

Faculty Mentor: Dr. David Mendoza-Cozatl, Plant Sciences

Funding Source: BioInformatics in Plant Sciences (BIPS)

Development of a yeast colony classification system using computer vision

Greg Glickert, Elizabeth D. Fallon, Christopher Zachary, and David G. Mendoza-Cozatl

Yeast one-hybrid (Y1H) and yeast two-hybrid (Y2H) systems are among the most common techniques to detect protein-DNA (Y1H) and protein-protein interactions (Y2H). These systems utilize a prey-bait system to detect interactions and our lab has implemented high-throughput Y1H and Y2H screenings by combining a transcription factor (TF) library containing 1,956 TFs from the model plant *Arabidopsis* and a microbial array pinning robot (RoToR). While our high throughput system is fully functional, the number of putative interactions that are able to be tested has outgrown our capabilities to identify positive interactions by hand. For instance, a Y1H screen probes 5,868 possible interactions per kb of DNA. This in turn creates a bottleneck if the results (i.e., colony size or color) are scored individually by researchers conducting the screen. To solve this issue, we have developed a colony classification system using computer vision and data processing libraries available in Python such as Open Computer Vision and NumPy for image analysis and SciPy/xlrd/Pandas and scikit-learn for data processing and formatting. Briefly, our pipeline first identifies colonies on individual yeast plates [96 or 384 formats], then the plate goes through an adaptive threshold so only one yeast colony remained in the image. To find the size of each colony, a connected components algorithm was used to find the area of each yeast colony. Next, color data was extracted by isolating individual colonies and separating the different color channels. Color and size data are then put into a data frame and exported into an Excel file. In the future, these data can then be used to train machine learning models made with Keras and scikit-learn to classify Y1H and Y2H hybrid with a higher degree of precision.