Code of Conduct & Scientific Integrity
Bailey, Kivlin & Schweitzer Labs
(ver. Spring 2022)

I. Code of Conduct

We welcome participants, students, and collaborators regardless of race, religion, gender identification, sexual orientation, age, or disability status. We honor the fact that diverse perspectives lead to diverse approaches and better science. It is through diverse perspectives and abilities that we can find answers to any given problem.

- Safety (physical and mental) and scientific integrity are the top priorities for the Bailey, Kivlin, and Schweitzer labs.
- All lab spaces are fair and equitable workplaces, where every member's opinions and questions are valued and respected.
- Lab Group meetings are a collaborative meeting, where all feedback is valid (however small or large a contribution). All perspectives in lab group are respected. Lab group is meant to provide constructive feedback on members' research and ideas, and in no way will be used to criticize the character of an individual. To ensure equitable conversations, be respectful of everyone's opportunity to speak (i.e., note if you are the person talking the most, and adjust).
- Research ethics (see Research Integrity below)
    - Always practice a high standard of personal and scientific integrity.
    - Use multiple checks and balances to ensure data integrity.
    - Hold each other accountable for upholding research ethics.
- Communication
    - All incoming graduate students, undergraduate researchers, and visitors will understand the goals of specific projects and overall goals of the labs through training upon entry to the lab.
    - Communication will occur through email, Slack and one-on-one conversation.
    - Communication among all lab members will be respectful, honest, and specific.
- Lab upkeep and maintenance
    - Only use a piece of lab equipment after you have been sufficiently trained in using that instrument.
    - Always clean up your workstation and put away materials when finished with a particular task, even if returning later or the next day.
    - Always fill out sign-in/out equipment sheet when borrowing <u>any</u> lab equipment.
    - Always notify lab PI's if equipment breaks or major chemical spills occur.

    - Notify all lab group members immediately of any ceiling leaks, alarms, or other mishaps.
- When problems arise among lab members or with PI's, voice your concerns immediately, and request a meeting with all parties. Consult with the UTK Ombudsperson when unsure of your rights or responsibilities.
- If major problems arise with projects, as they can (e.g., overfertilization that kills plants), please notify PI immediately so appropriate action can be taken quickly.

II. Research Integrity – Responsible Conduct of Research

Scientific integrity rests on research integrity. It is our duty as scientists to be as unbiased, transparent, careful, and honest as possible about our methods and data. As researchers, our shared values include trust, accountability, mutual respect, and fairness in all aspects of our research endeavors.

Guidelines for all research, whether in field, lab or greenhouse:

- Make and follow written protocols and explain data collection instructions thoroughly to all researchers before the work begins.
- Keep a detailed notebook where all notes, observations and documentation of methods and approaches are recorded. Note any unusual or abnormal samples or circumstances. Make hard copies of this regularly and store them in digital form.
- Prepare data sheets, equipment, and arrange for helpers ahead of time.
- Never fabricate, falsify or misrepresent data to meet a research objective.
- Communicate the importance/significance of the data if someone else is assisting in collecting or analyzing data.
- Ask questions if unsure about data collection, quality of data, or protocols. You are empowered to bring up and assist in making changes that are needed.

Field work

- **Prior to going to the field**, ensure that all workers are on the same page about goals (data needed, overall scientific aim), protocols (scientific, field safety), and requirements (clothing, gear, equipment). The person in charge should check in with each field worker at least once prior to departing for the field to make sure everyone has all required equipment and gear, and is aware of the goals and protocols.
- Develop a field work agenda/schedule (dates, times, equipment list).
- Follow field safety protocol (see separate document), including appropriate attire, first aid kit and always do research in pairs (more fun and safer) **with notification of when and where you are going and when you will return**.
- Geo-reference all sampled plots or specimens with GPS and make available to all collaborators on the project.

- Keep track of who collects the data or assign one person for all data collection when possible.
- Confirm data measurements with work partner(s) during data collection; calibrate approaches so data in the field are collected uniformly.
- Confirm units of measurements with work partner(s) during data collection (e.g., in meters or yards).
- Be clear about randomization (vs. haphazard choice of plots); use random number generator.
- Bring extra equipment (attach flagging tape to equipment if possible).
- Include a metadata sheet that contains description of column names in the main data sheet.
- Establish unique site and sample names.

Greenhouse work

- Complete project request form with Jeff in advance of any greenhouse project to reserve space, request greenhouse assistance and notify greenhouse managers of use/needs.
- Use only numbers on pot tags to prevent bias when collecting data (write in pencil and sharpie on opposite sides of the tag).
- Randomize placement of pots and re-randomize bi-weekly.
- Confirm data measurements with work partner during data collection.
- Confirm units of measurements with work partner during data collection.
- Keep track of who collects the data or assign one person for all of data collection.
- Reserve headhouse space with greenhouse manager using online form when processing samples in headhouse.
- Include metadata sheet that contains description of column names in main data sheet.
- Use “sign in sheet” to monitor watering, fertilizing, etc.
- Post and refer to greenhouse map that specifies project locations in the greenhouse.
- Make arrangements for others to take care of plants when you cannot do it yourself (in advance – let Jeff know of changes too).

Lab work

- **Prior to beginning lab work**, ensure that you and/or the person(s) working for/with you have been properly trained for the task at hand.
- Reserve equipment or make arrangements for lab space use for all equipment (create sign-up sheet if one does not exist).

- For tasks that require data collection that may be subjective (e.g., measuring stomatal width, measuring mycorrhizal colonization), designate the task to one person and have the task only be completed by that person.
- Create and maintain a metadata sheet that contains description of column names in the main data sheet.
- Label all samples and containers with date, name, contents/project.
- Maintain a digital inventory of all samples that includes all sample information and detailed location of all samples.
- Thoroughly clean lab space after each task is completed for that day.
- Use a lab equipment sign-up sheet to monitor when/how equipment is borrowed and returned.

Data curation/analysis

- **Keep a master file that is not manipulated**, labeled as “master” (copy and paste from master file or duplicate master file for analyses).
- If collecting data manually (not directly entered into computer), make sure to store the hard copy in secure location (scan hard copy and save in digital form).
- For each dataset, include a metadata sheet that contains description of column names in main data sheet, with units and equations (e.g., transformations, calculations from raw data).
- Create inventory of project file names (e.g., Excel file that contains description of each file with corresponding file names).
- Double check data is entered/calculated correctly in data sheets.
- Create and communicate quality control checks.
- For each variable/column, if values are coded, provide description of codes to facilitate understanding of data by all collaborators.
- Data cleaning. Make sure data is cleaned before analysis by checking text/numbers irregularities (e.g. one name written differently for the same variable, “Yes” and “yes”, “spore” and “spores”)
- Use “tidy data”: every column is a variable, every row is an observation, and every cell is a single value.
- Create excel workbooks for each research objective or question to facilitate navigating all data sheets.
- Make use of different sheets within the Excel workbook for different types of data to facilitate importation in R or other statistical software.
- Back up all data, name the folder with the date (YYYYMMDD) of last edits on online platforms (google drive, dropbox, icloud, etc.)
- Make all code reproducible by annotating scripts in R, ideally in R markdown format.

- Provide a heading to your R script: objective, authors (including source if adapted from someone's code), creation date, date of last edits.
- Create an outline of all the different steps. Make use of sections to differentiate all the steps (e.g. 1. Necessary packages, 2. Data importation 3. Data exploration, 4. Building the SEM mode, etc.)
- Write full comments. Make sure anyone (including the future you in 2 weeks or 1 year) can understand what, how and why of your code.
- If you need to troubleshoot your code, duplicate the problematic code, fix it and create the final working script. You may need to keep track of the problems encountered and how you fixed it.
- If you are using the same line of code many times (>twice) in your script, make it a local function. If many lines are repeated, create a local R package.
- Format your code as you may format your manuscript before submission (more here http://adv-r.had.co.nz/Style.html and https://www.britishecologicalsociety.org/wp-content/uploads/2017/12/guide-to-reproducible-code.pdf)
- When code is clean and ready to be made public, upload to publicly accessible data repository (e.g., GitHub page).
- Create folder for each project that contains "old files" (do not delete data files).
- When manipulating (**non-master**) files, label file name with your initials and date.
- Note software version during analysis.
- Create a results Excel or csv file that contains results tables from a particular analysis (rather than only keeping results in a software script).
- Use common sense and judgment to assess if outlier data points are true or errors.

Box 1. DATA MANAGEMENT PLAN
(example to show how data and samples should be stored and shared; NSF format)

1. Data Types and Sources. All team members (faculty, staff and students) will take the CITI training in Responsible Conduct of Research. We will make all plant trait, plant genomic, microbial sequence, meta-transcriptomic, metabolomic and soil biogeochemical data - in addition to all code for managing data and statistical analyses and summary files explaining how to implement this code - publicly available in the environmental data initiative (EDI) to enable any user to independently verify and validate our findings.
Research Products: We will use standard approaches to quantify the following research products.

Functional plant traits – UTK-EEB will generate, manage, and archive the following datasets:
Above- and belowground productivity, bud break and leaf senescence phenology, leaf traits, nutrient uptake rates and a range of other functional plant traits. Metadata for these trees is archived at EDI and will be linked to each dataset. Raw data files (.CSV) will be stored as a backup at UT.
Plant Genomics – Data associated with genetic and genomic analyses will include the following:
Unique identifiers for all specimens and electronic output files from microsatellite DNA fragment electrophoresis as well as ddRAD and transcript-based SNP sequencing; unique identifiers will be carried forward on all containers of all tissues and tissue extracts (e.g., genomic DNA) derived from whole specimens; specimens, tissues, tissue extracts; Data from genetic and genomic analyses will also include raw data files with corresponding metadata files generated from instrument analyses of molecular-grade products will be stored on the instrument computer(s) and archived on external hard-drives and a secure cloud-computing platform. Derivative electronic data files with corresponding metadata files will be stored on personal computers, external hard-drives, and a secure cloud-computing platform. Derivative data products also will be deposited in relevant public archives (e.g., Genbank, Dryad online database) to be publicly accessible upon peer-review publication
Soil Microbiome Meta-transcriptomics – We will use standard approaches to quantify soil microbiome genetics, and meta-transcriptomics and will generate, manage, and archive the following datasets: Bacterial 16S and Fungal ITS2 ribosomal sequences datasets, metatranscriptomics datasets, will be collated with all metadata, quality and formatting requirements consistent with accepted MIMARKS (Yilmaz et al. 2011) and MIMS (Field et al. 2008) standards, respectively. Bacterial and fungal amplicon data generated from high-throughput sequencing will initially be received in standard .fastq format and immediately stored. A copy of the raw data will undergo processing to remove low-quality and contaminant reads, cluster OTUs, and assign taxonomy. All information generated during these processing steps (including a detailed .txt log file) will be stored both on a hard disk and uploaded to a public data repository, Dryad. Standard .fastq files with associated metadata .csv file. Fastq files will contain all necessary information of the sequence identifier, the raw nucleotide sequence, and read quality score. Each amplicon dataset will be archived within the National Center for Biotechnology Information - Sequence Read Archive (NCBI-SRA - https://www.ncbi.nlm.nih.gov/sra). Metatranscriptomics sequence sets will additionally be uploaded and publically available within the Joint Genome Institutes Integrated Microbial Genomes (IMG - https://img.jgi.doe.gov/) and MG-RAST (http://metagenomics.anl.gov) system to facilitate initial data mining and comparison efforts by our team.

Abiotic Gradients and Metadata - UTK-EEB will manage, and archive the following datasets: Geolocation and bioclimatic data for all genotypes at all sites. Metadata for these trees is archived at UT and will be linked to each dataset. Raw data files (.CSV) will be stored at UT.

Soil Biogeochemistry –UTK-EEB will generate, manage, and archive the following datasets:

Available soil nutrients (NH_4^+, NO_3^-, PO_3^-), pH and other soil chemical properties will be collected from field studies. From the carbon and nitrogen mineralization experiments we will collect evolved CO_2 samples as well as soil ammonium and nitrate data. Metadata for these samples is archived at UT and will be linked to each dataset. Raw data files (.CSV) will be stored as a backup at UT.

Teaching materials: Teaching materials generated during the teaching exchange will be distributed during or immediately after instruction under the open access Creative Commons 0 (CC0) license allowing for educational reuse. Materials will be available via dedicated training websites but also deposited on GitHub, with citable DOIs provided by Zenodo. As best practices change, we will adapt to ensure the materials are reusable and available without any registration or login barriers. Teaching materials will include slides, exercises, notes and videos.

2. Content Standards. We will follow community best practices: We will store data according to community best practice, which is to store files as .csv and .txt files rather than .pdf, .xls. Associated with each set of files stored will be a metadata file, including an explanation of the experiments that produced the data, an explanation of abbreviations and column names used in the dataset, specific units that are reported, and methods used to measure a particular variable. These raw data files will be assimilated into a relational database that links all datasets by primary keys (e.g., individual ID, data of measurement, source sites, etc.) and that allows one to query the database for subsets of data to be used in a particular analysis. The database will be query-able via both SQL and MS Access. As part of the metadata associated with each data type, it will be linked to a file containing code that was used to analyze the data for particular publications. Next, in the same folder as the data files and metadata files, we will store the code for statistical analysis. Lastly, we will include a higher scale summary file explaining how the datasets are linked.

3. Access, Sharing, and Protection. Data will be stored in GenBank or Data Dryad repository (http://datadryad.org/), a free and open-access repository that is the current standard in many areas of ecology, genetics, and evolutionary biology, or in the National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov/guide/all/). Importantly, each dataset is given a Persistent Identifier, specifically a Digital Object Identifier (DOI), which is permanent and unique. This also follows the current best practices in the scientific field. Phylogenetic character matrices and trees from the soil microbial community analyses will be deposited with the

Open Tree of Life. Unique, concatenated, complete sequences from known microbial species from our studies will also be deposited in GenBank. Next-gen sequencing will go to SRA and all meta-transcriptomes will go to MG-RAST or IMG.

Metadata used to create journal figures or tables (e.g. data used in generating regressions, bar charts, FST outlier tests) will be made available along with the code from the statistical software used to generate the metadata, as Supplementary Materials to the journal article. As for the raw data (described in section 1), our plan will exceed the recommended storage for appendices or supplementary material (on the order of 1 gigabyte) for most journals, and so will be stored in NCBI, DataDryad, and/or GenBank, as explained above. The DOI for DataDryad entries and the GenBank and SRA accession numbers will be included in the journal articles. Lastly, in addition, we will add a record to Trace, UT's Institutional Repository (http://www.trace.tennessee.edu/), so that the data will also be associated with UT.

4. Provisions for Re-use. Redistribution and reuse will require a default license and will be available for re-use within two years of collection (NSF guidelines). We will use Creative Commons licensing (http://creativecommons.org/). Publications or software generated from projects funded by this grant will contain a statement such as: “These data were produced based on work supported by the National Science Foundation” and will include the grant number. Use of data generated from this grant must clearly recognize the developer and NSF in any subsequent publications or presentations. Similarly, remixing and adapting materials is permissible, but requires recognition of original authors.

5. Plans for Archiving and Preserving Data/Samples/Products. Data will be archived, indefinitely, as discussed above, and data will also be stored on UT file servers, thus providing triplicate copies of all data files. All specimens, tissues and archive-quality, molecular-grade derivatives (e.g., DNA) used for plant genetic and genomic analyses will be stored indefinitely in temperature-appropriate secured environments (e.g., lockable -20C or -80C freezers). Perishable products (e.g., PCR amplicons) will be stored until fully analyzed, after which they will be disposed of in a safe and secure manner. Information on all sample processing (e.g., date, method, outcome) will be recorded in hand-written narratives in laboratory notebooks housed in a secure location.

Dried soil samples will be stored at UT for five years post-publication; space limitations prevent them from being stored in perpetuity. Soil microbial DNA will be fully labeled and permanently archived in -20º C freezers. Soil RNA extractions will be stored at -80º C. Master files with all information and freezer locations are maintained to catalog the samples.