**Full Proposal Submitted to the Northwest Potato Research Consortium**

**Title: Automated identification of plant-parasitic nematodes of potato at genus and species level**

**Year Initiated: 2022-23. Current Year: 2022-23. Terminating Year: 2024.**

**Personnel & Cooperators:**

PIs involved include David Linnard Wheeler and Cynthia Gleason from Washington State University, Inga Zasada from the USDA-ARS, and Sam Chavoshi from AGNEMA. Sudha GC Upadhaya is a graduate student working with Dr. Wheeler. Both PIs from WSU will request funding.

**Funding Request for 2022-23: $67,784**

**Introduction: Problem Statement, Research Question(s) & Justification:**

Effective management of plant-parasitic nematodes requires early and accurate identification and quantification. Current nematode diagnosis relies heavily on morphology-based identification methods which demand highly skilled personnel and are time consuming. Similarly, molecular marker-based methods tend to be cost and resource intensive and are not available for all plant-parasitic nematodes. To help resolve this problem, a fast, accurate and efficient nematode identification tool is needed. Currently labs that provide nematode identification services do not possess the resources to produce fast and accurate results at scale. According to PI Chavoshi, who runs a nematode diagnostic clinic, one of the main bottlenecks in nematode identification is that it requires skilled personnel to look at nematodes individually and identify it. This process is not only time consuming and expensive in terms of labor costs, but also prone to inevitable human error. As such, automated tools that do not demand taxonomic expertise and generate reproducible results are needed. Such tools have been reliably developed with machine learning algorithms in various domains, from biomedicine to entomology (Martineau et al. 2017; McKinney et al. 2020). Most recently, a research group in Florida has begun to develop an identification tool for plant-parasitic nematodes that affect citrus (Buck 2021). This illustrates that nematologists in high value fruit crops realize the value of using machine learning algorithms to provide faster, cheaper, and more accurate nematode identification. Unfortunately, the nematode problems of citrus in Florida are not the same as potato in the PNW. The goal of this research proposal is to develop an automated method that would recognize key genera of plant-parasitic nematodes based on photos taken of extracted nematodes. We propose to develop our tool for nematodes that affect potato, such as root lesion (*Pratylenchus* spp.)*,* root knot (*Meloidogyne* spp.)*,* and stubby root (*Paratrichodorus* and *Trichodorus* spp.) nematodes(Hills et al. 2020; Zasada et al. 2018).

For this project, machine learning algorithms will be developed to identify common plant-parasitic nematode genera associated with potato production. Models will be built and trained on a large number of raw images of different nematode genera. The algorithms will learn to associate diagnostic features, specifically morphology (length, width, shape), texture, etc. Once trained, model performance will be validated on previously unseen images. Our goal in this proposal is to generate the baseline of data and validate the system. This research sets the stage for our long-term goal, in which these models will be developed into an automated detection and quantification system of plant-parasitic nematodes that threaten the potato industry that can be used by the industry. Such a tool will enable stakeholders to make faster and more informed nematode management decisions.

**Anticipated Benefits/Expected Outcomes/Information Transfer:**

Anticipated outcomes of this research include development and validation of machine learning algorithms for automated identification of three important plant-parasitic nematodes: root lesion, root knot, and stubby root nematodes. The research proposed here is the fundamental first step towards the development of technologies that can be transferred to diagnostic labs. We need to set-up the basics of our machine learning algorithms in order to reach our long-term goal. We envision that our developed technology will one day identify and quantify nematodes from an extracted soil sample. To achieve this goal, we need to start by training machine learning algorithms to recognize the three most problematic nematode groups in the PNW.

**Goal(s), Hypothesis, & Objectives**:

The goal of this project is to enhance the accuracy, efficiency, and reproducibility of nematode identification for potato growers in the Pacific Northwest. To accomplish this goal, we propose to build and validate machine learning algorithms that can quickly and accurately identify three plant-parasitic nematodes associated with potato production in the Pacific Northwest. The objectives of this study are to:

1. Capture microscopic images of common plant-parasitic nematodes.

2. Confirm the identify of our three nematode genera with traditional and molecular methods to validate our photo library.

3. Build and validate machine learning models that can identify nematodes.

**Procedures:**

For objective 1, pure cultures of each nematode genus already maintained in the Chavoshi, Gleason and Zasada laboratories will be used. To supplement our collection with nematodes other than the target nematodes, nematodes will be extracted from field soils or pot cultures with standard procedures (**Fig 1a**). Individual nematodes from pure cultures and soil samples will then be imaged with 200/100X magnification (**Fig 1b**). Objective 1 will be completed by Zasada, Gleason, and Chavoshi.

For objective 2, nematodes will be identified to genus with morphological features like shape of head, body length and shape, stylet length, shape of stylet knob etc. In total, images from four nematode genera including root lesion (*Pratylenchus* spp. (n = 1,000)), root knot (*Meloidogyne* spp. (n = 1,000)), and stubby root (*Paratrichodorus* spp. (n = 500) and *Trichodorus* spp, (n = 500)) nematodes will be captured (**Fig 1b**). Non-parasitic nematodes (n = 2,000) will also be imaged and serve as a control group. Different life stages of the nematodes including juveniles and adults (both male and female) will be included in image dataset. Captured images will include both full and partial (head and tail) body parts of juvenile/adult nematodes (**Fig 1c**). Like objective 1, objective 2 will be completed by Zasada, Gleason, and Chavoshi.

For objective 3, the labelled images from objective 2 will be used to build and validate machine learning algorithms. At least two separate machine learning algorithms will be developed and validated for genus level classification problems (**Fig 1d**). First, a subset of the images collected from objective 2 will be used to train the algorithms to associate the pixels in each image with the label or taxa within each image. If necessary, image augmentation will be completed to introduce variance into dataset. Mainly, convolution neural network (CNN), a type of machine learning algorithm which takes raw images as an input file, will be used. Several architecture types of CNN including ResNet, AlexNet, VGG etc. will be implemented to train the algorithms and their predictive performance will be evaluated. To test the accuracy of each algorithm, new images that have not yet been “seen” or learned will be presented to each algorithm. The models with the best accuracy will be selected for future use (**Fig 1e**). An accurate model is one that, for example, labels an image of a root lesion nematode as such. An inaccurate model will label and image of a root lesion nematode as a stubby root or root lesion nematode.

Altogether, the labelled images from Objective 1 will be compiled, and using the machine learning algorithms we develop, we will be able to scan the images and predict the genera of nematode present. The image analysis pipeline will be developed in the open-source programming languages, Python and R, and will be published online. Objective 3 will be completed by Upadhaya and Wheeler.

Diagram

Description automatically generated with medium confidence

**Figure 1.** Flow chart of experiment. (a) The collected nematodes will be identified at genus level and (b) imaged under a compound microscope. (c) Images including sample nematode images to be used to train machine learning algorithms. (d) The multilayered machine learning algorithm will be developed and optimized to extract and learn important features from each image. (e) Model performance will be evaluated using confusion matrix. Finally, the best performing model will be selected for nematode identification in future (e-bottom picture).

**Collaboration:** Wheeler and Upadhaya will perform image preprocessing, develop, and optimize machine learning algorithms, and run analysis. Gleason, Zasada, and Chavoshi will capture the microscopic images of plant-parasitic nematodes.

**Project Timeline:**

Objective 1 will be completed between the springs of 2022 and 2023.

Objective 2 will be completed between the winters of 2022 and 2023.

Objective 3 will be completed during the winter of 2023 and summer of 2024.

**Additional grant funding:** This project will serve to generate preliminary data for larger grants, like Washington State Department of Agriculture’s Specialty Crop Block Grant and USDA-NIFA Agriculture and Food Research Initiative (AFRI). The proposal will be scaled up to develop open-source platform for automated identification as well as quantification of plant parasitic nematodes.

**Literature Cited:**

Buck, B. 2021. AI may help UF researchers identify crop-destroying nematodes. UF|IFAS Blogs. <http://blogs.ifas.ufl.edu/news/2021/04/14/ai-may-help-uf-researchers-identify-crop-destroying-nematodes/>

Hills, K., Collins, H., Yorgey, G., McGuire, A. and Kruger, C. 2020. Improving soil health in Pacific Northwest potato production: a review. American Journal of Potato Research. 97:1-22.

Martineau, M., Conte, D., Raveaux, R., Arnault, I., Munier, D. and Venturini, G. 2017. A survey on image-based insect classification. Pattern Recognition. 65:273-284.

McKinney, S.M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., Back, T., Chesus, M., Corrado, G.S., Darzi, A. and Etemadi, M. 2020. International evaluation of an AI system for breast cancer screening. Nature. 577:89-94.

Zasada, I. A., Dandurand L-M., Gleason, C., Hagerty, C.H., and Ingham, R.E. 2018. Plant Parasitic Nematodes: Idaho, Oregon and Washington. Chapter 8. https://zasadalab-usda-ars-hcru.weebly.com/uploads/4/0/1/9/40199119/zasada\_et\_al\_bk\_ch\_2018.pdf

**Budget:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Wheeler | Gleason | Zasada | Chavoshi | **Total** |
| **1Salaries:** Faculty |  |  |  | 13,600 | 13,600 |
| Graduate student | 19,769 |  |  |  | 19,769 |
| Other students |  |  |  |  |  |
| Other labor |  | 8,754 | 4,616 | 7,500 | 20,870 |
| **2Employee Benefits (OPE):** Faculty |  |  |  |  |  |
| Graduate student | 2,491 |  |  |  | 2,491 |
| Other students |  |  |  |  |  |
| Other labor |  | 856 | 369 |  | 1,225 |
| Computing | 2,500 |  |  |  | 2,500 |
| Equipment |  |  |  |  |  |
| Travel: |  |  |  |  |  |
| Operating Expenses |  |  |  | 1,200 | 1,200 |
| Other Expenses |  | 500 | 500 | 5,129 | 6,129 |
| **Total** | 24,760 | 10,110 | 5,485 | 27,429 | 67,784 |

|  |
| --- |
| ¹Salary is to support graduate student for 0.5 FTE of 12 months at Wheeler's lab. |
| ²Benefits for graduate student are 12.6% of salary |

**Anticipated Total Requests in Coming Years: 2023-2024:** $72,000