Response to reviews for **ISMEJ-16-00575OA**
"Divergent extremes but convergent recovery of bacterial and archaeal soil communities to an ongoing subterranean coal mine fire"
by S-H Lee, J W Sorensen, KL Grady, TC Tobin, and A Shade

Thank you for your thoughtful consideration of the work! We indicate our response by the text beginning with ">>>". We provide full references at the end of the response document. Line numbers (L) refer to the revised line numbers in the tracked changes versions of the main text and supporting materials, which we provided as supplemental files for review only.

Comments from reviewer(s):

Referee #1 (Comments to the Author):

Report on Lee et al
This study makes use of a well documented environmental accident (the underground fire) to test some fundamental tenets about bacterial community assembly and stability. The scenario is intriguing in that it represents a press disturbance that is moving. The advantage of this is that the team can monitor resilience (i.e. recovery) as well as resistance.
All in all this is an outstanding contribution that uses high-throughput sequencing to test the fundamental principles of community assembly in bacteria (+ archaea), drift, dispersal etc. There are key insights into the resilience of the community as well as the more applied aspects of the effects of this type of disturbance. The methodology seems particularly thorough (I am not expert in the statistical tests), especially testing the robustness of the technical replicates and the statistical tests applied to the sequence data.
This is a particularly good example of the power of the high-throughput sequencing, not just to say that there are changes in microbial community structure but to advance ecological theories and understanding. The paper is very clearly presented and gets the messages across very well.

> >>>Thanks so much for the very positive comments on the work. We're glad that the reviewer found the work to be "**outstanding**" and "**thorough**."

Just a point for clarification, they refer to the 'rapid' recovery within 10-20 years. The term 'rapid' here should be justified – months would be rapid but years?

> >>> We apologize that we cannot find in the text where we specifically refer to the recovery as "rapid" – we did use the phrase "high resilience" on **L415** and have revised to remove the modifier.

Although not the subject of the study, the flip side of changes in microbial community structure and some soil chemistry (pH was lower in recovered sites) is that microbial functions are potentially altered. Just a sentence or two to remind readers that there would be potential effects on soil functions would round off the paper.

> >>> Thank you for this suggestion, and agreed. We have added these sentences on **L557**.

Referee #2 (Comments to the Author):

The manuscript presented by Lee et al. contains several positive aspects that stand out. First, I enjoyed the communication aspect, which articulates community assembly in line with the 'Vellend's conceptual synthesis in community ecology' (Vellend, 2010). Second, the study system is 'ideal' to test/describe the relative influences of these classes of processes on community assembly. And third, I have to emphasize that this is a well-written manuscript. It contains clear ideas presented in a well-structured manner.
With that said, I also have some major concerns about the manuscript.

> >>Thank you for the positive comments on the presentation and ecological foundation of the manuscript.

I listed some major drawbacks that 'weaken' the manuscript and to a certain degree, hinder the validity and appreciation of the findings: as it is presented throughout the text, it seems that all the analyses collate both bacteria and archaea. Is it correct to assume that? If yes, why did the authors chose such approach? These two domains are not related and are expected to respond differently across any given environmental gradient. This is of key importance for this study, as it presents a scenario of temperature disturbance.

> >>>Thank you for this comment. The primers used for this study were the Earth Microbiome Project (EMP) standard V4 primers (Caporaso et al. 2012), and they were chosen to allow for direct comparability between this dataset and the wealth of community data from other environments collected by the EMP. These primers target both archaeal and bacterial lineages, but are biased to not have as comprehensive of coverage for all known archaeal lineages (e.g., Quince et al. 2011; Parada et al. 2015). We hypothesized that archaeal lineages, in particular Crenarcheota, would be responsive to the fire because some of these lineages include thermophiles (e.g., Blöchl et al. 1997, Kvist et al. 2005). This is what we observed (Fig 3), and, following the reviewer's comments, we observed different responses to the temperature gradient between phyla and between OTUs, inclusive of both Bacteria and Archaea lineages (e.g., it was not that all bacterial lineages had one response and all archaeal lineages had a different one).

This manuscript lacks one major component to the story the authors are trying to convey, that is, the aboveground data (plant biomass / community composition / diversity). It is known that plants exert a high degree of selection and potentially drive a large proportion of microbial community turnover in soils. To simply neglect plant communities in the system can strongly bias the outcome of these analyses and, as such, the major conclusions of the article.

> >>>Thank you for this comment. We should have clarified that we purposefully collected cores only at bare (no vegetation) surface soil locations so that our communities were least influenced by local vegetation and were most directly comparable to one another, as the most thermal soils lacked any vegetation. We have added these clarifying details into the methods section (**L148**).

One of the weaker part of the manuscript lies in the sampling design and overall quantitative analytical approach. First, as chronosequences assume space-per-time substitutions (and there are known caveats associated with the use of the approach, see Walker et al., 2010), I would strongly recommend the consideration of temporal sampling within and across sites. However, the manuscript falls short in this aspect, as samples were collected at only one sampling time. This is also important for the analytical approach itself, as it provides short- and long- term successional dynamics, thus increasing the resolution scale of the analyses.

>>>We agree with the reviewer that temporal sampling will provide additional insights beyond the chronosequence analysis. The chronosequence design, despite its caveats as discussed by Walker et al. 2010, is a common and appropriate design for our question of long-term community assembly, and this design or similar has been used frequently for similar microbial ecology questions, including many published in *The ISME Journal* (e.g., Shütte et al. 2009; Kuramae et al. 2010; Brankatschk et al. 2011; Dini-Andreote et al. 2014; Nemergut et al. 2015). While we plan to continue to monitor these sites at the Centralia fire ecosystem to address some of these temporal questions in future work, the dataset that we have now is the complete chronosequence design. It will take multiple years to decades of follow-up to be able to expand the study to adequately observe temporal dynamics given the gradual fire movement (the Centralia research program is yet-unfunded, and the fire advances at a rate 5-7 m/yr). Aside from requiring multi-annual sampling and substantial effort and expense beyond the scope of the current chonosequence study, inclusion of an additional temporal study would greatly lengthen the work and reduce its focus (the current presentation of the chronosequence analysis is at/beyond page length requirements for the ISME Journal.).

>>>Both reviewers 1 and 3 have said that they feel the study, it its current design, provides "**key insights**" and is "**well-conceived**", respectively. Reviewer 3 additionally stated that the work is: "**strong in methodology in general and does a solid job of describing the uncovered community patterns.**"

Moreover, I do not understand the argument raised in pg 8 ln 245-249. I am against the idea of collating technical replicates to increase sequencing depth. I also do not agree with this need, as, to my view, all samples in this study have enormous coverages. This strategy cuts off the statistical validity of the manuscript. Replicates MUST always be shown and discussed with clarity.

>>>Thank you for this comment. We completely agree that replicates must always be shown and discussed with clarity. This is why we included a specific assessment of their variability and their contribution to dispersion in the ordination and variation in alpha diversity in **Supporting Figure 2** and **Supporting Table 2**, respectively. Additionally, we have made all data, including technical replicates completely available for independent analyses to be reproduced by others (see https://www.ncbi.nlm.nih.gov/sra/SRP082686, we have added the accession in **L220, L243**). We advocate for reproducible science, and we aim to be completely transparent with the replication. We note that reviewer 1 has commented as to the rigor of the treatment of replicates.

>>>There are two remaining issues raised by the reviewer that we are pleased to address: 1) that of including technical replicates in downstream analysis; and 2) that of collating technical replicates to improve observational effort.

>>>For point 1): It is statistically not advisable to use technical replicates in downstream analysis (e.g. for PERMANOVA or other hypothesis testing, linking environmental variables, etc) because this would artificially inflate the statistical power of the test (pseudo-replication, see Hurlburt et al. 1984 and citing works). Each technical replicate is not independent, and so they cannot be treated as independent samples (e.g., true biological replicates) in downstream analyses because they would provide redundant data.

>>>To address 2): To summarize our response for this point, pooling of technical replicates is supported by the literature, by our data for these particular soil communities, and by our scientific interest in the rare biosphere. First, Kennedy et al. 2014 (AEM) showed that the PCR step in sequence analysis contributes variability and suggests that combining independent PCR reactions (technical replicates, as we included here), results in improved observational effort for the community. We have followed their recommendations. Next, because the technical replicates have low variability (suggesting that technical methods were very reproducible, see Supporting Table 2), it is justified to combine them for downstream analysis. Related, our rarefaction analysis shows our soil communities would not reach complete asymptote with some of our soils if the technical replicates are not combined into one representative. This is typical for soils to require very deep sequencing efforts to capture its diversity. Aside from the limited conclusions drawn from incomplete sampling effort, there would also be potential to miss variability and dynamics attributed to rare biosphere members. Finally, our research group has special interest in the roles of the rare biosphere for community stability (e.g., Shade et al. 2014), and so we wanted to observe our communities as exhaustively as possible to capture the dynamics of the rare biosphere.

Last but not least, I am not confident about the quantitative approach provided. For instance, to quantitatively infer on the influence of selection by using only temperature and ordination analysis may not be the best strategy here. This also extends to the way dispersal was measured, 'the value of spatial distance for explaining differences in community structure' (pg13, L386-387). I am not an expert in such quantitative approaches, but by looking up to recent articles in the field I found Stegen et al., 2013; which (despite still not perfect) provides a more robust and reliable way to quantitatively describe such processes. To my view, simulation analyses and phylogenetic models are more robust ways to disentangle such processes in (microbial) community assemblages across env. gradients.

>>>We appreciate the comment on our quantitative approach, which was to use a suite of different but complementary analyses to partition the relative contributions of deterministic and stochastic processes to community patterns. There are several ways that these patterns could be analyzed with ecological statistics, and the approaches that we've chosen have ample and non-controversial support in the literature (please see below). We'd like to first discuss new analyses that we have added, and next to provide support for our original analyses. Specifically, we'd like to discuss how those approaches that we've originally chosen are appropriate and insightful.

>>>1. At the specific suggestion of reviewer 3, we have added an additional beta-null model approach to our current approach using both abundance (Bray Curtis) and phylogenetic weighted (Weighted UniFrac) null models (new **Figure 4**). We believe that

the addition of this simulation-based analysis (resampling to quantify a null distribution) addresses reviewer 2's specific request for simulation analysis.

>>>We admire the collection of the Stegen work and have cited many of these works in the current manuscript to set a conceptual foundation for partitioning stochastic and deterministic processes. However, there are several reasons that we chose not to include their exact approach in this work.

>>>The main limitation of the phylogenetic approach proposed by Stegen 2012/Dini-Andreote 2015 is that it makes the assumption that more closely related taxa are more ecologically similar (see p E1328, column 1, paragraph 4 in Dini-Andreote et al. 2015: "Using phylogenetic turnover to make ecological inferences requires that phylogenetic distances among taxa approximate differences in the ecological niches they occupy"). This assumption is not universally met, especially within-clades, as there are several works that have shown that even very closely related strains can have distinct ecological traits and niches (e.g., Rasko et al. 2008; Cordero et al. 2012, also works by Chisholm on Prochlorococcus ecotypes, etc.)

>>>A recent criticism of the Dini-Andreote et al. 2015 /Stegen et al. 2012/Stegen et al. 2013 and related work was put forth by Tucker et al 2016, which suggested that the approach used by Stegen and colleagues does not interpret to an ability to partition deterministic and stochastic variability, but rather to distinguish niche and neutral processes. The authors modified the original approach offered by Stegen et al 2012 and use simulations to show that stochastic and deterministic processes cannot be distinguished using beta-null modeling *sensu* Stegen 2012, but can comparatively distinguish niche and neutral processes*, if abundance-based measures of community beta diversity are applied*. Thus, we have instead used the Tucker et al. 2016 abundance-based approach and applied their updated interpretation (see new **Fig 4**, new **Fig 6** and **L477-493**). We also have extended the Tucker et al. 2016 approach to include use of the weighted phylogenetic resemblance (weighted UniFac), which is directly comparable with the other comparative (beta) diversity analyses used in our work, and also allows us to ask how incorporating phylogenetic information into the resemblance informs the outcome. To the best of our knowledge, this is the first time that UniFrac has been used in beta-null modeling. A benefit is that UniFrac calculations are less computationally expensive in R and more widely used than Beta MNTD in microbial ecology; we hope that the example and code will be useful for others and receive further development in the research community.

>>Another limitation to the Stegen and colleagues phylogenetic null modeling is in computing resources required for use with large trees (to calculate the inter-species phylogenetic branch lengths for Beta MNTD). This memory/processing limitation generally which hinders reproducibility of the code in R, and is likely why 454 datasets, datasets truncated to include only the most prevalent members, or low-richness datasets have been most successfully used with their approach. We are grateful for the hack that we received via reviewer 2 (from Jack Darcy) to improve the computational demand of the pairwise BetaMNTD calculation with large trees. The example provided by Mr. Darcy was estimated to work within ~190 hrs for one simulation with a 30K-tip tree; more than 999 permutations are needed to complete simulation for beta-null modeling. Because of the computational demand and because the method was not directly comparable with the other beta diversity analyses in our study, and because of the recent criticisms of the approach in the literature, we instead decided to move forward with the abundance-

based beta-null modeling as proposed by Tucker et al 2016 and suggested specifically by reviewer 3. We continue to explore and consider the BetaMNTD null approach, and to think about how to use a combination of complementary approaches to most completely address deterministic/stochastic and niche/neutral partitioning.

>>>2. Integrity of the original analyses. We would like to clarify that we've used a suite of environmental variables, not only temperature, in our analysis of deterministic processes, including percent moisture, organic matter, NO_3^-, NH_4^+, pH, SO_4, K, Ca, Mg, P, As, and Fe (see **L152** and **Supporting Figure 3**), and that linking environmental variables to changes in community structure is a valid and common deductive approach for assessing stochastic/deterministic processes with respect to space and/or the environment as discussed in the ecological literature (see Vellend et al. 2014 section on "Observational approaches: multivariate analysis" and Table 1 therein). We also have applied the variance partitioning methods suggested by Vellend et al. 2014 distinguish stochastic/deterministic processes (Table 1 therein).

>>> For the dispersal analysis, the explanatory value of space for differences in community structure has been reported previously using this and comparable methods in ecology (e.g., Vellend et al. 2014, Thompson and Townsend 2006, Cottenie et al. 2003, Whittaker et al. 2006).

>>> We note that the other two reviewers have commented very positively on our quantitative approaches, including their appropriateness and their rigor.

Other comments:
In pg 5, ln 133-134 the authors describe DNA yields to vary between 1.3 to 129 ng/µL (i.e. differing at two orders of magnitude). Later, in pg 9, Ln 279, they show that all soils contain the same order of magnitude, based on the bacterial 16S rRNA gene copy numbers. How do (can) they explain that? Is that an issue of bias in DNA extraction? Did they quantified bacteria and archaea separately? Moreover, there is no statistical support for Supplementary Figure 2, but there are clear differences in 16S rRNA copies across soil types. Also, it is not clear if qPCRs and cell counts were carried out for both bacteria and archaea separately.

>>>Thank you for this comment; we are happy to clarify. We apologize for the confusion about the yields of DNA reported in the text at (*original line number*) ln 133 and in (*new designation*) **Supporting Figure 4**. The disconnect was in the different units reported: **Supporting Figure 4** (note new designation, was old SFig2) was reported in units of ng/gram of dry soil (which is the most appropriate standardized units for soils, e.g. Frostegård et al. 1999) while the main text was reported in units of ng/uL; this is our mistake. For consistency, we have removed the concentration reported in ng/uL from the main text. The amount of 16S rRNA copies per gram of dry soil was consistent and not statistically different between fire classifications (all pairwise $p > 0.09$, student's t-test). Standardized masses of DNA were used for sequencing, as typical with library preparation (1,000 ng DNA at the Michigan State Genomics Core).

>>>Bacteria and archaea are not distinguishable in the qPCR of 16S rRNA genes using the EMP primers or in cell counts (bacterial and archaeal cells are indistinguishable by morphology), and so they were, necessarily, analyzed collectively. We have no evidence to suggest a particular bias in our DNA extraction protocol.

>>>Statistical support for **Supporting Figure 4** originally was reported in the main text (paragraph **L328**), but we have additionally added this text to the **Supporting Figure 4** figure legend (**L910**) and to our workflow R script (in workflow ~L111-119). As was true for the mass of DNA per gram of dry soil, 16S rRNA gene copies per gram of dry soil were not statistically different (all pairwise $p > 0.09$, student's t-test). Also, if a Bonferroni adjustment is applied (arguably, too conservative, see Moran et al. 2003), p at alpha= 0.05 should be $p < 0.05/3$ hypotheses or $p < 0.017$ to be significant.

>>>Quantitative PCR was performed on the same amplicons (with the same primers) as for the sequencing (EMP primers), and so it is not possible to distinguish bacteria and archaea with this assay because there is only one primer set used to target both kingdoms.

In addition, press disturbance, by definition, is expected to cause a shift in community abundance, isn't it? Even if at a short time scale. That contrasts with the idea presented in Pg 9 Ln 283-286. How can the authors explain this issue? This is one of the aspects not well-developed in the text. One line of evidence on that is the fact that 'hottest soils were more likely to have extreme or disparate values (of both community composition and soil contextual data). Here, I argue that press disturbance may cause a decrease in community diversity and abundance, thus potentially elevating the influence of drift. Can the authors comment on that?

>>> We agree with the reviewer that there are many definitions of disturbance (see Shade and Peters et al. 2012), and that not all definitions include a necessary kill-off of members, which is what we interpret the "shift in community abundance" comment to mean (please clarify if not). We have added text on **L128** to clarify our definition of disturbance.

>>>We are not observing the immediate impacts of the high temperatures on the community but rather the long-term outcomes of that press stressor, and so we consider this to be a long time scale relative to the generation times of microorganisms. A laboratory-scale experiment of warming would be needed to observe the immediate changes in community size or structure with the soils first exposure to the stressor (and, if ever funded, we have such exciting experiments planned –stay tuned!). Furthermore, the stressor progresses gradually in this ecosystem. The fire gradually advances along the coal seams (days to weeks to months), and as it advances, it gradually warms the surface soils, allowing for processes like dormancy and adaptation to "buy" time for taxa sensitive to high temperatures.

>>>We agree with the reviewer that the disturbance caused a decrease in diversity. In fact, we have presented evidence for this that included multiple diversity metrics (**Fig 1**). However, our cell counts and our qPCR results, which are independent methods of assaying community size, do not support that there was a statistically significant change in community size (number of cells/members; **Supporting Figure 4**).

Regarding the quantitative approach, more details must be provided in the description of the quantitative PCR (M&M) section. These concern the standard curve, replicates, reaction efficiency and post-calculations. The info is required for the clarity of communication.

>>> Thank you for this comment, and apologies for the omission of some of the qPCR

details – we added all of the information requested in supporting materials, in a new section called “Quantitative PCR Methods” (Supporting Methods, page 2, **Supporting Figure 8**).

Sequence processing: It is not clear to me why the approach initially clustered sequences using UPARSE and, later on, those sequences that did not matched the reference, were clustered using uclust (pg 7, ln 220). Please clarify it.

>>>This is the standard open-reference approach as published by Rideout et al. 2013. Since its publication, the open-reference approach has been widely used for OTU picking (method cited 73 times as of 27 Oct 2016; please see: https://scholar.google.com/scholar?cites=11445558021147656681&as_sdt=80000005&sciodt=0,23&hl=en).

>>>To explain further, open reference OTU picking aims to maximize use of all high quality sequences by applying consistent OTU definitions that are first assigned by matches to quality reference database and then by tractable *de novo* definitions that can be applied to subsequent studies. To summarize the introduction by Rideout et al. 2013: The alternatives to open-reference OTU picking are reference-based OTU picking and de novo OTU picking. Reference-based approaches remove any OTUs that do not hit to a reference database from the dataset. Thus, reference-based OTU picking by itself was not appropriate for our ecosystem because: 1) we hypothesized that we may to observe novel diversity and 2) soils taxa are diverse and are generally underrepresented in existing 16S rRNA databases. The de novo approach does not allow for tractable OTUs across datasets/analyses. Essentially, if we wanted to extend our dataset later, we would have to reassign OTUs “fresh” every time, which is not optimal. As described in Rideout et al., open-reference approach first matches representative sequences to a database, and assigns OTUs based on hits to the database. Then, any remaining sequences that have not hit the database and clustered de novo so that they can still be included in the analysis. The de novo OTU definitions are maintained for future OTU analyses. We have inserted clarifying text as to our choice of OTU picking approach (**L250**).

Pg 8. What is the difference between normalized and nonnormalized Unifrac distances? Are these the same of Weighted and Unweighted Unifrac? If yes, please remove such confusing terminology. If not, please explain the differences in detail.

>>>Thank you for this comment; we understand that it is confusing to have so many versions of the UniFrac distance available. The differences between normalized and non-normalized UniFrac distances have been previously discussed in the literature, and it is not the same as weighted an unweighted {e.g., Lozupone and Knight. 2005, Lozupone et al. 2007; and in this informative blog post by Wong: http://gloorlab.blogspot.com/2015/04/weighted-unifrac-in-all-its-forms_23.html). To briefly summarize, the originally proposed “non-normalized” UniFrac distances, unlike many other resemblances, are not bound by 0 and 1. The “normalized” procedure applies a correction to the original distances to fall within 0 to 1 and to improve intuitive interpretation and comparability to other resemblances. Normalization was reported to not fundamentally change the underlying results of analyses (please see Lozupone et al. 2007, Fig 2 therein). We have checked that the normalized and non-normalized agree

for our dataset (see **Supporting Table 3AB**). Both normalized and non-normalized values are computable in the QIIME toolkit and specifying which was computed is needed to reproduce the analysis. We have inserted the references as to the normalized/non-normalized discussion in the methods (**L293**). Please note that normalized and non-normalized are not the same as the weighted and unweighted: weighted metrics are based on relative abundances and unweighted on presence and absence of taxa.

P9 Ln 259-260: The R script is not available. As such, I could not access the neutral model(s) they claim to have used.

>>>We apologize for the confusion; all of our analysis scripts – even those in progress- are available on the "ShadeLab" GitHub repositories. Is the reviewer referring to the R script associated with the Burns et al. ISMEJ piece that codes the neutral model that we applied to our dataset? It was probably unclear that the "sncm.fit_function.R", script in our paper GitHub repository (https://github.com/ShadeLab/PAPER_LeeSorensen_inprep) was the script written by Burns (https://github.com/ShadeLab/PAPER_LeeSorensen_inprep/ ; the repository PATH for the Burns script is "R_analysis/MiscSourceScripts/"). We have improved organization for our scripts on GitHub, updated the repo links, and added clarifying language in the methods to point readers to this and other used scripts (**L302**) The Burns et al. 2013 neutral model script has been previously peer reviewed and has been since applied to other datasets by other groups (e.g., Venkataraman et al. 2015).

Pg 10 Ln 289-311: This whole paragraph seems completely out of the context. The way it is presented, it breaks the flow of the text. Please edit it accordingly for the sake of clarity.

>>> We have moved these two paragraphs to supporting results, as suggested, to improve flow.

Pg11 Ln 332: what is/are the effect(s) of the contamination of coal combustion products on these communities. Additional comments must be provided here.

>>>We have provided additional comments, based on the work of Janzen and Tobin-Janzen 2008 (starting **L372**).

Pg 11 Ln 333-336: This statement lacks statistical support (e.g. Mantel tests, Procrustes analysis).

>>>We have added Mantel and PROTEST results in the **Supporting Table 3B**; all $p < 0.001$ for both tests and all comparisons.

Pg 11 Ln 345: "Notably, … dynamics". I do not understand what does this sentence mean. It is vague and allows for different lines of interpretation. Please clarify.

>>>We apologize for the confusion. We have clarified this sentence to read (**L405**):

"Notably, soil fire history (estimated years since the local soil surface was first measured as hot, as reported by Elick 2011) was not correlated to community dynamics (**Supporting Table 4**)." We have also added explanation in the methods as to how fire history was estimated (**L156**): "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)."

Pg 12, Ln 372-383: This overall description is weak and can be moved out of the main text. It does not add much to the manuscript, unless main lines of causes and/or consequences can be raised out of the descriptive pattern.

>>>While we appreciate this suggestion, we respectfully disagree with the reviewer about excluding the phylum-level analysis. First, the observation of phylum-level patterns suggests an overarching, robust response that is consistently detectable across different taxonomic levels. Second, recent work has shown that some functions important for ecosystem services may be maintained at the phylum level for soil microorganisms (Morrissey et al. 2016), which suggests that phylum level patterns may provide insights into functional outcomes.

Pg 15 Ln 465-467: This question can only be answered by time-series analysis. This is likely due to an increase effect of drift after the disturbance event. Thus, it leads communities to be different of each other. As time proceeds, the strength of selection progressively emerges (by both soil physicochemical properties and dwelling plants). In this context, the microbial seed bank and local dispersion are key mechanisms. This point raises a question that cannot be answered by the data provided, given the limits of the experimental design. As such, this section must be kept concise and the authors must avoid to overstep beyond the outcomes that the article provides.

>>> We agree with the reviewer's general hypotheses offered in this comment. We had discussed these specific mechanisms (seed bank **L575**); we have eliminated local dispersal as a strong underlying driver, please see paragraph starting **L454**) and have attributed homogenizing selection from increased pH in post-fire soils (**L555**, **L650**). Thus, it seems that we have addressed all of the items about which the reviewer asks. While we agree that a time series analysis would provide additional insight (please see previous comment in response to the desire for additional time series – this would require years to decades of additional study), we respectfully disagree that our current study design does offer any insights to the questions posed. We have also revised the related subsequent sections to be more focused and cautious in our hypotheses.

Pg 15-16 Ln 474-489: Despite I like the lines here, I am still not entirely sure about them. To infer on such matter, manipulative experiments must be carried out. For instance, by transplanting seed banks or homogenizing them across sites and by limiting/manipulating dispersion across sites. As it is presented, this idea steps beyond the main outcome of the manuscript. As such, it requires a cautious revision.

>>>We have cautiously revised as suggested. Specifically, we have revised sections to focus on the evidence (direct and indirect) in our data and in the unique Centralia ecosystem for the priority effects/unmeasured extreme abiotic conditions hypothesis. We

have removed a paragraph discussing how stochastic versus responsive dormancy awakenings could contribute, which we agree cannot be addressed by the current study.

Figure 3 and Supplementary Figure 2: Statistics missing here.

>>>We have added our statistical results to the legends of **Fig 1** and **SFig4** (originally, Fig 3 and SFig2); the statistics were/are provided in the main text (**L363**, **L365**, **L331**).

Figure 5: What is the meaning of the group 'k__Bacteria; Others'? This is likely represented by a cluster of very distinct sequences that are not well affiliated, am I right? As such, there is no phylogenetic basis for such cluster. Then, my question is 'what is the meaning of showing it?

>>> This is a group of Bacterial OTUs that include high quality sequences yet are not able to be unidentified at the phylum level against the greengenes database taxonomy. The nomenclature is the typical output from taxonomic assignment using the QIIME workflow, where "k__" designates a "kingdom" –level taxonomy. We have altered this label to now read "Unidentified Bacteria" It is correct that, unlike the other designations, there is no phylogenetic basis for the cluster. The reason for showing it is to 1) demonstrate potential novel or under-described diversity in the Centralia ecosystem and 2) to represent the complete dataset, inclusive of well-identified taxa and not well-identified taxa. Including the unidentified lineages in the phylum-level breakdown is typical in the literature for these reasons. We have added text to the **Figure 3** (old Fig5) legend to clarify.

Figure 6: The heatmap is hard to read. What are the codes/affiliations for the rows? Please edit the figure with caution for clarity.

>>>We have improved the clarity of the heatmap by adding row and column labels, by augmenting the legend, and adding a neutral grid to distinguish the heat blocks. (This Figure is now **Figure 5**).

Table 1. The table should be better formatted. For instance the codes and taxonomic assignments are raw QIIME outputs. Edit them for clarity.

>>>We have improved the table as suggested.

Referee #3 (Comments to the Author):

This is a well-written and well-conceived study describing the response of soil bacteria and archaea to press disturbances from underground mine fires. This study is unique in the type and duration of disturbance being described and generates some interesting data. In its present form it will be of interest to community ecologists, particularly those interested in understanding community assembly in response to disturbance. The text could be strengthened a bit in the interpretation of drivers of multiple-equilibria across the fire-affected communities, more detail on this below.

>>>Thank you for the positive comments. We're glad the reviewer thinks that the study is "**unique**", "**interesting**" and "**well-conceived**" (and note that this last compliment is in contrast to the opinion of reviewer 2, who was critical of chronosequence analysis). We have strengthened the text as suggested, please see response below.

Minor Comments:
While the manuscript text is polished and relatively easy the figures do not always correspond properly with their numbering. E.g., Figure 1 in the materials for review is the site map, but the description for this appears under Figure 2; at one point in the manuscript the text references an ordination but points to the heat map in Figure 6. I was also momentarily confused by Figure 2 (Figure 1 by description) as it shows soil physical and chemical properties in relation to heat, but has the presumed independent variable on the vertical-axis instead of the horizontal-axis as a reader would expect.

>>>We apologize for the disconnect between the first two figures and their legends; this was our mistake. We have edited the legends in the text to be correctly assigned to their figures. We opted to put temperature on the y-axis to maintain easy comparability in y-axis ranges across panels and to make the most visually simple figure (switching axis would require more space because of full labeling required of both x and y axes for each panel in the grid). We were asked to remove figures from the main text by the editorial assistant, and so we have moved these figures to the supporting materials (now **SFig1** and **SFig 3**, respectively).

Major Comments:
I have only one major comment that would like further feedback from the authors on at this time. The manuscript is strong in methodology in general and does a solid job of describing the uncovered community patterns. However, it wraps up with a range of untested hypotheses for explaining the multiple equilibria of fire-affected communities. Perhaps an alternate approach invested more in conceptualizing the findings here with previous work on soil microbial responses to disturbances is a path toward a stronger contribution.

>>>Thank you for this comment, and we agree. Thus, we have integrated with Dini-Andreote et al. 2015 and Ferrenberg et al. 2013 to extend their conceptual framework to our study (**Fig 6**). We have added discussion of our conceptual framework beginning on L639. Our extension of the frame work integrates two ideas to extend the existing conceptual models: 1) the inclusion of a long phase encompassing the press disturbance, instead of disturbance represented as a single (pulse) point; and 2) the hypothesis of multiple equilibria (within the press disturbance phase).

An alternate modeling approach might allow for an easier integration with existing studies while serving to narrow the possible drivers of the multiple-equilibria. For example, neutral modeling can be completed using species abundance patterns as in the Sloan et al. (2007) type of approach used here, or through the use of beta-diversity based neutral models that account for the size of total and local species pools (e.g., Chase and Myers 2011). In my opinion, and perhaps the authors have points that will sway me, the present study seems to be better suited for neutral v niche assessments via the latter, beta-diversity approach. I say this with particular consideration for narrowing among the many processes that are being invoked as hypothetical explanations of the multiple equilibria of fire-affected communities.

The higher disparity in community composition in fire-affected soils would lead to higher levels

of observed beta diversity. Thus, interpreting this beta diversity in relation to expected levels of beta diversity given gamma and alpha diversity levels would be helpful. This is the approach used in Ferrenberg et al. (2013) that formed the basis of Dini-Andreote et al.'s (2015) model. I'm not arguing this is necessarily a better approach in all situations, but given that the community patterns presented in this study appear to fit with the patterns in Ferrenberg et al. (2013) who studied surface fire, it could be useful to see how observed values compare to expected values. If observed values far outpace expected, it would suggest an unmeasured deterministic force(s) such as competition or priority effects are at work shaping the fire-affected communities and this would narrow the hypothetical drivers of the observed multiple-equilibria. If the observed beta diversity is closer to the expected (which are based on randomizations) it would signal a stochastic influence in these communities possibly hinting at stochastic breaking of dormancy or legacies of past dispersal to seed banks for example. Understanding the ratio of these processes across the different communities would be insightful and allow you to place the results within the context of the three phase model proposed by Ferrenberg et al. (2013) and corroborated by Dini-Andreote et al (2015). Perhaps a more interesting outcome would be finding opposition to these models altogether. R code for implementing and recommendations for interpreting these models is available from Tucker et al. (2015): http://onlinelibrary.wiley.com/doi/10.1111/oik.02803/abstract;jsessionid=6CDDD35C8B73025CF55014A5E608E479.f04t03

>>>Thank you very much for this excellent suggestion, which we have incorporated into the revised work. Specifically, we have applied the Tucker et al. 2016 abundance weighted beta-null modeling approach to our data, and additionally have extended the approach to calculate beta-null expectations using the weighted UniFrac distance (incorporates phylogenetic breadth). From this new analysis, we found that the communities at the most extreme temperatures were relatively further from neutral expectations than the reference and recovered communities, suggesting a higher influence of niche processes at peak disturbance, and narrowing our hypothesis to be that priority effects or other deterministic factors largely are at play (**L477**, **L602**, **L627**, **Fig 4**). We have made new **Figures 4 and 6** to discuss the new results and our extension of the conceptual model. We appreciate feedback on the new analysis.

References cited

- Brankatschk R, Töwe S, Kleineidam K, Schloter M, Zeyer J. Abundances and potential activities of nitrogen cycling microbial communities along a chronosequence of a glacier forefield. The ISME journal. 2011 Jun 1;5(6):1025-37.
- Blöchl E, Rachel R, Burggraf S, Hafenbradl D, Jannasch HW, Stetter KO. Pyrolobus fumarii, gen. and sp. nov., represents a novel group of archaea, extending the upper temperature limit for life to 113 C. Extremophiles. 1997 Feb 1;1(1):14-21.
- Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M, Gormley N, Gilbert JA, Smith G, Knight R. 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J.
- Cordero OX, Wildschutte H, Kirkup B, Proehl S, Ngo L, Hussain F, Le Roux F, Mincer T, Polz MF. Ecological populations of bacteria act as socially cohesive units of antibiotic production and resistance. Science. 2012 Sep 7;337(6099):1228-31.
- Cottenie K, Michels E, Nuytten N, De Meester L. Zooplankton metacommunity structure: regional vs. local processes in highly interconnected ponds. Ecology. 2003 Apr 1;84(4):991-1000.

- Dini-Andreote F, e Silva MD, Triado-Margarit X, Casamayor EO, Van Elsas JD, Salles JF. Dynamics of bacterial community succession in a salt marsh chronosequence: evidences for temporal niche partitioning. The ISME journal. 2014 Oct 1;8(10):1989-2001.
- Dini-Andreote F, Stegen JC, van Elsas JD, Salles JF. Disentangling mechanisms that mediate the balance between stochastic and deterministic processes in microbial succession. Proceedings of the National Academy of Sciences. 2015 Mar 17;112(11):E1326-32.
- Frostegård Å, Courtois S, Ramisse V, Clerc S, Bernillon D, Le Gall F, Jeannin P, Nesme X, Simonet P. Quantification of bias related to the extraction of DNA directly from soils. Applied and environmental microbiology. 1999 Dec 1;65(12):5409-20.
- Hurlbert SH. Pseudoreplication and the design of ecological field experiments. Ecological monographs. 1984 Feb 1;54(2):187-211.
- Kennedy K, Hall MW, Lynch MD, Moreno-Hagelsieb G, Neufeld JD. Evaluating bias of Illumina-based bacterial 16S rRNA gene profiles. Applied and environmental microbiology. 2014 Sep 15;80(18):5717-22.
- Kuramae EE, Gamper HA, Yergeau E, Piceno YM, Brodie EL, DeSantis TZ, Andersen GL, van Veen JA, Kowalchuk GA. Microbial secondary succession in a chronosequence of chalk grasslands. The ISME journal. 2010 May 1;4(5):711-5.
- Kvist T, Mengewein A, Manzei S, Ahring BK, Westermann P. Diversity of thermophilic and non-thermophilic crenarchaeota at 80 C. FEMS microbiology letters. 2005 Mar 1;244(1):61-8.
- Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. Applied and environmental microbiology. 2005 Dec 1;71(12):8228-35.
- Lozupone CA, Hamady M, Kelley ST, Knight R. Quantitative and qualitative β diversity measures lead to different insights into factors that structure microbial communities. Applied and environmental microbiology. 2007 Applied and Environmental Microbiology. Mar 1;73(5):1576-85.
- Lozupone C, Lladser ME, Knights D, Stombaugh J, Knight R. UniFrac: an effective distance metric for microbial community comparison. The ISME journal. 2011 Feb 1;5(2):169.
- Nemergut DR, Knelman JE, Ferrenberg S, Bilinski T, Melbourne B, Jiang L, Violle C, Darcy JL, Prest T, Schmidt SK, Townsend AR. Decreases in average bacterial community rRNA operon copy number during succession. The ISME journal. 2015 Nov 13.
- Moran MD. Arguments for rejecting the sequential Bonferroni in ecological studies. Oikos. 2003 Feb 1;100(2):403-5.
- Morrissey EM, Mau RL, Schwartz E, Caporaso JG, Dijkstra P, van Gestel N, Koch BJ, Liu CM, Hayer M, McHugh TA, Marks JC. Phylogenetic organization of bacterial activity. The ISME journal. 2016 Mar 4.
- Parada AE, Needham DM, Fuhrman JA. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. Environmental microbiology. 2015 Nov 1.
- Quince C, Lanzen A, Davenport RJ, Turnbaugh PJ. Removing noise from pyrosequenced amplicons. BMC bioinformatics. 2011 Jan 28;12(1):1.
- Rasko DA, Rosovitz MJ, Myers GS, Mongodin EF, Fricke WF, Gajer P, Crabtree J, Sebaihia M, Thomson NR, Chaudhuri R, Henderson IR. The pangenome structure of Escherichia coli: comparative genomic analysis of E. coli commensal and pathogenic isolates. Journal of bacteriology. 2008 Oct 15;190(20):6881-93.
- 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.
- Schütte UM, Abdo Z, Bent SJ, Williams CJ, Schneider GM, Solheim B, Forney LJ. Bacterial succession in a glacier foreland of the High Arctic. The ISME journal. 2009 Nov 1;3(11):1258-68.
- Shade A, Jones SE, Caporaso JG, Handelsman J, Knight R, Fierer N, Gilbert JA. Conditionally rare taxa disproportionately contribute to temporal changes in microbial diversity. MBio. 2014 Aug 29;5(4):e01371-14.
- Stegen JC, Lin X, Konopka AE, Fredrickson JK. Stochastic and deterministic assembly processes in subsurface microbial communities. The ISME journal. 2012 Sep 1;6(9):1653-64.
- Stegen JC, Lin X, Fredrickson JK, Chen X, Kennedy DW, Murray CJ, Rockhold ML, Konopka A. Quantifying community assembly processes and identifying features that impose them. The ISME journal. 2013 Nov 1;7(11):2069-79.
- Thompson R, Townsend C. A truce with neutral theory: local deterministic factors, species traits and dispersal limitation together determine patterns of diversity in stream invertebrates. Journal of Animal Ecology. 2006 Mar 1;75(2):476-84.
- Tucker CM, Shoemaker LG, Davies KF, Nemergut DR, Melbourne BA. Differentiating between niche and neutral assembly in metacommunities using null models of β-diversity. Oikos. 2016
- Vellend M, Srivastava DS, Anderson KM, Brown CD, Jankowski JE, Kleynhans EJ, Kraft NJ, Letaw AD, Macdonald AA, Maclean JE, Myers-Smith IH. Assessing the relative importance of neutral stochasticity in ecological communities. Oikos. 2014 Dec 1;123(12):1420-30.
- Venkataraman A, Bassis CM, Beck JM, Young VB, Curtis JL, Huffnagle GB, Schmidt TM. Application of a neutral community model to assess structuring of the human lung microbiome. MBio. 2015 Feb 27;6(1):e02284-14.
- Whitaker RJ, Grogan DW, Taylor JW. Geographic barriers isolate endemic populations of hyperthermophilic archaea. Science. 2003 Aug 15;301(5635):976-8.