Use of untargeted metabolomics to analyse changes in extractable soil organic matter in response to long-term fertilisation

Tang, Sheng; Ma, Qingxu; Zhou, Jingjie; Pan, Wankun; Chadwick, David R.; Gregory, Andrew S.; Wu, Lianghuan; Jones, Davey L.

Biology and Fertility of Soils

DOI:
10.1007/s00374-023-01706-8

Published: 01/04/2023

Peer reviewed version

Use of untargeted metabolomics to analyse changes in extractable soil organic matter in response to long-term fertilisation

Sheng Tang[a,b], Qingxu Ma[a,b*], Jingjie Zhou[a], Wankun Pan[a,b], David R. Chadwick[b], Andrew S. Gregory[c], Lianghuan Wu[a], Davey L. Jones[b,d]

[a]*Ministry of Education Key Lab of Environmental Remediation and Ecosystem Health, College of Environmental and Resource Sciences, Zhejiang University, Hangzhou, 310058, China*

[b]*School of Natural Sciences, Bangor University, Gwynedd, LL57 2UW, UK*

[c]*Protecting Crops and the Environment, Rothamsted Research, Harpenden, Herts, AL5 2JQ, UK*

[d]*SoilsWest, Centre for Sustainable Farming Systems, Food Futures Institute, Murdoch University, Murdoch, WA 6105, Australia*

Email addresses:

Qingxu Ma: qxma@zju.edu.cn; Sheng Tang: tangsheng@zju.edu.cn; Wankun Pan: panwankun2017@163.com; Jingjie Zhou: 11814012@zju.edu.cn; David R. Chadwick: d.chadwick@bangor.ac.uk; Lianghuan Wu: finm@zju.edu.cn; Andrew S. Gregory: andy.gregory@rothamsted.ac.uk; Davey L. Jones: d.jones@bangor.ac.uk

*__Corresponding author__: Qingxu Ma

[contact details] qxma@zju.edu.cn

Abstract

This study aimed to explore the soil metabolic response to long-term fertiliser application and the effect of this response on the microbial community by taking advantage of the Woburn Organic Manuring Experiment (UK; operational since 1964). Untargeted metabolomes detected by gas chromatography-time of flight mass spectrometer/mass spectrometry (GC-TOFMS/MS) and ultra-high-pressure liquid chromatography-quadrupole time of flight mass-spectrometer/mass spectrometry (UHPLC-QTOFMS/MS) were used to explore which method better reflected soil microbe-accessible metabolites. Microbial community abundance was detected by high-throughput sequencing. We found that long-term farmyard manure application enhanced the soil's total and dissolved C and N contents. The metabolite content detected by GC-TOFMS/MS (TOF detector with a cold injection unit) had a negative linear correlation with soil organic matter, extractable organic nitrogen (N), and microbial carbon (C). Conversely, the metabolite content detected by UHPLC-QTOFMS/MS was positively correlated with soil organic matter, indicating that metabolites detected by UHPLC-QTOFMS/MS were the main components of soluble soil organic matter. More positive than negative correlations were observed between metabolites and bacterial (69.5%) and fungal (67.9%) taxa in the co-occurrence network. Among the bacterial taxa in the network, the family Planococcaceae and genus *Paenibacillus* showed the most correlations with metabolites. The choice of extraction and detection method affects the identity and number of metabolites

detected. Therefore, careful consideration is needed when selecting which methods to use. We demonstrated a strong correlation between soil metabolites and microbial community abundance. However, a deeper understanding of soil microbial function and metabolite formation, content, and decomposition is still needed.

Keywords: soil organic matter, dissolved organic matter, chemical fertiliser, farmyard manure, untargeted metabolomes

Introduction

Most of the C in the terrestrial biosphere is retained as soil organic matter (SOM), which originates from microbes, plants, and animals (Johnston et al. 2004). Soil microorganisms derive metabolites predominantly from SOM and its biomass turnover (Brown et al. 2021; Liang et al. 2019). Dissolved organic matter (DOM) is the most biologically-accessible component of SOM, playing a crucial role in C, N, and sulphur (S) cycling (Ma et al. 2020b, 2021a; Swenson et al. 2015). It contains a series of organic matter compounds such as carbohydrates, amino acids, hydroxyl acids, sugar acids, nucleosides, sterols, aromatics, amines, and miscellaneous compounds (Brown et al. 2021; Ma et al. 2022); and is in a constant state of flux driven by the microbial community and *in situ* metabolic activities (Mcleod et al. 2021; Schmidt et al. 2011). Therefore, understanding the composition and turnover of soil microbe-accessible substrates is crucial for exploring the complex dynamics of microbial communities and their nutrient cycling (Ma et al. 2020c; Zhu et al. 2022).

Fertiliser is an important field management intervention that strongly affects soil element content, nutrient cycling, and microbial community composition and function. Globally, agricultural production produces approximately seven billion tons of farmyard manure (FYM) each year (Thangarajan et al. 2013). Manure application to arable land can increase soil structural stability and nutrient levels, thereby enhancing soil C sequestration and biological activity (Maillard and Angers 2014). Partly substituting inorganic fertiliser with FYM can sustain agricultural productivity and

reduce environmental pollution (Hoyle and Fang 2018). FYM application strongly stimulates belowground biogeochemical processes: directly by adding large amounts of organic C and nutrients and indirectly by modifying biotic activity (Ma et al. 2018; Liu et al. 2020). Subsoil differs from topsoil in nutrient content, microbial biomass, community composition, bioavailability, age, and accessibility of soil C, which affect the rates of SOM decomposition (Cheng et al. 2017). In contrast to chemical fertilisers, which mainly affect only the topsoil, long-term FYM application generally improves the total and DOM content of both the topsoil and subsoil (Ma et al. 2020b; Yan et al. 2018). Additionally, it enhances the activities of enzymes such as β-glucosidase, protease, urease, and cellulase (Chang et al. 2010; Ma et al. 2020b). However, how the combined application of FYM and chemical fertiliser influences the soil metabolite composition is unclear.

A healthy and well-functioning soil system is vital for providing ecosystem services, especially food production in agricultural ecosystems (Liu et al. 2022; Wei et al. 2021). Metabolites in DOM are intermediates or products of enzymatic reactions, including organic acids, sugars, amino acids, and fatty acids. These are involved in microbial function, growth, and development. In addition to molecular methods of soil biological quality assessment, extracting and quantifying primary metabolites offer an alternative approach to better understanding belowground functions. The metabolic approach has been used extensively in plant biology (Hartman et al. 2020), biomedical science (Gupta et al. 2018), and research on the biochemical responses of microbes (Jozefczuk et al. 2014). However, its application in soil is limited, especially

under field conditions, and most studies have only focused on specific metabolites (Ma et al. 2021a; Warren, 2020). Recent studies have shown that the soil metabolome is sensitive and can reflect the functional responses of soil microbe communities to changes in their environment, such as fertiliser application, extreme drought, and dry-wet or freeze-thaw events (Brown et al. 2021; Miura et al. 2020).

Traditionally, soil DOM is quantified by extraction from soil samples using specific solutions (water, KCl, K_2SO_4, etc.) and subsequent analysis of its elemental composition using combustion or oxidization. However, the molecular composition cannot be detected using this method (Jones and Willet 2006). Untargeted metabolomics is rapidly gaining attention, but its results are highly dependent on the extraction method and detection instrument used. Gas chromatography/mass spectrometry (GC/MS) and liquid chromatography/mass spectrometry (LC/MS) are the most widely-used methods due to their broad analytical scope (alcohols, fatty acids, sterols, carbohydrates, amino acids, etc.), and availability of the spectral databases of various metabolites (Brailsford et al. 2019; Brown et al. 2021; Liu et al. 2021; Swenson et al. 2015). Other available methods include capillary electrophoresis/mass spectrometry (CE/MS) (Warren 2020) and Fourier transform ion cyclotron resonance/mass spectrometry (FTICR/MS) (Hirai et al. 2004), which are not extensively used. The compounds detected vary with the detection method used, and the method that most accurately reflects soil microbe-accessible metabolites is still unknown.

Microorganisms are the most sensitive soil quality indicators and respond

quickly to changes in soil DOM under chemical and organic fertiliser application (Ma et al. 2020b). A shift in microbial community composition indicates a change in the metabolism and function of the community in a soil ecosystem (McGuire and Treseder, 2010). Moreover, the microbial community strongly drives organic C and N utilisation and mineralisation (Ma et al. 2018). Nutrient (C, N and P) enrichment induces significant changes in the soil metabolite profile, as it changes microbial activity and its metabolic processes. A recent study based on UHPLC-MS/MS found that inorganic nutrient enrichment causes substantial shifts in both primary and secondary metabolism and changes in resource flow and soil functioning, and that the microbial community composition showed significant metabolic flexibility (Brown et al. 2022). C and N (together addition) generally increased peptide synthesis in soil, C and P addition increased the fatty acids synthesis, while glucose-C addition increased the synthesis of other carbohydrates (Brown et al. 2022). The systematic coupling of the microbial community and soil metabolomics can valuably improve our understanding of microbial strategies in response to environmental stress (Swenson et al. 2018). However, this presents a challenge given the large number of metabolites and complexity of the microbial community.

Therefore, in this field-based study, we aimed to improve our understanding of soil metabolic processes by exploring the response of soil metabolites to long-term fertiliser application, and to explore whether the metabolites extracted and detected by different methods can reflects soil organic compounds composition. We hypothesised that (1) the total DOM detected by traditional methods, GC-TOFMS/MS, and

UHPLC-QTOFMS/MS should be positively related to each other; (2) soil metabolomics and the microbial community are systematically coupled.

Materials and methods

Experimental site and treatments

Soil samples were collected in June 2018 from the long-term Woburn Organic Manuring experiment running since 1964 in Southeastern England (www.era.Rothamsted.ac.uk/WoburnFarm) to test the effects of organic manures and chemical fertilisers on soil fertility and crop production. The soil is derived from Lower Greensand parent material and is classified as a sandy loam-textured brown sand (10% clay, 6% silt, and 80% sand, excluding organic matter content). The soil samples were collected from three typical treatments that reflected current agronomic regimes: FYM applied at 25–50 t ha^{-1} y^{-1} for 28 y (high manure application, High-M), FYM applied at 10 t ha^{-1} y^{-1} for 16 y supplemented with chemical fertilisers (low manure application, Low-M), and chemical fertilisers only (No-M), with P and K inputs equivalent to 25–50 t ha^{-1} y^{-1} FYM. Each treatment consisted of four replicates. Each plot was 8.83 × 8.00 m with a 5-year arable rotation (since 2003 this has been spring barley and mustard, winter beans, winter wheat, forage maize, and mustard, and winter rye).

The treatment plots received chemical fertilisers or organic manures for three periods between 1964 and 2018. In the High-M treatment, FYM was applied from

1966–71, 1981–87, and 2003–18 (28 y in total). FYM was applied at 50 t ha^{-1} in the first two build-up periods and 25 t ha^{-1} in the final period. In the Low-M treatment, FYM was applied at 10 t ha^{-1} from 2003 onward (16 y in total). Before this, it received chemical fertilisers (P & K) equivalent to 7.5 t ha^{-1} straw input, containing approximately 30.8 kg N ha^{-1}. The No-M treatment received chemical fertilisers as N, P, and K at rates equivalent to High-M during the same years. Since 2003, the Low-M and No-M treatments received annual N (nitrochalk), P (triple superphosphate), and K and S (potassium sulphate) fertilisers at 165, 20, 83, and 36 kg ha^{-1}, respectively (equivalent annual rate for a 5-year crop rotation). All other aspects of agronomic management, including harvesting, tillage regime, herbicides and fungicides were consistent among the three treatments. Herbicides including spring-applied Atlantis (mesosulfuron-methyl + iodosulfuron-methyl-sodium, 3:0.6% w/w, Bayer CropScience Ltd, Cambridge, UK) at 400 mL ha^{-1}, Hiatus (thifensulfuron-methyl + tribenuron-methyl, 40:15% w/w, Rotam Global AgroSciences Ltd, Hong Kong) at 50 g ha^{-1}, and Sprinter (2,4-D as the dimethylamine and the monomethylamine salts, 700g L^{-1}, Nufarm Ltd, Otahuhu, Auckland, New Zealand) at 2 L ha^{-1}; fungicides including spring-applied Keystone (isopyrazam + epoxiconazole, 11.6:9.2% w/w, Agrichem, Yatala, Queensland, Australia) at 500 mL ha^{-1}, Folicur (tebuconazole, 25.9% w/w, Bayer CropScience Ltd, Cambridge, UK) at 800 mL ha^{-1}, and Cello (prothioconazole + tebuconazole + spiroxamine, 10.3:10.5:26.3% w/w, Bayer CropScience Ltd, Cambridge, UK) at 630 mL ha^{-1}. The total N, P, and S inputs during the build-up phase (1964–2018) under No-M were 2.46, 1.77, and 0.96 t, respectively.

The total C, N, P, and S inputs under High-M were 112.50, 5.80, 1.26, and 1.22 t, respectively, while that under the Low-M treatment was 14.10, 2.63, 1.69, and 1.00 t. Further details of the agronomic regime and experiment can be found in Ma et al. (2020b).

Winter rye (*Secale cereale* L.) was sown in the plots, and sampling was performed at the grain-filling stage in 2018. From each of four plots per treatment, the topsoil (0–23 cm plough layer) and subsoil (23–38 cm) samples were collected using a 2.5 cm diameter corer (18 cores per plot to make up one replicate). The soil was thoroughly mixed by hand and passed through a 5 mm sieve to remove roots, stones, and earthworms. The soil samples were then portioned into three parts: the first was stored at –80 °C to analyse soil metabolites and microbial community, the second was stored at 4 °C to assess soil microbial biomass, and the third was air-dried to determine basic soil properties.

Determination of soil properties

Basic soil properties were determined using traditional methods. Soil pH was determined at a 1:2.5 (v/v) soil: H_2O ratio. Total C and N were measured by dry combustion of finely milled soil using a CHN-2000 Analyser (Leco Co., St. Joseph, MI, USA). To determine the K_2SO_4 extractable C and N (total, organic, NO_3^-, and NH_4^+), 5 g of moist soil was extracted with 25 mL of 0.5 M K_2SO_4 for 30 min at 200 rpm, and centrifuged for 10 min at 12 000 × g at 25 °C. The dissolved organic C (DOC) and total dissolved N (TDN) in the extracts were detected using a multi N/C

2100S TOC-TN Analyser (Analytic Jena AG, Jena, Germany). The NO_3^- and NH_4^+ content in the extracts were detected colourimetrically using a microplate spectrophotometer (BioTek Instruments Inc., Winooski, VT, USA). Extractable organic N was calculated by subtracting the NO_3^- and NH_4^+ content from TDN. Soil microbial biomass C (MB-C) and N (MB-N) were determined using the $CHCl_3$ fumigation-extraction method (Vance et al., 1987). Organic C and N were extracted and detected from the fumigated soil in the same manner as from non-fumigated soil. MB-C and MB-N were calculated by a conversion factor of 2.22 for both C and N (Vong et al. 2003). The total soluble protein in the extracts was estimated by the acid hydrolysis of proteins in solution, and amino acids were subsequently determined, as described by Roberts and Jones (2008), and have been reported previously (Ma et al. 2020b). The 0.5 M K_2SO_4 extracts were passed through a 1 000 MW ultrafiltration membrane using an Amicon® stirred cell (Merck-Millipore, Billerica, MA, USA). To quantify the fraction of peptides and free amino acids. Amino acids in the flow-through were detected using the fluorometric OPAME method before and after acid hydrolysis with 6 M HCl (105 °C, 16 h) under N_2 (Jones et al. 2002).

Untargeted metabolomics detected by GC-TOFMS/MS

The soil samples stored at –80 °C were freeze-dried using an Edwards Super Modulyo freeze-drier (SciQuip Ltd., Shropshire, UK) for 3 d. The dried soil was ground using a ball mill (Retsch MM200, GmbH, Haan, Germany) to promote metabolite recovery from the microbial biomass (Wang et al. 2015). The samples were extracted by 3:3:2

(v/v/v) acetonitrile-isopropanol-water (Brailsford et al. 2019; Brown et al. 2021), as this extraction method can extract a broad range of metabolites. The untargeted metabolome was analysed at the UC Davis West Coast Metabolomics Facility using an automated linear exchange-cold injection system (ALEX-CIS) GC time of flight (TOF) MS (Brailsford et al. 2019; Brown et al. 2021). Briefly, 0.5 μL of the extracted solution was injected into an Rtx-5Sil MS capillary column (0.25 μm 95% dimethylsiloxane/5% diphenylpolysiloxane coating; 30 m length × 0.25 mm i.d.; Restek Corp., Bellefonte, PA, USA). This chromatography method yields excellent retention and separation of primary metabolite classes (amino acids, hydroxyl acids, carbohydrates, sugar acids, sterols, aromatics, nucleosides, amines, and miscellaneous compounds) with narrow peak widths of 2–3 s and very good within-series retention time reproducibility of better than 0.2 s absolute deviation of retention times. The GC thermal program was run at 50 °C for 1 min, then increased to 330 °C at 20 °C min^{-1}, and finally maintained at 330 °C for 5 min, with a He mobile phase. Upon elution, samples were injected into a Pegasus IV GC-TOF-MS (Leco Corp., St Joseph, MI, USA), using a mass resolution of 17 spectra s^{-1}, from 80–500 Da, at –70 eV ionisation energy and 1800 V detector voltage, with a 230 °C transfer line and 250 °C ion source (Withers et al. 2020). A mixture of internal retention index markers was prepared using fatty acidmethyl esters of C8, C9, C10, C12, C14, C16, C18, C20, C22, C24, C26, C28, and C30 linear chain length, dissolved in chloroform at concentrations of 0.8 mg mL^{-1} (C8–C16) or 0.4 mg mL^{-1} (C18–C30) as detailed in Fiehn et al. (2008). The raw data files were pre-processed directly after data acquisition and stored as

ChromaTOF-specific *.peg files. ChromaTOF v. 2.32 (Leco Corp.) was used for data pre-processing without smoothing, with a 3 s peak width, baseline subtraction just above the noise level, and automatic mass spectral deconvolution and peak detection at signal/noise levels of 5:1 throughout the chromatogram. Apex masses were reported for use in the BinBase algorithm. The results were exported to a data server with absolute spectra intensities and further processed by a filtering algorithm implemented in the metabolomics BinBase database, as shown in Withers et al. (2020). Both known and unknown compounds were analysed using MetaboAnalyst v4.0 (Chong et al., 2018; Xia and Wishart 2016). Prior to analysis, the data were $\log_{10}$ transformed and scaled by Pareto scaling (Chong et al. 2018).

Untargeted metabolomics detected by UHPLC-QTOFMS/MS

Complex lipid extraction was conducted using a modified bi-phasic method (Matyash et al. 2008), which is advantageous as the lipids are retained in the upper extraction phase, and the methyl tertiary-butyl ether (MTBE) solvent has a density lower than that of water. Compared to chloroform ($CHCl_3$), MTBE can be detected directly without the risk of contamination from the interphase or aqueous phase. Briefly, 225 µL of MeOH with internal standards was added to a 40 mg freeze-dried and ground soil sample and vortexed for 20 s; 750 µL MTBE was subsequently added and vortexed for 10 min. Samples were placed in a bead grinder for 30 s and then shaken for 6 min at 4 °C; 188 µL of MS-grade water was added, and the sample was centrifuged for 2 min at 14 000 × *g* at 4 °C. The upper phase was transferred to two

tubes (350 μL/tube), and one tube was evaporated to dryness using a SpeedVac. Dried extracts were re-suspended with a mixture of 1:9 toluene: MeOH (v/v) and an internal standard. The samples were analysed using an Agilent 1290 Infinity liquid chromatography (LC) system (G4220A binary pump, G4226A autosampler, and G1316C Column Thermostat) coupled to an Agilent 6530 MS (positive ion mode). Lipids were separated on an Acquity ultra high-pressure chromatography (UHPLC) CSH C18 column (1.7 μm; 100 × 2.1 mm) (Brown et al. 2021). The data were processed by the mass spectrometry-data independent analysis (MS-DIAL) software (Tsugawa et al. 2015), followed by data clean-up using the mass spectral feature list optimiser (MS-FLO) (Defelice et al. 2017). Peaks were annotated, and the MassHunter Quant software was applied to verify peak candidates (Brown et al. 2021). Valid and reproducible peaks were analysed using targeted MS/MS to increase overall peak annotations. In addition, nine internal standards were used to convert peak heights into good estimates of absolute (micromolar) concentrations for a range of biogenic amines typically detected in biofluids and tissues (shown in supporting materials). Notably, internal standards were included, but only for peak correction and quality control. Therefore, the data presented are qualitative, and the compounds were tentatively identified in line with typical untargeted analyses (Brown et al. 2021). This UHPLC-TOFMS/MS method reportedly yields an excellent retention and separation of acylcarnitines, trimethylamine oxide, cholines, betaines, S-adenosine methionine, S-adenosine-L-homocysteine, nucleotides and nucleosides, methylated and acetylated amines, di- and oligopeptides, while also yielding excellent retention and separation

of metabolite classes with narrow peak widths of 5–20s (biogenic amines, cationic compounds). The internal standards were D3-Creatinine (392 ng mL^{-1}), D9-Choline (50 ng mL^{-1}), D9-TMAO (49 ng mL^{-1}), D3-1-Methylnicotinamide (130 ng mL^{-1}), Valine-Tyrosine-Valine (146 ng mL^{-1}), D9-Betaine (151 ng mL^{-1}), D3-AC(2:0) (33 ng mL^{-1}), D3-Histamine, N-methylproline (31 ng mL^{-1}), D3-L-Carnitine (158 ng mL^{-1}), D3-Creatine (171 ng mL^{-1}), D5-L-Glutamine (1941 ng mL^{-1}), D3-DL-Glutamic acid (2426 ng mL^{-1}), D3-DL-Aspartic acid (9901 ng mL^{-1}), D4-Cystine (721 ng mL^{-1}), D4-Alanine (2847 ng mL^{-1}), D7-Arginine (743 ng mL^{-1}), D3-Asparagine (720 ng mL^{-1}), D5-Histidine (990 ng mL^{-1}), D10-Isoleucine (885 ng mL^{-1}), D10-Leucine (1856 ng mL^{-1}), D8-Lysine (681 ng mL^{-1}), D8-Methionine (495 ng mL^{-1}), D2-Ornithine (632 ng mL^{-1}), D8-Phenylalanine (743 ng mL^{-1}), D7-Proline (1274 ng mL^{-1}), D3-Serine (2475 ng mL^{-1}), D5-Threonine (1406 ng mL^{-1}), D8-Tryptophan (619 ng mL^{-1}), D8-Valine (5569 ng mL^{-1}).

Soil DNA extraction and sequencing of bacteria and fungi

Following the manufacturer's protocols, DNA from soil subsamples (0.5 g) was extracted using a FastDNA SPIN kit (MP Biomedicals, Irvine, CA, USA). A NanoDrop ND-1000 UV-Vis spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA) was then used to identify the concentrations and quality of the extracted DNA. Primers 515F-806R (Brown et al. 2021) for bacteria and ITS1F-ITS2 (Gardes and Bruns 2010) for fungi were used for amplification. The polymerase chain reaction products were sequenced using the Illumina Novaseq platform. Bacterial and

fungal sequence data were processed using an in-house pipeline (Kai et al. 2017). Sequences with a length exceeding 200 bp were retained for downstream analyses. Operational taxonomic units (OTUs) were clustered at a 97% similarity. We annotated the taxonomic data for representative sequences of bacteria and fungi using the SILVA (Quast et al. 2012) and UNITE (Nilsson et al. 2019) databases, respectively. A total of 1 790 490 and 1 616 428 high-quality bacterial and fungal sequences were generated with an average read count of 74 604 (55 781–85 255) and 67 351 (43 939–81 715) per sample, respectively.

Data and statistical analysis

All statistical analyses were performed using R (version 3.4.3). The metabolomics data were $\log_{10}$ transformed. Agglomerative hierarchical clustering analyses were performed for the metabolite concentration data under fertiliser treatment and soil depth according to Pearson correlation coefficients. The dendrograms were combined with heat maps generated based on the *z*-scores of metabolite concentrations. Principal component analysis (PCA) was performed to determine the relationship between fertiliser treatment and C, N, and metabolites at two soil depths. One-way ANOVA and Tukey *post-hoc* testing were used to assess the differences among the fertiliser treatments, and the Shapiro-Wilk test was used to check for normality; the topsoils and subsoils were analysed separately ($p < 0.05$). A random forest analysis was performed using the 'randomForest' R package of the Linear discriminant analysis effect size (LEfSe) on the Galaxy platform. The interaction between

metabolite concentrations and the microbial community was visualised using the 'psych' package in R and Gephi (http://gephi.github.io/).

Results

Effect of long-term fertiliser on soil properties

In the collected sandy soil samples, manure application increased the total and dissolved contents of C (Total C, DOC) and N, which increased with the FYM application rate (Fig. S1). Generally, the total and dissolved C and N contents were greater in the topsoil than in the subsoil. The peptide and amino acid contents were clustered with DOC. In addition, MB-C and MB-N were clustered with total C and N, SOM, and protein content.

Effects of long-term fertiliser on primary metabolites detected by GC-TOFMS/MS

The untargeted primary metabolomics analysis using GC-TOFMS/MS tentatively identified 186 compounds, of which 71 were previously identified. Among the known compounds, the concentrations of 33 compounds differed significantly between treatments ($p < 0.05$) (supporting materials). In contrast, the dissolved SOM content extracted by 3:3:2 (v/v/v) acetonitrile-isopropanol-water was generally lower in the subsoil than in the topsoil. There were two distinct responses: the concentrations in the first group decreased with long-term Low-M and High-M treatments and showed higher concentrations in the topsoil compared to those in the subsoil (n = 12); the

second group had higher concentrations in the subsoil than those in the topsoil (n = 59). The 50 most significant known metabolites revealed by ANOVA are presented in Fig. 1.

Effects of long-term fertiliser on primary metabolites detected by UHPLC-QTOFMS/MS

The curated complex lipid analysis identified 2 944 individual compounds, of which 144 were known (supporting materials). Among these previously identified compounds, the 90 that appeared in the highest concentrations were clustered into three groups:

(1) Compounds that appeared at higher concentrations in the topsoil than the subsoil and at higher concentrations under the No-M than the Low-M and High-M treatments (n = 35).

(2) Compounds that appeared at higher concentrations in the topsoil than in the subsoil and at the highest concentrations under the highest manure application (n = 24).

(3) Compounds with higher concentrations in the subsoil than in the topsoil (n = 31). The 50 most significant known metabolites revealed by ANOVA are presented in Fig. 2.

PCA analysis of soil properties and soil metabolomics

We observed a significant difference between the properties of the topsoil and subsoil

of the Low-M treatment and a large difference between the No-M and High-M treatments. The PCA indicated that the No-M and High-M treatments significantly influenced the soil metabolomes detected by GC-TOFMS/MS and UHPLC-QTOFMS/MS (Fig. 3).

A linear relationship between dissolved organic matter and metabolites detected by GC-TOFMS/MS and UHPLC-QTOFMS/MS

The metabolite profiles detected by GC-TOFMS/MS and UHPLC-QTOFMS/MS were inversely correlated (Fig. S2). Therefore, while the metabolites detected by UHPLC-QTOFMS/MS were positively correlated to SOM, EON (extractable organic nitrogen), and MB-C, those detected by GC-TOFMS/MS were inversely correlated (Fig. 4). In addition, several compounds such as tyrosine, glucose-1-phosphate, leucine, glutamine, and isoleucine were detected by both GC-TOFMS/MS and UHPLC-QTOFMS/MS, but only isoleucine detected by GC-TOFMS/MS was positively linked with that detected by UHPLC-QTOFMS/MS.

Response of bacterial and fungal communities to long-term organic and inorganic fertiliser application

The LEfSe analysis identified the microbial taxa that differed significantly between fertiliser regimes (Fig. 5). The High-M treatment had the most enrichment indicators (that were significant), whereas the Low-M treatment had the least. Among the bacteria, indicators belonged mainly to Proteobacteria, Actinobacteria, Firmicutes, and Acidobacteria, the predominant bacterial phyla (Fig. 5A). Particularly in the Low-

M treatment, Nitrospirae, which are involved in soil nitrification, were enriched. In the High-M treatment, the identified indicators included *Bacillus* and Proteobacteria, Actinobacteria, and Firmicutes. Among the fungi, the most prominent indicators were Ascomycota, Mucoromycota, and Aphelidiomycota, the predominant fungal phyla (Fig. 5B). Long-term high-rate manure application (High-M) significantly increased the abundance of Ascomycota, whereas long-term chemical fertiliser application (No-M) significantly enriched Mucoromycota.

Metabolites drive microbial community succession

The random forest analysis revealed the relative importance of metabolites in determining microbial community succession. The 15 most important metabolites are presented in Fig. 6. The most important driver of both bacterial and fungal community succession was 5'-methylthioadenosine (MTA). After that, N-epsilon-acetyllysine, gamma-glutamylleucine, histidine, and 3-indolepropionic acid correlated the most with the bacterial community (Fig. 6A). Fungal community succession correlated most strongly with 2'-O-methyladenosine, 1,4-cyclohexanedione, isobutyryl-L-carnitine, and corticosterone after MTA (Fig. 6B).

We constructed a co-occurrence network based on the LEfSe and random forest analysis results to further clarify the correlation between the microbial taxa and specific metabolites (Fig. 7). The 15 most important metabolites for the two communities and the identified indicators were selected to construct the co-occurrence network. There were more positive than negative correlations between bacterial taxa

and metabolites (69.5%) and fungal taxa and metabolites (67.9%) in the network. Among the bacterial taxa in the network, the family Planococcaceae and genus *Paenibacillus* showed the most correlations (8) with metabolites (Fig. 7A and Table S1). In the case of metabolites, gamma-glutamylleucine had the most correlations (20) with bacterial taxa. The fungal network was simpler, with fewer nodes and total degrees (Fig. 7B and Table S2) than the bacterial network. *Aspergillus caesiellus* and *Thermomyces lanuginosus* had the most correlations (8) with metabolites among the fungal taxa in the network, and MTA had the most links with fungal taxa.

Discussion

Effect of long-term organic and inorganic fertiliser application on soil organic matter

As expected, long-term FYM increased the stock of soil total and DOM directly by adding large amounts of organic C and nutrients and indirectly by increasing the microbial biomass (Liu et al. 2020; Ma et al. 2018). Microorganisms can rapidly utilise organic C, and the microbial necromass contributes greatly to SOC (soil organic C) sequestration, especially in soils supplemented with manure and that have an enhanced microbial biomass (Cui et al. 2020; Ma et al. 2020a; Wang et al. 2021). Based on the evaluation of glucosamine and muramic acid from bacterial and fungal necromasses, Wang et al. (2021) found that microbial necromass contributed to approximately half of the soil organic C in grassland and cropland soils. Therefore, the increased microbial biomass after FYM application could stimulate the formation

of SOM.

Long-term high FYM application increased the EON content in the subsoil but not in the topsoil, which was in direct contrast to the effect of the chemical fertilisers. We ascribe this to the blockage of sorption sites by organic acids and humic substances released from the manure (Haynes and Mokolobate 2001), which increases soluble organic N leaching to the subsoil (similar to that of soil soluble organic P) (Ma et al. 2020a). The sandy soil we studied has a lower adsorption ability compared to soils with high clay content; therefore, leaching has a greater effect on dissolved SOM content.

Effect of long-term organic and inorganic fertiliser application on soil metabolites detected by GC-TOFMS/MS

Besides the basic chemical and physical soil characteristics, metabolic profiles especially sugars, amino acids, and organic acids are an important indicator of soil quality and ecosystem function (Withers et al. 2020). Metabolites can be sensitive to changes in the soil environment directly related to the physicochemical properties and microbial community. The metabolomics data detected by GC-TOFMS/MS was negatively linked to dissolved organic C and N contents. Also, the total metabolite content detected by GC-TOFMS/MS and UHPLC-QTOFMS/MS were negatively correlated. While GC-TOFMS/MS can detect numerous primary metabolites, it is generally limited by its poor resolving power for highly labile metabolites and several N-containing metabolites, such as coelute and other sugar compounds with the same

m/z (Brown et al. 2021). Additionally, the samples were only detected by MS in positive ion mode; therefore, compounds only detectable in the negative mode were missed. Furthermore, some compounds, such as glycine betaine, are not amenable to derivatisation and hence are undetectable (Brown et al. 2021). Therefore, in this study, the compounds detected by GC-TOFMS/MS were not exhaustive, and the metabolomics data detected by GC-TOFMS/MS was negatively correlated to EON. The extraction solution might also greatly affect the metabolites detected. Extractions using 3:3:2 (v/v/v) acetonitrile-isopropanol-water reportedly cover a broad range of metabolites, which is still lower than that when using water or other solutions (Lee et al. 2012; Swenson et al. 2015). Likewise, when focusing on sterols and fatty acids, higher concentrations of organic solvent are needed, and aqueous solutions are better at extracting polar and small compounds due to the polar nature of the compounds (Swenson et al. 2015). Our results suggest that the metabolome detected by GC-TOFMS/MS might not accurately reflect the state of the soil and that UHPLC-QTOFMS/MS may yield more informative results in these sandy soils. However, this result is based on one study site, and the results may be different if focusing on different soils.

Effect of long-term organic and inorganic fertiliser application on soil metabolites detected by UHPLC-QTOFMS/MS

The selected compounds detected by UHPLC-QTOFMS/MS were clustered into three groups. The first group included compounds that were more concentrated in the

topsoil than the subsoil and more concentrated under chemical fertiliser application (No-M) than under low and high manure application (n = 35). Their lower concentration in the subsoil could be due to the higher absorption by soil particles, as limited compounds in the topsoil leached to the subsoil. The group comprised predominantly large molecular compounds, such as corticosterone, phenylacetamide, coniferylaldehyde, quinolone, nicotine, and hexadecylamine, which might have been derived as secondary metabolites from soil microorganisms after they utilised the nutrients from chemical fertilisers. The long-term use of chemical fertiliser might stimulate microorganisms to synthesise those compounds and assimilate the inorganic nutrients to adapt to the environmental changes caused by chemical fertiliser application. The second group of compounds had the highest concentration in the topsoil under high manure application. This group might have been derived from farmyard manure or microbial cycling. The last group had the highest concentration in the subsoil, either because they leached into the subsoil because of a lower adsorption ability, or because they were derived from microorganisms adapted to oxygen-deficient conditions in the subsoil (Ma et al. 2020a).

The metabolome detected by UHPLC-QTOFMS/MS was strongly correlated to total and dissolved SOM, indicating that UHPLC-QTOFMS/MS better reflected SOM content and composition, at least in this sandy bulk soil. In addition, the compounds were not strongly correlated to the dissolved organic C but were strongly correlated to extractable organic N. This might have been caused by the decoupling of C and N in

some compounds.

Correlations between soil metabolism and the bacterial community

Dissolved organic C, especially low molecular-weight compounds, including root exudates, could be utilised directly as C sources by soil microbes (Swenson et al. 2015). Therefore, soil metabolomics could improve our understanding of the coupling between organic/inorganic compounds and microbial communities in the soil (Johns et al. 2017). In this study, the most correlated factor for both bacterial and fungal community succession was MTA, followed by N-epsilon-acetyllysine, gamma-glutamylleucine, histidine, and 3-indolepropionic acid for the bacterial community (Fig. 6A) and 2'-O-methyladenosine, 1,4-cyclohexanedione, isobutyryl-L-carnitine, and corticosterone for the fungal community (Fig. 6B). MTA is a naturally occurring sulphur-containing nucleoside, indicating that S metabolism is important for the formation of microbial communities. Recently, S might have become a limiting element for microbial growth as a result of considerably decreased sulphur dioxide emissions following strict air-quality regulations, application of fertilisers with a limited S content, and a reduced S return via farmyard manure (Piotrowska-Długosz et al. 2017).

N-epsilon-acetyllysine is a derivative of the amino acid lysine, and A glutamyl-L-amino acid is obtained through formal condensation of the gamma-carboxy group of glutamic acid with the amino group of leucine. Indole-3-propionic acid is a bacterial metabolite that exerts antioxidant and neuroprotective activities. Most of

these metabolites are amino acid derivatives, which can be utilised by soil microorganisms, hence regulating microbial activity and/or changing microbial diversity (Ma et al. 2021b). Maltose and sucrose are low molecular compounds directly utilised as energy sources by microbes in the soil (Vives-Peris et al. 2020). In particular, organic acids and sugars are the main drivers of shifts in soil microbial communities in the rhizosphere and are positively or negatively correlated with the relative abundances of bacteria (Song et al. 2020; Swenson et al. 2015).

Our results showed that metabolite profiling and high-throughput sequencing could be successfully integrated. We found more positive correlations between bacterial taxa and metabolites (69.5%) and fungal taxa and metabolites (67.9%) than negative correlations in the co-occurrence network. The family Planococcaceae and genus *Paenibacillus* showed the most correlations with metabolites among the bacterial taxa in the network (Fig. 7A and Table S1). *Paenibacillus* is an important bacterium in bulk soil that plays an important role in N fixation, hormone production, siderophore secretion, and mineral nutrient activation (Li et al. 2021; Timmusk et al. 2005). In the rhizosphere, Proteobacteria are reported to be the main utilisers of plant root exudates (Haichar et al. 2008) and respond positively to low molecular-weight substances (Goldfarb et al. 2011). However, Bacteroidetes is not a dominant bacterial phylum in bulk soil but is found in high abundance in the rhizosphere (Aleklett et al. 2015). Therefore, it was not the dominant bacterial phylum in the tested bulk soil. In the case of metabolites, gamma-glutamylleucine had the most links (20) with bacterial

539 taxa.

540 Unlike the bacterial network, the fungal network was simpler, with fewer nodes
541 and lower total degrees (Fig. 7B and Table S2). Previous studies have demonstrated
542 that fungi tend to decompose recalcitrant SOC, such as lignin and cellulose, and
543 bacteria then utilise the fungal-derived products (de Boer et al. 2005). Among the
544 fungal taxa in the network, *Aspergillus caesiellus* and *Thermomyces lanuginosus* had
545 the most correlations (8) with metabolites. In addition, MTA was found with the most
546 degrees with the fungal taxa. Soil microbial community composition can be achieved
547 by high-throughput sequencing. However, the actual microbial functions, such as their
548 metabolism, are difficult to obtain with soil metagenome or amplicon sequencing
549 (Jansson and Hofmockel, 2018).

550 The soil metabolome was formed mainly of organic acids, sugars, and sugar
551 derivatives, which were largely negatively correlated with bacterial alpha-diversity.
552 Compared to sugars, organic acids accounted for more bacterial community
553 compositions at high taxonomic ranks, but this was reversed at the species and genus
554 levels. Keystone species in the co-occurrence network, such as *Microvirga*,
555 *Bryobacter*, and *Bradyrhizobium* were significantly correlated with organic acids and
556 sugars (Liu et al. 2020). We anticipate that these substrate-genome linkages could be
557 further evaluated and refined using other approaches. Stable isotope probing coupled
558 with labelled DNA sequencing (Orsi et al. 2016; Pepe-Ranney et al. 2016) and
559 integrated NanoSIMS and FISH imaging (Woebken et al. 2015; Fike et al. 2008) may

be used to examine the spatial localisation of microbes and their activities (Swenson et al. 2018). Complementary analyses of metabolic flux through real-time MS or NMR combined with stable isotopes may also offer a deeper understanding of metabolic network dynamics (Ina and David 2016; Jeong et al. 2017). A metabolomic profile alone cannot provide a complete understanding of interacting molecular pathways and their modes of regulation; the variation of metabolite levels cannot definitively infer functional change. Combining genomic and proteomic or transcriptomic results with metabolites may contribute toward a more holistic understanding of soil microbial function and regulation (Trauger et al. 2008).

Conclusions

We found that long-term farmyard manure application enhanced the total and dissolved soil contents of C and N. The metabolome detected by UHPLC-QTOFMS/MS was positively linearly correlated to SOM, EON, and MB-C, indicating that the metabolites detected by UHPLC-QTOFMS/MS reflect the soil organic matter content and composition. There were more positive correlations between bacterial and fungal taxa and metabolites than negative correlations in the network. The family Planococcaceae and genus *Paenibacillus* showed the most correlations with metabolites among the bacterial taxa in the network. Combining genomic and proteomic or transcriptomic results with metabolites may contribute toward a more holistic understanding of soil microbial function and regulation. It is impossible to extract all metabolites from soil, and the detected metabolites depend on the

extracting solution; therefore, a more detailed exploration of both extraction and detection methods that more accurately reflect the composition of soil compounds and their turnover is needed.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (32102488, 32172674); Zhejiang Provincial Natural Science Foundation of China (LZ23C150002); Zhejiang Key Research and Development Program (2022C02018, 2023C02016); the UK–China Virtual Joint Centre for Agricultural Nitrogen [grant number CINAg, BB/N013468/1], which is jointly supported by the Newton Fund, via UK BBSRC and NERC, and the Chinese Ministry of Science and Technology. The field experiment is maintained as part of the Rothamsted Long-Term Experiments National Capability, funded by BBSRC (BBS/E/C/000J0300). We thank the curators of the Electronic Rothamsted Archive (e-RA) for access to data from the Rothamsted Long-Term Experiments.

Conflict of interest

The authors declare no conflict of interest.

References

Aleklett K, Leff JW, Fierer N, Miranda H (2015) Wild plant species growing closely connected in a subalpine meadow host distinct root-associated bacterial communities. PeerJ, 3: e804.

Brailsford FL, Glanville HC, Golyshin PN, Marshall MR, Lloyd CE, Johnes PJ, Jones DL (2019) Nutrient enrichment induces a shift in dissolved organic carbon (DOC) metabolism in oligotrophic freshwater sediments. Sci Total Environ 690: 1131–1139.

Brown RW, Chadwick DR, Zang H, Jones DL (2021) Use of metabolomics to quantify changes in soil microbial function in response to fertiliser nitrogen supply and extreme drought. Soil Biol Biochem 160: 108351.

Brown RW, Chadwick DR, Bending GD, Collins CD, Whelton HL, Daulton E, Covington JA, Bull ID, Jones DL (2022) Nutrient (C, N and P) enrichment induces significant changes in the soil metabolite profile and microbial carbon partitioning. Soil Biol Biochem 172, 108779.

Chang EH, Chung RS, Tsai YH (2010) Effect of different application rates of organic fertilizer on soil enzyme activity and microbial population. Soil Sci Plant Nutr 53: 132–140.

Cheng L, Zhang N, Yuan M, Xiao J, Qin Y, Deng Y, Tu Q, Xue K, Van Nostrand JD, Wu L (2017) Warming enhances old organic carbon decomposition through altering functional microbial communities. ISME J: 11, 1825–1835.

Chong J, Soufan O, Li C, Caraus I, Li S, Bourque G, Wishart DS, Xia J (2018) MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. Nucleic Acids Res 46: 486–494.

Cui J, Zhu Z, Xu X, Liu S, Jones D, Kuzyakov Y, Shibistova O, Wu J, Ge T (2020) Carbon and nitrogen recycling from microbial necromass to cope with C:N stoichiometric imbalance by priming. Soil Biol Biochem 142: 107720.

de Boer W, Folman LB, Summerbell RC, Lynne B (2005) Living in a fungal world: impact of fungi on soil bacterial niche development. FEMS Microbiol Rev 29: 795-811.

Defelice BC, Mehta SS, Samra S, Ajka T, Wancewicz B, Fahrmann JF, Fiehn O (2017) Mass Spectral Feature List Optimizer (MS-FLO): a tool to minimize false positive peak reports in untargeted LC-MS data processing. Anal Chem 89: 3250–3255.

Fiehn O, Wohlgemuth G, Scholz M, Kind T, Lee DY, Lu Y, Moon S, Nikolau B (2008) Quality control for plant metabolomics: reporting MSI-compliant studies. Plant J 53: 691-704.

Fike DA, Gammon CL, Ziebis W, Orphan VJ (2008) Micron-scale mapping of sulfur

637 cycling across the oxycline of a cyanobacterial mat: a paired nanoSIMS and CARD-
638 FISH approach. ISME J 2: 749–759.

639 Gardes M, Bruns TD (2010) ITS primers with enhanced specificity for
640 basidiomycetes--application to the identification of mycorrhizae and rusts. Mol Ecol 2:
641 113–118.

642 Goldfarb KC, Karaoz U, Hanson CA, Santee CA, Bradford MA (2011) Differential
643 Growth Responses of Soil Bacterial Taxa to Carbon Substrates of Varying Chemical
644 Recalcitrance. Front Microbiol 2: 94.

645 Gupta L, Ahmed S, Jain A, Misra R (2018) Emerging role of metabolomics in
646 rheumatology. Int J Rheum Dis 21: 1468–1477.

647 Haichar FE, Marol C, Berge O, Rangel-Castro JI, Prosser JI, Balesdent J, Heulin T,
648 Achouak W (2008) Plant host habitat and root exudates shape soil bacterial
649 community structure. ISME J 2(12): 1221–1230.

650 Hartman S, Sasidharan R, Voesenek L (2020) The role of ethylene in metabolic
651 acclimations to low oxygen. New Phytol 229: 5–7.

652 Haynes RJ, Mokolobate MS (2001) Amelioration of Al toxicity and P deficiency in
653 acid soils by additions of organic residues: a critical review of the phenomenon and
654 the mechanisms involved. Nutr Cycl Agroecosystems 59: 47–63.

655 Hirai MY, Yano M, Goodenowe DB, Kanaya S, Kimura T, Awazuhara M, Arita M,
656 Fujiwara T, Saito K (2004) Integration of transcriptomics and metabolomics for
657 understanding of global responses to nutritional stresses in Arabidopsis thaliana. P
658 Natl Acad Sci USA 101: 10205–10210.

659 Hoyle FC, Fang Y (2018) Impact of agricultural management practices on the nutrient
660 supply potential of soil organic matter under long-term farming systems. Soil Till Res
661 175: 71–81.

662 Ina A, David M (2016) Advantages and pitfalls of mass spectrometry based
663 metabolome profiling in systems biology. Int J Mol Sci 17: 632.

664 Jansson JK, Hofmockel KS (2018) The soil microbiome-from metagenomics to
665 metaphenomics. Curr Opinion Microbiol 43: 162–168.

666 Jeong S, Eskandari R, Sun MP, Alvarez J, Keshari KR (2017) Real-time quantitative
667 analysis of metabolic flux in live cells using a hyperpolarized micromagnetic
668 resonance spectrometer. Sci Adv 3: e170034.

669 Johns CW, Lee AB, Springer TI, Rosskopf EN, Hong JC, Turechek W, Kokalis-
670 Burelle N, Finley NL (2017) Using NMR-based metabolomics to monitor the
671 biochemical composition of agricultural soils: A pilot study. Eur J Soil Biol 83: 98–
672 105.

673 Johnston CA, Groffman P, Breshears DD, Cardon ZG, Currie W, Emanuel W,

Gaudinski J, Jackson RB, Lajtha K, Nadelhoffer K (2004) Carbon cycling in soil. Front Ecol Environ 2: 522–528.

Jones DL, Willett VB (2006) Experimental evaluation of methods to quantify dissolved organic nitrogen (DON) and dissolved organic carbon (DOC) in soil. Soil Biol Biochem 38(5): 991–999.

Jones DL, Owen AG, Farrar JF（2002) Simple method to enable the high resolution determination of total free amino acids in soil solutions and soil extracts. Soil Biol Biochem 34: 1893–1902.

Jozefczuk S, Klie S, Catchpole G, Szymanski J, Willmitzer L (2014) Metabolomic and transcriptomic stress response of Escherichia coli. Mol Syst Biol 6: 364.

Kai F, Zhang Z, Cai W, Liu W, Xu M, Yin H, Wang A, He Z, Ye D (2017) Biodiversity and species competition regulate the resilience of microbial biofilm community. Mol Ecol 26 (21): 6170–6182

Lee DY, Bowen BP, Nguyen DH, Parsa S, Huang Y, Mao JH, Northen TR (2012) Low-dose ionizing radiation-induced blood plasma metabolic response in a diverse genetic mouse population. Radiat Res 178: 551–555.

Li Q, He X, Liu P, Zhang H, Chen S (2021) Synthesis of nitrogenase by Paenibacillus sabinae T27 in presence of high levels of ammonia during anaerobic fermentation. Appl Microbiol Biot 105 (7): 2889–2899

Liang C, Amelung W, Lehmann J, Matthias K (2019) Quantitative assessment of microbial necromass contribution to soil organic matter. Global Change Biol 25: 3578–3590

Liu K, Ding X, Wang J (2020) Soil metabolome correlates with bacterial diversity and co-occurrence patterns in root-associated soils on the Tibetan Plateau. Sci Total Environ 735: 139572.

Liu L, Wang T, Li S, Hao R, Li Q (2021) Combined analysis of microbial community and microbial metabolites based on untargeted metabolomics during pig manure composting. Biodegradation 32: 217–228.

Liu Q, Cornelius TA, Zhu Z, Muhammad S, Wei X, Johanna P, Wu J, Ge T (2022) Vertical and horizontal shifts in the microbial community structure of paddy soil under long-term fertilization regimes. Appl Soil Ecol 169: 104248

Liu S, Wang J, Pu S, Blagodatskaya E, Kuzyakov Y, Razavi BS (2020) Impact of manure on soil biochemical properties: A global synthesis. Sci Total Environ 745: 141003.

Ma Q, Xu M, Liu M, Cao X, Hill PW, Chadwick DR, Wu L, Jones DL (2022) Organic and inorganic sulphur and nitrogen uptake by co-existing grassland plant species competing with soil microorganisms. Soil Biol Biochem 168: 108627.

Ma Q, Kuzyakov Y, Pan W, Tang S, Chadwick DR, Wen Y, Hill PW, Macdonald A, Ge T, Si L, Wu L, Jones DL (2021a) Substrate control of sulphur utilisation and microbial stoichiometry in soil: Results of 13C, 15N, 14C, and 35S quad labelling. ISME J 15 (11): 3148–3158.

Ma Q, Tang S, Pan W, Zhou J, Chadwick DR, Hill PW, Wu L, Jones DL (2021b) Effects of farmyard manure on soil S cycling: Substrate level exploration of high- and low-molecular weight organic S decomposition. Soil Biol Biochem 160: 108359.

Ma Q, Wen Y, Ma J, Macdonald A, Hill PW, Chadwick DR, Wu L, Jones DL (2020a) Long-term farmyard manure application affects soil organic phosphorus cycling: A combined metagenomic and 33P/14C labelling study. Soil Biol Biochem 149: 107959.

Ma Q, Wen Y, Wang D, Sun X, Hill PW, Macdonald A, Chadwick DR, Wu L, Jones DL (2020b) Farmyard manure applications stimulate soil carbon and nitrogen cycling by boosting microbial biomass rather than changing its community composition. Soil Biol Biochem 144: 107760.

Ma Q, Wen Y, Pan W, Macdonald A, Hill PW, Chadwick DR, Wu L, Jones DL (2020c) Soil carbon, nitrogen, and sulphur status affects the metabolism of organic S but not its uptake by microorganisms. Soil Biol Biochem, 149, 107943.

Ma Q, Wu L, Wang J, Ma J, Zheng N, Hill PW, Chadwick DR, Jones DL (2018) Fertilizer regime changes the competitive uptake of organic nitrogen by wheat and soil microorganisms: An in-situ uptake test using 13C, 15N labelling, and 13C-PLFA analysis. Soil Biol Biochem 125: 319–327.

Maillard E, Angers DA (2014) Animal manure application and soil organic carbon stocks: a meta-analysis. Global Chang Biol 20: 666–679.

Matyash V, Liebisch G, Kurzchalia TV, Shevchenko A, Schwudke D (2008) Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics. J Lipid Res 49: 1137–1146.

Mcguire KL, Treseder KK (2010) Microbial communities and their relevance for ecosystem models: Decomposition as a case study. Soil Biol Biochem 42(4): 529–535.

Mcleod ML, Bullington L, Cleveland CC, Rousk J, Lekberg Y (2021) Invasive plant-derived dissolved organic matter alters microbial communities and carbon cycling in soils. Soil Biol Biochem 156: 108191.

Miura M, Hill PW, Jones DL (2020) Impact of a single freeze-thaw and dry-wet event on soil solutes and microbial metabolites. Appl Soil Ecol 153: 103636.

Nilsson RH, Larsson K, Taylor AFS, Bengtsson-Palme J, Jeppesen TS, Schigel D, Kennedy P, Picard K, Glöckner FO, Tedersoo L, Saar I, Kõljalg U, Abarenkov K (2019) The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. Nucleic Acids Res 47: 259–264.

748 Orsi WD, Smith JM, Liu S, Liu Z, Sakamoto CM, Wilken S, Poirier C, Richards TA,
749 Keeling PJ, Worden AZ, Santoro AE (2016) Diverse, uncultivated bacteria and
750 archaea underlying the cycling of dissolved protein in the ocean. ISME J 10 (9):
751 2158–2173.

752 Pepe-Ranney C, Koechli C, Potrafka R, Andam C, Eggleston E, Garcia-Pichel F,
753 Buckley DH (2016) Non-cyanobacterial diazotrophs mediate dinitrogen fixation in
754 biological soil crusts during early crust formation. ISME J 10 (2): 287–298.

755 Piotrowska-Długosz A, Siwik-Ziomek A, D Ugosz J, Gozdowski D (2017) Spatio-
756 temporal variability of soil sulfur content and arylsulfatase activity at a conventionally
757 managed arable field. Geoderma 295, 107–118.

758 Quast C, Pruesse E, Yilmaz P, Gerken J, Glckner FO (2012) The SILVA ribosomal
759 RNA gene database project: Improved data processing and web-based tools. Nucleic
760 Acids Res 41: 590–596.

761 Roberts P, Jones DL (2008) Critical evaluation of methods for determining total
762 protein in soil solution. Soil Biol Biochem 40: 1485–1495.

763 Schmidt M, Torn MS, Abiven S, Dittmar T, Guggenberger G, Janssens IA, Kleber M,
764 K Gel-Knabner I, Lehmann J, Manning D (2011) Persistence of soil organic matter as
765 an ecosystem property. Nature 478, 49–56.

766 Song Y, Li X, Yao S, Yang X, Jiang X (2020) Correlations between soil metabolomics
767 and bacterial community structures in the pepper rhizosphere under plastic
768 greenhouse cultivation. Sci Total Environ 728: 138439.

769 Swenson TL, Jenkins S, Bowen BP, Northen TR (2015) Untargeted soil metabolomics
770 methods for analysis of extractable organic matter. Soil Biol Biochem 80: 189–198.

771 Swenson TL, Karaoz U, Swenson JM, Bowen BP, Northen TR (2018) Linking soil
772 biology and chemistry in biological soil crust using isolate exometabolomics. Nat
773 Commun 9: 19.

774 Thangarajan R, Bolan NS, Guanglong T, Naidu R, Kunhikrishnan A（2013) Role of
775 organic amendment application on greenhouse gas emission from soil. Sci Total
776 Environ 465: 72–96.

777 Timmusk S, Grantcharova N, Wagner E, Gerhart H (2005) *Paenibacillus polymyxa*
778 invades plant roots and forms bioflims. Appl Environ Microb 71: 7292–7300.

779 Trauger SA, Kalisak E, Kalisiak J, Morita H, Weinberg MV, Menon AL, Poole FL 2nd,
780 Adams MW, Siuzdak G (2008) Correlating the transcriptome, proteome, and
781 metabolome in the environmental adaptation of a hyperthermophile. J Proteome Res 7:
782 1027–1035.

783 Tsugawa H, Cajka T, Kind T, Ma Y, Higgins B, Ikeda K, Kanazawa M,
784 VanderGheynst J, Fiehn O, Arita M (2015) MS-DIAL: data-independent MS/MS

deconvolution for comprehensive metabolome analysis. Nat Methods 12: 523–526.

Vance ED, Brookes PC, Jenkinson DS (1987) An extraction method for measuring soil microbial biomass C. Soil Biol Biochem 19: 703–707.

Vives-Peris V, De Ollas C, Gómez-Cadenas A, Pérez-Clemente RM (2020) Root exudates: from plant to rhizosphere and beyond. Plant Cell Rep 39: 3–17.

Vong PC, Dedourge O, Lasserre-Joulin F, Guckert A (2003) Immobilized-S, microbial biomass-S and soil arylsulfatase activity in the rhizosphere soil of rape and barley as affected by labile substrate C and N additions. Soil Biol Biochem 35: 1651–1661.

Wang B, An S, Liang C, Liu Y, Kuzyakov Y (2021) Microbial necromass as the source of soil organic carbon in global ecosystems. Soil Biol Biochem 162: 108422.

Wang X, Yang F, Zhang Y, Xu G, Liu Y, Tian J, Peng G (2015) Evaluation and optimization of sample preparation methods for metabolic profiling analysis of *Escherichia coli*. Electrophoresis 36: 2140–2147.

Warren CR (2020) Pools and fluxes of osmolytes in moist soil and dry soil that has been re-wet. Soil Biol Biochem 150, 108012.

Woebken D, Burow LC, Behnam F, Mayali X, Schintlmeister A, Fleming ED, Prufert-Bebout L, Singer SW, López Cortés A, Hoehler TM, Pett-Ridge J, Spormann AM, Wagner M, Weber PK, Bebout BM (2015) Revisiting N_2 fixation in Guerrero Negro intertidal microbial mats with a functional single-cell approach. ISME J 9: 485–496.

Wei L, Ge T, Zhu Z, Luo Y, Yang Y, Xiao M, Yan Z, Li Y, Wu J, Kuzyakov Y (2021) Comparing carbon and nitrogen stocks in paddy and upland soils: Accumulation, stabilization mechanisms, and environmental drivers. Geoderma 398: 115121.

Withers E, Hill PW, Chadwick DR, Jones DL (2020) Use of untargeted metabolomics for assessing soil quality and microbial function. Soil Biol Biochem 143: 107758.

Xia J, Wishart DS (2016) Using MetaboAnalyst 3.0 for Comprehensive Metabolomics Data Analysis. Curr Protoc Bioinform 55, 1–91.

Yan Z, Chen S, Dari B, Sihi D, Chen Q (2018) Phosphorus transformation response to soil properties changes induced by manure application in a calcareous soil. Geoderma 322: 163–171.

Zhu Z, Fang Y, Liang Y, Li Y, Liu S, Li Y, Li B, Gao W, Yuan H, Kuzyakov Y, Wu J, Richter A, Ge T (2022) Stoichiometric regulation of priming effects and soil carbon balance by microbial life strategies. Soil Biol Biochem 169: 108669

Figure captions

Fig. 1. Heat map of the 50 most significant known metabolites (detected by GC-TOFMS/MS) identified by ANOVA. Metabolites were clustered by Pearson correlation. The colour of squares linking metabolites to samples ranges from blue to red, indicating the number of standard deviations from the mean. No-M, chemical fertilisers without application of manure; Low-M, medium application rate of manure with chemical fertilisers; High-M, high application rate of only manure. 4-amino acid: 4-amino butyric acid; 4-hydro acid: 4-hydroxybenzoic acid; N-acety.: N-acetylmannosamine; UDP-N-acety.: UDP-N-acetylglucosamine; gly. alf. phos.: glycerol-alpha-phosphate; glucose-1-phos: glucose-1-phosphate; beta-mann.: beta-mannosylglycerate.

Fig. 2. Heat map of the 50 most significant known metabolites (detected by UHPLC-QTOFMS/MS) identified by ANOVA. Metabolites were clustered by Pearson correlation. The colour of squares linking metabolites to samples ranges from blue to red, indicating the number of standard deviations from the mean. No-M, chemical fertilisers without application of manure; Low-M, medium application rate of manure with chemical fertilisers; High-M, high application rate of only manure. Butyl: butylisopropylamine, N-N,N-dipro: N-(4-piperidinyl)-N,N-dipropylamine; 4-hydro: 4-hydroxy-1-(2-hydroxyethyl)-2,2,6,6-tetramethylpiperidine; 3-indol. Acid: 3-indoleacetic acid, arach. dopam.: arachidonyl dopamine; N,N-diethyl: N,N-diethyl-2-

841 aminoethanol; indole-3-carbox.: indole-3-carboxaldehyde; guanid. acid: 4-
842 guanidinobutyric acid; isobu. carni.: isobutyryl-L-carnitine; 1,1-dimet.: 1,1-dimethyl-
843 4-phenylpiperazinium; 4-amino. Acid: 4-aminobenzoic acid; 5-methy.: 5'-
844 methylthioadenosine; glycer.: glycerophosphocholine; atrazine-desis.: atrazine-
845 desisopropyl-2-hydroxy; 8-oxo-2-deoxy.: 8-oxo-2-deoxyadenosine; N-epsilon-acety.:
846 N-epsilon-acetyllysine; gamma-gluta.: gamma-glutamylleucine.

847 **Fig. 3.** Principal component analysis (PCA) of soil carbon and nitrogen content
848 detected by traditional methods (A), and metabolites detected by GC-TOFMS/MS (B)
849 and UHPLC-QTOFMS/MS (C) under long-term (1964–2018) manure and chemical
850 fertiliser applications. Prior to analysis, the data were $\log_{10}$ transformed. No-M,
851 chemical fertilisers without manure application; Low-M, medium application rate of
852 manure with chemical fertilisers; High-M, high application rate of only manure; T,
853 topsoil; S, subsoil.

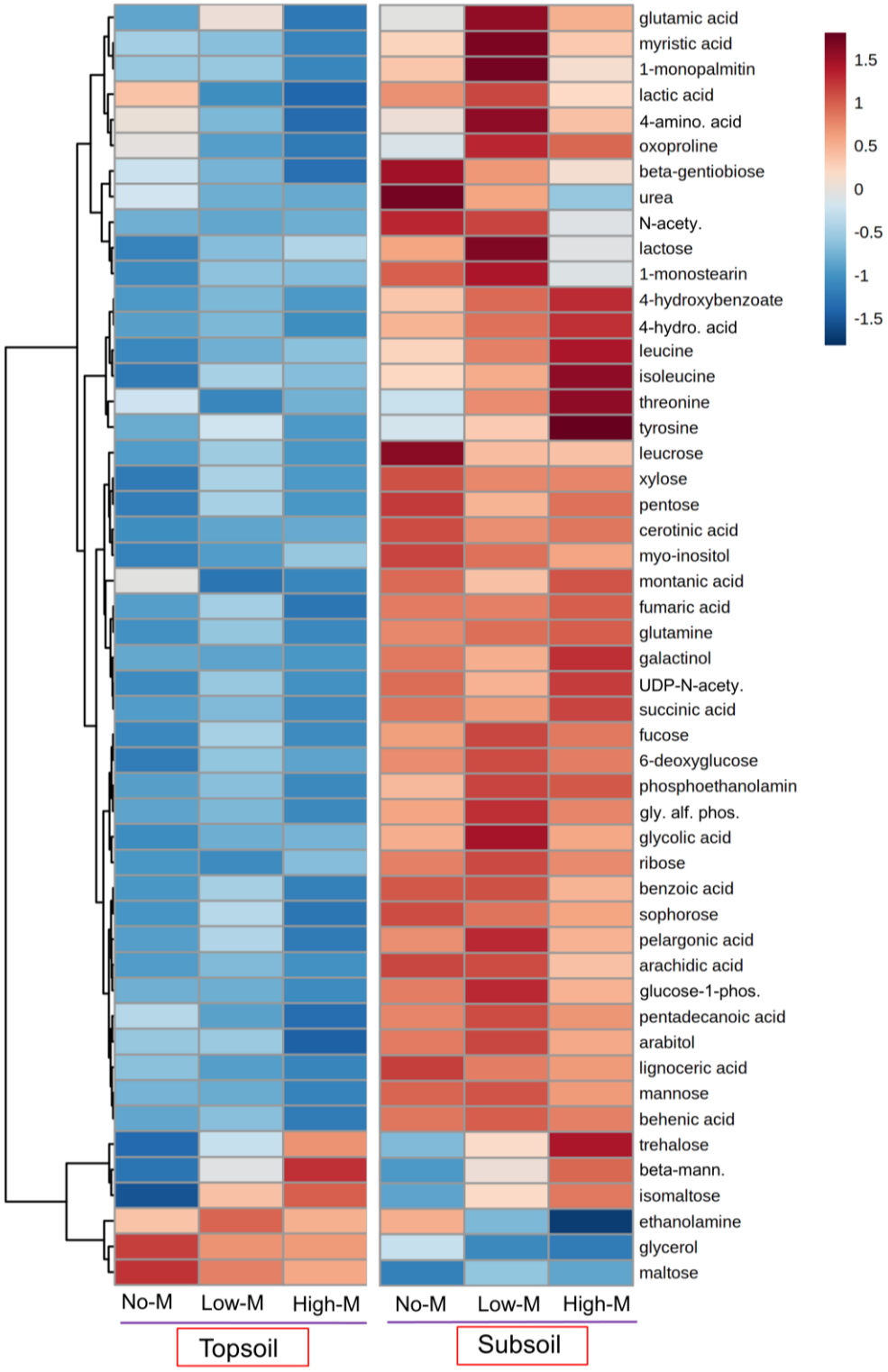
854 **Fig. 4.** Linear correlations of metabolites detected by GC-TOFMS/MS (A) and
855 UHPLC-QTOFMS/MS (B) with soil carbon and nitrogen content detected by
856 traditional methods under long-term (1964–2018) manure and chemical fertiliser
857 applications. DOC, dissolved organic carbon; SOM, soil organic matter; EON,
858 extractable organic N; MB-C, microbial biomass carbon; MB-N, microbial biomass
859 nitrogen.

860 **Fig. 5.** The response of bacterial (A) and fungal (B) communities at phylum to genus
861 levels to long-term organic and inorganic fertiliser application based on a linear

discriminant effect size analysis. Only taxa meeting a linear discriminant analysis significance threshold of LDA > 3 are shown and colour-coded. The six rings of the cladogram indicate the domain (d), phylum (p), class (c), order (o), family (f), and genus (g), from inside to outside.

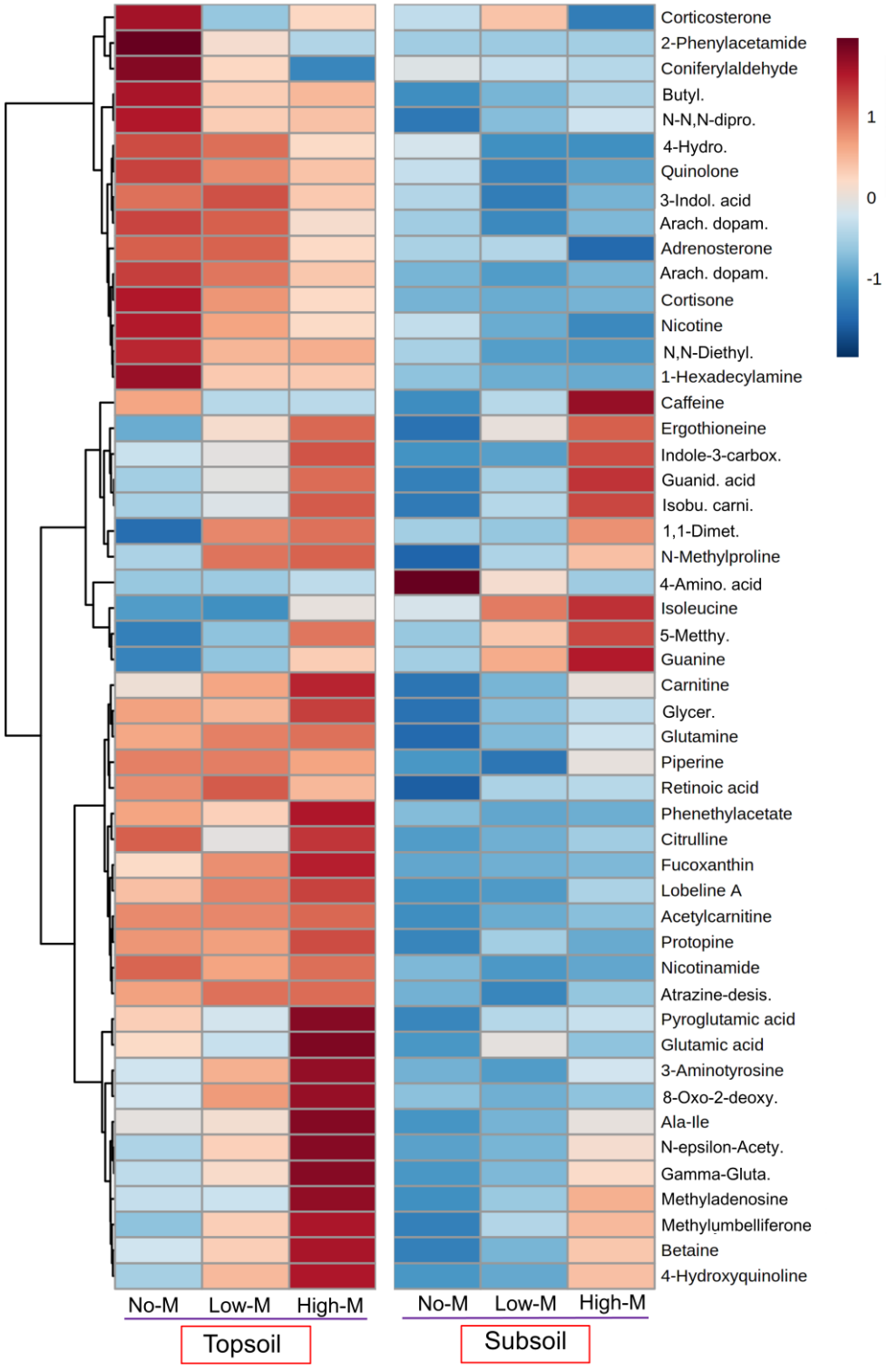
Fig. 6. Random forest analysis to determine factors affecting bacterial (A) and fungal (B) community succession. The metabolites detected by UHPLC-QTOFMS/MS were used in this analysis.

Fig. 7. Co-occurrence network of the metabolites and bacterial (A) and fungal taxa (B). The node size represented the degree in the network. Only significant Pearson correlation coefficients ($r > 0.8$ or $r < –0.8$ and $p < 0.05$) are shown. The metabolites detected by UHPLC-QTOFMS/MS were used in this analysis. Light purple and red lines indicate positive and negative correlations, respectively. Pink circles represent microorganisms, and green circles represent metabolites.



875

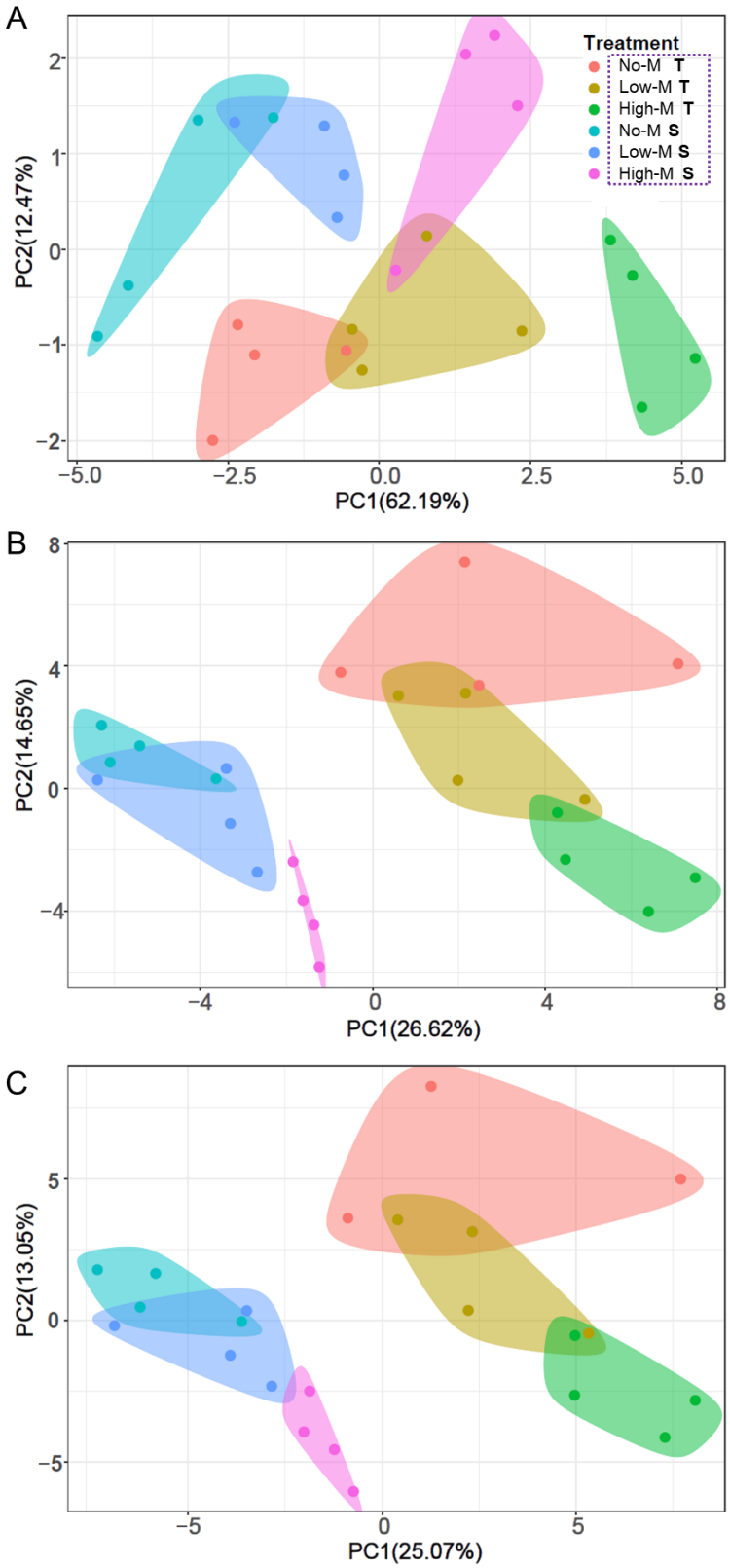
876 Figure 1



877

878 Figure 2

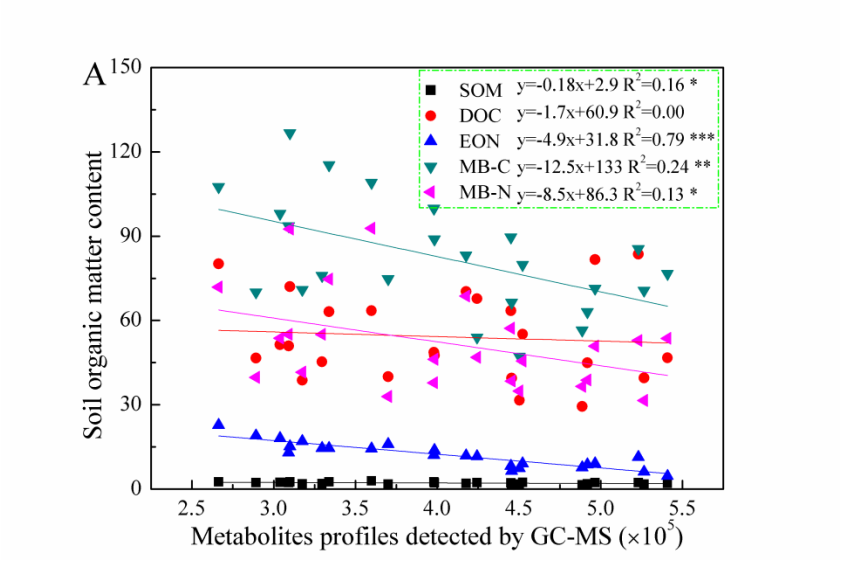
879



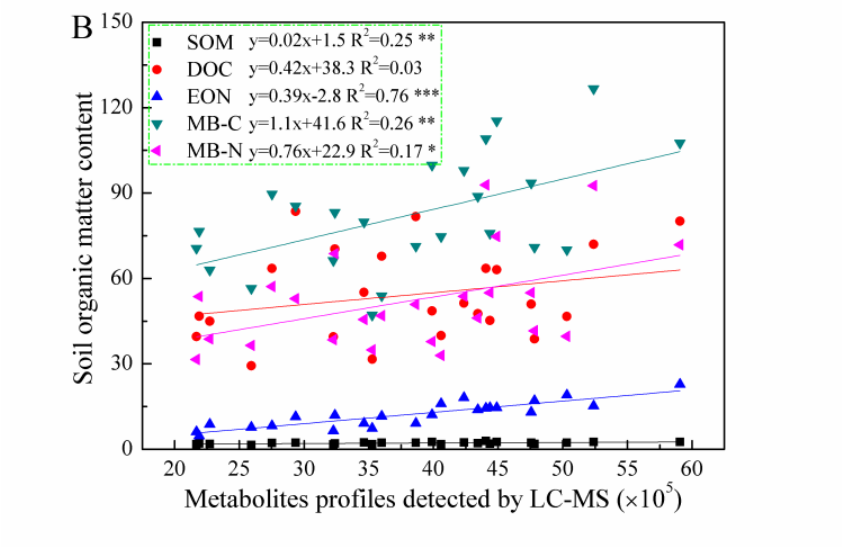
880

881 Figure 3

882



883



884 Figure 4

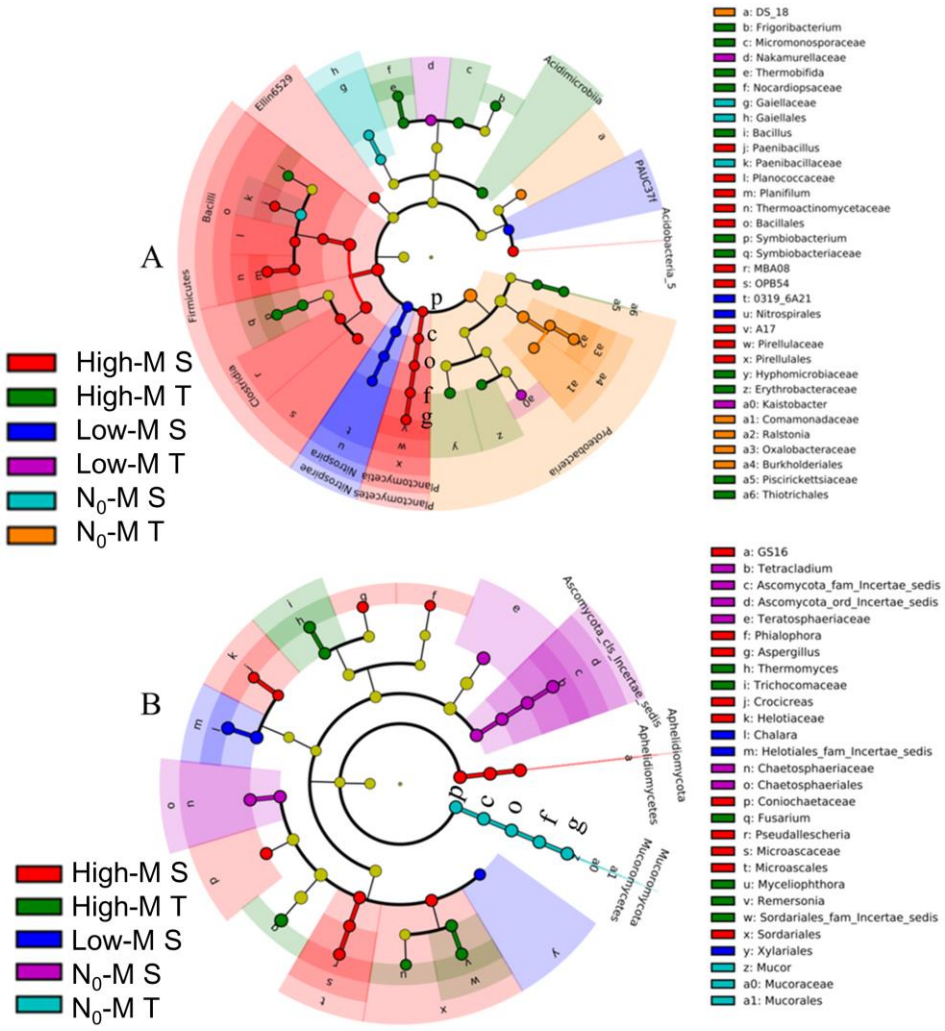

A
High-M S
High-M T
Low-M S
Low-M T
N0-M S
N0-M T
a: DS_18
b: Frigoribacterium
c: Micromonosporaceae
d: Nakamurellaceae
e: Thermobifida
f: Nocardiopsaceae
g: Gaiellaceae
h: Gaiellales
i: Bacillus
j: Paenibacillus
k: Paenibacillaceae
l: Planococcaceae
m: Planifilum
n: Thermoactinomycetaceae
o: Bacillales
p: Symbiobacterium
q: Symbiobacteriaceae
r: MBA08
s: OPB54
t: 0319_6A21
u: Nitrospirales
v: A17
w: Pirellulaceae
x: Pirellulales
y: Hyphomicrobiaceae
z: Erythrobacteraceae
a0: Kaistobacter
a1: Comamonadaceae
a2: Ralstonia
a3: Oxalobacteraceae
a4: Burkholderiales
a5: Piscirickettsiaceae
a6: Thiotrichales
B
High-M S
High-M T
Low-M S
N0-M S
N0-M T
a: GS16
b: Tetracladium
c: Ascomycota_fam_Incertae_sedis
d: Ascomycota_ord_Incertae_sedis
e: Teratosphaeriaceae
f: Phialophora
g: Aspergillus
h: Thermomyces
i: Trichocomaceae
j: Crocicreas
k: Helotiaceae
l: Chalara
m: Helotiales_fam_Incertae_sedis
n: Chaetosphaeriaceae
o: Chaetosphaeriales
p: Coniochaetaceae
q: Fusarium
r: Pseudallescheria
s: Microascaceae
t: Microascales
u: Myceliophthora
v: Remersonia
w: Sordariales_fam_Incertae_sedis
x: Sordariales
y: Xylariales
z: Mucor
a0: Mucoraceae
a1: Mucorales

Figure 5

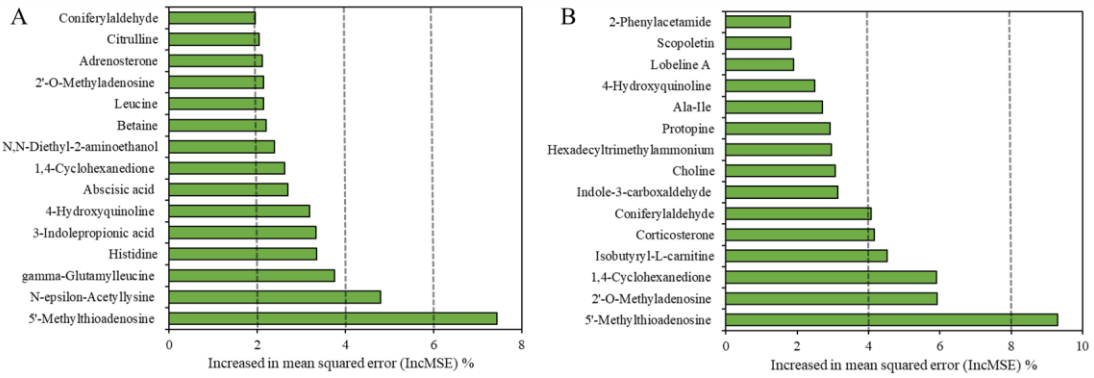
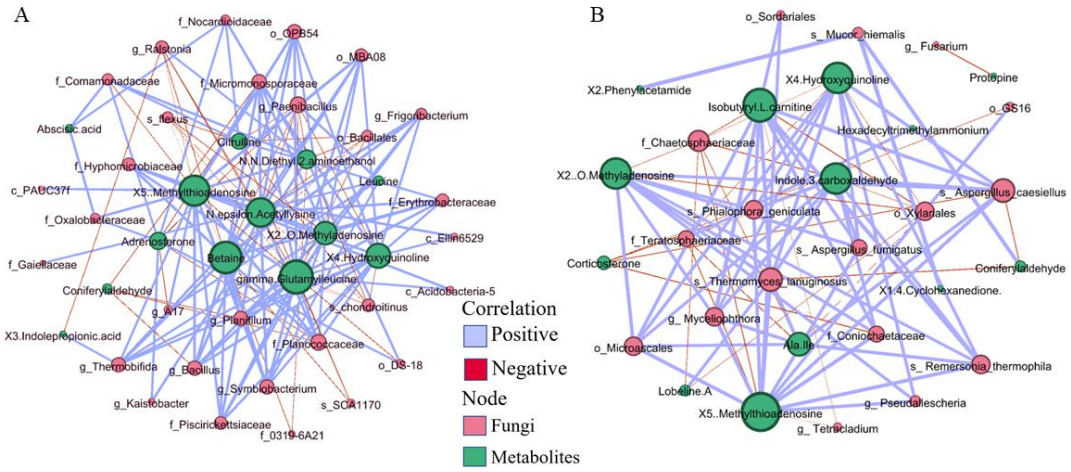

A
Coniferylaldehyde
Citrulline
Adrenosterone
2'-O-Methyladenosine
Leucine
Betaine
N,N-Diethyl-2-aminoethanol
1,4-Cyclohexanedione
Abscisic acid
4-Hydroxyquinoline
3-Indolepropionic acid
Histidine
gamma-Glutamylleucine
N-epsilon-Acetyllysine
5'-Methylthioadenosine
0
2
4
6
8
Increased in mean squared error (IncMSE) %
B
2-Phenylacetamide
Scopoletin
Lobeline A
4-Hydroxyquinoline
Ala-Ile
Protopine
Hexadecyltrimethylammonium
Choline
Indole-3-carboxaldehyde
Coniferylaldehyde
Corticosterone
Isobutyryl-L-carnitine
1,4-Cyclohexanedione
2'-O-Methyladenosine
5'-Methylthioadenosine
0
2
4
6
8
10
Increased in mean squared error (IncMSE) %

Figure 6



902

903 Figure 7

904