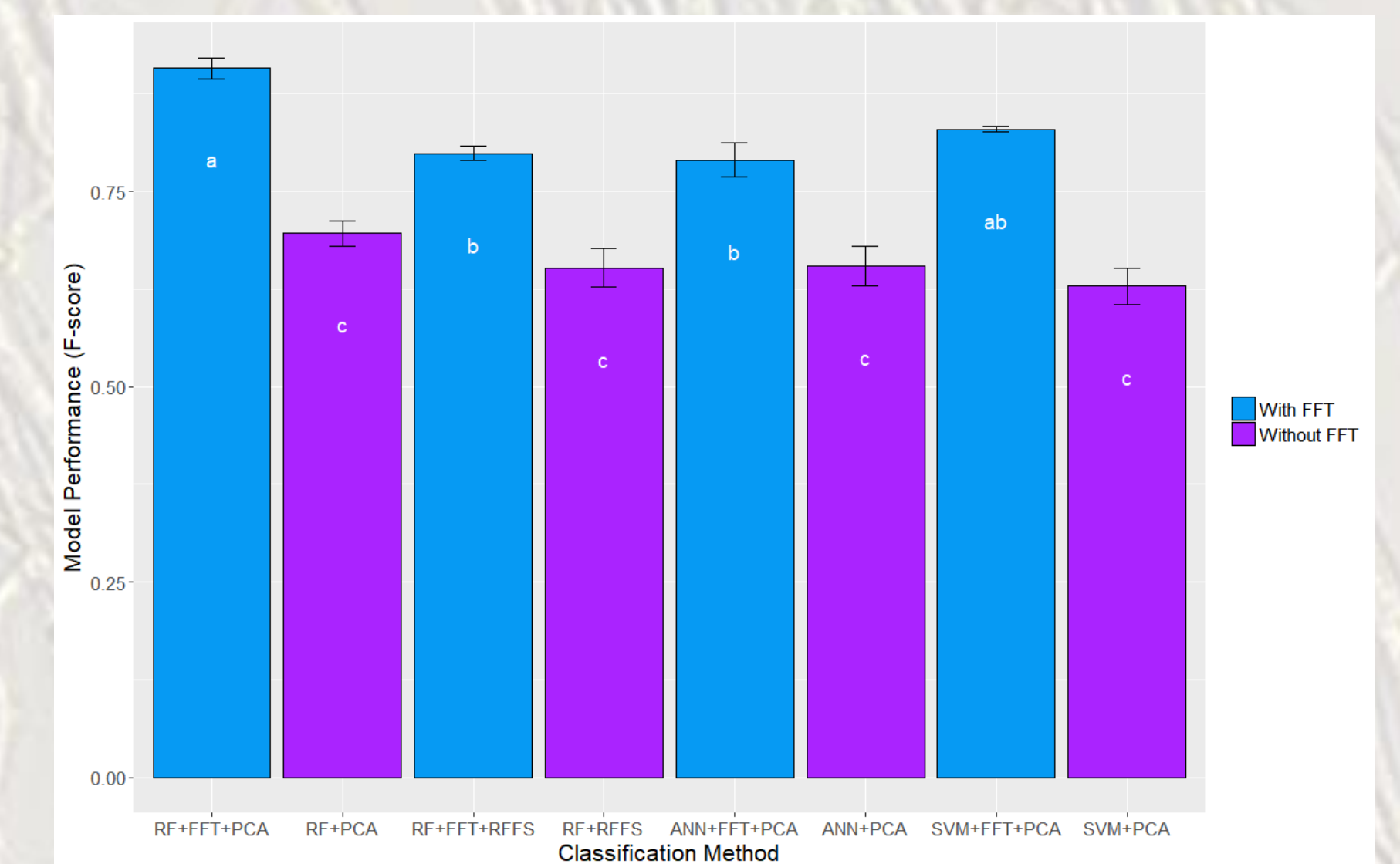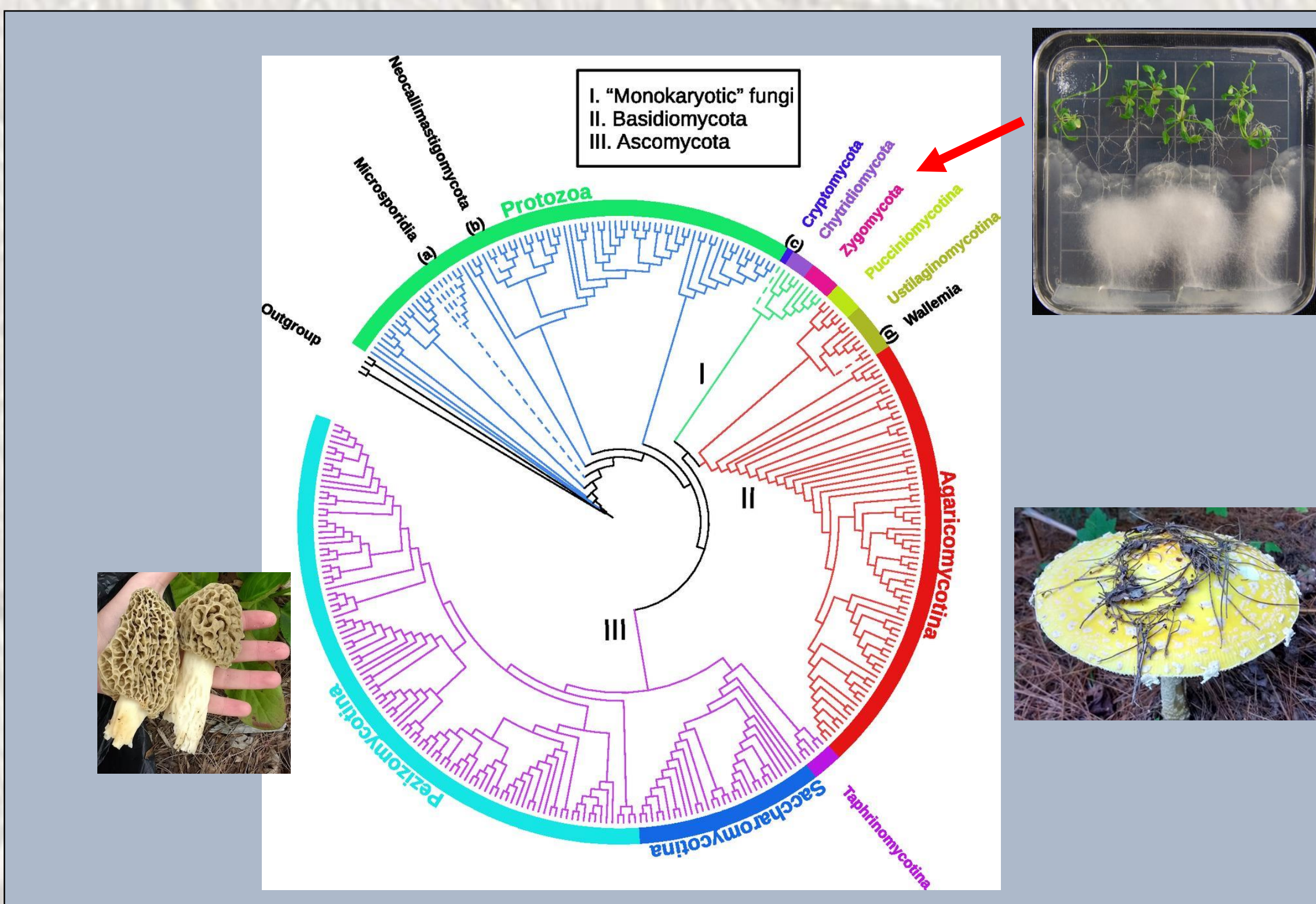- Best method combines FFT, PCA, and RF

Figure 4. Performance of Classification Methods. Performance significantly varies between methods ($F = 29.42$, $p = 3.25e^{-12}$), and letter annotations are different at $p < 0.05$ with Tukey's HSD test. The combination of Random Forest, Fast Fourier Transform, and Principal Component Analysis provides the best classification.

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
  - Use parallel computing to improve eliminate need for sampling

References

- Choi J, Kim SH. A genome Tree of Life for the Fungi kingdom. Proceedings of the National Academy of Sciences. 2017 Aug 29;114(35):9391-6.
- Mennitt D, Sherrill K, Fristrup 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.

Acknowledgements