Title:

Divergent 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, Centralia, resilience, deterministic, stochastic, niche, neutral

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, yet beta-null modeling suggested that communities at temperature extremes follow niche-based processes rather than null. We hypothesize that priority effects from responsive seed bank transitions may be key in explaining the multiple equilibria observed among communities at extreme temperatures. 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

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2446v2 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

longevity, press disturbances have the capacity to alter ecosystems beyond the possibility of recovery (e.g., Thrush *et al.*, 2009).

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; Vellend *et al.*, 2014; Tucker *et al.*, 2016; Ferrenberg *et al.*, 2013), 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). Together, 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.

Because diversification processes are relatively more important at evolutionary scales, Vellend et al. 2014 focused on the remaining processes of ecological selection, drift, and dispersal. They asserted that selection processes are deterministic, that drift processes are stochastic, and that dispersal processes can be either or both, depending on the situation (Vellend *et al.*, 2014). Tucker and colleagues provided clarity to the distinction between

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2446v2 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

deterministic/stochastic and niche/neutral processes, which are often used interchangeably. Niche/neutral refers to the ecological differentiation and equivalence of species, while deterministic/stochastic refers to non-probabilistic or probabilistic outcomes (Tucker *et al.*, 2016). Thus, neutrality concerns ecological equivalence of species, while stochasticity concerns demographic variability in birth, death, and dispersal.

We aimed to understand the responses of soil microbial communities to an anthropogenic press disturbance, and to apply the Vellend, 2010, Nemergut *et al.*, 2013, and Tucker *et al.*, 2016 conceptual frameworks 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 slowly (3-7 m/yr Elick, 2011) through underground coal seams. Depending on the depth of the coal bed, it burns at an estimated 46-69 m below the surface (Nolter and Vice, 2004; Elick, 2011). Heat, steam and combustion products vent upward from the fire through the overlying soils. The surface soil temperatures can exceed 80°C, scarring the landscape with dead vegetation that reveals the fire's subsurface trajectory. As steam and gasses pass through the overlying rock and soil, soil temperatures increase while soil chemical composition is altered by both spontaneous and microbial-mediated chemical reactions (Janzen and Tobin-Janzen, 2008). As the fire expands into new areas, it also retreats from some affected sites, which then recover to ambient temperatures (Elick, 2011; Nolter and Vice, 2004). Thus, the “end” of the disturbance can be delineated by temperature recovery. In this way, a chronosequence of fire-affected Centralia soils provides a space-for-time proxy of disturbance response and recovery.

Our research objectives were to understand the diversity and spatio-temporal dynamics of the surface soil bacterial and archaeal communities that have been impacted historically or are currently influenced by the ongoing subterranean coal mine fire in Centralia. We used a definition of disturbance response to include changes in member relative abundances as well as in composition. Previous work using terminal restriction fragment length polymorphism analysis showed that microbial diversity decreased at hotter sites, and that compositional changes were correlated with soil ammonium and nitrate concentrations (Tobin-Janzen *et al.*, 2005). We move

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2446v2 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

forward from this work to use high throughput sequencing of soil community 16S rRNA genes to quantify the community dynamics along a chronosequence of fire response and recovery. We specifically investigated the community assembly processes of selection, dispersal, and drift.

Materials and Methods

Study site, soil sampling, soil biogeochemistry and microbial community DNA extraction

We undertook fieldwork in Centralia (GPS: 46°46”24’N, 122°50”36W) on 5-6 October 2014. We collected surface soils to capture the expected maximum changes along a chronosequence of fire recovery (**Supporting Figure 1**). We sampled two fire fronts along gradients of historical fire activity. Fronts are trajectories of fire spread from the 1962 ignition site outward along near-surface coal seams (Elick, 2011). These fronts include surface soils that were previously hot and have cooled, as well as soils that are currently warmed by the ongoing fire. We collected soil from two unaffected, proximate sites as references, seven recovered sites along the gradient, and nine fire-affected sites (18 total soils), and these collections were distributed across both fire fronts. Soil samples were collected from the top 20 cm of surface soil (core diameter 5.1 cm), and were sieved through 4 mm stainless steel mesh. We collected cores only at bare surface soil locations (no vegetation) to minimize the influence of local vegetation and to maximize comparability between soils, as the thermal surface soils generally lacked vegetation. Collected soils were stored on ice up to 72 hr during transport to the laboratory, then stored at -80^{o}C pending further processing. The physico-chemical characteristics of each soil sample (percent moisture, organic matter (500°C), 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^{o}C for 2 days. Fire history was estimated as years since the surface soil was first hot from the fire, at each sampling location. Fire history observations were measured using either winter snow cover, aerial vegetation photography, or thermal infrared imagery, as collated and reported by Elick, 2011(Figure 3 therein). 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 DNA amount was standardized for sequencing to 1,000 ng/sample.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2446v2 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

156
157 *Soil cell counts*
158 Direct bacterial and archaeal cell counts were conducted on frozen soil samples based
159 on a protocol to separate cells from soil reported in (Portillo *et al.*, 2013). To dissociate the
160 microbial cells from soil particles, 10 g of soil was mixed with 100 mL of phosphate buffered
161 saline containing 0.5% Tween-20 (PBST). Soil samples were homogenized in a Waring blender
162 three times for 1 min each, followed by a 5 min incubation on ice. Slurries were centrifuged at
163 1000 x g for 15 min to concentrate soil particulates. Supernatants were set aside and stored at
164 4^{o}C, and the remaining soil pellets were re-suspended in 100 mL of fresh PBST and blended for
165 an additional 1 min. The soil slurry was then transferred to sterile 250 mL centrifuge bottles and
166 the blender was washed with an additional 25 mL of sterile PBST and added to the slurry before
167 centrifugation at 1000 x g for 15 min. All resulting supernatants for each site were combined,
168 then centrifuged at 10,000 x g for 30 min to pellet cells. Supernatants were discarded, and cell
169 pellets were re-suspended in 10 mL of sterile Milli-q water and 400 mL of 37% formaldehyde to
170 fix cells. 1 mL of cell suspension was then carefully layered over 500 µL of sterile Nycodenz
171 solution (0.8 g/mL in 0.85% NaCl), then centrifuged at 10,000 x g for 40 min. The upper layer
172 was then collected and cells were pelleted by centrifugation at 20,000 x g for 15 min, then
173 resuspended in 1 mL of sterile 0.85% NaCl. To dissociate remaining soil clumps, cell
174 suspensions were sonicated for 10 s in a sonicating water bath.
175 Cell suspensions were stained with DTAF ((5-(4,6-Dichlorotriazinyl) Aminofluorescein))
176 according to (Robertson *et al.*, 1999). DTAF-stained smears were visualized on a Nikon Eclipse
177 e800 microscope (Tokyo, Japan) equipped with a Photometrics Coolsnap Myo camera (Tuscon,
178 AZ, USA), and images were collected using Micro-Manager software (Edelstein *et al.*, 2014). Fiji
179 image analysis software was used to adjust background, thresholding, and to conduct particle
180 counts from images (Schindelin *et al.*, 2012). Briefly, background correction was completed
181 using an automated rolling ball subtraction with a 35-pixel radius, followed by automatic local
182 thresholding using the Bernsen method with a 12-pixel radius to convert greyscale images to
183 binary. Watershed segmentation was conducted to separate touching nuclei, then particles were
184 counted using the ImageJ *"*Analyze Particles*"* function, excluding anything smaller than 0.1
185 micron (Schneider *et al.*, 2012).
186
187 *Quantitative PCR*
188 We performed quantitative PCR (qPCR) using bacterial and archaeal 16S rRNA gene
189 universal primer sets (**Supporting Table 1;** Caporaso *et al.*, 2012). The reaction mixtures

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2446v2 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

consisted of 10 µL SYBR qPCR Master mix (Quanta Bioscience, Gaithersburg, MD, USA), 0.4 µL each of the forward and the reverse primers (0.4 pM), 2 µL of template DNA, and sterilized deionized water to adjust the final volume of 20 µL. The thermal profile was as follows: initial denaturation at 95°C for 10 s, followed by 40 cycles of denaturation at 95°C for 10 s, annealing at 50°C for 15 s, and extension at 72°C for 40 s. A final dissociation protocol (58°C to 94.5°C, increment 0.5°C for 10 s) was performed to ensure the absence of nonspecific amplicons. The reactions were conducted using the Bio-Rad iQ5 real time detection system (Bio-Rad, Hercules, CA, USA). Please see the supporting materials for more details as to the qPCR methods.

16S rRNA amplicon sequencing

For each of the 54 DNA samples (18 soils, each with three replicate DNA extractions) and mock community DNA, paired-end sequencing (150 base pair) was performed on the bacterial and archaeal 16S rRNA gene V4 hypervariable region using the Illumina MiSeq platform (Illumina, CA, USA; **Supporting Table 1**; Caporaso *et al.*, 2012). All of the sequencing procedures, including the construction of Illumina sequencing library using the Illumina TruSeq Nano DNA Library Preparation Kit, emulsion PCR, and MiSeq sequencing were performed by the Michigan State University Genomics Core sequencing facility (East Lansing, MI, USA) following their standard protocols. The Genomics Core provided standard Illumina quality control, including base calling by Illumina Real Time Analysis v1.18.61, demultiplexing, adaptor and barcode removal, and RTA conversion to FastQ format by Illumina Bcl2Fastq v1.8.4. Raw sequences were submitted to the GenBank SRA Accession SRP082686.

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

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2446v2 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

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. All sequences are available in NCBI's Short Read Archive (https://www.ncbi.nlm.nih.gov/sra/SRP082686).

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). We selected open-reference OTU picking because it allowed us to retain all high-quality sequences, even if they did not match to the reference database. In addition, we expected novel diversity in Centralia, and it was likely that many Centralia sequences would not hit to reference databases. Furthermore, we wanted to create consistent OTU definitions that could be tractable across this study and future work. In the open-reference OTU picking workflow, reference-based OTU clustering first was conducted using the usearch_global command to cluster sequences with 97% identity to the greengenes database (v 13.8, http://greengenes.secondgenome.com/downloads). Second, de novo OTU picking was performed for any sequences that did not hit the greengenes reference; the usearch command cluster_otus 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.

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

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2446v2 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

(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 2;** and alpha diversity among technical replicates provided in **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.*, 2007;Lozupone *et al.*, 2011) were calculated within QIIME. Non-normalized UniFrac distances can fall outside of 0 and 1, while normalized UniFrac distances are bound to 0 to 1; Lozupone *et al.*, 2007 reported no differences in overarching patterns in beta diversity between the nonnormalized and normalized UniFrac (Lozupone *et al.*, 2007), and we have found that this holds for our dataset (**Supporting Table 3**). 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 ("sncm.fit_function.R"); custom R scripts for beta-null model fitting written by Tucker *et al.*, 2016, Appendix 2 therein) modified by our group to include weighted UniFrac beta-null modeling; and ggplot and ggplots2 for plotting (Wickham, 2009). Our R script is available on GitHub ("R_analysis" repository in https://github.com/ShadeLab/PAPER_LeeSorensen_ISMEJ_2017)

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2446v2 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

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 (**Supporting Figure 3)**. 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. 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). 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 were within one order of magnitude of 16S rRNA copies per dry mass of soil with fire-affected soils having the highest copy numbers and recovered soils having the lowest, but there were no statistical differences among groups (**Supporting Figure 4A**, Student's t-test all pairwise $p \geq 0.09$). Total number of cells per dry mass of all soil ranged from 10^5 to 10^7 cells per gram of dry soil, but cell counts across fire classifications also were not statistically distinct (**Supporting Figure 4B**, Student's t-test all pairwise $p \geq 0.09$**)**. Together, these data indicate overall community size is relatively 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 (community size) among soils.

Sequencing efforts were near-exhaustive for these soils, as assessed by a clear asymptote achieved with rarefaction (**Supporting Figure 5**). A summary of sequencing efforts, as well as a discussion of reference-based and *de novo* OTU taxonomic assignments for fire-affected and recovered soils, are provided in supporting materials.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2446v2 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

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 lowest and most variable for fire-affected soils, and highest for reference sites (**Figure 1;** 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 ongoing surface contamination by coal combustion products, such as inorganic gases containing arsenic, selenium, ammonium, sulfur, and hydrogen sulfide, and organic toxins like polycyclic aromatic hydrocarbons (Janzen and Tobin-Janzen, 2008). Elements within inorganic gases mineralize and deposit around active vents (Janzen and Tobin-Janzen, 2008). Some coal combustion products, like volatile sulfur and nitrogen compounds, may enrich for microorganisms capable of using them, while other combustion products, like organic toxins, may decrease microbial community size or diversity (Janzen and Tobin-Janzen, 2008).

We used 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 resemblances revealed the same overarching patterns (all pairwise Mantel and PROTEST $p < 0.001$, **Supporting Table 3**), demonstrating that these patterns were very robust. However, weighted UniFrac distance provided the highest explanatory value (**Supporting Table 3**), suggesting that changes in both

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2446v2 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

356 phylogenetic breadth and the relative abundances of taxa are important for interpreting
357 community responses. As compared to recovered and reference sites, fire-affected soils were
358 distinct (PERMANOVA pseudo F = 16.10, R^2 = 0.50 and p = 0.001 on 1000 permutations) and
359 more variable in their community structure (difference in median dispersions = 0.53, p = 0.008;
360 **Figure 2**). Differences in surface soil temperature had most explanatory value on Axis 1 (77.1%
361 variance explained by Axis 1, temperature Axis 1 correlation = 0.97, p = 0.001, **Supporting**
362 **Table 4**), with nitrate and iron contributing; calcium and pH (and, to a lesser extent, soil
363 moisture) explained variation on Axis 2 (12.7% variance explained by Axis 2, **Supporting Table**
364 **4**). Notably, soil fire history (estimated years since the local soil surface was first measured hot
365 as reported by Elick, 2011) was not correlated to community dynamics (**Supporting Table 4**).

366 Fire-affected soils were more variable in their community structure across soils,
367 especially in soils at the most extreme temperatures observed (sites C13, C10 which were
368 >50°C at the time of sampling and were at the opposite ends of PCoA2). In contrast, recovered
369 soils were less variable, even though they spanned decades of difference in their years of peak
370 fire activity (the earliest impacted soils that we sampled were last recorded to be hot in 1980;
371 Elick, 2011). Also, recovered soils were very similar in community structure to reference soils.
372 These patterns show that Centralia soils achieve divergent community structures over the
373 transition from ambient to extreme conditions, but then generally converge towards a consistent
374 community structure after the fire subsides. These results also show resilience of soil
375 communities impacted by an extreme press disturbance, with recovery occurring within 10-20
376 years after the stressor subsided.

377 We observed a temperature "threshold" effect among fire-affected soils, and soils with
378 temperatures between 21 and 24.5°C (sites C06, C11, and C16) separated cleanly from soils
379 with temperatures greater than 30°C (**Figure 2**). To better understand the divergence in
380 community structure among fire-affected soils, we performed a PCoA with these communities
381 (**Supporting Figure 6A, Supporting Table 5**), and also a constrained analysis to ask what
382 variability remained after removing the influence of temperature (**Supporting Figure 6B,**
383 **Supporting Table 6**). Even after removing the influence of temperature, three discrete subsets
384 of fire-affected communities separated from each other along both axes, with C13 remaining as
385 an outlying point. C13 had very different calcium and pH than the other soils, and both of these
386 factors had high value in discriminating C13 from the other fire-affected soils (p = 0.092 and
387 0.014 respectively). There were no other measured abiotic factors that explained the divergence
388 among the fire-affected soils. In addition, the constrained axes had high explanatory value

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2446v2 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

(**Supporting Figure 6B,** combined axes 1 and 2 = 90.0% var. explained), suggesting that, given the measured conditions, there are additional processes beyond abiotic selection that explain the differences in these subsets.

We observed broad phylum-level changes in response to the fire (**Figure 3**, **Supporting Table 8**). Not all OTUs affiliated with particular phyla had identical responses; however, our analysis of phylum-level responses points to some general trends. In particular, fire-affected soils were enriched for members of Chloroflexi, Crenarcheaota and many lineages of unidentified Bacteria. As compared to the fire-affected soils, recovered soils also were enriched for Parvarchaeota, Bacteroidetes, Elusimicrobia, Gemmatimonadetes, Planctomycetes, Spirochaetes, TM6, and Verrucomicrobia suggesting that members affiliated with this these phyla are able to persist after the fire subsides. Acidobacteria also had an increase in recovered soils (but less significant, p = 0.10), presumably because of the decrease in soil pH observed post-fire (**Supporting Figure 3, pH panel: row 1, column 3**). Reference soils had higher representation of Proteobacteria and Verrucomicrobia, which suggests that members of these phyla may be sensitive to the fire.

Dispersal and drift

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 used two complementary approaches. First, 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 7, Supporting**

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2446v2 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

Table 7). Furthermore, we found a lower influence of dispersal (lower value of *m)* in the fire-affected sites (**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.

Next, we asked how observed differences in beta diversity deviate from null expectations. We used abundance-based beta-null approaches to distinguish niche and null processes according to Tucker *et al.*, 2016, and we extended their approach to also consider community differences in phylogenetic breadth by applying it to weighted UniFrac distances. In this comparative approach, deviations to and from a permuted null expectation (neutral) are used to interpret the relative influences of neutral and niche processes, respectively. All Centralia communities deviated from neutral, with reference and recovered soils falling closer to neutral expectations than fire-affected soils (**Figure 4A**). Fire-affected soils had statistically higher beta-null deviations than recovered soils (both $p < 0.05$ for Bray-Curtis and weighted UniFrac). In the fire-affected soils, there was a consistent increase in niche processes with increasing soil temperature, and the hottest sites deviated furthest from the neutral expectation (**Figure 4B**). Accounting for phylogenetic breadth (using weighted UniFrac distance, **Figure 4B** suggested relatively less deviation from neutral than accounting for abundance alone (using Bray-Curtis dissimilarity, **Figure 4B**), but both resemblances had similar trends (Pearson's $R= 0.71$, $p = 0.001$) and produced identical statistical outcomes. These abundance null deviation results agree with the Sloan neutral model because they suggest that unmeasured niche processes structure soil communities at temperature extremes.

Understanding community divergences at temperature extremes

To dig deeper into the differences in the three subsets of fire-affected soil (**Supporting Figure 6**) that were not well explained by measured abiotic selection, local dispersal, or drift as assessed by the Sloan neutral model of community assembly and beta-null modeling, 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. 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 soils (out of a possible 90, if each of the nine fire-affected soil harbored mutually exclusive membership across their top

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2446v2 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

453 10). These prevalent fire-affected OTUs spanned fourteen phyla or Proteobacteria classes,
454 included 30 *de novo* OTUs, and included seven taxa of unidentified Bacteria and two taxa of
455 unidentified Proteobacteria. Acidobacteria OTUs were detected among the top 10 for all fire-
456 affected soils, and eight of nine fire-affected soils included Chloroflexi among the top 10 OTUs.
457 In comparison, recovered soils included ten phyla or Proteobacteria classes among their
458 collective top 10, had no unidentified Bacteria or Proteobacteria, and included four *de novo*
459 OTUs. Acidobacteria and Alphaproteobacteria OTUs were among the top 10 for all recovered
460 soils, and six of the seven recovered soils also included Deltaproteobacteria. Together, these
461 results show that fire-affected soils were more divergent and diverse in their prevalent
462 membership than recovered soils.

463 An analysis of occurrence patterns of prevalent OTUs also showed greater divergence
464 among fire-affected soils than recovered (**Figure 5**), and further supported the distinction among
465 the subsets of fire-affected soils revealed by the constrained ordination (**Supporting Figure**
466 **6B**). Fire-affected soils had more OTUs within their collective most prevalent taxa, and were
467 more heterogeneous as shown by the wider range represented by the color scale and the more
468 divergent sample and OTU clustering. In fact, taxa that were among the top 10 in one fire-
469 affected soil were likely to be among the rare biosphere in another fire-affected soil, exhibiting
470 stark contrast in their abundances within these soils. However, most of the top 10 prevalent
471 OTUs were detected within every fire-affected soil (**Table 1, Figure 5**), suggesting that changes
472 in taxa relative abundances, rather than turnover in membership, were driving these patterns.

473 This dominance analysis helps to explain the lower fit of the neutral model, and the
474 relatively higher influence of niche processes with beta-null modeling, to fire-affected
475 communities. Outliers to the neutral model that were below detection (taxa that were present in
476 fewer sites than predicted given their relative abundance in the metacommunity) included these
477 many lineages that were prevalent in few fire-affected soils. Taxa that fall below their neutral
478 model prediction have been proposed to be “selected against” or particularly dispersal limited
479 (Burns 2015). However, in the Centralia extreme environment, we suggest these are taxa that
480 were most successful locally given the thermal disturbance.

481

482 *Community assembly processes given a press disturbance*

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

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2446v2 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

485 explanatory value for Centralia soil community dynamics. These communities first were
486 constrained by environmental filters imposed by the press disturbance, such as thermal
487 temperatures in fire-affected soils and low pH in recovered soils. The fire acts as a strong
488 environmental filter, resulting in decreased diversity and a very different phylogenetic
489 representation among the surviving lineages in fire-affected soils. These environmental filters,
490 such as changes in pH, likely alter the functions of the community as well as its composition.
491 However, even after removing the influence of temperature on fire-affected communities, the
492 communities fell into three distinct subsets that could not be explained by the physico-chemical
493 characteristics measured. Furthermore, neutral modeling, beta-null modeling and lack of spatial
494 autocorrelation suggests that these particular assessments for drift and dispersal processes
495 offer minimal explanation for fire-affected sites. Given the low explanatory value of unweighted
496 resemblances in describing patterns of comparative diversity (**Supporting Table 3**), and the
497 observation that many of the prevalent taxa detected in some fire-affected soils were rare in
498 other fire-affected soils (**Figure 5A**), we can also attribute these patterns to changes in the
499 relative abundances of taxa within a locality, rather than to changes in taxa turnover (differing
500 memberships). Thus, given that neither assessed selection, dispersal, nor drift processes, nor
501 their combination can provide a complete explanation for the divergence of fire-affected
502 communities, the questions remain: why are fire-affected soils so divergent from each other,
503 and how do they eventually manage to recover to the same post-disturbance community
504 structure?

505 One hypothesis is that the remaining variability in community structure of fire-affected
506 sites may be attributed to priority effects initiated from different local transitions between the
507 dormant seed bank and the active community. The proportion of dormant cells in soils is
508 estimated to be as high as 80% (Lennon and Jones, 2011), and the importance of dormancy for
509 microbial community assembly processes has been discussed at length (Nemergut *et al.*, 2013).
510 Specific to the Centralia coal mine fire disturbance, thermophiles are prime examples of
511 microbial seed bank members that often have been found in environments that are improbable
512 to permit their growth (e.g., Hubert *et al.*, 2009; McBee and McBee, 1956; Portillo *et al.*, 2012).

513 There are two aspects of seed banks that could help to explain Centralia community
514 divergences at temperature extremes: membership and dynamics. If each soil harbored a
515 different seed bank membership, different thermophilic taxa could become active and prevalent
516 in each fire-affected soil, and would manifest as drift influences. This scenario is not well-
517 supported by our data because we detect the dominant members of each fire-affected soil in the

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2446v2 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

other fire-affected soils, albeit in lower abundances. Alternatively, awakenings from the microbial seed bank (Buerger *et al.*, 2012) could result in priority effects at temperature extremes, in which the first fit 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 that are explained by niche processes. In addition, unknown nuances in local abiotic conditions at fire onset could also set communities onto parallel trajectories and result in multiple equilibria during the press, which would also be explained by niche processes. Our data indirectly support either of these last two scenarios, as the three separate clusters of fire-affected communities suggest multiple equilibria (**Supporting Figure 6B**). It could be that the most similar fire-affected communities began either from the same (or functionally equivalent) waking pioneer taxon, or from the same abiotic conditions (that are similar beyond reaching thermal temperatures), or from some combination of both, which initiated distinct trajectories towards each equilibrium.

Diversification is a fourth community assembly process discussed by Vellend, 2010 and Nemergut *et al.*, 2013. At ecological time scales, diversification was suggested by Vellend *et al.*, 2014 to have relatively lower influence than the other community assembly processes. 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 at distinct but parallel trajectories of diversification within a locality. However, we cannot reject the hypothesis that diversification processes also contribute to divergences in community structure at temperature extremes.

Conceptual model

Extending the conceptual models of Ferrenberg *et al.*, 2013 and Dini-Andreote *et al.*, 2015, we present a hypothesis of the assembly processes shaping communities before, during, and after an extreme press disturbance. Our model is based on our chronosequence trajectory for beta-null data presented in **Figure 4B**, and includes a phase encompassing the press disturbance, which extends beyond the representation of a pulse disturbance as a single time point as typical in previous conceptual models. Our model also incorporates a hypothesis of multiple transient equilibria within the press disturbance phase. We apply the advice of (Tucker

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2446v2 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

550 *et al.*, 2016) to not use the direction of the change from neutral (positive or negative) to infer
551 specific ecological processes.

552 We hypothesize that weak variable selection drives stability in heterogeneous Centralia
553 soil communities before the fire (reference sites in **Figure 4**; phase 1 in **Figure 6**). This is
554 additionally supported by the literature demonstrating generally high heterogeneity and diversity
555 in mature soil microbial communities (*e.g.,* O'Brien *et al.*, 2016). Next, strong environmental
556 filtering from thermal temperatures (homogeneous selection, phase 2) decreases community
557 diversity at the onset of the press disturbance. The lower diversity and prolonged disturbance
558 conditions permit priority effects initiated by taxa fit in the thermal environment (e.g.,
559 thermophiles waking from the seedbank), which set communities onto distinct deterministic
560 trajectories with multiple equilibria during the fire (phase 2). Alternatively, the distinct trajectories
561 and multiple equilibria could have been initiated by unmeasured nuances in abiotic conditions at
562 thermal onset. Finally, weak environmental filtering from increased soil acidity relaxes
563 communities back towards neutral in post-fire conditions (homogeneous selection, phase 3).

564 Regardless of the interim dynamics that resulted in community divergence to the
565 stressor, Centralia soils eventually recovered to a community structure very similar to reference
566 soils, and these community structures were explained by the ultimate post-fire soil environment.
567 Our results show that Centralia soil communities, though sensitive to this extreme, complex, and
568 arguably unnatural stressor, had near-complete return to pre-disturbance conditions, and were
569 resilient within ten to twenty years after the stressor subsides. We have no reason to suspect
570 that temperate soils in Centralia are exceptional as compared to other soils. Thus, these results
571 suggest that soils may have an intrinsic capacity for robustness to varied disturbances, even to
572 those disturbances considered to be "extreme", compounded, or incongruent with natural
573 conditions. Understanding the precise functional underpinnings of soil microbial community
574 resilience, including the roles of seed banks in determining that resilience, is a next important
575 step in predicting and, potentially, managing, microbial community responses to disturbances.

576

577 **Acknowledgements**

578

579 This work was supported by Michigan State University, with computing resources provided by
580 the Michigan State Institute for Cyber-Enabled Research. JWS acknowledges the Dr. C. A.

Reddy and Sasikala Reddy Award from the Michigan State Department of Microbiology and Molecular Genetics. We thank Trevor Grady for his graphic design work for Supporting Figure 1.

Conflict of Interest Statement

The authors declare no conflict of interest.

References

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

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

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

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

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

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

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

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

Desai C, Pathak H, Madamwar D. (2010). Advances in molecular and '-omics' technologies to gauge microbial communities and bioremediation at xenobiotic/anthropogen contaminated sites.

612 *Bioresour Technol* **101**: 1558–1569.

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

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

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

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

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

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

626 Ferrenberg S, O'Neill SP, Knelman JE, Todd B, Duggan S, Bradley D, *et al.* (2013). Changes in
627 assembly processes in soil bacterial communities following a wildfire disturbance. *Isme J* **7**:
628 1102–1111.

629 Fuentes S, Barra B, Gregory Caporaso J, Seeger M. (2015). From rare to dominant: A fine-
630 tuned soil bacterial bloom during petroleum hydrocarbon bioremediation. *Appl Environ Microbiol*
631 **82**: 888–896.

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

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

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

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

640 Li Y, Wen H, Chen L, Yin T. (2014). Succession of bacterial community structure and diversity in
641 soil along a chronosequence of reclamation and re-vegetation on coal mine spoils in China.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2446v2 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

PLoS One **9**. e-pub ahead of print, doi: 10.1371/journal.pone.0115024.

Lozupone CA, Hamady M, Kelley ST, Knight R. (2007). Quantitative and qualitative diversity measures lead to different insights into factors that structure microbial communities. *Appl Environ Microbiol* **73**: 1576–1585.

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

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

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

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

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

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

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

O'Brien SL, Gibbons SM, Owens SM, Hampton-Marcell J, Johnston ER, Jastrow JD, *et al.* (2016). Spatial scale drives patterns in soil bacterial diversity. *Environ Microbiol* **18**: 2039–2051.

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

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

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

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

Quadros PD de, Zhalnina K, Davis-Richardson AG, Drew JC, Menezes FB, Camargo FA d. O,

et al. (2016). Coal mining practices reduce the microbial biomass, richness and diversity of soil. *Appl Soil Ecol* **98**: 195–203.

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

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

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

Tucker CM, Shoemaker LG, Davies KF, Nemergut DR, Melbourne BA. (2016). Differentiating between niche and neutral assembly in metacommunities using null models of β-diversity. *Oikos* **125**: 778–789.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2446v2 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

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

Vellend M, Srivastava DS, Anderson KM, Brown CD, Jankowski JE, Kleynhans EJ, *et al.* (2014). Assessing the relative importance of neutral stochasticity in ecological communities. *Oikos* **123**: 1420–1430.

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.

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

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2446v2 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

Figures

Figure 1. Within-sample (alpha) diversity of fire-affected, recovered, and reference soils in Centralia for bacterial and archaeal community (**A**) Faith's phylogenetic diversity (all $p < 0.001$); (**B**) richness (total no. observed OTUs clustered at 97% sequence identity, all $p < 0.001$); and (**C**) Pielou's evenness (all p not significant).

Figure 2. Principal coordinate analysis (PCoA) based on weighted UniFrac distances of phylogenetic bacterial and archaeal community structure. Colors show the fire classification of the soil as fire-affected (red), recovered (yellow), or reference (green). The strength of statistically significant ($p < 0.10$) explanatory variables are shown with solid arrows.

Figure 3. Phylum-level responses to the Centralia coal mine fire. Mean relative abundance of phyla summarized within soil fire classifications (fire-affected, recovered, and reference). Unidentified Bacteria are a combination of OTUs unable to be assigned taxonomy at the phylum level, and are not a monophyletic group. "Phyla Below 0.01" are all OTUs assigned to phyla that collectively comprise less than 0.01 relative abundance in, and also are not a monophyletic group.

Figure 4. The relative changes in niche and neutral processes assessed using deviations from abundance-weighted beta-null models. Color gradient shows the soil temperature, as a proxy for disturbance intensity. (**A**) Abundance null deviations by fire classification. For both Bray-Curtis and weighted Unifrac resemblances, recovered and fire-affected communities had distinct null deviations (both $p < 0.05$); (**B**) Trajectory of beta-null deviations ranked by disturbance intensity from reference to fire-affected to recovered soils. Weighted UniFrac and Bray-Curtis trajectories are correlated ($p = 0.71$, $p = 0.001$).

Figure 5. Relative abundances of the collection of the most prevalent combined "top 10" taxa (rows) observed in (**A**) fire-affected or (**B**) recovered soils (columns) in Centralia. Color gradients indicate taxa relative abundances, with warm colors indicating prevalent taxa and cool colors indicating rare taxa within that soil. Note differences in color scale gradient between (**A**) and (**B**). Column labels are sample IDs, and OTU IDs are provided as row labels. OTU IDs that begin "OTU_dn" indicate that the taxon was clustered *de novo* in the open-reference OTU picking workflow; IDs that are numeric indicate that the taxon was assigned with high identity to a reference in the greengenes database. For reference-based OTUs, the numeric identifier corresponds to its representative sequence in the greengenes database. Top dendrograms

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2446v2 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

745 cluster soils that have similar community structure, and side dendrograms cluster OTUs that
746 have similar occurrence patterns.

747 **Figure 6**. Hypothesized conceptual model of Centralia community assembly following press
748 disturbance. Phase 1 represents the stable soil community pre-fire, and is characterized by
749 weak variable selection from typical soil heterogeneity and high community diversity. Because
750 the disturbance is a press, phase 2 occurs concurrent with the fire, when strong environmental
751 filters (homogenizing selection) imposed by the extreme conditions drive a sharp increase in
752 niche processes away from neutral conditions at the onset of the fire. Within phase 2, multiple
753 equilibria result from priority effects of pioneer taxa that are fit to survive in the extreme press
754 environment. Phase 3 is post-fire, characterized by relatively weak environmental filtering (e.g.,
755 increased in soil acidity) that relaxes communities towards neutral. Complete neutrality was not
756 observed in pre-fire or post-fire soils.

757

758

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2446v2 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

759 **Tables**

760 **Table 1**. Ten most abundant OTUs in fire-affected Centralia soils. OTUs (defined at 97%
761 sequence identity) were assigned to the most resolved taxonomic level possible; there were no
762 taxonomic assignments that could be made to these prevalent OTUs below the family level
763 (RDP Classifier confidence > 0.80).
764

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2446v2 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

765 **Supporting Figures**

766 **Supporting Figure 1.** Soil sampling sites at Centralia mine fire. In total, 18 surface soil samples
767 (5.08 cm x 20 cm PVC core) were collected along two fire fronts in Centralia, on 15/16 October
768 2014. Sampling sites encompass a gradient of historical fire activity (red flags: Fire-affected in
769 2014 (temperature > 21°C); yellow flags: recovered in temperature, post-fire; and green flags:
770 reference soils).

771 **Supporting Figure 2**. PCoA showing the variability among technical replicates. Three replicate
772 DNA extractions, amplifications and sequencing reactions were performed per soil, and these
773 sequences were subsequently pooled into one aggregate set of sequences to achieve deep
774 coverage of the community within each soil. Error bars are standard deviation around the mean
775 weighted UniFrac distance among technical replicates, each subsampled to an even 53,000
776 sequences per replicate.

777 **Supporting Figure 3.** Soil physical and chemical contextual data (x-axis) plotted against
778 temperature (y-axis). Color gradient shows the soil temperature, and symbols show soil fire
779 classification in October 2014 as fire-affected, recovered, or reference.

780 **Supporting Figure 4**. Quantification of (**A**) 16S rRNA copies per gram of dry soil and (**B**) cell
781 counts per gram of dry soil in fire-affected, recovered, and reference soils. 16S rRNA copies
782 were assessed using quantitative PCR, and cell counts were assessed using cell separation
783 from soil, staining and microscope imaging. There were no statistical differences in values
784 across fire classification for either measurement (all pairwise $p > 0.09$ with a student's t-test).

785 **Supporting Figure 5**. Centralia 16S rRNA amplicon sequencing effort assessed by
786 subsampling/rarefaction of (**A**) richness and (**B**) Faith's phylogenetic diversity with increasing
787 total number of sequences.

788 **Supporting Figure 6**. Divergences in fire-affected soils are not well explained by temperature.
789 **(A)** Principal coordinate analysis (PCoA) based on weighted UniFrac distances of phylogenetic
790 bacterial and archaeal community structure in fire-affected soils. The strength of statistically
791 significant ($p < 0.10$) explanatory variables are shown with blue arrows. **(B)** Constrained
792 analysis (CAP) based on weighted UniFrac distances, where the explanatory value of
793 temperature is removed from the analysis to understand the influence of the remaining
794 explanatory variables.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2446v2 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

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

801 **Supporting Figure 8**. Quantitative PCR standard curve for the amount of *E.coli* 16S rRNA
802 gene copies (cloned into plasmids) versus C_T values. The solid line is the regression (R^2 =
803 0.988). The error bars are the standard deviations obtained in three independent experiments.
804

805 **Supporting Tables**

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

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

813 **Supporting Table 3.** (A) Percent variation explained for PCoA axes 1 and 2 for nonnormalized
814 weighted and unweighted UniFrac, normalized weighted UniGrac, Sorensen-dice, and Bray-
815 Curtis distances/dissimilarities. Nonnormalized weighted UniFrac was chosen because it was
816 most informative in explaining the variance along the first two axes. (B) Pairwise resemblance
817 correlations calculated with Mantel and PROTEST. All $p < 0.001$ for all tests.

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

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

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2446v2 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

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

827 **Supporting Table 7**. Parameters and fits of neutral models, implemented as per Burns et al.
828 2015.

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

831

832

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2446v2 | CC BY 4.0 Open Access | rec: 21 Dec 2016, publ: 21 Dec 2016

Figure 1

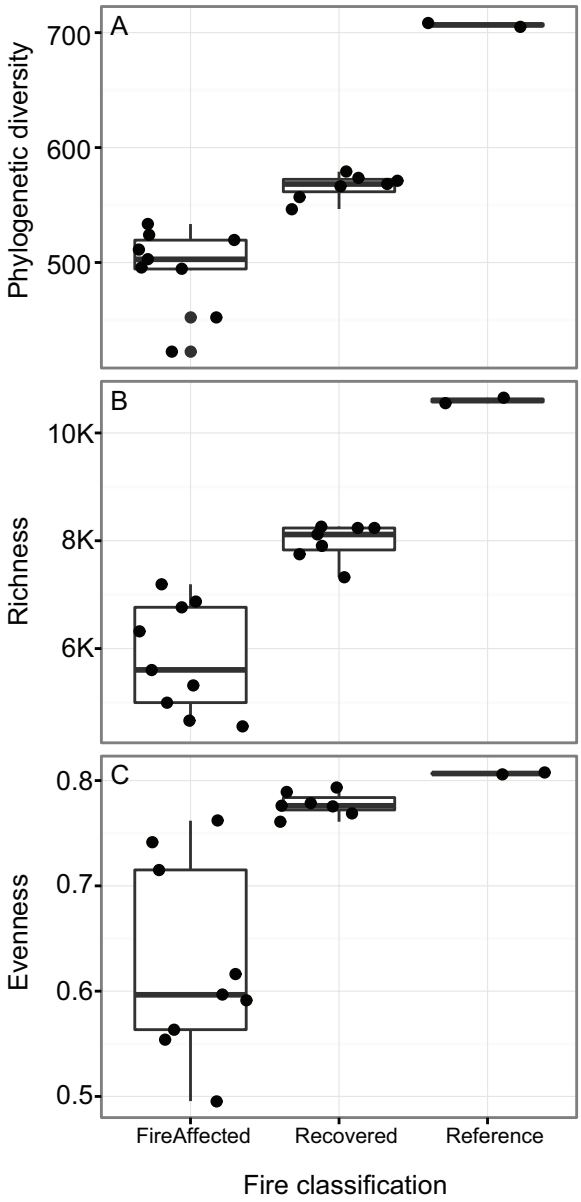


Figure 2

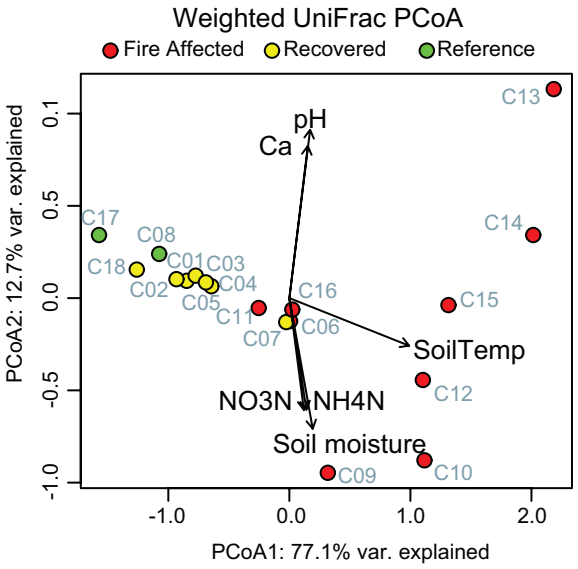


Figure 3

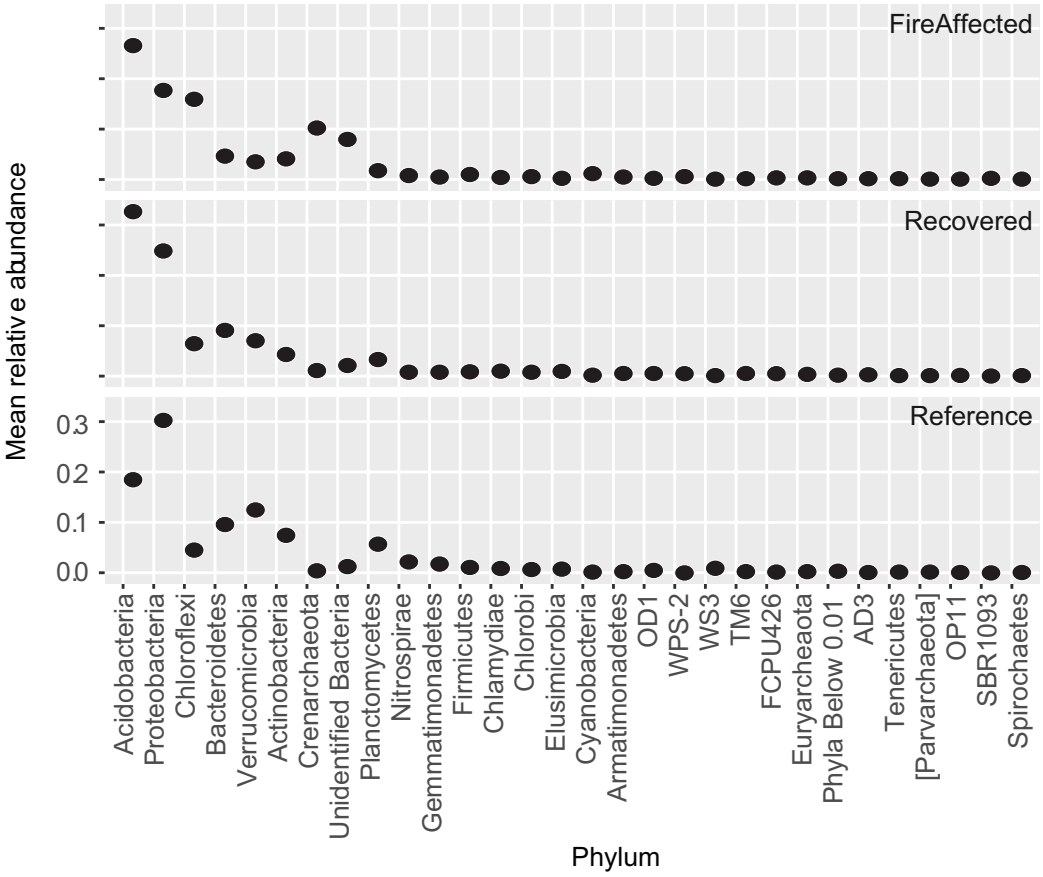


Figure 4

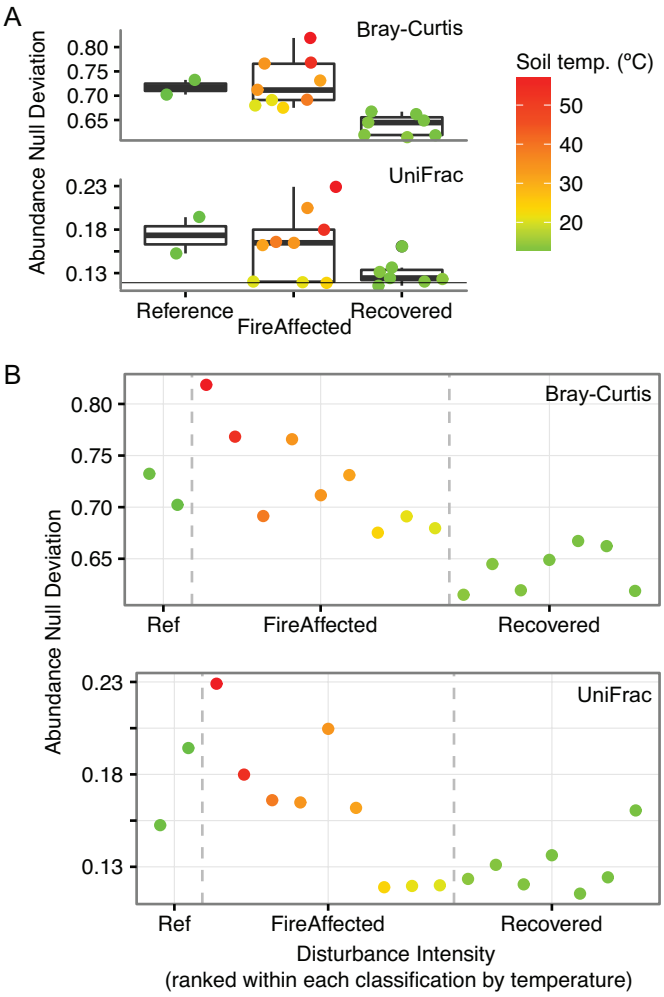


Figure 5

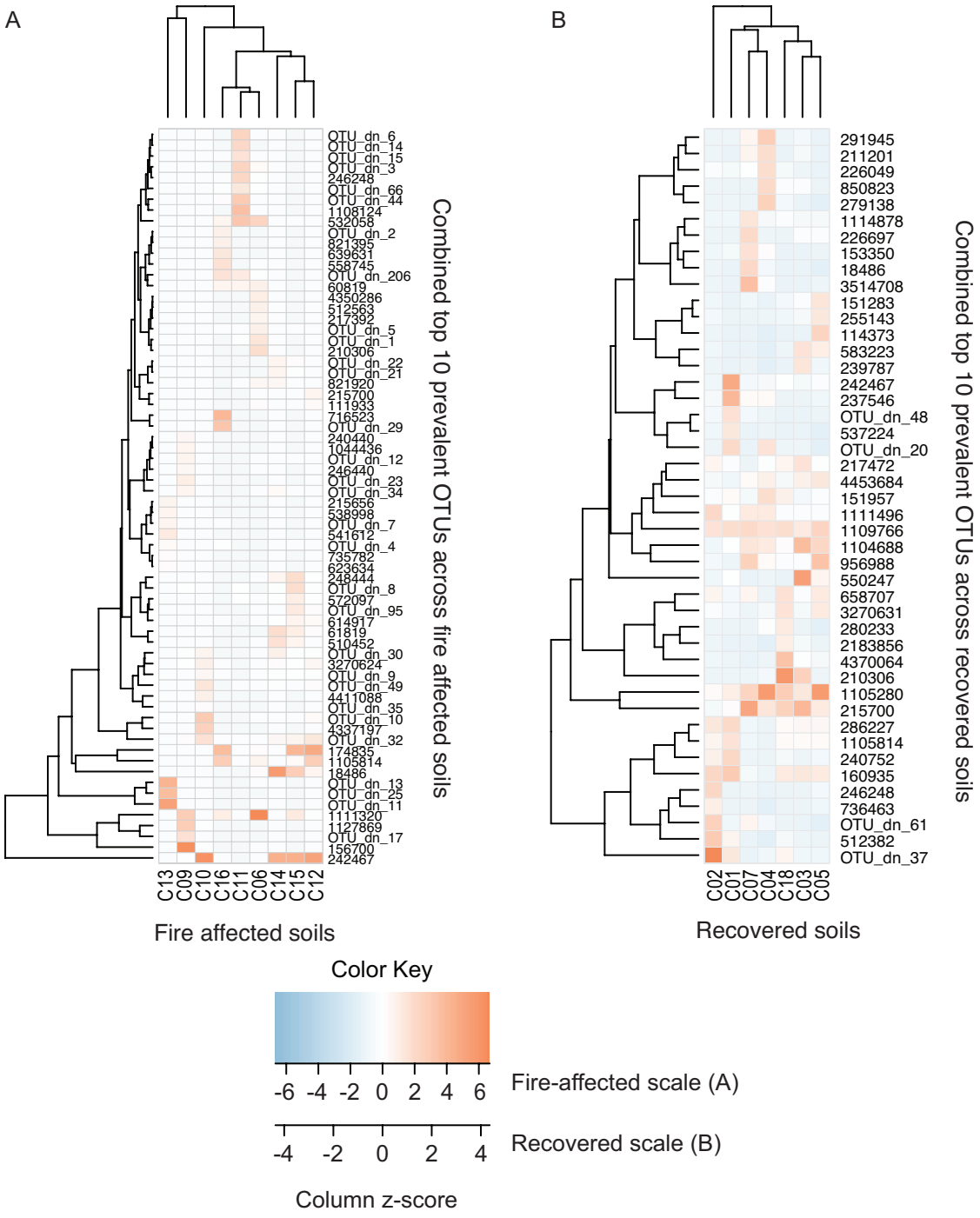
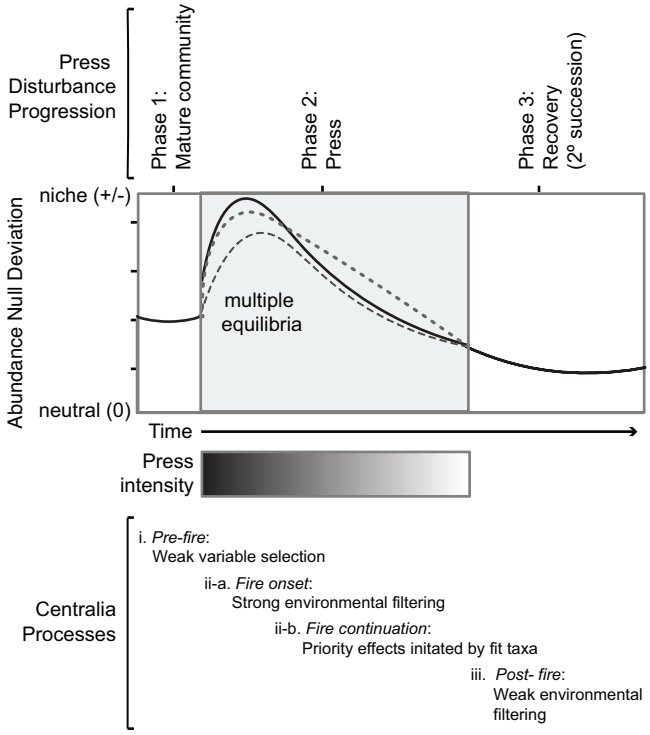

A
B
OTU_dn_6
OTU_dn_14
OTU_dn_15
OTU_dn_3
246248
OTU_dn_66
OTU_dn_44
1108124
532058
OTU_dn_2
821395
639631
558745
OTU_dn_206
60819
4350286
512563
217392
OTU_dn_5
OTU_dn_1
210306
OTU_dn_22
OTU_dn_21
821920
215700
111933
716523
OTU_dn_29
240440
1044436
OTU_dn_12
246440
OTU_dn_23
OTU_dn_34
215656
538998
OTU_dn_7
541612
OTU_dn_4
735782
623634
248444
OTU_dn_8
572097
OTU_dn_95
614917
61819
510452
OTU_dn_30
3270624
OTU_dn_9
OTU_dn_49
4411088
OTU_dn_35
OTU_dn_10
433717
OTU_dn_32
174835
1105814
18486
OTU_dn_13
OTU_dn_25
OTU_dn_11
1111320
1127869
OTU_dn_17
156700
242467
C13 C09 C10 C16 C11 C06 C14 C15 C12
Fire affected soils
Combined top 10 prevalent OTUs across fire affected soils
291945
211201
226049
850823
279138
1114878
226697
153350
18486
3514708
151283
255143
114373
583223
239787
242467
237546
OTU_dn_48
537224
OTU_dn_20
217472
4453684
151957
1111496
1109766
1104688
956988
550247
658707
3270631
280233
2183856
4370064
210306
1105280
215700
286227
1105814
240752
160935
246248
736463
OTU_dn_61
512382
OTU_dn_37
C02 C01 C07 C04 C18 C03 C05
Recovered soils
Combined top 10 prevalent OTUs across recovered soils
Color Key
-6 -4 -2 0 2 4 6
Fire-affected scale (A)
-4 -2 0 2 4
Recovered scale (B)
Column z-score

Figure 6



Tables

Table 1. Ten most abundant OTUs in fire-affected Centralia soils. OTUs (defined at 97% sequence identity) were assigned to the most resolved taxonomic level possible; there were no taxonomic assignments that could be made to these prevalent OTUs below the family level (RDP Classifier confidence > 0.80).

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%	Archaea; Crenarchaeota; MBGA
OTU_dn_1	2.5	100%	Bacteria; Chloroflexi; Ktedonobacteria;Thermogemmatisporales; Thermogemmatisporaceae;
OTU_dn_2	2.2	100%	Bacteria; Chloroflexi; Ktedonobacteria;Thermogemmatisporales Thermogemmatisporaceae
242467	2.0	100%	Bacteria; Acidobacteria; DA052;Ellin6513
174835	2.0	100%	Archaea; Crenarchaeota; Thermoprotei;YNPFFA; SK322
61819	1.7	100%	Bacteria; Acidobacteria; TM1
OTU_dn_17	1.5	78%	Bacteria; Proteobacteria; Deltaproteobacteria
215700	1.4	100%	Bacteria; Acidobacteria; Acidobacteriia;Acidobacteriales; Koribacteraceae
OTU_dn_8	1.3	100%	Bacteria
OTU_dn_3	1.2	100%	Bacteria