Environmental Science Advances

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *Environ. Sci.: Adv.*, 2025, **4**, 1877

Use of environmental forensics to assess land sustainability: a case study on a former coal mining site

Ibrahim A Salaudeen, Kate Randall *, Nabanita Basu, William Cheung, Catherine E. Nicholson, Justin J. Perry * and John R. Dean

A comprehensive environmental forensics investigation has been done on a former coal mining site in NE England, and now a country park used for recreation, but which lacks the ability to grow vegetation in certain areas. Initial mapping of the site was done using an unmanned aerial vehicle (UAV) with multispectral imaging (MSI) capability, followed by determination of 7 vegetation indices (VIs). The use of the VIs allowed a direct comparison between the two field sites and provided an indicator of vegetative stress. This was followed by field sampling and laboratory analyses using EX-XRF for metal analyses, soil property determination (pH, CEC and organic matter), metabolomic determination of the main soil metabolites using Hydrophilic Liquid Interaction Chromatography Hi-Resolution Mass Spectrometry, and a comprehensive investigation of soil bacteria and fungi using metagenomics. The results indicate how the soil environment of the top field has recovered to allow an abundance of flora in Spring and Summer, despite the soil having a low pH (4.0) and a high Pb concentration (94.0 mg kg^{-1}) but counterbalanced by the presence of natural plant and soil metabolites, and a high abundance of nutrient producing bacteria. In contrast, the bottom field is characterised by a sparse vegetation coverage on a harsher soil environment reminiscent of marshland, with a soil pH of 6.2, but a lower Pb concentration (58.4 mg kg^{-1}) contrasted with soil with a high sodium content (2050 mg kg^{-1}), the presence of man-made anthropogenic metabolites, and bacteria capable of undertaking soil remediation.

Received 23rd July 2025
Accepted 29th August 2025

DOI: 10.1039/d5va00229j

rsc.li/esadvances

Environmental significance

We report, for the first time, a multimodal environmental forensic investigation of a former coal mining site with respect to its impact on a current recreational country park after nearly 60 years of dormancy. Using conventional elemental analysis, selected soil properties, metabolomic profiling and environmental metagenomics has enabled identification of some key pollutant markers as well as the health of the soil in terms of microbial communities. This laboratory-based activity has been corroborated using an unmanned aerial vehicle with multispectral image camera to assess the important vegetation indices as indicators of vegetative stress. This comprehensive environmental forensics investigation has enabled a detailed study to be undertaken that has provided invaluable data for understanding the complexity of a current country park.

Introduction

Sustainable land management has emerged as a critical global priority due to increasing concerns over environmental degradation in post-industrial landscapes.[1] Industrial activities, particularly coal mining, have left indelible marks on ecosystems through soil acidification, heavy metal accumulation, and disruption of natural hydrological regimes.[2,3] These legacy impacts compromise both the ecological integrity and productive potential of affected lands, thus necessitating integrated approaches that combine environmental restoration with socio-economic revitalization.[4,5]

Contaminated sites stemming from coal mining operations present complex challenges that often exceed the remedial capabilities of traditional approaches.[6,7] The heterogeneous distribution of pollutants, coupled with the intricate history of industrial contamination, demands innovative assessment methods capable of accurately delineating pollutant sources and spatial variability.[8] In this context, emerging environmental forensic techniques which integrate advanced chemical analyses and geospatial mapping,[9] as well as metabolomics and metagenomics offer promising avenues for unravelling complex contamination histories and guiding effective remediation strategies.

School of Geography and Natural Sciences, Northumbria University, Newcastle upon Tyne, NE1 8ST, UK. E-mail: justin.perry@northumbria.ac.uk; kate.randall@northumbria.ac.uk

Environmental forensics is an interdisciplinary discipline that integrates analytical chemistry, geochemistry, and statistical methodologies to identify, characterize, and apportion contaminants in environmental matrices.[10,11] By employing a range of analytical techniques, such as several approaches in chromatography-mass spectrometry, elemental analysis, and multivariate analysis, this approach can provide a detailed reconstruction of contaminant sources and their subsequent migration pathways.[12,13] Such robust methodologies enable researchers to differentiate between natural background levels and pollution attributable to anthropogenic activities, thereby revealing complex contamination patterns that are often obscured in legacy industrial landscapes.[14] In the realm of land sustainability, environmental forensics serves as a powerful tool for assessing the long-term viability of contaminated sites. Its capacity for detailed source identification and spatial mapping of pollutants allows for a more informed evaluation of remediation needs and sustainable land use planning.[15,16] By generating precise contaminant profiles, environmental forensic techniques underpin strategic decision-making in environmental management, ensuring that remediation efforts are both targeted and effective, while simultaneously supporting policy frameworks aimed at restoring ecological and socio-economic functions.[11] Indeed, a multiparameter approach to the assessment of contamination may reveal that remediation is not prioritized or even required hence ensuring best use of resources.

The characterization of contaminated lands involves the use of a wide range of advanced analytical techniques to unravel the complexity of pollutant distribution and chemistry. Frequently analytical techniques, such as, variations of chromatography coupled with mass spectrometry are commonly employed to discern both chemical species and metabolite information with high sensitivity and specificity. This detailed chemical insight is essential for understanding the potential risks associated with contamination and, if necessary, for guiding subsequent remediation efforts.[17] In addition, the use of advanced metagenomics has enabled characterisation of the microbial and fungal communities present in soil, as indicators of soil health and diversity of habitat for flora. Complementing these laboratory-based techniques, is the use of multivariate statistical analysis which involves using tools, such as, principal component analysis and cluster analysis to enable the interpretation of complex environmental datasets, revealing hidden correlations and patterns in pollutant distribution.[18–20] Moreover, geospatial tools like remote sensing and Geographic Information System (GIS) mapping are invaluable for visualizing the spatial distribution of contaminants. Remote sensing provides a macroscopic view by detecting surface anomalies and vegetation stress, while GIS mapping integrates various data layers to identify contamination hotspots and track spatial variability.[21,22] Together, the integration of these techniques yields a comprehensive contamination profile that is key for the development of targeted remediation strategies and the advancement of sustainable land management practices.

This research presents a systematic and scientific evaluation of a former coal mining site to assess the impact on the environment of the various land-uses it has undergone historically up to its current state as a country park. This evaluation has involved the use of a wide range of complementary "next generation" analytical techniques relating to soil analyses including an assessment of soil properties, elemental analysis, environmental metabolomics (*via* Hydrophilic Liquid Interaction Chromatography Hi-Resolution Mass Spectrometry) and metagenomics coupled with the deployment of remote sensing and statistical data treatment to aid the development of a comprehensive investigation of the site.

Experimental

Field site and sampling

Weetslade Country Park is in North Tyneside, 8 km north of Newcastle upon Tyne (grid reference: NZ 258 722, 55° 2′ 37″ N, 1° 35′ 46″ W). The 43-hectare park site was opened in 2006 under the ownership of the Land Trust[23] but is managed by the Northumberland Wildlife Trust.[24] Previously, however, it was a coal colliery site that operated from 1903 until its closure on the 10th September 1966. At its peak of production an estimated 160 000 tons of coal was dug (1947). The coal was mainly graded for household use, with some used for steam generation and manufacturing. The deepest seams were 350 m below the current surface. After the demise of the colliery the site was left abandoned until the extensive landscaping in the early 2000's. Our investigation took place over spring/summer of 2023, with specific targeting of the north-facing embankment and its lower marshland. This site was chosen for investigation because of its potential for comparison of data from UAV observation with ground truthing chemical/microbial analysis, enabled by a supportive landowner and full accessibility to the site. In addition, from casual observation it was clear there was variation in vegetation over the site which could be investigated.

Soil sampling was done on the 19 and 27 June 2023 in two areas: the north-facing embankment (Top field) which contained freely draining soil and the bottom field (Fig. 1 and S1) which in comparison was marshy and relatively waterlogged. The top field was split into grid lines, marked A–L, with 6 sampling points per grid line equating to 72 individual grab samples. The bottom field was also split into grid lines, marked A–L, with (between 3 and) 6 sampling points per grid line equating to 65 individual grab samples. All sub-surface soil samples, between 2–10 cm depth, were collected using a stainless-steel trowel, and the sampling coordinates noted by a handheld GPS unit. To avoid cross-contamination, the trowel was cleaned with a new sterile wipe, between each sample. Dual soil samples were collected from each coordinate. One set of collected soil samples were placed in labelled collection bags *i.e.* Kraft sample bags and transported back to the laboratory in a refrigerated coolbox on the same day as collection for subsequent soil analysis. In the laboratory, the soil samples were air-dried to minimise loss of organic compounds (48 hours),[25] then ground with a mortar and pestle, and finally sieved through a 250 μm nylon mesh.[25] Finally, the sieved samples were subdivided into two: one portion was immediately frozen to reduce soil-microbial activity (−18 °C) for subsequent

Fig. 1 Contemporary Ordnance Map of Weetslade Country Park. [The red cross is representative of the field sampling on the north-facing embankment (NZ 255 726) while the blue star represents the field sampling area in the bottom field (NZ 256 727)]. © Crown copyright and database rights 2024 Ordnance Survey (AC0000851941) 1:25 000 Scale Colour Raster [TIFF geospatial data], Scale 1:25 000, Tiles: nz27_clipped, Updated: 1 June 2023, Ordnance Survey (GB), Using: EDINA Digimap Ordnance Survey Service, https://digimap.edina.ac.uk, Downloaded: 2024-01-18 12:00:57.457.

chromatographic analysis while the other portion was stored for analysis by energy dispersive XRF and soil parameters *i.e.* pH, cation exchange capacity, organic matter and soil moisture content. The other set of collected soil samples were placed in labelled collection bags and placed in a portable electric freezer (−18 °C). These samples were kept in a frozen state (−18 °C) prior to DNA extraction for metagenomic analysis.

Soil analysis by X-ray fluorescence spectrometry

Quantitative analysis of the collected soil samples was done by accurately weighing the ground soil sub-samples (approximately 4 g) followed by thorough mixing, with a Hoeschwax hydrocarbon binder (approximately 0.7 g) using a Retsch mixer mill MM 400 (Retsch, Germany) and finally pelletising by the ground mixture being pressed at 10-tonne pressure (Specac manual hydraulic press, Specac Ltd, Kent, UK). Care was taken to eliminate any cross-contamination between each soil sample. The benchtop ED-XRF instrument, Spectro Xepos (Spectro Analytical Instruments, Kleve, Germany), is equipped with a detector that is comprised of a microprocessor-controlled drift detector with Peltier cooling. An ED-XRF carousel which can accommodate 8 samples was used for sequential analysis of the pressed pellets.

Data processing was done using the Geochemistry Traces method, as supplied by the ED-XRF manufacturer (Spectro Analytical Instruments). Quality control of the ED-XRF data for multi-element analysis was done by analysis of 12 certified reference materials (CRMs), sourced due to their characterisation as soils or associated materials. The CRMs were: ERM-cc135a (contaminated brickworks soil) and LGC 6138 (coal-carbonisation site soil) (Laboratory of the Government Chemist, UK). BCR-723 (road dust) and BCR-176R (fly ash) (Institute for Reference Materials and Measurements, Belgium). BGS 102 (ironstone soil) (British Geological Survey, UK). GBW 07403 (a yellow-brown soil); GBW 07405 (yellow-red soil, polymetallic ore area); GBW 07406 (red soil, polymetallic ore area); GBW 07411 (soil); and GBW 07313 (marine sediment) (National Research Center for Certified Reference Materials (NRCCRM), China). SRM 2710 (Montana I soil) and SRM 1633a (coal fly ash) (National Institute of Standards and Technology, NIST, USA). Data from the CRMs was used to correct the sample element concentrations. For each element a lower limit of detection (LLOD) and limit of quantitation (LOQ) was determined using the standard deviation (σ) of the y-intercept and the average slope (s) of multiple linear regression equations ($n = 6$) as follows:

$$\text{LLOD} = 3.3\sigma/s \text{ and } \text{LOQ} = 10\ \sigma/s.$$

Environmental metabolomics by hydrophilic liquid interaction chromatography Hi-resolution mass spectrometry (data dependent analysis)

Metabolite sample extraction. Soil sub-samples (approximately 20 mg) were sonicated in methanol (1 ml) for 20 min in an ice water bath.[26] The extract was centrifuged (15 000 rpm, for 15 min at 4 °C), evaporated to dryness in a vacuum concentrator (Concentrator plus, Eppendorf UK Ltd, Stevenage, UK) at 20 mbar and 45 °C for 2 hours, prior to analysis. Prior to analysis, extracts were re-suspended in 100 µl of an acetonitrile : water (95 : 5% v/v) mixture with sonication (15 min) and filtered (0.22 µm, micro costar spin column) at 10 000 rpm for 10 min at 4 °C. Hydrophilic liquid interaction chromatography (HILIC) metabolite profiling of the soil extracts ($n = 5$ from each of Site A and Site B) was done on a Vanquish Liquid Chromatography chromatographic separation system connected to an ID-X High Resolution Mass Spectrometer (Thermo Scientific, Hemel Hempstead, UK). The metabolomic ESI^+ and ESI^- data sets were

processed *via* Compound Discoverer 3.2 using the following settings: untargeted metabolomic workflow, parent ion mass tolerance 10 ppm, alignment model adaptive curve, minimum intensity 3 signal to noise ratio (S/N) threshold 0.3 min, compound consolidation, retention time tolerance 0.3 min. Database matching were performed at MS2 level using Thermo Scientific online mzCloud databases (Hemel Hempstead, UK) with a similarity index of 70% or higher. The pooled quality control (QC) data was used to assess for instrumental draft and feature selection, the sum of the QC RSD variation across positive (ESI^+) ionization mode and negative (ESI^-) ionization mode was 2.62%. Each MS/MS ID metabolite which had an RSD variation of <25% was retained for multivariate analysis.[27,28]

Metabolomics data sets for the top and bottom fields in both positive (ESI^+) and negative (ESI^-) ionization modes were merged and normalised together. The data was auto scaled and $\log_{10}$ transformed prior to data visualisation. All data visualisation tools *i.e.* hierarchical cluster analysis, supervised partial least squares-discriminant analysis (PLS-DA) and unsupervised PCA were generated using metaboanalyst.[29]

Metagenomics

DNA extraction, quantification and quality. DNA was extracted from composite sample from each sample location. A 0.25 g sub-sample of soil was extracted using the DNeasy PowerSoil Kit (Qiagen, Hilden, Germany) following manufacturer's instructions. DNA quantification was determined *via* Quibit analysis (ThermoFisher). DNA quality was assessed using a NanoDrop™ One/OneC Microvolume UV-Vis Spectrophotometer, comparing the 260/280 nm and 260/230 nm ratios.

Shotgun metagenomic sequencing. Following quality control, extracted DNA was sent to the NU-OMICS DNA sequencing research facility for shotgun metagenomics using the Nextera XT approach.

Bioinformatic processing of microbial shotgun metagenomic data. The raw metagenome next-generation sequencing (NGS) fastq were ran through TrimGalore[30] using default parameters to remove adapters and low-quality bases. Trimmed files were then classified in Kraken 2 (ref. 31) using the plusPF_20240112 database and Braken[32] was used to estimate species abundance and then sample data was merged to create a species abundance table.

Pre-processing. From the metagenomic sequencing, taxonomic data was extracted, analysed and presented. R version 4.4.1 (ref. 33) was used to pre-process and conduct statistical analysis and graphically present the microbial datasets. The final operational taxonomic unit (OTU) abundance table with assigned taxonomy and sample metadata describing sample characteristics were imported into R. The metagenomic OTU abundance table data was subset into three independent datasets, consisting of (1) bacteria, (2) fungi and (3) archaea. For the bacterial NGS data, 178 OTUs classified as "unassigned bacteria" were removed. No "unassigned" OTUs were present within the archaeal and fungal datasets. Using the vegan package,[34] each OTU abundance table was then rarefied as the method of normalisation to standardise differential library sizes (sequencing depth per sample)[35,36] (Table S1). A pre-screen of sample sequencing depth per sample was used to determine the rarefication depth. This process is a trade-off between aiming to keep as many samples as possible, whilst minimising overall data loss. The datasets were therefore not simply rarefied to the lowest sample library, meaning some samples were lost during this process, that can be tracked back to low DNA concentration input. For bacteria, only one sample was lost, whereas for both fungi and archaea, 4 samples were lost (Table S1). Following rarefaction, OTU filtering was conducted, which involves removing OTUs with <3 read counts across all samples. Table S1 presents the number of final OTUs per microbial dataset. The phyloseq package[37] was then used in conjunction with ggplot[38] to generate and present α and β diversity community metrics.

Microbial community richness and evenness (α diversity). Bacterial, fungal and archaeal community richness and evenness were independently compared between field locations. Both richness and evenness data were analysed *via* generalised linear models using the glmmTMB package[39] and validated using the DHARMa package.[40] As richness scores (number of unique OTUs per sample) are count data, the negative binominal error distribution was applied. For evenness, the Simpson index was inversely transformed and analysed using the gaussian error distribution. Model and residual fit passed validation checks for all univariate microbial analysis. Figures presenting α diversity data show mean and standard error of the mean calculated with the emmeans package.[41]

Microbial community composition (β diversity). Differences in the overall composition of each microbial target group (bacteria fungi and archaea communities) between field location were assessed using multivariate negative binomial generalised linear models (MV-GLMs), with a negative binomial error term to account for overdispersion, a typical feature of microbial NGS datasets.[42] For each dataset, the cleaned OTU abundances were fitted to the model predictor (field location). Multivariate and univariate p values were obtained by log-likelihood ratio tests and model-free bootstrapping with 10 000 permutations of probability integral transform residuals and adjusted p values to account for multiple testing. MV-GLM is applicable to the analysis of microbial NGS data because it is possible to model changes in the relative abundance of individual OTUs whilst simultaneously looking at an overall community response. Residual plots were assessed for even spread to assess model fit.

Visualisation of microbial community compositional data are presented as PCoA plots after creating a Bray–Curtis distance matrix, allowing multivariate data to be viewed in a 2D space. The closer two points/samples are, the similar in composition the communities within each sample are.

Unmanned aerial vehicle

An unmanned aerial vehicle (UAV) with multispectral image (MSI) camera (DJI Phantom 4, Coptrz Ltd, Leeds, UK) was deployed for site investigation. The MSI camera consists of a 6-sensors, stabilized with a 3-axis gimbal, that cover the blue (450 ± 16 nm), green

(560 ± 16 nm), red (650 ± 16 nm), red edge (730 ± 16 nm), near-infrared (840 ± 26 nm) and RGB (visible) mode. Each camera lens has a field of view of 62.7°, a focal length of 5.74 mm, an autofocus set at ∝, and an aperture of f/2.2. During data capture the camera array is positioned perpendicular to the ground. Data was gathered across both the top and bottom fields on the 11 and 16 June 2023. For the top field, 2885 image files were gathered over 486 waypoints, as 16-bit TIF files corrected for ambient radiance values. The UAV speed was 5.0 m s^{-1} with an average height of 18.4 m for the 1865 m flight distance. For the bottom field, 2320 image files were gathered across 393 waypoints, as 16-bit TIF files corrected for ambient radiance values. The UAV speed was 5.0 m s^{-1} and had an average height of 18.4 m for the 1508 m flight distance. All flights were recorded with a resolution of 1.0 cm per px, a front overlap ratio of 75%, a side overlap ratio of 60% and a course angle of 90°. Specific weather conditions for both dates, relating to daytime temperature during flight, wind speed and direction were recorded a weather app (XTC weather), and recorded as 18–23 °C, 6–8 mph and in an easterly direction.

Data processing of the individual UAV images was done by merging and alignment using Agisoft Metashape Professional[43] to create a sparse point cloud. A 3D point cloud was initially created by using the GPS coordinates of each image, which allowed both precise alignment and the formation of solid mesh model. Finally, a composite orthomosaic image was created using the WGS 1984 Web Mercator coordinate system. Then, image analysis of each field site was investigated in the context of 7 different vegetation indices (VI's) (using ArcGIS Pro software). The seven VI's investigated were the normalised difference vegetation index (NDVI), red-edge normalised difference vegetation index ($NDVI_{RE}$), the red-edge simple ratio (SR_{RE}), the green normalised difference vegetation index (GNDVI), the soil adjusted vegetation index (SAVI), the transformed soil adjusted vegetation index (TSAVI) and the modified soil adjusted vegetation index (MSAVI). Each VI was selected due to its different approach for field investigation. The NDVI is one of the most common VI's used in agriculture and crop investigation as it has a sensitive response to green vegetation. However, it can be affected by soil brightness and soil colour.[44] The $NDVI_{RE}$ is used as a VI for estimating vegetation health using red-edge.[45] The SR_{RE} is used to assess the amount of healthy and stressed vegetation. The GNDVI can be deployed as an indicator for estimating photosynthetic activity,[46–48] whereas SAVI has been developed to minimize soil brightness using a correction factor, L (the default value for $L = 0.33$). In similarity, the TSAVI is also deployed with a correction for soil brightness by assuming the soil line has an arbitrary slope (s) and intercept (a) with the inclusion of an additional term to minimise the soil noise (X); the software default values for the terms are $s = 0.33$, $a = 0.5$ and $X = 1.5$.[49] Finally, the MSAVI, attempts to minimise the contribution of bare soil on the SAVI.[50] In addition, ArcGIS Pro was used to create raster profiles for Pb, pH, and Na distributions on both top and bottom fields.

$$\mathrm{NDVI} = \frac{(\mathrm{NIR} - \mathrm{Red})}{(\mathrm{NIR} + \mathrm{Red})} \quad (1)$$

$$\mathrm{NDVI_{RE}} = \frac{(\mathrm{NIR} - \mathrm{RE})}{(\mathrm{NIR} + \mathrm{RE})} \quad (2)$$

$$\mathrm{SR_{RE}} = \frac{(\mathrm{NIR})}{(\mathrm{RE})} \quad (3)$$

$$\mathrm{GNDVI} = \frac{(\mathrm{NIR} - \mathrm{Green})}{(\mathrm{NIR} + \mathrm{Green})} \quad (4)$$

$$\mathrm{SAVI} = \frac{(\mathrm{NIR} - \mathrm{Red})}{(\mathrm{NIR} + \mathrm{Red} + L)} \times (1 + L) \quad (5)$$

$$\mathrm{TSAVI} = \frac{(s \times (\mathrm{NIR} - s \times \mathrm{Red} - a))}{(a \times \mathrm{NIR} + \mathrm{Red} - a \times s + X \times (1 + s^2))} \quad (6)$$

$$\mathrm{MSAVI} = (0.5) \times (2(\mathrm{NIR} + 1) - \sqrt{((2 \times \mathrm{NIR} + 1)^2 - 8(\mathrm{NIR} - \mathrm{Red}))}) \quad (7)$$

Statistical analysis of VI's. For statistical analysis 137 vegetation index values from each VI were extracted within a 2 m radius from the GPS sampling point from the top and bottom fields. From the top field, 72 sampling points were identified, while from the bottom field 65 sampling points were identified. For each sampling point, 10 representative random VI values were selected from within the 2 m radius. The median of the 10 selected VI values was determined to represent the median value of the area under consideration. Using the methodology outlined, VI distributions for the top and bottom field for 7 indices, namely, NDVI, GNDVI, SAVI, Red-edge NDVI, TSAVI, Red-edge SR and MSAVI was computed. To select an appropriate statistical test for comparing the VI values of the top and bottom field for each of the 7 indices, normality tests were undertaken separately for the top and bottom field for each of the 7 indices. Shapiro Wilk was used to test normality of the distributions.[51] While for most of the indices, at least one of the fields (top and bottom) deviated from normal distribution (NDVI, SAVI, Red-edge NDVI, Red-edge SR, MSAVI), while for two indices (GNDVI and TSAVI) both fields followed a normal distribution at 0.05 level of significance. For the indices where distribution for at least one of the two fields deviated from normal distribution, the Mann Whitney test was used to compare groups. The indices where the distribution for both fields followed a normal distribution, an independent T-test was used to compare the top and the bottom field values.

Results and discussion

Quantitative metal analysis and soil properties

Initially the 137 soil samples were analysed for the elements Na, Mg, Al, Si, P, S, Cl, K, Ca, Sc, Ti, V, Cr, Mn, Fe, Ni, Cu, Zn, Ga, As, Rb, Sr, Y, Zr, Cs, Ba, La, Ce, Nd and Pb across both sites (top field, 72 samples and bottom field 65 samples) using EDXRF. Quality control of EDXRF analyses was done by analysis of the twelve CRMs. The results for the 30 elements, uncorrected, were analysed by PCA and PLS-DA. This approach identified a smaller group of elements (8) that had the highest variable importance

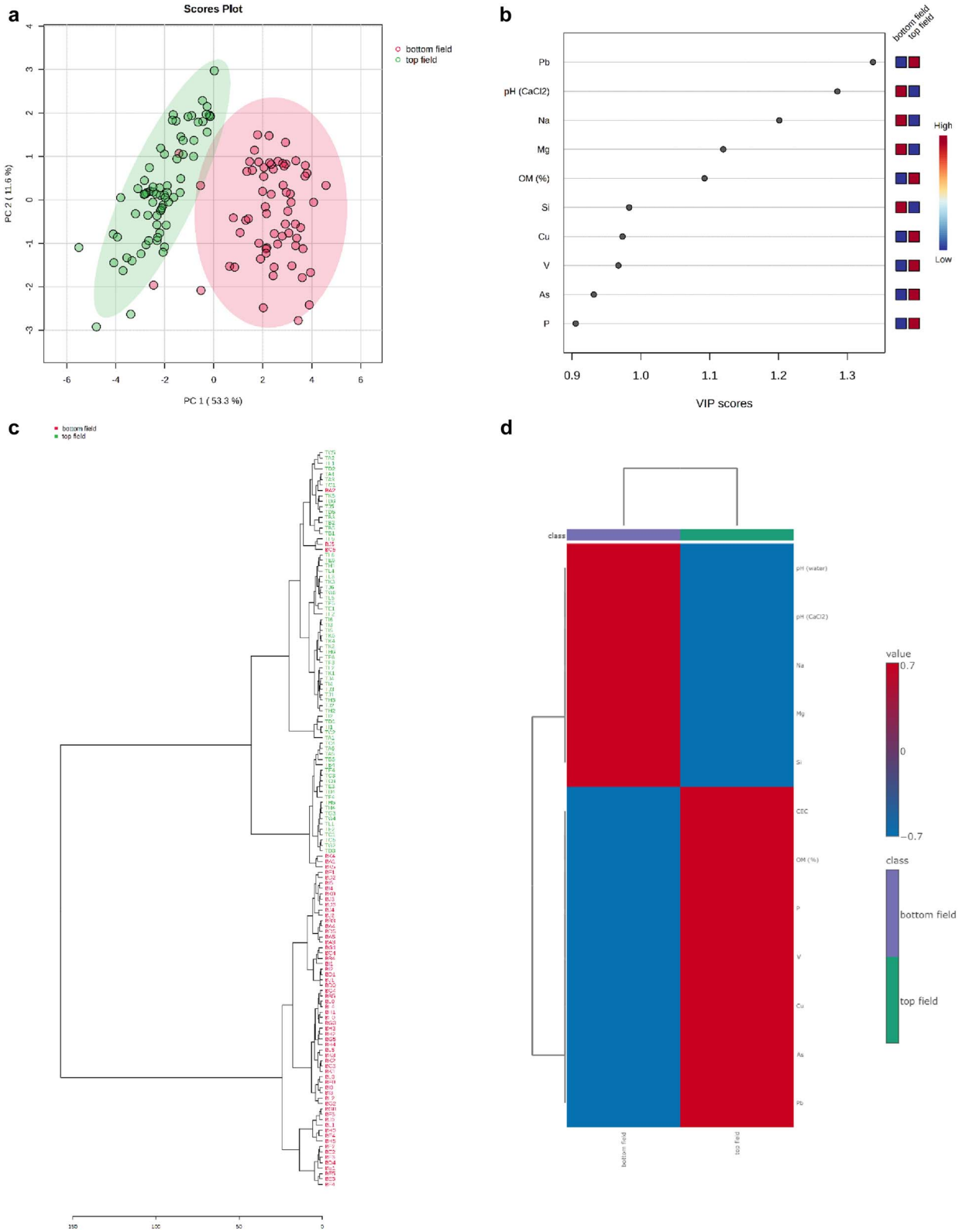

Fig. 2 Assessment of ED-XRF analysis for 8 elements and their associated soil properties from the top field and bottom field. (a) Principal component analysis, (b) partial least squares – discriminant analysis (PLS-DA) variable importance projection (VIP) scores for the top 10 variables, (c) the hierarchical clustering dendrogram* showing a measure of sample similarity, and (d) the associated dendogram*. *The dendrograms were created using the Euclidean measuring distance and the Ward clustering algorithm. Data scaled only for PCA.

in projection (VIP) score (>0.85) *i.e.*, Pb, Na, Mg, Zn, V, Si, As and P. Subsequent data analysis of these 8 elements was done after corrections using their CRM data. Excellent correlation was reported, for the 8 elements, between certified and measured concentrations (typically >0.93) were reported (Table S2), while typical LLODs varied between 2229 μg g^{-1} for Si and 1.2 μg g^{-1} for V, and LOQs varied between 6755 μg g^{-1} for Si and 3.7 μg g^{-1} for V. Intra-precision data was completed at the lower and higher concentration, per element with typical data varying between 0.07 and 3.9% RSD and 0.04–5.1% RSD ($n = 6$). The correct EDXRF data is shown in Table S3.

The pH (in water and $CaCl_2$), organic matter (OM%) and cation exchange capacity (CEC) for the soils from the top ($n = 72$) and bottom ($n = 65$) fields were determined using standard protocols (Table S3). The results are summarised as follows: the organic matter for the top and bottom fields expressed as the mean ± SD are 16.1 ± 2.9 and 10.3 ± 2.9, respectively. For the pH ($CaCl_2$) top field and bottom field expressed as the mean ± SD are 4.0 ± 0.6 and 6.2 ± 0.8, respectively. The CEC, for the top and bottom fields expressed as the mean ± SD are 20.9 ± 3.7 meq. per 100 g and 18.5 ± 6.4 meq. per 100 g, respectively. And finally, the pH (water) top field and bottom field, expressed as the mean ± SD, are 5.3 ± 0.7 and 7.1 ± 0.8, respectively.

The combined dataset (Table S3) was interrogated using statistical software (MetaboAnalyst version 6.0).[29] The Principal Component Analysis (PCA) of the data showed a significant clustering between the top and bottom fields, that were explained by 64.9% of the total variance (Fig. 2(a)). The VIP scores for the top 10 variables *i.e.* 8 elements and 2 soil properties *i.e.* pH($CaCl_2$) and % OM, highlighted significance at $x >$ 1.2 for Pb in the top field and the pH and Na content in the bottom field (Fig. 2(b)). The hierarchical clustering dendogram and associated dendogram, using the Euclidean measuring distance (Fig. 2(c)) and Ward's clustering algorithm (Fig. 2(d)) further highlight the clear difference between the top and bottom fields. Raster plots, using all the sample points in the top and bottom fields, of the data for Pb, pH($CaCl_2$) and Na are shown in Fig. 3. The Pb profile (Fig. 3(a)) across the top and bottom fields clearly shows the highest concentrations of Pb in the southeast region of the top field. While the plot of pH (Fig. 3(b)) shows the more neutral pH aspect of the soil across the bottom field, with a more acid pH across the top field. While the highest concentration of Na is recorded in the bottom field (Fig. 3(c)).

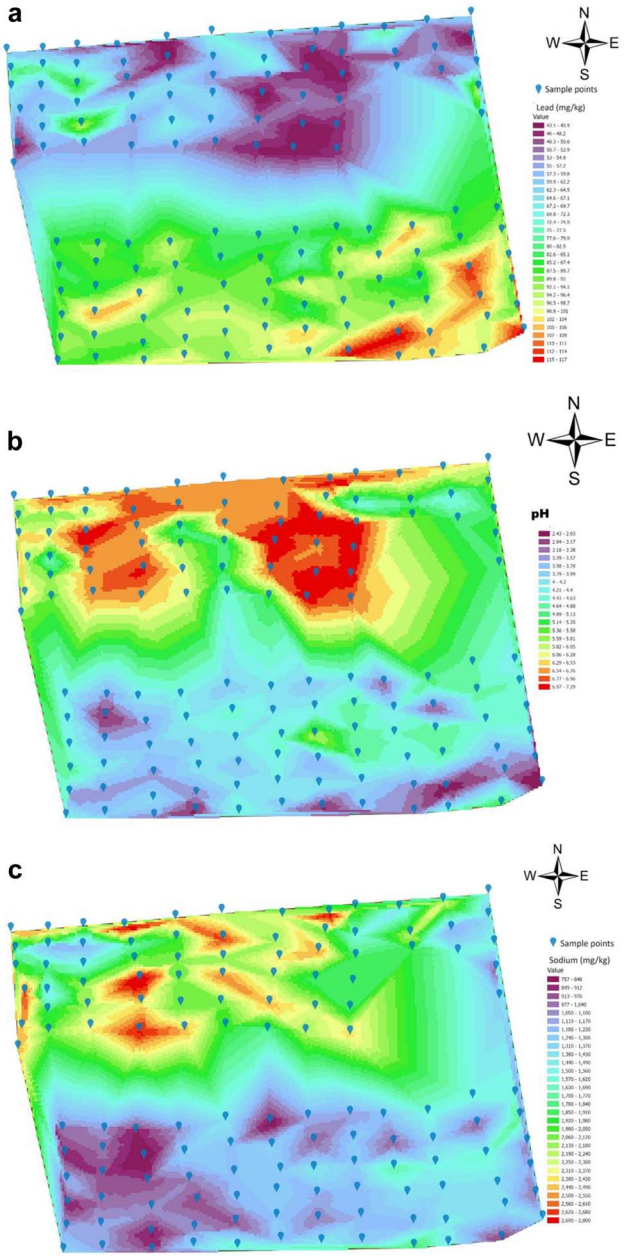

Fig. 3 Raster profiles for selected parameters across both the top and bottom field (a) Pb, (b) pH ($CaCl_2$), and (c) Na.

Metabolomic data analysis

The metabolomic profiling of soil samples (5 samples from both the top and bottom fields) identified 87 metabolites of organic compounds; 50 in positive mode (ESI^+) and 37 in negative mode (ESI^-), with a relative standard deviation of <3% within the 3-quality control (QC) samples. The PCA visualization of the 10 soil samples, five from the top field (sample number TA1, TE5, TG1, TH4 and TL6, corresponding as T = top field, letter = row, number = 1 nearest the top of the slope and 6 nearer the bottom of the slope) and 5 from the bottom field (sample number BA1, BE5, BH3, BK1 and BL4, corresponding as B = bottom field, letter = row, number = 1 closer to the bottom of the embankment and 5 as far as away from the embankment, but within the confines of the field parameter). Fig. 4(a) shows two distinct clusters that were explained by 47.3% of the total variance. The two clusters display a distinct phenotype, which are separated *via* a planar separation across the PC1 space. In addition, the hierarchical clustering dendrogram, using the distance measure of Spearman and the Ward clustering algorithm (Fig. 4(b)), further illustrates the distinct separation of the two groups (top field and bottom field) with the positioning of the pooled QC sample appropriately identified between the two groups.

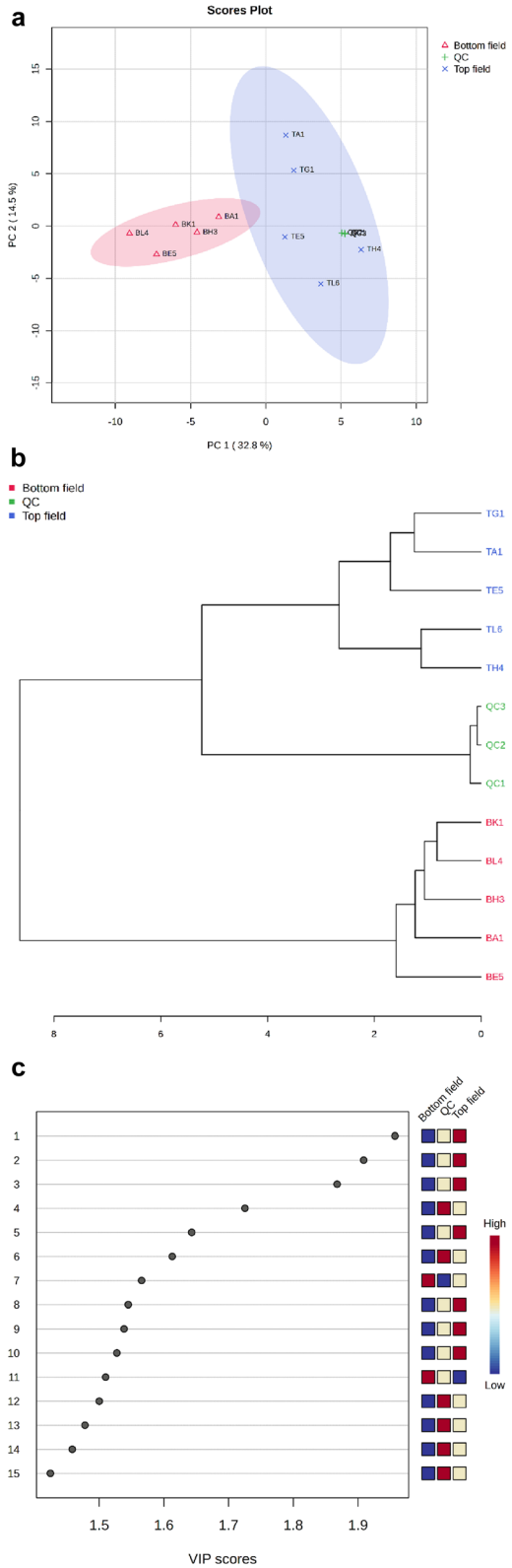

Fig. 4 Metabolomic analysis by hydrophilic liquid interaction chromatography mass spectrometry of soil samples from the top field and bottom field. (a) Principal component analysis of the 87 identified metabolites, (b) the hierarchical clustering dendrogram showing a measure of sample similarity, and (c) the partial least squares – discriminant analysis (PLS-DA) variable importance projection (VIP) scores for the top 15 metabolites. Note: 1 = 7,2′-dihydroxyflavanone; 2 = chrysoeriol; 3 = umbelliferone, 4 = purpurin; 5 = D-glucose; 6 = D-mannitol; 7 = dibutyl sebacate; 8 = 4-hydroxyquinoline; 9 = acridinone; 10 = 1,3-dihydroxyanthraquinone; 11 = diethyl-phthalate; 12 = D-sorbitol; 13 = maltotriose; 14 = isokaempferide; and 15 = betaine.

The VIP scores for the 15 most abundant metabolites which are statistically significant ($x > 1.4$), have been calculated (Fig. 4(c)). The metabolites with the highest VIP scores have a higher discriminating contribution, compared to those with the lowest VIP scores. The data for the top field has 7 important metabolites while the bottom field has 2 important metabolites (as highlighted in red), and with no coincident metabolites. The major metabolites (in terms of the highest VIP score) for the top field are: 7,2′-dihydroxyflavanone, chrysoeriol, umbelliferone, D-glucose, 4-hydroxyquinoline, acridinone, and 1,3-dihydroxyanthraquinone. Thereas the major metabolites in the bottom field are: dibutyl sebacate and diethyl phthalate (highest VIP score first). The top field metabolites are all natural products from biological processes.

In the case of 7,2′-dihydroxyflavanone and chrysoeriol, they are naturally occurring flavone compounds found in plants[52] and used for a number of biochemical processes including pigmentation and oxidation response.[53] Umbelliferone, a 7-hydroxycoumarin, is widely distributed in plants, and specifically *Rutaceae* and *Apiaceae* (Umbelliferae).[54] The carbohydrate, D-glucose, is the main source of energy in living organisms, and is also found in most fungi, bacteria, and plants.[55] While 4-hydroxyquinoline, acridinone and 1,3-dihydroxyanthraquinone (also known as purpuroxanthin or xanthopurpurin) are all natural product alkaloids, identifiable from plants.[56,57] In stark contrast the 2 key metabolites in the bottom field are man-made chemicals. Dibutyl sebacate and diethyl phthalate are both used as plasticizers in the production of plastics and many synthetic rubbers, especially nitrile rubber and neoprene.[58–60]

The average heat map response for the top 25 metabolites is shown in Fig. 5, while the heat map for all 87 metabolites is shown in Fig. S2, across the top and bottom fields. The distinctiveness of the metabolites from both sites is clear, with no overlap for the top 25 metabolites and minimal overlap for all metabolites, irrespective of whether using ESI^+ or ESI^-. Finally, violin plots (Fig. S3) show the relative abundance profiles for the 7 key metabolites in the top field and the 2 key metabolites in the bottom field. The occurrence of 7,2′-dihyroxyflavanone, chrysoeriol and umbelliferone as the three most distinguishing metabolites for the top site is strongly evidenced in the violin plots (Fig. S3(a(i–iii))) with negligible overlap in their profiles. There is less certainty in the occurrence of the metabolites (D-glucose, 4-hydroxyquinoline, acridinone, and 1,3-dihydroxyanthraquinone) as distinguishing features for the top site in Fig. S3(a(iv–vii)). In contrast, dibutyl sebacate, and diethyl phthalate illustrate their distinctiveness as identifiers for the bottom field (Fig. S3(b(viii–ix))). In all cases, the QC determined metabolites, confirm their contribution because of their extraction from pooled soil samples, by their position within the violin plots.

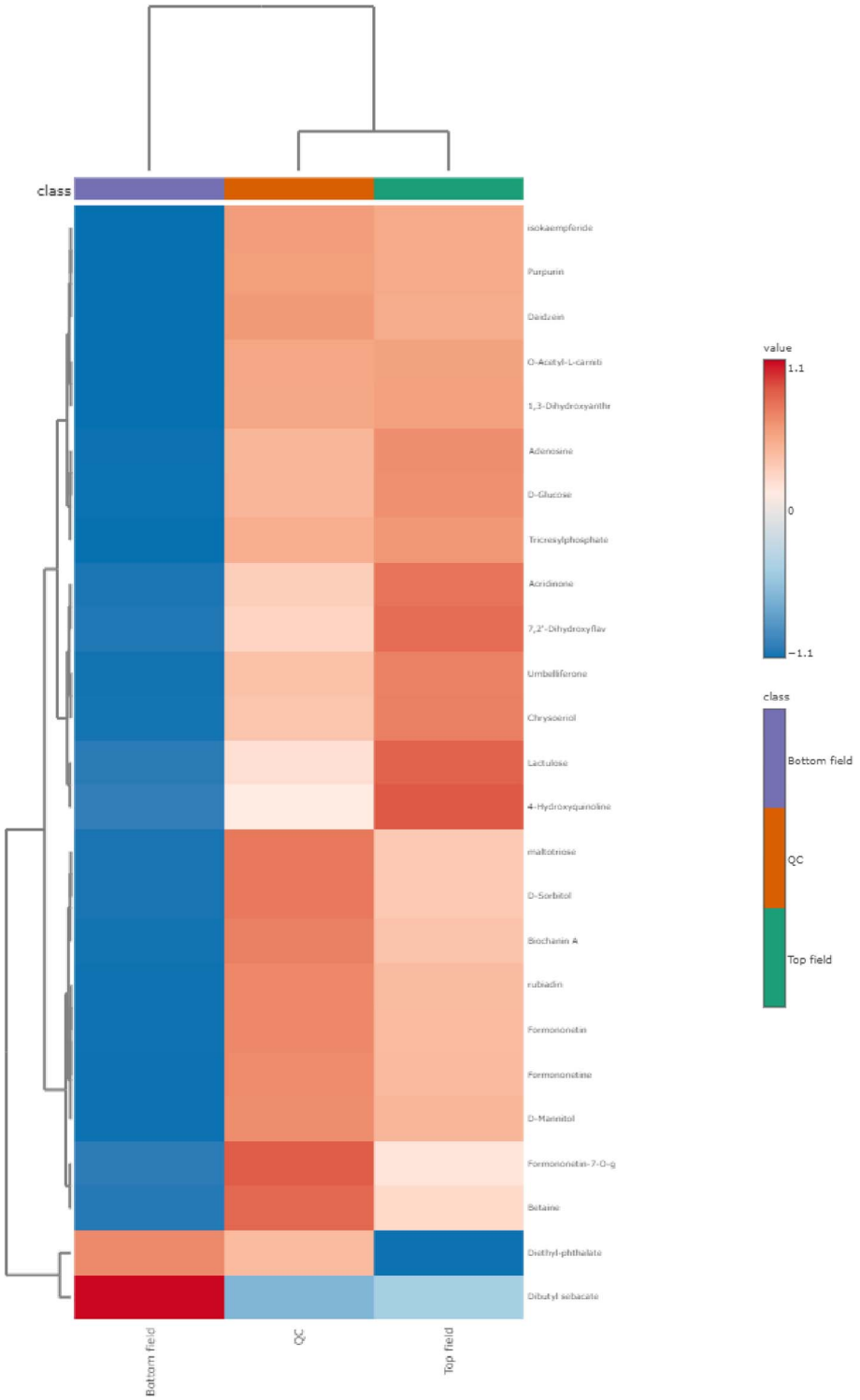

Fig. 5 Heatmap of the relative abundance of the top 25 metabolites.

Metagenomics data analysis

The microbial community composition (α diversity) analysis revealed that field location significantly shaped fungal richness and evenness, but for both bacteria and archaea community richness and evenness were unchanged (Fig. 6 and Table S4). The findings suggest that bacterial diversity in terms of both the number of taxa and their distribution remains relatively stable under the conditions in the study environment. This may further suggest that the communities may be resilient to factors affecting other microbial groups, like the fungi, possibly because of their better adaptation to the environmental conditions. However, the results show that the fungal richness and

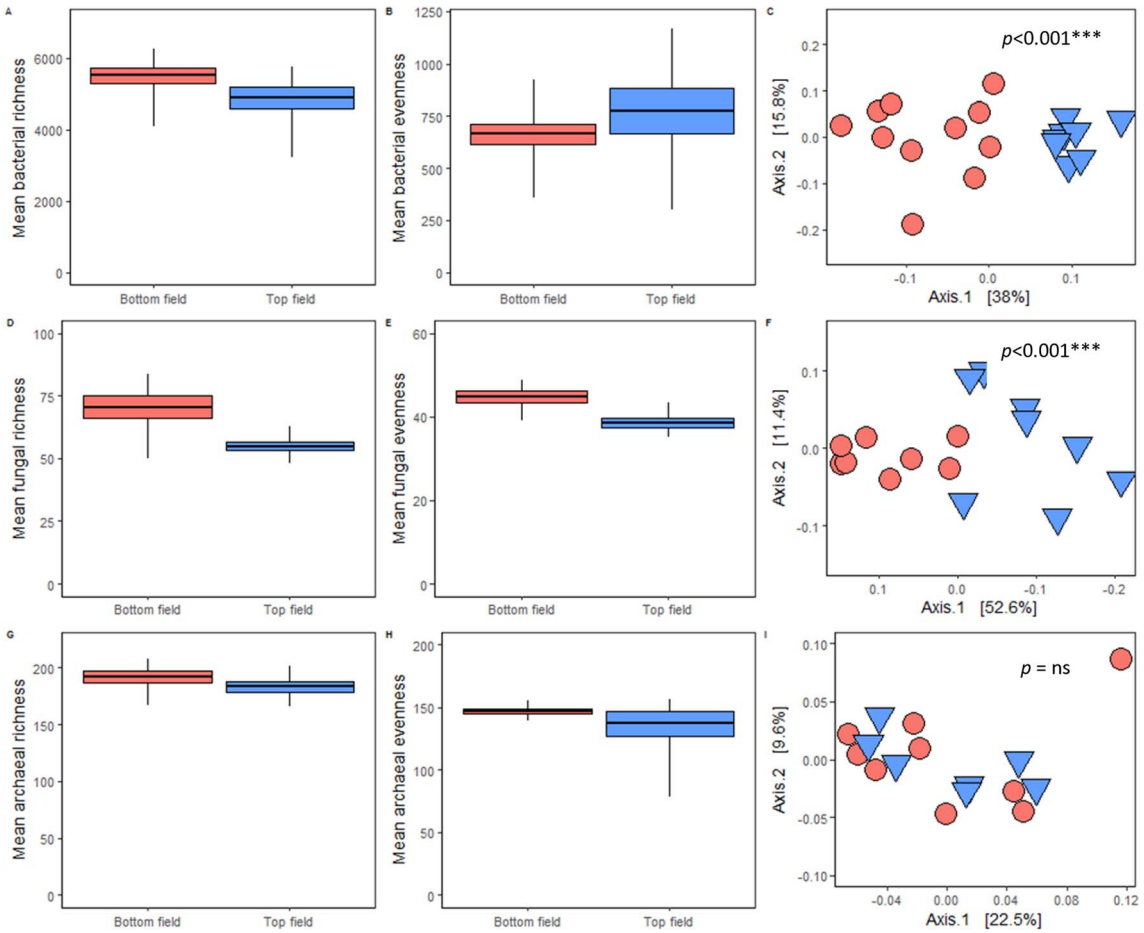

Fig. 6 Microbial alpha and beta diversity for bacterial, fungal and archaeal communities within top and bottom fields. (A) Mean bacterial OTU richness, (B) Mean bacterial evenness, (C) bacterial community composition, (D) mean fungal OTU richness, (E) mean fungal evenness, (F) fungal community composition, (G) mean archaeal OTU richness, (H) mean archaeal evenness, (I) archaeal community composition. Boxplots are presented as mean and standard error of the mean. Field location is distinguished by red to indicate the bottom field and blue to indicate the top field.

evenness are highly significant ($p < 0.001$). This is indicative of an uneven distribution of fungal taxa, with dominance by fewer species in the environmental conditions. These significant changes in the fungal diversity point to strong environmental or anthropogenic factors influencing fungal communities. This pattern often arises in response to environmental disturbances, such as pollution, habitat degradation, or altered soil chemistry.[61,62] Consistent with findings from this study, previous studies have reported that fungi are generally more sensitive to environmental disturbances compared to bacteria due to their reliance on specific organic substrates and symbiotic relationships.[63]

The overall composition of each microbial target group (bacterial, fungal and archaeal communities) between top and bottom samples revealed that there was a significant difference in the composition of bacterial and fungal communities between field location, whereas archaeal communities were indistinguishable (Table S5). The data suggests that bacterial and fungal communities are more responsive to environmental variation between the two sites, while archaeal communities appear relatively stable in the face of these differences. For bacterial β-diversity, the result indicates a highly significant variation between the top and bottom field locations. This aligns with findings from previous studies,[64,65] which demonstrate that bacterial communities are highly sensitive to environmental gradients like soil pH, organic matter content, and moisture. The significant difference in bacterial β-diversity observed here suggests that the environmental conditions at the top and bottom field sites, potentially influenced by factors like soil disturbance or moisture, are driving shifts in bacterial community composition. This variability may reflect the differing ecological niches available at each location, with bacterial communities adapting to the specific conditions of the top and bottom fields.

In addition, fungal β-diversity also shows a significant response to field location. Fungal communities are known to be highly sensitive to changes in soil properties, such as, organic matter content, nutrient availability, and moisture.[66,67] The significant difference in fungal β-diversity between the two field locations suggests that the fungal communities at the top and

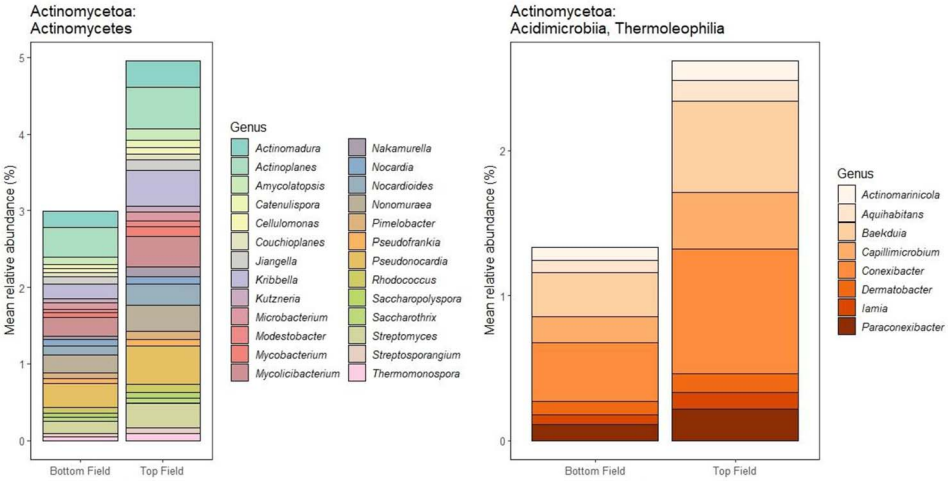

Fig. 7 Stacked bar charts of bacterial OTUs within the phylum Actinomycetoa with significant differences in relative abundances between the top and bottom fields. OTUs were extracted from the main dataframe and those of the same genus were agglomerated. For presentation purposes, genera were then subset to present those belonging to the classes, *Actinomycetes Acidimicrobiia* and *Thermolephilia*.

bottom fields are responding to different ecological conditions, potentially linked to the soil characteristics shaped by the site's history and current vegetation.

The differences in bacterial and fungal taxa were modelled against field location. Analysis of the microbial dataset shows some significant differences between the top and bottom fields depending upon the bacterial OTUs. It is interesting to note that for the phylum *Actinomycetoa*, that in general term the *actinomycetes*, *acidimicrobiia* and *thermolephilia* are all more abundant in the top field (as compared to the bottom field (Fig. 7)). Particularly more abundant, in the top field, are *Actinomycetes streptomyces*, *Actinomycetes pseudonocardia*, *Actinomycetes nonomuraea*, *Actinomycetes nocardiodide*, *Actinomycetes kribbella* and *Actinomycetes actinoplanes* (Fig. 7). For the phylum, *Actinomycetoa*, the *acidimicrobiia* and *thermoleophilia* are more dominant in the top field (Fig. 7). Particularly, *Thermoleophilia paraconexibacter*, *Thermoleophilia conexibacter*, *capillimicrobium*, and *Thermoleophilia baekduia*. These *Actinomycetota*, which are subdivided into the classes *Actinomycetes*, *Acidimicrobiia*, and *Thermoleophilia*, display distinct distributions. The better-drained embankment soils (Top field) favour the aerobic *actinomycetes* which thrive on recalcitrant organic matter, whereas the waterlogged marshland (Botton field) provides niches for taxa tolerant of lower oxygen and higher metal concentrations. The phylum *Actinomycetota* are a range of Gram-positive bacteria with a high guanine-cytosine (GC) content, two of the main nucleotide bases found in the nucleic acids DNA and RNA. The phylum is highly important to plant life due to their role in decomposing soil organic matter, tolerating heavy metals, and producing secondary metabolites.[68–70] Their presence is important for producing nutrient rich soils. Genera such as *Streptomyces*, *Nocardia*,

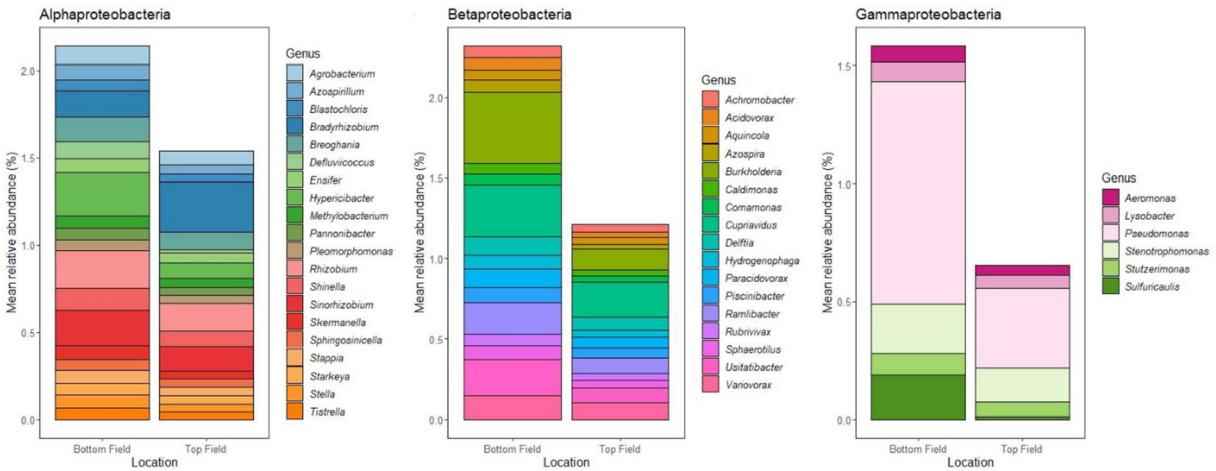

Fig. 8 Stacked bar charts of bacterial OTUs within the phylum Pseudomonadota (synonym Proteobacteria) with significant differences in relative abundances between the top and bottom fields. OTUs were extracted from the main dataframe and those of the same genus were agglomerated. For presentation purposes, genera were subset to present those belonging to the classes, *Alphaproteobacteria*, *Betaproteobacteria* and *Gammaproteobacteria*.

Rhodococcus, and *Micromonospora*, which are critical for nutrient cycling, pollutant degradation, and plant growth promotion,[71–74] occur in both fields, although their relative abundances vary according to localized environmental gradients.

In contrast, other bacteria within the phylum *Pseudomonadota* (synonym *Proteobacteria*) have shown more abundance in the bottom field (Fig. 8). Of note are members of the genus *Alphaproteobacteria*, which are especially more abundance in the bottom field of *Alphaproteobacteria sinorhizobium*, *Alphaproteobacteria rhizobium*, *Alphaproteobacteria hypericibacter* (Fig. 8). Interestingly *Alphaproteobacteria bradyrhizobium* is more abundant in the top field (Fig. 8). Within the genus *Betaproteobacteria*, the most abundant bacteria in the bottom field are *Betaproteobacteria usitatibacter*, *Betaproteobacteria ramlibacter*, *Betaproteobacteria cupriavidus*, and *Betaproteobacteria burkholderia* (Fig. 8). While in the genus *Gammaproteobacteria*, the most abundant bacteria in the bottom field are *Gammaproteobacteria sulfuricaulis*, *Gammaproteobacteria stenotrophomonas*, and *Gammaproteobacteria pseudomonas* Fig. 8.

The phylum *Pseudimonadota*, are a group of Gram-negative bacteria. Within the *Alphaproteobacteria* it was possible to identify the presence of *rhizobium*, a bacterium that fixes nitrogen. These important nitrogen fixing bacteria are endosymbiont within the root nodules of flowering plants. The *Betaproteobacteria* Gram-negative bacteria have an important role in soil with respect to maintaining pH, while one of the identified species (Fig. 8(b)), in the bottom field, *burkholderia* is a pathogenic bacterium. The Gram-negative *Gammaproteobacteria* identified were dominated by the presence of *Pseudomonas*, predominantly in the bottom field, but also in the top field. It has been found that some members of the Genus, *Pseudomonas*, can be used for bioremediation in the soil environment due to their ability to metabolise organic molecules.[75] Their higher abundance in the bottom field, could indicate their current role in bioremediation.

These classes, which include genera with varied metabolic capabilities, from nitrogen fixation and plant growth promotion to pollutant degradation and sulfur cycling, indicate that the soil microbial community is well-equipped for essential processes vital for ecosystem recovery in post-mining landscapes.[75–78] In particular, the waterlogged conditions in the marshland (Botton field) tend to select for anaerobic or facultative anaerobic taxa, while the well-drained embankment (Top field) supports aerobic, root-associated bacteria that enhance plant establishment and growth.

Additionally, amongst the bacteria OTUs is the significant abundance within the phylum Myxococcota, *Myxococcota anaeromyxobacter* in the bottom field (Fig. 9). This facultative anaerobic myxobacterium can survive and grow both in the presence or absence of oxygen. These Gram-negative bacteria are known for social predation and complex substrate degradation, playing a significant role in nutrient turnover.[79,80] Of interest is the presence of the *Thermodesulfobacteria*, and the genus *Desulfosarcina*. These Gram-negative, sulphate-reducing bacteria are known to thrive in extreme environments as characterised by high pressure and temperature.[81,82]

Overall, these integrated results suggest that the bacterial community in this post-mining landscape is both highly heterogeneous and functionally diverse, with distinct assemblages adapted to the contrasting conditions of the embankment (Top field) and marshland (bottom field). Environmental gradients including moisture, oxygen availability, nutrient levels, and contamination, are key drivers of microbial niche differentiation. This functional diversity underpins essential ecosystem processes such as organic matter decomposition, nutrient cycling, and contaminant transformation, all of which

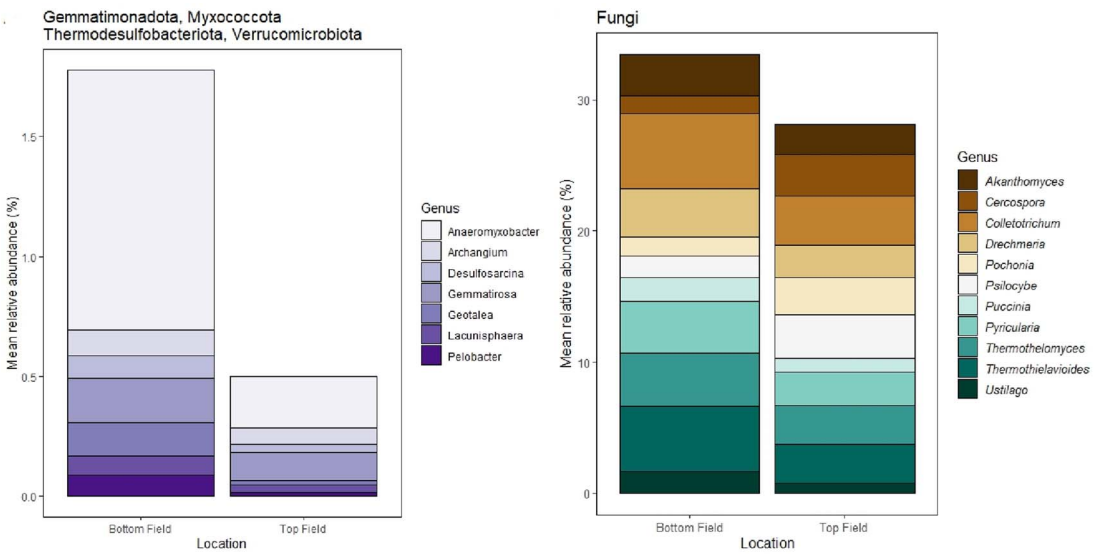

Fig. 9 Stacked bar charts representing bacterial OTUs belonging to additional phyla and all fungal taxa with significant differences in relative abundances between the top and bottom fields. OTUs were extracted from the main data frame and those of the same genus were agglomerated.

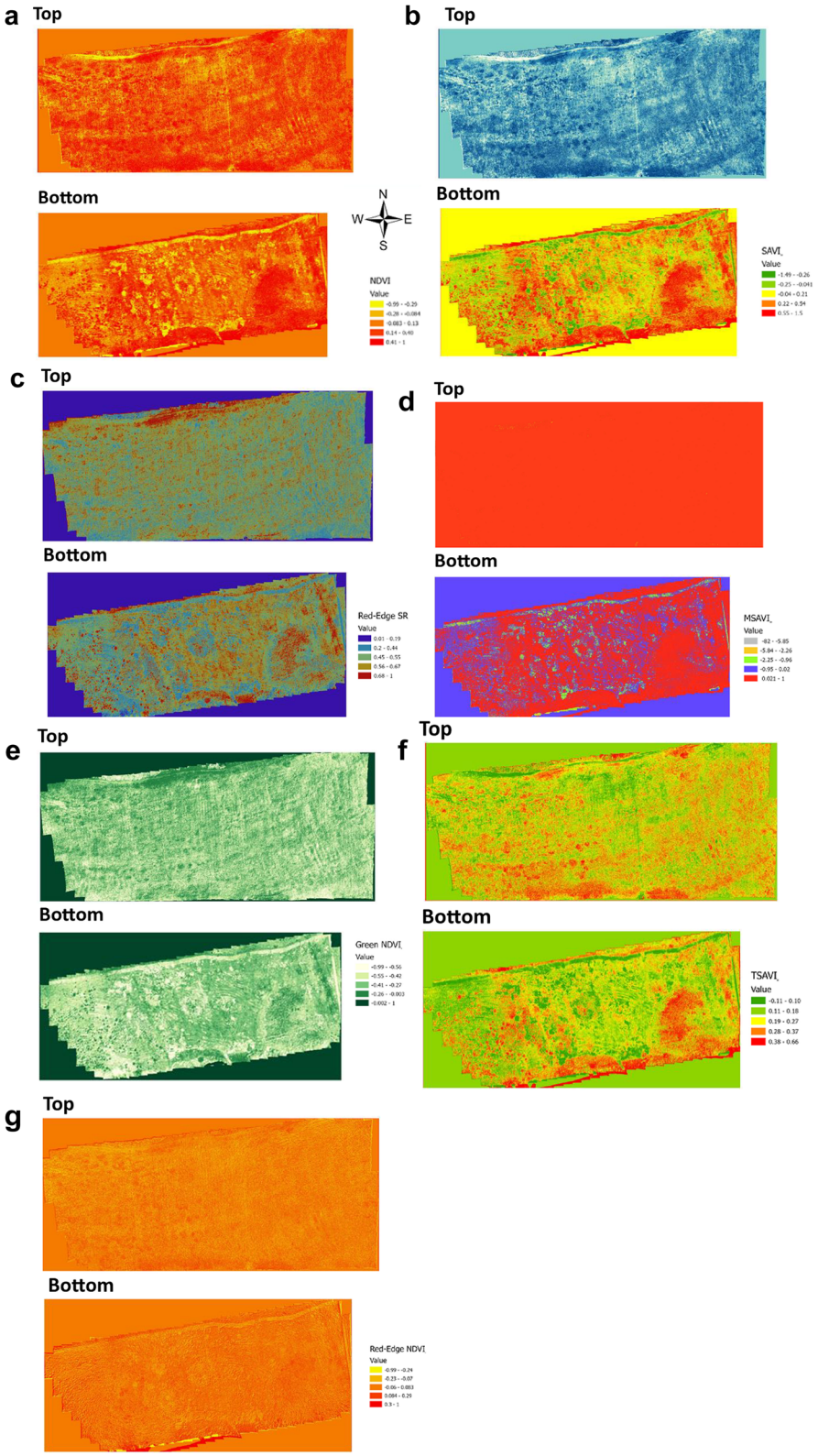

Fig. 10 Vegetation indices profiles using (a) NDVI, (b) SAVI, (c) Red-edge SR, (d) MSAVI, (e) GNDVI, (f) TASVI, and (g) Red-edge NDVI, for both the top and bottom fields.

are vital for the ongoing recovery and restoration of soil health in disturbed environments.[83]

Finally, the fungal taxa are evenly balanced in their occurrence between the top and bottom field with only minor variations in abundance (Fig. 9). One of the most abundant genera of fungi identified is *colletrotrichum*, which are symbionts to plants as endophytes (*i.e.* they live within the plant). The presence of *colletotrichum* highlights its role as a plant pathogen commonly associated with anthracnose diseases.[84,85] The persistence of *colletotrichum* in soil is particularly relevant in disturbed or reclaimed environments, as its presence is often associated with stressed or weakened vegetation. According to a previous report, *colletotrichum* species are known to produce micro-sclerotia, specialized survival structures that enable dormancy in the soil during unfavourable conditions such as winter or environmental stress, with these structures remaining viable for several years.[86] This suggests that its detection in this study may serve as an indicator of previous environmental pressures on plant health, potentially linked to residual contamination, nutrient imbalances, or other ecological stressors. Its prevalence in the bottom field, where marshland conditions could impose additional stress on plants, reinforces its role as a marker of environmental degradation. At the same time, its detection in the top field may reflect the early stages of ecological succession, where pioneer plant species susceptible to fungal pathogens are beginning to establish in the embankment soils. Another important genus of fungi present are the filamentous *Thermothielavioides* and *Thermothelomyces* which have the unusual characteristic of showing optimal growth above 45 °C.[87,88]

Remote sensing of field locations

The use of an unmanned aerial vehicle (UAV) with multispectral image (MSI) data has previously been shown[89] to be an effective tool to assess sub-surface contamination, and especially in mining contaminated landscapes.[90,91] The UAV-MSI data is initially processed to generate a variety of vegetation indices, which themselves can be used as indicators of vegetative stress. The generated VI profiles for both the top and bottom fields are shown in Fig. 10. By comparing the VI's, determined using the UAV-MSI and mathematical modelling, of the data allows a statistical interpretation (Table S6) of whether the vegetation was under stress.

Due to the non-normal distribution of the NDVI index for the top ($W = 0.96, p = 0.022$) and bottom ($W = 0.894, p < 0.001$) field, comparison of the mean NDVI value for the top and bottom field was undertaken using the Mann Whitney test. The top field ($N = 72$) has a larger mean rank (80.42) than the bottom field ($N = 65$) with mean rank (56.35). A statistically significant difference (at $\alpha = 0.05$) in the NDVI values for the top and bottom field was found ($U = 1518, p < 0.001$) (Table S6(a)). The results indicate that the top field has more vegetation than the bottom field, a difference that is found to be statistically significant.

Similar findings were recorded for the top ($N = 72$) and bottom field ($N = 65$) on SAVI ($U = 1492, p < 0.001$) (Table S6(b)), Red-edge SR ($U = 1390, p < 0.001$) (Table S6(c)) and MSAVI ($U = 1831, p < 0.028$) (Table S6(d)) indices/values.

The top ($N = 72$) and bottom ($N = 65$) field data for GNDVI values/indices follow normal distribution and have equal variances, hence a *t*-test (Table S6(e)) was undertaken for comparison. There is statistically significant difference in the GNDVI values for the top and bottom fields, $t(135) = -2.927, p = 0.004$, with the top field (Mean = 0.200, SD = 0.0625) having larger/higher values than the bottom field (Mean = 0.169, SD = 0.0606). From the result, statistically significant higher vegetation level in the top field can be inferred.

Though the TSAVI data for the top ($N = 72$) and bottom ($N = 65$) field follow normal distribution (Table S6(f)) , the equality of variance assumption was violated ($F = 5.525$, $p = 0.020$). Outcomes similar to GNDVI top and bottom field comparison was found for the TSAVI data, $t(110.17) = -2.726$, $p = 0.007$, with moderate effect size (Cohen's $d = -0.47$). As Red-edge NDVI data for the top ($N = 72$) and bottom ($N = 65$) field deviate from normal distribution (Table S6(g)), a Mann–Whitney test was undertaken. Unlike outcomes on other UAV-MSI data (NDVI, SAVI, Red-edge SR, MSAVI, GNDVI and TSAVI) for the same top and bottom field, the bottom field has a higher mean rank (89.91) than the top field (50.13). The mean rank of the bottom field being higher than the top field may be attributed to the strategy for estimating Red-edge NDVI indices, which might be distinctly different from the way other indices have been estimated. In coherence with results from other UAV-MSI data, statistically significant difference was found in the top and bottom field vegetation in Red-edge NDVI data ($U = 981, p < 0.001$).

After comparing all the potential VI's it was found that the best indicators of vegetative coverage under stress were represented by NDVI, GNDVI, SAVI, TSAVI, Red-edge SR and MSAVI.

Conclusion

Both the top and bottom fields contrast each other across all the analyses investigated. The presence of Pb in the top field 93.9 ± 8.2 mg kg^{-1} $n = 72$ (mean ± SD, mg kg^{-1}, n = number of determinations) is significantly different from Pb in the bottom field 58.4 ± 12.0 mg kg^{-1} $n = 65$. The presence of Pb largely in the top field, which was an integral part of the historical coal mining spoil heap, is testament to the known immobility[25,92] of this legacy anthropogenic element in aged soils. It should also be recognised that the Pb concentration, even at the highest concentrations do not exceed the current Environment Agency category 4 screening level concentration for a public open space (1300 mg kg^{-1})[93] though other categorizations are available for other land uses and locations. Based on the visual plant diversity (Table S7) which was identified and counted across both the top and bottom fields has provided an acknowledgement of the health of the top field. However, while Pb is present at a higher concentration in the top field it does not have a detrimental effect on plant growth. It is also noted that the Pb concentrations determined are lower than that reported (221.7 mg kg^{-1}) for the geographical area at Weetslade Country Park.[94] Further

details on the determination by XRF, and the associated statistical data treatment are available on their website.

The higher concentration of Na 2050 ± 430 mg kg^{-1}, $n = 65$ (mean ± SD, mg kg^{-1}, n = number of determinations) in the bottom field contrasts with the level in the top field (1204 mg kg^{-1}, $n = 72$). Also, the soil pH, using $CaCl_2$, 6.2 ± 0.8, $n = 65$ (mean ± SD, n = number of determinations), in the bottom field contrasts with the pH in the top field (4.0 ± 0.6, $n = 72$).

The geographical area of Weetslade Country Park has a typical Na concentration of 0.11% (0.2% Na_2O) (UKSO, 2016). This low concentration for Na is matched with the level of Na in the top field soil. Whereas the bottom field has a much higher concentration (typically by a factor of 2). The UKSO site[94] characterises (using Soilscapes for England and Wales) the Weetslade Country Park area generically as slowly permeable seasonally wet slightly acid but base rich loamy and clayey soils. Furthermore, the health of the top field, where the most abundant metabolites detected by HILIC, where metabolites of natural products remain are indicative of the soil richness. This is contrasted with the bottom field where the most abundant metabolites detected are man-made chemicals, which are themselves indicative of brownfield sites with a former industrial past.

The microbial investigation contributes significantly to the overall investigation of the site. The more dominant presence of the phylum *Actinomycetota* in the top field is indicative of the fertile soil and its ability to produce an abundance of flora in Spring and Summer. The presence, to some extent, of a similar abundance of the genus *Rhizobium* and its ability to fix nitrogen, is evident in the top field due to the significant flora coverage (Table (S7)) but is less evident, currently, in the bottom field which has a lack of flowering plants. The dominance, in the bottom field, of the genus *Betaproteobacteria* is testament to the role of these pH maintaining bacteria in establishing a more neutral pH (pH 6.2, in the bottom field compared to pH 4.0 in the top field). The presence of the bacteria from the genus *Gammaproteobacteria*, and specifically the abundance in both fields of *Pseudomonas* is indicative of the bioremediation that the soil is undergoing. Finally, the presence of the bacteria, *Thermodesulfobactereiota*, and the fungi, *Thermothielavioides* and *Thermothelomyces* are indicative of the sites former use as a colliery. These bacteria and fungi have the unusual characteristic of either thriving in extreme environments (*i.e.* coal mines) or showing optimal growth above 45 °C, respectively, all characteristic of the extreme environments from where coal was mined from seams on this site up to 350 m below ground.

The rapid deployment of a UAV-MSI could be used to assess potential vegetative coverage and stress prior to any fieldwork sampling and analyses. This useful approach has been previously deployed to assess contaminated land sites[89] as well as an indicator of agricultural crop development and yield prediction.[95–97] In this research, the deployment of a UAV-MSI on a former coal mining site, now operating as a country park for the past 17 years, has enabled a rapid differentiation between the two fields, in terms of their vegetation "health". This use and deployment of a UAV-MSI could be used alongside more traditional approaches for preliminary investigation of potential contaminated land sites, such as, historic maps and current and former land-owner searches to assist with decision-making. The further site investigation using next-generation analytical techniques has enabled the level of sub-surface contamination to be assessed, as well as indicate the next steps in terms of decontamination processes for the site.

In summary, each of the different types of environmental forensic analyses have contributed to an understanding of the complex soil environment on this former industrial coal mining site. The top site is characterised by an acidic soil, which while having a higher Pb concentration, has a lower sodium content, more natural organic metabolites, and a plethora of more nutrient producing bacteria. In contrast the bottom site, with its more neutral pH, lower Pb concentration, higher sodium content, and more man-made metabolites, is characterised by the presence of bacteria capable of undertaking soil remediation. The use of a UAV-MSI, and the generation and interpretation of vegetation indices, can provide a rapid tool, alongside more traditional approaches, for establishing a base-line environmental context prior to further on-site investigation.

Author contributions

Conceptualization: K. R., C. E. N., J. J. P., J. R. D. data curation: I. A. S., K. R., N. B., W. C., C. E. N., J. J. P., J. R. D. formal analysis: I. A. S., K. R., N. B., W. C., C. E. N., J. J. P., J. R. D. funding acquisition: J. R. D., J. J. P. investigation: I. A. S., K. R., C. E. N., J. R. D. methodology: I. A. S., K. R., N. B., W. C., C. E. N., J. J. P., J. R. D. project administration: K. R., C. E. N., J. J. P., J. R. D. resources: K. R., N. B., W. C., C. E. N., J. J. P., J. R. D. supervision: C. E. N., J. J. P., J. R. D. writing – original draft: I. A. S., K. R., N. B., W. C., C. E. N., J. J. P., J. R. D. writing – review & editing: K. R., N. B., W. C., C. E. N., J. J. P., J. R. D.

Conflicts of interest

The authors have no conflicts of interest to declare.

Data availability

Raw data from the metagenomic and metabolomic analyses can be accessed at this DOI; **https://doi.org/10.25398/rd.northumbria.29460011**.

All other data supporting this article have been included as part of the SI. Supplementary information: (a) photography of the two sites, (b) a summary of metabolite information from the metabolomics, (c) data wrt the microbial dataset, (d) XRF QC and experimental data, (e) statistical evaluation of the vegetation indices, (f) plant identification data. See DOI: **https://doi.org/10.1039/d5va00229j**.

Acknowledgements

The authors acknowledge the contribution of Jack Lightbown who participated in the soil field sampling. The NUOMICS facility, at Northumbria University, for the eDNA analyses. Access to Weetslade Country Park was granted by Mr Geoff Dobbins, Senior Estates officer (Reserves), Northumberland Wildlife Trust.

References

1 Y. Liu and X. Pan, Ecotope-based Urban Post-industrial Landscape Design, *IERI Procedia*, 2014, **9**, 185–189.
2 M. Tughyan, Impact of coal mining on environment and natural resources – causes and potential effects, *Int. J. Min. Sci.*, 2020, **6**(1), 32–36, DOI: 10.20431/2454-9460.0601003.
3 P. Gopinathan, T. Subramani, S. Barbosa and D. Yuvaraj, Environmental impact and health risk assessment due to coal mining and utilization, *Environ. Geochem. Health*, 2023, **45**, 6915–6922, DOI: 10.1007/s10653-023-01744-z.
4 K. Rodgers, I. McLellan, T. Peshkur, R. Williams, R. Tonner, C. W. Knapp, F. I. Henriquez and A. S. Husthouse, The legacy of industrial pollution in estuarine sediments: spatial and temporal variability implications for ecosystem stress, *Environ. Geochem. Health*, 2019, **42**, 1057–1068, DOI: 10.1007/s10653-019-00316-4.
5 K. C. Williams and J. C. Hoffman, Remediation to Restoration to Revitalization: Engaging Communities to Support Ecosystem-Based Management and Improve Human Wellbeing at Clean-up Sites, in *Ecosystem-Based Management. Ecosystem Services and Aquatic Biodiversity*, ed. T. O'Higgins, M. Lago and T. DeWitt, Springer, Cham, 2020, pp. 543–559, DOI: 10.1007/978-3-030-45843-0_27.
6 S. Nimbrana, P. Ranga and A. Malik, Bioremediation of Mining Sites: Sustainable Approach to Restore a Healthy Ecosystem, in *Heavy Metal Toxicity: Environmental Concerns, Remediation and Opportunities*, ed. R. P. Singh, P. Singh and A. Srivastava, Springer, Singapore, 2023, pp. 341–362, DOI: 10.1007/978-981-99-0397-9_16.
7 A. Rouhani, J. Skousen and F. M. G. Tack, An Overview of Soil Pollution and Remediation Strategies in Coal Mining Regions, *Minerals*, 2023, **13**(8), 1064, DOI: 10.3390/min13081064.
8 O. N. Bustos López, Role of Modern Innovative Techniques for Assessing and Monitoring Environmental Pollution, in *Bioremediation and Biotechnology*, ed. K. Hakeem, R. Bhat and H. Qadri, Springer, Cham, 2020, pp. 75–91, DOI: 10.1007/978-3-030-35691-0_4.
9 P. M. Mach and G. F. Verbeck, Analytical Methods and Trends in Environmental Forensics, in *Development and Environment*, ed. W. Burggren and B. Dubansky, Springer, Cham, 2018, DOI: 10.1007/978-3-319-75935-7_11.
10 R. D. Morrison, Critical review of environmental forensic techniques: part 1, *Environ. Forensics*, 2020, **1**(4), 157–173, DOI: 10.1006/enfo.2000.0017.
11 A. Souloutzoglou and A. Tasopoulou, The Methods and Techniques of Strategic Environmental Assessment. Comparative Evaluation of Greek and International Experience, *Sustainability*, 2020, **12**(8), 3310, DOI: 10.3390/su12083310.
12 D. L. Saber, D. Mauro and T. Sirivedhin, Applications of forensic chemistry to environmental work, *J. Ind. Microbiol. Biotechnol.*, 2005, **32**(11–12), 665–668, DOI: 10.1007/s10295-005-0013-z.
13 M. A. Z. Chahouki, Multivariate Analysis Techniques in Environmental Science, *Earth and Environmental Sciences*, ed. I. Ahmad Dar, InTech, 2011, pp. 539–564, DOI: 10.5772/26516, ISBN: 978-953-307-468-9, available from: http://www.intechopen.com/books/earth-and-environmental-sciences/multivariate-analysis-techniques-in-environmental-science.
14 E. Preziosi, G. Giuliano and R. Vivona, Natural background levels and threshold values derivation for naturally As, V and F rich groundwater bodies: A methodological case study in Central Italy, *Environ. Earth Sci.*, 2010, **61**(5), 885–897, DOI: 10.1007/s12665-009-0404-y.
15 W. Chen, Y. Cai, K. Zhu, J. Wei and Y. Lu, Spatial heterogeneity analysis and source identification of heavy metals in soil: a case study of Chongqing, Southwest China, *Chem. Biol. Technol. Agric.*, 2022, **2022**, 9–50, DOI: 10.1186/s40538-022-00313-3.
16 K. Kamboj, S. Sisodiya, A. K. Mathur, A. Zare and P. Verma, Assessment and Spatial Distribution Mapping of Criteria Pollutants, *Water, Air, Soil Pollut.*, 2022, **233**, 82, DOI: 10.1007/s11270-022-05522-y.
17 M. Llaver, E. F. Fiorentini, M. N. Oviedo, P. Y. Quintas and R. G. Wuilloud, Elemental Speciation Analysis in Environmental Studies: Latest Trends and Ecological Impact, *Int. Res. J. Publ. Environ. Health*, 2021, **18**(22), 12135, DOI: 10.3390/ijerph182212135.
18 S. M. Mudge, Multivariate and Geostatistical Methods in Environmental Forensics, in *Methods in Environmental Forensics*, CRC Press, 1st edn, 2008.
19 L. A. McGregor, C. Gauchotte-Lindsay, N. N. Daeid, R. Thomas and R. M. Kalin, Multivariate Statistical Methods for the Environmental Forensic Classification of Coal Tars from Former Manufactured Gas Plants, *Environ. Sci. Technol.*, 2012, **46**, 3744–3752.
20 S. Sutliff-Johansson, S. Pontér, A. Mäki, E. Engström, I. Rodushkin, P. Peltola and A. Widerlund, Groundwater environmental forensic investigation combining multivariate statistical techniques and screening analyses, *Environ. Forensics*, 2020, 1–15, DOI: 10.1080/15275922.2020.1850571.
21 H. I. Chaminé, A. J. S. C. Pereira and A. C. Teodoro, Remote sensing and GIS applications in earth and environmental systems sciences, *SN Appl. Sci.*, 2021, **3**, 870, DOI: 10.1007/s42452-021-04855-3.
22 A. Hussain, S. A. Bangroo and M. Muslim, Assessment of Soil Contamination Using Remote Sensing and Spatial Techniques, in *Geospatial Analytics for Environmental Pollution Modeling*, ed. F. Mushtaq, M. Farooq, A. B. Mukherjee and M. Ghosh Nee Lala, Springer, Cham, 2023, pp. 249–266, DOI: 10.1007/978-3-031-45300-7_10.
23 The Land Trust, https://thelandtrust.org.uk, accessed June 19, 2025.
24 Northumberland Wildlife Trust, https://www.nwt.org.uk, accessed June 19, 2025.
25 J. R. Dean, P. M. Amaibi, A. Okorie and J. A. Entwistle, A critical evaluation of the use and 'misuse' of As and Pb bioaccessibility data in human health risk assessments,

Environ. Res., 2020, **189**, 109915, DOI: 10.1016/j.envres.2020.109915.

26 H. Gould, W. Cheung, J. D. Finnigan, J. Muñoz-Muñoz, S. J. Charnock and G. B. Black, In search of complementary extraction methods for comprehensive coverage of the *Escherichia coli* metabolome, *Metabolites*, 2023, **13**, 1010, DOI: 10.3390/metabo13091010.

27 W. B. Dunn, W. Lin, D. Broadhurst, P. Begley, M. Brown, E. Zelena, A. A. Vaughan, A. Halsall, N. Harding, J. D. Knowles, S. Francis-McIntyre, A. Tseng, D. I. Ellis, S. O'Hagan, G. Aarons, B. Benjamin, S. Chew-Graham, C. Moseley, P. Potter, C. L. Winder, C. Potts, P. Thornton, C. McWhirter, M. Zubair, M. Pan, A. Burns, J. K. Cruickshank, G. C. Jayson, N. Purandare, F. C. Wu, J. D. Finn, J. N. Haselden, A. W. Nicholls, I. D. Wilson, R. Goodacre and D. B. Kell, Molecular phenotyping of a UK population: defining the human serum metabolome, *Metabolomics*, 2015, **11**(1), 9–26, DOI: 10.1007/s11306-014-0707-1.

28 O. Fiehn, D. Robertson, J. Griffin, *et al.*, The metabolomics standard initiative (MSI), *Metabolomics*, 2007, **3**, 175–178, DOI: 10.1007/s11306-007-0070-6.

29 *MetaboAnalyst 6.0*, 2025, available at: https://www.metaboanalyst.ca/.

30 F. Krueger, F. James, P. Ewels, E. Afyounian, M. Weinstein, B. Schuster-Boeckler, G. Hulselmans, *Sclamons, TrimGalore: v0.6.10*, 2023, https://zenodo.org/records/7598955.

31 D. E. Wood, J. Lu and B. Langmead, Improved metagenomic analysis with Kraken 2, *Genome Biol.*, 2019, **20**(257), 1–13, DOI: 10.1101/762302.

32 J. Lu, F. P. Breitwieser, P. Thielen and S. L. Salzberg, Bracken: estimating species abundance in metagenomics data, *PeerJ. Comput. Sci.*, 2017, **3**, e104, DOI: 10.7717/peerj-cs.104.

33 R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, 2021, https://www.R-project.org/.

34 F. J. Oksanen, F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P. Minchin, R. B. O'Hara, G. Simpson and P. Solymos, Vegan: Community Ecology Package, *R Package Version 2.4-4*, 2017, https://CRAN.R-project.org/package=vegan.1-298, https://www.researchgate.net/publication/323265820.

35 S. Weiss, Z. Z. Xu, S. Peddada, A. Amir, K. Bittinger, A. Gonzalez, C. Lozupone, J. R. Zaneveld, Y. Vázquez-Baeza, A. Birmingham, E. R. Hyde and R. Knight, Normalization and microbial differential abundance strategies depend upon data characteristics, *Microbiome*, 2017, **5**(27), 1–18, DOI: 10.1186/s40168-017-0237-y.

36 D. T. McKnight, R. Huerlimann, D. S. Bower, L. Schwarzkopf, R. A. Alford and K. R. Zenger, Methods for normalizing microbiome data: An ecological perspective, *Methods Ecol. Evol.*, 2019, **10**(3), 389–400, DOI: 10.1111/2041-210X.13115.

37 P. J. McMurdie and S. Holmes, Phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data, *PLoS One*, 2013, **8**(4), 1–11, DOI: 10.1371/journal.pone.0061217.

38 H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*, Springer-Verlag, DOI: 10.1007/978-3-319-24277-4, https://ggplot2.tidyverse.org.

39 M. E. Brooks, K. K. Kristensen, M. R. Darrigo, P. Rubim, M. Uriarte, E. Bruna and B. M. Bolker, Statistical modeling of patterns in annual reproductive rates, *Ecology*, 2019, **20**(20), 1–7, DOI: 10.1002/ecy.2706/suppinfo.

40 F. Hartig, DHARMa: Residual Diagnostics for Hierarchical (Multi-Level/Mixed) Regression Models, *R Package Version 0.4.7*, 2022, https://github.com/florianhartig/dharma.

41 R. Length, emmeans: Estimated Marginal Means, aka Least-Squares Means, *R Package Version 1.11.0*, 2024, https://rvlenth.github.io/emmeans/, https://rvlenth.github.io/emmeans/.

42 L. Wang, S. Wang and W. Li, RSeQC: quality control of RNA-seq experiments, *Bioinformatics*, 2012, **28**(16), 2184–2185, DOI: 10.1093/bioinformatics/bts356.

43 *Agisoft Metashape Professional (64 Bit) Software v.1.7.1*, Agisoft LLC, St. Petersburg, Russia. https://www.agisoft.com/.

44 J. W. Rouse, R. H. Haas, J. A. Schell and D. W. Deering, Monitoring vegetative systems in the great plains with ERTS, *NASA Spec. Publ.*, 1974, **351**(1), 309–317.

45 C. Evangelides and A. Nobajas, Red-Edge Normalised Difference Vegetation Index (NDVI705) from Sentinel-2 imagery to assess post-fire regeneration, *Remote Sens. Appl.: Soc. Environ.*, 2020, **17**, 100283, DOI: 10.1016/j.rsase.2019.100283.

46 J. Xue and B. Su, Significant remote sensing vegetation indices: A review of developments and applications, *J. Sens.*, 2017, 1353691, DOI: 10.1155/2017/1353691.

47 A. A. Gitelson, J. Y. Kaufman and M. N. Merzlyak, Use of a green channel in remote sensing of global vegetation from EOS-MODIS, *Rem. Sens. Environ.*, 1996, **58**(3), 289–298, DOI: 10.1016/S0034-4257(96)00072-7.

48 A. R. Huete, A soil-adjusted vegetation index (SAVI), *Rem. Sens. Environ.*, 1988, **25**(3), 295–309, DOI: 10.1016/0034-4257(88)90106-X.

49 F. Baret and G. Guyot, Potentials and limits of vegetation indices for LAI and APAR assessment, *Rem. Sens. Environ.*, 1991, **35**(2–3), 161–173, DOI: 10.1016/0034-4257(91)90009-U.

50 J. Qi, A. Chehbouni, A. R. Huete, Y. H. Kerr and S. Sorooshian, A modified soil adjusted vegetation index, *Rem. Sens. Environ.*, 1994, **48**(2), 119–126, DOI: 10.1016/0034-4257(94)90134-1.

51 B. W. Yap and C. H. Sim, Comparisons of various types of normality tests, *J. Stat. Comput. Simulat.*, 2010, **81**(12), 2141–2155, DOI: 10.1080/00949655.2010.520163.

52 S. Aboulaghras, N. Sahib, S. Bakrim, T. Benali, S. Charfi, F. E. Guaouguaou, N. E. Omari, M. Gallo, D. Montesano, G. Zengin, K. Taghzouti and A. Bouyahya, Health Benefits and Pharmacological Aspects of Chrysoeriol, *Pharmaceuticals*, 2022, **15**(8), 973, DOI: 10.3390/ph15080973.

53 J. J. Perry, L. Brown, E. Jurneczko, E. Ludkin and B. W. Singer, Identifying the plant origin of artists' yellow lake pigments by electrospray mass spectrometry,

Archaeometry, 2010, **53**(1), 164–177, DOI: **10.1111/j.1475-4754.2010.00530.x**.

54 O. Mazimba, Umbelliferone: Sources, chemistry and bioactivities review, *Bull. Fac. Pharm. Cairo Univ.*, 2017, **55**(2), 223–232, DOI: **10.1016/j.bfopcu.2017.05.001**.

55 D. Geiger, Plant glucose transporter structure and function, *Pfluegers Arch.*, 2020, **472**(9), 1111–1128, DOI: **10.1007/s00424-020-02449-3**.

56 T.-S. Wu, H. Furukawa, C.-S. Kuoh and K.-S. Hsu, Acridone alkaloids. Part 9. Chemical consistuents of Glycosmis citrifolia (wild) Lindl. Structures of novel linear pyranoacridones, furacridones, and other new acridone alkaloids, *J. Chem. Soc., Perkin Trans. 1*, 1983, 1681–1688, DOI: **10.1039/P19830001681**.

57 G. C. H. Derksen, H. A. G. Niederländer and T. A. van Beek, Analysis of anthraquinones in Rubia tinctorum L. by liquid chromatography coupled with diode-array UV and mass spectrometric detection, *J. Chromatogr., A*, 2002, **978**(1–2), 119–127, DOI: **10.1016/S0021-9673(02)01412-7**.

58 P. Khatri, D. Desai, N. Shelke and T. Minko, Role of plasticizer in membrane coated extended release oral drug delivery system, *J. Drug Delivery Sci. Technol.*, 2018, **44**, 231–243, DOI: **10.1016/j.jddst.2017.12.020**.

59 Dibutyl sebacate, 2025, **https://pubchem.ncbi.nlm.nih.gov/compound/7986**, last accessed June 19, 2025.

60 Diethyl phthalate, 2025, **https://pubchem.ncbi.nlm.nih.gov/compound/Diethyl-Phthalate**, (last accessed June 19, 2025).

61 V. Martirosyan, I. Stavi and T. Doniger, Fungal community dynamics in a hyper-arid ecosystem after 7 and 47 years of petroleum contamination, *Environ. Monit. Assess.*, 2024, **196**(1212), 1–13, DOI: **10.1007/s10661-024-13387-3**.

62 B. Yang, Z. Yang, K. He, W. Zhou and W. Feng, Soil Fungal Community Diversity, Co-Occurrence Networks, and Assembly Processes under Diverse Forest Ecosystems, *Microorganisms*, 2024, **12**(9), 1–22, DOI: **10.3390/microorganisms12091915**.

63 D. He, Z. Guo and W. Shen, Fungal Communities Are More Sensitive to the Simulated Environmental Changes than Bacterial Communities in a Subtropical Forest: The Single and Interactive Effects of Nitrogen Addition and Precipitation Seasonality Change, *Microb. Ecol.*, 2023, **86**, 521–535, DOI: **10.1007/s00248-022-02092-8**.

64 D. J. Van Horn, J. G. Okie, H. N. Buelow, M. N. Gooseff, J. E. Barrett and C. D. Takacs-Vesbach, Soil microbial responses to increased moisture and organic resources along a salinity gradient in a polar desert, *Appl. Environ. Microbiol.*, 2014, **80**(10), 3034–3043, DOI: **10.1128/AEM.03414-13**.

65 S. B. D. Diatta, L. N. Tall, Y. B. Ndour, M. Sembene and K. Assigbetsé, Composition and Diversity of Soil Bacterial Communities along an Environmental Gradient in the Sudano-Sahelian Region of Senegal, *Open J. Soil Sci.*, 2020, **10**(2), 58–89, DOI: **10.4236/ojss.2020.102004**.

66 H. Luo, C. Wang, K. Zhang, L. Ming, H. Chu and H. Wang, Elevational changes in soil properties shaping fungal community assemblages in terrestrial forest, *Sci. Total Environ.*, 2023, **900**, 1–12, DOI: **10.1016/j.scitotenv.2023.165840**.

67 X. Bai, E. Zhang, J. Wu, D. Ma, C. Zhang, B. Zhang, Y. Liu, Z. Zhang, F. Tian, H. Zhao and B. Wang, Soil fungal community is more sensitive than bacterial community to modified materials application in saline–alkali land of Hetao Plain, *Front. Microbiol.*, 2024, **15**, 1–14, DOI: **10.3389/fmicb.2024.1255536**.

68 J. T. T. Angelina and P. Nivedha, Beneficial role of actinomycetes in soil fertility and agriculture, *Int. J. Agro. Nutr. Pract.*, 2021, **1**(1), 13–16.

69 M. S. M. Selim, S. A. Abdelhamid and S. S. Mohamed, Secondary metabolites and biodiversity of actinomycetes, *J. Genet. Eng. Biotechnol.*, 2021, **19**(72), 1–13, DOI: **10.1186/s43141-021-00156-9**.

70 A. C. Rappazzo, A. Marchetta, C. Rizzo, M. Azzaro, W. R. L. Cairns, A. Lo Giudice and M. Papale, Enrichment, Isolation and Characterization of Heavy Metal-Tolerant Bacteria from Polar Lacustrine Sediments, *Microorganisms*, 2025, **13**(2), 1–13, DOI: **10.3390/microorganisms13020389**.

71 M. E. Trujillo, R. Riesco, P. Benito and L. Carro, Endophytic Actinobacteria and the Interaction of Micromonospora and Nitrogen Fixing Plants, *Front. Microbiol.*, 2015, **6**(1341), 1–15, DOI: **10.3389/fmicb.2015.01341**.

72 S. Khan, S. Srivastava, A. Karnwal and T. Malik, *Streptomyces* as a promising biological control agents for plant pathogens, *Front. Microbiol.*, 2023, **14**(1285543), 1–22, DOI: **10.3389/fmicb.2023.1285543**.

73 Y. Ma, J. Wang, Y. Liu, X. Wang, B. Zhang, W. Zhang, T. Chen, G. Liu, L. Xue and X. Cui, *Nocardioides*: "Specialists" for Hard-to-Degrade Pollutants in the Environment, *Molecules*, 2023, **28**(21), 7433, DOI: **10.3390/molecules28217433**.

74 M. Oubohssaine, L. Sbabou and J. Aurag, Potential of the plant growth-promoting rhizobacterium *Rhodococcus qingshengii* LMR356 in mitigating lead stress impact on *Sulla spinosissima* L, *Environ. Sci. Pollut. Res.*, 2024, **31**, 46002–46022, DOI: **10.1007/s11356-024-34150-8**.

75 R. Hayat, S. Ali, U. Amara, R. Khalid and I. Ahmed, Soil beneficial bacteria and their role in plant growth promotion: a review, *Ann. Microbiol.*, 2010, **60**, 579–598, DOI: **10.1007/s13213-010-0117-1**.

76 S. L. Ishaq, Plant-microbial interactions in agriculture and the use of farming systems to improve diversity and productivity, *AIMS Microbiol.*, 2017, **3**(2), 335–353, DOI: **10.3934/microbiol.2017.2.335**.

77 Q. Zhang, J. Ma, Y. Yang, Z. Luo, Y. Wang and F. Chen, Mining Subsidence-Induced Microtopographic Effects Alter the Interaction of Soil Bacteria in the Sandy Pasture, China, *Front. Environ. Sci.*, 2021, **9**(656708), 1–17, DOI: **10.3389/fenvs.2021.656708**.

78 N. S. Akimbekov, I. Digel, K. T. Tastambek, A. K. Marat, M. A. Turaliyeva and G. K. Kaiyrmanova, Biotechnology of Microorganisms from Coal Environments: From Environmental Remediation to Energy Production, *Biology*, 2022, **11**(9), 1306, DOI: **10.3390/biology11091306**.

79 C. L. Murphy, R. Yang, T. Decker, C. Cavalliere, V. Andreev, N. Bircher, J. Cornell, R. Dohmen, C. J. Pratt, A. Grinnell, J. Higgs, C. Jett, E. Gillette, R. Khadka, S. Mares, C. Meili, J. Liu, H. Mukhtar, M. S. Elshahed and N. H. Youssef, Genomes of novel Myxococcota reveal severely curtailed machineries for predation and cellular differentiation, *Appl. Environ. Microbiol.*, 2021, **87**, 1–20, DOI: **10.1128/AEM.01706-21**.

80 D. Zou, C. Zhang, Y. Liu and M. Li, Biogeographical distribution and community assembly of Myxococcota in mangrove sediments, *Environ. Microbiome*, 2024, **19**(47), 1–15, DOI: **10.1186/s40793-024-00593-2**.

81 I. Mujakic, K. Piwosz and M. Koblížek, Phylum gemmatimonadota and its role in the environment, *Microorganisms*, 2022, **10**(151), 1–17, DOI: **10.3390/microorganisms10010151**.

82 B. Baliyarsingh, B. Dash, S. Nayak and S. K. Nayak, Soil Verrucomicrobia and Their Role in Sustainable Agriculture, in *Advances in Agricultural and Industrial Microbiology*, ed. S. K. Nayak, B. Baliyarsingh, I. Mannazzu, A. Singh and B. B. Mishra, Springer, Singapore, 2022, pp. 105–124, DOI: **10.1007/978-981-16-8918-5_6**.

83 J. Bakr, A. Kompała-Bąba, W. Bierza, D. Chmura, A. Hutniczak, A. Błońska, T. Nowak, F. Magurno, A. M. Jagodziński and G. Woźniak, Taxonomic and functional diversity along successional stages on post-coalmine spoil heaps, *Front. Environ. Sci.*, 2024, **12**, 1–13, DOI: **10.3389/fenvs.2024.1412631**.

84 M. Sharma and S. Kulshrestha, Colletotrichum gloeosporioides: An anthracnose causing pathogen of fruits and vegetables, *Biosci. Biotechnol. Res. Asia*, 2015, **12**(2), 1–14, DOI: **10.13005/bbra/1776**.

85 Z. Guo, C. X. Luo, H. J. Wu, B. Peng, B. S. Kang, L. M. Liu, M. Zhang and Q. S. Gu, *Colletotrichum* Species Associated with Anthracnose Disease of Watermelon (*Citrullus lanatus*) in China, *J. Fungi*, 2022, **8**(8), 790, DOI: **10.3390/jof8080790**.

86 P. P. Than, H. Prihastuti, S. Phoulivong, P. W. Taylor and K. D. Hyde, Chilli anthracnose disease caused by Colletotrichum species, *J. Zhejiang Univ.–Sci. B*, 2008, **9**(10), 764–778, DOI: **10.1631/jzus.B0860007**.

87 S. Camargo, E. J. Mulinari, L. R. de Almeida, A. Bernardes, R. A. Prade, W. Garcia, F. Segato and J. R. C. Muniz, Functional and structural characterization of an α-L-arabinofuranosidase from Thermothielavioides terrestris and its exquisite domain-swapped β-propeller fold crystal packing, *Biochim. Biophys. Acta, Proteins Proteomics*, 2020, **1868**(12), 140533, DOI: **10.1016/j.bbapap.2020.140533**.

88 R. M. Berka, I. V. Grigoriev, R. Otillar, A. Salamov, J. Grimwood, I. Reid, N. Ishmael, T. John, C. Darmond, M.-C. Moisan, B. Henrissat, P. M. Coutinho, V. Lombard, D. O. Natvig, E. Lindquist, J. Schmutz, S. Lucas, P. Harris, J. Powlowski, A. Bellemare, *et al.*, Comparative genomic analysis of the thermophilic biomass-degrading fungi *Myceliophthora thermophila* and *Thielavia terrestris*, *Nat. Biotech.*, 2011, **29**, 922–927.

89 J. R. Dean, S. Ahmed, W. Cheung, I. Salaudeen, M. Reynolds, S. Bowerbank, C. E. Nicholson and J. J. Perry, Use of remote sensing to assess vegetative stress as a proxy for soil contamination, *Environ. Sci.: Processes Impacts*, 2024, **26**, 16117–6, DOI: **10.1039/D3EM00480E**.

90 L. Salgado, C. A. López-Sánchez, A. Colina, D. Baragaño, R. Forján and J. R. Gallego, Hg and As pollution in the soil-plant system evaluated by combining multispectral UAV-RS, geochemical survey and machine learning, *Environ. Pollut.*, 2023, **333**, 122066.

91 Y. Sun, S. Chen, H. Jiang, B. Qin, D. Li, K. Jia and C. Wang, Towards interpretable machine learning for observational quantification of soil heavy metal concentrations under environmental constraints, *Sci. Total Environ.*, 2024, **926**, 171931.

92 J. Entwistle, L. Bramwell, J. Wragg, M. Cave, E. Hamilton, A. Gardner and J. R. Dean, Investigating the geochemical controls on Pb bioaccessibility in urban agricultural soils to inform sustainable site management, *Geosciences*, 2020, **10**, 398, DOI: **10.3390/geosciences10100398**.

93 DEFRA, Department for Environment Food and Rural Affairs, Development of category 4 screening levels for assessment of land affected by contamination – SP1010, *Annex SP1010 Appendix H Lead*, 2013, **https://randd.defra.gov.uk/ProjectDetails?ProjectID=18341**, (last accessed June 19, 2025).

94 UKSO, *UK Soil Observatory*, 2016, **https://www.ukso.org**, (last accessed June 19, 2025).

95 S. Ahmed, C. E. Nicholson, S. R. Rutter, J. R. Marshall, J. J. Perry and J. R. Dean, Winter beans: the use of an unmanned aerial vehicle for monitoring and prediction of crop performance, *Open Sci. J.*, 2023, **8**(2), 1–16, **https://osjournal.org/ojs/index.php/OSJ/article/view/3336**.

96 S. Ahmed, C. E. Nicholson, S. R. Rutter, J. R. Marshall, J. J. Perry and J. R. Dean, Use of an unmanned aerial vehicle for monitoring and prediction of oilseed rape crop performance, *PLoS One*, 2023, **18**(11), e0294184, DOI: **10.1371/journal.pone.0294184**.

97 S. Ahmed, N. Basu, C. E. Nicholson, S. R. Rutter, J. R. Marshall, J. J. Perry and J. R. Dean, Use of machine learning for monitoring the growth stages of an agricultural crop, *Sustainable Food Technol.*, 2024, **2**, 104–125, DOI: **10.1039/D3FB00101F**.