Accounting for active populations provides a new perspective on the degree and mechanisms of soil microbiome resilience

Samuel E. Barnett[1], Ashley Shade[2*]

[1]Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI 48824, USA

[2] Université de Lyon, France; Research Group on "bacterial efflux and environmental resistance," CNRS, INRAe, Ecole Nationale Véterinaire de Lyon and Université Lyon 1, UMR 5557 Écologie Microbienne, 16 rue Dubois 69622 Villeurbanne Cedex, FRANCE

Corresponding author: Ashley Shade
UMR5557 - Ecologie Microbienne Lyon
Université Claude Bernard Lyon 1
Bât. Mendel, 5eme étage
16 Rue Dubois
69622 Villeurbanne cedex, FRANCE
+33472431377
ashley.shade@cnrs.fr

Classification: Biological Sciences; Microbiology

Keywords: Microbial ecology, soil, resilience, fire

Abstract

Soil microbiomes are essential mediators of ecosystem services, and their resistance and resilience to anthropogenic disturbances can have profound effects on soil function. However, we know little about the determinants of soil microbiome resilience, especially in the context of long-term (press) disturbances. Here, we leveraged the seven-decade-old underground coal seam fire in Centralia, Pennsylvania, USA, to examine surface soil bacterial community resilience to long-term heat disturbance. As the fire slowly advances along coal seams, formerly fire-affected soils cool to ambient temperature enabling geographically localized comparisons of impacted soils across a gradient of fire intensities with recovered and reference soils. To assess soil bacterial microbiome response and resilience to the fire, we executed a seven-year field study, annually sampling sites selected to capture the range of disturbance intensities. We quantified the dynamics of the bacterial communities over time and across the fire-impact gradient and additionally accounted for population activation by combining 16S rRNA with 16S rRNA gene amplicon sequencing. Over the years, the fire-affected sites cooled, and their microbiome structures recovered towards those of reference and recovered soils. However, the whole community demonstrated greater resilience than the active subset, suggesting that resilience could not solely be attributed to environmental selection and growth of active populations. We therefore suggest that regional dispersal, persistence, and revival from the local seed bank contribute to microbiome resilience. Importantly, this work indicates that DNA-based methods that do not account for activity may overestimate microbiome resilience, with implications for understanding microbiome functional recovery.

Significance statement

The climate crisis and anthropogenic pressure on the environment are fundamentally altering ecosystems and their functions. Many critical ecosystem functions, such as biogeochemical cycling, are facilitated by microbial communities. Understanding the responses and functional consequences of microbial communities to these long-term (*press*) disturbances requires consistent, interannual observations and a focus on the active microbial populations that contribute to the critical functions. This study leverages the seven-decade-old coal seam fire in Centralia, Pennsylvania, USA, to examine the resilience of soil bacterial microbiomes to an intense, long-term heat disturbance. It provides an interannual analysis of microbiome resilience along a fire-impact gradient and directly compares the dynamics of the active subset and the "whole" community (which also includes signals from dormant and deceased cells). This work provides evidence that immigration of inactive cells together with environmental selection can drive perceived microbiome recovery, and advances understanding of the mechanisms of environmental microbiome resilience.

Introduction

Soil microbial communities are vitally important to the biosphere's health by filling important roles in biogeochemical cycling (1) and plant health promotion (2, 3), among other ecosystem services. Recently, there has been a push to use microbial communities as indicators of soil and ecosystem health (4–7), particularly in contexts of changing environments and disturbance where the stability of microbiomes can provide insight into the short and long-term effects these events have on the ecosystem (4, 5). Community stability can be defined as the interplay between resistance (*i.e.*, the degree to which the community remains unchanged during and following a disturbance event) and resilience (*i.e.*, the rate at which the community returns to a pre-disturbance state) (8–10). Both microbial community resistance and resilience can vary substantially across soils and disturbances (10). The duration of a disturbance event, whether it is a pulse (*i.e.*, a discrete, short-term event) or press (*i.e.*, a long-term, continuous event), and whether the disturbance is compounded, may differentially affect the stability of the affected microbial community (9, 11). In light of the long-term environmental impacts of human activities such as industrial mining and climate change, microbial community stability under press disturbance is of particular concern in soil microbial and restoration ecology (9, 12, 13).

Soil microbial communities tend to be sensitive (low resistance) to intense press disturbances (9). However, the rate and degree of resilience of these communities can vary widely. The perceived variability in resilience may be due to complex, interacting, and poorly understood ecological and evolutionary mechanisms that determine resilience (9, 10), including the interplay of speciation, selection, and dispersal assembly processes (14, 15). Quantifications of microbiome resilience may also be complicated by the substantial pool of dormant microorganisms that persists in many environmental systems, but especially in soils (16).

Dormancy is a reversible state of extremely low metabolism (Lennon and Jones) that can allow bacteria to survive under poor conditions or during disturbance (17–19). Some dormant bacterial cells can remain viable for extended periods, and some can be selectively revived given conducive growth conditions (17, 18). Because ecosystem response to disturbance reflects the functions of active microbiome members (9, 20), it is important to understand and account for the dynamics of active members as a more direct predictor of ecosystem resilience and functional recovery.

The seven-decade-old coal mine fire in Centralia, Pennsylvania, USA, provides a unique case study to examine the resilience of soil microbial communities to an intense, long-term thermal disturbance (21). The temperatures of the surface soils above the coal seams that have active fire fronts are continuously higher than ambient soil temperatures (22). Additionally, coal combustion products percolate through the soil and open vents in the ground, altering soil chemistry at the surface (22–24). The gradual advancement of the fire along the coal seams allows for direct comparisons across a gradient of disturbance intensities that we defined by soil temperature and across fire-affected, decades-old recovered, and unaffected reference soils, all within a geographically localized area encompassing less than 0.35 km^2 (21). Furthermore, as disturbance intensity changes over time at a sampling site, community changes can be tracked and understood in the context of the dynamic warming and cooling that define this press disturbance.

The overarching goal of this study was to understand the capacity of soil bacterial communities to recover from an intense press disturbance and to identify underlying ecological mechanisms that support resilience. Our previous work investigated microbiome responses over the Centralia fire-impact gradient. It showed that the microbiomes of recovered soils converged

towards reference soils, thus demonstrating resilience, while the microbiomes of fire-affected soils maintained distinct structures and were highly variable (25). We subsequently conducted a follow-up laboratory study to control dispersal and to track activation of the Centralia soil microbiome to heating and found that both reactivation from dormancy and immigration together contributed to the observed resilience (26). These results motivated us to track activation dynamics in Centralia to understand the contributions of active members to the community disturbance response.

Here, we present a 7-year longitudinal study of the "whole" bacterial communities and the subset of active members (27, 28) annually sampled from our original fire-impact gradient of fire-affected, recovered, and reference sites in Centralia (Fig. 1A, Table S1). Whole bacterial communities were assessed using 16S rRNA gene amplicons, which include contributions of active, dormant, and recently deceased cells. The active subset of the whole community was determined by filtering the dataset to include only the taxa that met a conservative, population-level activity criterion based on their 16S rRNA to 16S rRNA gene ratios. Then, we applied several specific methods to quantify microbiome recovery, particularly time-lag and trajectory analysis (29, 30), where resilience is indicated by the slope of a time-lag regression and by the strength of directionality in trajectory analysis (e.g., toward an undisturbed state). Thus, we expected the strength of directionality and time-lag regression slope to correlate positively with disturbance intensity. Directionality and time lag were also compared across fire categories and between the whole and active partitions to determine their relative resilience. Building from previous findings (19, 25), we hypothesized that soil bacterial communities are resilient to long-term heat disturbance and that the observed resilience is largely attributable to directional

changes in the active populations, indicative of environmental selection and turnover as soils cool.

Results

Soil temperature and pH vary over fire classification

As expected, fire-affected soils were generally hotter than recovered and reference soils. Notably, the fire-affected soil temperatures decreased over the seven years, yet no consistent changes over time were observed in either recovered or reference sites (Fig. 1B, Table S2; linear mixed effects models). Soil pH did not change over time (Fig. S1A, Table S2; linear mixed effects models), though it did overall differ across fire classifications (ANOVA: DF = 2, F = 14.54, p-value < 0.001). pH was lower in recovered sites than either reference or fire-affected sites, and the pH range across fire-affected sites was large (Fig. S1B, posthoc Tukey's: p-values < 0.05).

Bacterial diversity changes across fire classification and over time

We used 16S rRNA gene and 16S rRNA amplicon sequencing of co-extracted DNA and reverse-transcribed RNA to assess community structure across our samples. DNA and RNA sequences were processed identically with shared operational taxonomic units (OTUs) generated *de novo* with 97% sequence identity. Non-bacterial OTUs were removed, and the reads were rarefied to the lowest depth within the analysis: 161,171 for DNA-only analysis (*i.e.*, whole community), 100,491 for DNA and RNA analysis (*i.e.*, active community subset).

To examine how bacterial communities varied over time within our sites, we examined the within-site phylogenetic diversity (*i.e.*, Faith's phylogenetic diversity). We found increased

diversity over time in fire-affected soils for both whole communities and active subsets but no increase in recovered or reference soils (Fig. S2; Table S3; linear mixed effects models). When time and fire classification were ignored, diversity increased as soil temperature decreased for whole communities and active subsets (Fig. 1C and D, Table S4; linear mixed effects models).

We next evaluated the variation in phylogenetic and compositional diversity between sites (*i.e.,* abundance-weighted UniFrac distance). Variation across whole bacterial communities was explained by fire classification, time, and their interaction (Fig. S3, Table 1; PERMANOVA blocked by Site ID). Fire classification explained the most variation (18%), while time and the interaction of fire classification and time explained about 6% and 4%, respectively. *Post hoc* tests within each fire class showed that time only explained variation in community structure in fire-affected sites (Fig S4, Table S5). *Post hoc* tests within each year showed fire classification explaining structural variation in all years, but the variation explained (R^2) decreased between 2015 and 2018 (Fig. S5, Table S5). Structural variation between fire classifications could be attributed to differences in beta-dispersion across fire classifications (ANOVA; df = 2, F = 34.722, p-value < 0.001). Dispersion across fire-affected soil was higher than recovered and reference soils (Table S6).

For the active subset of the bacterial community, fire classification and time, but not their interaction, explained variation in community structure (Fig. S6, Table 1; PERMANOVA blocked by Site ID). As for the whole community, fire classification explained more variation in community structure than time (12% and 6%, respectively). Similarly, beta dispersion varied across fire classifications for the active community subset (ANOVA; df = 2, F = 17.536, p-value < 0.001). Dispersion across fire-affected soil was again higher than recovered and reference soils, though the effect was smaller than for the whole community (Table S6).

Constrained correspondence analysis (CCA) demonstrated that the measured edaphic factors, particularly soil temperature, pH, and pH-associated properties (*e.g.*, Ca, Mg, and Fe concentration), explained a large portion of the variation in community structure for both whole communities (Fig. S7A; variation explained = 45.342%) and active subsets (Fig. S7B; variation explained = 35.789%). Variation between fire-affected and both reference and recovered sites was largely associated with soil temperature, as expected. In contrast, variation across and within the reference and recovered sites were primarily associated with pH (Fig. S7).

Heat intensity drives community trajectories

For both the whole bacterial communities and active subsets, time-lag analysis showed increasing community dissimilarity with an increasing difference by year for fire-affected sites but no relationship for either recovered or reference soils (linear mixed effects models; Fig. 2A and B, Table S7). Using maximum temperature as an indicator of overall disturbance intensity, we found that the slopes of the time lag relationships are positively correlated to the maximum soil temperature of the sites (Fig 2C and D; linear regression; whole community: $R^2 = 0.67$, p-value < 0.001, active subset: $R^2 = 0.30$, p-value = 0.011). Notably, the relationship between the time lag slope and maximum soil temperature was stronger for the whole community than the active subset.

Similar to the time-lag analysis, for both the whole communities and active subsets, there was higher directionality in the fire-affected sites than either recovered or reference sites, with no significant difference between the latter classifications (Fig. 3A and B; Kruskal-Wallis rank sum test; whole community: chi-squared = 9.04, df = 2, p-value = 0.01, active subset: chi-squared = 8.67, df = 2, p-value = 0.01; *post hoc* Dunn's tests). Community directionality was positively

correlated with maximum soil temperature, with a stronger relationship for the whole communities relative to the active subsets (Figs. 3C and D; linear regression; whole community: $R^2 = 0.55$, p-value < 0.001, active subset: $R^2 = 0.20$, p-value = 0.036).

To assess whether the bacterial communities were becoming more like the undisturbed reference communities as soils cooled, we tested whether the distance in community structure between disturbed soils (*i.e.*, fire-affected and reference soils) and recovered soils within years was associated with disturbance intensity, defined here as the difference in soil temperature to reference soil. We found that for both whole communities and active subsets, as the disturbance intensity decreased, the communities were less dissimilar (Fig 4; linear mixed effects models). However, the y-intercept was higher for the active subsets than whole communities, suggesting a greater difference in community structure even when disturbance intensity is null.

Active members contribute little to structural shifts in the whole community

We calculated the percentage of Bray-Curtis dissimilarity between years within individual sites attributable to the dynamics of active OTUs. To examine the influence of OTUs that become inactive or newly active over time, we ran this analysis twice using OTUs active in either the earlier or the later year of year pairing, respectively. Bray-Curtis dissimilarity was used in this case, rather than abundance-weighted UniFrac, because it can be effectively partitioned to a subset of OTUs as previously demonstrated (31). Surprisingly, for the majority, less than 50% of the structural dissimilarity in the whole community over time was attributable to active OTUs. For fire-affected sites, but not for recovered or reference sites, the dissimilarity attributed to active taxa within a site over time decreased as the period between sampling years increased (Fig. 5, Table S8; linear mixed effects models; p-values < 0.05). This trend was statistically

supported when selectively partitioning taxa active at the earlier and later timepoint in the year-pairing.

Interannual OTU activity dynamics vary across fire-affected and unaffected soils

We next examined the breakdown of OTUs into distinct yearly activity patterns (*i.e.,* sets of years in which they were identified to be active; Fig. 6 and Figs. S8 and S9). Only OTUs active in at least five samples were included in this analysis, and activity within a year was defined as activity in at least one site. Furthermore, reference and recovered sites were categorically combined as minimal change over time in active community subsets was observed for these microbiomes. For fire-affected sites (Fig. 6A), the richest activity patterns (*i.e.,* those encompassing the most OTUs) tended to have consistent activity in later sampling years (*e.g.,* 2018-2021) but not for early sampling years (*e.g.,* 2015-2017). This trend suggests that many observed OTUs may become active in later years for fire-affected sites, corresponding to cooling soil temperatures and recovery. For reference and recovered soils (Fig. 6B), the richest activity patterns had no clear annual progression or delineation. This lack of trend suggests more random activity patterns in the soils that experienced ambient temperatures. Activity groups were taxonomically diverse for both cases, with no obvious enrichment in taxa at the phylum level or class level for *Proteobacteria*.

Community assembly processes for whole communities differ across fire classifications

We used the beta-nearest taxon index (βNTI) (32, 33) and the abundance-weighted Raup-Crick index (RC_{bray}) (33) between sites within each year to determine the dominant processes driving whole bacterial community assembly within each fire classification. Due to the

computational requirements for these calculations, DNA samples were first rarefied to 10,000 reads. While the majority of within-year, site-to-site comparisons indicate dominance of homogeneous selection (βNTI < -2), phylogenetic turnover among fire-affected sites tended to be higher than among recovered or reference sites (Fig. 7A; Kruskal-Wallis rank sum test: DF = 2, chi-squared = 68.399, p-value < 0.001; post hoc Dunn test) with many site pairs demonstrating weak selection (-2 ≤ βNTI ≤ 2) or variable selection (βNTI > 2). This pattern was consistent across seven years (Fig. 7D and S10). For these site pairs with weak selection, homogenizing dispersal (RC_{bray} < -0.95), followed by drift alone (-0.95 ≤ RC_{bray} ≤ 0.95), tended to be the dominant assembly process (Fig. 7D).

To ask if two of the most notable edaphic soil factors were associated with potential selection pressures in this system, we examined linear relationships between βNTI and scaled between-site differences in soil pH and temperature and their interaction between sites agnostic of fire classification. Both edaphic factors showed statistically significant positive relationships to βNTI (Fig. 7B and C, Table S9; linear regression: R^2 = 0.5031, p-value < 0.001). Since we used scaled measures for differences in pH and temperature, we can directly compare their coefficients to examine their relative influence on phylogenetic turnover. We found that the coefficient for the difference in pH (2.63) is much larger than that of the difference in temperature (0.42), suggesting a greater influence of pH on phylogenetic turnover than on temperature. There was no statistically significant interaction between these two factors suggesting unrelated influences of temperature and pH in selection (Table S9).

Discussion

We used long-term microbiome surveys across soils from Centralia, Pennsylvania, USA, to examine the resilience of whole bacterial communities and active subsets impacted by heat press disturbance from a subterranean coal seam fire. Over seven years, we observed general cooling of the fire-affected sites (Fig. 1B), indicating a continuous reduction in disturbance intensity. Cooling was not unexpected, as the fire is predicted to move deeper underground with its advance, following the coal seam incline (22). While heat intensity decreased in most sites, some other fire-induced environmental changes continued. For example, elevated local CO_2 concentrations and steam venting were sometimes observed even when soils were only a few degrees Celsius above ambient temperature (Table S1). While fire disturbance was ongoing, the observable change in intensity provided a unique opportunity to observe the transition and recovery of the microbial communities during both the early stages of recovery (*i.e.*, cooling fire-affected sites) and later stages (*i.e.*, recovered sites).

Temporal trends in bacterial diversity, both within-sample (alpha) and between-sample (beta) revealed strong evidence for the resilience of the microbiomes in fire-affected sites and both the whole communities and in the active subsets (Figs. 1C-D, S2-S7). In comparison, communities in recovered and reference sites were relatively less dynamic and more stable over the years. Applying the concepts from the stability basin analogy for community resilience (29), both directionality and time-lag analyses (Figs. 2 and 3) suggested that the fire-affected soils were shifting from their high-temperature, disturbed community state to a relatively stable, recovered state. Comparatively, the recovered and reference soils maintained a relatively stable state over time. Further, the recovered state for the fire-affected and recovered soils appears to be like the reference soils that did not experience any heating from the fires (Fig. 4). These findings

expand and reinforce the previous results from the single sampling timepoint in Centralia, that also indicated recovery of bacterial communities in Centralia soils (25).

Dispersal is an important microbiome assembly mechanism for soil (34–37). Across our Centralia sites, when selection was weak ($-2 < \beta NTI < 2$), which occurred almost exclusively across fire-affected sites, homogenizing dispersal ($RC_{bray} < -0.95$) was the dominant assembly process (Fig. 6). Laboratory-based manipulations of reference soils from Centralia have demonstrated that dispersal enhances the resilience of microbial communities to heat disturbance as compared to controls without dispersal (26). Dispersal likely explains the increased within-site diversity observed as fire-affected soils cool (*i.e.*, new taxa adding to the local species pool). However, even when dispersal is high, immigrant taxa may not necessarily be active or become established. Rather their DNA may remain extractable either in intact cells (dormant or dead) or as extracellular DNA (eDNA) (38–40). There have been proposed models whereby dormancy potentially aids in microbial dispersal capability (*e.g.*, microbial conveyor belt (41)), though more evidence is needed to support these hypotheses (42). During the 45-week mesocosm experiment that explicitly controlled dispersal (26), we observed that post-disturbance activation of some of the dispersed taxa was possible and contributed to overall resilience. Building from these collective results, we present an updated working hypothesis to integrate dispersal and dormancy processes and suggest one potential mechanism of soil microbiome resilience. We hypothesize that taxa lost to disturbance are progressively replaced by dispersed members from the regional species pool (*i.e.*, undisturbed and recovered soils) as thermal selection weakens but likely maintains dormancy while conditions remain suboptimal. This process effectively rebuilds and accumulates the dormant seed bank but does not contribute immediately to functional recovery. When conditions become favorable, members of the seed bank can activate and

progress along a recovery trajectory. Re-establishing the microbial seed bank after a disturbance provides a management strategy that may be useful in bioremediation and active restoration of soils (43, 44), but much more work is needed to test and understand these proposed mechanisms.

Patterns of OTU activity over time in fire-affected sites (Fig. 6) demonstrated that many bacteria are likely to be activating (either after dispersal or from the local seed bank) in the fire-affected soils over the last few years as the soils have cooled. Of the top 20 interannual activity patterns, in terms of the number of OTUs displaying the same pattern of activity, most tend to have detectable activity only in the later sampling years. In other words, many active taxa become detectably active after one to three years of no detectable activity in fire-affected soil. This is best exemplified by the second most numerous activity pattern that includes activity in the years 2016-2021 but not 2015. In combined reference and recovered sites, on the other hand, the top 20 patterns appear to have relatively random years where activity is not detected. Most of these top 20 activity patterns include activity in the first few sampling years. Upon examination of the phylum-level taxonomic breakdown of these activity patterns, we see that the most numerous patterns are comprised of highly diverse phyla (Fig. 6). This taxonomic diversity suggests that activation and perhaps recovery of the active community subset is indiscriminate across the phylogenetic tree, or at least at high taxonomic levels.

Notably, the temporal changes within the whole community were greater than in the active subset, indicating differences in resilience and recovery between the active and whole communities. For example, time explains a greater portion of the structural variation for whole communities than for active subsets, and no interaction was found between time and fire classification for those active communities (Table 1). Similarly, the relationships between heat intensity (*i.e.*, maximum temperature) and both time lag slopes and structural directionality were

weaker for active subsets than the whole communities. When considering variation in the whole community structure within a site over time, active OTUs tended to explain a smaller percentage of the community dissimilarity as the timescale increased, but only for fire-affected sites (Fig. 5). This important finding indicates that over the long term (*i.e.*, multiple years), changes in the structure of the active community subset accounts for less of the change in overall community structure in recovering soils relative to more stable communities. Thus, putting all these results together, we propose that while the overall bacterial community is largely resilient, the active populations may recover slower. From this conclusion and the evidence of this study, we hypothesize that dispersal enables bacterial taxa from the regional species pool to arrive and possibly establish in the fire-affected soils as they cool, with a portion of immigrants remaining dormant though persistent as disturbance intensity slowly wanes.

pH and associated edaphic properties are well-known drivers of bacterial community diversity and soil assembly (35, 45–48). Apart from temperature, edaphic factors associated with soil pH explained structural differences in the community (e.g., Ca, Mg, Fe, etc.; Fig. 2) (43). Standardized phylogenetic turnover between sites (βNTI), a key indicator of selection (49, 50), correlated to differences in soil pH (Fig 6), indicating pH-related selection pressures across Centralia soils (35, 47). βNTI was also correlated to differences in soil temperature between sites. However, this relationship was weaker than that of soil pH, and the interaction between these edaphic factors was not statistically significant, suggesting that the selection pressure of these factors differs (Table S9). From these trends in phylogenetic turnover, we propose that soil pH and related factors are selection drivers maintaining bacterial communities across soils in Centralia but that active coal fires alter the dominant assembly processes either by imposing alternate selection pressures (*e.g.*, selection for thermotolerant traits) or by increasing

stochasticity (*e.g.*, dispersal and drift) among sites. Other press disturbances, such as land use intensification, have been shown to interact with soil pH in driving variation in soil bacterial community assembly (35). The acidic soils in the recovered and some fire-affected sites are likely the result of the deposition of combustion and mining products such as ammonium, nitrate, or sulfur (23–25). Soil acidification may be a legacy effect of mining (51, 52) and coal fires in this area, potentially creating alternative stable states among the recovered communities rather than a comprehensive, single recovery state (53).

Other mechanisms that are difficult to disentangle from our hypothesized model may also be contributing to the patterns we observed. First, as observed across seasons from our HOBO temperature monitors in five sites (Fig. S11), cyclical disturbance intensity patterns may produce temporary conditions suitable for the growth or maintenance activity of otherwise fire-sensitive taxa (11). Second, as soils are extremely diverse (54) non-detection of rare taxa is a distinct possibility despite our deep sequencing, as demonstrated by the rarefaction curves (Fig. S12). Third, extracellular DNA (eDNA) may contribute to the divergence in the whole community and active subset dynamics between fire-affected and unheated soils because eDNA in soil degrades more rapidly at higher temperatures (55). However, a large influence of eDNA on the observed interannual patterns is not expected due to the short lifespan of nucleic acids in soils in general.

Broadly, our findings and hypothesized model for bacterial community recovery from press disturbance have implications for understanding ecosystem recovery to large, long-term stress. We interpret the observed relatively high resilience of the whole community with somewhat lower resilience of the active subset to indicate that, while microbial community functional potential may recover rapidly along with the demographic recovery of members lost to the disturbance, the realized activity of these communities may take longer to recover because

it is harbored within the dormant pool. As only the active subset of the community performs ecosystem functions and services, it is likely that biologically mediated soil processes, such as biogeochemical cycling, may remain altered for some period after disturbance intensity wanes, especially for functions with low redundancy among populations. If ecosystem function is used to measure ecosystem recovery (12), then whole bacterial community surveys will provide a biased view of the recovery progress. However, our findings provide a positive outlook on the recovery potential of heavily anthropogenically disturbed soils. With the growing importance of restoration ecology in the context of anthropogenetic press disturbances and climate change, understanding the stability and potential for recovery of soil microbial communities will help to predict changes to vital ecosystem functions and develop methods of microbiome restoration.

Materials and Methods

Study site, soil collection, and processing

Centralia is located in Columbia County, Pennsylvania, overlying an anthracite seam of the Buck Mountain coal bed (22). This coal seam has been continuously burning since 1962, with at least two fire fronts moving at about two m/yr. While the coal seam and fire lie about 46 m below the surface, heat from the fire and combustion gasses make their way to the surface, significantly increasing surface soil temperatures in some areas and producing observable steam vents (22). Based on previous experiments, we selected 23 sites along the two main fire fronts (19, 25). Soils were collected from each site from 2015 to 2021 in the first or second week of October. Four sites were removed from the experiment before analysis due to exhaustion of available soil, a landslide, or exceedingly low DNA concentrations in the collected soil.

Where possible, 20 cm deep and 5.1 cm in diameter soil cores were collected using ethanol-sterilized PVC soil corers. When coring was impossible, primarily in vents, the soil was collected with an ethanol-sterilized trowel. Collected soil was homogenized and sieved to 4 mm. Soil for nucleic acid extraction was immediately transferred to a sterile 50 ml falcon tube and flash-frozen in liquid N_2. The time from soil collection to flash freezing was generally under 4 minutes. Flash-frozen soils were stored at -20˚C or on ice while transported to the lab and then stored at -80˚C until nucleic acid extraction. The soils used for soil chemistry analyses were transferred to a Whirl-Pak bag and stored on ice or at 4˚C while transported to the lab. Soil moisture was measured gravimetrically with ~10 g of soil. The Soil and Plant Nutrition Lab at Michigan State University (East Lansing, MI, USA; https://www.canr.msu.edu/spnl) processed all other soil chemistry analyses according to standard protocols. Just before soil collections, soil core temperature and air temperatures were measured using ethanol-sterilized electronic thermometers, and CO_2 venting measurements were measured using an EGM-5 portable CO_2 gas analyzer (PP Systems). Daily soil temperatures were monitored at five sites (Table S1) using HOBO 64K pendant Temperature data loggers (UA-001-64; ONSET) buried ~10cm below the surface near the coring site and recovered each year.

DNA/RNA extraction and processing

DNA and RNA were coextracted from 0.25 g of soil using the Griffiths et al. phenol-chloroform protocol (56), with 0.1 mm zirconium bead containing BeadBug homogenizer tubes (#Z763764; Benchmark Scientific) for bead beating. DNA and RNA aliquots of extracted nucleic acid were stored at -80˚C for further processing. DNA and RNA processing was performed within a month of extraction to prevent excessive RNA degradation. DNA

concentration was determined using Qubit dsDNA BR Assay Kit (#Q32853; Invitrogen). If possible, DNA aliquots were diluted to ~2 ng/µl DNA, and RNA aliquots were diluted to ~200 ng/µl DNA before further processing. RNA concentration was determined using Qubit RNA HS Assay Kit (#Q32852; Invitrogen). DNA was degraded using the TURBO DNA-free kit (#AM1907; Invitrogen) following the rigorous DNase treatment protocol for the RNA aliquot. cDNA was generated from RNA using the SuperScript III Reverse Transcriptase kit (#18080093; Invitrogen) following the standard protocol. Both mock community (57) and nuclease-free water controls were performed for each batch of nucleic acid extraction, and all controls were processed along with samples.

Amplification and sequencing of DNA and cDNA were performed by the Research Technology Support Facility at Michigan State University (East Lansing, MI, USA). Amplification via PCR was accomplished using primers targeting the V4 16S rRNA gene region with 515f = GTGCCAGCMGCCGCGGTAA and 806r = GGACTACHVGGGTWTCTAAT (58). The PCR mixture consisted of 7.5 µl 2X Dream Taq master mix (#K1071; Thermo Scientific), 6.5 µl 0.5 µM primer mix, and 1 µl template DNA or cDNA. PCR was performed using the following protocol: 95˚C for 3 min, 30 cycles of 95˚C for 45 s, 50˚C for 60 s, 72˚C for 90 s, and finally 72˚C for 10 min. During amplification, dual indexing barcodes were added, and samples were pooled into two libraries following cleanup. Sequencing was performed on an Illumina MiSeq with the 2x250 paired-end V2 kit. Triplicate sequencing runs were performed for both DNA libraries, while single runs were performed for both RNA libraries.

Sequence processing

Raw, demultiplexed sequencing reads are available through the NCBI short read archive (BioProject PRJNA973689). Individually for each sequencing run, raw reads were first trimmed using cutadapt (v1.18) (59) to remove remaining primer sequences. Paired-end reads were then merged using PEAR (v0.9.6) (60) with a maximum length of 600 bp and a minimum overlap of 20 bp. Paired reads with a maximum expected error greater than one and/or containing unknown bases were removed using usearch (v10.0.240_i86linux64) (61). Filtered sequences were aligned to the SILVA SEED reference alignment (v132), and poorly aligned sequences were removed along with those containing homopolymers with a length of 8 or greater using mothur (v1.40.4) (62). At this point, quality-filtered sequences from all sequencing runs, including DNA and cDNA libraries, were pooled into a master dataset. Final operational taxonomic units (OTU) were then clustered *de novo* at 97% sequence identity using usearch (61) with chimera removal. Representative OTU sequences in FASTA format are available through figshare (DOI: 10.6084/m9.figshare.23060354). OTU taxonomy was assigned based on the SILVA reference database (v132) using sintax from vsearch (v2.15.2_linux_x86_64) (63) with a probability cutoff of 0.8. Following the removal of non-bacterial OTUs (*i.e.*, *Archaea*, mitochondria, chloroplast, and unclassified domain), a phylogenetic tree was generated using FastTree (v2.1.10) (64) with *Sulfolobus* (X90478) included as an outgroup aligned with ssu-align (v0.1.1) (65). The sample metadata, OTU table, taxonomy table, and phylogenetic tree were combined into a phyloseq (66) object available through figshare (DOI: 10.6084/m9.figshare.23060354). The sequence processing code is available at https://github.com/ShadeLab/Centralia_RNA_DNA_multiyear.

We defined bacterial OTUs based on 16S rRNA gene sequence clusters at 97% sequence identity, including DNA and RNA sequence sets. OTUs were used rather than ASVs to avoid the

potential artificial splitting of bacterial populations into multiple clusters (67, 68) and to account for potential increased errors due to the additional PCR step in reverse transcribing RNA.

Statistical analyses.

All statistical analyses were conducted using R (version 4.1.1) (69). The analysis code is available at https://github.com/ShadeLab/Centralia_RNA_DNA_multiyear. Change in soil temperature and pH over time was examined within fire classifications with linear mixed effects (LME) models with package nlme (70) using either soil core temperature or pH as the response variable, time as the fixed effect, and site ID as the random effect. Testing broad differences in soil pH across fire classifications was accomplished using ANOVA and post hoc Tukey's test.

OTU counts from triplicate libraries were added together for DNA samples, and samples with less than 100,000 reads were removed. RNA samples with less than 10,000 reads were also removed. Before whole community analysis (DNA samples only), entire sites with less than three years of adequate DNA sequencing were removed, and the resulting OTU table was rarefied to 161,171 reads to account for the wide range of sequencing depths. For active bacterial community analysis, DNA samples without a corresponding well-sequenced RNA sample and entire sites with less than three years of adequate DNA and RNA sequencing were removed, and the resulting OTU table was rarefied to 100,491 reads. OTUs found in an RNA sample but not the corresponding DNA sample were given a count of 1 in the DNA sample (27, 28). OTUs with an RNA: DNA ratio ≥ 1 in each paired RNA and DNA sample were then defined as active (27, 28). Further analysis of the active community subset used the modified OTU tables from the DNA samples but only included active OTUs (*i.e.*, zero count wherever DNA and cDNA count was zero or RNA: DNA ratio was < 1).

DNA and cDNA sequence set rarefaction produced 50,242 OTUs across both datasets, 38,020 (75.7%) of which were considered active in at least one sample. For each sample, 50.5 ± 0.6% of these active OTUs could be considered phantom (*i.e.*, cDNA reads found, but no reads in the corresponding DNA samples after rarefaction). However, this phantom percentage drops to 35.6 ± 0.7% when considering unrarefied DNA reads and 2.1 ± 0.1% when considering the entire DNA dataset regardless of the sample (Fig. S13), providing evidence that many phantom taxa within the rarefied dataset are likely biological sequences and not entirely artifacts.

Alpha diversity was measured as Faith's phylogenetic diversity indices. Change in alpha diversity over time was examined within fire classifications with LME models, using the diversity indices as the response variable, time as the fixed effect, and site ID as the random effect. Again, differences in alpha diversity across soil temperatures were examined regardless of fire classification using LME models with the diversity indices as the response variable, soil core temperature as the fixed effect, and site ID as the random effect.

Beta diversity was measured using abundance-weighted UniFrac distance calculated with package phyloseq (66). Variation in community structure across fire classification, sampling year, and their interaction was performed using PERMANOVA blocked by site ID. *Post hoc* analysis of variation in community structure across fire classifications within each year or across years within each fire class was also performed using PERMANOVA blocked by site ID and adjusting p-values for multiple comparisons using the Benjamini-Hochberg procedure. Variation in community structure explained by edaphic factors and temperature was examined using a constrained correspondence analysis with package phyloseq (66). Stepwise model variable selection was accomplished with function ordistep from the vegan package.

Time-lag analysis was performed separately for each fire classification using LME models with samples paired within each site, abundance-weighted UniFrac distance across sites as the response variable, the square root of timepoint difference as the fixed effect, and site ID as the random effect. Slopes of simple time-lag linear regressions for each site were then regressed to maximum soil temperature at each site, regardless of fire classification, using a linear model. Community trajectory analyses were performed using package ecotraj (71, 72). Trajectory directionality was first compared across fire classifications using a Kruskall-Wallis test and *post hoc* pairwise Dunn's tests with p-values adjusted for multiple comparisons using the Benjamini-Hochberg procedure. Trajectory directionality was also regressed to maximum soil temperature at each site, regardless of fire classification, using a linear model.

We calculated the percentage of the Bray-Curtis dissimilarity across sampling time points within each site attributed to active taxa using methods from Shade *et al.* 2014 (31). In this case, whole community OTU tables were those of the DNA samples modified for RNA: DNA ratio analyses, including phantoms and 100,491 read rarefaction. This analysis was run separately for OTUs active in the first timepoint of each timepoint pair (*e.g.*, 2016 when comparing a site between 2016 and 2020) and in the second timepoint of each pair (*e.g.*, 2020 when comparing a site between 2016 and 2020). The percentage of the Bray-Curtis dissimilarity attributed to active taxa was compared across fire classifications using a Kruskall-Wallis test and *post hoc* pairwise Dunn's tests with p-values adjusted for multiple comparisons using the Benjamini-Hochberg procedure. These attributed percentages were also examined across time differences between samplings within each fire class using LME models with the percentage of Bray-Curtis dissimilarity as the response variable, the difference in sampling year as the fixed effect, and site ID as the random effect.

Upset plots were made using custom code based on the R package UpSetR (73). For these plots, reference and recovered sites were combined as an overall “no-fire” group, and all sites within a year were combined for the new fire-affected and no-fire groups. Thus, an OTU was considered active within a specific year if it was active in at least one site within that year. Only OTUs found to be active in at least five years were included in this analysis.

Community assembly processes contributing to the structure of the whole bacterial communities were accomplished using calculations of βNTI, a standardized measure of between-sample phylogenetic turnover indicative of deterministic assembly processes, and RC_{bray}, a standardized measure of species turnover between samples that combined with βNTI is indicative of stochastic community assembly processes (49, 50). Before calculations, the DNA sample OTU table was rarefied to 10,000 reads to reduce computational load. βNTI and RC_{bray} calculations were conducted for site pairs within a year, not across years. βNTI was calculated using the big data calculation method from package iCAMP (74). For site pairs, $\beta NTI < -2$ indicates assembly explained by homogeneous selection, $\beta NTI > 2$ indicates assembly explained by variable selection, and $|\beta NTI| \leq 2$ indicates assembly explained by stochastic process or weak selection. For those site pairs with $|\beta NTI| \leq 2$, RC_{bray} was calculated using code modified from Barnett *et al*. 2020 (35). For these site pairs, $RC_{bray} < -0.95$ indicates assembly explained by homogenizing dispersal, $RC_{bray} > 0.95$ indicates assembly explained by limited dispersal, and $|RC_{bray}| < 0.95$ indicates assembly explained by drift alone. βNTI was performed using the phylogenetic tree including all OTUs regardless of sampling year, assuming all OTUs recovered are present in some capacity within the regional species pool. In contrast, RC_{bray} used separate metacommunity abundance profiles for each sampling year to account for potential yearly changes in the structure of the regional species pool. βNTI values for site pairs within each fire

classification were compared across these classes using a Kruskall-Wallis test and *post hoc* pairwise Dunn's tests with p-values adjusted for multiple comparisons using the Benjamini-Hochberg procedure. βNTI values for site pairs were examined in relation to the differences in soil temperature, the difference in soil pH, and their interaction using a linear model after scaling and centering both factors to make them directly comparable.

Acknowledgments

We thank Tammy Tobin-Janzen for Centralia site consultation and field support and Keara L Grady for superb coordination of fieldwork logistics and organization of multi-year soil procurement and processing. We additionally thank SH Lee, JW Sorensen, TK Dunivin, JL Chodkowski, N Stopnisek, and M Mechan Llontop for their enthusiastic participation in the fieldwork. This work was supported by the U.S. National Science Foundation CAREER award #1749544 to AS.

References

1. E. L. Madsen, Microorganisms and their roles in fundamental biogeochemical cycles. *Curr. Opin. Biotechnol.* **22**, 456–464 (2011).
2. T. R. Turner, E. K. James, P. S. Poole, The plant microbiome. *Genome Biol.* **14**, 209 (2013).
3. R. Mendes, P. Garbeva, J. M. Raaijmakers, The rhizosphere microbiome: significance of plant beneficial, plant pathogenic, and human pathogenic microorganisms. *FEMS Microbiol. Rev.* **37**, 634–663 (2013).
4. A. H. C. van Bruggen, *et al.*, One Health - Cycling of diverse microbial communities as a connecting force for soil, plant, animal, human and ecosystem health. *Sci. Total Environ.* **664**, 927–937 (2019).
5. A. Dubey, *et al.*, Soil microbiome: a key player for conservation of soil health under changing climate. *Biodivers. Conserv.* **28**, 2405–2429 (2019).
6. J. Lehmann, D. A. Bossio, I. Kögel-Knabner, M. C. Rillig, The concept and future prospects of soil health. *Nat. Rev. Earth Environ.* **1**, 544–553 (2020).
7. R. C. Wilhelm, H. M. van Es, D. H. Buckley, Predicting measures of soil health using the microbiome and supervised machine learning. *Soil Biol. Biochem.* **164**, 108472 (2022).
8. S. L. Pimm, The complexity and stability of ecosystems. *Nature* **307**, 321–326 (1984).
9. A. Shade, *et al.*, Fundamentals of Microbial Community Resistance and Resilience . *Front. Microbiol.* **3** (2012).
10. B. S. Griffiths, L. Philippot, Insights into the resistance and resilience of the soil microbial community. *FEMS Microbiol. Rev.* **37**, 112–129 (2013).
11. P. Laurent, G. B. S., L. Silke, Microbial Community Resilience across Ecosystems and

Multiple Disturbances. *Microbiol. Mol. Biol. Rev.* **85**, e00026-20 (2021).

12. J. A. Harris, Measurements of the soil microbial community for estimating the success of restoration. *Eur. J. Soil Sci.* **54**, 801–808 (2003).

13. R. D. Bardgett, T. Caruso, Soil microbial community responses to climate extremes: resistance, resilience and transitions to alternative states. *Philos. Trans. R. Soc. B Biol. Sci.* **375**, 20190112 (2020).

14. M. Vellend, Conceptual Synthesis in Community Ecology. *Q. Rev. Biol.* **85**, 183–206 (2010).

15. D. R. Nemergut, *et al.*, Patterns and Processes of Microbial Community Assembly. *Microbiol. Mol. Biol. Rev.* **77**, 342–356 (2013).

16. Z. T. Aanderud, S. E. Jones, N. Fierer, J. T. Lennon, Resuscitation of the rare biosphere contributes to pulses of ecosystem activity. *Front. Microbiol.* **6**, 24 (2015).

17. S. E. Jones, J. T. Lennon, Dormancy contributes to the maintenance of microbial diversity. *Proc. Natl. Acad. Sci.* **107**, 5881 LP – 5886 (2010).

18. J. T. Lennon, S. E. Jones, Microbial seed banks: the ecological and evolutionary implications of dormancy. *Nat. Rev. Microbiol.* **9**, 119–130 (2011).

19. P. J. Kearns, A. Shade, Trait-based patterns of microbial dynamics in dormancy potential and heterotrophic strategy: case studies of resource-based and post-press succession. *ISME J.* **12**, 2575–2581 (2018).

20. S. D. Allison, J. B. H. Martiny, Resistance, resilience, and redundancy in microbial communities. *Proc. Natl. Acad. Sci.* **105**, 11512–11519 (2008).

21. A. Shade, Understanding Microbiome Stability in a Changing World. *mSystems* **3**, e00157-17 (2018).

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

23. T. Tobin-Janzen, *et al.*, Nitrogen changes and domain Bacteria ribotype diversity in soils overlaying the Centralia, Pennsylvania underground coal mine fire. *Soil Sci.* **170** (2005).

24. C. Janzen, T. Tobin-Janzen, "Microbial Communities in Fire-Affected Soils" in *Microbiology of Extreme Soils*, P. Dion, C. S. Nautiyal, Eds. (Springer Berlin Heidelberg, 2008), pp. 299–316.

25. S.-H. Lee, J. W. Sorensen, K. L. Grady, T. C. Tobin, A. Shade, Divergent extremes but convergent recovery of bacterial and archaeal soil communities to an ongoing subterranean coal mine fire. *ISME J.* **11**, 1447–1459 (2017).

26. J. W. Sorensen, A. Shade, Dormancy dynamics and dispersal contribute to soil microbiome resilience. *Philos. Trans. R. Soc. B Biol. Sci.* **375**, 20190255 (2020).

27. E. Blagodatskaya, Y. Kuzyakov, Active microorganisms in soil: Critical review of estimation criteria and approaches. *Soil Biol. Biochem.* **67**, 192–211 (2013).

28. B. A. W., K. P. J., S. Ashley, 16S rRNA/rRNA Gene Ratios and Cell Activity Staining Reveal Consistent Patterns of Microbial Activity in Plant-Associated Soil. *mSystems* **4**, e00003-19 (2019).

29. K. A. Lamothe, K. M. Somers, D. A. Jackson, Linking the ball-and-cup analogy and ordination trajectories to describe ecosystem stability, resistance, and resilience. *Ecosphere* **10**, e02629 (2019).

30. B. E. Beisner, D. T. Haydon, K. Cuddington, Alternative stable states in ecology. *Front. Ecol. Environ.* **1**, 376–382 (2003).

31. A. Shade, *et al.*, Conditionally Rare Taxa Disproportionately Contribute to Temporal

Changes in Microbial Diversity. *MBio* **5**, e01371-14 (2014).

32. J. C. Stegen, *et al.*, Quantifying community assembly processes and identifying features that impose them. *Isme J.* **7**, 2069 (2013).

33. F. Dini-Andreote, J. C. Stegen, J. D. van Elsas, J. F. Salles, Disentangling mechanisms that mediate the balance between stochastic and deterministic processes in microbial succession. *Proc. Natl. Acad. Sci.* **112**, E1326 LP-E1332 (2015).

34. M. B. N. Albright, J. B. H. Martiny, Dispersal alters bacterial diversity and composition in a natural community. *ISME J.* **12**, 296–299 (2018).

35. S. E. Barnett, N. D. Youngblut, D. H. Buckley, Soil characteristics and land-use drive bacterial community assembly patterns. *FEMS Microbiol. Ecol.* **96**, fiz194 (2020).

36. S. E. Evans, L. P. Bell-Dereske, K. M. Dougherty, H. A. Kittredge, Dispersal alters soil microbial community response to drought. *Environ. Microbiol.* **22**, 905–916 (2020).

37. K. E. Walters, *et al.*, Routes and rates of bacterial dispersal impact surface soil microbiome composition and functioning. *ISME J.* (2022) https:/doi.org/10.1038/s41396-022-01269-w.

38. E. M. Morrissey, *et al.*, Dynamics of extracellular DNA decomposition and bacterial community composition in soil. *Soil Biol. Biochem.* **86**, 42–49 (2015).

39. P. Carini, *et al.*, Relic DNA is abundant in soil and obscures estimates of soil microbial diversity. *Nat. Microbiol.* **2**, 16242 (2017).

40. J. T. Lennon, M. E. Muscarella, S. A. Placella, B. K. Lehmkuhl, How, when, and where relic DNA affects microbial diversity. *MBio* **9**, e00637-18 (2018).

41. M. Mestre, J. Höfer, The Microbial Conveyor Belt: Connecting the Globe through Dispersion and Dormancy. *Trends Microbiol.* **29**, 482–492 (2021).

42. M. J. Choudoir, K. M. DeAngelis, A framework for integrating microbial dispersal modes into soil ecosystem ecology. *iScience* **25**, 103887 (2022).

43. B. Moreira-Grez, *et al.*, Reconditioning Degraded Mine Site Soils With Exogenous Soil Microbes: Plant Fitness and Soil Microbiome Outcomes . *Front. Microbiol.* **10** (2019).

44. W. L. King, T. H. Bell, Can dispersal be leveraged to improve microbial inoculant success? *Trends Biotechnol.* **40**, 12–21 (2022).

45. N. Fierer, R. B. Jackson, The diversity and biogeography of soil bacterial communities. *Proc. Natl. Acad. Sci.* **103**, 626–631 (2006).

46. A. A. Malik, B. C. Thomson, A. S. Whiteley, M. Bailey, R. I. Griffiths, Bacterial Physiological Adaptations to Contrasting Edaphic Conditions Identified Using Landscape Scale Metagenomics. *MBio* **8**, e00799-17 (2017).

47. B. M. Tripathi, *et al.*, Soil pH mediates the balance between stochastic and deterministic assembly of bacteria. *ISME J.* **12**, 1072–1083 (2018).

48. J. B. Yavitt, C. A. Roco, S. J. Debenport, S. E. Barnett, J. P. Shapleigh, Community Organization and Metagenomics of Bacterial Assemblages Across Local Scale pH Gradients in Northern Forest Soils. *Microb. Ecol.* (2020) https:/doi.org/10.1007/s00248-020-01613-7.

49. J. C. Stegen, X. Lin, A. E. Konopka, J. K. Fredrickson, Stochastic and deterministic assembly processes in subsurface microbial communities. *ISME J.* **6**, 1653–1664 (2012).

50. J. C. Stegen, X. Lin, J. K. Fredrickson, A. E. Konopka, Estimating and mapping ecological processes influencing microbial community assembly. *Front. Microbiol.* **6**, 370 (2015).

51. K. C. Rice, J. S. Herman, Acidification of Earth: An assessment across mechanisms and

scales. *Appl. Geochemistry* **27**, 1–14 (2012).

52. H. Wang, J. Zhang, L. Zhang, J. Wang, Z. Xu, Gas Emission and Soil Chemical Properties Associated with Underground Coal Fires, Wuda Coalfield, Inner Mongolia, China. *Nat. Resour. Res.* **29**, 3973–3985 (2020).

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

54. M. R. Rondon, R. M. Goodman, J. Handelsman, The Earth's bounty: assessing and accessing soil microbial diversity. *Trends Biotechnol.* **17**, 403–409 (1999).

55. S. H. Sirois, D. H. Buckley, Factors governing extracellular DNA degradation dynamics in soil. *Environ. Microbiol. Rep.* (2018) https:/doi.org/10.1111/1758-2229.12725.

56. R. I. Griffiths, A. S. Whiteley, A. G. O'Donnell, M. J. Bailey, Rapid Method for Coextraction of DNA and RNA from Natural Environments for Analysis of Ribosomal DNA- and rRNA-Based Microbial Community Composition. *Appl. Environ. Microbiol.* **66**, 5488–5491 (2000).

57. J. Colovas, A. F. Bintarti, M. E. Mechan Llontop, K. L. Grady, A. Shade, Do-it-Yourself Mock Community Standard for Multi-Step Assessment of Microbiome Protocols. *Curr. Protoc.* **2**, e533 (2022).

58. J. J. Kozich, S. L. Westcott, N. T. Baxter, S. K. Highlander, P. D. Schloss, Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl. Environ. Microbiol.* **79**, 5112–5120 (2013).

59. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10 (2011).

60. J. Zhang, K. Kobert, T. Flouri, A. Stamatakis, PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30**, 614–620 (2014).

61. R. C. Edgar, Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).

62. P. D. Schloss, *et al.*, Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**, 7537–7541 (2009).

63. T. Rognes, T. Flouri, B. Nichols, C. Quince, F. Mahé, VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**, e2584 (2016).

64. M. N. Price, P. S. Dehal, A. P. Arkin, FastTree: Computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26**, 1641–1650 (2009).

65. E. P. Nawrocki, "Structural RNA Homology Search and Alignment using Covariance Models," Washington University in Saint Louis. (2009).

66. P. J. McMurdie, S. Holmes, phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* **8**, e61217 (2013).

67. S. I. Glassman, J. B. H. Martiny, Broadscale Ecological Patterns Are Robust to Use of Exact Sequence Variants versus Operational Taxonomic Units. *mSphere* **3**, e00148-18 (2018).

68. P. D. Schloss, Amplicon Sequence Variants Artificially Split Bacterial Genomes into Separate Clusters. *mSphere* **6**, e00191-21 (2021).

69. R Core Team, R: A language and environment for statistical computing (2018).

70. J. Pinheiro, D. Bates, S. DebRoy, D. Sarkar, R Core Team, nlme: Linear and Nonlinear Mixed Effects Models (2020).

71. M. De Cáceres, *et al.*, Trajectory analysis in community ecology. *Ecol. Monogr.* **89**, e01350 (2019).

72. A. Sturbois, *et al.*, Extending community trajectory analysis: New metrics and representation. *Ecol. Modell.* **440**, 109400 (2021).

73. J. R. Conway, A. Lex, N. Gehlenborg, UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* **33**, 2938–2940 (2017).

74. D. Ning, *et al.*, A quantitative framework reveals ecological drivers of grassland microbial community assembly in response to warming. *Nat. Commun.* **11**, 4717 (2020).

Community type	Factor	Degrees of freedom	Sum of Squares	R^2	F value	p-value
Whole community	Fire Classification	2	0.497	0.180	12.811	0.001
	Year	6	0.157	0.057	1.350	0.001
	Fire Classification: Year	12	0.110	0.040	0.474	0.002
	Residual	103	1.997	0.723		
	Total	123	2.762	1.000		
Active community	Fire Classification	2	0.343	0.124	6.793	0.001
	Year	6	0.158	0.057	1.042	0.002
	Fire Classification: Year	12	0.184	0.067	0.608	0.18
	Residual	82	2.068	0.751		
	Total	102	2.753	1.000		

Table 1: PERMANOVA results comparing whole bacterial community or active subset structure across samples based on abundance weighted UniFrac distance blocked by site ID.

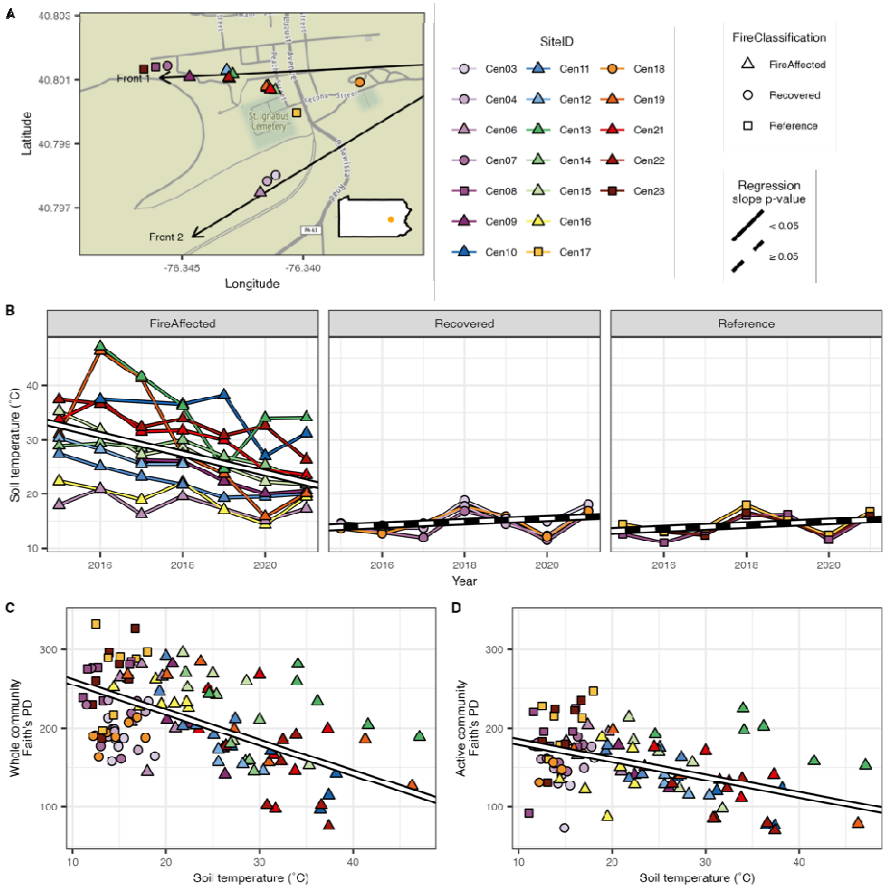

Figure 1: Across the fire-affected sites, soils tended to cool over time, and along with this cooling, bacterial alpha diversity increased. A) Map of collection sites in Centralia, Pennsylvania, USA. B) Tracking of soil temperature (˚C) over the seven sampling years for each site showed general soil cooling in fire-affected sites. C) Whole bacterial community phylogenetic diversity (Faith's PD) decreased as soil core temperature increased. D) Active bacterial community subset phylogenetic diversity decreased as soil core temperature increased. For all plots, colored points and lines indicate site and fire classification. For plots B-D, white and black lines indicate linear mixed-effects regressions using site ID as the random effect, with solid white lines indicating statistically significant slopes (p-value < 0.05) and dashed lines indicating non-statistical significance (p-value ≥ 0.05). Full statistics are found in tables S2-S4, and phylogenetic diversity trends over time are found in Fig. S2.

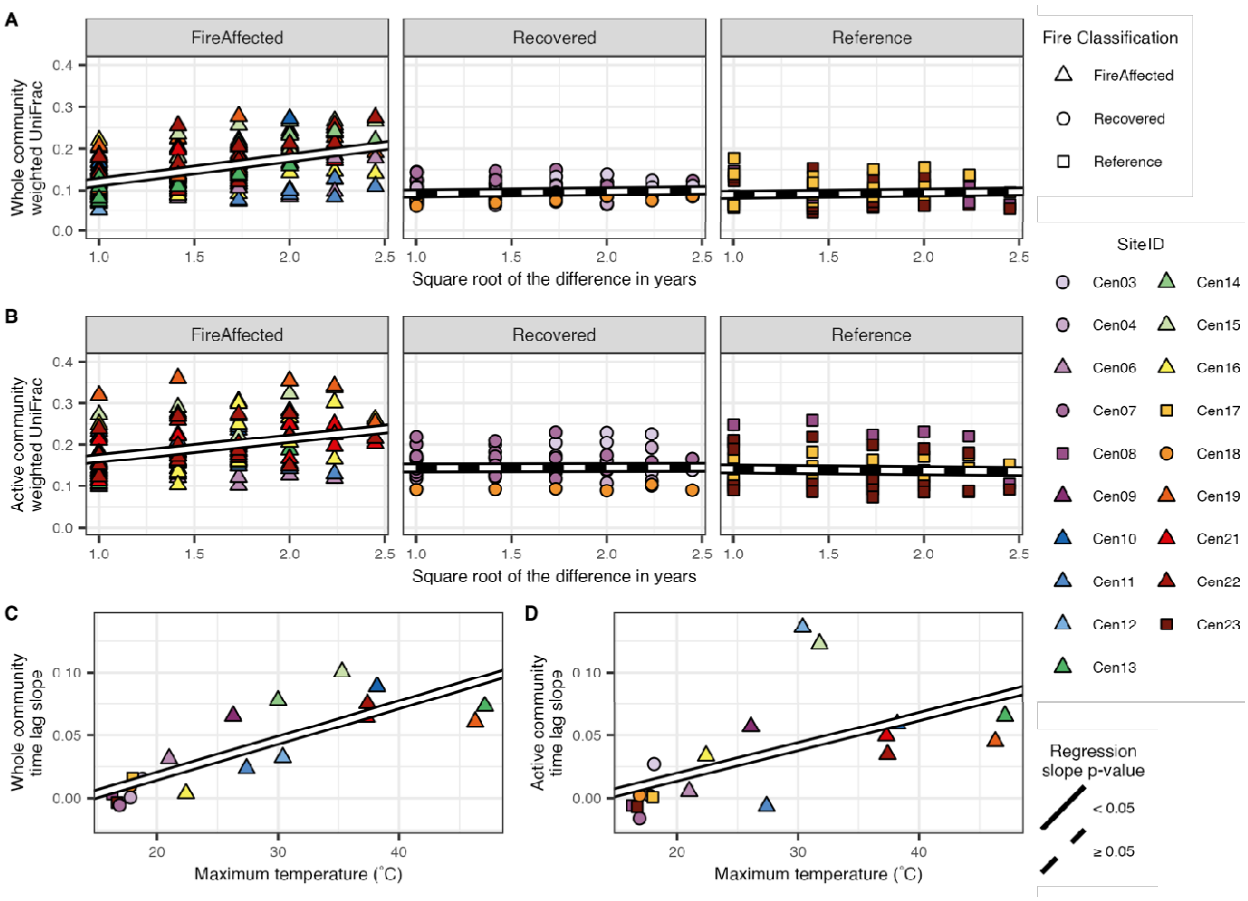

Figure 2: Time-lag analysis demonstrated that variation in both whole bacterial community and active subset structure between samplings of the same site increased as the time difference between sampling increased. However, this trend was only statically significant for fire-affected sites. A) Whole community time-lag analysis. B) Active community subset time-lag analysis. C) Slopes of whole community time-lag regressions made individually for sites increased as maximum soil temperature (proxy of disturbance intensity) increased across sites. D) Slopes of active community subset time-lag regressions made individually for sites increased as maximum soil temperature increased across sites. Points indicate site ID and fire classification. White and black lines indicate linear mixed-effects regressions using site ID as the random effect (A and B) and linear regression (C and D) with solid white lines indicating statistically significant slopes (p-value < 0.05) and dashed lines indicating non-statistical significance (p-value ≥ 0.05). Full statistics for the linear mixed effects regressions are provided in Table S7.

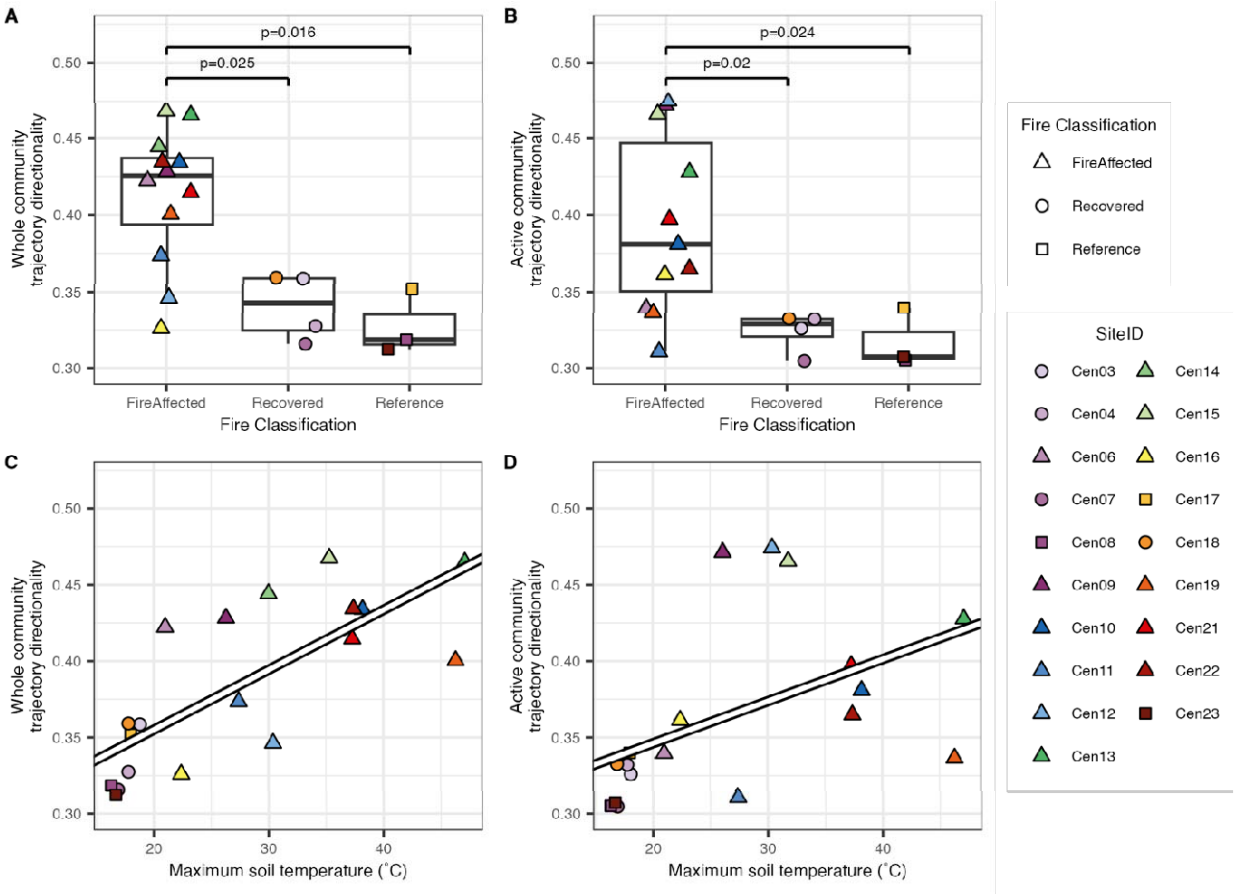

Figure 3: Trajectory analysis demonstrated that directionality in both whole bacterial community (A) and active subset (B) structural shifts over time were generally highest in fire-affected soils compared to recovered and reference soils. Post-hoc pairwise Dunn's test p-values < 0.05 are displayed above statistical significance indicator bars. Directionality in both whole bacterial community (C) and active subset (D) structural shifts increased as maximum soil temperature within sites increased. White lines indicate linear regressions, which had statistically significant slopes (p-value < 0.05). For all plots, points indicate site ID and fire classification.

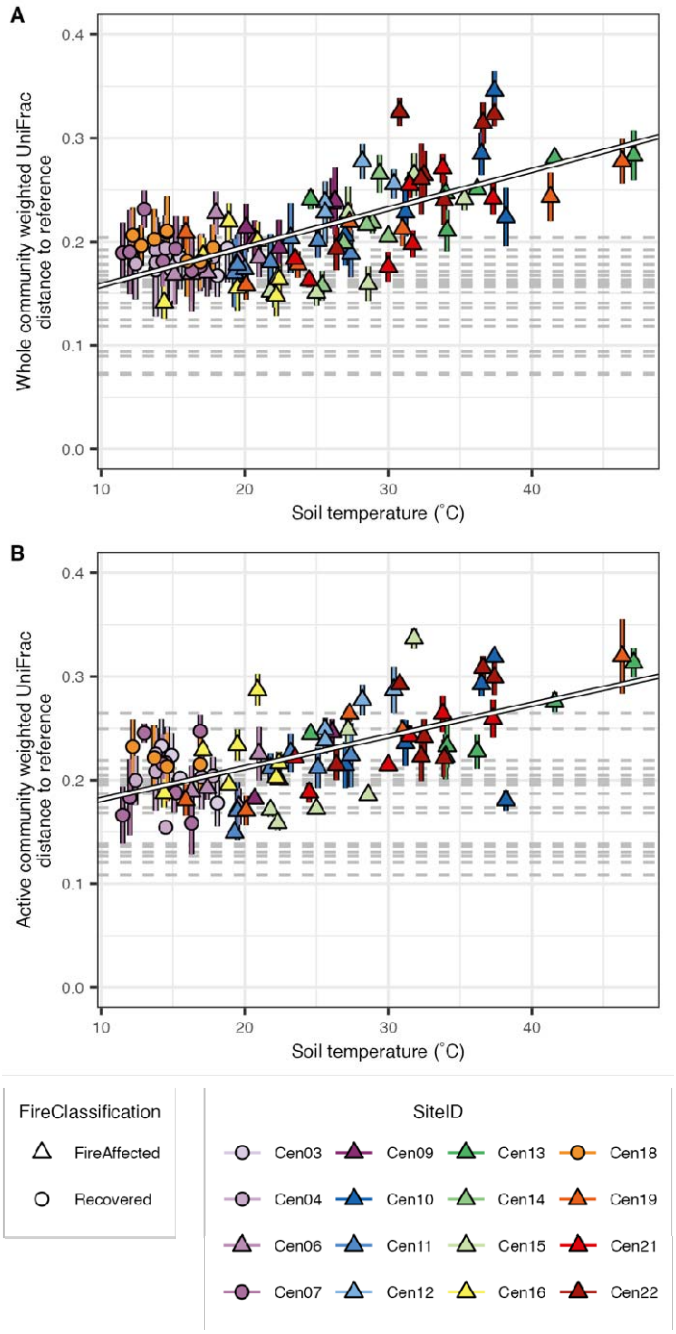

Figure 4: For both (A) whole bacterial communities and (B) active subsets, warmer soils were more structurally distinct from the undisturbed reference soils. Structural dissimilarity was calculated as the mean weighted UniFrac distance between fire-affected and recovered sites and the three reference sites paired within the sampling year. White lines indicate linear mixed-effects regressions using site ID as the random effect, which had statistically significant slopes (p-value < 0.05). Points indicate site ID and fire classification (fire affected or recovered,) and error bars represent the standard error for the distance between a tested sample and the three paired reference samples.

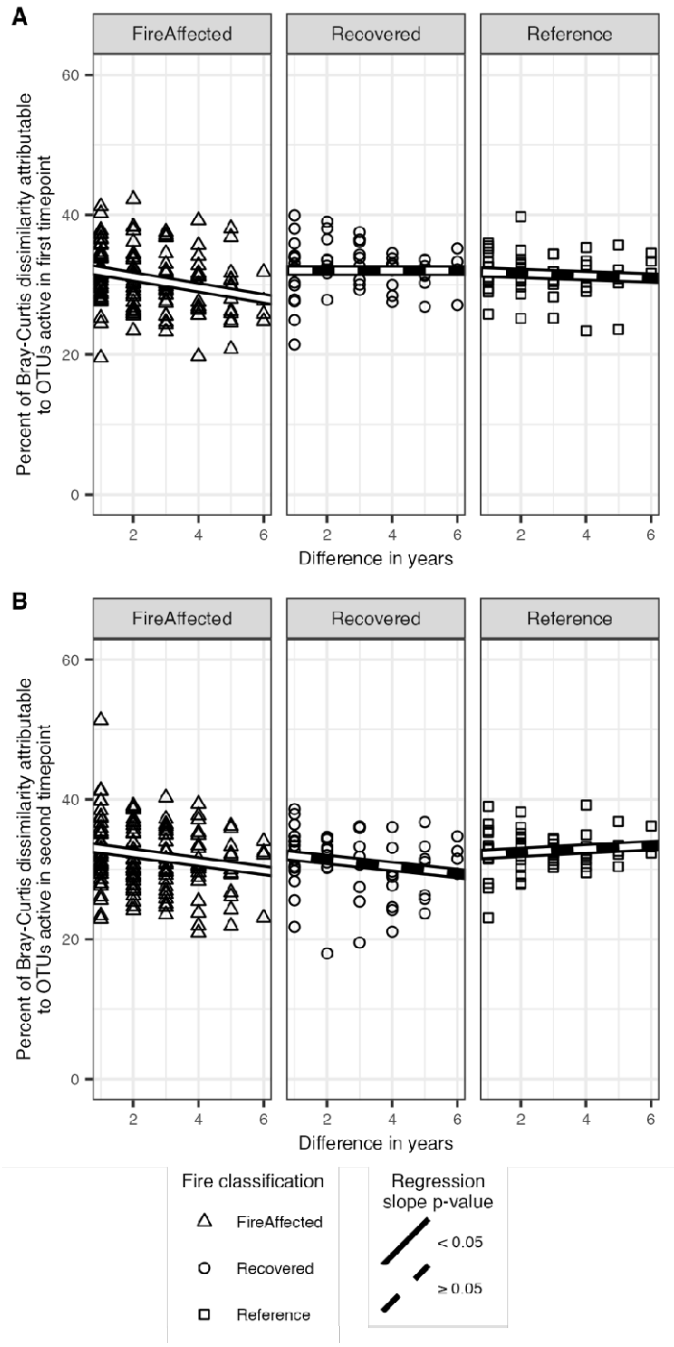

Figure 5: For fire-affected sites, the whole bacterial community structural dissimilarity attributable to the active subset decreases as the time between sampling time points increases. In other words, less of the whole community dissimilarity is attributable to the active subset across larger timescales for fire-affected sites but not for recovered or reference sites. This holds across site pairs for OTUs active in (A) the early timepoint (*e.g.*, 2016 when comparing 2016 and 2020) and in (B) the later timepoint (*e.g.*, 2020 when comparing 2016 and 2020). Structural dissimilarity over time is measured by the Bray-Curtis dissimilarity between years within a site, partitioning OTUs active in either the early or late time points of each pair. White and black lines

indicate linear mixed-effects regressions using site ID as the random effect, with solid white lines indicating statistically significant slopes (p-value < 0.05) and dashed lines indicating non-statistical significance (p-value ≥ 0.05).

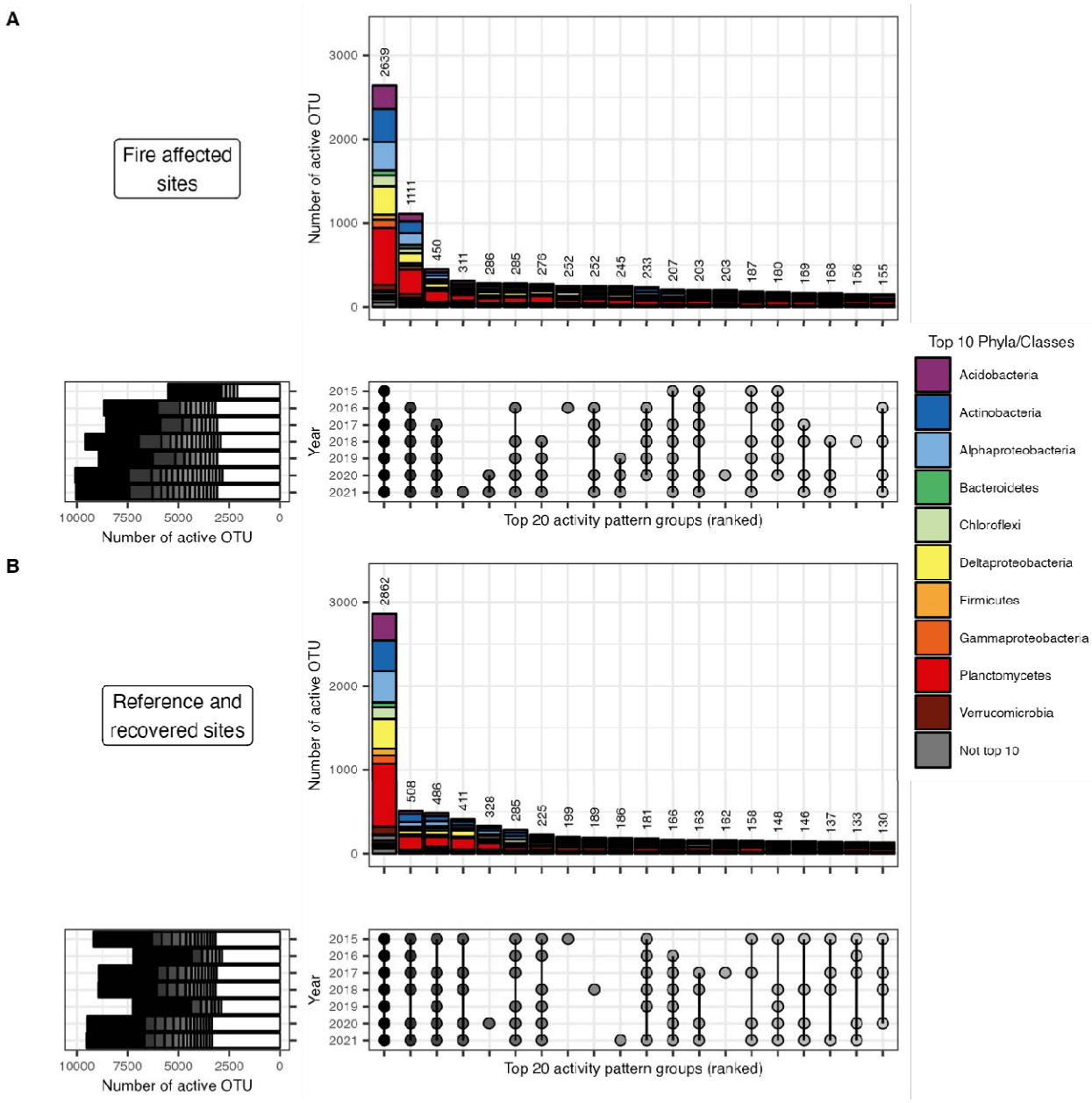

Figure 6: UpSet plots of the top 20 (as determined by OTU membership) activity across years groups for (A) fire-affected sites and (B) combined recovered and reference sites. All sites sampled each year were combined such that OTUs were defined as active if they had an RNA: DNA ≥ 1 in at least one site within that year. The bottom right plots show the activity patterns of the top 20 groups, with dots indicating activity within that year and groups ordered by OTU membership. The top vertical bar chart shows the number of OTUs within each activity group, colored by the top 10 phyla (class for *Proteobacteria*). The bottom left horizontal bar chart

shows the number of active OTUs within each year, colored by their membership in the top 20 activity groups. White sections of the bars indicate OTUs from the other 107 groups (see Figs. S8 and S9 for all 127 activity groups). For the fire-affected sites (A), most of the top 20 activity groups do not have detectable activity in the first few sampled years, while most of the top 20 activity groups for reference and recovered sites (B) have detectable activity during those years.

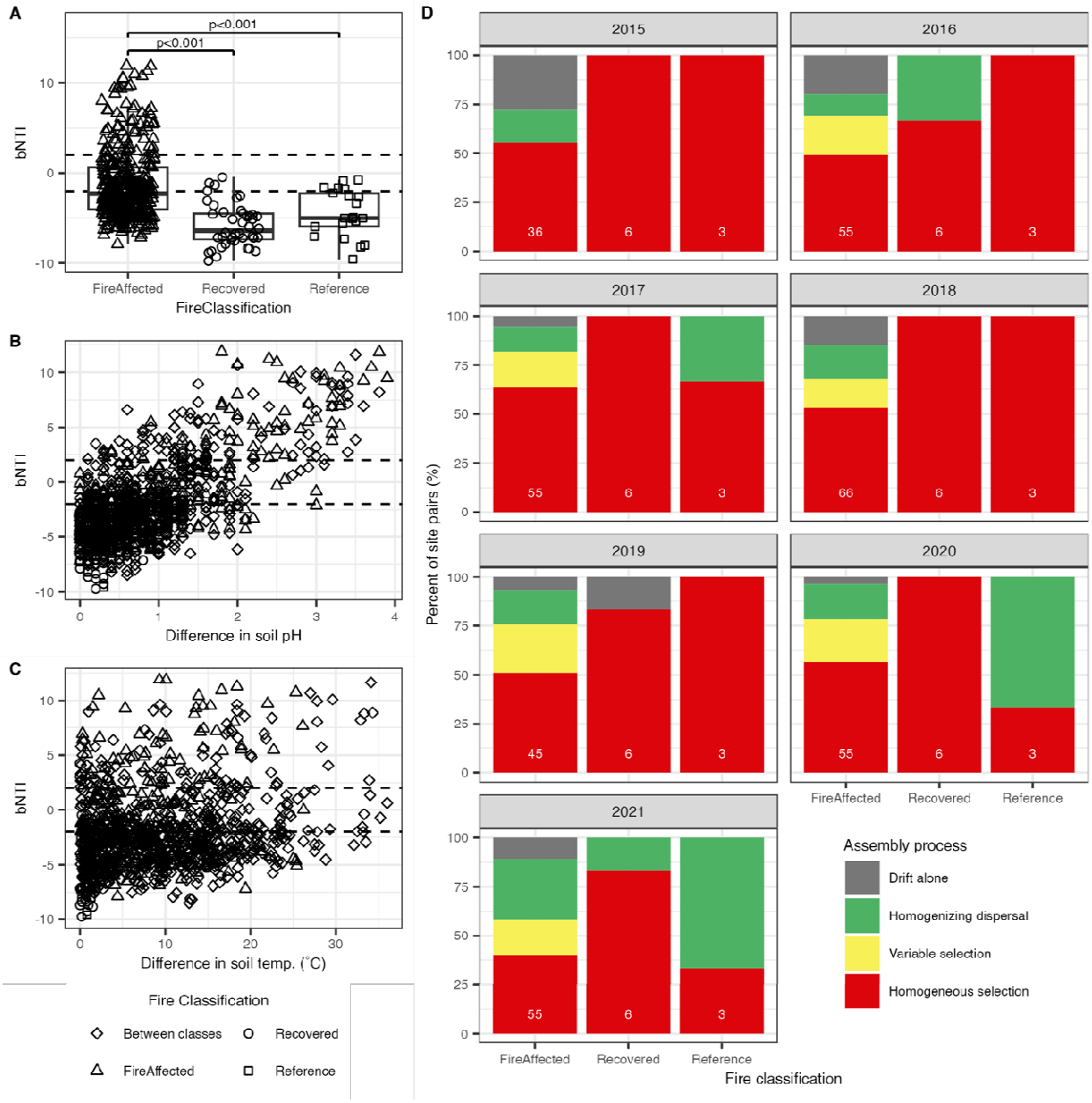

Figure 7: whole bacterial community assembly dynamics across sites as measured with standardized phylogenetic and structural turnover. A) βNTI is generally higher between fire-affected sites than recovered or reference sites whose site pairs almost always have a βNTI < 2, indicating homogeneous selection. Post hoc pairwise Dunn's test p-values < 0.05 are displayed

above statistical significance indicator bars. B) Regardless of fire classification, βNTI tended to increase as the difference in soil pH between sites increased, indicating pH-related selection pressures in these soils. C) Regardless of fire classification, βNTI tended to increase as the difference in soil temperature between sites increased, indicating some temperature-related selection pressures in these soils. Linear regressions, including these two relationships and their interaction, are provided in Table S9. D) Percentage of site pairs within each fire classification and year for which community assembly can be attributed to the five assembly processes examined with βNTI and RC_{bray}. The numbers of site pairs are indicated by white numbers at the bottom of each bar. All site pairs were within each year, not across years.