Phenomics pipeline for high-throughput image analysis of *in vitro* plant development

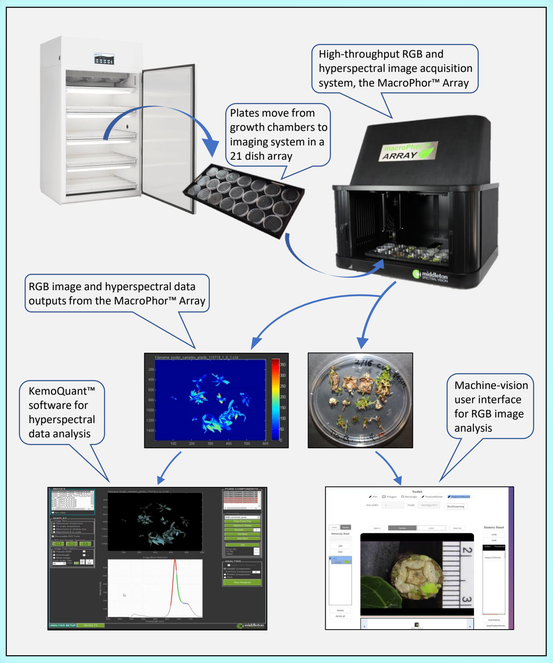
Anna Magnuson[1], Chris Draves[2], Howland Jones[2], Natalie Crothers[2], Gabor Kemeny[2], David Braumann[2], Michael Nagle[1], Cathleen Ma[1], Brett Pierce[1], Ritesh Mewalal[1], Jialin Yuan[3], Nihar A. Doshi[3], Fuxin Li[3], Yuan Jiang[4] and Steven H. Strauss[1]

[1]Department of Forest Ecosystems & Society, Oregon State University, Corvallis OR; [2]Middleton Spectral Vision, Middleton WI; [3]Department of Computer Science, Oregon State University, Corvallis OR; [4]Statistics Department, Oregon State University, Corvallis OR

Abstract

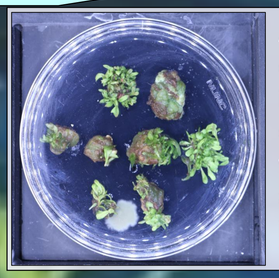
Collecting quantitative data on *in vitro* plant material has long been difficult as it relies on scoring systems that are subjective and time consuming. Oregon State University has partnered with Middleton Spectral Vision (MSV, a Wisconsin-based company) to develop a phenomics pipeline that can efficiently generate high-throughput precision data. The work is part of an NSF-Plant Genome Research Program funded study to conduct genome wide associate studies (GWAS) to identify genes that determine the rate of regeneration and transformation in poplar (*Populus*). The pipeline begins with an RGB and hyperspectral imaging system that is coupled with two routes of data extraction: KemoQuant™, a software developed by MSV that can analyze spectral data, and a machine-vision system to identify and quantify tissues from RGB and ultimately hyperspectral images.

Pipeline overview



Custom imaging system — The MacroPhor™ Array

The MacroPhor™ Array is designed to capture RGB and hyperspectral images across a 21 Petri dish configuration. Hyperspectral images contain a full 400-800nm emission spectrum for each pixel which allows for early and precise detection of fluorescence signals including GFP, and quantitative analysis of spectral components within a sample.

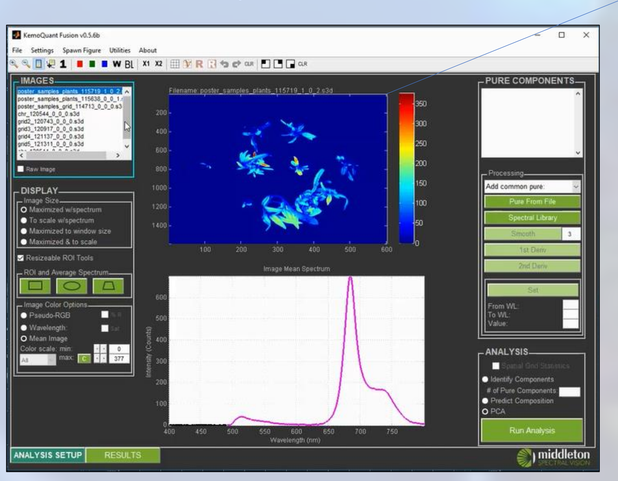


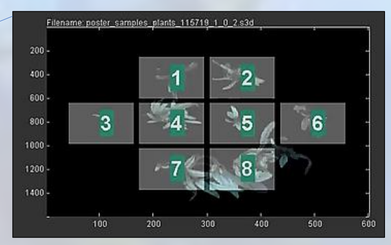
An LED light ring was developed to allow for RGB images without producing glare on the Petri dish lid (lights are visible at edges of Petri dishes, and ring is visible in chamber image to the middle-left).

A 488nm wavelength laser scans over the area of a Petri dish to detect GFP, chlorophyll, and other fluorescence signatures in the sample.

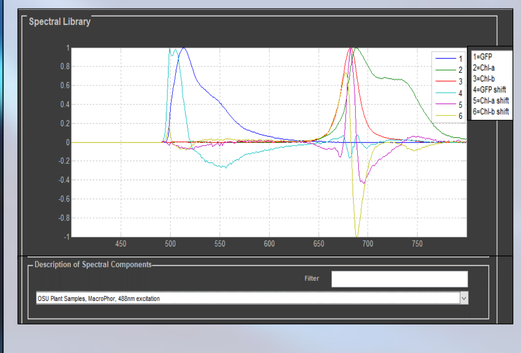
Hyperspectral analysis

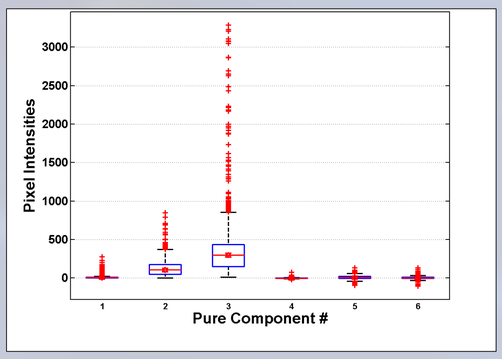
KemoQuant™ uses multivariate curve resolution to detect pure component fluorescent signals within spectra and automatically exports data on signal intensities and component ratios.





A custom grid overlay system was developed so that KemoQuant™ will recognize tissue explants on a Petri dish and output their data individually.

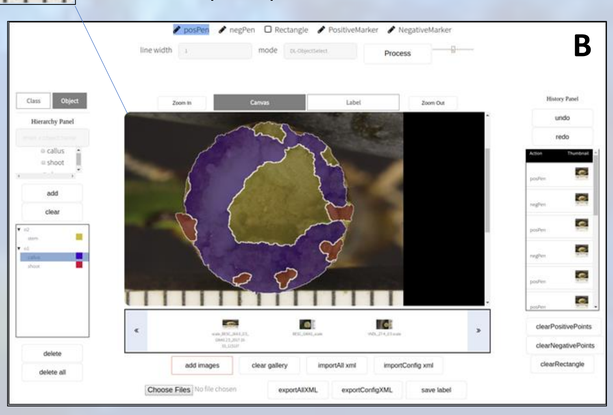


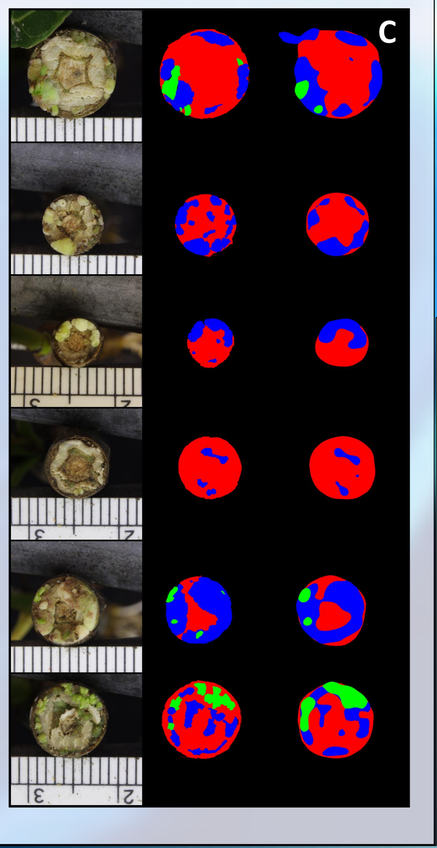


Pure components for an OSU sample include GFP, chlorophyll a, chlorophyll b, and a unique signature shift for each signal.

Machine vision

(A) Raw RGB image of *in vivo* poplar stem regenerating callus and shoots. Using a custom GUI (graphic user interface) seen in (B), it was manually annotated to train the machine vision algorithm, which is based on a deep neural network (DNN) model.





(C) Results from the first round of DNN training. The middle column shows the user annotated image to its left, and the right column contains the results from the DNN model output. The DNN model returns an estimated probability for each pixel as to whether it believes the annotation it chose is correct—its current estimate is above 90%, but as this consists mostly of background, we are working on a larger training set to further refine and increase resolution.

Current applications

- The Middleton Spectral Vision system extracts unique spectral signatures and quantifies the amount of fluorescent proteins and chlorophyll species in OSU tissue samples
- Hyperspectral imaging will aid in a pilot study to determine the effectiveness of the developmental genes *LEC1* and *EBB1* to increase resolution of GWAS for shoot regeneration in poplar
- Hyperspectral imaging will allow us to replace the standard GFP reporter with an auxin-responsive GFP reporter and cytokinin-responsive YFP reporter by early detection of signals
- KemoQuant™ will help determine the expression of the GFP and YFP reporters, and thus their ability to yield biologically informative data about causes of variation in regeneration and transformation
- High-throughput RGB imaging with the MacroPhor™ Array will speed and increase the accuracy of phenotype acquisition
- Training on the DNN model is ongoing and will ultimately allow us to obtain high-throughput data for an *in vivo* GWAS study of variation in rootability that is currently in progress

Acknowledgements

We thank the National Science Foundation Plant Genome Research Program for support (IOS # 1546900, Analysis of genes affecting plant regeneration and transformation in poplar), and members of the Tree Biosafety and Genomics Research Cooperative at OSU for its long term support of the Strauss laboratory.