Title:

Stochastic extremes but convergent recovery of bacterial and archaeal soil communities to an ongoing subterranean coal mine fire

Authors:

Sang-Hoon Lee[1,2a], Jackson W Sorensen[1a], Keara L Grady[1], Tammy C Tobin[3], and Ashley Shade[1,4]*

Affiliations:

[1]Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing MI 48840

[2]School of Civil, Environmental, and Architectual Engineering, Korea University, Seoul, South Korea

[3]Department of Biology, Susquehanna University, Selinsgrove PA 17870

[4]Program in Ecology, Evolutionary Biology, and Behavior, Michigan State University, East Lansing, MI, 48840

[a] contributed equally

*correspondence: shadeash@msu.edu

Keywords:

press disturbance, microbial diversity, thermophile, soil, community assembly, extremophile, anthropogenic, Centralia, resilience

Running title:

Stochastic extremes, convergent recovery

Abstract

Press disturbances are stressors that are extended or ongoing relative to the generation times of community members, and, due to their longevity, have the potential to alter communities beyond the possibility of recovery. They also provide key opportunities to investigate ecological resilience and to probe biological limits in the face of prolonged stressors. The underground coal mine fire in Centralia, Pennsylvania has been burning since 1962 and severely alters the overlying surface soils by elevating temperatures and depositing coal combustion pollutants. As the fire burns along the coal seams to disturb new soils, previously disturbed soils return to ambient temperatures, resulting in a chronosequence of fire impact. We used 16S rRNA gene sequencing to examine bacterial and archaeal soil community responses along two active fire fronts in Centralia, and investigated the influences of assembly processes (selection, dispersal and drift) on community outcomes. The hottest soils harbored the most variable and divergent communities, despite their reduced diversity. Recovered soils converged toward similar community structures, demonstrating resilience within 10-20 years and exhibiting near-complete return to reference communities. Measured soil properties (selection), local dispersal, and neutral community assembly models could not explain the divergences of communities observed at temperature extremes. We hypothesize that transitions between the seed bank and the active community, which would manifest as drift processes, are key in explaining these divergences. These results suggest that soils generally have an intrinsic capacity for robustness to varied disturbances, even to press disturbances considered to be “extreme”, compounded, or incongruent with natural conditions.

Introduction

Human interactions with and alterations of environmental systems are important components of global change (Allen *et al.*, 2014). Anthropogenic disturbances are outcomes of human activity, and include land-use and land-cover changes, pollution, dispersal of invasive species, and over-harvesting of native animal or plant populations (Vitousek *et al.*, 2008). Anthropogenic disturbances are typically classified as press disturbances, as they often impact multiple generations of organisms within their ecosystems (Bender *et al.*, 1984). Because of their longevity, press disturbances have the capacity to alter ecosystems beyond the possibility of recovery (e.g., Thrush *et al.*, 2009).

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2446v1 | CC BY 4.0 Open Access | rec: 13 Sep 2016, publ: 13 Sep 2016

Within every ecosystem, microbial communities underpin biogeochemical processes, sustain the bases of food webs, and recycle carbon and nutrients. In some situations of anthropogenic disturbance, such as pollution, native microbial communities also can provide bioremediative functions to support ecosystem recovery (Ruberto *et al.*, 2009; Desai *et al.*, 2010; Ma *et al.*, 2016; Fuentes *et al.*, 2015). Because of their foundational roles in driving important ecosystem processes, understanding how microbial communities respond to press disturbance can provide insights into the potential for ecosystems to recover. It may also help to uncover mechanisms by which environmental microbial communities may be managed to improve ecosystem outcomes. A better understanding of microbial responses to press disturbances, including examples of communities that have recovered or shifted to an alternative stable state, is necessary to move toward the goal of microbial community management (Shade and Peter *et al.*, 2012).

Recent work has highlighted the importance of understanding the relative contributions of community assembly processes to community changes (e.g., Vellend, 2010; Nemergut *et al.*, 2013; Dini-Andreote *et al.*, 2015; Evans *et al.*, 2016), and these processes can also be informative for understanding community changes after a disturbance (e.g., secondary succession; Dini-Andreote *et al.*, 2015). According to Vellend, 2010, community assembly can be summarized by four major processes: dispersal, diversification, drift, and selection. *Dispersal* is the movement of individuals between localities, *diversification* is the generation of new genetic variation (which can lead to speciation), *drift* encompasses the stochastic processes resulting in fluctuations in member abundances (e.g. births and deaths), and *selection* refers to deterministic fitness differences among taxa driven by abiotic conditions or biotic interactions (as summarized by Nemergut *et al.*, 2013). These processes complement and interact to drive community patterns, and together provide a foundation on which to build a predictive theoretical framework for microbial community ecology.

We aimed to understand the responses of soil microbial communities to an anthropogenic press disturbance, and to apply the Vellend, 2010 and Nemergut *et al.*, 2013 conceptual framework of community assembly for interpretation of patterns. The town of Centralia, Pennsylvania is the site of an underground coal mine fire that has been burning since 1962. It is one of thousands of coal mine fires burning in the world today (Melody and Johnston, 2015), which are inconspicuously common anthropogenic disturbances. However, the Centralia fire is especially long-lived, and, after efforts to extinguish it failed, it was left to burn until it self-extinguished (Nolter and Vice, 2004). The fire is expected to burn slowly until the coal reserves have been consumed. The fire currently underlies more than 150 acres and continues to spread

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2446v1 | CC BY 4.0 Open Access | rec: 13 Sep 2016, publ: 13 Sep 2016

92 slowly (3-7 m/yr Elick, 2011) through underground coal seams. Depending on the depth of the
93 coal bed, it burns at an estimated 46-69 m below the surface (Nolter and Vice, 2004; Elick,
94 2011). Heat, steam and combustion products vent upward from the fire through the overlying
95 soils. The surface soil temperatures can exceed 80°C, scarring the landscape with dead
96 vegetation that reveals the fire's subsurface trajectory. As steam and gasses pass through the
97 overlying rock and soil, soil temperatures increase while soil chemical composition is altered by
98 both spontaneous and microbial-mediated chemical reactions (Janzen and Tobin-Janzen,
99 2008). As the fire expands into new areas, it also retreats from some affected sites, which then
100 recover to ambient temperatures (Elick, 2011; Nolter and Vice, 2004). Thus, the “end” of the
101 disturbance can be delineated by temperature recovery. In this way, a chronosequence of fire-
102 affected Centralia soils provides a space-for-time proxy of disturbance response and recovery.

103 Our research objectives were to understand the diversity and spatio-temporal dynamics
104 of the surface soil bacterial and archaeal communities that have been impacted historically or
105 are currently influenced by the ongoing subterranean coal mine fire in Centralia. Previous work
106 using terminal restriction fragment length polymorphism analysis showed that microbial diversity
107 decreased at hotter sites, and that compositional changes were correlated with soil ammonium
108 and nitrate concentrations (Tobin-Janzen *et al.*, 2005). We move forward from this work to use
109 high throughput sequencing of soil community 16S rRNA genes to quantify the community
110 dynamics along a chronosequence of fire response and recovery. We specifically investigated
111 the community assembly processes of selection, dispersal, and drift.

112

113 **Materials and Methods**

114 *Study site, soil sampling, soil biogeochemistry and microbial community DNA extraction*
115 We undertook fieldwork in Centralia (GPS: 46°46”24’N, 122°50”36W) on 5-6 October 2014. We
116 collected surface soils to capture the expected maximum changes along a chronosequence of
117 fire recovery (**Figure 1**). We sampled two fire fronts along gradients of historical fire activity.
118 Fronts are trajectories of fire spread from the 1962 ignition site outward along near-surface coal
119 seams (Elick, 2011). These fronts include surface soils that were previously hot and have
120 cooled, as well as soils that are currently warmed by the ongoing fire. We collected soil from two
121 unaffected, proximate sites as references, seven recovered sites along the gradient, and nine
122 fire-affected sites (18 total soils), and these collections were distributed across both fire fronts.
123 Soil samples were collected from the top 20 cm of surface soil (core diameter 5.1 cm), and were

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2446v1 | CC BY 4.0 Open Access | rec: 13 Sep 2016, publ: 13 Sep 2016

sieved through 4 mm stainless steel mesh. Collected soils were stored on ice up to 72 hr during transport to the laboratory, then stored at -80°C pending further processing. The physico-chemical characteristics of each soil sample (percent moisture, organic matter 500, NO_3^-, NH_4^+, pH, SO_4, K, Ca, Mg, P, As, and Fe) were assayed by the Michigan State Soil and Plant Nutrient Laboratory according to their standard protocols (East Lansing, MI, USA, http://www.spnl.msu.edu/). Gravimetric soil moisture was measured after drying the soil at 80°C for 2 days. Soil community DNA was extracted from 0.25 g of soil in three technical replicates using the MoBio Power Soil DNA Isolation Kit according to the manufacturer's protocol (MoBio, Solana Beach, CA, USA). The concentration of the extracted DNA was measured using the Qubit® dsDNA BR Assay Kit (Life Technologies, NY, USA), and ranged from 1.3 to 129 ng/μL (average and standard deviation = 33.0 ± 29.9 ng/μL).

Soil cell counts

Direct bacterial and archaeal cell counts were conducted on frozen soil samples based on a protocol to separate cells from soil reported in (Portillo *et al.*, 2013). To dissociate the microbial cells from soil particles, 10 g of soil was mixed with 100 mL of phosphate buffered saline containing 0.5% Tween-20 (PBST). Soil samples were homogenized in a Waring blender three times for 1 min each, followed by a 5 min incubation on ice. Slurries were centrifuged at 1000 x g for 15 min to concentrate soil particulates. Supernatants were set aside and stored at 4°C, and the remaining soil pellets were re-suspended in 100 mL of fresh PBST and blended for an additional 1 min. The soil slurry was then transferred to sterile 250 mL centrifuge bottles and the blender was washed with an additional 25 mL of sterile PBST and added to the slurry before centrifugation at 1000 x g for 15 min. All resulting supernatants for each site were combined, then centrifuged at 10,000 x g for 30 min to pellet cells. Supernatants were discarded, and cell pellets were re-suspended in 10 mL of sterile Milli-q water and 400 mL of 37% formaldehyde to fix cells. 1 mL of cell suspension was then carefully layered over 500 μL of sterile Nycodenz solution (0.8 g/mL in 0.85% NaCl), then centrifuged at 10,000 x g for 40 min. The upper layer was then collected and cells were pelleted by centrifugation at 20,000 x g for 15 min, then resuspended in 1 mL of sterile 0.85% NaCl. To dissociate remaining soil clumps, cell suspensions were sonicated for 10 s in a sonicating water bath.

Cell suspensions were stained with DTAF ((5-(4,6-Dichlorotriazinyl) Aminofluorescein)) according to (Robertson *et al.*, 1999). DTAF-stained smears were visualized on a Nikon Eclipse e800 microscope (Tokyo, Japan) equipped with a Photometrics Coolsnap Myo camera (Tuscon, AZ, USA), and images were collected using Micro-Manager software (Edelstein *et al.*, 2014). Fiji

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2446v1 | CC BY 4.0 Open Access | rec: 13 Sep 2016, publ: 13 Sep 2016

158 image analysis software was used to adjust background, thresholding, and to conduct particle
159 counts from images (Schindelin *et al.*, 2012). Briefly, background correction was completed
160 using an automated rolling ball subtraction with a 35-pixel radius, followed by automatic local
161 thresholding using the Bernsen method with a 12-pixel radius to convert greyscale images to
162 binary. Watershed segmentation was conducted to separate touching nuclei, then particles were
163 counted using the ImageJ *"*Analyze Particles*"* function, excluding anything smaller than 0.1
164 micron (Schneider *et al.*, 2012).
165
166 *Quantitative PCR*
167 We performed quantitative PCR (qPCR) using bacterial and archaeal 16S rRNA gene
168 universal primer sets (**Supporting Table 1;** Caporaso *et al.*, 2012). The reaction mixtures
169 consisted of 10 µL SYBR qPCR Master mix (Quanta Bioscience, Gaithersburg, MD, USA), 0.4
170 µL each of the forward and the reverse primers (0.4 pM), 2 µL of template DNA, and sterilized
171 deionized water to adjust the final volume of 20 µL. The thermal profile was as follows: initial
172 denaturation at 95^{o}C for 10 s, followed by 40 cycles of denaturation at 95^{o}C for 10 s, annealing
173 at 50^{o}C for 15 s, and extension at 72^{o}C for 40 s. A final dissociation protocol (58^{o}C to 94.5^{o}C,
174 increment 0.5^{o}C for 10 s) was performed to ensure the absence of nonspecific amplicons. The
175 reactions were conducted using the Bio-Rad iQ5 real time detection system (Bio-Rad, Hercules,
176 CA, USA). The detailed methods for the calculation of real fluorescent signal intensity and the
177 creation of standard curves were described in previous studies (Kim and Cho, 2010; Lee *et al.*,
178 2013).
179
180 *16S rRNA amplicon sequencing*
181 For each of the 54 DNA samples (18 soils, each with three replicate DNA extractions) and mock
182 community DNA, paired-end sequencing (150 base pair) was performed on the bacterial and
183 archaeal 16S rRNA gene V4 hypervariable region using the Illumina MiSeq platform (Illumina,
184 CA, USA; **Supporting Table 1**; Caporaso *et al.*, 2012). All of the sequencing procedures,
185 including the construction of Illumina sequencing library using the Illumina TruSeq Nano DNA
186 Library Preparation Kit, emulsion PCR, and MiSeq sequencing were performed by the Michigan
187 State University Genomics Core sequencing facility (East Lansing, MI, USA) following their
188 standard protocols. The Genomics Core provided standard Illumina quality control, including
189 base calling by Illumina Real Time Analysis v1.18.61, demultiplexing, adaptor and barcode
190 removal, and RTA conversion to FastQ format by Illumina Bcl2Fastq v1.8.4. Raw sequences
191 were submitted to the GenBank SRA Accession SRP082686.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2446v1 | CC BY 4.0 Open Access | rec: 13 Sep 2016, publ: 13 Sep 2016

To estimate sequencing error, mock community DNA was prepared from six different type strains (*D. radiodurans* ATCC13939, *B. thailandensis* E264, *B. cereus* UW85, *P. syringae* DC3000, *F. johnsoniae* UW101, *E. coli* MG1655). The genomic DNA from these type strains were extracted separately using the EZNA Bacterial DNA Kit (Omega Bio-tek, GA, USA) according to the manufacturer's protocol, and then quantified using the Qubit® dsDNA BR Assay Kit (Life Technologies, NY, USA). Each isolates' 16S rRNA sequence was amplified using universal 27F and 1492R primers. Amplification was performed with the GoTaq Green Master Mix (Promega) with the following reaction conditions: 0.4uM each primer, 20-200 ng template, 12.5ul 2X GoTaq Green Mastermix and nuclease free water to 25 uL final volume. The products were visualized on 1% agarose gels before being cleaned using the Promega Wizard SV Gel and PCR Cleanup System per manufacturer's instructions. Cleaned amplification products were sequenced using the 27F and 1492R primers using the ABI Prism BigDye Terminator Version 3.1 Cycle kit at Michigan State's Genomics Research Technology Support Facility (https://rtsf.natsci.msu.edu/genomics/). Forward and reverse reads were merged using the merger tool in the EMBOSS (V. 6.5.7) package (Rice *et al.*, 2000). Based on the DNA concentration, size of genomic DNA, and 16S rRNA gene copy number, the final mixture contained 100,000 copies of 16S rRNA gene from each strain. The mock community was sequenced alongside the 54 soils' metagenomic DNA.

Sequence processing

Paired-end sequence merging, quality filtering, denoising, singleton-sequence removal, chimera checking, and open-reference Operational Taxonomic Unit (OTU) picking were conducted using a UPARSE workflow v8.1 (Edgar, 2013; Edgar and Flyvbjerg, 2014). Open-reference OTU picking was modified for compatibility with the UPARSE pipeline but proceeded as described for open-reference workflows (Rideout *et al.*, 2014). First, reference-based OTU clustering was conducted using usearch_global command to cluster sequences with 97% identity to the greengenes database (v 13.8, http://greengenes.lbl.gov). Second, de novo OTU picking was performed for any sequences that did not hit the greengenes reference; the uclust command was used to cluster sequences at 97% identity (this step includes chimera checking). The reference-based and de novo OTUs were combined together to create the final dataset. Finally, to reduce the potential effects of candidate contaminant sequences, any sequences in the final dataset that matched 100% to a database of extraneous sequences (found in the mock community) were removed.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2446v1 | CC BY 4.0 Open Access | rec: 13 Sep 2016, publ: 13 Sep 2016

Additional analyses were performed with QIIME v. 1.9.1 (Caporaso *et al.*, 2010b), including alignment with PyNAST (Caporaso *et al.*, 2010a), taxonomic assignment with the RDP Classifier (Wang *et al.*, 2007), tree building with FastTree (Price *et al.*, 2009), subsampling/rarefaction to an equal sequencing depth, and within and comparative diversity calculations (e.g., UniFrac ,Lozupone and Knight, 2005). Sequences identified as Chlorophyta, Streptophyta (i.e., Chloroplasts) and Mitochondria were removed before subsampling to an even sequencing depth. Our sequence analysis workflow and computing notes are available on GitHub (https://github.com/ShadeLab/PAPER_LeeSorensen_inprep/blob/master/Sequence_analysis/MockCommunityWorkflow.md). We used the UPARSE workflow (with the recommended 10% divergence filter) for error rate calculation using the mock community (http://drive5.com/usearch/manual/upp_tut_misop_qual.html).

Ecological statistics

We first assessed the reproducibility of evenly-sequenced technical replicates (DNA extraction and sequencing replicates), and found that replicates were similar to one another in measures of within-sample (alpha) and comparative diversity (beta diversity). The average and standard deviation of weighted (nonnormalized) UniFrac distances between replicates was 0.319 ± 0.126 with a range from 0.105 to 1.29 (maximum distance between different samples was 4.49; **Supporting Figure 1** and **Supporting Table 2**). Given the low technical variability, unrarefied technical replicates were collapsed into one combined set of sequences for each soil core to provide more exhaustive sequencing of each soil; these collapsed samples were subsampled to an even sequencing depth (321,000 sequences per soil), and singleton OTUs (observed only once in the dataset) were removed before proceeding with analysis. Within sample-diversity of species richness, Faith's phylogenetic diversity (whole tree method), and comparative diversity of weighted and unweighted UniFrac distance (nonnormalized and normalized, Lozupone *et al.*, 2011) were calculated within QIIME. The data were then moved into the R environment for statistical analyses. Briefly, we used vegan functions for multivariate hypothesis testing, fitting environmental vectors to ordinations (envfit), constrained ordination (capscale), and Mantel tests (mantel) and to calculate Pielou's evenness (Oksanen *et al.*, 2011); the cmdscale function (stats) for principal coordinates analysis; custom code of neutral models of community assembly (Sloan *et al.*, 2007) as written and implemented by (Burns *et al.*, 2015); and ggplot and ggplots2 for plotting (Wickham, 2009). Our R script is available on GitHub

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2446v1 | CC BY 4.0 Open Access | rec: 13 Sep 2016, publ: 13 Sep 2016

(https://github.com/ShadeLab/PAPER_LeeSorensen_inprep/blob/master/R_analysis/Centralia2014_AmpliconWorkflow.R)

Results and discussion

Soil physical-chemical characteristics and microbial population size

We measured a suite of contextual data for each sampling site, and asked whether any of those data were correlated with surface soil temperature (**Figure 2)**. Centralia soils generally represented a wide range of soil chemistry. We did not find strong correlations between measured contextual data and temperature, with the exception of correlations with ammonium and nitrate (Pearson's R = 0.50 and 0.54, respectively; $p < 0.05$). This finding supports previous work in Centralia showing that ammonium and nitrate were elevated at active vents (Tobin-Janzen *et al.*, 2005). In addition, the pH of recovered sites was consistently lower than reference sites (mean pH = 4.4 and 5.9, respectively), and the hottest soils were more likely to have extreme or disparate values. Notably, in two previous reports, soil ammonium, nitrate, and sulfur concentrations were not necessarily correlated with absolute soil temperature values at Centralia, nor to proximity to an active vent; though extreme or disparate chemistry values were sometimes observed at hot sites, values comparable to unaffected sites were also routinely observed (Tobin-Janzen *et al.*, 2005; Janzen and Tobin-Janzen, 2008). Thus, the authors suggested that duration of fire impact, whether the fire was advancing or receding from the site, and other complex environmental factors were likely contributing.

All soils had the same order of magnitude of 16S rRNA copies per dry mass of soil (**Supporting Figure 2A**, Student's t-test all pairwise $p > 0.20$), with fire-affected soils having the highest copy numbers and recovered soils having the lowest. Total number of cells per dry mass of all soil were within the same order of magnitude, ~10^7 cells per gram of dry soil (**Supporting Figure 2B**, Student's t-test all pairwise $p > 0.09$**)**. Together, these data indicate overall community size is stable across the fire gradient and that any changes in community structure along the fire gradient are due to changes in member abundances rather than to differences in the total number of individuals among soils.

Sequencing summary

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2446v1 | CC BY 4.0 Open Access | rec: 13 Sep 2016, publ: 13 Sep 2016

Sequencing efforts were exhaustive for these soils, as assessed by a clear asymptote achieved with rarefaction (**Supporting Figure 3**). After quality filtering, our 16S rRNA amplicon dataset produced 5,778,000 high-quality reads (5,776,626 sequences after omitting singletons OTUs) with a UPARSE-calculated error rate of 0.469%. In total, we observed 28,220 OTUs (26,846 when omitting singleton OTUs) defined at 97% sequence identity; approximately one-third of OTUs were defined based on high-identity matches to the greengenes v13.8 reference database (8,967 OTUs; 8,794 when omitting singleton OTUs), while two-thirds were defined *de novo* after unsuccessful attempts to match the database (19,253 OTUs; 18,052 when omitting singleton OTUs).

Coal mine fire ecosystems are sources of novel microbial functions, including reported aerobic nitrogen fixation (Ribbe *et al.*, 1997) and novel antibiotics (Wang *et al.*, 2014b, 2014a). Furthermore, thermophiles are of interest for bioprospecting for natural products such as thermally-stable enzymes (e.g., for biomass deconstruction from lignocellulosic crops (Blumer-Schuette *et al.*, 2014) and novel antibiotics (Garg *et al.*, 2012). The large number of *de novo* OTUs in this soil dataset suggests that Centralia soils also harbor substantial undescribed microbial diversity and functions. We observed a broad range of 65 phyla in Centralia soils. Among the *de novo* lineages of interest were several archaeal taxa tentatively identified as Crenarcheaota and Parvarcheaota, and several minor bacterial lineages tentatively assigned as TM6, TM7, OD1, OP11, LD1, WPS-2, and WS-3. A 16S rRNA clone library and T-RFLP study of three soil microbial communities that were each proximate to active coal seam vents in China also reported a proportionally large number of Crenarcheaota among detected archaeal clones (Zhang *et al.*, 2013), suggesting that these may be common inhabitants of soils impacted by long-term fires.

Selection

To understand the influence of selection (deterministic) processes on community responses, we used surface soil temperatures measured in 2014 to designate categorical groups of communities according to their fire classification. Soils classified as reference and recovered had temperatures between 12 and 15°C (ambient air temperature was 13.3°C at the time of soil collection), while soils classified as fire-affected had temperatures ranging from 21 to 58°C. We hypothesized that within-sample diversity would be lower in fire-affected soils because of the extreme environmental filter of high temperatures, which we expected to result in lower richness and less phylogenetic breadth. Faith's phylogenetic diversity and OTU richness both were

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2446v1 | CC BY 4.0 Open Access | rec: 13 Sep 2016, publ: 13 Sep 2016

lowest and most variable for fire-affected soils, and highest for reference sites (**Figure 3;** Student's t-test all pairwise $p < 0.001$). Pielou's evenness had a similar trend, with fire-affected soils having lower evenness than other soils, suggesting that there are a small number of highly dominant OTUs in the fire-affected soils (all pairwise $p > 0.05$, not significant). These results generally agree with studies investigating soil microbial diversity after coal mine reclamation in China and Brazil, respectively, where the most recovered/reconstructed soils (20 years post-mining in Li *et al.*, 2014) and 19 years of reconstruction in Quadros *et al.*, 2016) had highest within-sample diversity and were most comparable to reference sites. Centralia soils are expected to share similar contamination from coal extraction with these mine reclamation soils, but also are distinct because of their thermal conditions and additional contamination by coal combustion products.

We used nonnormalized weighted UniFrac distance to assess comparative community diversity across the fire categories. Weighted UniFrac distance was chosen after considering multiple taxonomic and phylogenetic, and weighted and unweighted metrics. All ordinations revealed the same overarching patterns, demonstrating that these patterns were very robust. However, weighted UniFrac distance provided the highest explanatory value (**Supporting Table 3**), suggesting that changes in both phylogenetic breadth and the relative abundances of taxa are important for interpreting community responses. As compared to recovered and reference sites, fire-affected soils were distinct (PERMANOVA pseudo $F = 16.10$, $R^2 = 0.50$ and $p = 0.001$ on 1000 permutations) and more variable in their community structure (difference in median dispersions = 0.53, $p = 0.008$; **Figure 4**). Differences in surface soil temperature and fire history had most explanatory value on Axis 1 (77.1% variance explained), with nitrate and iron contributing; calcium and pH (and, to a lesser extent, soil moisture) explained variation on Axis 2 (12.7% variance explained, **Supporting Table 4**). Notably, soil fire history was not correlated to community dynamics.

Fire-affected soils were more variable in their community structure across soils, especially in soils at the most extreme temperatures observed (sites C13, C10 which were >50°C at the time of sampling and were at the opposite ends of PCoA2). In contrast, recovered soils were less variable, even though they spanned decades of difference in their years of peak fire activity (the earliest impacted soils that we sampled were last recorded to be hot in 1980; Elick, 2011). Also, recovered soils were very similar in community structure to reference soils. These patterns show that Centralia soils achieve divergent community structures over the transition from ambient to extreme conditions, but then generally converge towards a consistent

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2446v1 | CC BY 4.0 Open Access | rec: 13 Sep 2016, publ: 13 Sep 2016

355 community structure after the fire subsides. These results also show high resilience of soil
356 communities impacted by an extreme press disturbance, with recovery occurring within 10-20
357 years after the stressor subsided.

358 We observed a temperature "threshold" effect among fire-affected soils, and soils with
359 temperatures between 21 and 24.5°C (sites C06, C11, and C16) separated cleanly from soils
360 with temperatures greater than 30°C (**Figure 4**). To better understand the divergence in
361 community structure among fire-affected soils, we performed a PCoA with these communities
362 (**Supporting Figure 4A, Supporting Table 5**), and also a constrained analysis to ask what
363 variability remained after removing the influence of temperature (**Supporting Figure 4B,**
364 **Supporting Table 6**). Even after removing the influence of temperature, three discrete subsets
365 of fire-affected communities separated from each other along both axes, with C13 remaining as
366 an outlying point. C13 had very different calcium and pH than the other soils, and both of these
367 factors had high value in discriminating C13 from the other fire-affected soils ($p = 0.092$ and
368 0.014 respectively). There were no other measured abiotic factors that explained the divergence
369 among the fire-affected soils. In addition, the constrained axes had high explanatory value
370 (**Supporting Figure 4B,** combined axes 1 and 2 = 90.0% var. explained), suggesting that there
371 are additional processes beyond selection that explain the differences in these subsets.

372 We observed broad phylum-level changes in response to the fire (**Figure 5A**,
373 **Supporting Table 8**). Not all OTUs affiliated with particular phyla had identical responses;
374 however, our analysis of phylum-level responses points to some general trends. In particular,
375 fire-affected soils were enriched for members of Chloroflexi, Crenarcheaota and many lineages
376 of unidentified Bacteria. As compared to the fire-affected soils, recovered soils also were
377 enriched for Parvarchaeota, Bacteroidetes, Elusimicrobia, Gemmatimonadetes,
378 Planctomycetes, Spirochaetes, TM6, and Verrucomicrobia suggesting that members affiliated
379 with this these phyla are able to persist after the fire subsides. Acidobacteria also had an
380 increase in recovered soils (but less significant, $p = 0.10$), presumably because of the decrease
381 in soil pH observed post-fire (**Figure 2**). Reference soils had higher representation of
382 Proteobacteria and Verrucomicrobia, which suggests that members of these phyla may be
383 sensitive to the fire.

384

385 *Dispersal and drift*

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2446v1 | CC BY 4.0 Open Access | rec: 13 Sep 2016, publ: 13 Sep 2016

To investigate the relative importance of local dispersal, we assessed the value of spatial distance for explaining differences in community structure. If local dispersal were important, we would expect that soils in close proximity would have more similar community structures than soils that are distant from one another. We found no relationship in the measured spatial distances between soil collection sites and their corresponding differences in community structure for all sites (Mantel p = 0.66 on 999 permutations), nor for recovered sites only (after removing the fire-affected sites from analysis; Mantel p = 0.135 on 999 permutations). The lack of evidence for spatial autocorrelation suggests that local dispersal is not a key factor shaping community structure in Centralia soils.

To explore the relative importance of drift in fire-affected and recovered soils, we fitted a neutral model of community assembly. The model predicts taxon frequencies as a function of their metacommunity log abundances, which is one method to consider the influence of drift with the influence of dispersal (calculated as an immigration term, *m*, to the model). The neutral model fit better to the recovered sites than to fire-affected sites (R-squared = 0.53, 0.12 respectively; **Supporting Figure 5, Supporting Table 7**). Furthermore, we found a lesser influence of dispersal (lower value of *m)* in the fire-affected sites regardless of whether the complete set of taxa from the regional metacommunity (taxa from all sites, including fire-affected, recovered, and reference) was used as the regional sample pool or whether the taxa detected only in fire-affected sites were used as the regional sample pool (**Supporting Table 7**). These differences in fit and generally minimal influence of dispersal suggest that neutral processes play a more minor role in the microbial community assembly of fire-affected sites than they do in the recovered sites.

Understanding community divergences at temperature extremes

To dig deeper into the differences in the three subsets of fire-affected soil that were not well explained by abiotic selection, local dispersal, or drift as assessed by the Sloan neutral model of community assembly (**Supporting Figure 4**), we asked if there were notable differences in their dominant memberships. Fire-affected soils generally had more variability and greater phylogenetic breadth in their dominant membership than recovered soils, and each fire-affected subset harbored an exclusive membership among their most prevalent taxa, supporting an influence of drift on community outcomes. We examined the top 10 prevalent taxa from each of the nine fire-affected soils. Collectively, there were 68 unique top 10 OTUs in fire-affected

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2446v1 | CC BY 4.0 Open Access | rec: 13 Sep 2016, publ: 13 Sep 2016

soils (out of a possible 90, if each of the nine fire-affected soil harbored mutually exclusive membership across their top 10). These prevalent fire-affected OTUs spanned fourteen phyla or Proteobacteria classes, included 30 *de novo* OTUs, and included seven taxa of unidentified Bacteria and two taxa of unidentified Proteobacteria. Acidobacteria OTUs were detected among the top 10 for all fire-affected soils, and eight of nine fire-affected soils included Chloroflexi among the top 10 OTUs. In comparison, recovered soils included ten phyla or Proteobacteria classes among their collective top 10, had no unidentified Bacteria or Proteobacteria, and included four *de novo* OTUs. Acidobacteria and Alphaproteobacteria OTUs were among the top 10 for all recovered soils, and six of the seven recovered soils also included Deltaproteobacteria. Together, these results show that fire-affected soils were more divergent and diverse in their prevalent membership than recovered soils.

An analysis of occurrence patterns of the OTUs detected among the top 10 also showed greater divergence among fire-affected soils than recovered, and further supported the distinction among the subsets of fire-affected soils revealed by the constrained ordination (**Figure 6**). Fire-affected soils had more OTUs within their collective most prevalent taxa, and were more heterogeneous as shown by the wider range represented by the color scale and the more divergent sample and OTU clustering. In fact, taxa that were among the top 10 in one fire-affected soil were likely to be among the rare biosphere in another fire-affected soil, exhibiting stark contrast in their abundances within these soils. However, most of the top 10 prevalent OTUs were detected within every fire-affected soil, suggesting that changes in taxa relative abundances, rather than turnover in membership, were driving these patterns.

This dominance analysis helps to explain the lower fit of the neutral model to fire-affected soils. Outliers to the neutral model that were below detection (taxa that were present in fewer sites than predicted given their relative abundance in the metacommunity) included these many lineages that were prevalent in few soils. Taxa that fall below their neutral model prediction have been proposed to be “selected against” or particularly dispersal limited (Burns 2015). However, in the Centralia extreme environment, these were taxa that were most successful locally given the thermal disturbance.

Community assembly processes and the role of the seed bank in disturbance responses

Centralia soil communities were sensitive to the coal mine fire, and changed substantially from reference conditions. Selection processes, specifically abiotic soil conditions, offered high

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2446v1 | CC BY 4.0 Open Access | rec: 13 Sep 2016, publ: 13 Sep 2016

explanatory value for Centralia soil community dynamics. These communities first were constrained by environmental filters imposed by the press disturbance, such as high temperatures in fire-affected soils and low pH in recovered soils. The fire acts as a strong environmental filter, resulting in decreased diversity and a very different phylogenetic representation among the surviving lineages in fire-affected soils. However, even after removing the influence of temperature on fire-affected communities, they fell into three distinct subsets that could not be explained by the physico-chemical characteristics measured in this study. Furthermore, neutral model fits and lack of spatial autocorrelation suggests that these particular assessments for drift and dispersal processes offer minimal explanation for fire-affected sites. Given the low explanatory value of unweighted resemblances in describing patterns of comparative diversity (**Supporting Table 3**), and the observation that many of the prevalent taxa detected in some fire-affected soils were rare in other fire-affected soils (**Figure 6A**), we can also attribute these patterns to changes in the relative abundances of taxa, rather than to changes in taxa turnover (differing memberships). Thus, given that neither assessed selection, dispersal, nor drift processes, nor their combination can provide a complete explanation for the divergence of fire-affected communities, the question remains: why are fire-affected soils so divergent from each other, and yet eventually manage to recover to the same post-disturbance community structure?

We hypothesize that the remaining variability in community structure of fire-affected sites may be attributed to fluctuations between the dormant seed bank and the active community. Given that proportion of dormant cells in soils is estimated to be near 80% (Lennon and Jones, 2011), we posit that seed bank fluctuations contribute to drift processes that are not well-quantified with current methods, but are nonetheless indirectly supported by the results of this study.

The general importance of dormancy for microbial community assembly processes has been proposed previously by Nemergut *et al.*, 2013, and the unique conditions of Centralia may offer opportunity to investigate dormancy processes *in situ*. There are two general aspects of seed banks that could help to explain Centralia community divergences at temperature extremes: membership and dynamics. If each soil harbored a different seed bank membership, different thermophilic taxa could become active and prevalent in each fire-affected soil. This scenario is not well-supported by our data because we detect the dominant members of each fire-affected soil in the other fire-affected soils, albeit in lower abundances. Alternatively, stochastic awakenings from the microbial seed bank (Buerger *et al.*, 2012) could result in

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2446v1 | CC BY 4.0 Open Access | rec: 13 Sep 2016, publ: 13 Sep 2016

"priority effects" at temperature extremes, in which the first microorganisms to wake after the fire's local onset have important influence over the community's ultimate trajectory (e.g., Fukami, 2015). In our chronosequence study, the outcome of priority effects would appear as divergent community structures at high temperatures, therefore manifesting as unexplained drift. Our data indirectly support this scenario, as the three separate clusters of fire-affected communities (**Supporting Figure 4B**) hint that consistent trajectories are possible. It could be that the most similar fire-affected communities started from the same waking pioneer taxon.

Interactions between active and dormant memberships could be manifested in part as within-site membership fluctuations, as we observed in this study. Given that awakenings from the microbial seed bank may be either responsive to a cue or stochastic (e.g., Lennon and Jones, 2011; Buerger *et al.*, 2012), newly active seed bank members may contribute either to selection or drift processes; yet, these contributions would be very difficult to partition and may remain unexplained because common statistical methods do not explicitly consider them (e.g. Sloan *et al.*, 2007). Because of analysis challenges as well as challenges in observing environmental microbial seed banks, there is a knowledge gap in understanding the roles of seed bank members for community responses to disturbances (though, see Aanderud *et al.*, 2015 for a key example of responses of rare or dormant taxa to soil re-wetting events). Thermophiles are prime examples of seed bank members that often have been found in environments that are improbable to permit their growth (e.g., Hubert *et al.*, 2009; McBee and McBee, 1956; Portillo *et al.*, 2012). Thus, a greater understanding of thermophile contributions to temperate soil seed banks and their strategies for resuscitation may provide insights into the general implications of dormant and active members for microbial community robustness to stressors.

Diversification is a fourth community assembly process discussed by Vellend, 2010 and Nemergut *et al.*, 2013. We do not directly address diversification in this study, focusing instead on ecological processes. Aside from a consistent observation of Acidobacteria and Chloroflexi among the dominant taxa in fire-affected soils, there is no evidence that different but closely related lineages are most prevalent across all fire-affected soils, which may have hinted to distinct but parallel trajectories of diversification within a locality. However, at this time, we cannot reject the hypothesis that diversification processes also contribute to divergences in community structure at temperature extremes.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2446v1 | CC BY 4.0 Open Access | rec: 13 Sep 2016, publ: 13 Sep 2016

515 *Conclusions*

516 Regardless of the interim dynamics that resulted in community divergence to the
517 stressor, Centralia soils eventually recovered to a community structure very similar to reference
518 soils, and these community structures were explained by the ultimate post-fire soil environment.
519 Thus, our results show that Centralia soil communities, though sensitive to this extreme,
520 complex, and arguably unnatural stressor, had near-complete return to pre-disturbance
521 conditions, and were resilient within ten to twenty years after the stressor subsides. We have
522 no reason to suspect that temperate soils in Centralia are exceptional as compared to other
523 soils. Thus, these results suggest that soils may have an intrinsic capacity for robustness to
524 varied disturbances, even to those disturbances considered to be “extreme”, compounded, or
525 incongruent with natural conditions. Understanding the precise functional underpinnings of soil
526 microbial community resilience, including the roles of seed banks in determining that resilience,
527 is a next important step in predicting and, potentially, managing, microbial community responses
528 to disturbances.

529

530 **Acknowledgements**

531

532 This work was supported by Michigan State University, with computing resources provided by
533 the Michigan State Institute for Cyber-Enabled Research. JWS acknowledges the Dr. C. A.
534 Reddy and Sasikala Reddy Award from the Michigan State Department of Microbiology and
535 Molecular Genetics. We thank Trevor Grady for his graphic design work for Figure 1.

536

537 **References**

538 Aanderud Z, Jones S, Fierer N, Lennon JT (2015). Resuscitation of the rare biosphere
539 contributes to pulses of ecosystem activity *Front Microbiol* **6**: 1–11.

540 Allen MR, Barros VR, Broome J, Cramer W, Christ R, Church JA, *et al.* (2014). IPCC Fifth
541 Assessment Synthesis Report-Climate Change 2014 Synthesis Report. *IPCC Fifth Assess*
542 *Synth Report-Climate Chang 2014 Synth Rep* pages: 167.

543 Bender EEA, Case TJT, Gilpin ME. (1984). Perturbation Experiments in Community Ecology:
544 Theory and Practice. *Ecology* **65**: 1–13.

545 Blumer-Schuette SE, Brown SD, Sander KB, Bayer EA, Kataeva I, Zurawski J V., *et al.* (2014).

546 Thermophilic lignocellulose deconstruction. *FEMS Microbiol Rev* **38**: 393–448.

547 Buerger S, Spoering A, Gavrish E, Leslin C, Ling L, Epstein SS. (2012). Microbial scout
548 hypothesis, stochastic exit from dormancy, and the nature of slow growers. *Appl Environ*
549 *Microbiol* **78**: 3221–3228.

550 Burns AR, Stephens WZ, Stagaman K, Wong S, Rawls JF, Guillemin K, *et al.* (2015).
551 Contribution of neutral processes to the assembly of gut microbial communities in the zebrafish
552 over host development. *Isme J* **in press**: 1–10.

553 Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. (2010a).
554 PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* **26**: 266–
555 267.

556 Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, *et al.* (2010b).
557 QIIME allows analysis of high-throughput community sequencing data. *Nature* **7**: 335–336.

558 Caporaso JG, Lauber CL, Walters W a, Berg-Lyons D, Huntley J, Fierer N, *et al.* (2012). Ultra-
559 high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME*
560 *J* **6**: 1621–1624.

561 Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, *et al.*
562 (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample.
563 *Proc Natl Acad Sci U S A* **108**: 4516.

564 Desai C, Pathak H, Madamwar D. (2010). Advances in molecular and '-omics' technologies to
565 gauge microbial communities and bioremediation at xenobiotic/anthropogen contaminated sites.
566 *Bioresour Technol* **101**: 1558–1569.

567 Dini-Andreote F, Stegen JC, van Elsas JD, Salles JF. (2015). Disentangling mechanisms that
568 mediate the balance between stochastic and deterministic processes in microbial succession.
569 *Proc Natl Acad Sci* **112**: E1326–E1332.

570 Edelstein AD, Tsuchida M a, Amodaj N, Pinkard H, Vale RD, Stuurman N. (2014). Advanced
571 methods of microscope control using μManager software. *J Biol Methods* **1**: 10.

572 Edgar RC. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads.
573 *Nat Methods* **10**: 996–8.

574 Edgar RC, Flyvbjerg H. (2014). Error filtering, pair assembly and error correction for next-
575 generation sequencing reads. *Bioinformatics* **31**: 3476–3482.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2446v1 | CC BY 4.0 Open Access | rec: 13 Sep 2016, publ: 13 Sep 2016

576 Elick JM. (2011). Mapping the coal fire at Centralia, Pa using thermal infrared imagery. *Int J*
577 *Coal Geol* **87**: 197–203.

578 Evans S, Martiny JB, Allison SD. (2016). Effects of dispersal and selection on stochastic
579 assembly in microbial communities. *ISME J* 1–10.

580 Fuentes S, Barra B, Gregory Caporaso J, Seeger M. (2015). From rare to dominant: A fine-
581 tuned soil bacterial bloom during petroleum hydrocarbon bioremediation. *Appl Environ Microbiol*
582 **82**: 888–896.

583 Fukami T. (2015). Historical contingency in community assembly : integrating niches, species
584 pools, and priority effects. *Annu Rev Ecol Evol Syst* **46**: 1–23.

585 Garg N, Tang W, Goto Y, Nair SK, van der Donk W a. (2012). Lantibiotics from Geobacillus
586 thermodenitrificans. *Proc Natl Acad Sci U S A* **109**: 5241–5246.

587 Hubert C, Loy a., Nickel M, Arnosti C, Baranyi C, Bruchert V, *et al.* (2009). A Constant Flux of
588 Diverse Thermophilic Bacteria into the Cold Arctic Seabed. *Science (80-)* **325**: 1541–1544.

589 Janzen C, Tobin-Janzen T. (2008). Microbial Communities in Fire-Affected Soils. In:
590 *Microbiology of Extreme Soils*. Springer, pp 299–316.

591 Kim H-J, Cho J-C. (2010). Simple and rapid detection of Listeria monocytogenes in fruit juice by
592 real-time PCR without enrichment culture. *Food Control* **21**: 1419–1423.

593 Lee SH, Park JH, Kang HJ, Lee YH, Lee TJ, Park HD. (2013). Distribution and abundance of
594 Spirochaetes in full-scale anaerobic digesters. *Bioresour Technol* **145**: 25–32.

595 Lennon JT, Jones SE. (2011). Microbial seed banks: the ecological and evolutionary
596 implications of dormancy. *Nat Rev Microbiol* **9**: 119–130.

597 Li Y, Wen H, Chen L, Yin T. (2014). Succession of bacterial community structure and diversity in
598 soil along a chronosequence of reclamation and re-vegetation on coal mine spoils in China.
599 *PLoS One* **9**. e-pub ahead of print, doi: 10.1371/journal.pone.0115024.

600 Lozupone C, Knight R. (2005). UniFrac: a new phylogenetic method for comparing microbial
601 communities. *Appl Environ Microbiol* **71**: 8228–8235.

602 Lozupone C, Lladser ME, Knights D, Stombaugh J, Knight R. (2011). UniFrac: an effective
603 distance metric for microbial community comparison. *ISME J* **5**: 169–172.

604 Ma Y, Rajkumar M, Zhang C, Freitas H. (2016). Beneficial role of bacterial endophytes in heavy
605 metal phytoremediation. *J Environ Manage* **174**: 14–25.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2446v1 | CC BY 4.0 Open Access | rec: 13 Sep 2016, publ: 13 Sep 2016

606 MCBEE RH, MCBEE VH. (1956). The incidence of thermorphilic bacteria in arctic soils and
607 waters. *J Bacteriol* **71**: 182–187.

608 Melody S, Johnston F. (2015). Coal mine fires and human health: What do we know? *Int J Coal*
609 *Geol* **152**: 1:14.

610 Nemergut DR, Schmidt SK, Fukami T, O'Neill SP, Bilinski TM, Stanish LF, *et al.* (2013).
611 Patterns and Processes of Microbial Community Assembly. *Microbiol Mol Biol Rev* **77**: 342–356.

612 Nolter M a, Vice DH. (2004). Looking back at the Centralia coal fire: a synopsis of its present
613 status. *Int J Coal Geol* **59**: 99–106.

614 Oksanen AJ, Blanchet FG, Kindt R, Minchin PR, Hara RBO, Simpson GL, *et al.* (2011). vegan :
615 community ecology package. *R Packag version 115-1*. http://cran.r-project.org/, http://vegan.r-
616 forge.r-project.org.

617 Portillo MC, Leff JW, Lauber CL, Fierer N. (2013). Cell size distributions of soil bacterial and
618 archaeal taxa. *Appl Environ Microbiol* **79**: 7610–7617.

619 Portillo MC, Santana M, Gonzalez JM. (2012). Presence and potential role of thermophilic
620 bacteria in temperate terrestrial environments. *Naturwissenschaften* **99**: 43–53.

621 Price MN, Dehal PS, Arkin AP. (2009). FastTree: Computing large minimum evolution trees with
622 profiles instead of a distance matrix. *Mol Biol Evol* **26**: 1641–1650.

623 Quadros PD de, Zhalnina K, Davis-Richardson AG, Drew JC, Menezes FB, Camargo FA d. O,
624 *et al.* (2016). Coal mining practices reduce the microbial biomass, richness and diversity of soil.
625 *Appl Soil Ecol* **98**: 195–203.

626 Ribbe M, Gadkari D, Meyer O. (1997). N2 Fixation by Streptomyces thermoautotrophicus
627 Involves a Molybdenum-Dinitrogenase and a Manganese-Superoxide Oxidoreductase That
628 Couple N2Reduction to the Oxidation of Superoxide Produced from O2by a Molybdenum-CO
629 Dehydrogenase. *J Biol Chem* **272**: 26627–26633.

630 Rice P, Longden I, Bleasby A. (2000). EMBOSS: The European Molecular Biology Open
631 Software Suite. *Trends Genet* **16**: 276–277.

632 Rideout JR, He Y, Navas-Molina JA, Walters WA, Ursell LK, Gibbons SM, *et al.* (2014).
633 Subsampled open-reference clustering creates consistent, comprehensive OTU definitions and
634 scales to billions of sequences. *PeerJ* **2**: e545.

635 Robertson GP, Coleman DC, Bledsoe C (eds). (1999). Standard Soil Methods for Long-Term

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2446v1 | CC BY 4.0 Open Access | rec: 13 Sep 2016, publ: 13 Sep 2016

Ecological Research. Oxford University Press: Cary, NC, USA.

Ruberto L, Dias R, Lo Balbo A, Vazquez SC, Hernandez EA, Mac Cormack WP. (2009). Influence of nutrients addition and bioaugmentation on the hydrocarbon biodegradation of a chronically contaminated Antarctic soil. *J Appl Microbiol* **106**: 1101–1110.

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, *et al.* (2012). Fiji: an open source platform for biological image analysis. *Nat Methods* **9**: 676–682.

Schneider C a, Rasband WS, Eliceiri KW. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**: 671–675.

Shade A, Peter H, Allison SD, Baho D, Berga M, Buergmann H, *et al.* (2012). Fundamentals of microbial community resistance and resilience. *Front Microbiol* **3**: 417.

Sloan WT, Woodcock S, Lunn M, Head IM, Curtis TP. (2007). Modeling taxa-abundance distributions in microbial communities using environmental sequence data. In: Vol. 53. *Microbial Ecology*. pp 443–455.

Thrush SF, Hewitt JE, Dayton PK, Coco G, Lohrer AM, Norkko A, *et al.* (2009). Forecasting the limits of resilience: integrating empirical research with theory. *Proc R Soc B Biol Sci* **276**: 3209–3217.

Tobin-Janzen T, Shade A, Marshall L, Torres K, Beblo C, Janzen C, *et al.* (2005). Nitrogen Changes and Domain Bacteria Ribotype Diversity in Soils Overlying the Centralia, Pennsylvania Underground Coal Mine Fire. *Soil Sci* **170**: 191–201.

Vellend M. (2010). Conceptual synthesis in community ecology. *Q Rev Biol* **85**: 183–206.

Vitousek PM, Mooney HA, Lubchenco J, Melillo JM. (2008). Human domination of Earth's ecosystems. In: *Urban Ecology: An International Perspective on the Interaction Between Humans and Nature*. pp 3–13.

Wang Q, Garrity GM, Tiedje JM, Cole JR. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* **73**: 5261–5267.

Wang X, Elshahawi SI, Shaaban KA, Fang L, Ponomareva L V., Zhang Y, *et al.* (2014a). Ruthmycin, a new Tetracyclic Polyketide from Streptomyces sp. RM-4-15. *Org Lett* **16**: 456–459.

Wang X, Shaaban KA, Elshahawi SI, Ponomareva L V, Sunkara M, Copley GC, *et al.* (2014b).

666 Mullinamides A and B, new cyclopeptides produced by the Ruth Mullins coal mine fire isolate
667 Streptomyces sp. RM-27-46. *J Antibiot (Tokyo)* **67**: 571–5.

668 Wickham H. (2009). ggplot2: elegant graphics for data analysis. Springer: New York.

669 Zhang T, Xu J, Zeng J, Lou K. (2013). Diversity of prokaryotes associated with soils around
670 coal-fire gas vents in MaNasi county of Xinjiang, China. *Antonie van Leeuwenhoek, Int J Gen*
671 *Mol Microbiol* **103**: 23–36.

672

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2446v1 | CC BY 4.0 Open Access | rec: 13 Sep 2016, publ: 13 Sep 2016

673 **Figures**

674 **Figure 1.** Soil physical and chemical contextual data (x-axis) plotted against temperature (y-
675 axis). Color gradient shows the soil temperature, and symbols show soil fire classification in
676 October 2014 as fire-affected, recovered, or reference.

677 **Figure 2.** Soil sampling sites at Centralia mine fire. In total, 18 surface soil samples (5.08 cm x
678 20 cm PVC core) were collected along two fire fronts in Centralia, on 15/16 October 2014.
679 Sampling sites encompass a gradient of historical fire activity (red flags: Fire-affected in 2014
680 (temperature > 21°C); yellow flags: recovered in temperature, post-fire; and green flags:
681 reference soils).

682 **Figure 3.** Within-sample (alpha) diversity of fire-affected, recovered, and reference soils in
683 Centralia for bacterial and archaeal community (**A**) Faith's phylogenetic diversity; (**B**) richness
684 (total no. observed OTUs clustered at 97% sequence identity); and (**C**) Pielou's evenness.

685 **Figure 4**. Principal coordinate analysis (PCoA) based on weighted UniFrac distances of
686 phylogenetic bacterial and archaeal community structure. Colors show the fire classification of
687 the soil as fire-affected (red), recovered (yellow), or reference (green). The strength of
688 statistically significant ($p < 0.10$) explanatory variables are shown with solid arrows.

689 **Figure 5**. Phylum-level responses to the Centralia coal mine fire. Mean relative abundance of
690 phyla summarized within soil fire classifications (fire-affected, recovered, and reference).

691 **Figure 6**. Relative abundances of the collection of the most prevalent "top 10" taxa (rows)
692 observed in any (**A**) fire-affected or (**B**) recovered soil in Centralia. Warm colors show
693 prevalence within a site (columns), and cool colors show rarity. Note differences in color scale
694 gradient between (**A**) and (**B**).

695

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2446v1 | CC BY 4.0 Open Access | rec: 13 Sep 2016, publ: 13 Sep 2016

696 **Tables**

697

698 **Table 1**. Ten most abundant OTUs in fire-affected Centralia soils.

OTU ID	Cumulative % abundance (out of total No. sequences in fire-affected samples)	% occurrence (out of 9 warm or venting fire-affected soils)	Taxonomic assignment
111933	5.5%	100%	k__Archaea; p__Crenarchaeota; c__MBGA; o__; f__; g__; s__
OTU_dn_1	2.5	100%	k__Bacteria; p__Chloroflexi; c__Ktedonobacteria; o__Thermogemmatisporales; f__Thermogemmatisporaceae; g__; s__
OTU_dn_2	2.2	100%	k__Bacteria; p__Chloroflexi; c__Ktedonobacteria; o__Thermogemmatisporales; f__Thermogemmatisporaceae; g__; s__
242467	2.0	100%	k__Bacteria; p__Acidobacteria; c__DA052; o__Ellin6513; f__; g__; s__
174835	2.0	100%	k__Archaea; p__Crenarchaeota; c__Thermoprotei; o__YNPFFA; f__SK322; g__; s__
61819	1.7	100%	k__Bacteria; p__Acidobacteria; c__TM1; o__; f__; g__; s__
OTU_dn_17	1.5	78%	k__Bacteria; p__Proteobacteria; c__Deltaproteobacteria
215700	1.4	100%	k__Bacteria; p__Acidobacteria; c__Acidobacteriia; o__Acidobacteriales; f__Koribacteraceae; g__; s__
OTU_dn_8	1.3	100%	k__Bacteria
OTU_dn_3	1.2	100%	k__Bacteria

699

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2446v1 | CC BY 4.0 Open Access | rec: 13 Sep 2016, publ: 13 Sep 2016

700 **Supporting Figures**

701 **Supporting Figure 1**. PCoA showing the variability among technical replicates. Three replicate
702 DNA extractions, amplifications and sequencing reactions were performed per soil, and these
703 sequences were subsequently pooled into one aggregate set of sequences to achieve deep
704 coverage of the community within each soil. Error bars are standard deviation around the mean
705 weighted UniFrac distance among technical replicates, each subsampled to an even 53,000
706 sequences per replicate.
707
708 **Supporting Figure 2**. Quantification of (**A**) 16S rRNA copies and (**B**) cell counts in fire-
709 affected, recovered, and reference soils. 16S rRNA copies were assessed using quantitative
710 PCR, and cell counts were assessed using cell separation from soil, staining and microscope
711 imaging.

712 **Supporting Figure 3**. Centralia 16S rRNA amplicon sequencing effort assessed by
713 subsampling/rarefaction of (**A**) richness and (**B**) Faith's phylogenetic diversity with increasing
714 total number of sequences.

715 **Supporting Figure 4**. Divergences in fire-affected soils are not well explained by temperature.
716 **(A)** Principal coordinate analysis (PCoA) based on weighted UniFrac distances of phylogenetic
717 bacterial and archaeal community structure in fire-affected soils. The strength of statistically
718 significant ($p < 0.10$) explanatory variables are shown with blue arrows. **(B)** Constrained
719 analysis based on weighted UniFrac distances, where the explanatory value of temperature is
720 removed from the analysis to understand the influence of the remaining explanatory variables.

721 **Supporting Figure 5.** Neutral models of community assembly (abundance v. occurrence) for
722 (**A**) the total community ("All", n= 18), **(B)** fire-affected soils ("Fire_Affected", n=9) and **(C)**
723 recovered soils ("Recovered" n=7). Red symbols show OTUs that had higher abundance than
724 their prediction, and blue symbols show OTUs that had lower abundance than their prediction.
725 The thick yellow line is the neutral model prediction, and the thin yellow lines show a 95%
726 confidence interval around the prediction.

727

728 **Supporting Tables**

729 **Supporting Table 1.** Primers used in this study.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2446v1 | CC BY 4.0 Open Access | rec: 13 Sep 2016, publ: 13 Sep 2016

730 **Supporting Table 2**. Mean and standard deviation ("sd") phylogenetic diversity and number of
731 OTUs ("richness) across technical sequencing replicates for the un-collapsed dataset (rarefied
732 to 53,000 sequences per sample). Three replicate DNA extractions, amplifications and
733 sequencing reactions were performed per soil, and, after calculating the technical variability,
734 these sequences were pooled into one aggregate set of sequences to achieve deep coverage
735 of the community within each soil.

736 **Supporting Table 3.** Percent variation explained for PCoA axes 1 and 2 for nonnormalized
737 weighted and unweighted UniFrac, normalized weighted UniGrac, Sorensen-dice, and Bray-
738 Curtis distances/dissimilarities. Nonnormalized weighted UniFrac was chosen because it was
739 most informative in explaining the variance along the first two axes.

740 **Supporting Table 4**. Explanatory value of soil contextual data to changes in Centralia soil
741 community structure along PCoA axes for the all soils. Factors significant at $p < 0.10$ are in
742 bold.

743 **Supporting Table 5**. Explanatory value of soil contextual data to changes in Centralia soil
744 community structure along PCoA axes for the fire-affected soils. Factors significant at $p < 0.10$
745 are in bold.

746 **Supporting Table 6**. Explanatory value of soil contextual data to changes in Centralia soil
747 community structure along the constrained PCoA axes for the fire-affected soils, after removing
748 the influence of temperature. Factors significant at $p < 0.10$ are in bold.

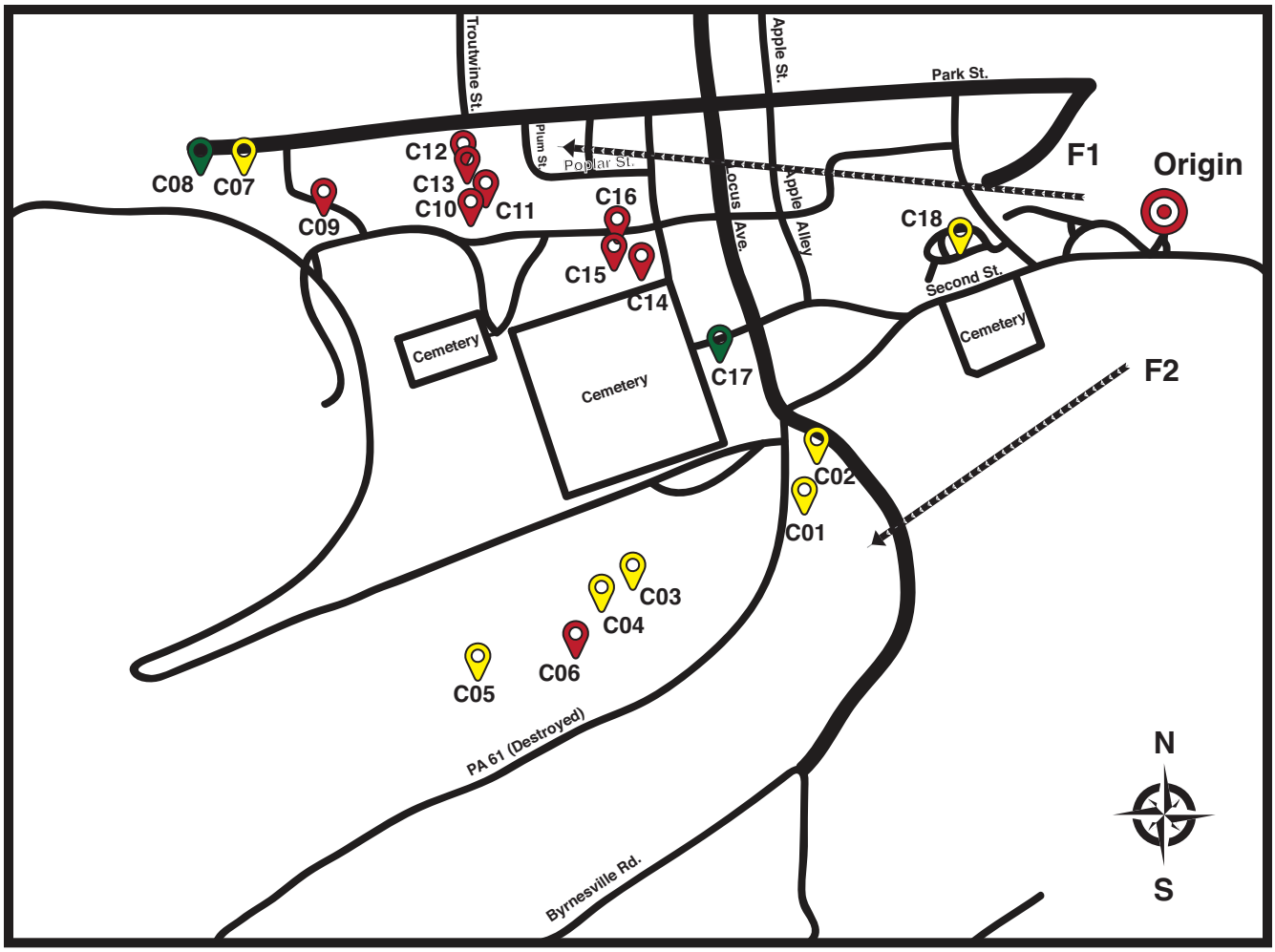
749 **Supporting Table 7**. Parameters and fits of neutral models, implemented as per Burns et al.
750 2015.

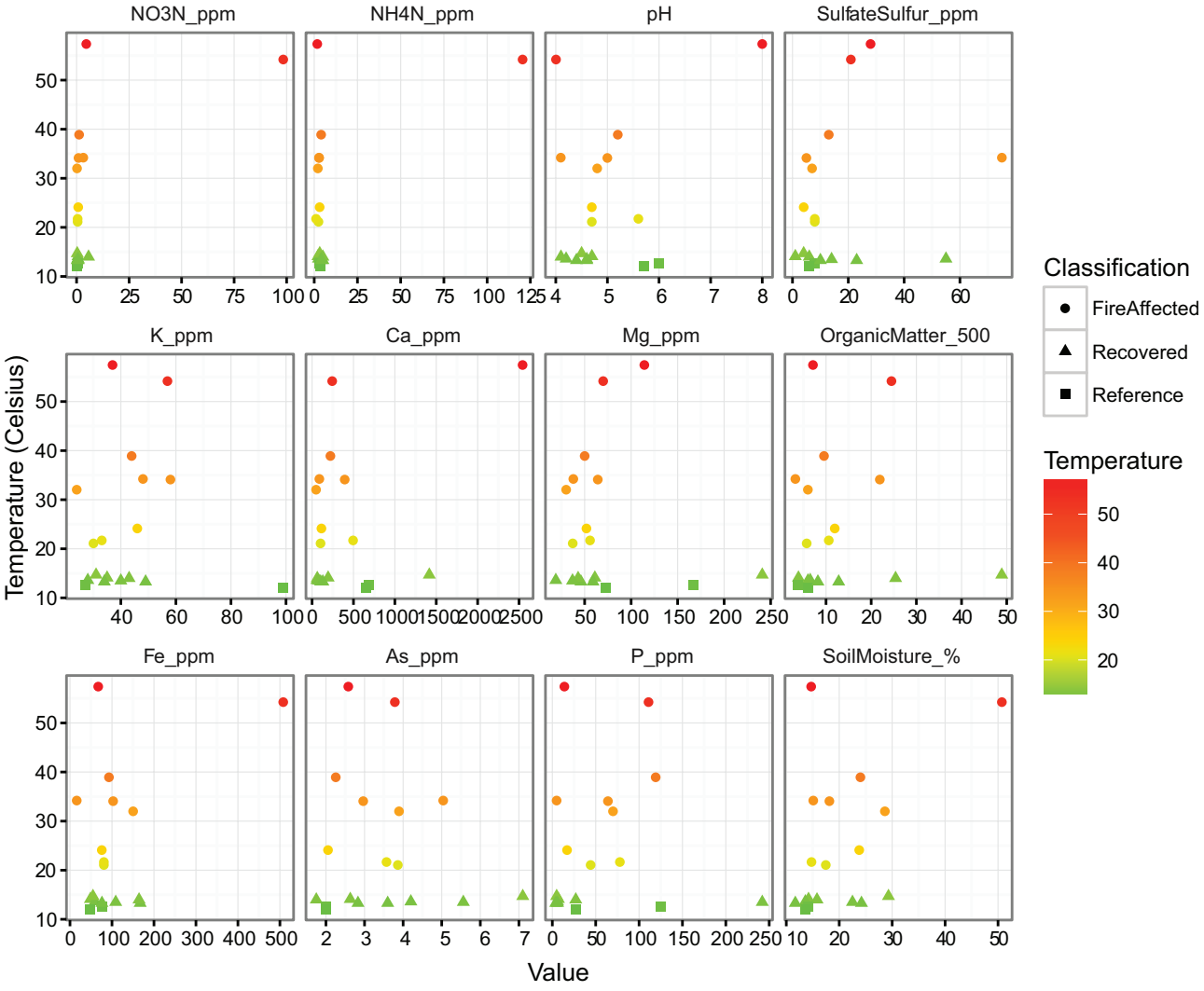
751 **Supporting Table 8**. Welch's t-tests comparing the mean relative abundances of phyla across
752 fire-affected and recovered soils. Bold values are significant at $p < 0.05$.

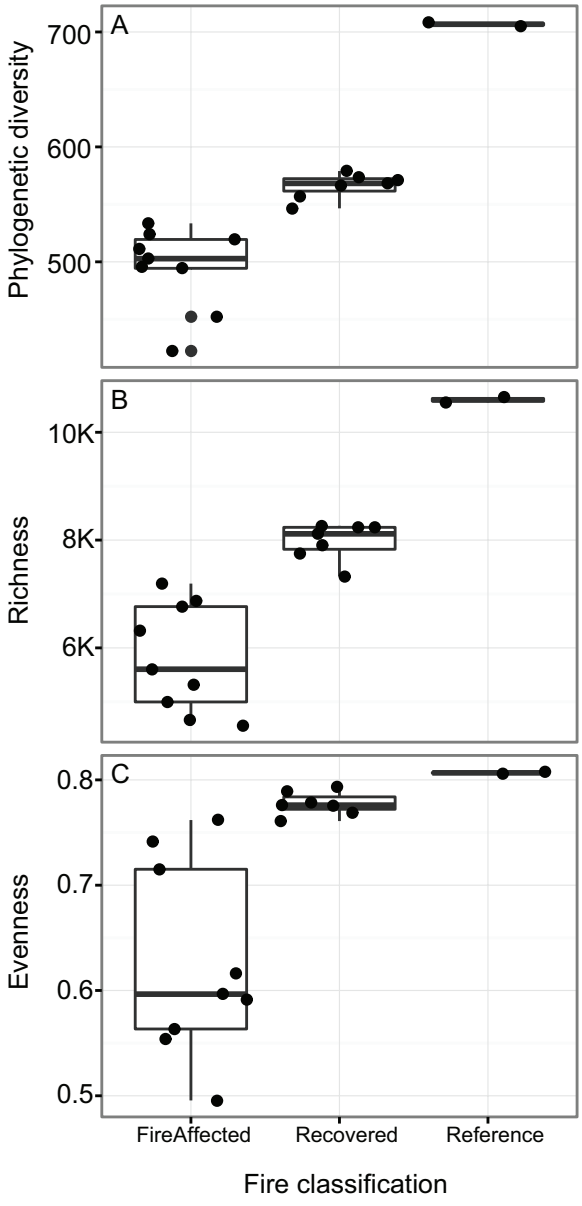
753

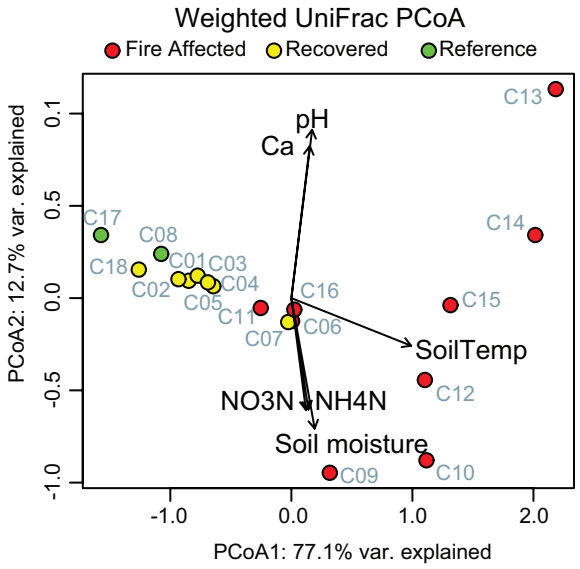
754

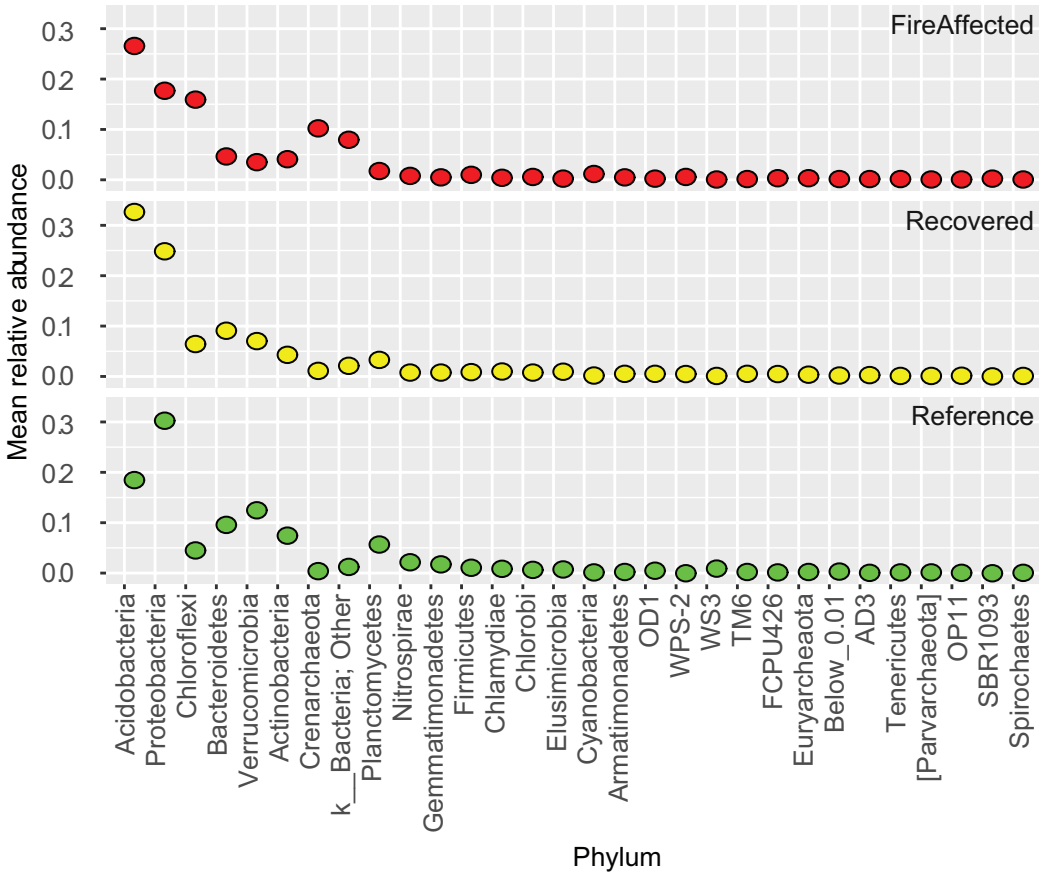
PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2446v1 | CC BY 4.0 Open Access | rec: 13 Sep 2016, publ: 13 Sep 2016

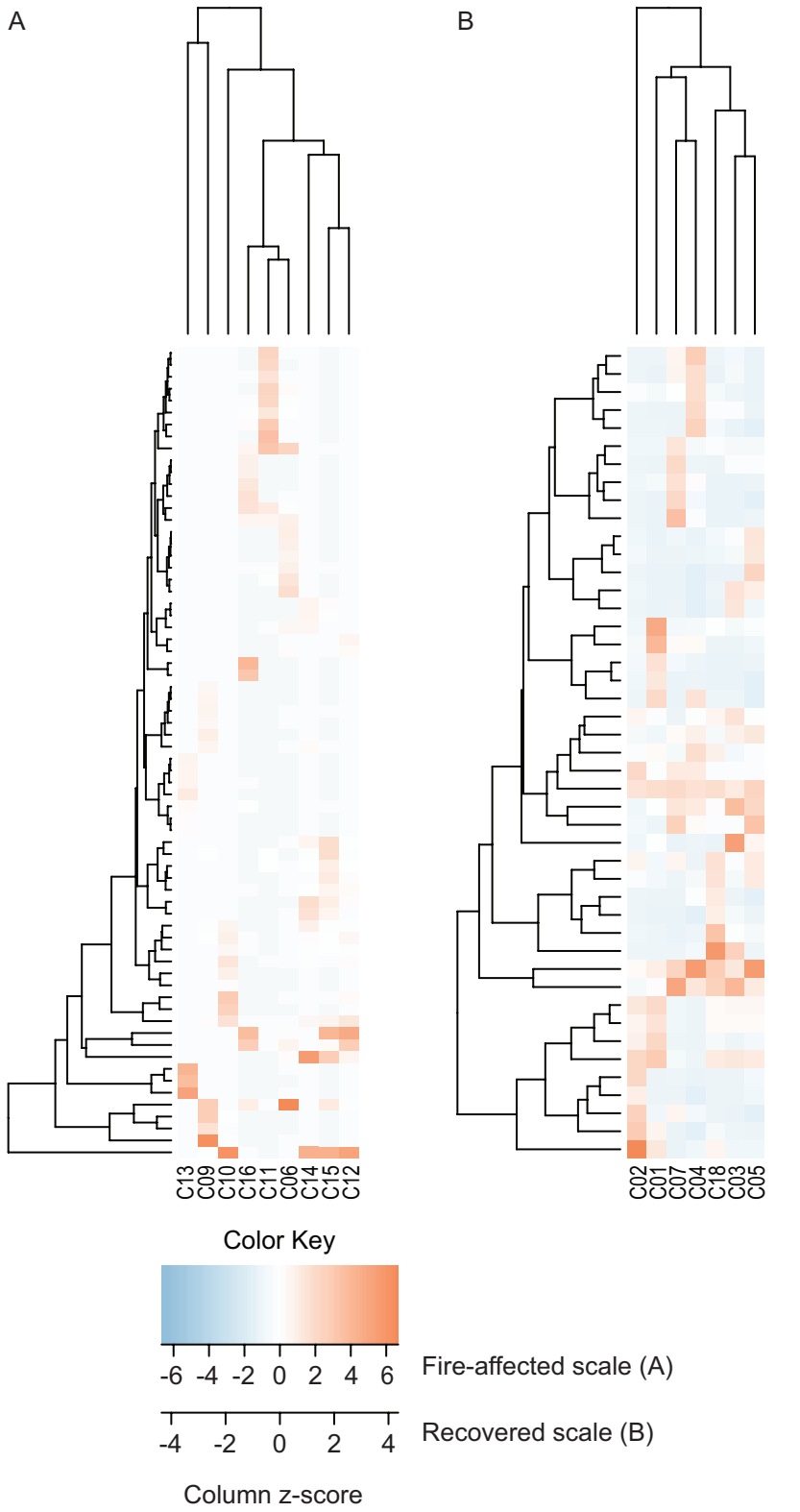



NO3N_ppm
NH4N_ppm
pH
SulfateSulfur_ppm
K_ppm
Ca_ppm
Mg_ppm
OrganicMatter_500
Fe_ppm
As_ppm
P_ppm
SoilMoisture_%
Temperature (Celsius)
Value
Classification
FireAffected
Recovered
Reference
Temperature
50
40
30
20








A
B
C13
C09
C10
C16
C11
C06
C14
C15
C12
C02
C01
C07
C04
C18
C03
C05
Color Key
-6 -4 -2 0 2 4 6
Fire-affected scale (A)
-4 -2 0 2 4
Recovered scale (B)
Column z-score