What Does Reproducible Research Mean for Plant Pathology?

A. H. Sparks, E. M. Del Ponte, N. J. Grünwald, and Z. Foster

2017-04-23

1. Adam H. Sparks: Centre for Crop Health, University of Southern Queensland, Toowoomba, Qld 4350, Australia
2. Emerson M. Del Ponte: Universidade Federal de Viçosa, Viçosa, MG, Brasil
3. Zachary S. L. Foster: Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR 97330, USA
4. Niklaus J. Grünwald: Horticultural Crops Research Laboratory, USDA Agricultural Research Service, Corvallis, OR 97330, USA

# ABSTRACT

Abstracts are mandatory and limited to one 200 word paragraph.

# MAIN TEXT

Reproducibility and replicability in scientific research have once again been highlighted recently (Nature 2016; Baker 2016) as an issue.

Patil et al. (2016) have provided several definitions to clarify the concepts surrounding reproducibility and replicability. For the purposes of this paper we follow the definitions as given by Patil et al. (2016).

* Why reproducible research

## BEST METHODS FOR REPRODUCIBLE RESEARCH

* Provide definitions (provide definitions for terms used so it's clear)
* Data
* Data formatting (flat files; use Comma Chameleon, Table Tool, others?)
* Data storage (don't edit raw data files; use file permissions to prevent changes to raw data files, use data bases where possible and appropriate; etc.)
* When publishing
* Provide data
* Provide code
* Using GitHub for code (and small data?)
* Using Figshare or Zenodo vs a lab website (DOIs, other reasons)

## WHAT IS THE STATE OF REPRODUCIBLE RESEARCH IN PLANT PATHOLOGY?

* Madden et al. (2015) supply an *e-****X****tra*\* with reproducible examples for readers.
* Duku et al. (2016) provide models, data and code, (<http://adamhsparks.github.io/MICCORDEA/>) necessary to replicate the entire study modelling the effects of climate change on rice bacterial blight and rice leaf blast in Tanzania.
* Sparks et al. (2011, 2014) provide models, data and code, (<http://adamhsparks.github.io/Global-Late-Blight-MetaModelling/>) necessary to replicate model development and the subsequent the study on the effects of climate change on potato late blight.
* Del Ponte provides data and a reproducible report that explain in details all steps of the analysis and the R codes for conducting a meta-analysis for assessing heterogeneity in relationship between white mold incidence and soybean yield and between incidence and soybean tied.
* Example from Grünwald lab:
* paper <http://apsjournals.apsnet.org/doi/full/10.1094/PHYTO-12-14-0350-FI>
* github repo <https://github.com/grunwaldlab/Sudden_Oak_Death_in_Oregon_Forests>
* Other examples from plant pathology providing e-Xtras or supplemental material

## RANDOM SAMPLING OF ARTICLES FROM THE TOP 20 PLANT PATHOLOGY JOURNALS

The 21 plant pathology discipline journals were selected by the authors as representations of discipline-based journals where others in the field of phytopathology are likely to publish were used to create a database of journals from which to randomly select articles for inspection. Two hundred articles were randomly selected from 2012 to 2016 from a list of randomly selected pages assigned to a randomised list of the 21 journals (Sparks et al. 2017) where the page number fell within an article for the given journal. In cases where an article was not suitable, *e.g.*, a review or otherwise not related to plant pathology, the next article was selected until a suitable article was found. Notes regarding the selection of articles can be found in the file, XXXX, available in this paper's repository. The pages list was numbered from page one and went to 150. This was done since some journals restart their numbering with each issue and also ensures that the journal is more likely to have a page number corresponding to the randomly generated value. This also assumes that there is no effect or bias on reproducibility based on the time of year that an article was published, since most journals start with page number one at the beginning of the year.

The list of journals was saved as a comma separated value (CSV) file and imported into R (R Core Team 2017).

## DISCUSSION

## ACKOWLEDGEMENTS

### LITERATURE CITED

Baker, M. 2016. Is there a reproducibility crisis? Nature. 533:453–454.

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. Climatic Change. 135:569–583 Available at: <http://dx.doi.org/10.1007/s10584-015-1580-2>.

Madden, L. V., Shah, D. A., and Esker, P. D. 2015. Does the *P* value have a future in plant pathology? Phytopathology. 105:1400–1407.

Nature. 2016. Reality check on reproduciblity. Nature. 533:437.

Patil, P., Peng, R. D., and Leek, J. 2016. A statistical definition for reproducibility and replicability. bioRxiv. Available at: <http://biorxiv.org/content/early/2016/07/29/066803>.

R Core Team. 2017. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at: <https://www.R-project.org/>.

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 Available at: <http://www.esajournals.org/doi/abs/10.1890/ES11-00128.1>.

Sparks, A. H., Forbes, G. A., Hijmans, R. J., and Garrett, K. A. 2014. Climate change may have limited effect on global risk of potato late blight. Global Change Biology.:3621–3631 Available at: <http://dx.doi.org/10.1111/gcb.12587>.

Sparks, A. H., Ponte, E. M. D., Foster, Z., and Grünwald, N. J. 2017. Reproducible-research-in-plant-pathology. Available at: <https://github.com/adamhsparks/Reproducible-Research-in-Plant-Pathology> [Accessed ].