

Utah State University

DigitalCommons@USU

Funded Research Records

11-13-2019

Collaborative Proposal: MSA: Predicting the effects of nitrogen deposition on the soil carbon sink with a continental-scale experiment

Bonnie G. Waring

Utah State University, bonnie.waring@usu.edu

Follow this and additional works at: https://digitalcommons.usu.edu/funded_research_data



Part of the [Biology Commons](#)

Recommended Citation

Waring, Bonnie G., "Collaborative Proposal: MSA: Predicting the effects of nitrogen deposition on the soil carbon sink with a continental-scale experiment" (2019). *Funded Research Records*. Paper 125.

https://digitalcommons.usu.edu/funded_research_data/125

This Grant Record is brought to you for free and open access by DigitalCommons@USU. It has been accepted for inclusion in Funded Research Records by an authorized administrator of DigitalCommons@USU. For more information, please contact rebecca.nelson@usu.edu.



Data Management and Access Plan

Data provenance

Our proposed experiment consists of replicated N fertilization experiments conducted within 20 plots distributed across ten NEON core terrestrial sites. Materials to construct soil cores will be shipped to all participating NEON sites at the start of the experiment. Each core will have a unique alphanumeric ID that indicates the NEON site to which it is deployed, the replicate plot to which it belongs, the treatment to which that core is assigned, and a replicate number. Thus, a complete core ID would take the format “OSBS-P1-Nfert-4,” for example. We will use a similar system to identify samples taken as part of initial site surveys (e.g. “OSBS-P1-initial-4.”) This unique ID will be associated with all measurements taken on soils within the core, allowing us to trace the provenance of each individual sample analyzed throughout the experiment.

Sample storage

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 and field 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, two main classes of data will be generated:

- **Soil biogeochemical data.** For both initial soil characterizations and for core harvests, we will measure:
 - soil pH (measured at the core scale) and texture (measured at the plot scale)
 - 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, geographic coordinates of the site of collection, climatic information pertinent to the site of collection, and details about any experimental treatments applied.

- **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 (site/date of soil collection, treatments applied in the field, 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.

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.

At monthly intervals, all data will be uploaded to a Dropbox folder accessible by all Co-Investigators, students, and research assistants. Prior to gaining access to this Dropbox, all project staff will receive appropriate training on appropriate policies for submission and curation of data within the team folder.

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 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>) and the Department of Energy Joint Genome Institute's Integrated Microbial Genomes and Microbiomes (IMG/M) database (<https://img.jgi.doe.gov/>).

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, 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 NSF funding in a single convenient location.

References

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