

The Riffomonas Reproducible Research Tutorial Series

Patrick D Schloss¹

¹ Department of Microbiology & Immunology, University of Michigan, Ann Arbor, MI USA

DOI: [10.21105/jose.00013](https://doi.org/10.21105/jose.00013)

Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Submitted: 08 May 2018

Published: 30 August 2018

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC-BY](#)).

Summary

The [Riffomonas Reproducible Research tutorial series](#) is a collection of tutorials that focuses on the improvement of reproducible data analysis for those doing microbial ecology research. Although the materials focus on issues in microbial ecology, the principles are broadly applicable. Each tutorial presents broad concepts and how they are related to reproducibility as well as applied practice using specific tools that are designed to foster reproducibility. In addition to activities that are specific to individual tutorials, the series also uses data from Kozich et al. (Kozich, Westcott, Baxter, Highlander, & Schloss, 2013) to create an example that is developed over the course of fourteen tutorials. The tutorial series consists of 14 lessons that are available as HTML-based presentations with embedded presenter notes taken from live demonstrations of the content that are available as videos hosted on YouTube.

Statement of Need

The design and structure of the tutorials in the series give scientists new to the field of microbial ecology research the background and tools that they need to implement a reproducible analysis of their data. Although there are a growing number of commentaries speak to a “reproducibility crisis” in science and microbiology in particular (Casadevall, Ellis, Davies, McFall-Ngai, & Fang, 2016; Collins & Tabak, 2014; Garijo et al., 2013; Noble, 2009; Ravel & Wommack, 2014; Schloss, 2018), there has not been a concentrated effort to improve reproducibility within microbiology. Although the results of these studies are exciting, the interdisciplinary nature of the field means that the practitioners are not experts in all areas. This is especially true in the analysis of the large datasets that are being generated using high throughput sequencing.

Content

The fourteen tutorials focus on issues related to documentation, transparency, openness, and automation. The primer by Noble (Noble, 2009) motivates much of the material on project organization, documentation, and automation. Because of the unique nature of computational research relative to bench science, considerable attention is also given to using version control to document the evolution of a project and how to best work individually as well as with others to foster reproducibility. Although the themes are generally tool-agnostic, the practical implementation of these tools is prominent in the materials and when specific tools are used, other options are mentioned. Specific tools that are included in the tutorials include:

- Documentation ([markdown](#), [rmarkdown](#) (Allaire et al., 2018), R (R Core Team, 2018), [make](#), [git](#))
- Organization ([bash](#), [HPC/AWS](#), [git](#))
- Automation ([bash](#), R (R Core Team, 2018), [make](#))

- Transparency ([ORCID](#), [FigShare](#), [git](#), [GitHub](#), open source licensing)
- Collaboration ([git](#), [GitHub](#), open source licensing)

The focus of this tutorial series is on developing proficiency with a set of tools to foster reproducible research. Therefore although these tools have many uses, the materials focus on those features of the tools that support greater reproducibility. Specifically, the following tools are used in the tutorials, but learners should not finish the materials expecting to be proficient in using them. Although proficiency in these tools is not required, background in these tools will be helpful.

- R (R Core Team, 2018)
- `mothur` (Kozich et al., 2013)

Throughout the tutorial there are exercises that encourage the participants to discuss topics with their mentors and research group, activities where the participants are asked to engage resources at a deeper level. By demonstrating a meaningful level of participation in the tutorial series, participants receive a “virtual badge” and will be listed on the Reproducible Research Tutorial Series Honor Roll, which provides a certification of their training. The material is presented in an escalating manner such that over multiple tutorials participants are introduced to a topic, then given partial exposure and deeper exposure, and finally work with the tools proficiently.

The tutorial series is available as a series of slide decks written in [markdown](#) and are rendered using the [remark framework](#). The content is available under the CC-BY-SA 4.0 license. The slides are available on the [tutorial series GitHub repository](#). Accompanying the slides are YouTube-hosted videos where the author presents the material and does live-coding demonstrations to help participants engage with the material. In addition to links to the videos on the tutorial home page, there is a YouTube playlist that aggregates the 14 tutorials as the [Riffomonas Reproducible Research Tutorial Series](#). Transcripts from these videos are available as presenter notes for each of the slide decks by pressing ‘p’ when the slides are open.

Acknowledgements

This work was supported by a grant from the National Institutes of Health (R25 GM116149).

References

- Allaire, J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., Wickham, H., et al. (2018). Rmarkdown: Dynamic documents for r. Retrieved from <https://CRAN.R-project.org/package=rmarkdown>
- Casadevall, A., Ellis, L. M., Davies, E. W., McFall-Ngai, M., & Fang, F. C. (2016). A framework for improving the quality of research in the biological sciences. *mBio*, 7(4), e01256–16. doi:[10.1128/mbio.01256-16](https://doi.org/10.1128/mbio.01256-16)
- Collins, F. S., & Tabak, L. A. (2014). NIH plans to enhance reproducibility. *Nature*, 505(7485), 612–613. doi:[10.1038/505612a](https://doi.org/10.1038/505612a)
- Garijo, D., Kinnings, S., Xie, L., Xie, L., Zhang, Y., Bourne, P. E., & Gil, Y. (2013). Quantifying reproducibility in computational biology: The case of the tuberculosis drugome. (C. A. Ouzounis, Ed.) *PLOS ONE*, 8(11), e80278. doi:[10.1371/journal.pone.0080278](https://doi.org/10.1371/journal.pone.0080278)
- Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K., & Schloss, P. D. (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing

- amplicon sequence data on the MiSeq illumina sequencing platform. *Applied and Environmental Microbiology*, 79(17), 5112–5120. doi:[10.1128/aem.01043-13](https://doi.org/10.1128/aem.01043-13)
- Noble, W. S. (2009). A quick guide to organizing computational biology projects. (F. Le-witter, Ed.) *PLOS Computational Biology*, 5(7), e1000424. doi:[10.1371/journal.pcbi.1000424](https://doi.org/10.1371/journal.pcbi.1000424)
- Ravel, J., & Wommack, K. (2014). All hail reproducibility in microbiome research. *Microbiome*, 2(1), 8. doi:[10.1186/2049-2618-2-8](https://doi.org/10.1186/2049-2618-2-8)
- R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Schloss, P. D. (2018). Identifying and overcoming threats to reproducibility, replicability, robustness, and generalizability in microbiome research. (J. Ravel, Ed.) *mBio*, 9(3), e00525–18. doi:[10.1128/mbio.00525-18](https://doi.org/10.1128/mbio.00525-18)