TOS SCIENCE DESIGN FOR TERRESTRIAL MICROBIAL DIVERSITY

PREPARED BY	ORGANIZATION	DATE
Lee Stanish	FSU	12/16/2015
J. Jacob Parnell	FSU	07/25/2013
Kali Blevins	FSU	07/25/2013

APPROVALS	ORGANIZATION	APPROVAL DATE
Mike Stewart	PSE	
Andrea Thorpe	SCI	

RELEASED BY	ORGANIZATION	RELEASE DATE
Judy Salazar	CM	

See configuration management system for approval history.

The National Ecological Observatory Network is a project solely funded by the National Science Foundation and managed under cooperative agreement by NEON, Inc. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Change Record

REVISION	DATE	ECO #	DESCRIPTION OF CHANGE
A	08/22/2014	ECO-02185	Initial release
B	12/16/2015		1. Revision to science design to incorporate the additional measurements for microbial biomass. 2. Modification to Figure 2 to account for addition of microbial biomass measurement. 3. Updated text referring to number of cores collected per sample from 3 to 1 or more. 4. Redefined a soil sample from an entire core to a single horizon. 5. Removed metatranscriptomic analysis

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

1 DESCRIPTION

1.1 Purpose

NEON design documents are required to define the scientific strategy leading to high-level protocols for NEON subsystem components, linking NEON Grand Challenges and science questions to specific measurements. Many NEON *in situ* measurements can be made in specific ways to enable continental-scale science rather than in ways that limit their use to more local or ecosystem-specific questions. NEON strives to make measurements in ways that enable continental-scale science to address the Grand Challenges. Design Documents flow from questions and goals defined in the NEON Science Strategy document, and inform the more detailed procedures described in Level 0 (L0; raw data) protocol and procedure documents, algorithm specifications, and Calibration/Validation (CalVal) and maintenance plans.

1.2 Scope

This document defines the rationale and requirements for terrestrial microbial diversity in the NEON Science Design.

1.3 Acknowledgments

The design of the terrestrial microbial diversity sampling for NEON described herein is the result of invaluable input from the Terrestrial Microbial Technical Working Group, including Rachel Gallery (University of Arizona), Kathryn Docherty (Western Michigan University), Greg Caporaso (Northern Arizona University), Gary King (Louisiana State University), James M. Tiedgje (Michigan State University), Eric Triplett (University of Florida), Lydia Zeglin (Kansas State University), Linda Kinkel (University of Minnesota), Diana Nemergut (Duke University), Chris Blackwood (Kent State University), Noah Fierer (University of Colorado), and Michael Allen (University of California, Riverside).

2 RELATED DOCUMENTS AND ACRONYMS

2.1 Applicable Documents

Applicable documents contain information that shall be applied in the current document. Examples are higher level requirements documents, standards, rules and regulations.

AD[01]	NEON.DOC.000001	NEON Observatory Design
AD[02]	NEON.DOC.001282	Introduction to the TOS Science Designs
AD[03]	NEON.DOC.000913	TOS Science Design for Spatial Sampling Design
AD[04]	NEON.DOC.002652	NEON Level 1, Level 2 and Level 3 Data Products
AD[05]	NEON.DOC.000906	TOS Science Design for Terrestrial Biogeochemistry
AD[06]	NEON.DOC.000914	TOS Science Design for Plant Biomass and Productivity
AD[07]	NEON.DOC.014048	TOS Protocol and Procedure: Soil Physical, Chemical, and Microbial Measurements

2.2 Reference Documents

Reference documents contain information complementing, explaining, detailing, or otherwise supporting the information included in the current document.

RD [01]	NEON.DOC.000008	NEON Acronym List
RD [02]	NEON.DOC.000243	NEON Glossary of Terms
RD [03]		
RD [04]		

2.3 External References

External references contain information pertinent to this document, but are not NEON configuration-controlled. Examples include manuals, brochures, technical notes, and external websites.

ER [01]	
ER [02]	
ER [03]	

2.4 Acronyms

Acronym	Definition
PLFA	Phospholipid Fatty Acid
DNA	Deoxyribonucleic Acid
mRNA	Messenger Ribonucleic Acid
rRNA	Ribosomal Ribonucleic Acid
PCR	Polymerase Chain Reaction
qPCR	Quantitative Polymerase Chain Reaction
ITS	Internal Transcribed Spacer
bp	Base Pair
CEC	Cation Exchange Capacity
OM	Organic Matter

3 INTRODUCTION

3.1 Overview of the Observatory

The National Ecological Observatory Network (NEON) is a continental-scale ecological observation platform for understanding and forecasting the impacts of climate change, land use change, and invasive species on ecosystems. NEON is designed to enable users, including scientists, planners and policy makers, educators, and the general public, to address the major areas in environmental sciences, known as the Grand Challenges (**Error! Reference source not found.**). NEON infrastructure and data products are strategically aimed at those aspects of the Grand Challenges for which a coordinated national program of standardized observations and experiments is particularly effective. The open access approach to the Observatory's data and information products will enable users to explore NEON data in order to map, understand, and predict the effects of humans on the earth and understand and effectively address critical ecological questions and issues. Detailed information on the NEON design can be found in AD[01], AD[02].

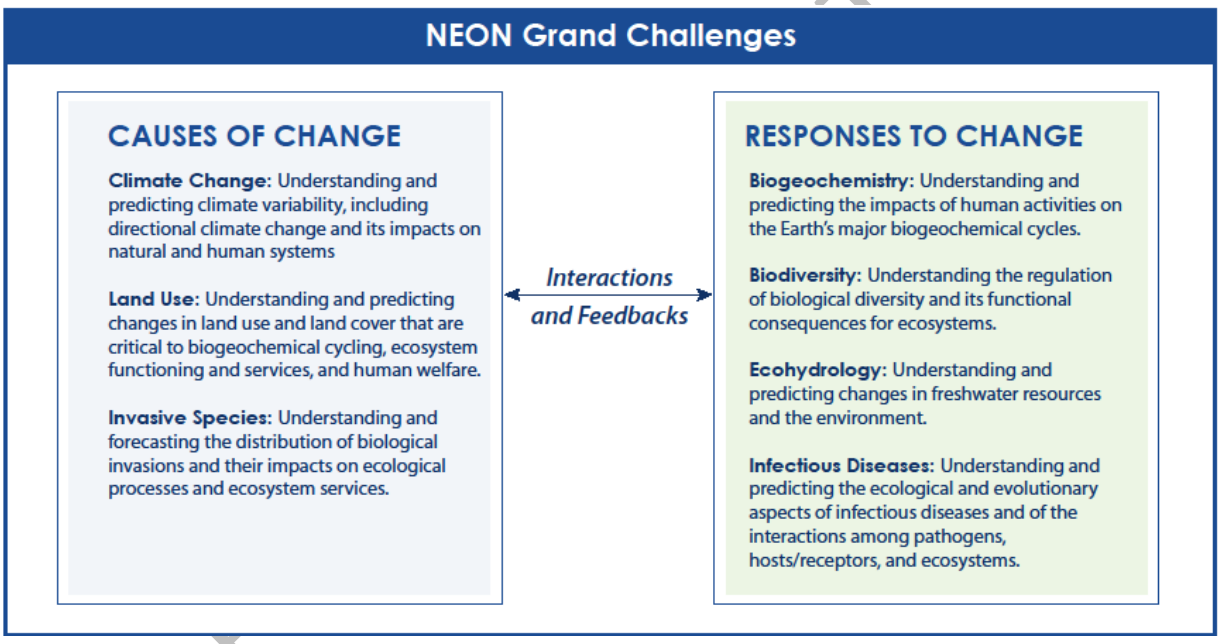

Figure 1. The Grand Challenges in Ecology as defined by the National Research Council (2001) and expanded by NEON.

3.2 Components of the Observatory

There are five components of the Observatory: the Airborne Observation Platform (AOP), Terrestrial Instrument System (TIS), Aquatic Observation System (AOS), Aquatic Instrument System (AIS), and Terrestrial Observation System (TOS). Collocation of measurements associated with each of these components will allow for linkage and comparison of data products. For example, remote sensing data provided by the Airborne Observation Platform (AOP) will link diversity and productivity data collected on individual plants and stands by the Terrestrial Observation System (TOS), and flux data captured by instruments on the tower (TIS) can be linked to satellite-based remote sensing. For additional information on these systems, see Schimel et al. 2007, and Keller et al. 2008.

3.3 The Terrestrial Observation System (TOS)

The NEON TOS will quantify the impacts of climate change, land use, and biological invasions on terrestrial populations and processes by sampling key groups of organisms (sentinel taxa), infectious disease, soil biogeochemistry, and nutrient fluxes across system interfaces (air, land, and water) (AD[01], AD[02]). The sentinel taxa were selected to include organisms with varying life spans and generation times, and wide geographic distributions to allow for standardized comparisons across the continent. Many of the biological measurements will enable inference at regional and continental scales using statistical or process-based modeling approaches. The TOS sampling design captures heterogeneity representative of each site to facilitate this inference when possible. Plot and organism-scale measurements will also be coordinated with the larger-scale airborne measurements, which provide a set of synergistic biological data products at the site scale. Details of these design elements and algorithms can be found in individual design documents available through the NEON website (www.NEONscience.org).

The standardization of protocols across all sites is key to the success of NEON (and its novelty) and must be maintained at all sites through time. Thus, although specific techniques may be required at some sites (e.g., due to different vegetation types), protocols have been developed to ensure data comparability. These details can also be found in individual design documents available through the NEON website (www.NEONscience.org).

The TOS Science Designs define the scientific strategies leading to high-level sampling designs for NEON sentinel taxa, terrestrial biogeochemistry, and infectious disease, linking NEON Grand Challenges and science questions to specific measurements (AD[02]). The TOS Spatial Sampling Design document describes the sampling design that collocates observations of the components of the TOS (AD[03]). TOS Science Design documents were developed following input from the scientific community, including discipline-specific Technical Working Groups, and the National Science Foundation (AD[02]). Science Designs will be reviewed periodically to ensure that the data collected by NEON are those best suited to meet the requirements of the observatory (AD[01]), are (to the extent possible) consistent with standards used by the scientific community, and fit within the scope of NEON. Additional information on the development and review process can be found in AD[02].

4 INTRODUCTION TO THE TERRESTRIAL MICROBIAL DIVERSITY SAMPLING DESIGN

4.1 Background

Microorganisms are critical drivers of biogeochemical processes that influence global climate, water quality, and atmospheric composition (Vitousek et al., 1997; Canadell et al., 2007; Galloway et al., 2008; Gilbert 2009; Conley et al., 2009). Shifts in microbial assemblages in response to environmental change will potentially affect the biogeochemical cycles they mediate (Allison & Martiny 2008). Since most biogeochemical cycles are interconnected in a complex network of feedback relationships (Bardgett et al., 2008; Falkowski et al., 2008; Finzi et al., 2011), changes in microbial communities can have profound impacts on ecosystem services.

Although microorganisms play a key role in response to global change, most ecosystem models consign microbial inputs to 'black box' status (Andren & Balandreau, 1999) where inputs and outputs are based on rate equations with little consideration for spatiotemporal community dynamics and actual function (Docherty & Gutknecht 2011; Todd-Brown et al., 2011; Treseder et al., 2011). Current global biogeochemical models are generally based on microbial processes that have been measured from microorganisms in equilibrium (Schimel 2001), but models that consider seasonal transitions, plant invasions, climate change, and land-use management by definition include dynamic microbial assemblages.

Microbes can respond rapidly and at microscopic scales to environmental changes. As such, data are lacking that capture microbial structure and function at the spatial and temporal scales necessary to model and predict microbial dynamics at larger scales and over long periods of time. In the past decade, transcendent studies began to unlock the 'black box' of microbial ecology. Localized studies of microbial community dynamics have shed light on the physiological mechanics associated with ecosystem services (Nemergut et al., 2005; Iv et al., 2011; Yergeau et al., 2009). Meanwhile regional- to continental-scale studies explored habitat drivers that could have impacts at multiple scales (Dinsdale et al., 2008; Lozupone & Knight 2007), but for which spatial and temporal resolution are not accounted. Similarly, short-term explorations provide clues to ecosystem dynamics (Bardgett et al., 2005, Wittebolle et al., 2009), but are unable to monitor the impact of long-term oscillations. On the other hand, long term monitoring (Ramirez et al., 2010; Aber & Magill 2004) is generally at a coarse temporal resolution that fails to capture episodic, rare events that yield important information about ecosystem stability in response to short-term changes.

In order to refine the role of microbial community dynamics in ecosystem models, biodiversity metrics that link structure (richness, evenness and diversity) to function (both potential and active) need to be developed in the context of appropriate spatial—ranging from local to continental—and temporal—seasonal to decadal—scales (Raes & Bork 2008).

4.2 NEON's Contribution

The National Ecological Observatory Network (NEON) provides a platform to consider the role of microorganisms in ecosystem processes that combines standardized long-term seasonal monitoring at dozens of sites dispersed in different ecological regions across the continental United States, including Hawaii, Alaska, and Puerto Rico. In total, standardized and coordinated measurements will occur at 20 core sites that will remain for the duration of the project (30 years), and up to 27 relocatable sites that are intended to measure local ecological processes and phenomena over shorter (5-10 year) time scales. The goals of NEON are to:

- monitor ecological changes with respect to invasive species, changes in land use and climate change;
- provide an open and accessible data and specimen repository for free; and
- facilitate research using NEON resources

As part of NEON's open access policy, physical samples collected over the life of the observatory will be archived and made available for additional analyses. Extensive long-term coordinated ecosystem monitoring will allow investigators to understand and forecast patterns of ecological change at local, regional, and continental scales (Kao et al., 2012).

The NEON microbial ecology program is designed to provide data highlighting changes in microbial properties through space and time. This strategy rests on two principal design elements: 1) how microbial communities change, and 2) how those changes can be observed. Existing work demonstrates that environmental change can modify microbial assemblages in various ways (Figure 2). Microbes can respond to change through ecological processes of succession, adaptation, and shifts in abundance corresponding to changes in (or maintenance of) microbial function (e.g. Allison & Martiny 2008). These changes are observed through measurements of microbial biodiversity that reflect the different responses (Figure 2; Raes & Bork 2008).

As environmental conditions change over the next few decades due to altered climate, land-use practices, invasive species, or other local events such as wild fire or extreme weather, the multiple dimensions of microbial biodiversity analyzed in conjunction with other biotic and abiotic measurements will provide improved understanding of microbial assemblage function in ecosystems biology. The dominant factors that affect microbial biodiversity can be summarized by soil biogeochemistry and vegetation type (Van Der Heijden et al., 2008; Harris, 2009). Consequently, microbial assemblage structure and function data will be collected in coordination with soil biogeochemical measurements and in the context of vegetation and plant productivity. Environmental drivers of microbial assemblage structure (Fuhrman et al., 2006) such as aboveground biomass, vegetation type, and soil moisture, can be determined from a larger scale context using aerial imaging (Lefsky et al., 2002). Thus, suites of measurements describing soil properties collected throughout NEON will help to model microbial assemblages on a continental scale. By collecting measures of microbial biodiversity (assemblage structure and function) in conjunction with soil biogeochemistry, plant productivity, and a suite of

environmental measurements, NEON data will provide keys to understanding the dynamics of microbial assemblages in a changing environment and their influence on ecosystem processes.

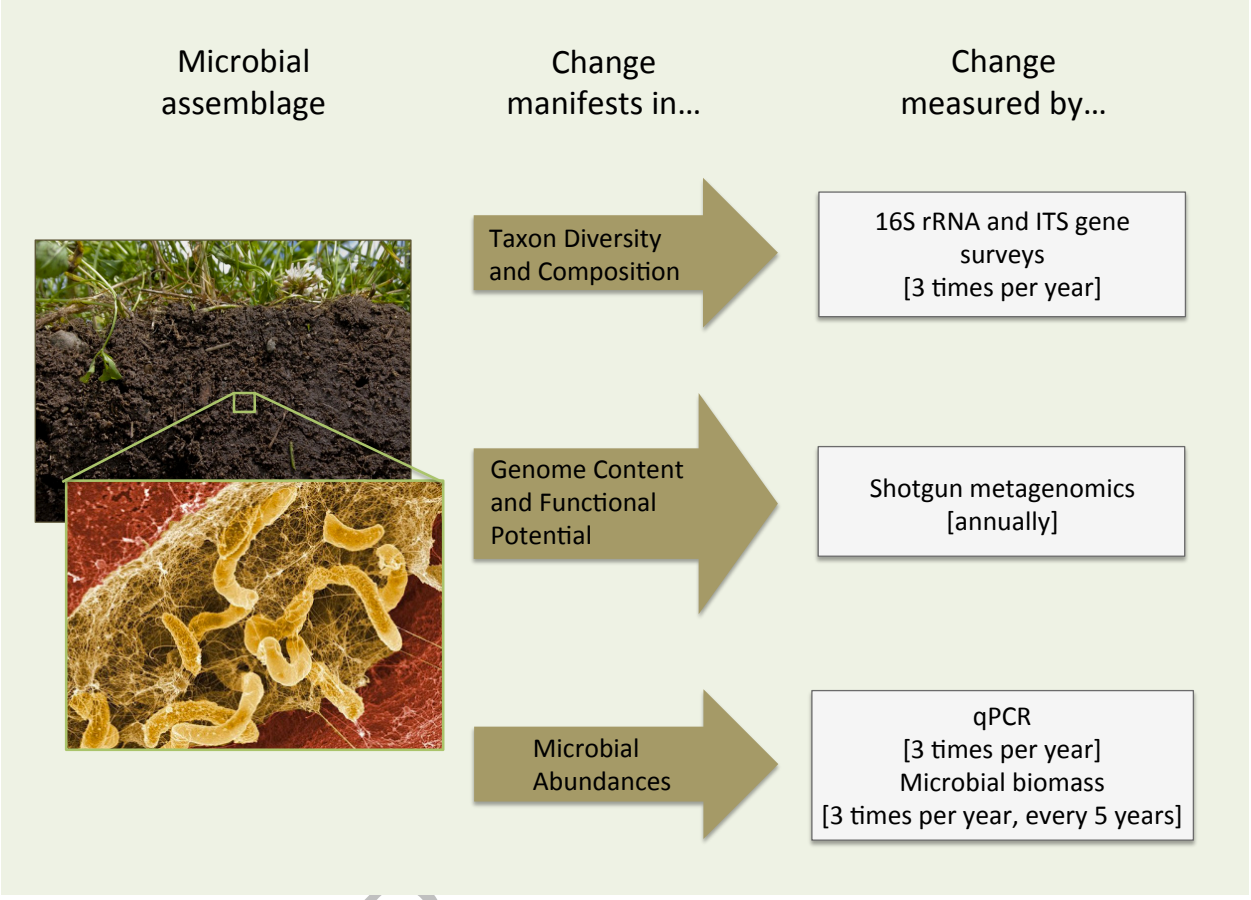

Figure 2. NEON's proposal for measuring potential effects of environmental change on microbial assemblages.

4.3 Purpose and Scope

This document discusses the sampling scheme and underlying design strategies and rationale associated with NEON's soil microbial ecology plan. This document includes a brief discussion of the framework of measurements to be collected followed by a description of sampling strategies and then provides detailed information on methodology.

5 SAMPLING FRAMEWORK

NEON's priority for measurement methods aligns with strategies of other affiliate organizations (e.g., Earth Microbiome Project, Terragenome Project, Global Ocean Survey). Specifically, nucleic acid extraction procedures and minimal information of metadata associated with soil samples collected for microbial assemblage analyses must be standardized (Yilmaz et al., 2011). Specific lists of information associated with genomes, metagenomes, and marker gene sequencing can be found at http://gensc.org/gc_wiki/index.php/MIxS_Compliance. Furthermore, co-location of microbial sampling

with other ecological and abiotic measurements, specifically plant productivity AD[06] and soil biogeochemistry AD[05], will help to shed light on drivers of ecosystem processes and provide points of reference for potential scale-up exercises.

5.1 Science Requirements

This science design is based on Observatory science requirements that reside in NEON's Dynamic Object-Oriented Requirements System (DOORS). Copies of approved science requirements have been exported from DOORS and are available in NEON's document repository, or upon request.

5.2 Data Products

Execution of the protocols that stem from this science design produces samples and/or generates raw data satisfying NEON Observatory scientific requirements. These data and samples are used to create NEON data products and are documented in the NEON Level 1, Level 2, and Level 3 Data Products Catalog (AD[04]).

5.3 Priorities and Challenges for Terrestrial Microbial Diversity

Understanding the geographic turnover or variation in microbial assemblages is crucial to understanding their ecology and evolution. In many cases, the variation in microbial assemblages is principally driven by environmental factors (Bell et al 2005; Fuhrman et al 2006), and the variation occurs at the sub-centimeter (Woyke et al 2006), meter (Baker et al 2009), and kilometer scales (Whitaker et al 2003; Green & Bohannan 2006; Tringe et al 2005; Rusch et al 2007; DeLong et al 2006). In general, environmental conditions and dispersal limitation cause ecological assemblages to become increasingly different with increased distance (Nekola and White 1999). Although community changes across all distances are important, due to the large spatial scale and long-term nature of NEON, measurements will take place at meter-scale resolution.

It is critical to recognize that sequencing technology advances rapidly, and current approaches for measuring microbial diversity may become obsolete. NEON will regularly evaluate its methodologies against new methodologies with the goals of maintaining consistency in data quality and comparability and maximizing scientific utility.

6 SAMPLING DESIGN FOR TERRESTRIAL MICROBIAL DIVERSITY

6.1 Sampling Design for Terrestrial Microbial Diversity

NEON soil microbial community analytical measurements will be standardized with the aquatic microbial measurements collected in freshwater systems throughout the observatory. NEON will have aquatic sites that include streams and shallow lakes. The sampling design for aquatic microbial ecology is not presented here, but the methods for assessing microbial assemblages (DNA/mRNA extraction, and sequencing) shall be identical.

The terrestrial microbial sampling strategy consists of a two-phase approach to determine spatiotemporal variation in microbial assemblages (Figure 3). The first phase will optimize the timing and frequency of sampling, while Phase 2 implements the long-term sampling schedule optimized during Phase 1. The spatial layout is founded on NEONs spatially-balanced random design AD[03], which stratifies plots by vegetation type.

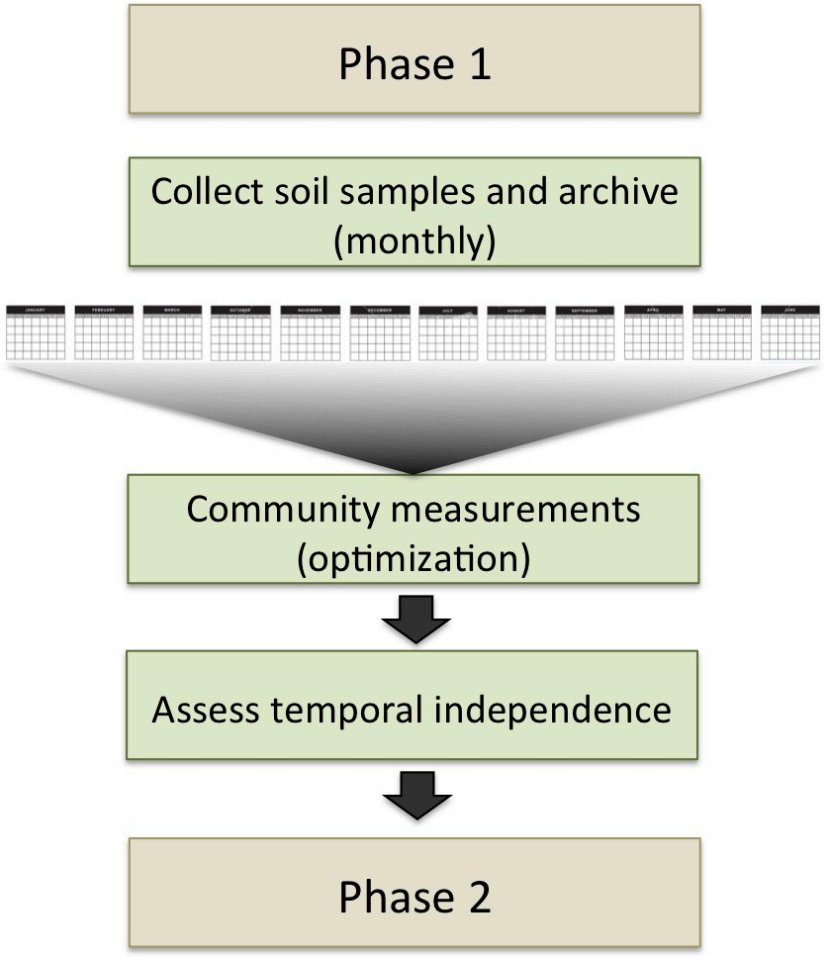

Figure 3. Temporal design of microbial sampling strategy. The temporal component consists of 2 phases designed to determine optimal sampling frequency.

During Phase 1, samples were collected every month that the soil remains unfrozen for up to two years (Figure 3). Up to 10 plots were sampled at each site with plots located both in the tower airshed and distributed throughout the site (see AD[03]). The number of plots and the exact measurements collected varied depending on field season. Phase 1 measurements included 16S rRNA gene and ITS sequencing, q-PCR, metagenomics and community mRNA profiling. These measurements provided information on the seasonal variation in microbial community composition and activity. The results from Phase 1 will be used to determine the time points throughout the year that maximize seasonal variation.

In combination with results from two pilot studies (discussed below), Phase 2 will implement the spatial and temporal scale sampling for the duration of the Observatory (Figure 2).

Soil sampling is carried at 10-15, 40x40 m plots (Figure 4). These plots are located both within the instrumented tower airshed AD[02] and are distributed throughout each site. Each plot is divided into four subplots to provide spatial balance across the plot, and in 3 of the subplots, a random location is selected for sampling. Soil sampling will follow the same protocol used to collect soil for biogeochemical analyses, with the exception that microbial samples will be immediately frozen on dry ice following collection (AD[07]). After the soil sample is homogenized, a subset of the sample will be used for biogeochemical analysis, another subset will be used for microbial molecular measurements, and every 5 years a subset will be used for microbial biomass measurement. At the domain facility, soils for molecular analysis will be kept on dry ice or in a -80C freezer until shipped, and soils for microbial biomass analysis will be stored refrigerated.

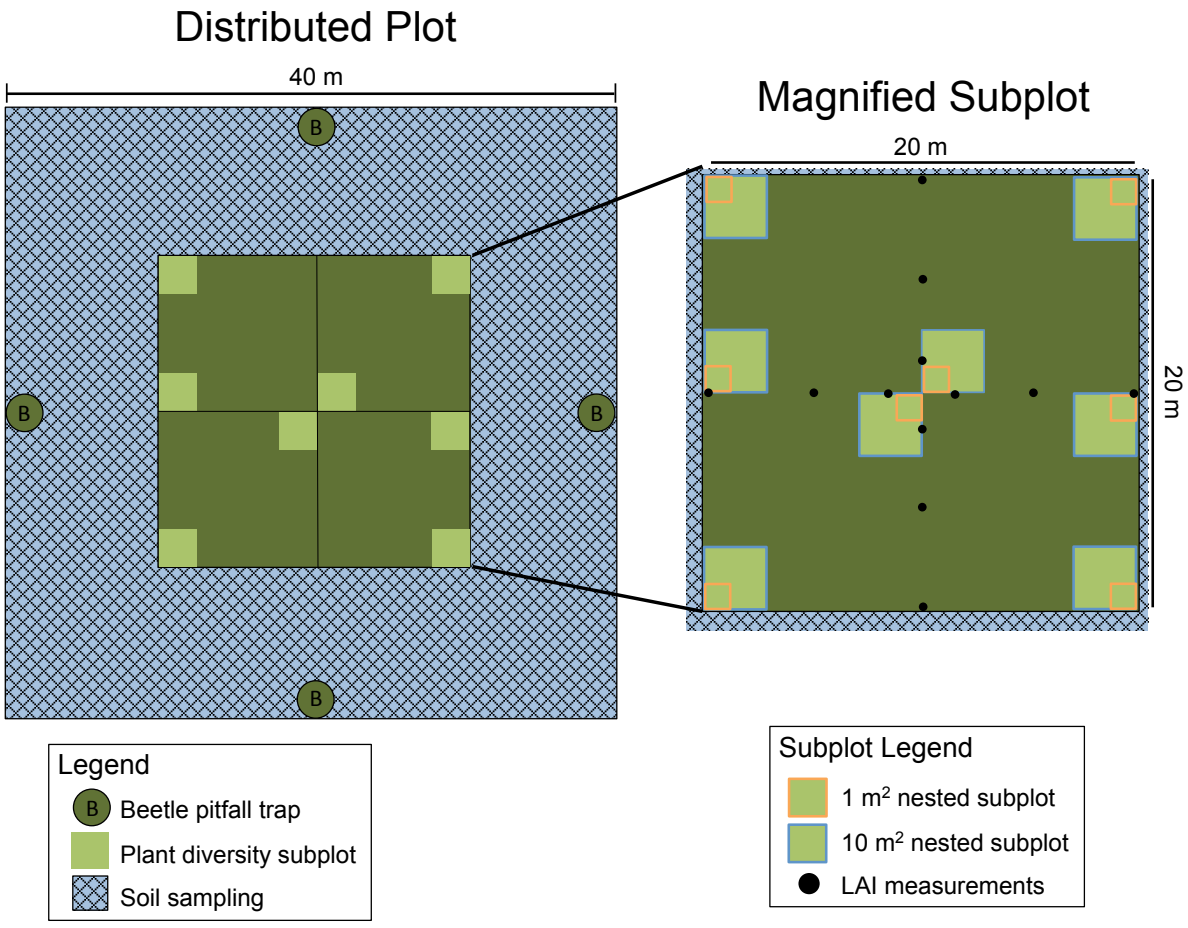

Figure 4. General layout for distributed plots. Sampling for beetle diversity, plant diversity, and leaf area index (LAI) measurements may also occur at soil plots.

The field protocol(s) used by NEON for collecting soil cores to analyze physical properties, biogeochemical constituents, and microbial assemblages follows the protocols presented in the Soil

Science Society of America *Methods of Soil Analysis* texts (Sparks et al., 1996; Dane and Topp, 2002). Soils are inherently spatially heterogeneous. Thus, three samples are collected per plot at up to 12 plots across a site in order to capture variability at multiple scales (e.g., sub-plot, plot, transect, airshed, site). Plot locations in which soil samples will be collected vary for each NEON site and are based on vegetation type and the size of the tower airshed. Soil core coordinates are randomly generated within each plot and will be made available for the field crew to upload to the handheld PDA and/or print to datasheets prior to collection. Field crews will use high-resolution GPS units to locate plots and soil core coordinates will be provided as randomized X, Y-coordinates within the plot. Three sets of coordinates will be provided for each plot for sample collection at each time point. A soil sample consists of a single soil horizon, broadly defined as either organic or mineral. Up to 3 cores are collected within 0.5 m of the given coordinates in order to obtain sufficient material for the entire sample suite. If sample coordinates are obstructed, field technicians can move the sample location to the next random location provided. The sample is then homogenized for biogeochemical and microbial subsampling.

NEON microbial soil sampling will sample to a maximum depth of 30 cm. During most sampling bouts, only the top horizon (organic or mineral) shall be processed for microbial analyses. During a joint biogeochemical sampling bout, however, both the organic and mineral horizons will be sampled concurrently, if both are present. It is critical that the locations from which soil samples are collected have not been disturbed prior to sampling. Due to the temporal component of the soil microbe analyses, soil core collection from one site should be done as close in time as possible (preferably all cores will be collected within 1-2 days of each other).

Upon collection, samples will be shipped on dry ice to NEON for analytical preparation or directly to contract facilities for nucleic acid extraction and sequencing, and for archival. NEON's calibration and validation department will also randomly send a standard sample to determine the sequencing contract laboratory's reproducibility.

NEON has tested the microbial science design by conducting two pilot studies. The first prototype effort took place over the 2009-2010 field sampling season (Docherty et al., 2015). In this study, 408 samples were collected and analyzed from 4 domains that represent a broad latitudinal gradient with soils with unique characteristics, climates, temperature, and precipitation. Specifically, samples were collected from Utah (Domain 15; July 8, July 15, July 17, October 3, and March 2), Florida (Domain 3; June 29, July 10, July 16, October 15, February 15, and March 1), Hawai'i (Domain 20; July 2, August 13, August 24, October 20, and February 11), and Alaska (Domain 19; June 28, August 8, and August 29) (Table 1).

Table 1. Location and general metadata associated with samples collected for the spatiotemporal prototype

Location	Veg. type	Lat	Long	Elev. (m)	MAP (mm)	MAT (°C)	Soil pH
Alaska	Boreal Forest/Taiga	65.15	-147.5	290	260	-3	5.1
Utah	Grassland/Shrubland	40.18	-112.4	1676	274	8.9	7.8
Florida	Tropical Dry forest	29.69	-81.9	46	750	20	5.2
Hawai'i	Tropical moist forest	19.93	-155.2	1167	2500	15.5	6.1

At each site, a sample grid measuring 150 m x 300 m overall containing eight 75 m x 75 m cells was established within the expected airshed of the instrument tower. Within each of the resulting eight cells,

three sampling sites were selected by randomly assigned GPS coordinates to establish the first sample; additional samples were collected within plots for the duration of the study. A set of 3 cores encompassing the 0-10 cm depth interval beneath the litter layer was collected within each sampling cell at each time point. In addition, sub-sets of the three cores corresponding to each cell were combined for a composite sample representative of each cell. Five 10-gram sub-samples were obtained from each sample and labeled and frozen at -80 ˚C.

DNA was extracted from all soil samples as described previously by Fierer and Jackson (2006) using the PowerSoil DNA isolation kit from MoBio Laboratories (Carlsbad, CA) following Earth Microbiome Project standard protocols. Universal primers for the 16S rRNA gene, described by Hamady et al. (2008), were used to PCR amplify this gene from all soils in order to examine the diversity of bacteria and Archaea. The 18S rRNA gene was amplified using a set of universal primers designed by Knight and Fierer. All DNA sequencing was done by 454 pyrosequencing using the titanium protocol and analyzed using QIIME software.

In addition to rRNA gene sequencing, initial pilot strategies used a targeted nitrogen fixation functional gene approach (*nifH* genes) to determine functional diversity. Reads from 201 samples were filtered through initial barcode matching and quality control steps using the RDP pyro initial process tool. A total of 1.15 million reads matched the barcodes, and 89.33% passed the initial process filtering. Following FrameBot frameshift correction (99.97% of the sequences passed and the average frameshift rate was 0.16 per sequence), sequences were aligned with HMMER3 and clustered using the complete-linkage clustering algorithm in the RDP myClust tool.

Microbial assemblages were compared with other biotic and abiotic information collected from each sample including the exact location of each sample based on GPS coordinates, a description of the environment where the sample was originally obtained, soil type, horizon, where each sample is in relation to the NEON tower, pH, water content, cation exchange capacity (including calcium magnesium, sodium, and potassium), percent organic matter, total carbon, total nitrogen, total biomass determined by phospholipid fatty acid analysis, chloride, nitrite, bromide, nitrate, phosphate, and sulfate.

The second pilot study involved collection of samples from the Harvard Forest site in 2012. A total of 32 samples were collected from ten 20 x 20 m plots distributed throughout different vegetation types. At each plot, 3-5 samples were collected; each sample was divided by soil horizon for separate analysis. In two of the plots, 5 samples were collected and both the organic horizon and mineral horizon microbial communities were analyzed.

Microbial sequence data were processed using the QIIME pipeline (Caporaso et al 2011) following the protocols and standards used by the Earth Microbiome Project (http://www.earthmicrobiome.org/). A total of 23,376 unique bacterial/archaeal sequences at 97% sequence similarity were identified from all samples, providing ample measures of biodiversity. The average number of sequences per sample was 1826 with a minimum of 22 sequences and a maximum of 3318.

6.1.1 Sampling Methods

The analysis of all major taxonomic groups shall be accomplished by targeting the 16S rRNA sequence for bacteria and archaea. PCR primers that have been developed to target the V3-V4 region and shown to yield optimal community clustering will be used (Caporaso et al 2011). For fungal communities, the nuclear ribosomal transcribed spacer region (ITS) will be targeted for amplification and sequenced using paired-end read sequencing for maximum resolution. The region spanning ITS1, 5.8S and ITS2 can be amplified by universal primers in the conserved flanking regions of SSU and LSU and has been proposed as a barcode for fungi (Begerow et al 2010). Despite differences in read length and sequencing protocols, there is a 90% overlap in taxonomic composition between 454 sequencing and Illumina-based sequencing (Luo et al., 2012). Due to the increased accuracy and number of reads, as well as decreased cost of sequencing, the Illumina MiSeq platform shall be used for 16S and ITS sequencing, generating 15-20,000 paired-end reads of 300 bp per sample.

NEON will also shotgun sequence the **metagenome of microbial communities** using the Illumina HiSeq2000 platform to generate 5-10 million 300 bp paired-end reads per sample.

Microbial abundance shall be determined using quantitative PCR targeting the bacterial 16S and fungal ITS rRNA genes. **Estimates of microbial biomass** shall also be made using the chloroform fumigation method, or a similarly informative method such as PLFA analysis, which has been shown to reveal functionally relevant relationships between microbial activities and environmental conditions (Docherty et al., 2015).

6.1.1 Spatial Distribution of Sampling

Initial prototype 16S community sequence data suggest that at the 150 m x 300 m grid scale, total carbon is a dominant driver of microbial diversity and correlates significantly with the principal component that captures the most variation at each site (Figure 4). However, between sites pH correlates with the principal component and individual sites are closely grouped in principal component space. These data suggest that samples collected at the 150 m x 300 m grid scale may be influenced by spatial autocorrelation and that the difference in within site variation and between site variation may be too great for effective scaling. Consequently, the collection of soil samples using plots distributed across vegetation types throughout the site is justified.

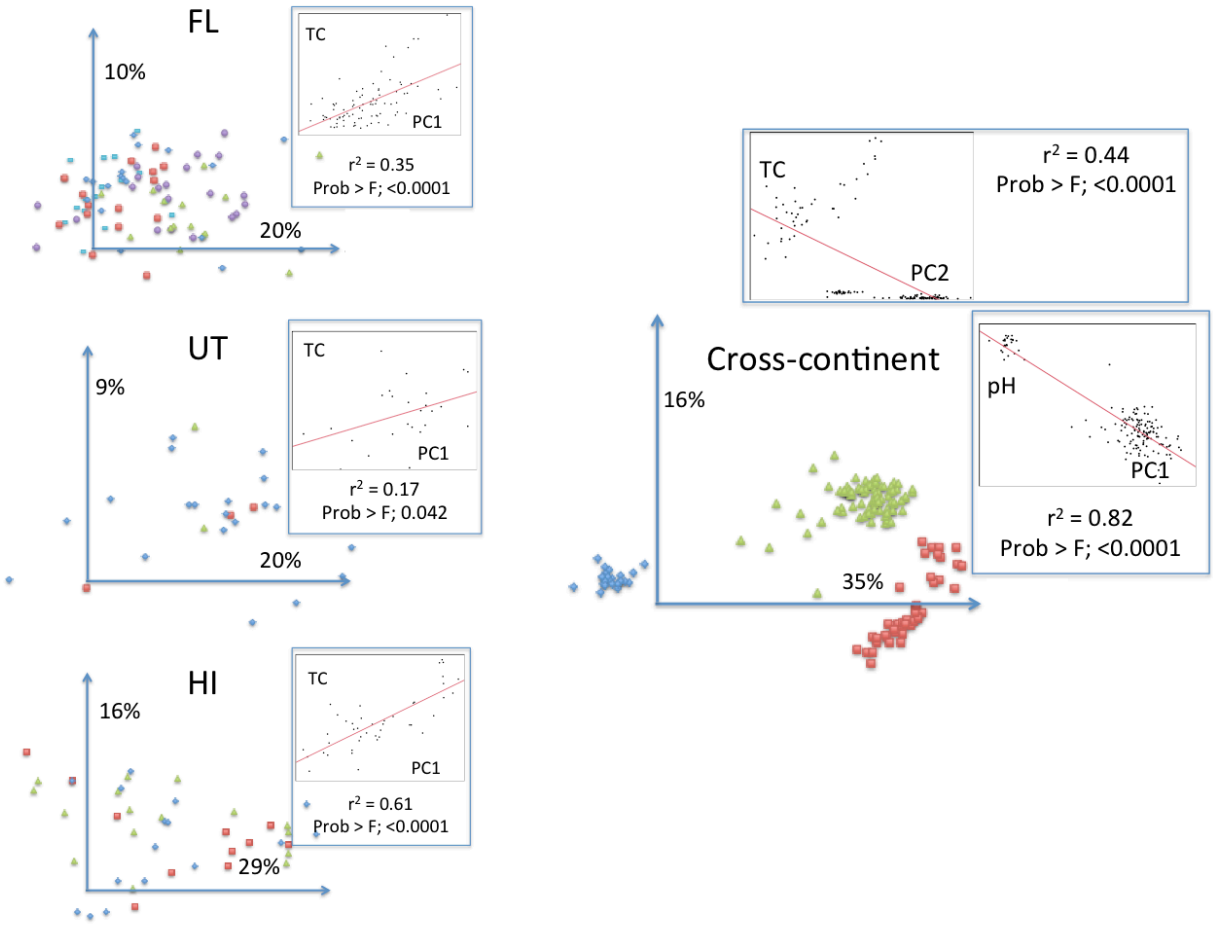

Figure 5. Principal component analysis of microbial communities. Individual sites (left) colors represent different sample collection dates. All sites (right) colors represent individual sites (blue = Utah; Green = Florida; Red = Hawai'i). Principal component analysis (x-axis is PC1, y-axis is PC2) on Operational Taxonomic Unit tables based on 97% cutoff. Insets display correlation between component scores (x-axis) and best fit environmental variable (y-axis; total carbon as percent sample as C).

Previous studies have shown that sample size can influence microbial community composition (Ellingsøe and Johnson 2002, Ranjard et al. 2003), with larger sample sizes producing more robust results with less variability, and smaller sample sizes allowing for detection of less abundant taxa. We examined the effect of sample size on 16S taxon diversity by comparing OTU recovery from samples representing individual soil cores and from a composite sample of 3 combined cores, all from the same plot. Our data show that the abundances of higher abundance OTUs (>10 reads) correlate strongly between individual and composite samples, while the correlation breaks down in less abundant OTUs (Figure 5). Based on these results and the suggestion by Ranjard et al. (2002), the soil samples used for taxonomic measurements will be based on individual cores (or composite cores within 1 meter of each other, when required to obtain sufficient material) in order to capture a more complete diversity inventory, while metagenomic analyses will be conducted on samples composited at the plot level for robust detection of gene composition.

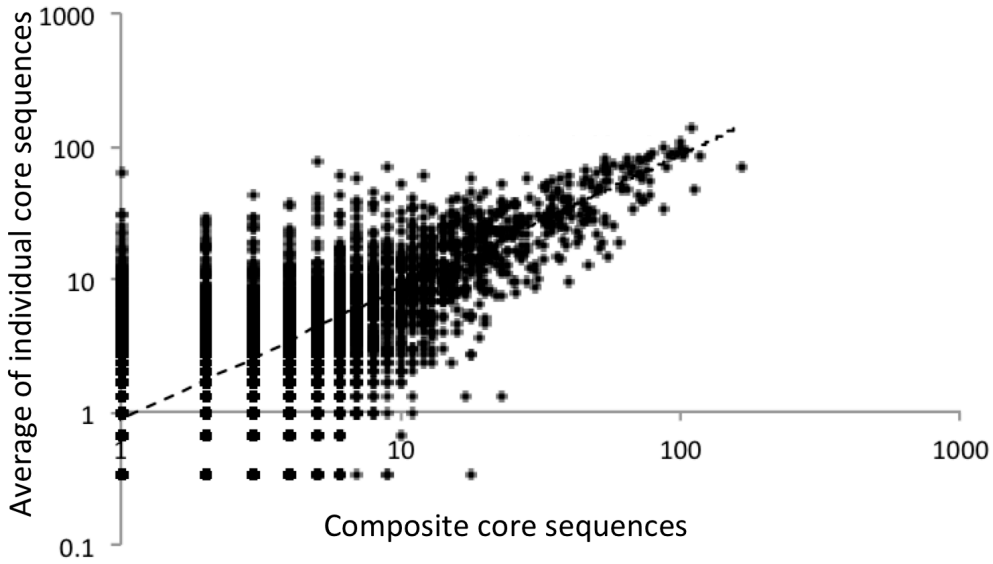

Figure 6. Correlation between composite core 16S sequences of 97% cutoff taxonomic groups and individual core sequences. Note, axes are log-scaled.

In the second prototype, microbial communities from different vegetation types were determined. Harvard Forest has a clear distinction between mineral and organic horizons and these differences are an important factor in defining microbial diversity (Figure 6).

In addition to 16S sequencing, Harvard Forest sample soil mRNA was extracted, enriched, and sequenced using an Illumina HiSeq 2000 platform. These sequences are limited to organic horizon samples from a disturbed plot primarily with herbaceous vegetation, a successional shrubland plot, and four plots with mature forest of different plant diversity. The number of functional gene reads sequenced varied from 40,000 to 484,000 across samples. Principal component analysis suggests that over 80% of the variation is accounted for in the top leading variable (>90% in two components; Figure 7). Cluster analysis suggests that the vegetation type plays a critical role in structuring microbial communities based on the expressed functional genes.

Data from the pilot study suggest that collection of samples within a confined area such as the 300 m x 150 m grid design is limited in spatial independence with respect to microbial community diversity (Table 3). This information in conjunction with Harvard Forest prototype (Figure 7) and recent metagenomic data (Antonopoulos et al. submitted) suggests adopting a spatially balanced design that is stratified by vegetation type (Figure 4; AD[04]).

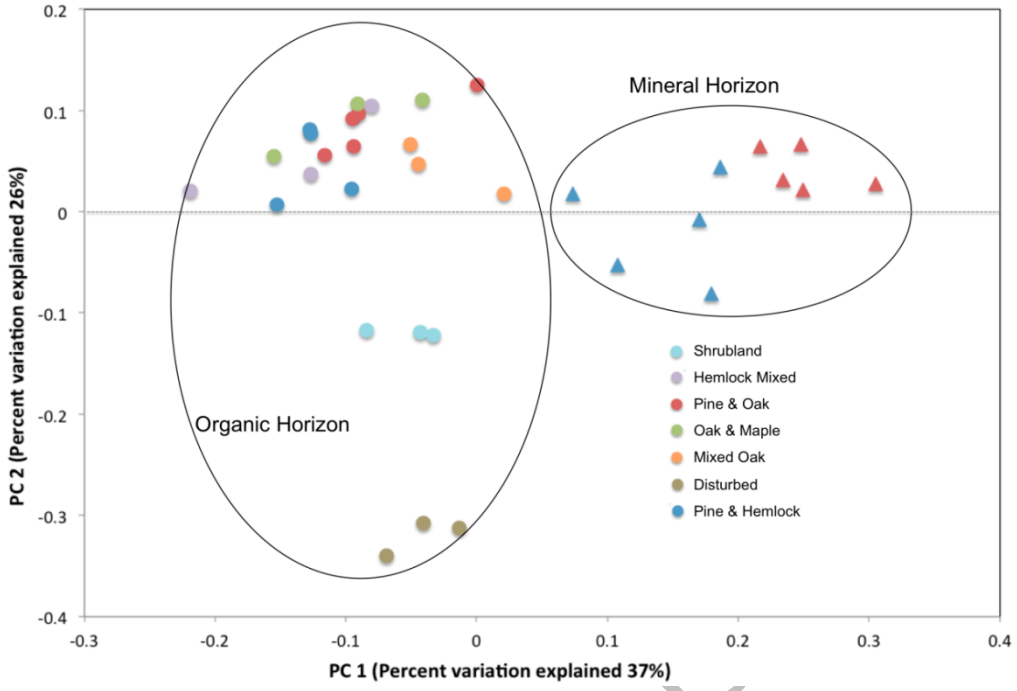

Figure7. Principal component analysis of 16S rRNA gene sequence of Harvard Forest microbial communities.

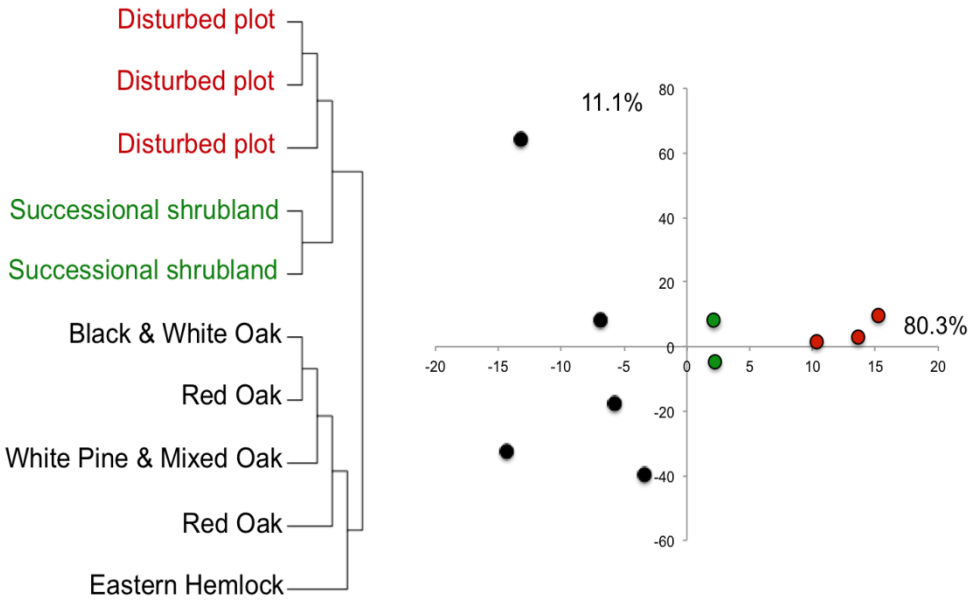

Figure 8. Functional diversity of Harvard Forest soil microbial communities using soil metatranscriptomics. Cluster analysis of microbial community gene expression (left) shows grouping by high-level vegetation type. Principal component analysis (right) suggests that over 80% of the variation is described by the leading variable.

The amount of soil extracted for template DNA in effect represents sampling effort. As such, the diversity of genes eventually sequenced from the template DNA could be influenced by the mass of soil extracted. To test this hypothesis, DNA was extracted from 4 grams of soil (in 0.33 g portions) from each of three Florida and three Utah soil samples, and combined to give the equivalents of template DNA from 0.33, 0.67, 1.0, 2.0 and 4.0 g of soil from each sample.

Similar to soil mass, the amount of template DNA amplified is another aspect of sample size. As such, the diversity of genes eventually sequenced from the template DNA could be influenced by the amount of template amplified. To test this hypothesis, DNA was extracted from 4 grams of soil from each of three Florida and three Utah soil samples and diluted to perform PCR with 1, 5, 10, 25 and 50 ng of template per 20 µl reaction.

An examination of results for individual samples showed an increase in observed OTUs with increasing soil mass for three of the six samples. Species accumulation curves showed that the Florida sample from 0.33 g of soil was less species rich than samples from greater amounts. There were no differences among the Florida samples from other soil masses or among any of the Utah samples from varying soil masses. Two of the Florida samples had the lowest Chao1 values (0.33 and 0.66 g of soil), suggesting very low biodiversity in sandy soils; there were no correlations between soil mass extracted and Chao1 for the Utah samples (see details in Wang et al., 2013).

For the Florida samples, no amplification occurred with 1 or 5 ng of template per reaction, suggesting a minimum sampling effort in these soils. Otherwise, no consistent increase in species richness with increasing template concentration was detected.

These pilot study data suggest that sampling will:

- Be collected from plots distributed in different vegetation types throughout each site. The number of plots will range from 4 to 10 depending on vegetation diversity, site size, and available resources.
- Be composite of up to 3 cores collected within 1 m of each other. Composite samples may be necessary to obtain sufficient material for both microbial and biogeochemical assays.
- Encompass the surface horizon. In cases of a distinct organic horizon, this will be sampled separate from the mineral horizon. Where no organic horizon exists, sampling will include surface up to 30 cm depth.
- Provide enough material that nucleic acids be extracted from sufficient soil so as to eliminate undersampling and underrepresentation of microbial diversity within a sample. This analysis suggests that no less than 1 g of soil be used for measuring gene diversity to minimize undersampling.
- The data also suggest that for some samples, such as for two of the Florida samples, diversity estimates increase with increasing amounts of template used, from 10 to 50 ng per reaction. Preferring to err on the side of caution, NEON shall use 50 ng of template per reaction for PCR-based microbial measurements, where available.

6.1.3 Temporal Distribution of Sampling

Initial pilot data at the 16/18S level suggest low seasonal variability, particularly when compared with spatial variability (Table 2). When attributing biotic or abiotic factors to microbial community diversity, seasonal variation is significant only in the 16S rRNA genes of the community from the Florida samples (Table 3).

Table 2. Power –law distance-decay models for 16S and 18S diversity

		z-value	R^2	*p*-value

Spatial				
	16S	-.105	0.40	**<0.001**
	18S	-.032	0.24	**0.012**
Temporal				
	16S	-.0113	0.001	0.41
	18S	-.006	0.003	0.87

* z-value represents the exponent of the power-law function

Table 3. Mantel correlation between within-site environmental variation and community diversity

Site	Taxon	Environmental Variables*	Taxonomic Interaction**	Spatial Distribution	Temporal Distribution
Florida	16S rRNA	.305/**.0003**	0.088	-0.023	.091/**.022**
	18S rRNA	.113/**.004**		0.011	-0.05
Alaska	16S	-0.126	0.31/**.05**	-0.082	-0.009
	18S	0.078		0.025	-0.083
Utah	16S	0.013	0.1	0.01	0.05
	18S	-0.003		-0.025	0.107
Hawaii	16S	.154/**.008**	.156/**.005**	.096/**.004**	-0.058
	18S	0.03		0.041	-0.039

*Environmental variables include pH, %OM, CEC, soil moisture, Mg, Ca, K, Na, total C, total N, biomass (PLFA), Cl, NO_3-N, SO_4, and PO_4.

**Taxonomic interaction is the correlation of 16S and 18S distance matrices.

The first number is the correlation coefficient, the second bolded number is the *p*-value. *P*-values are only reported for significant ($p \leq 0.05$) values.

Measurements for microbial community diversity using 16S/ ITS rRNA sequencing and microbial abundance/biomass will be based on NEON's 2-phase approach for optimizing temporal frequency of sampling. Specifically, for the first two years, samples will be collected at monthly intervals to determine total seasonal variability. Temporal sampling frequency will be optimized to account for maximum seasonal variation. The proposed analysis of seasonal variation will be to place 12 monthly sampling data in an NMDS plot and highlight the 3 samples that are furthest from each other. Each year will provide 2 replicates to give an idea of an overall optimum sampling frequency.

Potential functional diversity and community genome composition will be measured using metagenomics analysis once at each plot during the summer months, when plants at many sites reach

peak greenness. These data will allow for long-term changes in functional potential and microbial speciation to be observed. Details for site-specific sample timing will be included in protocol appendices.

6.1.4 Logistics and Adaptability

The approaches outlined here for measuring microbial processes reflect the best practices currently employed by the broad community of microbial ecologists. It is important to recognize, however, that the field of microbial ecology is evolving, and NEON anticipates that the best practices will change over the course of the observatory. NEON intends to remain on the forefront of the methods and analyses employed for measuring the various aspects of microbial community structure and function. This will be accomplished by continual interaction with and input from the soil microbial ecology community in developing and adopting standardized protocols. Substantial changes in methodology driven by the community of microbial ecologists could warrant future iterative examination of measurement methods.

7 REFERENCES

Aber, J.D., A.H. Magill. 2004. Chronic nitrogen additions at the Harvard Forest (USA): the first fifteen years of a nitrogen saturation experiments. Forest Ecology and Management 196:1-5.

Allison, S.D., J.B.H. Martiny. 2008. Resistance, resilience, and redundancy in microbial communities. Proceedings of the National Academy of Sciences USA 105(Suppl. 1):11512-11519.

Andren O., J. Balandreau. 1999. Biodiveristy and soil functioning – from black box to can of worms? Applied Soil Ecology 13:105-108.

Baker, K.L., S. Langenheder, G.W. Nicol, D. Ricketts, K. Killham, C.D. Campbell, J.L. Prosser. 2009. Environmental and spatial characterization of bacterial community composition in soil to inform sampling strategies. Soil Biology and Biochemistry

Bardgett, R.D., W.D. Bowman, R. Kaufmann, S.K. Schmidt. 2005. A temporal approach to linking aboveground and belowground ecoloty. Trends in Ecology and Evolution 20:634-641.

Bardgett R.D., C. Freeman, N.J. Ostle. 2008. Microbial contributions to climate change through carbon cycle feedbacks. ISME Journal 8: 805-814.

Bell, T., J.A. Newman, B.W. Silverman, S.L. Turner, A.K. Lilley. 2005. The contribution of species richness and composition to bacterial services. Nature 436:1157-1160.

Begerow, D., H. Nilsson, M. Unterseher, W. Maier. 2010. Current state and perspectives of fungal DNA barcoding and rapid identification procedures. Applied Microbiology and Biotechnology 87:99-108.

Canadell, J.G., C. Le Quere, M.R. Raupach, C.B. Field, E.T. Buitenhuis, P. Ciais, T.J. Conway, N.P. Gillett, R.A. Houghton, G. Marland. 2007. Contributions to accelerating atmospheric CO_2 growth from economic activity, carbon intensity, and efficiency of natural sinks. Proceedings of the National Academy of Sciences USA 104:18866-18870.

Caporaso, J.G., C.L. Lauber, W.A. Walters, D. Berg-Lyons C.A. Lozupone, P.J. Turnbaugh, N. Fierer, R. Knight. 2007. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proceedings of the National Academy of Sciences USA 108:4516-4522.

Conley, D.J., H.W. Paerl, R.W. Howarth, D.F. Boesch, S.P. Seitzinger, K.E. Havens, C. Lancelot, G.E. Likens. 2009. Controlling eutrophication: nitrogen and phosphorus. Science 323:1014-1015.

Dane, J.H., G.C. Topp (eds.). 2002. Methods of Soil Analysis. Part 4 –Physical Methods.

DeLong, E.F., C.M. Preston, T. Mincer, V. Rich, S.J. Hallam, N-U. Frigaard, A. Martinez, M.B. Sullivan, R. Edwards, B.R. Brito, S.W. Chisholm, D.M. Karl. 2006. Community genomics among stratified microbial assemblages in the ocean's interior. Science 311:496-503.

Dinsdale, E.A., R.A. Edwards, D. Hall, F. Angly, M. Breitbart, J.M. Bruic, M. Furian, C. Desnues, M. Haynes, L. Li, L. McDaniel, M.A. Moran, K.E. Nelson, C. Nilsson, R. Olson, J. Paul, B.R. Brito, Y. Ruan, B.K.

Swan, R. Stevens, D.L. Valentine, R.V. Thurber, L. Wegley, B.A. White, F. Rohwer. 2008. Functional metagenomic profiling of nine biomes. Nature 452:629-632.

Docherty, K.M., J.L.M. Gutknecht. 2012. The role of environmental microorganisms in ecosystem response to global change: Current state of research and future outlooks. Biogeochemistry 109:63-83.

Docherty, K.M., H.M. Borton, N. Espinosa, M. Gebhardt, J. Gil-Loaiza, J.L.M. Gutknecht, P.W. Maes, et al. 2015. Key edaphic properties largely explain temporal and geographic variation in soil microbial communities across four biomes. PloS One 10 (11): e0135352.

Falkowski, P.G., T. Fenchel, E.F. Delong. 2008. The microbial engines that drive Earth's biogeochemical cycles. Science 320:1034-1039.

Fierer, N., R.B. Jackson. 2006. The diversity and biogeography of soil bacterial communities. Proceedings of the National Academy of Sciences USA 103:626-631.

Finzi, A.C., J.J. Cole, S.C. Doney, E.A. Holland, R.B. Jackson. 2011. Research frontiers in the analysis of coupled biogeochemical cycles. Frontiers in Ecology and the Environment 9:74-80.

Fuhrman, J.A., I. Hewson, M.S. Schwalbach, J.A. Steele, M.V. Brown, S. Naeem. 2006. Annually reoccurring bacterial communities are predictable from ocean conditions. Proceedings of the National Academy of Sciences USA 103:13104-13109.

Galloway, J.N., A.R. Townsend, J.W. Erisman, M. Bakunda, Z. Cai, J.R. Freney, L.A. Martinelli, S.P. Seitzinger, M.A. Sutton. 2008. Transformation of the nitrogen cycle: Recent trends, questions, and potential solutions. Science 320:889-892.

Gilbert, J.A., D. Field, P. Swift, L. Newbold, A. Oliver, T. Smyth, P. Somerfield, S. Huse, I. Joint. 2009. Seasonal succession of microbial communities in the Western English Channel using 16S rDNA-tag pyrosequencing. Environmental Microbiology 11(12), 3132–3139.

Green, J., B.J.M. Bohannan. 2004. Spatial scaling of microbial biodiversity. Trends in Ecology and Evolution 21:501-507.

Hamady, M., J.J. Walker, J.K. Harris, N.J. Gold, R. Knight. 2008. Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. Nature Methods 5:235-237.

Harris, J. 2009. Soil microbial communities and restoration ecology: facilitators or followers? Science 325:573-574.

Iv, L. C. Latta, M. Baker, T. Crowl, J. J. Parnell, B. Weimer, D. B. DeWald, M. E. Pfrender. 2011. Species and Genotype Diversity Drive Community and Ecosystem Properties in Experimental Microcosms. Evolutionary Ecology 25: 1107–25.

Kao, R.H., C.M. Gibson, R.E. Gallery, C.L. Meier, D.T. Barnett, K.M. Docherty, K.K. Blevins, P.D. Travers, E. Azuaje, Y.P. Springer, K.M. Thibault, V.J. McKenzie, M. Keller, L.F. Alves, E.S. Hinckley, J. Parnell, D. Schimel. 2012. NEON terrestrial field observations: designing continental-scale, standardized sampling. Ecosphere 3:art115.

Keller, M., D.S. Schimel, W.W. Hargrove, F.M. Hoffman. 2008. A continental strategy for the National Ecological Observatory Network. Frontiers in Ecology and the Environment 6:282-284.

Lefsky, M.A., W.B. Cohen, G.G. Parker, D.J. Harding. 2002. Lidar remote sensing for ecosystem studies. BioScience 52:20-30.

Lozupone, C.A., R. Knight. 2005. UniFrac: a new phylogenetic method for comparing microbial communities. Applied and Environmental Microbiology 71:8228-8235.

Lozupone, C.A., R. Knight. 2007. Global patterns in bacterial diversity. Proceedings of the National Academy of Sciences USA 104:11436-11440.

National Research Council. 2001. Grand Challenges in Environmental Sciences. 107 pp. The National Academies Press, Washington, D.C.

Nekola J.C., P.S. White. 1999. The distance decay of similarity in biogeography and ecology. Journal of Biogeography 26:867-878.

Nemergut, D.R., E.K. Costello, A.F. Meyer, M. Pescador, M. Weintraub, S.K. Schmidt. 2005. Structure and function of alpine and arctic soil microbial communities. Research in Microbiology 156:775-784.

Raes, J., P. Bork. 2008. Molecular eco-systems biology: towards an understanding of community function. Nature Reviews Microbiology 6:693-699.

Ramirez, K.S., C.L. Lauber, R. Knight, M.A. Bradford, N. FIerer. 2010. Consistent effects of nitrogen fertilization on soil bacterial communities in contrasting systems. Ecology 91:3463-3470.

Rusch, D.B., A.L. Halpern, G. Sutton, K.B. Heidelberg, S. Williamson, S. Yooseph et al. 2007. The *Sorcerer II* global ocean samling expedition: northwest Atlantic through eastern tropical Pacific. PLoS Biology 5:e77.

Schimel, D., W. Hargrove, F. Hoffman, and J. MacMahon. 2007. NEON: a hierarchically designed national ecological network. Frontiers in Ecology and the Environment 5:59–59.

Schimel, J. 2001. Soil microbiology, ecology and biochemistry for the 21st century, in EA Paul (ed) Soil microbiology, Ecology, and Biochemistry. Academic Press.

Sparks, D.L., A.L. Page, P.A. Helmke, R.H. Loeppert, P.N. Soltanpour, M.A. Tabatai, C.T. Johnston, M.E. Sumner (eds.). 1996. Methods of Soil Analysis. Part 3 – Chemical Methods.

Todd-Brown, K.E.O., F.M. Hopkins, S.N. Kivlin, J.M. Talbot, S.D. Allison. 2012. A framework for representing microbial decomposition in coupled climate models. Biogeochemistry 109:19-33.

Treseder, K.K., T.C. Balser, M.A. Bradford, E.L. Brodie, E.A. Dibinsky, V.T. Eviner, K.S. Hofmockel, J.T. Lennon, U.Y. Levine, B.J. MacGregor, J. Pett-Ridge, M.P. Waldrop. 2011. Integrating microbial ecology into ecosystem models: challenges and priorities. Biogeochemistry DOI 10.1007/s10533-011-9636-5.

Tringe, S.G., C. von Mering, A. Kobayashi, A.A. Salamov, K. Chen, H.W. Chang, M. Podar, J.M. Short, E.J. Mathur, J.C. Detter, P. Bork, P. Hugenholtz, E.M. Rubin. 2005. Comparative metagenomics of microbial communities. Science 308:554-557.

Van der Heijden, M.G.A., R.D. Bardgett, N.M. van Straalen. 2008. The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. Ecology Letters 11:296-310.

Vitousek, P.M., H.A. Mooney, J. Lubchenco, J.M. Melillo. 1997. Human domination of Earth's ecosystems. Science 277:494-499.

Wang, Q., Quensen, J. F., Fish, J. A., Lee, T. K., Sun, Y., Tiedje, J. M., & Cole, J. R. 2013. Ecological patterns of nifH genes in four terrestrial climatic zones explored with targeted metagenomics using FrameBot, a new informatics tool. *mBio*, *4*(5), e00592-13.

Whitaker, R.J., D.W. Grogan, J.W. Taylor. 2003. Geographic barriers isolate endemic populations of hyperthermophilic archaea. Science 301:976-978.

Wittebolle, L., M. Marzonrati, L. Clement, A. Balloi, K. Heylen, P. De Vos, D. Daffonchio, W. Verstraete, N. Boon. 2009. Initial species evenness favours functionality under selective stress. Nature 458:623-626.

Woyke, T., G. Xie, A. Copeland, J.M. Gonzalez, C. Han, H. Kiss, J.H. Saw, P. Senin, C. Yang, S. Chatterji, J-F. Cheng, J.A. Eisen, M.E. Sieracki, R. Stepanauskas. 2006. Assembling the marine metagenome, one cell at a time. PLos ONE 4:e5299.

Yergeau, E., M. Arbour, R. Brousseau, D. Juck, J.R. Lawrence, L. Masson, L.G. Whyte, C.W. Greer. 2009. Microarray and real-time PCR analyses of the response of high Arctic soil bacteria to hydrocarbon pollution and bioremediation treatments. Applied and Environmental Microbiology 75:6258-6267.

Yilmaz, P., R. Kottmann, D. Field, R. Knight, J.R. Cole, et al. 2011. Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. Nature Biotechnology 29:415-420.