

Kelly Sovacool

✉ sovacool@umich.edu
🌐 sovacool.dev
in [kelly-sovacool](https://www.linkedin.com/in/kelly-sovacool)
🐙 [kelly-sovacool](https://github.com/kelly-sovacool)

Bioinformatician building and applying open source tools for microbiome research, and contributing to data science education along the way.

Education

2018-present **PhD Bioinformatics**, *Dept. of Computational Medicine and Bioinformatics*, University of Michigan
○ Advisor: Patrick D. Schloss

2014-2018 **BS Biology**, *Dept. of Biology*, University of Kentucky
○ Minor: Computer Science

Skills

Languages & Tools R, Python, C++, Bash, Snakemake, git, GitHub, R Markdown, Jupyter, Quarto, LaTeX, conda/mamba, Docker, Singularity, SLURM, Unix CLI.

Research supervised machine learning, data visualization, reproducible manuscripts, gut microbiome composition, amplicon sequence analysis, metagenomics, bulk RNA-seq.

Software package maintenance, test-driven development, continuous integration, documentation, collaboration & peer review, high performance computing.

Research Experience

2019-present **Graduate Student Researcher**, *Schloss Lab, Dept. of Microbiology and Immunology*, University of Michigan

- Develop, benchmark, and maintain bioinformatics workflows and software packages in R & Python.
- Build machine learning pipelines for human gut microbiome classification problems in colorectal cancer and *C. difficile* infection.
- Publish findings as reproducible papers.
- Collaborate with other scientists on microbiome projects and mentor junior lab members.

2018-2019 **Rotation Student Researcher**, *Program in Biomedical Sciences*, University of Michigan

- Worked in several different bioinformatics labs to find the best mentorship fit.

2015-2018 **Undergraduate Lab Assistant**, *Moseley Bioinformatics Lab, Dept. of Molecular and Cellular Biochemistry*, University of Kentucky

- Developed a Python package for comparative genomics. Identified sets of orthologous and paralogous gene products to facilitate collinearity analysis and detection of gene duplication events.

- 2016-2018 **BIO395 Independent Research Student**, *Weisrock Lab, Dept. of Biology*, University of Kentucky
- Developed bash scripts and a SNP calling pipeline in Snakemake.
 - Population structure analysis of the *Ambystoma tigrinum* species complex.
 - Bayesian species delimitation of the *Desmognathus fuscus* species complex.
- 2015-2016 **Undergraduate Lab Assistant**, *Jaromczyk Lab, Dept. of Computer Science*, University of Kentucky
- Maintained the *Epichloë festucae* genome project database.
 - Analyzed RNA-seq data of *Chenopodium quinoa* and coffee ringspot virus.

Teaching Experience

- Jan-Apr 2023 **Graduate Student Instructor**, *BIOINF 576: Tool Development for Bioinformatics. Dept. of Computational Medicine & Bioinformatics*, University of Michigan
- Develop curriculum and teach students the principles of software development in R & Python. Topics: software design, implementation, testing, documentation, issue tracking, peer review, and release.
- 2019-2022 **Facilitator & Capstone Project Mentor**, *Girls Who Code at U-M DCMB*, University of Michigan
- Weekly Club during the school year and week-long Summer Experience for high schoolers to learn Python for data science
- 2018-present **Workshop Instructor & Helper**, *U-M Carpentries*, University of Michigan
- 2-day Software Carpentry workshops teaching computational skills for reproducible research
- Jun 2022 **Instructor**, Virtual
- Intro to R & RNA-Seq Workshop for ASM Microbe conference attendees
- Apr 2019 **DNA Day Ambassador**, *Michigan DNA Day*, Pioneer High School, Ann Arbor, MI
- Epigenetics & Scientific Journeys
- Mar 2019 **Workshop helper**, *Graduate Society of Black Engineers and Scientists*, University of Michigan
- Data Visualization with Python Workshop
- Mar 2019 **Capstone Activity Leader**, *Females Excelling More in Math, Engineering, & the Sciences*, University of Michigan
- Binary Numbers through Ozobots with GWC at U-M DCMB
- 2012-2018 **Tutor**, *freelance*
- for high school and college students in Biology, Calculus, Chemistry, Computer Science, and Bioinformatics.

Service

- 2019-2022 **Executive Committee Member: Open Source Technical Lead**, *Girls Who Code at U-M DCMB*
- Plan, apply for funding, develop curriculum, and maintain resources to teach introductory Python programming & data science to young women+.
 - Facilitate collaborative development and maintenance of our open source teaching resources.
 - Organize our year-round Club and annual Data Science Summer Experience for high school women+.
- 2019-2022 **Organizer and Maintainer**, *U-M Carpentries*
- Co-lead development & maintenance of a curriculum for workshops teaching programming skills for reproducible research.
 - Maintain the website, develop curriculum, and organize workshops.
 - Collaborate with U-M Women in Science and Engineering to organize workshops for women+.
- 2021-2023 **Mentor**, *Schloss Lab*
- Mentor an undergraduate student in building reproducible machine learning models to predict *C. difficile* infection severity from gut microbiome composition.
- 2021-2022 **CoderSpaces co-host**, *U-M ISR Data Science Hub*
- Hold office hours at a weekly virtual help session for data science practitioners
- 2021 **Peer reviewer**
- PLOS ONE (1)
- 2019-2021 **Graduate Student Coordinator**, *U-M Data Analysis Networking Group*
- Organize monthly meetings & a one-day symposium for researchers to sharpen their data analysis skills.
 - Apply for funding through a Rackham Interdisciplinary Workshop grant.
- 2009-present **Volunteer Sound Engineer**, *various churches and non-profit organizations*
- Mix FOH, monitors, recording, and live stream.

Open Source Contributions

Software

mikropml, *User-Friendly R Package for Supervised Machine Learning Pipelines*

- Co-author and maintainer

schttools, *Schloss Lab tools for reproducible microbiome research*

- Co-author and maintainer

mikropml snakemake workflow, *Snakemake template for building reusable and scalable machine learning pipelines with mikropml*

- Co-author and maintainer

mothur, *Command-line tool for processing microbial amplicon sequence data*

- Contributor

mothur snakemake workflow, *Snakemake template for reproducible microbial amplicon sequence analysis with mothur*

- Co-author

Curricula

U-M Software Carpentry, *Intro to R, the Unix shell, and git for workshops on reproducible research.*

○ Co-author and maintainer

Girls Who Code at U-M DCMB, *Intro to Python for Data Science via Live-Coding for Girls Who Code clubs.*

○ Co-author and maintainer

Code Clubs, *Short coding tutorials for lab meetings*

○ Contributor

U-M DANG!, *repro-packs: Organizing projects for reproducibility and headache prevention.*

○ Author

Intro to R & RNA-seq, *Workshop for 2022 ASM Microbe attendees*

○ Contributor

Presentations

Talks

Nov 2022 **Bioinformatics Student Research Hour**. Predicting *C. difficile* infection severity from the taxonomic composition of the gut microbiome. University of Michigan

Feb 2022 **Seminar for the KG Jebsen Center for Genetic Epidemiology**. Intro to git & GitHub. (Virtual) Norwegian University of Science and Technology

Mar 2021 **Bioinformatics Student Research Hour**. OptiFit: a fast method for fitting amplicon sequences to existing OTUs. (Virtual) University of Michigan

Apr 2018 **Systems Biology and Omics Integration Seminar**. Developing a Global Homology Analysis for Comparative Genomics. University of Kentucky

Posters

Jun 2022 **ASM Microbe**. Predicting the severity of *C. difficile* infections from the taxonomic composition of the gut microbiome. Washington, DC

Jun 2020 **ASM Microbe**. OptiFit: a fast method for fitting amplicon sequences to existing OTUs. Virtual

Apr 2018 **Showcase for Undergraduate Scholars**. Developing a Global Homology Analysis for Comparative Genomics. University of Kentucky

Apr 2018 **National Conference on Undergraduate Research**. Developing a Global Homology Analysis for Comparative Genomics. University of Central Oklahoma

Apr 2016 **Showcase for Undergraduate Scholars**. Processing RNA-seq Reads of Plants Infected with the Coffee Ringspot Virus. University of Kentucky

Apr 2016 **UT-KBRIN Bioinformatics Summit**. Processing RNA-seq Reads of Plants Infected with the Coffee Ringspot Virus. Cadiz, KY

Apr 2015 **Showcase for Undergraduate Scholars**. The Effect of Meditation on Performance. University of Kentucky

Awards

Grants and Fellowships

- 2022 **Conference Travel Grant.** Rackham Graduate School, University of Michigan (\$900)
- 2020-2021 **Rackham Interdisciplinary Workshop Grant.** Rackham Graduate School, University of Michigan (\$500)
- 2020 **Conference Travel Grant.** Rackham Graduate School, University of Michigan (\$800)
- 2019-2020 **Rackham Interdisciplinary Workshop Grant.** Rackham Graduate School, University of Michigan (\$500)
- 2019-2021 **NIH T32 Bioinformatics Training Program Fellow.** Bioinformatics Graduate Program, University of Michigan
- 2017 **Oswald Research & Creativity Award, 2nd place in Biological Sciences.** Office of Undergraduate Research, University of Kentucky (\$200)
- 2017 **Summer Research Grant.** Office of Undergraduate Research, University of Kentucky (\$2,000)
- 2014-2018 **Presidential Scholarship.** University of Kentucky (out-of-state full tuition)

Honors

- 2018 **Graduated Cum Laude with Departmental Honors in Biology.** University of Kentucky
- 2018 **Biology Undergraduate Research Award Nominee.** University of Kentucky
- 2014-2018 **Lewis Honors College.** University of Kentucky

Continuing Education

- Jan 2020 **Building Tidy Tools workshop at rstudio::conf.** Rstudio, PBC, San Francisco, CA
- Dec 2019 **Winter School in Research Software Engineering.** US Research Software Sustainability Institute, Seattle, WA
- Jun 2019 **Software Carpentry Instructor Training.** The Carpentries, University of Michigan
- May 2019 **PyCon Education Summit & Conference.** The Python Software Foundation, Cleveland, OH
- Apr 2019 **MICROBIOL 612.2.** Riffomonas Minimal R Workshop, University of Michigan

Publications

* Indicates co-first author

9. Armour CR, **Sovacool KL**, Close WL, Topçuoğlu BD, Wiens J, Schloss PD. Sep 2022. Streamlined implementation of a machine learning model to classify screen relevant neoplasia using reference-based OTU clustering. *bioRxiv*. 10.1101/2022.09.01.506299

8. Barron MR, **Sovacool KL**, Abernathy-Close L, Vendrov KC, Standke AK, Bergin IL, Schloss PD, Young VB. Jul 2022. Intestinal Inflammation Reversibly Alters the Microbiota to Drive Susceptibility to *Clostridioides Difficile* Colonization in a Mouse Model of Colitis. *mBio*. 10.1128/mbio.01904-22
7. **Sovacool KL**, Westcott SL, Mumphrey MB, Dotson GA, Schloss PD. Feb 2022. OptiFit: An Improved Method for Fitting Amplicon Sequences to Existing OTUs. *mSphere*. 10.1128/msphere.00916-21
6. *Lapp Z, ***Sovacool KL**, Lesniak N, King D, Barnier C, Flickinger M, Krüger J, Armour CR, Lapp MM, Tallant J, Diao R, Oneka M, Tomkovich S, Anderson JM, Lucas SK, Schloss PD. Jan 2022. Developing and Deploying an Integrated Workshop Curriculum Teaching Computational Skills for Reproducible Research. *JOSE*. 10.21105/jose.00144
5. *Duda M, ***Sovacool KL**, Farzaneh N, Nguyen VK, Haynes SE, Falk H, Furman KL, Walker LA, Diao R, Oneka M, Drotos AC, Woloshin A, Dotson GA, Kriebel A, Meng L, Thiede SN, Lapp Z, Wolford BN. Dec 2021. Teaching Python for Data Science: Collaborative Development of a Modular & Interactive Curriculum. *JOSE*. 10.21105/jose.00138
4. *Topçuolu BD, *Lapp Z, ***Sovacool KL**, Snitkin E, Wiens J, Schloss PD. May 2021. Mikropml: User-Friendly R Package for Supervised Machine Learning Pipelines. *JOSS*. 10.21105/joss.03073
3. Everson KM, Gray LN, Jones AG, Lawrence NM, Foley ME, **Sovacool KL**, Kratovil JD, Hotelling S, Hime PM, Storfer A, Parra-Olea G, Percino-Daniel R, Aguilar-Miguel X, O'Neill EM, Zambrano L, Shaffer HB, Weisrock DW. Apr 2021. Geography Is More Important than Life History in the Recent Diversification of the Tiger Salamander Complex. *PNAS*. 10.1073/pnas.2014719118
2. Hagan AK, Lesniak NA, Balunas MJ, Bishop L, Close WL, Doherty MD, Elmore AG, Flynn KJ, Hannigan GD, Koumpouras CC, Jenior ML, Kozik AJ, McBride K, Rifkin SB, Stough J, **Sovacool KL**, Sze MA, Tomkovich S, Topcuoglu BD, Schloss PD. Aug 2020. Ten Simple Rules to Increase Computational Skills among Biologists with Code Clubs. *PLoS Comput Biol*. 10.1371/journal.pcbi.1008119
1. Goodin MM, Farman M, Inocencio H, Jang C, Jaromczyk JW, Moore N, **Sovacool KL**. Aug 2016. Processing RNA-Seq Data of Plants Infected with Coffee Ringspot Virus. *BMC Bioinform*. 10.1186/s12859-016-1154-y