8-8-2019

Leveraging synthetic root-soil systems to quantify relationships between plant traits and the formation of soil organic carbon

Bonnie G. Waring
Utah State University, bonnie.waring@usu.edu

Appendix 6: Data Management Plan

Physical sample storage and data provenance
Our proposed laboratory incubation experiment consists of 648 synthetic soil microcosms and 216 field soil controls. Each microcosm will have a unique alphanumeric ID that indicates the treatments to which it is assigned. This unique ID will be associated with all measurements taken on soils within the microcosm, allowing us to trace the provenance of each individual sample analyzed throughout the experiment.

Physical samples generated as part of this work (soils, DNA extracts, microbial biomass extracts) will be stored in the Waring Lab at USU for at least 2.5 years after the date relevant manuscripts are published. Each physical sample will have a unique reference number that corresponds to the data provenance scheme described above. All laboratory notebooks related to this work will be permanently stored in the Waring lab, according to the lab's established Standard Operating Protocol.

Data types and sources
During the course of this research, three main classes of data will be generated:

- **Soil biogeochemical data.** For each synthetic soil microcosm, we will measure:
  - soil CO_2 fluxes, quantified biweekly
  - soil carbon pools (microbial biomass C, soil organic C in particulate and mineral-associated fractions)
  - soil nitrogen pools (microbial biomass N, soil organic N in particulate and mineral-associated fractions, inorganic N)
  - the isotopic signatures of each C and N pool described above
  - soil extracellular enzyme activities

  These data will be stored in CSV file format. Each observation will carry the unique alphanumeric IDs described above, and will be linked to a metadata file reporting the sample type, date of collection, and details about any experimental treatments applied. Any raw data files used to generate the CSVs (e.g. outputs of the Shimadzu Gas Chromatograph software) will also be archived, and the file paths will be indicated in the metadata file.

- **Soil metagenomics data.** The proposed work will generate large metagenomics sequence libraries; sequences will be stored in FASTQ file format, and metadata in CSV file format. We will ensure that metadata files comply with the Genome Standards Consortium's Minimum Information about Metagenome Sequences (MIMS) and the recommendations of ten Hoopen et al (2017). Thus, at a minimum, associated metadata will report the following: sample-level information (date of collection, treatments applied in the lab, associated soil parameters); experiment-level information (sequencing platform and library preparation methods used); and analysis information (sufficient to replicate workflows for functional and taxonomic analysis of sequence reads). Additionally, quantitative PCR will be conducted using 16S and ITS primers; data from these qPCR runs will also be stored in native file format, and data summaries in CSV file format.

- **Model simulation code and results.** We will conduct 72 model simulation exercises using three biogeochemical models: MEND (Wang et al. 2013), MIMICS (Wieder et al. 2014), and CORPSE (Sulman et al. 2014). Model simulations will be conducted in the R

programming language. Data associated with the model simulation exercises will consist of:

- Model source code, sufficient to replicate all 72 simulations (see note below)
- Parameter files required to run simulations
- Data files used for model validation and evaluation
- Model simulation results, including raw outputs and associated figures
- Meta-data, including help files and other documentation

Model code, parameter files, simulation results, and metadata will be stored in a GitHub repository managed by the PI.

Note: Original MEND, MIMICS, and CORPSE model code will need to be modified by our team at USU in order to run/validate/compare the models in a common testbed framework, and it is this modified code that will be pushed to GitHub. I will solicit permission from MEND, MIMICS, and CORPSE model authors before pushing code to a public repository. MIMICS and CORPSE model code are already publically available on GitHub, and these repositories (as well as the associated publications) will be extensively cited in model documentation. In the event that the original model authors do not want model code made publically available, only parameter files, data files, model simulation results, and meta-data will be linked to a public repository. These data, along with the model descriptions in original publications (Wang et al. 2013, Wieder et al. 2014, Sulman et al. 2014) should be sufficient to reproduce the analyses.

Finally, our proposed work involves **methods development**, i.e. testing and validation of experimental approaches to create synthetic root/soils system. We will make these protocols publically available to the scientific community by publishing an open-access article in the *Journal of Visualized Experiments.* This journal works with authors to script and film instructional videos that illustrate the scientific methods described in the associated peer-reviewed manuscripts.

Sharing and preservation

Data sharing among the project team

As data is generated, it will be backed up in at least three locations: the data generator's computer (e.g. that belonging to the graduate student or research assistant), the PI's laboratory computer, and in the cloud via the PI's USU Box account. The PI will maintain responsibility for ensuring that all datasets are appropriately stored.

Data sharing with the public

All empirical data generated by this research will be made freely available to the public following quality control and analysis at the following repositories:

- Biogeochemical datasets will be uploaded to the Environmental System Science Data Infrastructure for a Virtual Ecosystem (**ESS-DIVE**) data archive (https://ess-dive.lbl.gov) and Dryad Digital Repository (http://www.datadryad.org)
- All sequences generated from the metagenomics study will be deposited in the NCBI Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra), the Department of Energy Joint Genome Institute's Integrated Microbial Genomes and Microbiomes (IMG/M) database (https://img.jgi.doe.gov/), and **ESS-DIVE**.
- Model code and model simulation results will be stored in a public repository on GitHub (https://github.com/). Although code will be updated throughout the course of the research described here, final versions associated with resulting publications will be archived in a separate, permanent sub-repository.

Data will be uploaded to these repositories at the time of manuscript submission or one year following completion of the associated experiment, whichever comes first.

All data generated as part of this grant will also be stored on Utah State University's publically accessible Digital Commons Platform. DigitalCommons@USU supports all file types and formats. Files are provided with persistent URLs, and if needed, USU Library staff can obtain DOIs for datasets. The system is able to produce license and copyright statement as needed, and creates standard citations. All files are backed up at multiple sites, including cloud storage. Preservation copies are stored in Amazon Web Services, with redundant storage across multiple facilities and are regularly verified for integrity of data using checksums. Moreover, the Digital Commons indexes publications and datasets by grant number, allowing the public to access any material produced with DOE funding in a single convenient location.

References

Sulman BS, Phillips RP, Oishi AC, Shevliakova E, Pacala SW. 2014. Microbe-driven turnover offsets mineral-mediated storage of soil carbon under elevated CO_2. *Nature Climate Change* 4: 1099-1102.

ten Hoopen P, Finn RD, Bongo LA, et al. 2017. The metagenomics data life-cycle: standards and best practices. *GigaScience* 6: gix047.

Wang G, Post WM, Mayes MA. 2013. Development of microbial-enzyme mediated decomposition model parameters through steady-state and dynamic analyses. *Ecological Applications* 21: 255-272.

Wieder WR, Grandy AS, Kallenbach CM, Bonan GB. 2014. Integrating microbial physiology and physio-chemical principles in soils with the MIcrobial-MIneral Carbon Stabilization (MIMICS) model. *Biogeosciences* 11: 3899-3917.