

# Schuyler D. Smith, Ph.D.

✉ [schuyler.d.smith@gmail.com](mailto:schuyler.d.smith@gmail.com)    [github.com/schuyler-smith](https://github.com/schuyler-smith)    [linkedin.com/in/schuyler-ds](https://www.linkedin.com/in/schuyler-ds)  
 [schuyler-smith.github.io](https://github.com/schuyler-smith)    [Google Scholar](#)    +1 (413) 212-9110    Durham, NC

## Education

**Ph.D. Bioinformatics and Computational Biology** Dec 2021  
Iowa State University | Ames, Iowa  
Dissertation: *Characterization of environmental microbiomes impacted by Iowa agriculture*

**M.S. Quantitative Genetics and Plant Breeding** May 2015  
Texas A&M University | College Station, Texas  
Thesis: *Molecular Characterization of the Texas Maize Breeding Program*

**B.S. Genetics** Dec 2012  
Iowa State University | Ames, Iowa

## Experience

**Senior Computational Scientist, Bioinformatics** Jan 2022 – Present  
**Nutrien Ag Solutions (Waypoint Analytical) | Soil Biome | Champaign, Illinois**

- Designed and deployed end-to-end NGS pipelines, scaling metagenomic data processing to tens of thousands of samples.
- Implemented cloud-based NextFlow workflows on AWS/GCP, reducing runtimes by 40% and ensuring reproducibility.
- Engineered custom C++ tools for high-throughput lab data processing and QC, accelerating turnaround times.
- Authored R and Python packages for microbiome analytics.
- Built an automated soil-biome reporting platform, streamlining client deliverables.
- Developed ML-based predictive models linking microbiome metrics to product placement decisions.
- Created RAG-based GenAI chatbot for recommendations built into product pipeline.
- Designed and maintained gRPC APIs, enabling seamless integration with downstream applications.
- Project lead for development and implementation of metagenomic database.
- Delivered stakeholder-ready analyses, visualizations, and presentations to inform strategic initiatives.
- Mentored 5+ junior scientists and interns, fostering skills in bioinformatics workflows and best practices for data science.

**Graduate Research Assistant – Ph.D.** Jan 2017 – Dec 2021

**Iowa State University | GERMS Laboratory | Ames, Iowa**

- Developed reproducible pipelines for 16S rRNA and shotgun metagenomics, supporting high-throughput microbiome studies.
- Authored R packages enabling scalable microbiome analytics, adopted by collaborators and cited in peer-reviewed publications.
- Published research on antibiotic-resistance gene dispersal, advancing knowledge in environmental microbiology.
- Identified microbial predictors of harmful algal blooms, informing water-quality management strategies.
- Conducted large-scale marker detection studies, improving methodology for microbiome gene identification.

- Mentored junior lab members in bioinformatics and experimental methods.
- Instructed Carpentries workshops, training hundreds of students and professionals in reproducible computing practices.

#### **Graduate Research Assistant – Ph.D.**

*Jun 2015 – Sep 2016*

##### **University of Wisconsin – Madison | Potato Genetics Lab | Madison, Wisconsin**

- Developed NGS and SNP-calling pipelines for autotetraploid crops, supporting population-scale genomic analyses.
- Created R packages implementing ML-based genotyping, increasing throughput by 25%.

#### **Biological Science Technician – Internship**

*Jun 2014 – Dec 2014*

##### **USDA-ARS | Arid-Land Agricultural Research Center | Maricopa, Arizona**

- Evaluated high-throughput image-based phenotyping systems to enhance crop trait measurement.

#### **Graduate Research Assistant – M.S.**

*Jan 2013 – May 2015*

##### **Texas A&M University | Maize Genetics Lab | College Station, Texas**

- Published genetic marker characterization of Texas maize germplasm, informing breeding program decisions.

#### **Maize Breeding Intern**

*May 2012 – Sep 2012*

##### **Monsanto Company | Huxley Research Station | Huxley, Iowa**

#### **Maize Product Trait Development Intern**

*Jun 2011 – Dec 2011*

##### **DuPont Pioneer | Willmar Research Station | Willmar, Minnesota**

## **Applicable Skills**

**Programming & Scripting:** Python, R, C++, Shell, SQL

**Bioinformatics & Genomics:** NGS, 16S rRNA, Metagenomics, Microbiome Analysis, qPCR Pipelines, GBS, GWAS, RNA-Seq, WGS, Variant Calling

**Machine Learning & Modeling:** TensorFlow, scikit-learn, Supervised learning: regression, random forests, gradient boosting, SVMs; Unsupervised learning: clustering, PCA, t-SNE; Predictive modeling; Network analysis (igraph)

**Data Analysis & Visualization:** pandas, data.table, ggplot2, seaborn, plotly, R Markdown, Jupyter

**Workflow & Cloud:** NextFlow, Snakemake, Airflow, Kubernetes, AWS, GCP, SLURM

**DevOps / CI/CD:** Docker, Singularity/Apptainer, Git, GitHub Actions, Travis CI, Jenkins

**Dashboards & Reporting:** Shiny, Dash, Power BI

**Web & Document Tools:** Markdown, HTML/CSS,  $\LaTeX$

**Image & Data Analysis Tools:** ImageJ

**Databases & APIs:** MySQL, PostgreSQL, SQLite, gRPC

**Operating Systems:** Linux (Desktop/Server), Windows, MacOS, Proxmox VE

## **Select Projects**

### **Soil Laboratory Quality Control Automation**

*Python, scikit-learn, regression, random forest, XGBoost*

Implemented machine learning models to detect failed samples in laboratory pipelines, reducing manual review time and improving client turnaround.

- Scalable Metagenomics Pipeline** *NextFlow, Airflow, AWS, Docker, DIAMOND, Kraken, MEGAN*  
Developed a reproducible pipeline for large-scale soil microbiome sequencing and annotation, reducing runtime by 40% and enabling processing of terabyte-scale datasets.
- qPCR Quality Control Tool** *C, Python, Dash, plotly, AWS, GCP, Docker*  
Designed and created a CLI C-software for automated quality control of qPCR data, reducing manual review time and improving data accuracy for downstream analysis. Then developed an interactive dashboard for visualizing qPCR results and trends over time.
- Soil-Biome Report** *Python, Figma, Docker, AWS, gRPC, PostgreSQL*  
Worked with team to design a PDF report for customers who wanted soil-biome analysis. Authored the python package implemented in gRPC for processing the biome data for the report format.
- Soil Biological Product Recommendation System** *Python, Dash, Docker, GCP, Gemini, RAG, GenAI*  
Built a recommendation platform integrating RAG and GenAI to suggest soil biological products based on microbiome profiles and crop type, enabling more data-driven and customized client decisions.
- Product Trial Analyses** *R, R Markdown, ggplot2, Jupyter Notebooks*  
Analyzed agricultural product trial datasets and generated reproducible reports for internal teams and external stakeholders, supporting product development and strategy decisions.

## Selected Publications

- Smith, S. D., Velásquez-Zapata, V., & Wise, R. P. (2025). *Ngpint v3: A containerized orchestration python software for discovery of next-generation protein-protein interactions*. *Bioinformatics*, 41(6).
- Velásquez-Zapata, V., Elmore, J. M., Patel, S., Smith, S. D., Fuerst, G., & Wise, R. P. (2024). *Effector-host protein networks in an integrated barley-powdery mildew interactome*. *Plant and Animal Genome Conference/PAG 31 (January 12-17, 2024)*.
- Velásquez-Zapata, V., Smith, S. D., Surana, P., Chapman, A., & Wise, R. (2023). *Transcriptome-based host epistasis and pathogen co-expression in barley-powdery mildew interactions* [Preprint]. *bioRxiv*.
- Smith, S. D., Choi, J., Ricker, N., Yang, F., Hinsla-Leasure, S., Soupir, M. L., Allen, H. K., & Howe, A. (2022). *Diversity of antibiotic resistance genes and transfer elements-quantitative monitoring (darte-qm): A method for detection of antimicrobial resistance in environmental samples*. *Communications Biology*, 5(1), 216.
- Smith, S. D. (2019). *Phylosmith: An r-package for reproducible and efficient microbiome analysis with phyloseq-objects*. *Journal of Open Source Software*.
- Smith, S. D., Colgan, P., Yang, F., Rieke, E., Soupir, M., Moorman, T., Allen, H., & Howe, A. (2019). *Investigating the dispersal of antibiotic resistance associated genes from manure application to soil and drainage waters in simulated agricultural farmland systems*. *PLoS One*, 14(9), e0222470.

## Workshops Taught

- |   |             |
|---|-------------|
| Data Carpentry: Genomics   Oklahoma State University   online   | 2021        |
| Software Carpentry: Shell, Git, R for Reproducible Scientific Analysis   University of Idaho   online | 2021        |
| Data Carpentry: Ecology   Merck & Co.   online  | 2020        |
| Data Carpentry: Genomics   George Washington University   Washington, D.C.                            | 2020        |
| Data Carpentry: Ecology   Merck & Co.   Upper Gwynedd, PA   | 2019        |
| Girls in Science Initiative   Science Center of Iowa   Des Moines, IA                                 | 2019        |
| Introduction to Data Analysis   Iowa State University   Ames, IA                                      | 2017 & 2018 |
| Iowa State BCB Data Analysis Language Workshops Introduction to Unix                                  | 2017 & 2018 |
| Iowa State BCB Data Analysis Language Workshops Introduction to Python                                | 2017        |
| Explorations in Data Analyses for Metagenomic Advances in Microbial Ecology                           | 2017 & 2018 |