SURF Grant Proposal:

Systems Biology, Data Analysis, and Soil Diversity

Principle Investigator: XXXXXX
Faculty Mentor: Dr. Jill Stewart University of North Carolina at Chapel Hill

Project Dates: May 4 – August 14

Funding Requested: $2000.00

Specific Aims:

The goal of this study is to discover the statistical interactions amongst the microbial populations of small farm unit soils. This study will broaden the holistic understanding of soil systems and provide me with the tools and experiences necessary to prepare myself for future research in this area. In order to address this goal, I intend to carry out three specific aims for this study:

1. I will sample the majority of a small farm system's soil diversity by collecting approximately 600 soil samples collected from 61 acre Efland, North Carolina farm, Fickle Creek Farm.
2. I will quantify the microbial presence of twelve different bacterial species, substrate induced respiration, pH, basal respiration, and nitrogen fixation within each soil sample.
3. I will rewrite systems genomic statistical modeling R code functions to infer causal relationships between the microbial contents and other factors quantified in Specific Aims 2.

Background and Significance:

As researchers develop and exploit new technologies, scientific exploration into some of the world's most complex questions becomes increasingly feasible. Advanced drug treatment technologies enable individuals to survive previously fatal diseases, have more children, and live longer. In 2013, the United Nations projected that the world population would reach 9.6 billion by 2050. Wes Jackson, the president of the Land Institute, addresses this idea and asks, "Can we keep ourselves fed? Can we save the stuff, the soil and water, of which we are made? All of us are just stopovers between soil

and soil" (2011). The answer to these questions is fairly simple; we must maximize soil health and sustainably yield more crops by using technology to understand the systems in which our food is grown.

Understanding the complexity behind soil systems is not nearly as straightforward. Just as we cannot tweak the expression of one magical gene and cure obesity, we cannot introduce a pesticide or herbicide in a soil system and dramatically improve crop yield; the systems must be thoroughly and holistically investigated. Similar to the gene interactions networks and environmental facts that regulate human physiological conditions such as obesity, complex microbial menageries within soil systems cultivate crop abundance. Existing systems biology and data collection technologies could help us understand the microbial interaction pathways within these complex systems that improve soil health and sustainably increase crop output. In other words, just as technology led to the food-deprived situation we may soon be facing, its full exploitation can provide a solution.

The microbes within soil systems are extremely complex, depending on an intricate combination of living and non-living factors such as plant roots, soil organic matter, Nitrogen based compounds, temperature, pH, geochemical molecules, water, and gases like carbon dioxide and oxygen. Since these conditions vary amongst soil samples, one would expect each sample to have a unique quantifiable microbial population (Bakker 2014, Pascault 2010, Koressaar 2009). Computational analysis techniques such as microbial fingerprinting and spectrometry permit the collection and analysis of data collected from large data samples throughout eleven consecutive months (Pascualt 2010). Koressaar et al.'s MultiMPrimer3 method identifies PCR primers within microbial

species that can be used to determine whether or not microbes of a specific phylogeny are present within a soil sample, and *in situ* hybridization technology can quantify these populations (2009).

The impact of soil quality on human health demands a greater understanding of the interactions within soil. Soil microbial quantification technologies, small farm units with wide varieties of soil diversities, and statistical models permitting causal inference should be utilized to meet this demand.

Methods:

The first step to this study requires the collection of hundreds of soil samples from a small farm unit. Fickle Creek Farm's 61-acre farm plot conducts a wide variety of practices with unique effects on the soil, including permaculture, agroforestry, and rotational grazing. These processes, respectively, cycle nutrients throughout the farm, permit interactions between agricultural and forestry systems, and introduce microbes from animal feces into the soil. Approximately ten evenly spaced soil samples will be collected from each acre of farmland, generating a dataset that captures most of the diversity present within the small farm system and has enough statistical power to generate statistically significant results. At the time of collection, I will place approximately 5 mL of soil from each location in an individual test tube, and record the soil temperature, air humidity, time, and other qualitative observations from the collection site. I will bring these samples back to a lab and perform appropriate lab techniques to generate a dataset with the figures collected at the field site and the quantified microbial contents, pHs, basal respirations, and nitrogen fixations of the soil samples. The quantities of twelve microbes, *N.gonorrhoeae* (*N.gonorrhoeae FA 1090*),

Helicobacter pylori (*H.pylori 26695* and *H.pylori J99*), *Mycoplasma genitalium* (*M.genitalium G37*), *Listeria monocytogenes* (*L.monocytogenes EGD* and *L.monocytogenes str. 4b F2365*) and *Candida albicans* (*C.albicans SC5314* and *C.albicans WO-1*), will be collected using standard *in situ* hybridization technology.

During the time spent waiting for machines to run, I will begin rewriting and modifying systems genomics statistical modeling code to utilize during the data analysis segment of this study. Some of these models include QTLnet and cape, two models developed to infer causal interactions between quantitative trait loci (QTLs), genes, and phenotypes. After finalizing the dataset, I will analyze the data using correlation analyses to find candidate interactions to focus on, and then regression analyses, modified QTLnet analyses, and modified cape analyses to infer causality between the variables within the dataset.

Timeline:

Weeks 1-2: May 4 – 15	Collect Soil Samples
Weeks 3-8: May 18 – June 26	Quantification / Rewriting Code for statistical models
Weeks 9-13: June 29 – July 31	Data analysis
Weeks 14-15: August 3 – 14	Develop and finalize deliverables

Budget:

Gas / Transportation	$150
Lab Equipment	$150
On-Campus Summer Housing	$1700

Preliminary Work Experience:

As a 2014 student research fellow at the Jackson Laboratory, I developed and carried out a ten-week systems genomics research project analyzing gene-to-gene and gene-to-phenotype interactions using R code and its various statistical modeling packages. Beginning with a colossal dataset that quantified the expression of 16,677

genes in seven different tissues collected from 519 mice, I concluded that *Smarca2* downregulates *Ucp1* expression in adipose tissue and am currently working on publishing a paper on my findings. To analyze the dataset generated for this study, I intend to follow similar data analysis methods I have used in past research. In addition to these skills, I plan to implement the statistical knowledge I am acquiring through my job as a data technician in training with NC State University's Center for Environmental Farming Systems. I currently have an A average in BIOL 101 and the lab course at UNC Chapel Hill, and took courses last year at the NC School of Science and Mathematics on biochemistry, organic chemistry, and molecular genomics, where I learned laboratory techniques such as PCR, protein purification techniques, and gel running.

Final Products and Dissemination Plan:

With the vast diversity captured within this dataset, I plan on to inferring significant causal relationships within the soil of Fickle Creek Farm. I intend to write and attempt to publish a paper highlighting my findings.

Word Count:

1181

Works Cited:

Bakker, M. G., et al. (2014). Diffuse Symbioses: Roles of Plant-Plant, Plant-Microbe and Microbe-Microbe Interactions in Structuring the Soil Microbiome. *Molecular ecology*, 23(6), 1571-1583.

Koressaar, T., et al. (2009). Automatic Identification of Species-Specific Repetitive DNA Sequences and their Utilization for Detecting Microbial Organisms.

Bioinformatics, 25(11), 1349-1355.

Pascault, N., et al. (2010). In Situ Dynamics of Microbial Communities during Decomposition of Wheat, Rape, and Alfalfa Residues. *Microbial Ecology,* 60(4), 816-828.

United Nations. (2013). World Population Projected to Reach 9.6 Billion by 2050 with Most Growth in Developing Regions, Especially Africa – Says UN. *UN Press Release*.

Wes Jackson. (2010). *Between soil and soil*. Madison: The Progressive, Inc.

List of Scholarships and Honors:

- Bruton College Fellows Scholarship
- NC Education Lottery Scholarship
- Honors Carolina Scholar
- Carolina Research Scholar
- Buckley Public Service Scholar
- Achieve Carolina Scholar