Soil DNA chronosequence analysis shows bacterial community re-assembly following post-mining forest rehabilitation

Article

Accepted Version

Research Article

Soil DNA chronosequence analysis shows bacterial community re-assembly following post-mining forest rehabilitation

Running headline: **post-mining changes to soil bacterial communities**

Shawn D. Peddle[1], Andrew Bissett[2], Ryan J. Borrett[3], Paul Bullock[4], Michael G. Gardner[1,5], Craig Liddicoat[1,6], Mark Tibbett[7,9], Martin F. Breed[1], Siegfried L. Krauss[8,9]

[1]College of Science and Engineering, Flinders University, Bedford Park, SA, Australia. [2]CSIRO Oceans and Atmosphere, Hobart, Tasmania, Australia. [3]Environmental and Conservation Sciences, Murdoch University, Murdoch, WA, Australia. [4]South32 Worsley Alumina. [5]Evolutionary Biology Unit, South Australian Museum, Adelaide, SA, Australia. [6]School of Public Health, The University of Adelaide, Adelaide, Australia. [7]Department of Sustainable Land Management & Soil Research Centre, School of Agriculture, Policy and Development, University of Reading, Berkshire, United Kingdom. [8]Kings Park Science, Department of Biodiversity, Conservation and Attractions, Perth, WA, Australia. [9]School of Biological Sciences, The University of Western Australia, Crawley, WA, Australia.

Correspondence author: Shawn Peddle, Tel: +61 08 8201 2113, Email: shawn.peddle@flinders.edu.au College of Science and Engineering, Flinders University, Bedford Park, SA, Australia.

Author contributions: SP, MB, SK conceived and designed the research; SP, SK, RB, AB collected the data; SP, RB, CL, MB, SK, AB analysed the data; SP wrote the manuscript with contributions from all other authors.

ORCIDs: SDP 0000-0003- 3464-3058, AB 0000-0001-7396-1484, RJB 0000-0001-8663-0844, MGG 0000-0002-8629-354X, CL 0000-0002-4812-7524, MT 0000-0003-0143-2190, MFB 0000-0001-7810-9696, SLK 0000-0002-7280-6324

Key words: Bauxite mining, completion criteria, ecological restoration, eDNA, microbiome, recovery trajectory, soil biodiversity, soil microbiology

IMPLICATIONS FOR PRACTICE

- Consideration of soil microbiota in mine site rehabilitation and restoration is important for returning functional, self-sustaining biodiverse ecosystems and improving restoration practices.
- Bacterial community variation can be high among reference sites which highlights the need for appropriate sampling design in assessing soil microbial recovery trajectories.
- Our study shows how changes in bacterial communities across a restoration chronosequence can be routinely monitored to provide insights into the recovery of soil microbiota towards restoration targets.

48 **ABSTRACT**

49 Mining activities modify both above- and below-ground ecological communities,
50 presenting substantial challenges for restoration. The soil microbiome is one of these
51 impacted communities and performs important ecosystem functions but receives
52 limited focus in restoration. Sequencing soil DNA enables accurate and cost-
53 effective assessment of soil microbiota, allowing for comparisons across land use,
54 environmental, and temporal gradients. We used amplicon sequencing of the
55 bacterial 16s rRNA gene extracted from soil samples across a 28-year post-mining
56 rehabilitation chronosequence to assess soil bacterial composition and diversity
57 following rehabilitation at a bauxite mine in Western Australia's jarrah forest. We
58 show that while bacterial alpha diversity did not differ between reference and
59 rehabilitated sites, bacterial community composition changed dramatically across the
60 chronosequence, suggesting strong impacts by mining and rehabilitation activities.
61 Bacterial communities generally became increasingly similar to unmined reference
62 sites with time since rehabilitation. Soil from sites rehabilitated as recently as 14
63 years ago did not have significantly different communities to reference sites. Overall,
64 our study provides evidence indicating the recovery of soil bacterial communities
65 towards reference states following rehabilitation. Including several ecological
66 reference sites revealed substantial natural variability in bacterial communities from
67 within a single mine site. We urge future restoration chronosequence studies to
68 sample reference sites that geographically span the restored sites and/or are
69 spatially paired with restored sites to ensure this variability is captured and to
70 improve any inferences on recovery.

INTRODUCTION

The global mining sector is reliant on access to mineral deposits and expansions into intact biodiverse ecosystems (Stevens & Dixon 2017). In Australia, it is estimated that mining has impacted approximately 10 million hectares of land (Grant 2009). Mining activities extensively modify landscapes, directly impacting on both above- (e.g., animal, plant) and below-ground ecological communities (e.g., soil microbiota) (Banning et al. 2011; Stevens & Dixon 2017; Kneller et al. 2018). These often-severe ecosystem impacts present challenges in restoring or rehabilitating biodiverse and functional ecosystems (Doley et al. 2012; Tibbett 2015). Indeed, as the ecological impacts of mining continue to grow, so does the need for improved understanding of how best to repair the damage done.

Restoration projects have tended to focus on recreating aboveground plant communities, often overlooking soil biodiversity (Heneghan et al. 2008; Farrell et al. 2020). However, there is increasing attention paid to soil biodiversity and plant-soil-biota relationships, which has largely been enabled by DNA sequencing technologies (Breed et al. 2019). The important role of soil in ecological restoration has long been known, especially regarding physical and chemical processes such as nutrient cycling and soil formation (Heneghan et al. 2008; Kardol & Wardle 2010). However, soil microbiota (i.e., communities of bacteria, archaea, eukaryotes) and their interactions within the soil system and with aboveground biota have received less attention (Harris 2009; Eisenhauer et al. 2017; Mendes et al. 2019). The question of whether soil microbial communities recover following aboveground revegetation is still unclear, with some observational studies finding a transition towards reference ecosystem states (e.g., Barber et al. 2017; Gellie et al. 2017),

while others have found either that recovery had stalled (e.g., Farrell et al. 2020; Lem et al. 2022) or recovery is dependent on organism or topsoil handling methods (Van der Heyde 2020). This presents a problem since soil microbiota are highly diverse and functionally important ecosystem components and therefore understanding their ecology and responses to both impacts and restoration or rehabilitation is integral to ecosystem restoration (Cameron et al. 2018; Delgado-Baquerizo et al. 2020).

Surface strip mining results in strong and long-lasting impacts on soil biotic and abiotic properties, including decreases in soil microbial activity and organic matter content, and changes in pH and salinity levels (George et al. 2010; Lewis et al. 2010; Sheoran et al. 2010; Banning et al. 2011). These impacts can be driven by the removal and stockpiling of topsoils for extended periods of time, which can expose soils to high temperatures and subsequent drying (Golos & Dixon 2014). Although best practice for the rehabilitation of surface mining is to directly return topsoil, and where necessary, stockpile soil for the shortest time possible (Rokich et al. 2000; Tibbett 2010; Lewis et al. 2010; Spain et al 2015) In reality, topsoils are still routinely stockpiled for extended periods before being used to restore mine sites (Golos & Dixon 2014; Ngugi et al. 2018). While the intent of this ‘direct return’ process is to limit the impact of the mining process on soil properties, how the biological properties of soil respond following direct return of topsoil and subsequent rehabilitation is still unclear.

While the potential use of soil microbiota as an ecosystem indicator is beginning to be explored as part of an interrelated matrix of biotic and abiotic ecosystem

components (Muñoz-Rojas 2018; Tibbett et al. 2019), cause and effect relationships regarding the response of soil microbiota post-rehabilitation and specific drivers of any recovery remains a notable knowledge gap (Lem et al. 2022). A pragmatic approach to begin to understand changes in microbiota with rehabilitation has been to use space as a proxy for time using a rehabilitation chronosequence design (i.e., sampling across a series of similar sites with different times since rehabilitation), and there are examples of this type of study in a post-mining context (Ngugi et al. 2018; Schmid et al. 2020; van der Heyde et al. 2020). Chronosequence studies provide an efficient approach to study the effect of time as an alternative to long-term longitudinal sampling (Walker et al. 2010). However, spatial and temporal confounding factors (e.g., spatial and/or temporal variation in soil, climate and rehabilitation methods), can impact inferences made from chronosequence studies (Pickett 1989; Fleming 1999). With variation of soil microbial communities being so scale dependant (Martiny et al. 2011; Fierer 2017), how spatial variability among reference sites impacts inferences from these chronosequence studies, and particularly regarding rehabilitation targets and completion criteria (Manero et al. 2021), needs to be assessed.

Recent advances in DNA sequencing technologies have enabled improved assessments of whole communities of soil microbiota compared to historical culture-dependent methods (Thompson et al. 2017; Breed et al. 2019; Berg et al. 2020; Nkongolo & Narendrula-Kotha 2020). One such method is to use high-throughput sequencing to generate amplicon datasets, which can be used to compare the diversity and composition of targeted microbial groups (e.g., bacteria) across different environmental conditions, locations, land uses, rehabilitation interventions,

146 or chronosequences to determine how soil microbial diversity and community
147 composition may be impacted (Fierer et al. 2012; Thompson et al. 2017; Breed et al.
148 2019; Tedersoo et al. 2019). Here, we used sequencing of the bacterial 16s rRNA
149 gene from soils collected across a 28-year rehabilitation chronosequence to
150 investigate the recovery trajectories of soil bacterial communities with time since
151 rehabilitation at a bauxite mine site in Western Australia's northern jarrah
152 (*Eucalyptus marginata*) forest. Given the extreme impact of the bauxite mining
153 process on pre-disturbance soils (George et al. 2010), we expect the bacterial
154 communities in the newly rehabilitated sites to be least similar to those of reference
155 sites, with a successional trend of increasing similarity with time. We address the
156 following research questions:

157 1. Does soil bacterial diversity and community composition differ between
158 rehabilitated sites and unmined reference sites?
159 2. How variable are bacterial communities across multiple reference sites that
160 geographically span the mine site?
161 3. Does the soil bacterial community change through the chronosequence with
162 communities in older rehabilitated sites becoming more like those found in
163 reference sites?
164 4. Do changes in soil bacterial communities associate with changes in soil
165 abiotic properties across the chronosequence?

166 Our study improves understanding of changes in soil bacterial communities over
167 time following mine site rehabilitation and helps to enable rehabilitation practitioners
168 to better consider soil bacteria in their interventions. Further, we highlight the
169 variation of bacterial communities across our six reference sites pointing to the need

to account for spatially dependent factors through appropriate experimental design and reference site selection in chronosequence-based restoration studies.

METHODS

Study site and soil sampling

This study was conducted at the Worsley Alumina mine in southwest Western Australia (Fig. 1) where bauxite has been mined since 1984. Mining and rehabilitation work are ongoing, with approximately 5900 hectares of land cleared for mining and 3200 hectares rehabilitated to date. The mine is located in northern jarrah (*Eucalyptus marginata*) forest within the Southwest Australian Floristic Region, an international biodiversity hotspot (Myers et al. 2000). The northern jarrah forest is a dry sclerophyllous open forest or woodland dominated by jarrah (*E. marginata*) and marri (*Corymbia calophylla*) trees with a diverse understory dominated by species from the Fabaceae, Asteraceae, Proteaceae, and Myrtaceae families (Koch & Samsa 2007). Soils within the mine are sandy-gravel, lateritic (high in aluminium and iron), nutrient poor, and slightly acidic. The mine site has a Mediterranean-type climate with dry hot summers and cool wet winters and a mean annual rainfall of 505 mm (Australian Bureau of Meteorology, 2021).

The mining process at this site first involves removal of all vegetation and topsoil, then overburden is stripped away to access the bauxite ore. Long term (>3 months) topsoil storage for rehabilitation is limited where possible. Instead, the preferred practice is for the ‘direct return’ of topsoil from donor locations (e.g., newly mined areas) to a previously mined area. Following bauxite extraction, mined areas are first contoured to reflect surrounding topography using non-ore and gravel material, and

195 then topsoil is spread to a minimum depth of 10 cm before being furrowed and
196 seeded with a mix of local native plant species. This plant species mix has increased
197 from 40 species in 1994 to over 200 by 2015 to better represent the diverse natural
198 floral diversity of the sites prior to disturbance.
199
200 For our study, soil sampling occurred between October and December 2019 as part
201 of the Australian Microbiome (AM) Initiative, following the protocols of the Biomes of
202 Australian Soil Environments (Bissett et al. 2016). Sample sites were chosen, as far
203 as practicable, to provide an even distribution of sampling locations covering the
204 spatial extent of mining activities and sites of varying rehabilitation age (Fig. 1). We
205 were also conscious of the need to evenly distribute sampling locations to limit the
206 effect of spatial autocorrelations. Six uncleared reference sites that were largely
207 embedded within and throughout the mine area were selected to compare with the
208 rehabilitated sites, and to capture natural spatial variation in bacterial communities
209 across the mine site. Restored sites included: two from 1991, four from 1996, two
210 from 1999, two from 2002, two from 2005, one from 2007, three from 2011, and
211 three from 2017 (n = 25 sites in total). Sites rehabilitated in 2017 were rare within the
212 main mine area, forcing samples to be taken from two sites rehabilitated in 2017 and
213 an adjacent reference site, from a spatially separate area approximately 4km away
214 from the main sampling sites.
215
216 In each site, soil was sampled from two depths (0-10 cm and 20-30 cm) where each
217 sample represented a composite from nine subsamples systematically chosen to
218 represent site heterogeneity within 25 x 25 m plots. The nine subsamples from each
219 depth were pooled into a sterile plastic bag, and then homogenised. From each

pooled sample, a 500 g subsample of soil was taken for physicochemical analysis and a 50 mL subsample for DNA extraction. Soil chemical analyses were performed at CSBP Laboratories (Perth, Western Australia) to quantify soil organic carbon, ammonium, potassium, sulphur, calcium, pH, nitrate, and phosphorous. The 50 mL sample was frozen on-site and sent packed on dry ice to the Australian Genome Research Facility (AGRF) in Adelaide, South Australia for DNA extraction (described below). Each replicate had GPS coordinates and a panoramic photograph taken (Fig. S1).

DNA extraction, sequencing, and bioinformatics

DNA was extracted from each sample in triplicate using the Qiagen DNeasy Powerlyzer Powersoil Kit following manufacturer's instructions and quantified fluorometricly. Soil bacterial 16S rRNA was amplified using the 27F (Lane 1991) and 519R (Lane et al. 1985) primer set before sequencing (300bp PE) on the Illumina MiSeq platform. Sequence data used for this work was generated by the Australian Microbiome using their amplicon analysis workflow (Bissett et al. 2016) (https://www.australianmicrobiome.com/protocols/16sanalysisworkflow/) and were downloaded as amplicon sequence variant abundance tables from the AM portal (12 Aug. 2020) (https://www.australianmicrobiome.com/; samples 102.100.100/138358-138407). Paired end reads were merged using Flash2 (Magoč & Salzberg 2011), merged sequences were then further screened to remove those with ambiguities, long homopolymer runs, or too short/long using Mothur screen.seqs (Schloss et al. 2009). Reads passing filter were dereplicated and denoised to zero radius operational taxonomic units (zOTU) using the UNOISE3 algorithm (Edgar 2016) in USEARCH (Edgar 2010). All reads were then mapped to zOTUs to construct a

zOTU by read count table. zOTUs were assigned taxonomy with the RDP Bayesian classifier (Wang et al. 2007) and the SILVA v132 rRNA database (Quast et al. 2013; Yilmaz et al. 2014; Glöckner et al. 2017). zOTUs not classified as “Bacteria” or classified as “Bacteria_unclassified” at the phylum level were discarded, along with those classified “Mitochondria” or “Chloroplast”. zOTUs which did not occur in at least two samples were also discarded to avoid unrepresentative taxa.

Data analyses

R version 4.0.2 (R Core Team, 2020) was used for all downstream statistical analyses. Rarefaction curves were generated comparing observed zOTU richness and Shannon diversity against sample sequence read depth to assess if sample diversity was adequately represented by read depth, as well as to determine an appropriate read depth for rarefaction (Fig.S2). Two samples (one 20-30 cm deep reference site and one 20-30 cm deep 2017 site) were found to have low sequence read depths (80 and 28,854 reads respectively) and were removed from analysis. The remaining samples were rarefied to the lowest remaining sample read depth (n = 54,840 reads) using the *rarefy_even_depth* function in *Phyloseq* (McMurdie & Holmes 2013) to ensure unbiased comparisons across samples. zOTUs that were not present in at least two samples were discarded to avoid non-representative taxa.

Bacterial diversity and community composition

We calculated observed bacterial zOTU richness, and estimated Chao1 richness, Gini-Simpson (Simpson), and Shannon-Weiner (Shannon) diversity indices using *phyloseq* to assess any differences in sample (*alpha*) diversity through the

chronosequence. These diversity data were compared across year of rehabilitation separately for each depth using permuted analysis of variance with the *aovp* function in *lmperm* v2.1.0 (Wheeler et al. 2016) with 5000 permutations.

To explore differences in bacterial community composition across the chronosequence, variation in bacterial community composition (beta diversity) across depth and year of rehabilitation was visualised with non-metric multi-dimensional scaling (NMDS) ordinations of Bray-Curtis distances from the rarefied zOTU abundances using *ordinate* in *phyloseq*. Differences in bacterial community composition across depth and year of rehabilitation were assessed using permuted multivariate analysis of variance (PERMANOVA) implemented with the *adonis2* method in *vegan* (Oksanen et al. 2013). To account for the repeated measure of two depths in soil sampling, we implemented a nested design with our PERMANOVAs with the *setBlock* function to constrain the permutations by a dummy variable accounting for depth as a repeated measure. Homogeneity of group dispersions was tested with the *betadisper* function in *vegan*.

To evaluate the trajectory of bacterial communities in rehabilitated sites towards reference sites and establish how varied bacterial communities are among multiple unmined reference sites, we used Bray-Curtis distances to assess the 'similarity to reference' for each sample. This involved calculating similarity values (i.e., 100%*(1 – distance)), for each sample to all reference samples, including each reference sample to all other references (Liddicoat et al. 2022). The distribution of similarity to reference values across the different years of rehabilitation were then displayed as a series of boxplots. A Kruskal-Wallis multiple comparison test was used to determine

whether the similarity to reference of samples changed with year of rehabilitation, and any significant differences between rehabilitation years were identified using post-hoc Dunn tests with Bonferroni correction to adjust *p* values for multiple comparisons. Heatmaps of the relative abundances of bacterial phyla, class, and order from non-rarefied zOTU data created with the *plot_heatmap()* function in *phyloseq* were used to visualise if any particular taxa were driving changes in community composition through the chronosequence for each depth.

Soil chemical associations

Associations between bacterial community composition and scaled (i.e., mean-centred and divided by the standard deviation) soil chemical variables across the chronosequence were visualised and assessed with constrained correspondence analysis (CCA) with the *ordiR2step()* function in *vegan* separately for each depth. Highly correlated (>0.75) variables were identified (ammonium and potassium at 0-10 cm and calcium at 20-30 cm) and removed using the *findCorrelation()* function in *caret* (Kuhn 2015). Model-selected soil variables were tested for significance with permuted ANOVA with 999 permutations. Nitrate and phosphorous variables were not included in analysis as they returned below-threshold measurements for multiple samples. Differences in each soil chemical variable across the chronosequence were assessed with Kruskal-Wallis tests, and Dunn post-hoc tests with Bonferroni corrections and visualised in a series of scatterplots for each soil depth.

Spatial autocorrelation

We investigated the association between bacterial community composition (using Bray-Curtis ecological distances) and geographic distances between replicates to

test for the presence of spatial autocorrelation. Here, we used Haversine distance matrices for each depth using the *distm* function in *geosphere* (Hijmans et al. 2017), which calculates the distance between every sample based on a spherical land surface from GPS coordinates. The relationship between the Haversine distance matrix and Bray-Curtis distance matrix was examined via a Mantel test in *vegan* using the *spearman* method with 9,999 permutations.

RESULTS

A total of 4,192,984 bacterial 16s rRNA reads were generated across the 50 samples, which spanned the two soil depths across the 28-year rehabilitation chronosequence. There were 70,199 unique bacterial zOTUs identified with a mean of 83,859 ±19,546 SD sequence reads per sample (Table 1). Following quality filtering and rarefaction to the lowest sample read depth of 54,840 reads, 65,098 unique zOTUs remained for analysis across the remaining 48 samples.

Bacterial diversity and community composition

Bacterial community composition varied significantly by soil depth and year of rehabilitation (Fig. 2; PERMANOVA: depth df=1, F=7.170, p=0.005; year df=8, F=2.3462.02, p=0.005). Community composition in rehabilitation sites became increasingly similar to reference sites with time since rehabilitation (Fig. 3). Bray-Curtis similarity to reference values showed significant variation (Kruskal-Wallis: 0-10 cm p<0.001, 20-30 cm p<0.001) and post-hoc Dunn tests with Bonferroni correction, at both the 0-10 cm and 20-30 cm depths (Fig. 3), indicated younger rehabilitation sites were significantly different to reference sites, while older rehabilitation sites were not different to reference sites. The median among reference site similarity

(similarity of each reference site to all other reference sites) was 40% in 0-10 cm soils and 47% in the deeper 20-30 cm soils. At both depths, younger sites that differed to reference sites and had 10-15% lower median similarity to reference values than the among reference site similarity values. Year of rehabilitation had no effect on observed zOTU richness, Chao1 estimated, Simpson, or Shannon diversity metrics at either soil sample depth (permuted ANOVA: $p > 0.05$ in each case; Table 2, Fig. S3). Heatmaps of bacterial phylum, class, and order for both sample depths are presented as supplementary data in Figures S6-S11.

Soil chemical associations

At the 0-10 cm depth, CCA model selection indicated bacterial community composition associated with both pH, which decreased with age, and organic carbon, which increased with age (Fig. 4, Fig. S4). Tests of significance of the terms indicated by CCA with showed no evidence of significance for pH (permuted ANOVA: df=1, F=1.205, p=0.062) but strong evidence for organic carbon (permuted ANOVA: df=1, F=2.07, p=0.001). Although not associated with changes in bacterial communities across the chronosequence, calcium, potassium, sulphur, and ammonium all saw increases with age in the 0-10 cm soil profile (Fig. S4). At the deeper 20-30 cm depth pH was the only CCA model selected soil variable that associated with bacterial communities across the chronosequence (permuted ANOVA: df=1, F=1.349, p=0.014, Fig. 4) and pH decreased with age (Fig. S4). No soil chemical variable significantly varied by year of rehabilitation at either sample depth following Bonferroni corrections for multiple tests (p>0.05 in all cases, Fig. S4).

Spatial autocorrelation

370 Analysis of Bray-Curtis ecological distances representing bacterial community
371 composition, and the geographic distances between samples showed a significant,
372 though weak, spatial autocorrelation (Mantel test: r=0.231, p=0.012; Fig. 4a)
373 indicating that geographic distance between samples associated with differences in
374 bacterial community composition. To explore if this spatial autocorrelation was being
375 driven by three sites that were geographically separate from all other sites (Fig. 1),
376 we removed these and reran the Mantel test which resulted in no significant
377 correlation (Mantel test: r=0.081, p=0.162; Fig. 4b).
378
379 **DISCUSSION**
380 Here we quantified variation in soil bacterial communities across a 28-year
381 rehabilitation chronosequence following rehabilitation of a bauxite mine site in
382 Western Australian jarrah forest. There was a clear association of bacterial
383 community composition with age of rehabilitation, where older sites were more like
384 reference sites than younger sites. In the shallow soils (0-10cm), we found strong
385 evidence of bacterial community composition in sites rehabilitated as recently as
386 2002 (17 years old) being as similar to the reference sites as the reference sites
387 were to each other. In the deeper soils (20-30cm), this trajectory towards reference
388 site bacterial community composition appeared somewhat slower, potentially
389 exacerbated by the higher among reference median similarity of 47% compared to
390 40% in the shallow soils. These biologically important trends with increasing age
391 reflect a successional transition in the structure of bacterial communities, where
392 communities in rehabilitated sites increasingly resembled those from unmined
393 reference sites with increased time since rehabilitation. Although community
394 composition was associated with rehabilitation age, we observed no effect of

rehabilitation age or soil depth on bacterial alpha diversity. Our findings show that while the mining process impacts bacterial communities even with direct return of topsoil, these communities can respond rapidly to environmental changes following rehabilitation. This relatively rapid change can provide an early indication of ecosystem recovery trajectories moving toward the reference ecosystem (Banning et al. 2011; Yan et al. 2019). Together, these results indicate that ecologically important soil bacterial communities are on a trajectory towards recovery following rehabilitation techniques applied at the Worsley Alumina bauxite mine.

We observed associations between changes in soil pH with bacterial community structure at both sampled soil depths. Globally, soil pH is among the strongest drivers of soil bacterial community composition at local and broad spatial scales (Fierer 2017). However, these effects may not be as clear across narrower ranges of pH. At both depths soil pH decreased with time since rehabilitation, but in the deeper soils pH trended away from the pH of reference sites. This negative association between pH with time since rehabilitation at this soil depth could be impacting on deeper soil bacterial community composition and may be a barrier to future bacterial community recovery. This highlights the importance of targeting ideal site-specific soil pH levels for microbiota in mine rehabilitation and may be an avenue to investigate the potential to shorten recovery time frames by optimising soil pH earlier in the rehabilitation process.

Soil organic carbon also associated with bacterial communities in the 0-10 cm soil samples, increasing with time since rehabilitation and becoming more like the reference sites. Like pH, soil organic carbon is one of the most important abiotic

420 factors in structuring soil bacterial community composition (Fierer 2017) and soil
421 organic carbon is a key indicator of soil quality (Muñoz-Rojas 2018). Soil organic
422 carbon is expected to accumulate more rapidly in surface soils compared to deeper
423 soils due to the build-up of detritus on the surface. This change in abiotic soil
424 properties is likely to have driven the rapid development of contrasting bacterial
425 community composition depth profiles between our two sample depths to some
426 degree.
427
428 Soil chemical properties have large effects on the composition and diversity of soil
429 bacterial communities (Fierer 2017; Bahram et al. 2018; Delgado-Baquerizo &
430 Eldridge 2019). Although bauxite mining is known to impact soil abiotic properties
431 such as calcium, phosphorous, potassium, and aluminium (George et al. 2010;
432 Lewis et al. 2010), we did not observe significant differences in these variables
433 across different years of rehabilitation in our chronosequence. The absence of
434 differences in these abiotic properties could potentially be explained by decreased
435 impacts of direct soil return procedures that are employed at this site, compared to
436 more common soil stockpiling practices. However, more research is required to
437 better understand the factors that are driving the observed soil chemistry variation.
438
439 Although depth explained more variation in bacterial community composition than
440 time since rehabilitation, the recovery trajectory with time is similar across both
441 depths with bacterial communities in rehabilitated sites becoming increasingly like
442 reference sites with time. Even with the homogenisation of soil that occurs with direct
443 return or storage of topsoils, our youngest sites still developed depth profiles in as
444 short as two years following rehabilitation. This stratification of bacterial composition

445 across soil depth is thought to result from differential availability of macronutrients
446 and organic carbon and/or differing environmental gradients across soil depths
447 (Allison et al. 2007), and both these trends are supported by our results that
448 generally show lower nutrient levels in the deeper soils. Our results support recent
449 soil genomic research that show variation in bacterial community composition
450 between soil depths as well as directional trends of community composition with time
451 since rehabilitation (Gellie et al. 2017; Yan et al. 2019).While these results indicate a
452 recovery trajectory of bacterial communities returning to their pre-disturbance
453 condition with time since rehabilitation, ascertaining whether soil edaphic variables,
454 aboveground plant communities, or other factors are specifically driving this recovery
455 is still unclear.
456
457 While chronosequence studies can be used to examine the effect of time following
458 rehabilitation, changes to rehabilitation practices over time such as soil handling or
459 revegetation seed mixes can confound conclusions from these studies. The northern
460 jarrah forest where our site is situated is characterised by an overstory dominated by
461 jarrah (*E. marginata*) and marri (*Corymbia calophylla*) tree species, and most of the
462 regions floristic diversity is found in the understory and groundcover (Koch & Samsa
463 2007). Seed mixes at our sites have changed over time, with over 200 species
464 directly seeded in the youngest sites compared to 40 species directly seeded in our
465 oldest sites. While these changing practises introduce confounding factors into
466 chronosequence studies, the direct return of topsoil does also return the native
467 seedbank that will help reduce the impact of differences in species directly seeded.
468 However, in contrast to soil bacterial community trajectories, vegetation communities
469 in recently restored sites more closely resemble reference sites than do older

restored sites (George et al 2009). The cause and consequences of this apparent disassociation between above and below ground trajectories requires further investigation.

In natural soil systems, succession in bacterial communities is thought to be initially stochastic before becoming increasingly deterministic (Dini-Andreote et al. 2015), where soil properties (particularly pH, availability of soil carbon, and nitrogen) and plant-soil feedbacks drive succession (Fierer 2017). Succession in soil bacterial communities in human-altered systems, such as in response to rehabilitation interventions following mining and agriculture, are less understood with only a handful of recent soil genomic studies showing patterns of compositional differences in bacterial communities following rehabilitation (e.g. Barber et al. 2017; Gellie et al. 2017; Ngugi et al. 2018; Yan et al. 2018; Schmid et al. 2020; van der Heyde et al. 2020). None of these previous studies however address the degree of variation in bacterial communities among reference sites, or how this potential variation can impact what we determine to be an appropriate rehabilitation target to which we should be comparing rehabilitated sites against.

Our results clearly indicate bacterial community composition in older rehabilitated sites was as similar to community composition in reference sites as the communities in individual reference sites are to the other reference sites. This similarity however highlights the low degree of among-reference site similarity in bacterial communities and highlights the need for future studies to better consider this high degree of variation. Our reference-to-reference comparison showed a median similarity of 40% at the zOTU level. While this degree of similarity is influenced by the analysis

methods (i.e., ecological distance measures, sequence grouping or clustering approaches, multiplexing, denoising), we do not aim to establish here if zOTU levels of resolution provide the most appropriate indication of progress towards the reference target. We conduct a comprehensive methodological investigation using soil bacterial community data in chronosequence studies that explore these points in detail in Liddicoat et al. (2022).

Here, we sampled six reference sites embedded within both the mine and our rehabilitated sites to provide an indication of the variation present in the bacterial community among reference sites in general. This among-reference site variation confirms expectations from known associations between soil bacterial communities and soil physical and chemical characteristics, and how both these factors can vary spatially (Green & Bohannan 2006; Neupane et al. 2019). Previous studies using a chronosequence design to explore changes in soil microbial communities following mine site rehabilitation have only sampled limited numbers of reference sites (e.g., 1-3 sites) (Ngugi et al. 2018; Schmid et al. 2020; van der Heyde et al. 2020) and none have reported on the variation present among reference sites. With spatial variation of bacterial communities being so scale dependant (Fierer 2017), the degree of variation among reference sites will likely impact interpretations of communities being used as the target. We recommend future studies that assess recovery trajectories of soil microbiota with a chronosequence design capture spatial variation among reference sites by sampling many reference sites that geographically span the study site. This reference site selection should be done to ensure an adequate representation of soil and vegetation community heterogeneity across the study site and, where possible, pair rehabilitated sites with a nearby reference site to maximise

similarity between sites and therefore increase the chances of isolating the effect of interest.

While our study design included an even spatial distribution of our rehabilitation sites with reference sites throughout the mine area, we observed a significant positive correlation between bacterial community dissimilarity and geographic distance. This association was largely driven by three sample sites, and when these sites were removed there was no longer a significant association, supporting our conclusion of an effect of time, rather than space, on bacterial communities across the chronosequence. This spatial effect on bacterial community composition is likely to be driven not only by our spatial outliers but also by associations between soil abiotic properties (e.g., pH, potassium or other unsampled soil parameters) and microbial community composition (Martiny et al. 2011). As geographic distance between samples increased, so do changes in soil properties. This environmentally driven spatial variation of soil microbial communities highlights the need for appropriate experimental designs that limit spatial confounders where practicable or address their ecological implications. Furthermore, to experimentally test cause-effect relationships in a rehabilitation context, either experiments need to be embedded into rehabilitation sites (Gellie et al. 2018) or longitudinal studies need to be done to conclusively ascertain temporal changes in soil microbiota (Lem et al. 2022). Also, to investigate any potential return of key bacterial-mediated ecological functions, future studies should incorporate shotgun metagenomic data to directly ascertain changes in functional gene abundances as inferring any functional changes from 16S data alone is problematic (Sun et al. 2020).

While we found no difference in bacterial alpha diversity across the chronosequence we do note this lack of difference may have been caused by our observed zOTU richness not reaching the species asymptote in all samples. Previous studies have shown a variety of bacterial diversity changes with rehabilitation, including higher diversity in younger sites before peaking in moderately aged sites and then diversity reductions towards reference sites (Barber et al. 2017; Sun et al. 2017). These previously published diversity patterns were explained as resulting from an initial disturbance, followed by rapid expansion of generalist and opportunistic taxa, before niche specific taxa begin to establish as the vegetation community re-establishes (Kardol & Wardle 2010; Liddicoat et al. 2019). However, these trends are by no means universal, and similar to our results, other studies have shown no change or significant differences in soil bacterial alpha diversity attributable to age across a chronosequence (Gellie et al. 2017; Yan et al. 2019; Schmid et al. 2020). These discrepancies in the response of soil bacterial alpha diversity to rehabilitation makes predicting a response of soil bacterial diversity *a priori* difficult. Soil microbial diversity has been shown to have links to aboveground biodiversity and ecosystem services and functions (Fierer et al. 2012; Bardgett & Van Der Putten 2014; Prober et al. 2015; Bender et al. 2016). However, higher diversity does not necessarily reflect greater ecological integrity than lower diversity, and neither does it imply greater or improved functionality (Shade 2017). To assess any change in bacterial functions, future studies would benefit from incorporating assessments of functional gene abundances through time following restoration. The initial topsoil disturbance in mining and any prolonged topsoil storage can negatively impact on soil microbial diversity, potentially reducing functionality. In this case however, the mine's direct

569 return approach has potentially limited the impact of degrading processes on soil
570 microbiota reducing impacts on soil diversity.
571
572 In conclusion, by using high-throughput amplicon sequencing of the bacterial 16s
573 rRNA gene, we show a clear recovery trajectory in soil bacterial communities
574 following post-mining rehabilitation as well as high variability among reference sites.
575 This among-reference variability highlights the need for restoration chronosequence
576 studies to sample several reference sites that geographically span the rehabilitation
577 site and/or are spatially paired with rehabilitated sites to improve inferences of a
578 recovery trajectory. Our results provide further evidence of the association between
579 soil pH and bacterial community composition and suggest further research is needed
580 to determine if recovery timeframes can be improved by modifying soil pH early in
581 the rehabilitation process. Overall, our study provides a robust perspective of how
582 environmental DNA can be used as a monitoring tool within an improved
583 chronosequence design to assess the recovery trajectory of degraded ecosystems
584 following restoration interventions.

Acknowledgements:

We would like to acknowledge the contribution of the Australian Microbiome consortium in the generation of data used in this publication. The Australian Microbiome initiative is supported by funding from Bioplatforms Australia and the Integrated Marine Observing System (IMOS) through the Australian Government's National Collaborative Research Infrastructure Strategy (NCRIS), Parks Australia through the Bush Blitz program funded by the Australian Government and BHP, and the CSIRO. We would like to acknowledge the contribution of the Biomes of Australian Soil Environments (BASE) consortium in the generation of data used in this publication. The BASE project was supported by funding from Bioplatforms Australia through the Australian Government National Collaborative Research Infrastructure Strategy (NCRIS). We thank Research England grant "Policy change to halt biodiversity loss and restore sustainable ecosystems after mining" who helped support MT. We also thank Janet Anthony and Luisa Ducki for assistance with soil sampling and Reghan Mann at South32 for logistic assistance and two anonymous reviewers whose comments greatly improved the manuscript.

REFERENCES

Allison VJ, Yermakov Z, Miller RM, Jastrow JD, Matamala R (2007) Using landscape and depth gradients to decouple the impact of correlated environmental variables on soil microbial community composition. Soil Biology & Biochemistry 39:505-516

Australian Bureau of Meteorology (2021) Climate statistics for Australian locations http://www.bom.gov.au/climate/averages/tables/cw_010917.shtml (accessed Aug. 02, 2021)

Bahram M, Hildebrand F, Forslund SK, Anderson JL, Soudzilovskaia NA, Bodegom PM, Bengtsson-Palme J, Anslan S, Coelho LP, Harend H (2018) Structure and function of the global topsoil microbiome. Nature 560:233-237

Banning NC, Gleeson DB, Grigg AH, Grant CD, Andersen GL, Brodie EL, Murphy D (2011) Soil microbial community successional patterns during forest ecosystem restoration. Applied Environmental Microbiology 77:6158-6164

Barber NA, Chantos-Davidson KM, Amel Peralta R, Sherwood JP, Swingley WD (2017) Soil microbial community composition in tallgrass prairie restorations converge with remnants across a 27-year chronosequence. Environmental Microbiology 19:3118-3131

Bardgett RD, Van Der Putten WH (2014) Belowground biodiversity and ecosystem functioning. Nature 515:505-511

Bender SF, Wagg C, Van Der Heijden MG (2016) An underground revolution: biodiversity and soil ecological engineering for agricultural sustainability. Trends in Ecology & Evolution 31:440-452

Berg G, Rybakova D, Fischer D, Cernava T, Vergès M-CC, Charles T, Chen X, Cocolin L, Eversole K, Corral GH (2020) Microbiome definition re-visited: old concepts and new challenges. Microbiome 8:1-22

Bissett A, Fitzgerald A, Meintjes T, Mele PM, Reith F, Dennis PG, Breed MF, Brown B, Brown MV, Brugger J (2016) Introducing BASE: the Biomes of Australian Soil Environments soil microbial diversity database. GigaScience 5:s13742-13016-10126-13745

Breed MF, Harrison PA, Blyth C, Byrne M, Gaget V, Gellie NJ, Groom SV, Hodgson R, Mills JG, Prowse TA (2019) The potential of genomics for restoring ecosystems and biodiversity. Nature Reviews Genetics 20:615-628

Cameron EK, Martins IS, Lavelle P, Mathieu J, Tedersoo L, Gottschall F, Guerra CA, Hines J, Patoine G, Siebert J (2018) Global gaps in soil biodiversity data. Nature Ecology & Evolution 2:1042-1043

Delgado-Baquerizo M, Eldridge DJ (2019) Cross-biome drivers of soil bacterial alpha diversity on a worldwide scale. Ecosystems 22:1220-1231

639 Delgado-Baquerizo M, Reich PB, Trivedi C, Eldridge DJ, Abades S, Alfaro FD,
640 Bastida F, Berhe AA, Cutler NA, Gallardo A (2020) Multiple elements of soil
641 biodiversity drive ecosystem functions across biomes. Nature Ecology & Evolution
642 4:210-220

643 Dini-Andreote F, Stegen JC, Van Elsas JD, Salles JF (2015) Disentangling
644 mechanisms that mediate the balance between stochastic and deterministic
645 processes in microbial succession. Proceedings of the National Academy of
646 Sciences 112:E1326-E1332

647 Doley D, Audet P, Mulligan D (2012) Examining the Australian context for post-
648 mined land rehabilitation: reconciling a paradigm for the development of natural and
649 novel ecosystems among post-disturbance landscapes. Agriculture, Ecosystems &
650 Environment 163:85-93

651 Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST.
652 Bioinformatics 26:2460-2461

653 Edgar RC (2016) UNOISE2: improved error-correction for Illumina 16S and ITS
654 amplicon sequencing. bioRxiv:081257

655 Eisenhauer N, Antunes PM, Bennett AE, Birkhofer K, Bissett A, Bowker MA, Caruso
656 T, Chen B, Coleman DC, De Boer W (2017) Priorities for research in soil ecology.
657 Pedobiologia 63:1-7

658 Farrell HL, Barberán A, Danielson RE, Fehmi JS, Gornish ES (2020) Disturbance is
659 more important than seeding or grazing in determining soil microbial communities in
660 a semiarid grassland. Restoration Ecology 28:S335-S343

661 Farrell HL, Léger A, Breed MF, Gornish ES (2020) Restoration, Soil Organisms, and
662 Soil Processes: Emerging Approaches. Restoration Ecology 28:S307-S310

663 Fierer N (2017) Embracing the unknown: disentangling the complexities of the soil
664 microbiome. Nature Reviews Microbiology 15:579

665 Fierer N, Leff JW, Adams BJ, Nielsen UN, Bates ST, Lauber CL, Owens S, Gilbert
666 JA, Wall DH, Caporaso JG (2012) Cross-biome metagenomic analyses of soil
667 microbial communities and their functional attributes. Proceedings of the National
668 Academy of Sciences 109:21390-21395

669 Fleming RA (1999) Statistical advantages in, and characteristics of, data from long-
670 term research. The Forestry Chronicle 75:487-489

671 Gellie NJ, Breed MF, Mortimer PE, Harrison RD, Xu J, Lowe AJ (2018) Networked
672 and embedded scientific experiments will improve restoration outcomes. Frontiers in
673 Ecology and the Environment 16:288-294

674 Gellie NJ, Mills JG, Breed MF, Lowe AJ (2017) Revegetation rewilds the soil
675 bacterial microbiome of an old field. Molecular Ecology 26:2895-2904

676 George SJ, Kelly R, Greenwood P, Tibbett M (2009) Is soil carbon being
677 sequestered along a reconstructed biodiverse Australian Jarrah Forest

chronosequence following bauxite mining? Mine Closure 2009: Proceedings of the Fourth International Conference on Mine Closure 137-144 Australian Centre for Geomechanics

George S, Kelly R, Greenwood P, Tibbett M (2010) Soil carbon and litter development along a reconstructed biodiverse forest chronosequence of South-Western Australia. Biogeochemistry 101:197-209

Glöckner FO, Yilmaz P, Quast C, Gerken J, Beccati A, Ciuprina A, Bruns G, Yarza P, Peplies J, Westram R, Ludwig W (2017) 25 years of serving the community with ribosomal RNA gene reference databases and tools. Journal of Biotechnology 261:169-176

Golos PJ, Dixon KW (2014) Waterproofing topsoil stockpiles minimizes viability decline in the soil seed bank in an arid environment. Restoration Ecology 22:495-501

Grant C (2009) State and Transition models for mining restoration in Australia. New Models for Ecosystem Dynamics and Restoration:280-294

Green J, Bohannan BJ (2006) Spatial scaling of microbial biodiversity. Trends in Ecology & Evolution 21:501-507

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

Heneghan L, Miller SP, Baer S, Callaham Jr MA, Montgomery J, Pavao-Zuckerman M, Rhoades CC, Richardson S (2008) Integrating soil ecological knowledge into restoration management. Restoration Ecology 16:608-617

Hijmans RJ, Williams E, Vennes C, Hijmans MRJ (2017) Package 'geosphere'. Spherical Trigonometry 1:7

Kardol P, Wardle DA (2010) How understanding aboveground–belowground linkages can assist restoration ecology. Trends in Ecology & Evolution 25:670-679

Kneller T, Harris RJ, Bateman A, Muñoz-Rojas M (2018) Native-plant amendments and topsoil addition enhance soil function in post-mining arid grasslands. Science of The Total Environment 621:744-752

Koch JM, Samsa GP (2007) Restoring Jarrah forest trees after bauxite mining in Western Australia. Restoration Ecology 15:S17-S25

Kuhn M (2015) Caret: classification and regression training. Astrophysics Source Code Library. ascl-1505.

Lane D (1991) 16S/23S rRNA sequencing. Nucleic acid techniques in bacterial systematics. 115-175

Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML, Pace NR (1985) Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. Proceedings of the National Academy of Sciences 82:6955-6959

Lem AJ, Liddicoat C, Bissett A, Cando-Dumancela C, Gardner MG, Peddle SD, Watson CD, Breed MF (2022) Does revegetation cause soil microbiota recovery? Evidence from revisiting a revegetation chronosequence six years after initial sampling. Restoration Ecology e13635.

Lewis DE, White JR, Wafula D, Athar R, Dickerson T, Williams HN, Chauhan A (2010) Soil functional diversity analysis of a bauxite-mined restoration chronosequence. Microbial Ecology 59:710-723

Liddicoat C, Weinstein P, Bissett A, Gellie NJ, Mills JG, Waycott M, Breed MF (2019) Can bacterial indicators of a grassy woodland restoration inform ecosystem assessment and microbiota-mediated human health? Environment International 129:105-117

Liddicoat C, Krauss SL, Bissett A, Borrett RJ, Ducki LC, Peddle SD, Bullock P, Dobrowolski MP, Grigg A, Tibbett M, Breed MF (2022) Next generation restoration metrics: Using soil eDNA bacterial community data to measure trajectories towards rehabilitation targets. Journal of Environmental Management 310:114748

Magoč T, Salzberg SL (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27:2957-2963

Manero A, Standish R, Young R (2021) Mine completion criteria defined by best-practice: A global meta-analysis and Western Australian case studies. Journal of Environmental Management 282:111912

Martiny JB, Eisen JA, Penn K, Allison SD, Horner-Devine MC (2011) Drivers of bacterial β-diversity depend on spatial scale. Proceedings of the National Academy of Sciences 108:7850-7854

Mcmurdie PJ, Holmes S (2013) phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLOS ONE 8:e61217

Mendes MS, Latawiec AE, Sansevero JB, Crouzeilles R, Moraes LF, Castro A, Alves-Pinto HN, Brancalion PH, Rodrigues RR, Chazdon RL (2019) Look down—there is a gap—the need to include soil data in Atlantic Forest restoration. Restoration Ecology 27:361-370

Muñoz-Rojas M (2018) Soil quality indicators: critical tools in ecosystem restoration. Current Opinion in Environmental Science & Health 5:47-52

Myers N, Mittermeier RA, Mittermeier CG, Da Fonseca GA, Kent J (2000) Biodiversity hotspots for conservation priorities. Nature 403:853-858

Neupane S, Goyer C, Zebarth BJ, Li S, Whitney S (2019) Soil bacterial communities exhibit systematic spatial variation with landform across a commercial potato field. Geoderma 335:112-122

Ngugi MR, Dennis PG, Neldner VJ, Doley D, Fechner N, Mcelnea A (2018) Open-cut mining impacts on soil abiotic and bacterial community properties as shown by restoration chronosequence. Restoration Ecology 26:839-850

754 Nkongolo K, Narendrula-Kotha R (2020) Advances in monitoring soil microbial
755 community dynamic and function. Journal of Applied Genetics:1-15

756 Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'hara R, Simpson GL,
757 Solymos P, Stevens MHH, Wagner H (2013) Package 'vegan'. Community ecology
758 package, version 2:1-295

759 Pickett ST (1989) Space-for-time substitution as an alternative to long-term studies.
760 Pages 110-135 In: Long-term studies in ecology. Springer

761 Prober SM, Leff JW, Bates ST, Borer ET, Firn J, Harpole WS, Lind EM, Seabloom
762 EW, Adler PB, Bakker JD (2015) Plant diversity predicts beta but not alpha diversity
763 of soil microbes across grasslands worldwide. Ecology Letters 18:85-95

764 Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner
765 FO (2013) The SILVA ribosomal RNA gene database project: improved data
766 processing and web-based tools. Nucleic Acids Research 41:D590-D596

767 R Core Team (2020). R: A language and environment for statistical computing. R
768 Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org
769 (accessed Nov. 25, 2020)

770 Rokich DP, Dixon KW, Sivasithamparam K, Meney KA (2000) Topsoil handling and
771 storage effects on woodland restoration in Western Australia. Restoration Ecology
772 8:196-208

773 Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski
774 RA, Oakley BB, Parks DH, Robinson CJ (2009) Introducing mothur: open-source,
775 platform-independent, community-supported software for describing and comparing
776 microbial communities. Applied and Environmental Microbiology 75:7537-7541

777 Schmid CA, Reichel R, Schröder P, Brüggemann N, Schloter M (2020) 52 years of
778 ecological restoration following a major disturbance by opencast lignite mining does
779 not reassemble microbiome structures of the original arable soils. Science of The
780 Total Environment:140955

781 Shade A (2017) Diversity is the question, not the answer. The ISME Journal 11:1-6

782 Sheoran V, Sheoran A, Poonia P (2010) Soil reclamation of abandoned mine land by
783 revegetation: a review. International Journal of Soil, Sediment and Water 3:13

784 Spain AV, Tibbett M, Hinz DA, Ludwig JA, Tongway DJ (2015) The mining-
785 restoration system and ecosystem development following bauxite mining in a
786 biodiverse environment of the seasonally dry tropics, Northern Territory, Australia.
787 Pages 159-227 In: Tibbett M (ed.) Mining in ecologically sensitive landscapes. CRC
788 Press, USA

789 Stevens J, Dixon K (2017) Is a science-policy nexus void leading to restoration
790 failure in global mining? Environmental Science & Policy 72:52-54

Sun S, Li S, Avera BN, Strahm BD, Badgley BD (2017) Soil bacterial and fungal communities show distinct recovery patterns during forest ecosystem restoration. Applied Environmental Microbiology 83:e00966-00917

Sun S, Jones RB, Fodor AA (2020) Inference-based accuracy of metagenome prediction tools varies across sample types and functional categories. Microbiome 8:1-9

Tedersoo L, Drenkhan R, Anslan S, Morales-Rodriguez C, Cleary M (2019) High-throughput identification and diagnostics of pathogens and pests: overview and practical recommendations. Molecular Ecology Resources 19:47-76

Thompson LR, Sanders JG, Mcdonald D, Amir A, Ladau J, Locey KJ, Prill RJ, Tripathi A, Gibbons SM, Ackermann G (2017) A communal catalogue reveals Earth's multiscale microbial diversity. Nature 551:457-463

Tibbett M (2010) Large-scale mine site restoration of Australian eucalypt forests after bauxite mining: soil management and ecosystem development. Pages 309-326 In: Ecology of industrial pollution. Cambridge University Press UK

Tibbett M (2015) Mining in ecologically sensitive landscapes: concepts and challenges. Mining in Ecologically Sensitive Landscapes; Tibbett, M., Ed.; CSIRO Publishing: Clayton, VIC, Australia:3-6

Tibbett M, Gil-Martínez M, Fraser T, Green ID, Duddigan S, De Oliveira VH, Raulund-Rasmussen K, Sizmur T, Diaz A (2019) Long-term acidification of pH neutral grasslands affects soil biodiversity, fertility and function in a heathland restoration. Catena 180:401-415

Van Der Heyde M, Bunce M, Dixon K, Wardell-Johnson G, White N, Nevill P (2020) Changes in soil microbial communities in post mine ecological restoration: Implications for monitoring using high throughput DNA sequencing. Science of The Total Environment 749:142262

Walker LR, Wardle DA, Bardgett RD, Clarkson BD (2010) The use of chronosequences in studies of ecological succession and soil development. Journal of Ecology 98:725-736

Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Applied and Environmental Microbiology 73:5261-5267

Wheeler B, Torchiano M, Torchiano MM (2016) Package 'lmPerm'. R package version:1-1

Yan D, Bissett A, Gellie N, Mills JG, Lowe AJ, Breed MF (2019) Soil bacterial community differences along a coastal restoration chronosequence. Plant Ecology:1-17

Yan D, Mills JG, Gellie NJ, Bissett A, Lowe AJ, Breed MF (2018) High-throughput eDNA monitoring of fungi to track functional recovery in ecological restoration. Biological Conservation 217:113-120

831 Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, Schweer T, Peplies
832 J, Ludwig W, Glöckner FO (2014) The SILVA“and "All-species Living Tree Project
833 ”LTP)" taxonomic frameworks. Nucleic Acids Research 42:D643-D648
834

TABLES AND FIGURES

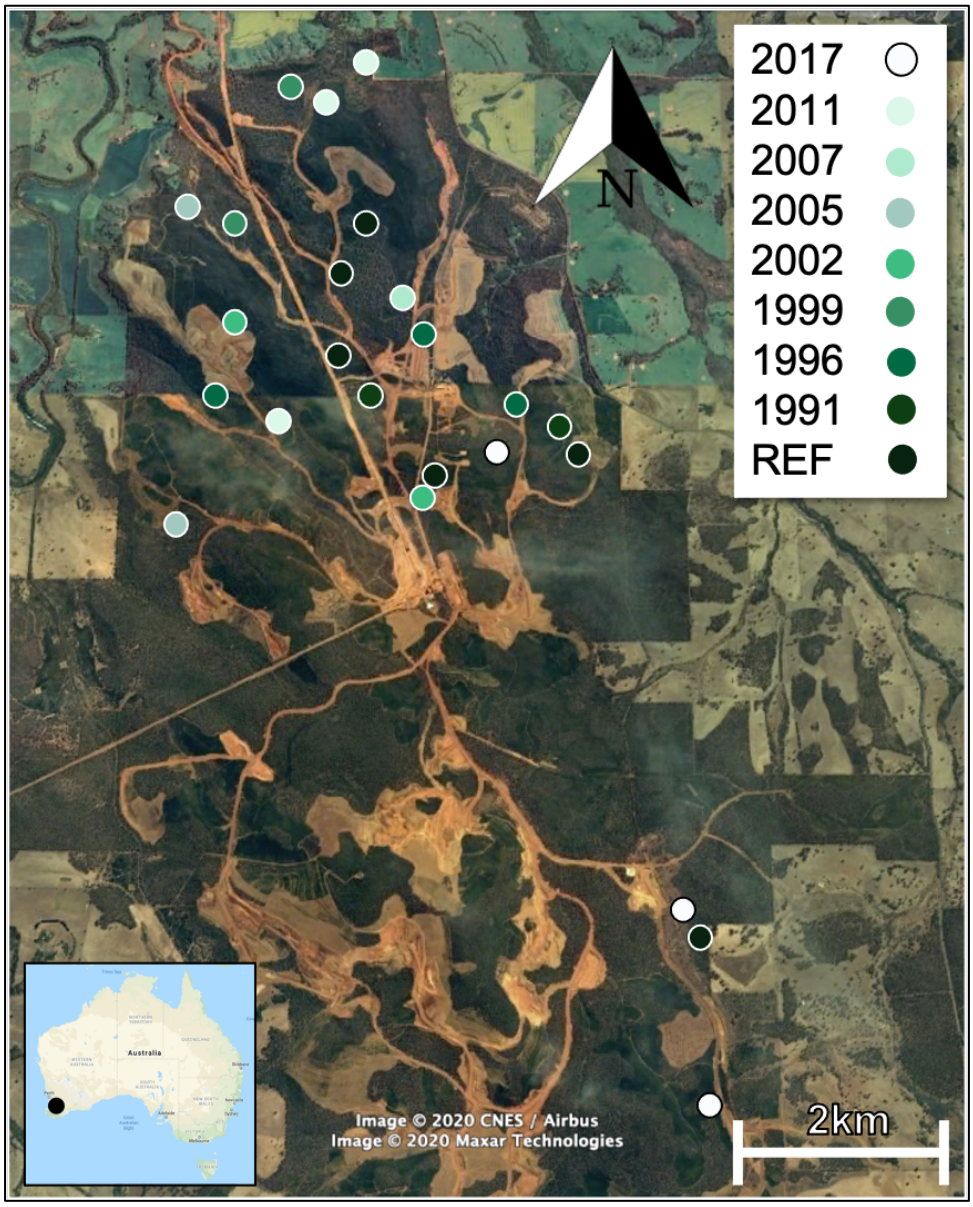
Table 1. Mean (±SD) amplicon sequence variant (zOTU) abundance by year of rehabilitation and depth. *Standard deviation was not calculated for 2007 with only one sample from each depth.

Year of rehabilitation	Samples (n)	Depth (cm)	Mean zOTU abundance (±SD)
Reference	6	0-10	84,660 ±14,314.35
1991	2	0-10	59,787 ±6,843.38
1996	3	0-10	89,332 ±12,136.41
1999	2	0-10	86,825 ±4,585.59
2002	2	0-10	78,984 ±7,860.91
2005	2	0-10	86,476 ±19,240.36
2007	1*	0-10	116,520 *
2011	3	0-10	82,815 ±6,373.19
2017	3	0-10	80,934 ±5,380.64
Reference	6	20-30	77,613 ±41,010.35
1991	2	20-30	77,149 ±5,621.5
1996	3	20-30	91,201 ±9,412.71
1999	2	20-30	90,964 ± 379.01
2002	2	20-30	99,775 ±33,844.25
2005	2	20-30	85,143 ±163.34
2007	1*	20-30	86,817 *
2011	3	20-30	87,982 ±4,723.58
2017	3	20-30	69,488 ±34,863.61

840 **Table 2.** Mean (±SD) amplicon sequence variant (zOTU) richness and diversity of bacterial communities assessed with permuted
841 analysis of variance at South 32's Worsley Bauxite mine, Western Australia. *2007 (n=1) was excluded from statistical analysis for
842 both depths.

Year of rehabilitation	Samples (n)	Depth (cm)	zOTU Richness (±SD)		Diversity (±SD)	
			Observed	*Chao 1*	*Shannon*	*Simpson*
Reference	6	0-10	12576.0 ±1771.1	19695.7 ±3754.2	8.34 ±0.25	0.998 ±0.0003
1991	2	0-10	13928.5 ±487.2	23335.8 ±510.1	8.51 ±0.09	0.999 ±0.0001
1996	3	0-10	13156.5 ±597.5	21527.9 ±969.9	8.40 ±0.10	0.998 ±0.0004
1999	2	0-10	15764.5 ±637.1	25034.2 ±207.7	8.82 ±0.17	0.999 ±0.0002
2002	2	0-10	12436.5 ±2448.7	18038.6 ±6015.8	8.52 ±0.08	0.999 ±0.0003
2005	2	0-10	10595.0 ±4736.2	14763.1 ±7808.9	8.20 ±0.70	0.998 ±0.0013
2007*	1*	0-10	14140.0 *	22922.9 *	8.58 *	0.999 *
2011	3	0-10	11744.0 ±2739.6	16349.4 ±4921.6	8.47 ±0.32	0.999 ±0.0003
2017	3	0-10	12825.0 ±514.1	19235.4 ±402.2	8.49 ±0.12	0.998 ±0.0004
***P* values**			Df=8, *p*=0.371	Df=8, *p*=0.136	Df=8, *p*=0.497	Df=8, *p*=0.627
Reference	6	20-30	12993.6 ±2390.4	18389.5 ±5026.0	8.39 ±0.19	0.999 ±0.0002
1991	2	20-30	15152.0 ±1630.6	23301.9 ±3041.1	8.46 ±0.13	0.999 ±<0.0001
1996	3	20-30	16331.2 ±2189.7	24477.7 ±4478.3	8.62 ±0.23	0.999 ±0.0004
1999	2	20-30	16193.5 ±392.4	23343.4 ±752.1	8.66 ±0.24	0.998 ±0.0008
2002	2	20-30	13864.5 ±7428.2	20227.7 ±14273.4	8.48 ±0.61	0.999 ±0.0002
2005	2	20-30	11238.0 ±1195.0	14129.2 ±3158.1	8.39 ±0.06	0.999 ±0.0001
2007*	1*	20-30	16000 *	23211.3 *	8.64 *	0.999 *
2011	3	20-30	14858.3 ±2812.4	20633.8 ±5448.7	8.66 ±0.27	0.999 ±0.0001
2017	3	20-30	13821.0 ±1548.6	19487.1 ±5208.2	8.55 ±0.03	0.999 ±<0.0001
***P* values**			Df=8, *p*=0.675	Df=8, *p*=0.659	Df=8, *p*=0.763	Df=8, *p*=0.847

843



844

845 **Figure 1.** Map of sampling sites from the rehabilitation chronosequence at South
846 32's Worsley bauxite mine in southwest Western Australia. Circles indicate sampling
847 sites, with colour representing year of rehabilitation. Soil was sampled from two
848 depths (0-10 cm and 20-30 cm) at each site.

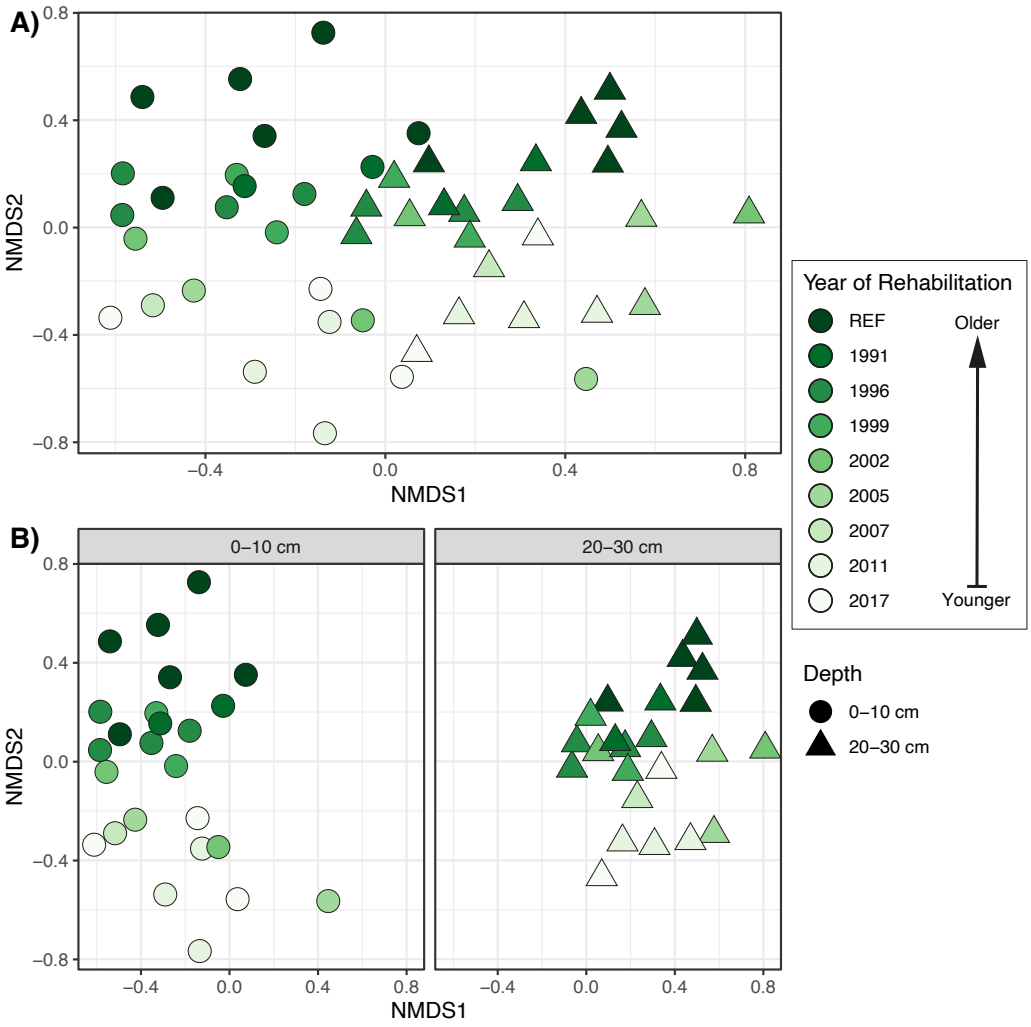


Figure 2. Non-metric multidimensional scaling (NMDS) ordinations of Bray-Curtis distance matrices indicating bacterial community composition across the rehabilitation chronosequence at Worsley Alumina, Western Australia. (A) ordination of samples from both 0-10 cm and 20-30 cm soil depths and (B) ordinations of each depth separately. Ordinations indicate a general convergence of bacterial community composition with increasing age towards the reference sites at both soil depths.

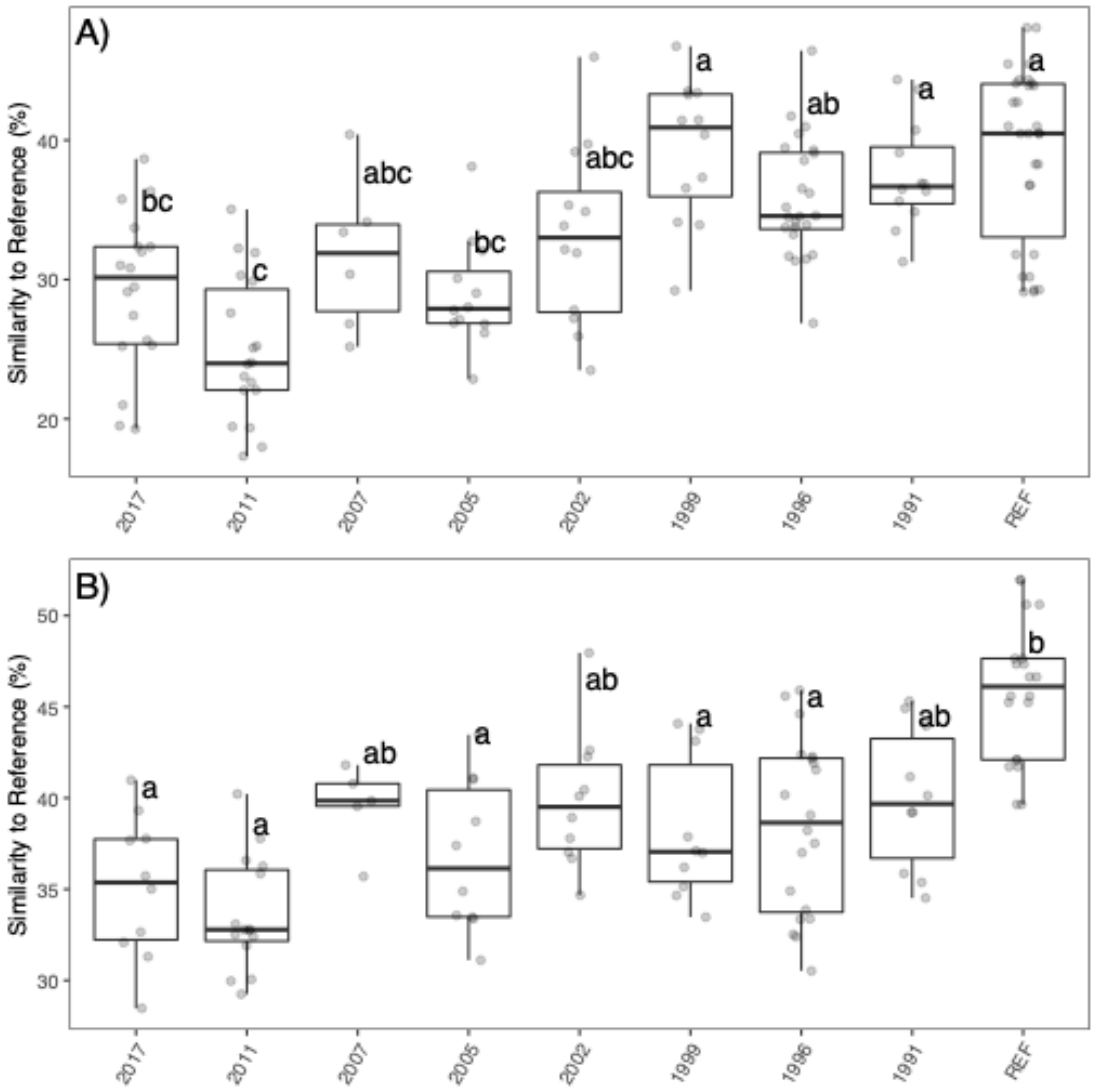


Figure 3. Boxplot indicating similarity to reference of Bray-Curtis distances for each sample at (A) 0-10 cm soil depth and (B) 20-30 cm soil depth. Horizontal lines indicate 25th, 50th (median), and 75th percentile of similarities to reference and vertical lines represent 95% confidence intervals. Kruskal-Wallis tests indicated significant differences (p<0.05) between years of rehabilitation at both depths and Dunn post hoc tests with Bonferroni correction indicated younger rehabilitated sites are different to references and older rehabilitated sites are comparable to references. Groups not sharing a letter are significantly different (2017, 2011, and 2005 are

866 significantly different to reference sites at the 0-10 cm depth and 2017, 2011, 2005,

867 1999, and 1996 are different to reference sites at the 20-30 cm depth).

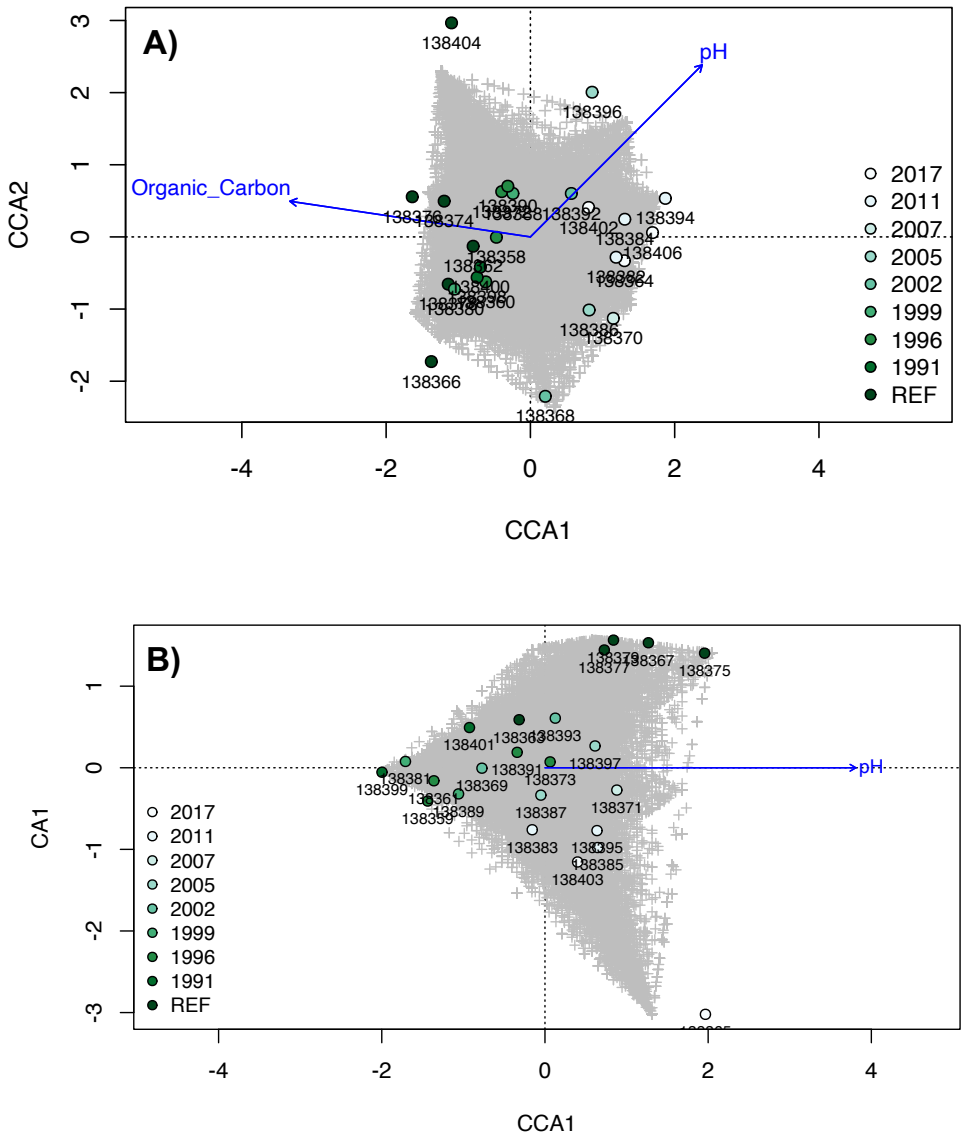
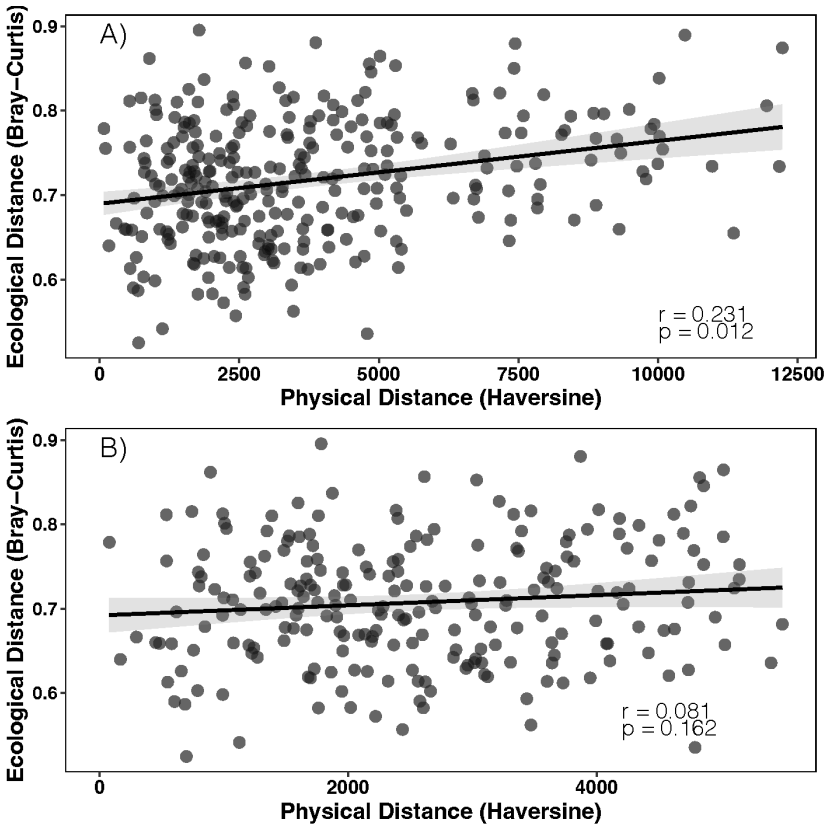


Figure 4: Constrained correspondence analyses (CCA) between bacterial community composition (Bray-Curtis dissimilarity) and associated soil chemical variables at (A) 0-10 cm depth and (B) 20-30 cm depth. Blue arrows indicate direction of influence of soil variable on bacterial communities.



874

875 **Figure 5** Scatterplot of the association between the distance between samples
876 (Haversine distance matrix) and bacteria community composition (Bray-Curtis
877 distance matrix), showing Mantel test statistics. (A) shows a significant correlation
878 present with all sites included, and (B) shows no significant correlation with three
879 geographically separate sites removed indicating these three sites are driving the
880 spatial autocorrelation.

881