RESOURCE ANNOUNCEMENT

hagis, an R Package Resource for Pathotype Analysis of *Phytophthora sojae* Populations Causing Stem and Root Rot of Soybean

Austin G. McCoy,[1,†] Zachary Noel,[1,2] Adam H. Sparks,[3] and Martin Chilvers[1,2]

[1] Department of Plant, Soil and Microbial Sciences, Michigan State University, East Lansing, MI 48824, U.S.A.
[2] Program in Ecology, Evolutionary Biology, and Behavior, Michigan State University
[3] University of Southern Queensland, Centre for Crop Health, Toowoomba, Qld 4350, Australia

Abstract

Phytophthora sojae is a significant pathogen of soybean worldwide. Pathotype surveys for *Phytophthora sojae* are conducted to monitor resistance gene efficacy and determine if new resistance genes are needed. Valuable measurements for pathotype analysis include the distribution of susceptible reactions, pathotype complexity, pathotype frequency, and diversity indices for pathotype distributions. Previously the Habgood-Gilmour Spreadsheet (HaGiS), written in Microsoft Excel, was used for data analysis. However, the growing popularity of the R programming language in plant pathology and desire for reproducible research made HaGiS a prime candidate for conversion into an R package. Here we report on the development and use of an R package, hagis, that can be used to produce all outputs from the HaGiS Excel sheet for *P. sojae* or other gene-for-gene pathosystem studies.

Uniform and healthy stand establishment is essential to maximizing soybean (*Glycine max*) yield. Oomycetes such as *Phytophthora sojae* constitute a significant threat to stand establishment and yield. *Phytophthora sojae* has been managed primarily via deployment of single resistance genes in commercial soybean cultivars, which interact with *P. sojae Avr* gene products to confer resistance (Anderson et al. 2015). Genetic resistance to *P. sojae* is the most economical form of control for *P. sojae*, as it confers season-long protection to noncompatible pathotypes (Dorrance et al. 2016). However, *P. sojae* pathotype surveys need to be regularly conducted to determine shifts in pathotypes over time and to provide recommendations for effective resistance genes. Although state-wide pathotype surveys have been conducted for the past 60 years in the United States, there has been no significant advance in pathotype analysis since the development of the Habgood-Gilmour spreadsheet (HaGiS), written in Microsoft Excel, in 1999 (Herrmann et al. 1999; Kaufmann and Gerdemann 1958).

Phytophthora sojae pathotype surveys monitor the efficacy of soybean resistance genes in relation to one or more *P. sojae* populations. In doing so, large sets of virulence data are generated, potentially for hundreds of isolates (Dorrance et al. 2016). Using such large datasets within the HaGiS Excel-based program can be cumbersome and time-intensive

Austin G. McCoy, Zachary Noel, and Adam H. Sparks contributed equally to this work.

[†]Corresponding author: A. G. McCoy; mccoyaus@msu.edu

The author(s) declare no conflict of interest.

Accepted for publication 14 August 2019.

Table 1. Example of tabular output from the hagis program, using the summarize_gene() at 60% susceptibility cut-off function[a]

Gene	No. susceptible	% Pathogenic
Susceptible	21	100
1a	21	100
1b	15	71.43
1c	20	95.24
1d	16	76.19
1k	18	85.71
2	14	66.67
3a	5	23.81
3b	20	95.24
3c	4	19.05
4	5	23.81
5	13	61.9
6	11	52.38
7	21	100

[a] This function produces a detailed table displaying the number of isolates each gene is not effective against as well as offering a percentage of the isolates tested that are pathogenic on each gene.

to transfer the data into and perform analysis. The R statistical programming language (R Core Team 2019) offers the ability to work with large datasets in an easy and efficient manner, without the additional data-entry steps that the HaGiS Excel program requires, while treating the virulence data as read-only, thereby further reducing the chance for errors.

R has become widely used in plant-pathology studies, due to its open-source framework and amenability to conduct reproducible research (Bergna et al. 2018; Duku et al. 2016; Sparks et al. 2011; Wallace et al. 2018). Using an R package for analyzing pathotype survey data can replicate all analyses provided by Excel-based programs. It allows users to create reproducible research and more detailed visualizations, as well as allowing the plant-pathology community to actively contribute to and build upon this code for future studies. For instance, McCoy and Noel (2018) produced R scripts to conduct these analyses, originally performed with HaGiS, which were used to create the hagis R package (McCoy et al. 2019).

For ease of use, the package uses a single argument format, which works in all hagis functions. Users provide their own data in the form of a spreadsheet, CSV, or text file, specifying the proper fields for analysis. Functions are provided to calculate pathotype complexity and summarize the distribution of reactions for each gene tested. Simple, Shannon, Simpson, Gleason, and evenness diversity indices are calculated for the pathotype dataset. Outputs from these analyses are given in publication-ready graphics or tables and can be further modified by the user (Table 1).

R language offers many advantages to Excel-based data analysis, such as reproducibility and user customization. Furthermore, hagis takes advantage of the data.table package (Dowle and Srinivasan 2019) to efficiently handle large datasets, such as those produced through *P. sojae* pathotype surveys, rapidly and efficiently. Significantly, hagis provides the first development in *P. sojae* pathotype analysis in 20 years. While hagis was developed to support *P. sojae* pathotype surveys, it was designed to work with any pathotype analyses of gene-for-gene pathosystems to determine effective resistance genes in management.

The package source code, including the Rmarkdown code for this paper, more information, and instructions on how to use hagis can be found in the Zenodo records database. The package can be downloaded and installed from the Comprehensive R Archive Network (CRAN) website and is released under the MIT license.

Author-Recommended Internet Resources

CRAN website: https://CRAN.R-project.org/package=hagis
data.table CRAN website: https://CRAN.R-project.org/package=data.table
GitHub hagis page: https://openplantpathology.github.io/hagis
Zenodo hagis v3.0.0: https://zenodo.org/record/3378007

Literature Cited

Anderson, R. G., Deb, D., Fedkenheuer, K., and McDowell, J. M. 2015. Recent progress in RXLR effector research. Mol. Plant-Microbe Interact. 28: 1063-1072.

Bergna, A., Cernava, T., Rändler, M., Grosch, R., Zachow, C., and Berg, G. 2018. Tomato seeds preferably transmit plant beneficial endophytes. Phytobiomes J. 2:183-193.

Dorrance, A. E., Kurle, J., Robertson, A. E., Bradley, C. A., Giesler, L., Wise, K., and Concibido, V. C. 2016. Pathotype diversity of *Phytophthora sojae* in eleven states in the United States. Plant Dis. 100:1429-1437.

Dowle, M., and Srinivasan, A. 2019. data.table: Extension of 'data.frame'.

Duku, C., Sparks, A. H., and Zwart, S. J. 2016. Spatial modelling of rice yield losses in Tanzania due to bacterial leaf blight and leaf blast in a changing climate. Clim. Change 135:569-583.

Herrmann, A., Lower, C. F., and Schachtel, G. A. 1999. A new tool for entry and analysis of virulence data for plant pathogens. Plant Pathol. 48:154-158.

Kaufmann, M. J., and Gerdemann, J. W. 1958. Root and stem rot of soybean caused by *Phytophthora sojae* n. sp. Phytopathology 48:201-208.

McCoy, A. G., and Noel, Z. A. 2018. AGmccoy/Phytopthora-sojae-Pathotype-analysis: Beta-release of Phytophthora sojae pathotype analysis code (December 25, 2018). Zenodo. http://doi.org/10.5281/zenodo.2526326

McCoy, A. G., Noel, Z. A., Sparks, A. H., Kamvar, Z. N., and Chilvers, M. I. 2019. openplantpathology/hagis v3.0.0. (August 26, 2019). Zenodo. http://doi.org/10.5281/zenodo.3378007

R Core Team. 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Sparks, A. H., Forbes, G. A., Hijmans, R. J., and Garrett, K. A. 2011. A metamodeling framework for extending the application domain of process-based ecological models. Ecosphere 2:art90.

Wallace, J. G., Kremling, K. A., Kovar, L. L., and Buckler, E. S. 2018. Quantitative genetics of the maize leaf microbiome. Phytobiomes J. 2:208-224.