Project title: Data Integration to investigate the decline of Soil Microbial Biomass and Aquatic Microbial Cell Density and its causes

Project abstract:

Soil microbes contribute to climate change through nutrient cycling and greenhouse gas flux. For example, soil microbes produce extracellular enzymes to break down detritus to produce carbon dioxide, exclusively carry out methane production and oxidation processes, aid plant nitrogen uptake, and drive nitrogen cycling. Therefore, how much microbes inhabit soils is a predominant factor controlling carbon cycling in terrestrial ecosystems. Soil microbes are not in a steady state; rather, they are a group of living individuals that change over time in a changing environment. In a recent study in my laboratory, we found soil microbial biomass declined from 1992 to 2013, published in Nature Communication in 2022. However, the reason for this decline remains elusive due to a lack of mechanistic understanding. Modeling approaches are a valuable supplement to field experiments in understanding microbial processes in terrestrial ecosystems. My lab has been developing the CLM-Microbe model to represent microbial processes over the past ten years. This project will expand our existing data and integrate the data with the CLM-Microbe model to examine the causes of the decline of soil microbial biomass. With this project, the student will be trained in various skills: meta-analysis, data quality control, data visualization, biostatic analysis, and modeling techniques.

Application type: One undergraduate mentee

Project description:

<u>Project Overview</u>

A recent collaborative study in my group reported a declining soil microbial biomass across the globe from 1992 to 2013 due to climate warming. Later, a "Matters Arising" was accepted for publishing regarding the declining soil microbial biomass and its causes. The manuscript has been published, and the response has been accepted for publication in Nature Communication. This debate inspires further research to address this question, which is also a part of my funded NSF CAREER project – Integrating a microbial data system with a microbial-explicit model to investigate microbial biogeochemistry. In this proposed effort, my group will hire an undergraduate student who has registered to be an undergraduate in my group to work on this topic in Spring 2024. The overarching goal is to prepare the data and model to understand the mechanisms of declined microbial biomass in the past three decades (1992-2020). Over the 10 weeks, the student will be trained to do five tasks: 1) developing a framework for a meta analysis, 2) compilating observational data from literature and online resources; 3) controlling the data quality by double-check the unit, data sources, cross-validation; 4) exploratory analysis with a biostatistics approach and data visualization, and 5) gaining basic skills of microbial model application. To make a reasonable timeline over the 10

weeks, we will separate the project into five tasks that take 1-3 weeks each. At the end of the project, the data will serve for my Ph.D. student to carry out a model application to examine the decline of soil microbial biomass and its causes in the past three decades (extended to include 1992-2020). The major novelty of this project is integrating data with a model to understand the mechanism for the declining microbial biomass.

Mentorship and training plan

The undergraduate student will be mentored by one of my Ph.D. students and me through the 10 weeks. The mentoring of undergraduate students focuses on engagement with the scientific literature (searching and identifying literature, formatting microbial data, and data visualization with R programming) and communication skills; meanwhile, training opportunities on data extraction, quality control, data analysis, and visualization will be provided. These practices blend microbial ecology concepts with quantitative skills to promote students' understanding of the interdisciplinarity of STEM fields. Over the past five years, my lab has accumulated many R scripts for all those data processing, analysis, and visualization archived in my GitHub repositories. It will be provided as training tools for the students. With the support of my NSF CAREER project, I have trained four PhD students, one master's student, six undergraduate students, and six high school students on data compilation and modeling techniques. These successful mentorship and training activities provide a solid foundation for the student in this project. More specific skills the student receives are shown in the next section of planned activities.

Planned activities

The undergraduate student will be trained to carry out five activities, listed below with details. 1) In the first activity, the student will learn how to apply the scientific method to address the question of declining microbial biomass over a decadal scale. The student will be guided to read the previous publication in my lab and learn how to develop a hypothesis, plan data collection, and potentially data analysis targeting hypothesis testing.

2) In the second activity, the student will be trained to compile data from published literature and other online datasets. As this activity expands our previous dataset

(https://doi.org/10.3334/ORNLDAAC/1264), the student will adopt our previous procedure to finish this task.

3) In this third activity, the student will be trained to do data quality control, primarily focusing on unit conversion, data gap filling, and data cross-validation. This will also follow previous procedures developed in my lab which have been archived in my GitHub account (https://github.com/email-clm).

4) In this activity, the student will learn how to do exploratory analysis. As the identified student has just taken the Biostatistics course (BIO215) in my class in Fall 2023, this provides a good opportunity to implement and practice what he learned.

5) As the final data will be used to integrate with the model, the student will be provided an opportunity to learn the modeling technique, which will provide information for the student to prepare the data for model use in the next step.

Timeline

Date	Tasks – major deliverable / milestone	Training skills
June 3rd	Student initialized the project	
June 7th	Student finished the overall framework of the proposed effort	Scientific method, overall idea of meta analysis
June 21st	Finished the expansion of the existing dataset by compilating recent 5 years of data from literature and online resources	searching and identifying literature; data extraction; major microbial variables and their meaning
June 28th	Finished the data quality control and unit conversion	quality control; unit conversion, data standardization
July 12th	Finished the exploratory data analysis by using biostatistics approach	Biostatistics analysis, data exploratory analysis, average, standard derivation, standard error
July 26th	Data-model integration	R programming, CLM-Microbe model

August 9th	**Presentation for undergraduate research symposium in Fall 2024**	**Organizing the research output; presentation skills**

Besides these identified major milestones over the 10 weeks, I will have a one-on-one meeting with the student weekly. The meeting will help the student to identify any problems with the tasks and provide immediate assistance. Meanwhile, during the weekly meeting, my Ph.D. students and postdoctoral researchers will have opportunities to interact with the students to provide suggestions for data analysis and modeling techniques.