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

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:

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 fertilizer application and the effect of this response on the microbial community, taking advantage of the Woburn Organic Manuring Experiment (UK), which has been running 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. The microbial community abundance was detected by high-throughput sequencing. We found that long-term farmyard manure application enhanced the total and dissolved C and N contents in the soil. The metabolite content detected by GC-TOFMS/MS (TOF detector with a cold injection unit) had a linear negative correlation to soil organic matter, extractable organic nitrogen (N), and microbial carbon (C). Conversely, the metabolite content detected by UHPLC-QTOFMS/MS was positively correlated to them, indicating the metabolites detected by UHPLC-QTOFMS/MS is the main components of soil soluble organic matter. More positive than negative correlations were observed between metabolites and bacterial (69.5%) or 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. Extraction and detection methods are affected by not only the variety but also the number of detected metabolites. Careful

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

Keywords: soil organic matter, dissolved organic matter, chemical fertilizer, 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). Dissolved organic matter (DOM) is the most biologically-accessible component of SOM, playing a crucial role in C, N, and sulfur (S) cycling (Ma et al. 2020b, 2021a; Swenson et al. 2015). Soil microorganisms derive metabolites predominantly from SOM and its biomass turnover (Brown et al. 2021; Liang et al. 2019). DOM 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). DOM 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).

Fertilization is one of the most important field management interventions that strongly affects soil element content, nutrient cycling, and the microbial community. Annually, agricultural production produces about seven billion tons of farmyard manure (FYM) globally (Thangarajan et al. 2013). Manure application can increase soil structural stability and nutrient levels, enhancing soil C sequestration and biological activity in arable land (Maillard and Angers, 2014). Partly substituting inorganic fertilizer 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 (Liu et al. 2020; Ma et al. 2018). 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 fertilizers, long-term FYM application generally improves the total and DOM content of 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 FYM and chemical fertilizer application 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 growth, development, and function. 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 in the biochemical response of a microbial species (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 soil microbial functional responses to changes in their environment, such as fertilizer application, extreme drought, and dry-wet or freeze-thaw events (Brown et al. 2021; Miura et al. 2020).

Traditionally, soil DOM quantification is achieved 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 accessible and widely-used methods due to their broad analytical scope (alcohols, fatty acids, sterols, carbohydrates, amino acids, etc), low operation cost, and availability of 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 most suitable method that 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 fertilizer 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 the utilization and mineralization of organic C and N (Ma et al. 2018). Nutrient (C, N and P) enrichment induces significant changes in the soil metabolite profile and microbial C partitioning. A recent study based on UHPLC-MS/MS found that inorganic nutrient enrichment causes substantial shifts in both secondary and primary metabolism and changes in resource flow and soil functioning, and the microbial community composition showed significant metabolic flexibility (Brown et al. 2022). The systematic coupling of the microbial community and soil metabolomics can provide valuable information to improve our understanding of microbial strategies in response to environmental stress (Swenson et al. 2018). However, the elucidation of this link is difficult due to a large number of metabolites and the 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 fertilizer application. We hypothesised that (1) the total DOM detected by traditional methods, GC-TOFMS/MS and UHPLC-QTOFMS/MS, should be positive 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 fertilizers 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 fertilizers (low manure application, Low-M), and chemical fertilizers only (No-M), with P and K inputs equivalent to 25–50 t ha^{-1} y^{-1} FYM. Each treatment consisted of four replications. 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, winter rye) since 1964.

The treatment plots received chemical fertilizers or organic manures for three periods between 1964 and 2018. In the High-M treatment, FYM was applied from 1964–72, 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 fertilizers (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 fertilizers as N, P, and K at rates equivalent to High-M during the same years. Since 2003, all treatments

161 received annual N (nitrochalk), P (triple superphosphate), and K and S (potassium
162 sulfate) fertilizers at 165 (equivalent annual rate for a 5-year crop rotation), 20, 83, and
163 36 kg ha^{-1}, respectively. All other aspects of agronomic management, including
164 harvesting, tillage regime, herbicides (including spring-applied Atlantis at 400 mL ha^{-1},
165 Hiatus at 50 g ha^{-1}, and Sprinter at 2 L ha^{-1}) and fungicides (including spring-applied
166 Keystone at 500 mL ha^{-1}, Folicur at 800 mL ha^{-1}, and Cello at 630 mL ha^{-1}), were
167 consistent among the three treatments. The total N, P, and S inputs during the build-up
168 phase (1964–2018) under No-M were 2.46, 1.77, and 0.96 t, respectively. The total C,
169 N, P, and S inputs under High-M were 112.50, 5.80, 1.26, and 1.22 t, respectively, while
170 that under the Low-M treatment was 14.10, 2.63, 1.69, and 1.00 t. Further details of the
171 agronomic regime and experiment can be found in Ma et al. (2020b).

172 From each of four plots per treatment, the topsoil (0–23 cm plow layer) and subsoil
173 (23–38 cm) samples were collected using a 2.5 cm diameter corer (18 cores per plot to
174 make up one replicate). Winter rye (*Secale cereale* L.) was sown in the plots in 2018,
175 and sampling was performed at the grain-filling stage. The soil was thoroughly mixed
176 by hand and passed through a 5 mm sieve to any remove roots, stones, and earthworms.
177 The soil samples were then portioned into three parts: the first was stored at –80 °C to
178 analyse soil metabolites and microbial community; the second was stored at 4 °C to
179 assess soil microbial biomass; and the third was air-dried to determine basic soil
180 properties.

181 *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 colorimetrically 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). Total soluble protein in the extracts was estimated by the acid hydrolysis of proteins in solution with the subsequent determination of the amino acids as described by Roberts and David (2008) and have been reported previously (Ma et al. 2020b). To quantify the fraction of peptides and free amino acids, 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). 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 running 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

225 of 17 spectra s^{-1}, from 80–500 Da, at –70 eV ionization energy and 1800 V detector
226 voltage, with a 230 °C transfer line and 250 °C ion source (Withers et al. 2020). A
227 mixture of internal retention index markers was prepared using fatty acidmethyl esters
228 of C8, C9, C10, C12, C14, C16, C18, C20, C22, C24, C26, C28, and C30 linear chain
229 length, dissolved in chloroform at concentrations of 0.8 mg ml^{-1} (C8–C16) or 0.4 mg
230 ml^{-1} (C18–C30) as detailed shown in Fiehn et al. (2008). The Raw data files are
231 preprocessed directly after data acquisition and stored as ChromaTOF-specific *.peg
232 files. ChromaTOF vs. 2.32 (Leco Corp.) is used for data preprocessing without
233 smoothing, with a 3 s peak width, baseline subtraction just above the noise level, and
234 automatic mass spectral deconvolution and peak detection at signal/noise levels of 5:1
235 throughout the chromatogram. Apex masses are reported for use in the BinBase
236 algorithm. The results are exported to a data server with absolute spectra intensities and
237 further processed by a filtering algorithm implemented in the metabolomics BinBase
238 database as shown in Withers et al. (2020). Both known and unknown compounds were
239 analysed using MetaboAnalyst v4.0 (Chong et al., 2018; Xia and Wishart, 2016). Prior
240 to analysis, the data were log_{10} transformed and scaled by Pareto scaling (Chong et al.
241 2018).

242 *Untargeted metabolomics detected by UHPLC-QTOFMS/MS*

243 Complex lipid extraction was conducted using a modified bi-phasic method (Matyash
244 et al. 2008), which is advantageous as the lipids are retained in the upper extraction
245 phase, and the Methyl tertiary-butyl ether (MTBE) solvent has a density lower than that

246 of water. Compared to chloroform ($CHCl_3$), MTBE can be detected directly without the
247 risk of contamination from the interphase or aqueous phase. Briefly, 225 µL of MeOH
248 with internal standards was added to a 40 mg freeze-dried and ground soil sample and
249 vortexed for 20 s; 750 µL MTBE was subsequently added and vortexed for 10 min.
250 Samples were placed in a bead grinder for 30 s and then shaken for 6 min at 4 °C; 188
251 µL of MS-grade water was added, and the sample was centrifuged for 2 min at 14 000
252 × *g* at 4 °C. The upper phase was transferred to two tubes (350 µL/tube), and one tube
253 was evaporated to dryness using a SpeedVac. Dried extracts were re-suspended with a
254 mixture of 1:9 toluene: MeOH (v/v) and an internal standard. The samples were
255 analysed using an Agilent 1290 Infinity liquid chromatography (LC) system (G4220A
256 binary pump, G4226A autosampler, and G1316C Column Thermostat) coupled to an
257 Agilent 6530 MS (positive ion mode). Lipids were separated on an Acquity ultra high-
258 pressure chromatography (UHPLC) CSH C18 column (1.7 µm; 100 × 2.1 mm) (Brown
259 et al. 2021). The data were processed by the mass spectrometry-data independent
260 analysis (MS-DIAL) software (Tsugawa et al. 2015), followed by data clean-up using
261 the mass spectral feature list optimizer (MS-FLO) (Defelice et al. 2017). Peaks were
262 annotated and the MassHunter Quant software was applied to verify peak candidates
263 (Brown et al. 2021). To increase overall peak annotations, valid and reproducible peaks
264 were analysed using targeted MS/MS. In addition, 9 internal standards were used to
265 convert peak heights into good estimates of absolute (micromolar) concentrations for a
266 range of biogenic amines typically detected in biofluids and tissues (shown in
267 supplement 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 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 an excellent retention and separation of metabolite classes (biogenic amines, cationic compounds) with narrow peak widths of 5–20s.

Soil DNA extraction and sequencing of bacteria and fungi

DNA from soil subsamples (0.5 g) was extracted using a FastDNA SPIN kit (MP Biomedicals, Irvine, CA, USA) following the manufacturer's protocols. 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 (from 55 781 to 85 255) and 67 351 (from 43 939 to 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 fertilizer 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 fertilizer 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 fertilizer treatments, and the Shapiro-Wilk test was used to check for normality; the top- and sub-soils 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 (Afgan et al. 2018). 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 fertilization 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 fertilization 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 ($p < 0.05$) (supplement materials). In contrast to what was observed for the total and DOC content, the dissolved SOM content extracted by 3:3:2 (v/v/v) acetonitrile-isopropanol-water was generally smaller in the subsoil than that 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 greater concentrations in the topsoil compared to those in the subsoil (n = 12); the second group had greater concentrations in the subsoil than those in the topsoil (n = 59). The top 50 most significant known metabolites identified by ANOVA are presented in Fig. 1.

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

The curated complex lipid analysis identified 2 944 individual compounds, of which 144 were known (supplement materials). Among these previously identified

compounds, 90 were selected due to their high concentrations. The selected compounds were clustered into three groups: (1) compounds that appeared at greater concentrations in the topsoil than that in the subsoil, showing greater concentrations under the No-M than the Low-M and High-M treatments (n = 35); (2) compounds that appeared at greater concentrations in the topsoil than that in the subsoil, showing the greatest concentrations under the greater manure application (n = 24); and (3) compounds with greater concentrations in the subsoil than that in the topsoil (n = 31). The top 50 most significant known metabolites identified by ANOVA are presented in Fig. 2.

PCA analysis of soil properties and soil metabolomics

We observed a significant difference between the properties of topsoil and subsoil of the Low-M treatment, and large differences between No-M and High-M treatments. 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).

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, those detected by GC-TOFMS/MS were inversely correlated to SOM, EON, and MB-C contents (Fig. 4). In addition, several compounds were detected by both GC-TOFMS/MS and UHPLC-QTOFMS/MS, such as tyrosine, glucose-1-phosphate, leucine, glutamine, and isoleucine, 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 fertilizer application

LEfSe analysis was performed to identify the microbial taxa that differed significantly between fertilization regimes (Fig. 5). The most significant enrichment indicators were identified in the High-M treatment, while the least was in the Low-M treatment. 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 mineral fertilizer application (No-M) significantly enriched Mucoromycota.

Microbial community succession is driven by metabolites

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 clarify further 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 fertilizer application on soil organic matter

Expectedly, 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). Organic C can be rapidly utilised by microorganisms, and the microbial necromass contributes greatly to SOC (soil organic C) sequestration, especially in soils supplemented with manure having enhanced microbial biomass (Cui et al. 2020; Ma et al. 2020a; Wang et al. 2021). Microbial necromass contributed to approximately half of the soil organic C in grassland and cropland soils, based on the evaluation of glucosamine and muramic acid from bacterial and fungal necromasses (Wang et al. 2021). Therefore, the increased microbial biomass after FYM application can stimulate the formation of SOM. Long-term high FYM application increased the EON content in the subsoil but not in the topsoil, which is in direct contrast to the effect of chemical fertilizers. 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 smaller 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 fertilizer application on soil metabolites detected by GC-TOFMS/MS

Besides the basic chemical and physical soil characteristics, metabolic profiles are an important indicator of soil quality and ecosystem function (Withers et al. 2020). Metabolites can be sensitive to changes in the soil environment condition, 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. 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, and therefore the compounds only detectable in the negative mode were missed. Furthermore, some compounds, such as glycine betaine, are not amenable to derivatization, 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, greater 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. 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 fertilizer 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 mineral fertilizer application (No-M) than under low and high manure application (n = 35). Their smaller concentration in the subsoil could be due to the greater absorption by soil particles, as limited compounds in the topsoil leached to the subsoil. The group was comprised of almost all large molecular compounds, such as corticosterone, phenylacetamide, coniferylaldehyde, quinolone, nicotine, and hexadecylamine, which might have derived as secondary metabolites from soil microorganisms after they utilised the nutrients from chemical fertilizers. The long-term use of chemical fertilizer might stimulate microorganisms to synthesize those compounds and assimilate the inorganic nutrients to adapt to the environmental changes caused by mineral fertilizer application. The second group of compounds had the greatest 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 greatest concentration in the subsoil, either because they leached into the subsoil because of a smaller adsorption ability, or because they were derived from special microorganisms in the subsoil adapted to oxygen-deficient conditions (Ma et al. 2020a).

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

464 *Correlations between soil metabolism and the bacterial community*

465 Dissolved organic C, especially low molecular-weight compounds, including root
466 exudates, could be utilised directly as C sources by soil microbes (Swenson et al. 2015).
467 Therefore, soil metabolomics could improve our understanding of the coupling between
468 organic/inorganic compounds and microbial communities in the soil (Johns et al. 2017).
469 In this study, the most correlated factor for both bacterial and fungal community
470 succession was MTA, followed by N-epsilon-Acetyllysine, gamma-Glutamylleucine,
471 Histidine, and 3-Indolepropionic acid for the bacterial community (Fig. 6A); and 2'-O-
472 Methyladenosine, 1,4-Cyclohexanedione, Isobutyryl-L-carnitine, and Corticosterone
473 for the fungal community (Fig. 6B). MTA is a naturally occurring sulfur-containing
474 nucleoside, indicating that S metabolism is important for the formation of microbial
475 communities. Recently, due to a considerable decrease in sulfur dioxide emissions
476 following strict air-quality regulations, the application of fertilizers with a limited S
477 content, and a reduced S return via farmyard manure, S might be a limiting element for
478 microbial growth (Piotrowska-Długosz et al., 2017). N-epsilon-Acetyllysine is a

derivative of the amino acid lysine. 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 the microbial community (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 were the main utilizers of plant root exudates (Haichar et al. 2008). They responded positively to low molecular-weight substances (Goldfarb et al. 2011). However, Bacteroidetes is not a dominant bacterial phylum in bulk soil and

500 is found in high abundance in the rhizosphere (Aleklett et al. 2015). Therefore, it was
501 not the predominant bacterial phylum in the tested bulk soil. In the case of metabolites,
502 gamma-Glutamylleucine had the most links (20) with bacterial taxa.

503 Unlike the bacterial network, the fungal network was simpler, with lower nodes
504 and lower total degrees (Fig. 7B and Table S2). Previous studies have demonstrated that
505 fungi tend to decompose recalcitrant SOC, such as lignin and cellulose, and bacteria
506 then utilize the fungal-derived products (de Boer et al. 2005). Among the fungal taxa in
507 the network, the species *Aspergillus caesiellus* and *Thermomyces lanuginosus* had the
508 most correlations (8) with metabolites. In addition, MTA was found with the most
509 degrees with the fungal taxa. Soil microbial community structures can be achieved by
510 high-throughput sequencing. However, the actual microbial functions, such as their
511 metabolism, are difficult to obtain with soil metagenome or amplicon sequencing
512 (Jansson and Hofmockel, 2018).

513 The soil metabolome was mainly formed of organic acids, sugars, and sugar
514 derivatives, which were widely negatively correlated with bacterial alpha-diversity.
515 Compared to sugars, organic acids accounted more for bacterial community
516 compositions at high taxonomic ranks, while this was reversed at the species and genus
517 levels. Keystone species in co-occurrence network, such as *Microvirga*, *Bryobacter*,
518 and *Bradyrhizobium* were significantly correlated with organic acids and sugars (Liu et
519 al. 2020). We anticipate that these substrate-genome linkages could be further evaluated
520 and refined using other approaches. Stable isotope probing coupled with labelled DNA

sequencing (Orsi et al. 2016; Pepe-Ranney et al. 2016) and integrated NanoSIMS and FISH imaging (Woebken et al. 2015; Fike et al. 2008) may be used to examine the spatial localization 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. The combination of 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 linear and positively 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. The combination of 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 which more accurately reflect the composition of soil compounds and its turnover is needed.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (32102488, 32172674), 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.

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

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

572 Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ,
573 Fierer N, Knight R (2011) Global patterns of 16S rRNA diversity at a depth of millions
574 of sequences per sample. P Natl Acad Sci USA 108: 4516–4522.

575 Chang EH, Chung RS, Tsai YH (2010) Effect of different application rates of organic
576 fertilizer on soil enzyme activity and microbial population. Soil Sci Plant Nutr 53: 132–
577 140.

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

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

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

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

590 Fiehn O, Wohlgemuth G, Scholz M, Kind T, Lee DY, Lu Y, Moon S, Nikolau B (2008)
591 Quality control for plant metabolomics: reporting MSI-compliant studies. Plant J 53:
592 691-704.

593 Fike DA, Gammon CL, Ziebis W, Orphan VJ (2008) Micron-scale mapping of sulfur
594 cycling across the oxycline of a cyanobacterial mat: a paired nanoSIMS and CARD-
595 FISH approach. ISME J 2: 749-759.

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

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

601 Gupta L, Ahmed S, Jain A, Misra R (2018) Emerging role of metabolomics in

rheumatology. Int J Rheum Dis 21: 1468–1477.

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

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

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

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

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

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

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

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

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

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.

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

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

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

648 Liang C, Amelung W, Lehmann J, Matthias K (2019) Quantitative assessment of
649 microbial necromass contribution to soil organic matter. Global Change Biol 25: 3578-
650 3590

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

654 Liu K, Ding X, Wang J (2020) Soil metabolome correlates with bacterial diversity and
655 co-occurrence patterns in root-associated soils on the Tibetan Plateau. Sci Total Environ
656 735: 139572.

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

660 Liu S, Wang J, Pu S, Blagodatskaya E, Kuzyakov Y, Razavi BS (2020) Impact of
661 manure on soil biochemical properties: A global synthesis. Sci Total Environ 745:
662 141003.

663 Ma Q, Kuzyakov Y, Pan W, Tang S, Chadwick DR, Wen Y, Hill PW, Macdonald A, Ge
664 T, Si L, Wu L, Jones DL (2021a) Substrate control of sulphur utilisation and microbial
665 stoichiometry in soil: Results of 13C, 15N, 14C, and 35S quad labelling. ISME J 15
666 (11): 3148-3158.

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

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

673 Ma Q, Wen Y, Wang D, Sun X, Hill PW, Macdonald A, Chadwick DR, Wu L, Jones DL
674 (2020b) Farmyard manure applications stimulate soil carbon and nitrogen cycling by
675 boosting microbial biomass rather than changing its community composition. Soil Biol
676 Biochem 144: 107760.

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

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

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

686 Matyash V, Liebisch G, Kurzchalia TV, Shevchenko A, Schwudke D (2008) Lipid
687 extraction by methyl-tert-butyl ether for high-throughput lipidomics. J Lipid Res 49:
688 1137–1146.

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

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

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

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

700 Orsi WD, Smith JM, Liu S, Liu Z, Sakamoto CM, Wilken S, Poirier C, Richards TA,
701 Keeling PJ, Worden AZ (2016) Diverse, uncultivated bacteria and archaea underlying
702 the cycling of dissolved protein in the ocean. ISME J 10 (9): 2158-2173.

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

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

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

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

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

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

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

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

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

727 Trauger SA, Kalisak E, Kalisiak J, Morita H, Weinberg MV, Menon AL, Poole FL 2nd,
728 Adams MW, Siuzdak G (2008) Correlating the transcriptome, proteome, and
729 metabolome in the environmental adaptation of a hyperthermophile. J Proteome Res 7:
730 1027–1035.

731 Tsugawa H, Cajka T, Kind T, Ma Y, Higgins B, Ikeda K, Kanazawa M, VanderGheynst
732 J, Fiehn O, Arita M (2015) MS-DIAL: data-independent MS/MS deconvolution for
733 comprehensive metabolome analysis. Nat Methods 12: 523–526.

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

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

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

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

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

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

748 Woebken D, Burow LC, Behnam F, Mayali X, Schintlmeister A, Fleming ED, Prufert-
749 Bebout L, Singer SW, López Cortés A, Hoehler TM, Pett-Ridge J, Spormann AM,
750 Wagner M, Weber PK, Bebout BM (2015) Revisiting N_2 fixation in Guerrero Negro

751 intertidal microbial mats with a functional single-cell approach. ISME J 9: 485–496.

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

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

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

759 Yan Z, Chen S, Dari B, Sihi D, Chen Q (2018) Phosphorus transformation response to
760 soil properties changes induced by manure application in a calcareous soil. Geoderma
761 322: 163–171.

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

765 **Figure captions**

766 **Fig. 1.** Heatmap of the 50 most significant known metabolites (detected by GC-
767 TOFMS/MS) identified by ANOVA. Metabolites were clustered by Pearson correlation.
768 The color of squares linking metabolites to samples ranges from blue to red, indicating
769 the number of standard deviations from the mean. No-M, mineral fertilizers without
770 application of manure; Low-M, medium application rate of manure with mineral
771 fertilizers; High-M, high application rate of manure with mineral fertilizers. 4-amino
772 acid: 4-amino butyric acid; 4-hydro acid: 4-hydroxybenzoic acid; N-acety.: N-
773 acetylmannosamine; UDP-N-acety.: UDP-N-acetylglucosamine; gly. alf. phos.:
774 glycerol-alpha-phosphate; glucose-1-phos: glucose-1-phosphate; beta-mann.: beta-
775 mannosylglycerate.

776 **Fig. 2.** Heatmap of the 50 most significant known metabolites (detected by UHPLC-
777 QTOFMS/MS) identified by ANOVA. Metabolites were clustered by Pearson

778 correlation. The color of squares linking metabolites to samples ranges from blue to red,
779 indicating the number of standard deviations from the mean. No-M, mineral fertilizers
780 without application of manure; Low-M, medium application rate of manure with
781 mineral fertilizers; High-M, high application rate of manure with mineral fertilizers.
782 Butyl: Butylisopropylamine; N-N,N-dipro: N-(4-Piperidinyl)-N,N-dipropylamine; 4-
783 Hydro: 4-Hydroxy-1-(2-hydroxyethyl)-2,2,6,6-tetramethylpiperidine; 3-indol. Acid: 3-
784 Indoleacetic acid; Arach. Dopam.: Arachidonyl dopamine; N,N-Diethyl: N,N-Diethyl-
785 2-aminoethanol; Indole-3-carbox.: Indole-3-carboxaldehyde; Guanid. acid: 4-
786 Guanidinobutyric acid; Isobu. Carni.: Isobutyryl-L-carnitine; 1,1-Dimet.: 1,1-
787 Dimethyl-4-phenylpiperazinium; 4-amino. Acid: 4-Aminobenzoic acid; 5-Methy.: 5'-
788 Methylthioadenosine; Glycer.: Glycerophosphocholine; Atrazine-desis.: Atrazine-
789 desisopropyl-2-hydroxy; 8-Oxo-2-deoxy.: 8-Oxo-2-deoxyadenosine; N-epsilon-Acety.:
790 N-epsilon-Acetyllysine; gamma-Gluta.: gamma-Glutamylleucine.

791 **Fig. 3.** Principal component analysis (PCA) of soil carbon and nitrogen content
792 detected by traditional methods (A), and metabolites detected by GC-TOFMS/MS (B)
793 and UHPLC-QTOFMS/MS (C) under long-term (1964–2018) manure and mineral
794 fertilizer applications. Prior to analysis, the data were log10 transformed. No-M,
795 mineral fertilizers without manure application; Low-M, medium application rate of
796 manure with mineral fertilizers; High-M, high application rate of manure with mineral
797 fertilizers; T, topsoil; S, subsoil.

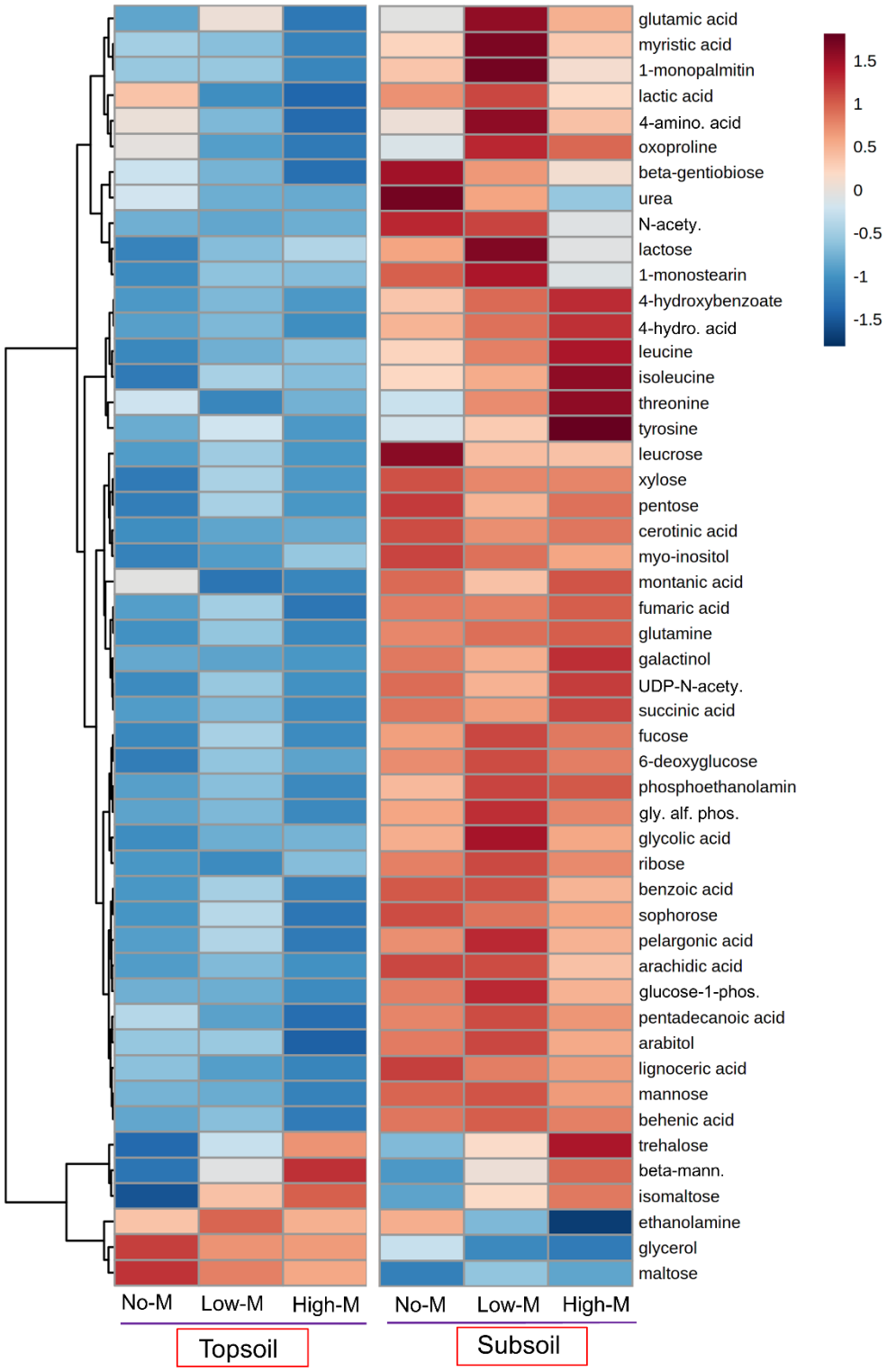
798 **Fig. 4.** Linear correlations of metabolites detected by GC-TOFMS/MS (A) and

799 UHPLC-QTOFMS/MS (B) with soil carbon and nitrogen content detected by
800 traditional methods under long-term (1964–2018) manure and mineral fertilizer
801 applications. DOC, dissolved organic carbon; SOM, soil organic matter; EON,
802 extractable organic N; MB-C, microbial biomass carbon; MB-N, microbial biomass
803 nitrogen.

804 **Fig. 5.** The response of bacterial (A) and fungal (B) communities at phylum to genus
805 levels to long-term organic and inorganic fertilizer application based on a linear
806 discriminant effect size analysis. Only taxa meeting a linear discriminant analysis
807 significance threshold of LDA > 3 are shown and color-coded. The six rings of the
808 cladogram indicate the domain (d), phylum (p), class (c), order (o), family (f), and genus
809 (g), from inside to outside.

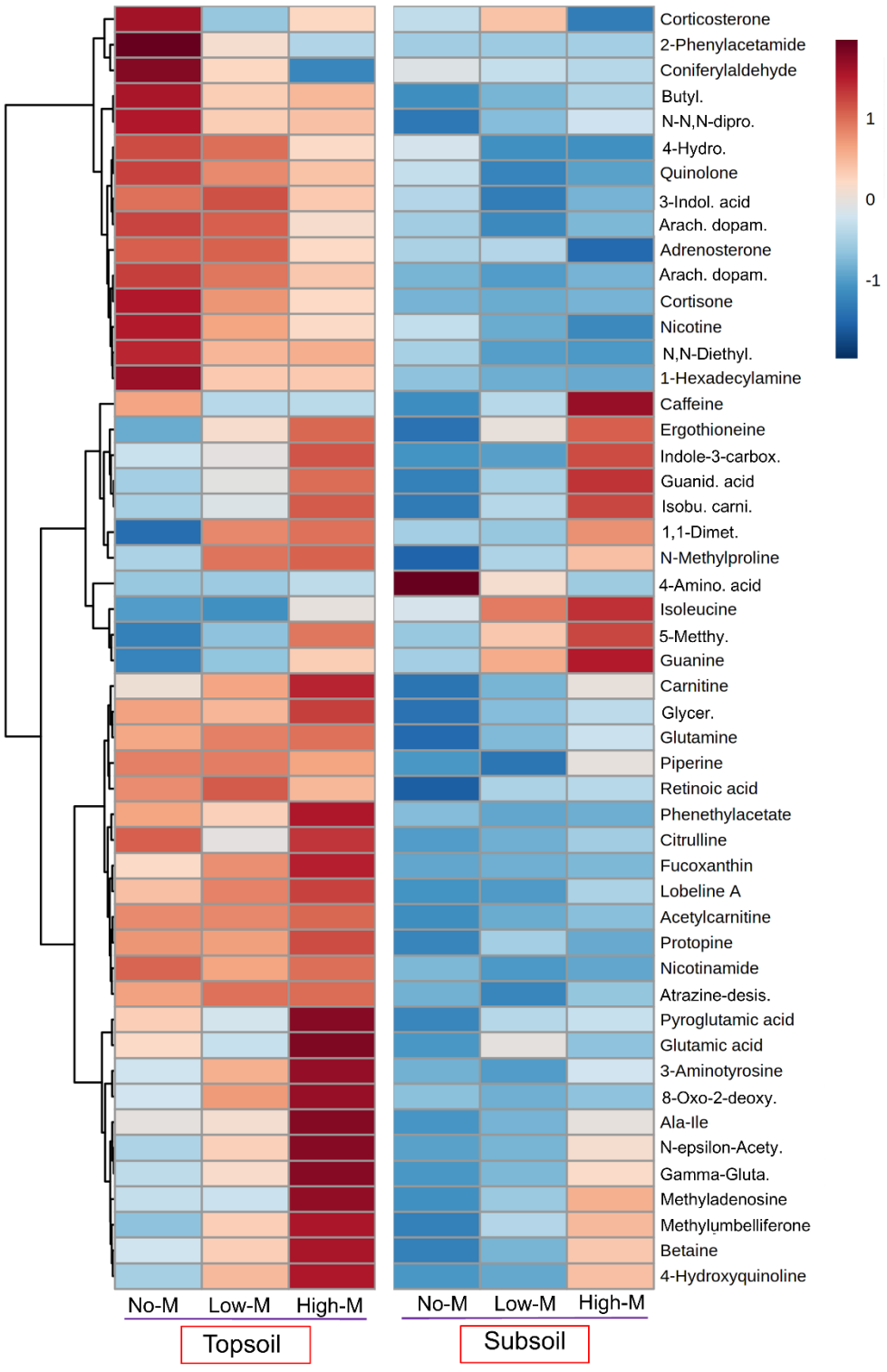
810 **Fig. 6.** Random forest analysis to determine factors affecting bacterial (A) and fungal
811 (B) community succession. The metabolites detected by UHPLC-QTOFMS/MS was
812 used in this analysis.

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



819

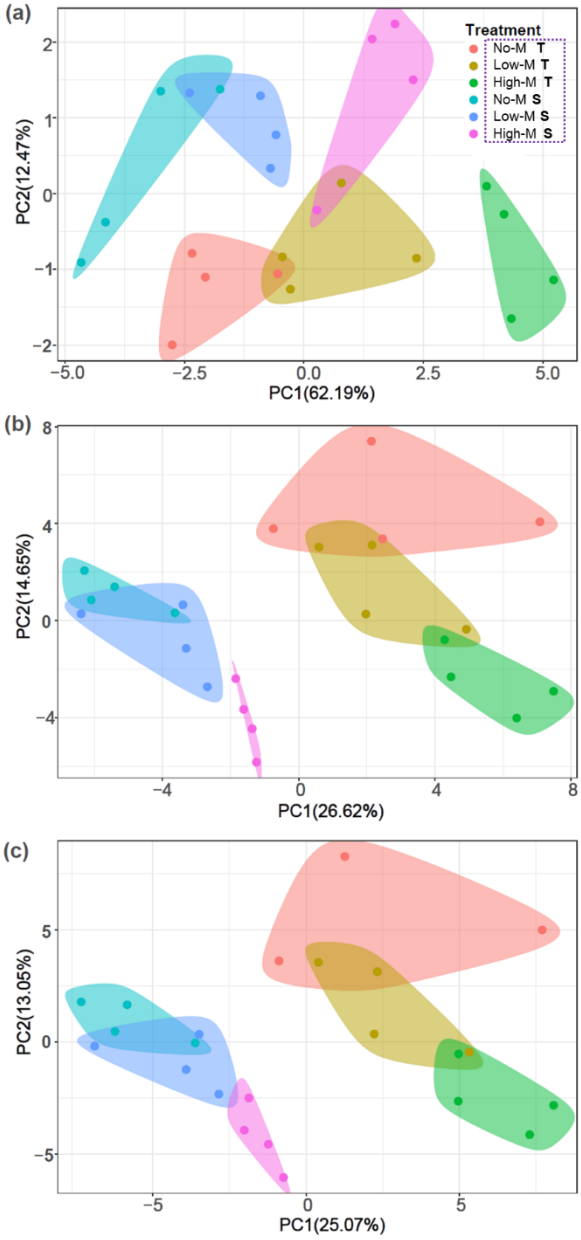
820 Figure 1



821

822 Figure 2

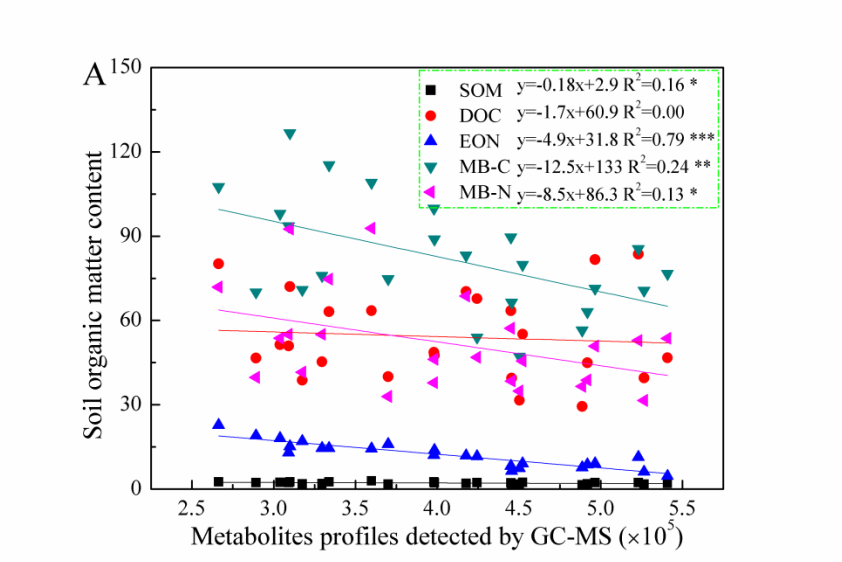
823



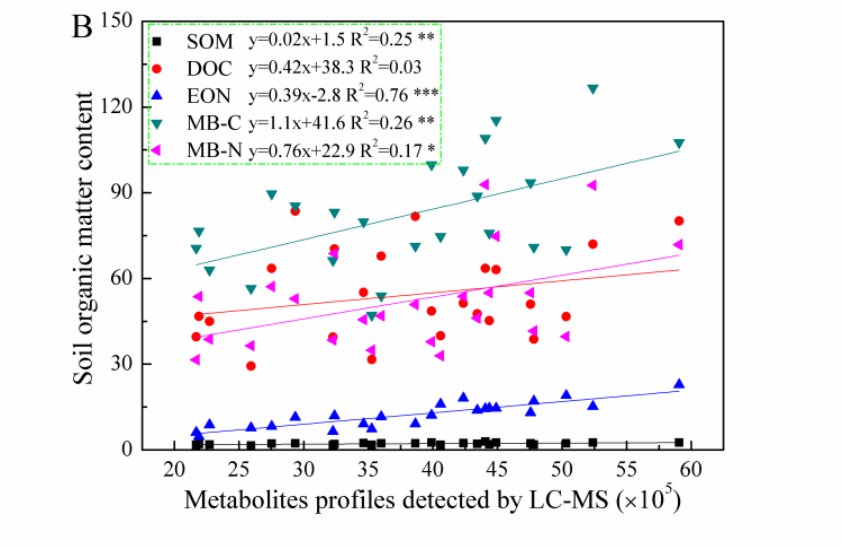
824

825 Figure 3

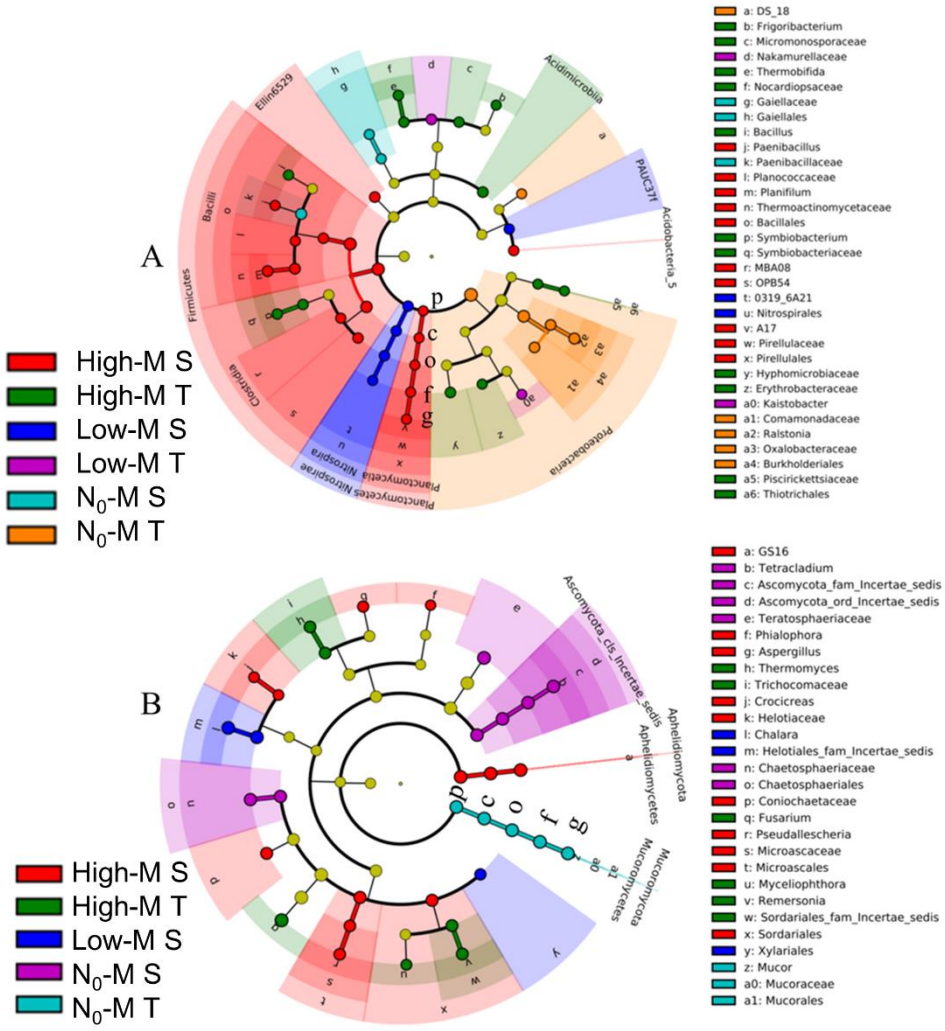
826



827



828 Figure 4



829

830 Figure 5

831

832

833

834

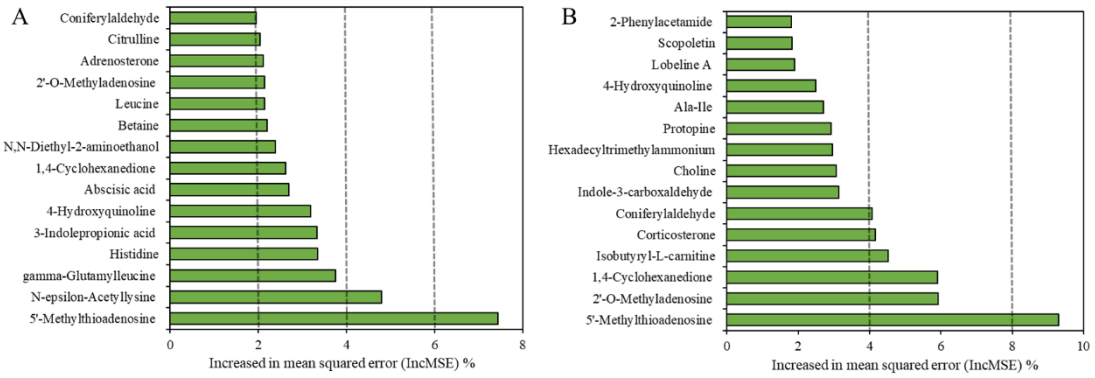
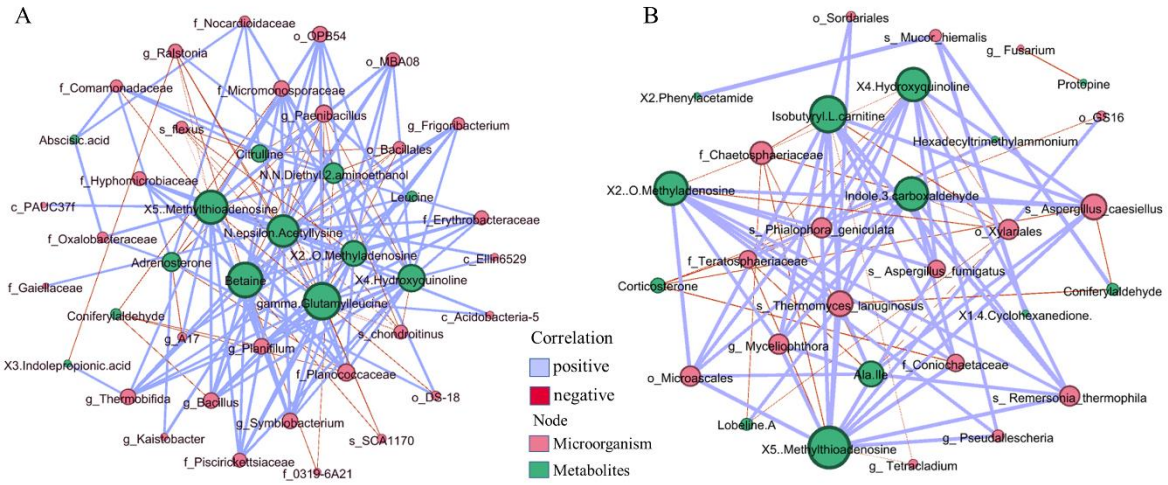

A
Coniferylaldehyde
Citrulline
Adrenosterone
2'-O-Methyladenosine
Leucine
Betaine
N,N-Diethyl-2-aminoethanol
1,4-Cyclohexanedione
Abscisic acid
4-Hydroxyquinoline
3-Indolepropionic acid
Histidine
gamma-Glutamylleucine
N-epsilon-Acetyllysine
5'-Methylthioadenosine
0
2
4
6
8
Increased in mean squared error (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



845

846 Figure 7

847

Supporting material

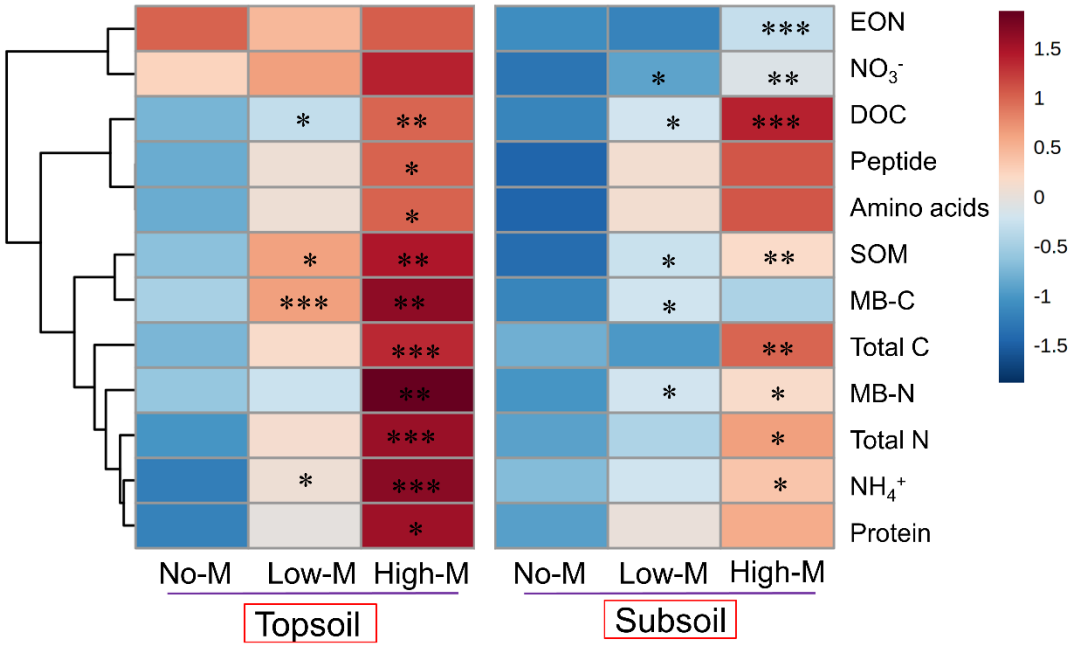


Figure S1

Fig. S1. Heatmap based on hierarchical clustering of the effects of long-term (1964–2018) manure and mineral fertilizer applications on total and dissolved contents of C and N in the topsoil and subsoil. Contents that significantly decreased are displayed in blue, whereas elements that significantly increased are displayed in red. The brightness of each colour corresponds to the magnitude of the difference compared with average values. Clustering of compounds according to Pearson correlation coefficients is depicted by the dendrogram on the left. Differences between treatments (compared with No-M) in the topsoil and subsoil were separately analysed using *t*-tests. DOC, dissolved organic carbon; SOM, soil organic matter; EON, extractable organic N; MB-C, microbial biomass carbon; MB-N, microbial biomass nitrogen; No-M, mineral

fertilizers only; Low-M, medium application rate of manure with mineral fertilizers; High-M, high application rate of manure with mineral fertilizers. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$.

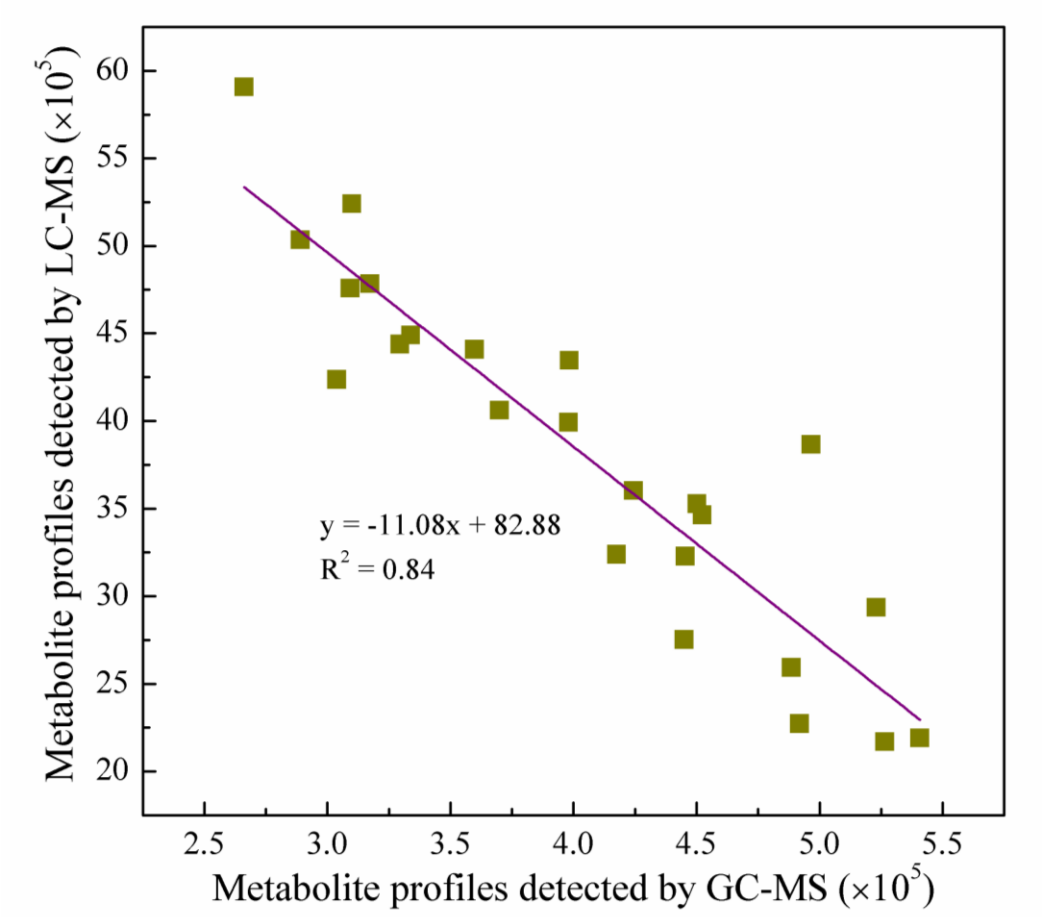


Fig. S2. Linear regression of metabolites detected by GC-TOFMS/MS and UHPLC-QTOFMS/MS under long-term (1964–2018) manure and mineral fertilizer applications.

869 **Table S1** Degree of bacterial taxa and metabolites in the co-occurrence network.

Node	Label	Degree
Bacteria	f_Planococcaceae	8
Bacteria	g_Paenibacillus	8
Bacteria	f_Micromonosporaceae	7
Bacteria	g_Bacillus	7
Bacteria	g_Planifilum	7
Bacteria	g_Symbiobacterium	7
Bacteria	g_Thermobifida	7
Bacteria	o_OPB54	7
Bacteria	f_Erythrobacteraceae	6
Bacteria	f_Hyphomicrobiaceae	6
Bacteria	f_Piscirickettsiaceae	6
Bacteria	g_Ralstonia	6
Bacteria	o_MBA08	6
Bacteria	s_chondroitinus	6
Bacteria	f_Comamonadaceae	5
Bacteria	g_Frigoribacterium	5
Bacteria	o_Bacillales	5
Bacteria	s_Flexus	5
Bacteria	f_Nocardioidaceae	4
Bacteria	f_Oxalobacteraceae	4
Bacteria	g_A17	3
Bacteria	o_DS-18	3
Bacteria	s_SCA1170	3
Bacteria	c_Acidobacteria-5	2
Bacteria	c_Ellin6529	2
Bacteria	f_0319-6A21	2
Bacteria	g_Kaistobacter	2
Bacteria	c_PAUC37f	1
Bacteria	f_Gaiellaceae	1
Metabolites	gamma-Glutamylleucine	20
Metabolites	Betaine	19
Metabolites	5'-Methylthioadenosine	18
Metabolites	N-epsilon-Acetyllysine	17
Metabolites	4-Hydroxyquinoline	14
Metabolites	2'-O-Methyladenosine	13
Metabolites	N,N-Diethyl-2-aminoethanol	10
Metabolites	Adrenosterone	9
Metabolites	Citrulline	8
Metabolites	Coniferylaldehyde	4
Metabolites	Leucine	4
Metabolites	Abscisic acid	3
Metabolites	3-Indolepropionic acid	2

870

871 **Table S2** Degree of fungal taxa and metabolites in the co-occurrence network.

0	Label	Degree
Fungi	s_Aspergillus_caesiellus	8
Fungi	s_Thermomyces_lanuginosus	8
Fungi	f_Chaetosphaeriaceae	7
Fungi	g_Myceliophthora	6
Fungi	o_Microascales	6
Fungi	o_Xylariales	6
Fungi	s_Phialophora_geniculata	6
Fungi	s_Remersonia_thermophila	6
Fungi	f_Coniochaetaceae	5
Fungi	f_Teratosphaeriaceae	5
Fungi	s_Aspergillus_fumigatus	5
Fungi	g_Pseudallescheria	3
Fungi	s_Mucor_hiemalis	3
Fungi	g_Tetracladium	2
Fungi	o_GS16	2
Fungi	o_Sordariales	2
Fungi	g_Fusarium	1
Metabolites	5'-Methylthioadenosine	14
Metabolites	Isobutyryl-L-carnitine	12
Metabolites	Indole-3-carboxaldehyde	11
Metabolites	2'-O-Methyladenosine	11
Metabolites	4-Hydroxyquinoline	11
Metabolites	Ala-Ile	8
Metabolites	Corticosterone	4
Metabolites	Coniferylaldehyde	3
Metabolites	Lobeline A	3
Metabolites	Hexadecyltrimethylammonium	1
Metabolites	Protopine	1
Metabolites	1,4-Cyclohexanedione	1
Metabolites	2-Phenylacetamide	1

872