COPAS Vision™—An integral component in the Smart Soil Organism Detector of soil nematodes.

*Summary of **nematode** research presented in Camila C. Filgueiras, Biosensors and Bioeletronics, https://doi.org/10.1016/j.bios.2022.114417*

Objective

Understanding soil biodiversity is of critical importance in the study of soil ecology. However, the numbers of microorganisms in a small area of soil make it very difficult to accurately survey a single environmental sample let alone a larger, more informative collection of samples.

This study asks: can the COPAS Vision cytometry and imaging instrument (utilizing a 500um flow channel) be used to collect large data sets on a high throughput scale "such that a technician can run sample and receive a detailed output of number and type of organisms in a matter of minutes?"

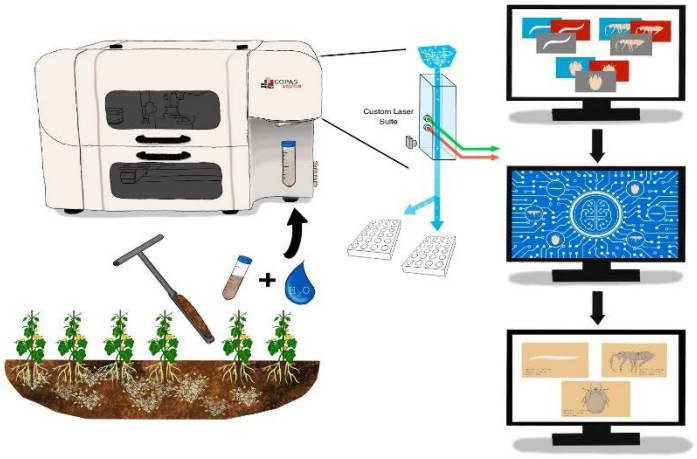

Figure 1. *Smart SOD. The Smart Soil Organism Detector. COPAS Vision (top left corner) equipped with multi-laser suite and high resolution imaging of objects in the flow channel is used to collect cytometry and imaging data. This data was applied to a machine learning pipeline trained to recognize soil organisms. Live organisms can then be separated, identified, and used in subsequent analysis.*

Introduction

The COPAS family of research instrumentation comprises a collection of large particle cytometers utilizing at least 1 excitation laser and up to 8 channels of fluorescence collection. Unique to COPAS instrumentation is the Profiling feature which graphically plots the fluorescence intensity changes along the length of the object as it passes through the laser(s). Large objects up to 1.5mm in diameter can be analyzed for physical and fluorescence characteristics and gently dispensed into a multi-well plate or other collection container for further investigation or reuse. The COPAS Vision also has equipped a camera to take an image of the object inside the flow channel. This image accompanies the cytometry data and can be analyzed using Union Biometrica's FlowPilot software or other image analysis tools.

Utilizing machine learning tools to process the massive amounts of imaging and cytometry data from the COPAS Vision, the Filgueiras group created what they term the Smart Soil Organism Detector (Smart SOD, figure 1) to automatically evaluate soil samples for their unique composition of substrate, nematode and microarthropod constituents.

Methods

Control samples containing soil substrate and separately *S. feltiae* entomopathogenic nematodes (EPN) infective juveniles (IJs) (hosted by waxworms and collected on white traps) were generated and run on the COPAS Vision (500um flow channel) to tune the Smart SOD system to identify: 1) the substrate background, 2) live nematodes, 3) dead nematodes (propidium iodide stained), and 4) nematode cuticles as discreet populations. Images collected by Vision were used to confirm individual object as well as population identities. Cohorts were split into training and test sets (80:20). The training set was used to train a multi-class support vector machine tuned using 5-fold, 5 repeated cross validation. Evaluation for out-of-sample performance was determined for the test set--both S. feltiae and tests and those repeated with *H. bacteriophora* produced similar results and 100% out of sample accuracy.

Additional nematode strains (14 in total) were run individually to build a machine learning protocol utilizing 24 total features using the Uniform Manifold Approximation and Projection in two dimensions. Data was overlaid with both strain and genera labels to visually examine differences between the individuals. Training set data was further cleaned, and processed to remove near zero variance predictors; predictors were also scaled and rebalanced to account for imbalanced culture constraints. These sets were then used to evaluate additional supervised machine learning models for accurate detection of different nematode strains. K-nearest Neighbors (KNN), Support Vector Machines (SVM), and Gradient Boosted Trees (XGBoost) used a 5-fold cross validation for hyperparameter tuning and 10-fold cross validation for in-samples performance evaluation. Further tuning using grid search racing with ANOVAs was conducted. The best model was trained from the entire training set and applied to the test. How well the trained model performed was assessed by out-of sample accuracy, specificity, and sensitivity. Additionally, ROC, precision-recall, confusion matrices, and variable importance measures were extracted from the final model's performance.

Results

H. bacteriophora entomopathogenic nematode infective juveniles segregated distinctively from sand and soil particles in the representative soil sample substrate utilizing cytometry parameters of Time of Flight (TOF-length of object signal), extinction (EXT-optical density) from use of 488nm laser. See Figure 2 at right.

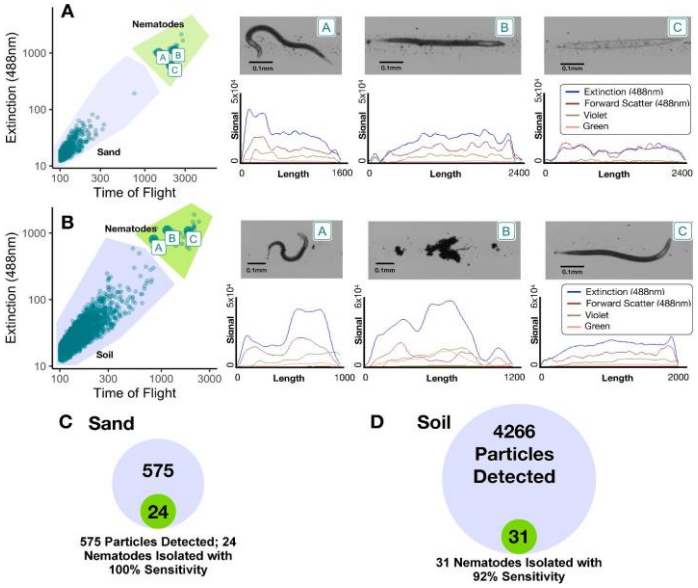

Figure 2. *Effectively separating nematodes from sand and soil substrate. A) Nematodes separate from sand particles based on time of flight in the flow cytometer flow cell and extinction at 488 nm. Each individual point represents a particle detected by the flow cytometer. Uppercase letters correspond to nematodes imaged by the COPAS Vision instrument with corresponding profiles along their length in multiple wavelengths as recorded by the instrument. B) Nematodes also separate from soil particles along similar dimensions. C) Sensitivity of nematode extraction from sand. D) Sensitivity of nematode extraction from soil.*

Similarly, live nematodes, dead nematodes and shed cuticle were separable utilizing differences in emission in the FarRed channel of fluorescence collection (emission collected at 680nm wavelength). See figure 3 below.

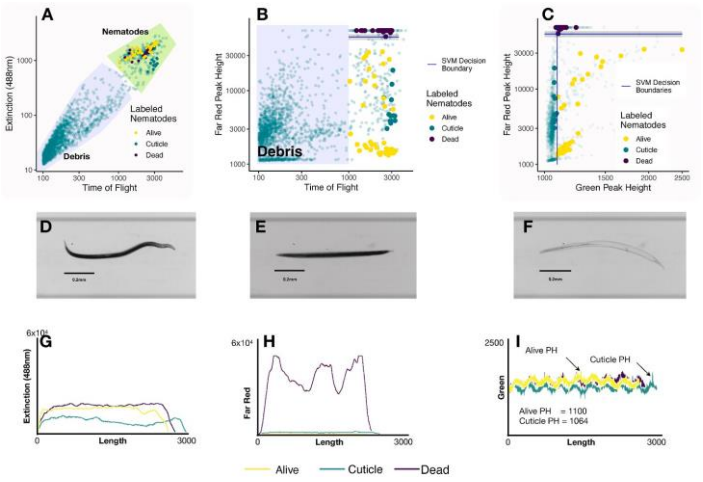

Figure 3. *Instrumentation coupled with supervised machine learning can separate live nematodes, dead nematodes, and nematode cuticles. A) Nematodes were separated from soil debris and a subsample imaged and labeled. B) Support Vector Machines (SVMs) separate alive and dead nematodes along the FarRed Peak Height Axis. C) SVMs separate nematode cuticle from alive and dead nematodes. For A-C, Points represent individual particles detected by the instrument. Horizontal and surrounding shaded region denote SVM decision boundaries and 95% confidence intervals respectively. D) Alive S. feltiae entomopathogenic nematode infective juvenile. E) Dead S. feltiae entomopathogenic infective juvenile, F) Cuticle discarded by S. feltiae. G) Similar body length extinction (at 488 nm) profiles of alive S. feltiae, dead S, feltiae and S, feltiae cuticle from d-f respectively. H) Disparate body length profiles in the FarRed spectrum. Dead S. feltiae register order of magnitudes higher in the Far Red spectrum due to propidium iodide staining. I) Green peak height is used by SVMs to distinguish cuticles from live nematodes. Although the differences between alive S. feltiae and S. feltiae cuticles are small, it is enough to produce 100% out-of-sample accuracy.*

Finally, separate clusters of 3 distinct genera of entomopathogenic nematodes (*Heterorhabditis*, *Steinernema*, and *Caenorhabditis*) emerged from the data collected from 13 relevant strains. Features most important to distinguishing the different strains were not size and density but rather information in the violet and green spectrum: violet peak height and peak width (collected at 445nm) ranked in the top ten features for classification. See Figures 4 below and Figure 5 next page.

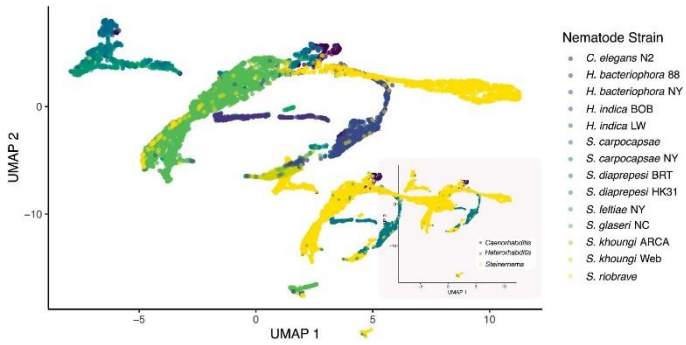

Figure 4. *Multi-dimensional differences between nematode individuals projected into two dimensions using uniform manifold approximation and projection. Each nematode was scanned by the instrument up to 8,000 times along its body length and interrogated with a series of lasers of different wavelength generating 43 different features. Following projection into two-dimensional space, points were labeled either with strain (full figure) or genus (inset).*

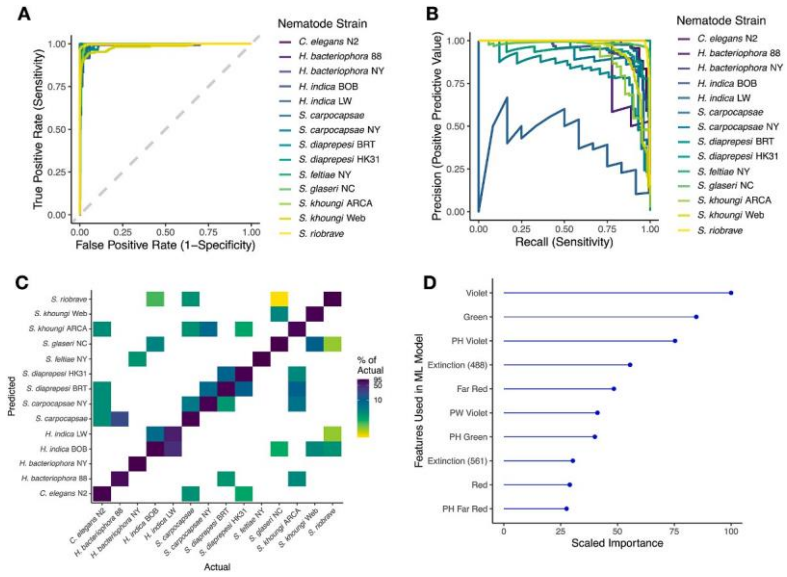

Figure 5. *Performance of gradient boosted supervised classification models for identifying nematode strains using Smart SOD sensors. A) Receiver operating characteristic by strain. B) Precision-recall curves by strain. C) Confusion matrix for strain classification where the color (note log scale) corresponds to the percentage of actual (True) individuals falling in that class (ie. Percentages were calculated along vertical (true) bins, not horizontally). D) Variable importance measures from the classification model. Importances are scaled so that the most important feature is at 100. Only the top ten most important features are shown.*

Conclusions

The authors state "The Smart SOD system is capable of extracting and distinguishing soil organisms from their substrate, isolating live organisms, accurately identifying microorganism strains, and accurately identifying classes and major groups of arthropods. Moreover, the Smart SOD system is capable of accomplishing all of this in a high-throughput manner."

Its abilities to identify and distinguish live, dead, and parts of shed nematodes allows the researcher to better probe questions about the health and complexity that current genomics techniques lack. What's more, the system's ability to identify different strains based on the fluorescence and imaging, means the system can be used to shed light on cuticle construction and adaptation to soil environment. Surveillance of soil in this way may also uncover new organisms whose image is automatically collected as well. And because organisms are handled gently, any individual or group type can be dispensed for further analysis or to propagate new lines for research.

Work performed in 2022, by:

Filgueiras, C[1], Kim, Y[2], Wickings, K[3], Borai, F[4], Duncan, L[5], Willett, D[6]

[1]University of North Carolina, Ashville, NC USA

[2]Union Biometrica, Holliston MA, USA

[3]Cornell University, Cornell AgriTech, Geneva, NY USA4

[4]University of Florida, Wimauma, FL USA

[5]University of Florida, Lake Alfred, FL USA

[6]North Carolina State University, Asheville, NC USA