Received: 12 September 2022 | Revised: 29 May 2023 | Accepted: 5 June 2023

DOI: 10.1111/ejss.13392

RESEARCH ARTICLE

European Journal of Soil Science WILEY

Assessing the impact of interfering organic matter on soil metaproteomic workflow

Matthias Waibel[1,2] | **Kevin McDonnell**[1] | **Maria Tuohy**[1] | **Sally Shirran**[3] | **Sylvia Synowsky**[3] | **Barry Thornton**[2] | **Eric Paterson**[2] | **Fiona Brennan**[4] | **Florence Abram**[1]

[1]School of Biological and Chemical Sciences, Ryan Institute, University of Galway, Galway, Ireland

[2]The James Hutton Institute, Aberdeen, UK

[3]Biomedical Sciences Research Centre, University of St Andrews, St Andrews, UK

[4]Teagasc, Johnstown Castle, Wexford, Ireland

Correspondence
Florence Abram, School of Biological and Chemical Sciences, Ryan Institute, University of Galway, Galway H91 TK33, Ireland.
Email: florence.abram@nuigalway.ie

Funding information
Irish Research Council for Science, Engineering and Technology; National University of Ireland, Galway; University of Galway College of Science; Irish Research Council, Grant/Award Number: GOIPG/2016/1215; Rural and Environment Science and Analytical Services Division of the Scottish Government

Abstract

Soil organic matter (SOM) is biologically, chemically, and physically complex. As a major store of nutrients within the soil, it plays an important role in nutrient provision to plants. An enhanced understanding of SOM utilisation processes could underpin better fertiliser management for plant growth, with reduced environmental losses. Metaproteomics can allow the characterisation of protein profiles and could help gain insights into SOM microbial decomposition mechanisms. Here, we applied three different extraction methods to two soil types to recover SOM with different characteristics. Specifically, water-extractable organic matter, mineral-associated organic matter and protein-bound organic matter were targeted with the aim to investigate the metaproteome enriched in those extractions. As a proof-of-concept, replicated extracts from one soil were further analysed for peptide identification using liquid chromatography followed by tandem mass spectrometry. We employed a framework for mining mass spectra for both peptide assignment and fragmentation pattern characterisation. Different extracts were found to exhibit contrasting total protein and humic substance content for the two soils investigated. Overall, water extracts displayed the lowest humic substance content (in both soils) and the highest number of peptide identifications (in the soil investigated) with the most frequent peptide hits associated with diverse substrate/ligand binding proteins of Proteobacteria and derived taxa. Our framework also highlighted a strong peptidic signal in unassigned and unmatched spectra, information that is currently not captured by the pipelines employed in this study. Taken together, this work points to specific areas for optimisation in chromatography and mass spectrometry to adequately characterise SOM-associated metaproteomes.

KEYWORDS
chemical fractionation, grassland soil, metaproteomics, soil organic matter

1 | INTRODUCTION

Soil organic matter (SOM) constitutes a pivotal element of soils and their genesis, also representing the largest store of nutrients supporting plant growth (Bot & Benites, 2007; Montgomery, 2007). Against the backdrop of tighter environmental and economic constraints, understanding SOM dynamics could allow for better matching of fertiliser applications with plant requirements (Ashman & Puri, 2013; Paul, 2014). SOM is hard to characterise, however, due to the dynamic nature of its stabilisation and decay processes, soil matrix micro-scale spatial arrangement, and aqueous supramolecular aggregation of bio-organic molecules (Baveye & Wander, 2019; Duval et al., 2005; Kleber et al., 2007; Lehmann & Kleber, 2015; Piccolo, 2001; Pot et al., 2022; Schlüter et al., 2022; Sutton & Sposito, 2005). In addition, bio-organic molecules can undergo attenuation by substitutions of functional groups, resulting in complex non-repetitive (super-)mixtures (Brown et al., 2016; Hertkorn et al., 2007; Stenson, 2008). In principle, SOM derives from dead organic matter (OM) of partially or strongly decomposed plant material and from microbial products and necromass (Angst et al., 2021; Baveye & Wander, 2019; Cotrufo et al., 2013; Kuzyakov, 2010; Liang et al., 2019; Paul, 2016; Sutton & Sposito, 2005; Tan, 2014). Nutrient release is mostly dependent on biological agents and their degree of control on degradation/mineralisation, which is in turn dependent on chemical, physical and ecological persistence factors of SOM, including physical separation from decomposers, occlusion, biochemical stabilisation and energy limitation (Allison et al., 2010; Cotrufo et al., 2013; Dungait et al., 2012; Fontaine & Barot, 2005; Kemmitt et al., 2008; Kuzyakov et al., 2009; Lehmann & Kleber, 2015; Paterson, 2009; Paul, 2016; Rillig et al., 2007). Confluence of microbes in narrow crevices, pores, or hotspots such as biofilms, and their joint effort to produce enzymes may lead to SOM degradation (Ekschmitt et al., 2005; Spohn & Kuzyakov, 2014). Such transformations could be driven by specific enzymatic activities related to the extracellular breakdown of celluloses, hemicelluloses, lignins, proteins, and amino-sugars, but a wider variety of proteins and mechanisms are likely to be involved, especially with regard to stable SOM pools (Burns et al., 2013; Burns & Dick, 2002; Wallenstein & Burns, 2011). Indeed, the molecular composition of SOM moieties likely dictates the suite of proteins involved in their transformations.

Much research effort has been driven toward developing extractants to recover specific SOM fractions (Bremner & Lees, 1949; Helfrich et al., 2007; Nichols & Wright, 2006; Tan, 2014; Von Lützow et al., 2007). Different extractants have been found to systematically bias soil protein extraction (Bastida et al., 2018; Greenfield et al., 2018; Masciandaro et al., 2008). Particulate organic matter (POM) and mineral-associated organic matter (MAOM) fractions have been related to turnover times of conceptual soil organic C pools (Just et al., 2021; Peralta et al., 2022; Zimmermann et al., 2007). To date, knowledge gaps exist with regard to the relative importance of protein functions and taxonomy pertaining to the nature of extracted SOM fractions. SOM extraction using sodium pyrophosphate $Na_4P_2O_7$ (NaPPi) may enrich for MAOM (Von Lützow et al., 2007). When buffered at neutrality, NaPPi acts as a mild extractant that was previously reported not to lyse cells and to preserve enzymatic activity (Nannipieri, 2006). Water-extractable organic matter (WEOM) has been widely used for the determination of active/labile SOM and microbial/enzymatic activity (Grosso et al., 2014; Kaiser et al., 2010; Rennert et al., 2007). Furthermore, water extraction could in principle also encompass the recovery of extracellular enzymes (Schulze et al., 2005). Protein extraction buffers designed by Moore et al. (2012) employ denaturants and chaotropic salts to solubilise proteins by interfering with protein–protein/−mineral/-OM binding.

Metaproteomics is the study of the protein complement expressed by microbial communities and encompasses taxonomic and functional protein assignments and as such links microbes to function (Abram, 2015; Keller & Hettich, 2009; Verberkmoes et al., 2009). The potential of metaproteomics for the characterisation of taxonomy and functions recovered from contrasting labile and mineral-associated SOM fractions was previously shown (Schulze et al., 2005), but more efforts are required to enable extracellular metaproteomics as this could potentially capture early SOM degradation steps

Highlights

- Different protein extractions led to variations in protein yields and humic substance content.
- Water extraction resulted in the lowest humic substance contamination and the highest number of peptide identifications.
- Unassigned and unmatched spectra from soil samples harboured strong peptide signature in mass spectrometry, despite lack of peptide matches.
- Existing proteomic pipelines require tailoring to soil samples.

(Bastida et al., 2018; Johnson-Rollings et al., 2014; Keiblinger et al., 2016). Technical challenges for soil metaproteomics typically include low protein yield (particularly for extracellular proteins), negative impact of co-extracted soil mineral and humic substances on protein yield and peptide matching, as well as the lack of candidate protein entries (genomic or protein) for data-dependent, uninterpreted mass spectrometry-based searches (Keller & Hettich, 2009; Nannipieri, 2006; Renella et al., 2014). Peptide matching in proteomics is a statistical process and the trade-off between sensitivity and specificity presents a conundrum (Cottrell, 2011). For example, a larger database including more protein sequences may not perform better as the chance of getting a false positive just by combinatorial variations alone increases (ibid.). In practice, the choice of database has been found to greatly impact on the number of successful matches (Muth, Kolmeder, et al., 2015; Tanca et al., 2013). Soil extracts with many co-ionised contaminants and adverse ion interactions have presented challenges for untargeted biomolecule detection in natural organic matter (NOM) research (Hockaday et al., 2009; Novotny et al., 2014), and those effects present further complications for metaproteomics (Arenella et al., 2014; Qian & Hettich, 2017).

The aim of our study was to characterise proteins recovered using three different extraction methods: (i) focusing on WEOM; (ii) targeting MAOM and (iii) aiming to solubilise proteins from multiple complexes (Moore et al., 2012). We did not employ dedicated cell lysis steps, such as bead beating, boiling or in-tip sonication, as an attempt to enrich for extracellular proteins, acknowledging however that cell lysis could not possibly be fully prevented (Barnard et al., 2013; Brown, 1976; Rojas et al., 2018). Two mineral grassland soils with high organic matter content were used in this study (Table S1). As a proof-of-concept, extracts from a brown earth soil with a locally characterised grassland plant community were further analysed, as such soil type is considered to be representative of European granitic soils (Stahr et al., 2016). Liquid chromatography followed by tandem mass spectrometry (LC–MS/MS) was conducted using a joint framework for peptide assignment and fragmentation pattern analysis. We set out to address the following research objectives:

1. To assess the impact of extraction method on total protein and humic substance contents as well as extract fragmentation profile.
2. To evaluate the effect of extraction method on peptide assignments (hits).

2 | MATERIALS AND METHODS

2.1 | Soil properties and extractions

Soils were sampled to 10 cm depth, sieved to 2 mm, flash frozen in liquid nitrogen and stored at −80°C. Proteins were extracted from 10 g fresh-weight soil in a randomised block design with triplicate technical replicates distributed into three blocks, each containing the three extractants, blank controls, and two soils (Table S2). Blank controls did not contain any soil but were only made up of extractants, which were taken through the entire extraction process. Extractants were added at a 3:1 ratio to 10 g fresh-weight soil in 50 mL polypropylene tubes. The extractants were (i) EDTA-type (Moore et al., 2012) containing 6 M urea, 2 M thiourea, 0.01 M Tris–HCl at pH 7.4, 10% (v/v) glycerol, 0.002 M tributyl phosphine, 2% (w/v) CHAPS, and 1 mM EDTA (autoclaved), (ii) water extractant made up of filter sterilised (0.2 μm) milli-Q grade water at pH 7.5 (no pH adjustment was required), and (iii) 0.1 M NaPPi solution (tetrasodium pyrophosphate, $Na_4P_2O_7$) modified from Bremner and Lees (1949) and made in 0.1 M Gomori-type buffer (8.5% K_2HPO4 and 91.5% KH_2PO4; Joseph & David, 2001) at pH 5.8 to a final mixture at pH 7.5, and filter sterilised (0.2 μm). All extractants were supplemented with protease inhibitor Pepstatin A (Sigma Aldrich) to 4.8 μM working concentration and protease inhibitor cocktail (Roche cOmplete) just before extraction. Extractants were warmed to 37°C, mixed with soil and treated in a sonication water bath (Decon FS100b) for 5 min to break up soil aggregates and release proteins from the soil matrix (Ogunseitan, 2006). From this point on, samples were processed on ice. Suspensions were centrifuged to separate soil particles and supernatants, filtered (0.45 μm PES, Sarstedt filter discs), and 20 mL loaded onto ultrafiltration concentrators (Sartorius Vivaspin 20, 3000 Da MWCO, PES). Retentates were concentrated further with smaller ultrafiltration concentrators (Sartorius Vivaspin 2). Extracts were reconstituted in molecular grade water (FisherSci) and precipitated with sodium deoxycholate (Na-Doc) and trichloroacetic acid (TCA) precipitation (Bensadoun & Weinstein, 1976). Pellets were resuspended in ice-cold 0.1 M NaOH and fine sediment was removed by centrifugation at 10,000 *g* for 15 min (Tan, 2014). Sample buffer was exchanged for 10 mM Tris–HCl pH 8.5 buffer (filter sterilised) and samples were concentrated in ultrafiltration concentrators (Sartorius Vivaspin 2). Final concentrates were sub-aliquoted prior to further analysis.

2.2 | Total protein and humic substance quantification

For the Lowry assay, aliquots of final extracts were initially diluted 100 times with 2% (w/v) Na_2CO_3 in 0.1 M NaOH (Lowry et al., 1951) and samples were diluted into the range of the assay standards. Lowry reactive substances were quantified using a microplate adaptation of the modified Lowry assay (Mccarthy & Tuohy, 2011). In addition, samples and blank controls were also prepared without copper, as "absorbance blinds", to partition the total Lowry absorbance into protein and humic substance-derived portions (Frolund et al., 1995; Redmile-Gordon et al., 2013). Lowry absorbance was measured at 680 nm and quantified against standard concentrations of bovine serum albumin between 0 to 200 μg mL^{-1}, which were prepared in 0.1 M NaOH and 2% Na_2CO_3.

2.3 | Gel-based, gradient reverse-phased nanoflow UHPLC nESI+ QqTOF for metaproteomics and chemical analysis

Based on total protein and humic substances quantification determined with the Lowry assay, concentrated extracts were normalised to 30 μg total protein content with 10 mM Tris–HCl buffer, incubated with LDS sample buffer (Invitrogen) and 50 mM DTT (Invitrogen), and loaded on 4–12% gradient Bis-tris gel (Invitrogen). Samples were run ca. 1 cm using MES-SDS as running buffer (Invitrogen). Gels were washed briefly with molecular grade water and fixed in 40% (v/v) methanol, 10% (v/v) acetic acid water solution (Joseph & David, 2001). Lanes were excised and gel chunks were cut into small pieces in 50% (v/v) methanol in water. Gel pieces were washed with 25 mM ammonium bicarbonate followed by reduction with 10 mM DTT in 25 mM ammonium bicarbonate. After reduction, the gel pieces were alkylated with 100 mM iodoacetamide in 25 mM ammonium bicarbonate. Gel pieces were washed twice with 25 mM ammonium bicarbonate before double trypsin digest with each 50 μL trypsin (4 ng μL^{-1}). The gel plugs were extracted twice using 50 μL 5% (v/v) formic acid and extracts were combined and dried by vacuum centrifugation (SpeedVac) prior to resuspension in 20 μL loading buffer (LP) phase (98% water/ 2% acetonitrile (ACN)/ 0.05% trifluoroacetic acid (v/v/v)). Four μL of extract were loaded on an Acclaim PepMap 100 C_{18} trap column (L = 20 mm) with nanoLC Ultra 2D plus loading pump and nanoLC as-2 autosampler (Eksigent). Loading on the trap column and washing for 10 min was done at a flow rate of 5 μL min^{-1} of LP phase. The trap column was then switched in line with the analytical column (Acclaim PepMap 100 C_{18}, d = 75 μm, L = 150 mm) and peptides were eluted at 300 nL min^{-1} flow rate with an elution gradient of increasing ACN by increasing proportion of phase B (2% water/ 98% ACN/ 0.1% formic acid (v/v/v)) and decrease of phase A (98% water/ 2% ACN/ 0.1% formic acid (v/v/v)) (Figure S1). The eluate was sprayed via nano capillary and electrospray ionisation source (ESI) in positive ion mode (Heating = 120°C, capillary voltage = +1500 V, cone voltage not set) into a QqTOF style tandem mass spectrometer setup (TripleTOF 5600+, AB Sciex Pte. Ltd., Foster City, CA, U.S.A.) and analysed in Information Dependent Acquisition (IDA) mode, performing 120 ms of MS scan and mass selection range between 5 to 1250 *m/z*, followed by 80 ms MS/MS analyses on the 20 most intense peaks seen by MS. Washes were performed between samples, using the elution profile shown in Figure S1. To address low-performance issues, wash cycles and MS recalibrations were increased, and all samples re-analysed on LC–MS/MS (Table S2).

2.4 | Data analysis

Instrument recorded data were subjected to signal processing (PeakView, ABSciex) and converted to Mascot peak list files (*.mgf)* for Mascot search (Matrix Science Inc., Boston, MA, U.S.A.). The files were analysed with Mascot Version 2.6.0, against the large NCBIprot (NCBI) database (187,857,634 sequences; NCBIprot_20190208.fasta) or Mascot Version 2.7.0.7 against smaller and non-redundant SwissProt (SWP) database (564,277 sequences; SwissProt_2021_01.fasta). Trypsin was set as the cleavage enzyme with a maximum of one missed cleavage, cysteine with carbamidomethyl as fixed modification, and methionine oxidation as a variable modification. Peptide mass tolerance was set at 20 ppm and MS/MS mass tolerance at ±0.05 Da. The raw, peak and result files (*.wiff. scan*, *.mgf*, *.mzIdentML*, *.csv)* are available at the ProteomeXchange Consortium (identifier PXD034783).

Peak lists (*.mgf*) were also searched with Metaproteome Analyser (MPA local) software (Heyer et al., 2019; Muth, Behne, et al., 2015) against SWP. For all pipelines, 1% and 5% false discovery rate (FDR) settings were used. Mascot search results and information on spectra (fragmentation events) were exported as *.csv* files and were classified into four categories: (i) assigned spectra correspond to accepted queries with peptide-spectrum-match (PSM), where peptide expectation value met the significance threshold for the FDR criterion; (ii) anchored spectra correspond to PSM not meeting the FDR threshold, but belonging to accepted protein-families (Koskinen et al., 2011); (iii) unassigned spectra were PSMs not

meeting the FDR threshold and not belonging to accepted protein families; and (iv) unmatched spectra were those without PSM. Assigned and anchored spectra were further classified into contaminants, and non-contaminants (quality controlled, QC), as well as microbial hits (MO). Eukaryote hits were subjected to BLAST search to assess similarity to contaminants. Fragment ion series m/z for each spectrum was extracted from Mascot report. Fragment ions from each of the samples were then analysed using modulo plots developed by McDonnell et al. (2023). MS/MS spectra were grouped into assigned, unassigned and unmatched categories for each extraction type as described earlier. The observed mass-to-charge ratios (m/z) of the fragment ions were then plotted against the m/z modulo 1. The distributions were compared to the expected distribution of peptide-like fragments (McDonnell et al., 2023). Data processing steps are outlined in Figure S2 and the corresponding code is available on Github (https://github.com/waibel-123/PSManalysis).

Phylogenetic and functional metaproteome of unique Mascot assigned MO at 1% FDR setting (Koskinen, 2021) was analysed with Unipept Metaproteome analysis tool (Mesuere et al., 2015), using Unipept desktop with UniProt TrEMBL 2020.01 as database (Gurdeep Singh et al., 2018). For visualisation of taxonomy and functional relationships, Chord plots were made using Chordomics package (McDonnell et al., 2020).

2.5 | Statistical methods

To test for significant effects ($p < 0.05$), ANOVA, two-way Scheirer–Ray–Hare test, or Kruskal–Wallis test were used respectively where model conditions were met (Dytham, 2011). Significant differences at $p < 0.05$ were assessed with Tukey HSD as post-hoc test or Wilcoxon rank-sum test in *R* v.4.05 (Team R, 2019). Distribution and density of spectral counts were shown in histograms overlaid with density plots using the ggplot2 package (Wickham et al., 2019), where bin width corresponds to windows of 2.5 min of the retention time.

3 | RESULTS

3.1 | Protein and humic substance content of extracts

Protein and humic substance content were determined on the processed extracts and then related back to the starting amount of soil (Figure 1). Extraction methods produced contrasting extracts in two respects: (i) extractant was found to be a significant factor determining total protein content, and (ii) extractant affected total humic substance content, which was significantly highest in NaPPi extracts. For both measures, extractant was overriding soil as a factor (Tables S3 and S4).

3.2 | LC–MS/MS spectra and untargeted chemical analysis

LC–MS/MS, total recorded spectral counts and their classification varied greatly with sample type in the Athenry soil analysed (Figure 2 and Figure S3). Increasing the number of LC wash cycles in addition to MS recalibration increased spectral counts for water samples but not for NaPPi and EDTA samples (data not shown). In all samples, assigned QC spectra tended to elute over the elution gradient in phase I and II, especially in water samples, while unmatched spectra concentrated in most hydrophobic conditions (Figure S4). It must be noted that blank extractant controls, which should theoretically contain very few peaks, displayed a high frequency of unmatched spectra. Overall, unassigned and unmatched spectra were the most frequently detected spectral type (Figure 2, Figure S3, Table S5). Low-quality unassigned and anchored spectra exhibited significantly reduced fragment numbers when compared to assigned spectra (Figures S5–S7). Furthermore, the differences in elution time of bi-modal unassigned peaks indicated compositional differences in spectra, which were not accessible by peptide-spectrum matching. To investigate what these compositional differences might be, we applied the methodology described by McDonnell et al. (2023). As most of the mass of molecules is attributable to their constituent nucleons (protons and neutrons), singly charged MS/MS ions can be expected to appear at integer values in m/z space. However, differences in binding energy mean that the average mass per nucleon differs between molecules depending on their chemical formula. Figure 3 shows the m/z values of the fragment ions plotted against the m/z modulo 1. Peptide fragments have a relatively consistent ratio of carbon, hydrogen, oxygen, nitrogen and sulphur due to being composed of amino acids. This means that peptide fragments appear approximately at integer multiples of 1.0005 Da in MS/MS spectra (McDonnell et al., 2023). This average nucleon mass is indicated by a red line with a slope of 0.0005 in Figure 3. Multiply charged peptide fragments will appear as parallel lines to this slope (McDonnell et al., 2023). As can be seen in Figure 3, almost all fragment ions from the water replicate 1 appear to come from peptide origins, including those classified as unassigned and unmatched. This was replicated for all

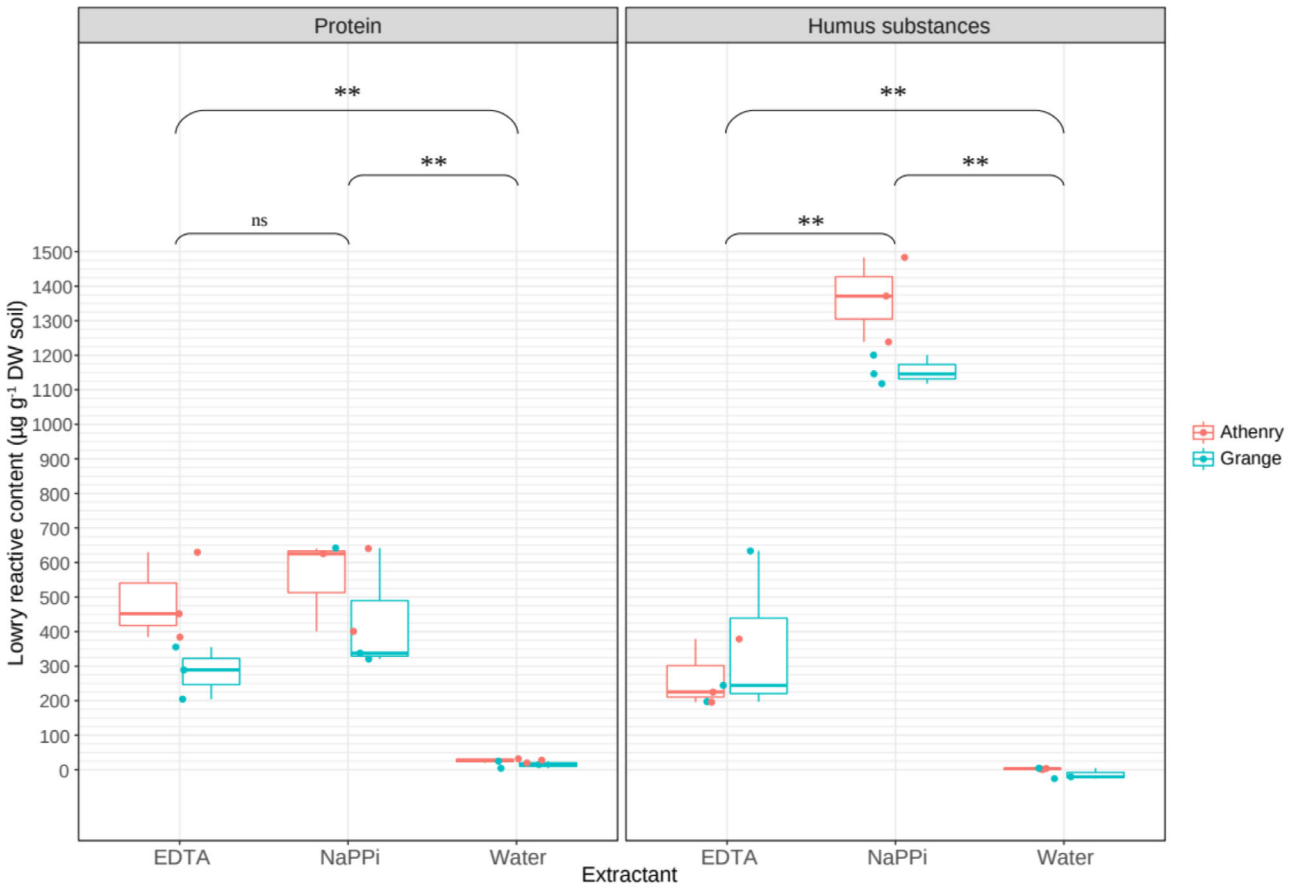

FIGURE 1 Protein and humic content of Athenry and Grange soil extracts. Yields are reported after processing and cleaning and are related to dry weight (DW) equivalent amount of soil. Following two-way Scheirer–Ray–Hare test result that extractant was the single significant factor, soil factor was bulked and Kruskal–Wallis test determined again that extractant was a significant factor for protein and humic content, respectively ($\chi^2 > 11$, $p < 0.005$). Differences in protein and humic substance content with extractant were assessed for significance with Wilcoxon rank-sum test adjusted for multiple testing at $p < 0.005$ significance level (**) or were not significant (ns).

other water replicate samples (Figure S8), EDTA replicate samples (Figure S10) and all NaPPi replicate samples with the exception of unmatched peaks from NaPPi replicated 1 (Figure S9). This indicates that all the corresponding MS spectra contain peptide-like information that is not currently captured. Interestingly, control samples exhibited different behaviour (Figures S8–S10). The distribution of fragment ions in unmatched and unassigned spectra from control samples appears on a steeper slope. This indicates that the fragments have a heavier average nucleon than peptides. This could be due to a higher ratio of hydrogen or a lower ratio of heavier elements such as oxygen when compared to peptides. Furthermore, the distribution of ions creates well-defined streaks in the plots indicating the ratio of constituent elements in this subset of fragments is consistent. Overall, this indicates that unmatched and unassigned spectra from control samples likely contain molecules other than peptides. Moreover, it is worth noting that the unmatched hydrophobic spectra also contain many fragments that do not appear to belong to either slope. These fragments are of unknown origin and are mainly present in the *m/z* range of 500–1700 but span the entire *m/z* modulo 1 axis.

3.3 | Metaproteomics

Metaprotein functional and taxonomic assignments of contrasting extraction methods from one soil type (Athenry soil) were determined using Mascot with NCBI database, which generated the most microbial PSMs (Table S6). Even though the number of total PSMs varied greatly between extract types (ibid.), peptide hits were most frequently assigned to the phylum Proteobacteria (Figures S11–S13; Table S7). Overall, water extracts yielded the most assigned QC peptide hits (Table S6). Most hits were associated with proteins of diverse substrate/ligand binding proteins of ABC transport systems

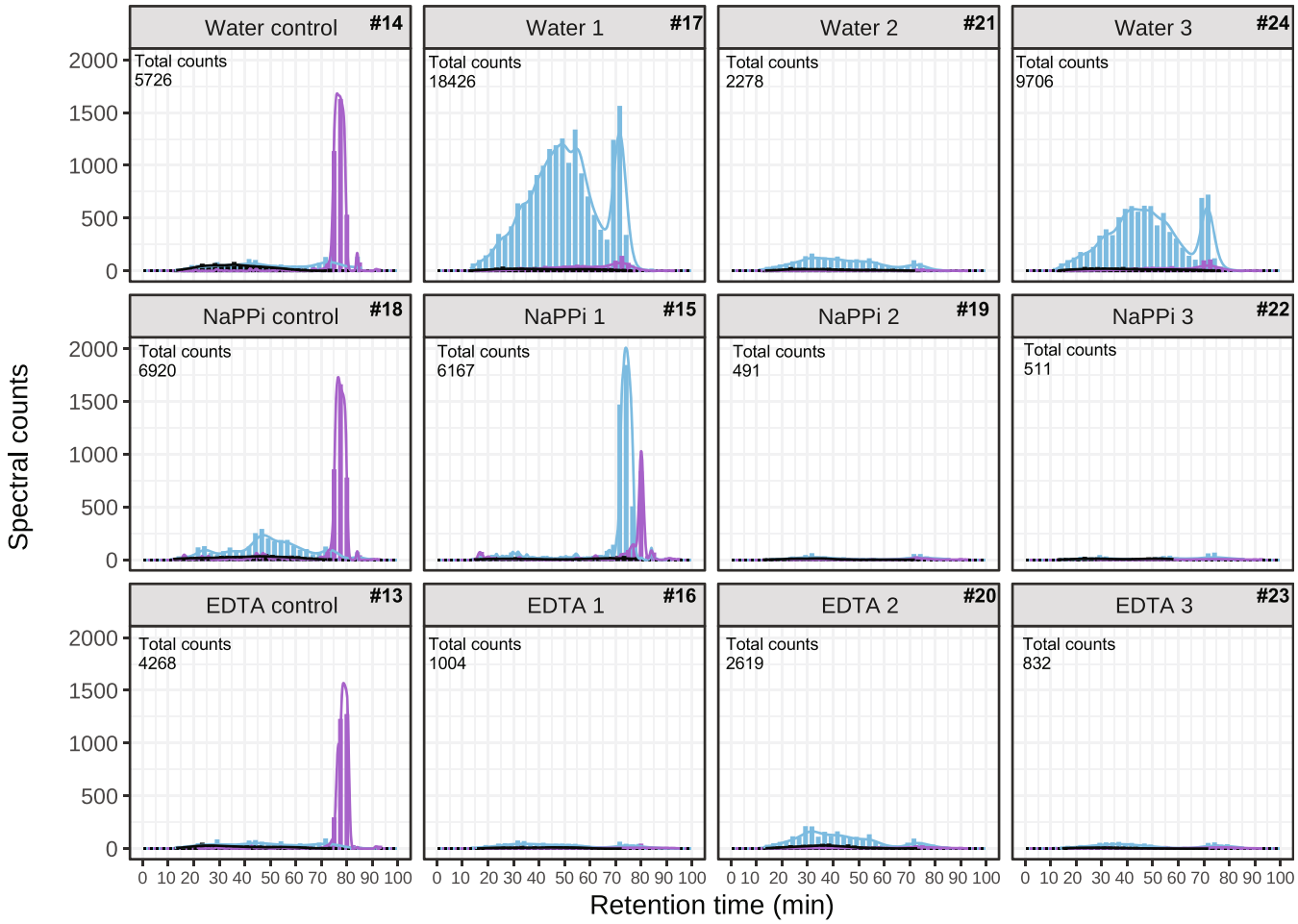

FIGURE 2 Spectral counts (parent ion fragmentation events) and colour-coded categories over retention time in LC–MS/MS. Light blue colour indicates unassigned spectra, purple indicates unmatched, and black indicates all other categories (assigned QC, assigned contaminants, anchored QC, and anchored contaminants). There were 3 replicates per extractant indicated with suffix 1, 2, 3. Blank extractant control replicates were bulked for LC–MS/MS. The order in which the samples were run is indicated by #, starting with #13. Spectral classifications are based on results from Mascot with NCBI database at 1% false discovery rate (FDR) setting. Spectral classifications (light blue, purple and black categories) are shown separately in Figure S3.

(Figure S11, Tables S7 and S8). Further peptide hits were also associated with the outer membrane, such as porins, and hits to NMT1/THI5 family or SsuA/THI5-like, as well as bacterial alkaline-phosphatase-like or sulfatase function.

4 | DISCUSSION

4.1 | Impact of extraction method on extracts composition

The three extraction methods employed here resulted each in different protein and humic substance content (Figure 1). Protein content was previously shown to differ with extractant (Bastida et al., 2018; Greenfield et al., 2018; Kanerva et al., 2013; Masciandaro et al., 2008). Here, we show that total humic substance content was significantly higher in NaPPi extracts, confirming the use of this extractant for the effective liberation of OM substances from OM-mineral associations (MAOM) (Bakina & Orlova, 2012; Bremner & Lees, 1949; Masciandaro et al., 2008). EDTA extracts also exhibited significantly higher total humic substance content when compared to water extracts, likely explained by EDTA chelation disrupting Ca^{2+} bridges between enzyme-humic substance complexes (Nannipieri, 2006). EDTA has also been associated with increased co-extraction of humic substances in DNA extractions (ibid.). Protein yields of NaPPi and EDTA extracts were similar (Figure 1), and this could indicate, that both EDTA and NaPPi extractants lysed cells, hence encompassing microbial intracellular protein content. Alternatively, the high total protein content in NaPPi extracts could also be the

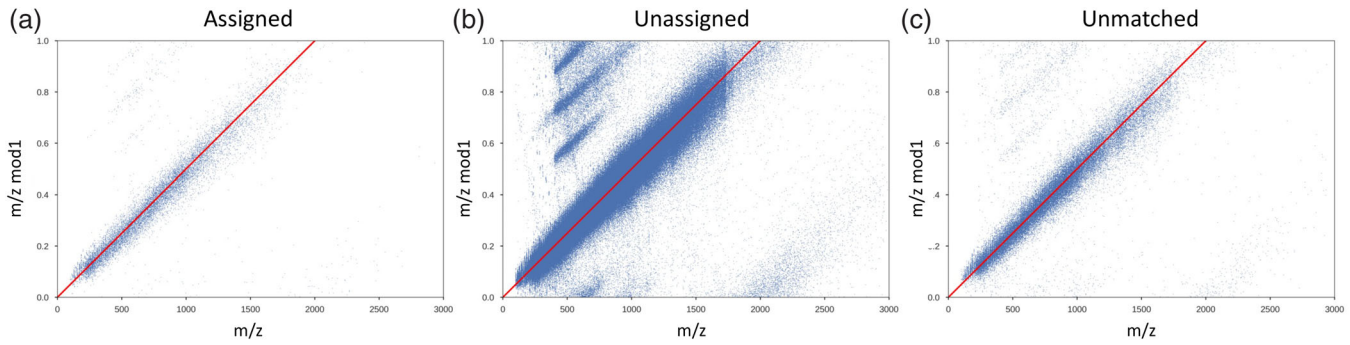

FIGURE 3 Distribution of *m/z* values versus *m/z* modulo 1 for assigned, unassigned and unmatched peaks from water 1 sample shown in Figure 2. The expected average nucleon mass of peptide fragments (1.0005 Da) is indicated by a red line with a slope of 0.0005.

result of the large pool of soil-stabilised N/protein (Kanerva et al., 2013; Nannipieri, 2006). The much lower total protein content of water extracts might indicate that cell lysis was limited. While EDTA and NaPPi extracts both exhibited significantly higher total humic substance contents than water extracts (Figure 1), this also indicates that NaDoc-TCA precipitation did not successfully separate proteins from other soil organic substances. Furthermore, repeated cleaning steps only resulted in approaching a ratio of 1:1 total protein to humic substances and were at the cost of reduced protein recovery (data not shown). Despite this, NaDoc-TCA is regarded as a standard biochemistry preparation to precipitate proteins from dilute solutions (Bensadoun & Weinstein, 1976), and TCA precipitation is commonly applied to soil protein extracts (Keiblinger et al., 2016), even though it was also observed to co-precipitate other soil organic substances (Qian & Hettich, 2017). Workflows using phenol/chloroform-isoamyl alcohol and phenol phase content precipitation with ammonium acetate in methanol have led to higher number of peptide hits with LC–MS/MS (Keiblinger et al., 2012; Quinn et al., 2022; Thorn et al., 2019), perhaps due to their ability to reduce total soil organic substances, which is important for LC–MS/MS, as discussed later. This cleaning method, however, has been argued to lead to protein composition bias (Keiblinger et al., 2016; Qian & Hettich, 2017). It is worth noting that the aqueous phase is typically removed when using a phenol/chloroform-isoamyl alcohol method, whereas here we demonstrate that water extracts contain important protein signals (Figure 3, Figure S8).

We report varying performance of contrasting soil fractions on LC–MS/MS as well as high incidence of unmatched spectra in blank controls. When examining the order in which the samples were run on LC–MS/MS (Table S2), we could identify soil substance carry-over (Figure 2 and Figure S3). This points to the adverse effects of soil substances carryover on ion acquisition processes (Hockaday et al., 2009; Novotny et al., 2014). Humic substances may bind strongly to C_{18}-coated adsorbents and may not be released efficiently, depending on the elution profile (Trubetskaya et al., 2015; Trubetskoi & Trubetskaya, 2015). However, they can generally be eluted in a reverse-phased LC system and ionised in ESI. Indeed, fulvic acid, humic acid, and other NOM substances are regularly investigated with such setups (Brown et al., 2016; Koch et al., 2008; Kujawinski et al., 2002; Reemtsma & These, 2003; Stenson, 2008). For relatively clean water samples, the elution profile used here covered relevant peptide elution conditions well (Figure S4). Soil extracts with unfavourable protein to humic substance ratio were previously reported to be challenging to analyse with LC–MS/MS due to adverse ion interactions (e.g., charge neutralisation) and can even lead to column clogging (Arenella et al., 2014; Qian & Hettich, 2017). This may explain some low spectral counts for EDTA and NaPPi extracts even in most hydrophobic elution conditions (Figure 2, Figure S3). Considering the fundamental difficulties in efficiently separating proteins from complex soil organic mixtures, more elaborate separation techniques need to be investigated such as for example the one proposed by Qian and Hettich (2017), and/or those employed in NOM research that involves capillary zone electrophoresis and free-flow electrophoresis coupled to MS (Garrison et al., 1995; Gaspar et al., 2010; Keuth et al., 1998; Schmitt-Kopplin & Kettrup, 2003), and ion mobility chamber (Lu et al., 2018; Xia et al., 2006). Our results indicate that broadening the very hydrophobic elution phase could perhaps extend the spectral acquisition time of unmatched spectra, especially for NaPPi extracts where this peak was concentrated (Figure 2, Figures S3 and S4). Fundamentally, to avoid carry-over between soil samples, we recommend extending the LC cleaning time with high ACN or methanol proportions beyond the increased wash cycles already employed.

The ion fragmentation pattern analysis (Figure 3, Figures S8–S10) provided important clues regarding the

nature of the most frequent spectra detected in water extracts, which belong to the unassigned category (Figure 2, Figure S3). Although the latter exhibited bimodal peaks, separated by hydrophobicity, they were found to exhibit strong peptide signature (Figure 3 and Figure S8). This was surprising but in line with previous reports of the hydrophobic nature of WEOM (Rennert et al., 2007), as well as elution under more hydrophilic conditions (Trubetskaya et al., 2015, Trubetskoi & Trubetskaya, 2015). Despite displaying a strong peptide signal, unassigned spectra did not meet the Mascot significance threshold, perhaps due to co-elution of competing ions, which could have led to target ion (peptide fragment ladder) suppression (Arenella et al., 2014, Qian & Hettich, 2017). Indeed, unassigned and anchored spectra exhibited significantly reduced fragment numbers compared to their assigned counterparts (Figures S5–S7).

Unmatched peaks of blank controls and NaPPi sample replicate 1, eluting in the most hydrophobic elution fractions exhibited a distribution of fragment ions along a steeper slope than that expected for peptidic signal, indicating the presence of molecules other than peptides (Figures S8–S10). Furthermore, those unmatched spectra also contained many fragments of unknown origin (i.e., not belonging to either slope), which may correspond to humic substances co-extracted particularly during NaPPi and EDTA extractant (Figure 1). Their presence in the blank controls could be resulting from sample carryover between runs as NaPPi and EDTA soil samples were run first in LC–MS/MS (Table S2). Strikingly, all extractions led to replicate samples containing strong peptide signals including spectra classified as unassigned and even unmatched (Figure 3, Figures S8–-S10). This clearly highlights the need for further developments to allow soil metaproteomics to reach its full potential.

4.2 | Metaproteomic composition in contrasting soil extracts

We sought to characterise the soil metaprotein composition in contrasting soil extracts. Despite exhibiting the lowest extracted specific protein content (Figure 1), water extracts yielded the most assigned QC peptide hits (Table S6), likely due to the low MS/MS co-occurrence of other compounds (Figure 3) and a favourable LC elution profile (Figure 2, Figure S3) tailored toward clean matrix proteomics. Furthermore, water extracts harbour the greater unrealised potential for assigned hits derived from its large pool of unassigned spectra, displaying strong peptide signature (Figure 3, and Figure S8). Although repeated LC washing and MS calibration (Table S2) improved results in water extracts (data not shown), this may not have fully alleviated the impact of soil substance carryover on water extract peptide identification.

Peptides were most frequently assigned to the phylum Proteobacteria and this was in agreement with other soil metaproteomic studies, including both those targeting extracellular proteins (Bastida et al., 2018; Johnson-Rollings et al., 2014) and those encompassing cell lysis steps (Keiblinger et al., 2012; Mattarozzi et al., 2017; Quinn et al., 2022; Thorn et al., 2019). Although Gram-negative bacteria, may be regarded to be more susceptible to mechanical lysing conditions, due to their typically thinner peptidoglycan layer, Proteobacteria were found to exhibit population stability to soil drying and rewetting stress similarly to, for example, Firmicutes (Barnard et al., 2013). This could possibly indicate that soil bacteria have adapted to variable soil water status and that the outer membrane of Gram-negative bacteria can also confer substantial load-bearing stability (Rojas et al., 2018). Technically, peptide matches may be attributed to several different proteins and functions in a taxonomically and functionally diverse environment, and this was shown here by reporting on maximum 3 functions per peptide (Figures S11–S13; Table S7; Gurdeep Singh et al., 2018). In water extracts, the most frequent hits had functions involved in diverse substrate/ligand binding proteins of ABC transport systems (Figure S11, Tables S7 and S8). This likely indicates the importance of plant-derived low molecular weight organic compounds for bacterial survival in soil under the influence of grassland rhizospheres. Indeed, ABC systems are typically highly regulated (Davidson et al., 2008), and root exudation is a primary source of sugars, amino acids and other labile substrates (Kuzyakov, 2010; Paterson, 2009; Paul, 2016). Elements of bacterial ABC systems can also be involved in osmosensing, DNA repair, and in export systems (Davidson et al., 2008), and this could link to the importance of plant-microbe signalling and quorum sensing in the rhizosphere (Hawkes et al., 2007). Further peptide hits to NMT1/THI5 family or SsuA/THI5-like, may be potentially involved in the biosynthesis of hydroxymethylpyrimidine, binding of vitamine for ABC transport, or sulfonate binding and transport (Blum et al., 2021). Apart from these outer membrane-associated hits, further peptide hits were also associated with export system protein families such as TAT signal peptide, ribonuclease Z/Hydroxyacylglutathione hydrolase-like, (metallo)-beta-lactamase and lamin tail domain (Blum et al., 2021; Lee et al., 2006; Mans et al., 2004; Wickner & Schekman, 2005). Despite this, free-extracellular enzymes, that is, fungal enzymes, and other fungal proteins were rarely detected and may have been limited by the low number

of assigned peptide hits (Johnson-Rollings et al., 2014), and the bioinformatic challenge in identifying eukaryotic peptides in metaproteomic settings (Choudhary et al., 2001; Cottrell, 2011; Muth, Kolmeder, et al., 2015). Bacterial alkaline-phosphatase-like or sulfatase were also identified (Table S8). This function was denoted as arylsulfatase with EMBL GO annotation, a well-studied enzyme, which makes sulphate esters, the dominant form of S and aromatic and sugar sulphates in soil, available for plant uptake via sulfatase mineralisation (Kertesz, 2000; Kertesz et al., 2007). Arylsulfatases have been reported to be associated with membrane, intracellular, extracellular, or periplasmic space (Cregut et al., 2013; Kertesz, 2000). Overall, microbial peptide hits indicated that WEOM extracts were largely composed of proteins from the extra- or ectocellular space.

5 | STUDY LIMITATIONS AND CONCLUSIONS

We combined proteomic and fragmentation pattern analysis and showed that contrasting soil extracts exhibit strong molecular peptide signature on LC–MS/MS. This proof-of-concept study highlights the potential for soil extract fractionation prior to proteomics but further research is required to validate such an approach. Extractions should be carried out on a range of soils with different textures, mineralogy, OM quality and land management. The workflow presented here could also benefit from further optimisation and include for example pre-cleaning steps as well as de novo peptide matching strategy. The number of peptides identified would also likely greatly benefit from the availability of a relevant metagenomic database (Jouffret et al., 2021). Our experimental strategy could also be validated using NOM reference materials (Suwanee River, Elliott Soil) and employ higher resolution mass spectrometers for target molecule matching and molecular formula assignment. Overall, this study identified specific areas of focus for improving soil metaproteomics and proposed a framework for mining mass spectra for both peptide assignment and fragmentation pattern characterisation.

AUTHOR CONTRIBUTIONS

Matthias Waibel: Conceptualization; investigation; funding acquisition; writing – original draft; visualization; methodology; writing – review and editing. **Kevin McDonnell:** Investigation; writing – original draft; methodology; visualization; writing – review and editing. **Maria Tuohy:** Conceptualization; methodology; supervision; writing – review and editing. **Sally Shirran:** Methodology; writing – review and editing. **Sylvia Synowsky:** Methodology; writing – review and editing. **Barry Thornton:** Funding acquisition; writing – review and editing; supervision. **Eric Paterson:** Funding acquisition; writing – review and editing; supervision. **Fiona Brennan:** Funding acquisition; writing – review and editing; supervision. **Florence Abram:** Conceptualization; funding acquisition; writing – original draft; writing – review and editing; visualization; supervision.

ACKNOWLEDGEMENTS

The authors are grateful to Jim Grant (Teagasc) for his help with experimental design, and Clelton Santos (University of Galway) for his help on protein biochemistry, The authors thank Richard Hewison (JHI) for grassland plant identifications and David Wall (Teagasc) for sampling of Grange soil. We are also grateful to Liz Shaw (University of Reading) and Alexandre de Menezes (University of Galway) for their helpful comments on an earlier version of this work. Matthias Waibel was funded by the University of Galway College of Science and the Irish Research Council under GOIPG/2016/1215. The James Hutton Institute receives funding support from the Rural and Environment Science and Analytical Services Division of the Scottish Government. Open access funding provided by IReL.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interests.

DATA AVAILABILITY STATEMENT

Raw, peak and result files (.wiff.scan, *.mgf*, *.mzIdentML*, *.csv)*, gel picture and experimental design are available at the ProteomeXchange Consortium via the PRoteomics IDEntifications database (PRIDE) partner repository under the dataset identifier PXD034783. Code is available on github: https://github.com/waibel-123/PSManalysis.

WEB-ENHANCED OBJECTS

The interactive Chordomics plots are available under: https://waibel-m-123.shinyapps.io/matthias_data

ORCID

Matthias Waibel https://orcid.org/0000-0002-2505-8917
Florence Abram https://orcid.org/0000-0001-7522-9158

REFERENCES

Abram, F. (2015). Systems-based approaches to unravel multi-species microbial community functioning. *Computational and Structural Biotechnology Journal*, *13*, 24–32.

Allison, S. D., Weintraub, M. N., Gartner, T. B., & Waldrop, M. P. (2010). Evolutionary-economic principles as regulators of soil enzyme production and ecosystem function. In *Soil Enzymology* (pp. 229–243). Springer.

Angst, G., Mueller, K. E., Nierop, K. G., & Simpson, M. J. (2021). Plant-or microbial-derived? A review on the molecular

composition of stabilized soil organic matter. *Soil Biology and Biochemistry*, *156*, 108189.

Arenella, M., Giagnoni, L., Masciandaro, G., Ceccanti, B., Nannipieri, P., & Renella, G. (2014). Interactions between proteins and humic substances affect protein identification by mass spectrometry. *Biology and Fertility of Soils*, *50*(3), 447–454.

Ashman, M., & Puri, G. (2013). *Essential soil science: A clear and concise introduction to soil science*. John Wiley & Sons.

Bakina, L., & Orlova, N. (2012). Special features of humus acids extraction from soils by sodium pyrophosphate solutions of different alkalinity. *Eurasian Soil Science*, *45*(4), 392–398.

Barnard, R. L., Osborne, C. A., & Firestone, M. K. (2013). Responses of soil bacterial and fungal communities to extreme desiccation and rewetting. *The ISME Journal*, *7*(11), 2229–2241.

Bastida, F., Jehmlich, N., Torres, I., & García, C. (2018). The extracellular metaproteome of soils under semiarid climate: A methodological comparison of extraction buffers. *Science of the Total Environment*, *619*, 707–711.

Baveye, P. C., & Wander, M. (2019). The (bio) chemistry of soil humus and humic substances: Why is the "new view" still considered novel after more than 80 years? *Frontiers in Environmental Science*, *7*, 27.

Bensadoun, A., & Weinstein, D. (1976). Assay of proteins in the presence of interfering materials. *Analytical Biochemistry*, *70*(1), 241–250.

Blum, M., Chang, H., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., Nuka, G., Paysan-Lafosse, T., Qureshi, M., & Raj, S. (2021). The InterPro protein families and domains database: 20 years on. *Nucleic Acids Research*, *49*(D1), D344–D354.

Bot, A., & Benites, J. (2007). *The importance of soil organic matter: Key to drought-resistant soil and sustained food production*. Food and Agriculture Organisation (FAO).

Bremner, J., & Lees, H. (1949). Studies on soil organic matter: Part II. The extraction of organic matter from soil by neutral reagents. *The Journal of Agricultural Science*, *39*(3), 274–279.

Brown, A. (1976). Microbial water stress. *Bacteriological Reviews*, *40*(4), 803–846.

Brown, T. A., Jackson, B. A., Bythell, B. J., & Stenson, A. C. (2016). Benefits of multidimensional fractionation for the study and characterization of natural organic matter. *Journal of Chromatography A*, *1470*, 84–96.

Burns, R. G., Deforest, J. L., Marxsen, J., Sinsabaugh, R. L., Stromberger, M. E., Wallenstein, M. D., Weintraub, M. N., & Zoppini, A. (2013). Soil enzymes in a changing environment: Current knowledge and future directions. *Soil Biology and Biochemistry*, *58*, 216–234.

Burns, R. G., & Dick, R. P. (2002). *Enzymes in the environment: Activity, ecology, and applications*. CRC Press.

Choudhary, J. S., Blackstock, W. P., Creasy, D. M., & Cottrell, J. S. (2001). Matching peptide mass spectra to EST and genomic DNA databases. *Trends in Biotechnology*, *19*, 17–22.

Cotrufo, M. F., Wallenstein, M. D., Boot, C. M., Denef, K., & Paul, E. (2013). The microbial efficiency-matrix stabilization (MEMS) framework integrates plant litter decomposition with soil organic matter stabilization: Do labile plant inputs form stable soil organic matter? *Global Change Biology*, *19*(4), 988–995.

Cottrell, J. S. (2011). Protein identification using MS/MS data. *Journal of Proteomics*, *74*(10), 1842–1851.

Cregut, M., Piutti, S., Slezack-Deschaumes, S., & Benizri, E. (2013). Compartmentalization and regulation of arylsulfatase activities in Streptomyces sp., microbacterium sp. and Rhodococcus sp. soil isolates in response to inorganic sulfate limitation. *Microbiological Research*, *168*(1), 12–21.

Davidson, A. L., Dassa, E., Orelle, C., & Chen, J. (2008). Structure, function, and evolution of bacterial ATP-binding cassette systems. *Microbiology and Molecular Biology Reviews*, *72*(2), 317–364.

Dungait, J. A., Hopkins, D. W., Gregory, A. S., & Whitmore, A. P. (2012). Soil organic matter turnover is governed by accessibility not recalcitrance. *Global Change Biology*, *18*(6), 1781–1796.

Duval, J. F., Wilkinson, K. J., van Leeuwen, H. P., & Buffle, J. (2005). Humic substances are soft and permeable: Evidence from their electrophoretic mobilities. *Environmental Science & Technology*, *39*(17), 6435–6445.

Dytham, C. (2011). *Choosing and using statistics: A biologist's guide* (3rd ed.). Wiley-Blackwell.

Ekschmitt, K., Liu, M., Vetter, S., Fox, O., & Wolters, V. (2005). Strategies used by soil biota to overcome soil organic matter stability—Why is dead organic matter left over in the soil? *Geoderma*, *128*(1–2), 167–176.

Fontaine, S., & Barot, S. (2005). Size and functional diversity of microbe populations control plant persistence and long-term soil carbon accumulation. *Ecology Letters*, *8*(10), 1075–1087.

Frolund, B., Griebe, T., & Nielsen, P. (1995). Enzymatic activity in the activated-sludge floc matrix. *Applied Microbiology and Biotechnology*, *43*(4), 755–761.

Garrison, A., Schmitt, P., & Kettrup, A. (1995). Capillary electrophoresis for the characterization of humic substances. *Water Research*, *29*(9), 2149–2159.

Gaspar, A., Harir, M., Hertkorn, N., & Schmitt-Kopplin, P. (2010). Preparative free-flow electrophoretic offline ESI-Fourier transform ion cyclotron resonance/MS analysis of Suwannee River fulvic acid. *Electrophoresis*, *31*(12), 2070–2079.

Greenfield, L. M., Hill, P. W., Paterson, E., Baggs, E. M., & Jones, D. L. (2018). Methodological bias associated with soluble protein recovery from soil. *Scientific Reports*, *8*(1), 1–6.

Grosso, F., Temussi, F., & de Nicola, F. (2014). Water-extractable organic matter and enzyme activity in three forest soils of the Mediterranean area. *European Journal of Soil Biology*, *64*, 15–22.

Gurdeep Singh, R., Tanca, A., Palomba, A., Van Der Jeugt, F., Verschaffelt, P., Uzzau, S., Martens, L., Dawyndt, P., & Mesuere, B. (2018). Unipept 4.0: Functional analysis of metaproteome data. *Journal of Proteome Research*, *18*(2), 606–615.

Hawkes, C. V., Deangelis, K. M., & Firestone, M. K. (2007). Root interactions with soil microbial communities and processes. In Z. G. Cardon & J. L. Whitbeck (Eds.), *The rhizosphere: An ecological perspective*. Elsevier.

Helfrich, M., Flessa, H., Mikutta, R., Dreves, A., & Ludwig, B. (2007). Comparison of chemical fractionation methods for isolating stable soil organic carbon pools. *European Journal of Soil Science*, *58*(6), 1316–1329.

Hertkorn, N., Ruecker, C., Meringer, M., Gugisch, R., Frommberger, M., Perdue, E., Witt, M., & Schmitt-Kopplin, P. (2007). High-precision frequency measurements: Indispensable tools at the core of the molecular-level analysis of complex systems. *Analytical and Bioanalytical Chemistry*, *389*(5), 1311–1327.

Heyer, R., Schallert, K., Büdel, A., Zoun, R., Dorl, S., Behne, A., Kohrs, F., Püttker, S., Siewert, C., & Muth, T. (2019). A robust and universal metaproteomics workflow for research studies and routine diagnostics within 24 hours using phenol extraction, FASP digest and the MetaProteomeAnalyzer. *Frontiers in Microbiology*, *10*, 1883.

Hockaday, W. C., Purcell, J. M., Marshall, A. G., Baldock, J. A., & Hatcher, P. G. (2009). Electrospray and photoionization mass spectrometry for the characterization of organic matter in natural waters: A qualitative assessment. *Limnology and Oceanography: Methods*, *7*(1), 81–95.

Johnson-Rollings, A. S., Wright, H., Masciandaro, G., Macci, C., Doni, S., Calvo-Bado, L. A., Slade, S. E., Plou, C. V., & Wellington, E. M. (2014). Exploring the functional soil-microbe interface and exoenzymes through soil metaexoproteomics. *The ISME Journal*, *8*(10), 2148–2150.

Joseph, S., & David, W. R. (2001). *Molecular cloning: A laboratory manual*. Cold Spring Harbor.

Jouffret, V., Miotello, G., Culotta, K., Ayrault, S., Pible, O., & Armengaud, J. (2021). Increasing the power of interpretation for soil metaproteomics data. *Microbiome*, *9*, 195. https://doi.org/10.1186/s40168-021-01139-1

Just, C., Poeplau, C., Don, A., Van Wesemael, B., Kögel-Knabner, I., & Wiesmeier, M. (2021). A simple approach to isolate slow and fast cycling organic carbon fractions in central European soils—Importance of dispersion method. *Frontiers in Soil Science*, *1*, 692583.

Kaiser, M., Wirth, S., Ellerbrock, R., & Sommer, M. (2010). Microbial respiration activities related to sequentially separated, particulate and water-soluble organic matter fractions from arable and forest topsoils. *Soil Biology and Biochemistry*, *42*(3), 418–428.

Kanerva, S., Smolander, A., Kitunen, V., Ketola, R., & Kotiaho, T. (2013). Comparison of extractants and applicability of MALDI–TOF-MS in the analysis of soil proteinaceous material from different types of soil. *Organic Geochemistry*, *56*, 1–9.

Keiblinger, K. M., Fuchs, S., Zechmeister-Boltenstern, S., & Riedel, K. (2016). Soil and leaf litter metaproteomics-a brief guideline from sampling to understanding. *FEMS Microbiology Ecology*, *92*(11), fiw180.

Keiblinger, K. M., Wilhartitz, I. C., Schneider, T., Roschitzki, B., Schmid, E., Eberl, L., Riedel, K., & Zechmeister-Boltenstern, S. (2012). Soil metaproteomics–comparative evaluation of protein extraction protocols. *Soil Biology and Biochemistry*, *54*, 14–24.

Keller, M., & Hettich, R. (2009). Environmental proteomics: A paradigm shift in characterizing microbial activities at the molecular level. *Microbiology and Molecular Biology Reviews*, *73*(1), 62–70.

Kemmitt, S., Lanyon, C., Waite, I., Wen, Q., Addiscott, T., Bird, N. R., O'Donnell, A., & Brookes, P. (2008). Mineralization of native soil organic matter is not regulated by the size, activity or composition of the soil microbial biomass—A new perspective. *Soil Biology and Biochemistry*, *40*(1), 61–73.

Kertesz, M. A. (2000). Riding the sulfur cycle–metabolism of sulfonates and sulfate esters in gram-negative bacteria. *FEMS Microbiology Reviews*, *24*(2), 135–175.

Kertesz, M. A., Fellows, E., & Schmalenberger, A. (2007). Rhizobacteria and plant sulfur supply. *Advances in Applied Microbiology*, *62*, 235–268.

Keuth, U., Leinenbach, A., Beck, H. P., & Wagner, H. (1998). Separation and characterization of humic acids and metal humates by electrophoretic methods. *Electrophoresis*, *19*(7), 1091–1096.

Kleber, M., Sollins, P., & Sutton, R. (2007). A conceptual model of organo-mineral interactions in soils: Self-assembly of organic molecular fragments into zonal structures on mineral surfaces. *Biogeochemistry*, *85*(1), 9–24.

Koch, B. P., Ludwichowski, K., Kattner, G., Dittmar, T., & Witt, M. (2008). Advanced characterization of marine dissolved organic matter by combining reversed-phase liquid chromatography and FT-ICR-MS. *Marine Chemistry*, *111*(3–4), 233–241.

Koskinen, V. (2021). Unipept and mascot. Blog Post https://www.matrixscience.com/blog/unipept-and-mascot.html#comments

Koskinen, V. R., Emery, P. A., Creasy, D. M., & Cottrell, J. S. (2011). Hierarchical clustering of shotgun proteomics data. *Molecular & Cellular Proteomics*, *10*(6), M110.003822.

Kujawinski, E. B., Freitas, M. A., Zang, X., Hatcher, P. G., Green-Church, K. B., & Jones, R. B. (2002). The application of electrospray ionization mass spectrometry (ESI MS) to the structural characterization of natural organic matter. *Organic Geochemistry*, *33*(3), 171–180.

Kuzyakov, Y. (2010). Priming effects: Interactions between living and dead organic matter. *Soil Biology and Biochemistry*, *42*(9), 1363–1371.

Kuzyakov, Y., Blagodatskaya, E., Blagodatsky, S., & Blagodatsky, S. (2009). Comments on the paper by Kemmitt et al.(2008)'Mineralization of native soil organic matter is not regulated by the size, activity or composition of the soil microbial biomass–A new perspective'[Soil Biology & Biochemistry 40, 61–73]: The biology of the Regulatory Gate. *Soil Biology and Biochemistry*, *41*(2), 435–439.

Lee, P. A., Tullman-Ercek, D., & Georgiou, G. (2006). The bacterial twin-arginine translocation pathway. *Annual Review of Microbiology*, *60*, 373–395.

Lehmann, J., & Kleber, M. (2015). The contentious nature of soil organic matter. *Nature*, *528*(7580), 60–68.

Liang, C., Amelung, W., Lehmann, J., & Kästner, M. (2019). Quantitative assessment of microbial necromass contribution to soil organic matter. *Global Change Biology*, *25*(11), 3578–3590.

Lowry, O. H., Rosebrough, N. J., Farr, A. L., & Randall, R. J. (1951). Protein measurement with the Folin phenol reagent. *Journal of Biological Chemistry*, *193*, 265–275.

Lu, K., Gardner, W. S., & Liu, Z. (2018). Molecular structure characterization of riverine and coastal dissolved organic matter with ion mobility quadrupole time-of-flight LCMS (IM Q-TOF LCMS). *Environmental Science & Technology*, *52*(13), 7182–7191.

Mans, B., Anantharaman, V., Aravind, L., & Koonin, E. V. (2004). Comparative genomics, evolution and origins of the nuclear envelope and nuclear pore complex. *Cell Cycle*, *3*(12), 1625–1650.

Masciandaro, G., Macci, C., Doni, S., Maserti, B. E., Leo, A. C., Ceccanti, B., & Wellington, E. (2008). Comparison of extraction methods for recovery of extracellular β-glucosidase in two different forest soils. *Soil Biology and Biochemistry*, *40*(9), 2156–2161.

Mattarozzi, M., Manfredi, M., Montanini, B., Gosetti, F., Sanangelantoni, A. M., Marengo, E., Careri, M., & Visioli, G. (2017). A metaproteomic approach dissecting major bacterial

functions in the rhizosphere of plants living in serpentine soil. *Analytical and Bioanalytical Chemistry*, *409*(9), 2327–2339.

Mccarthy, T., & Tuohy, M. G. (2011). A multi-step chromatographic strategy to purify three fungal endo-β-glucanases. In *Protein chromatography* (pp. 497–524). Springer.

Mcdonnell, K., Howley, E., & Abram, F. (2023). Critical evaluation of the use of artificial data for machine learning based de novo peptide identification. *Computational and Structural Biotechnology Journal*, *21*, 2732–2743.

Mcdonnell, K., Waters, N., Howley, E., & Abram, F. (2020). Chordomics: A visualization tool for linking function to phylogeny in microbiomes. *Bioinformatics*, *36*(4), 1309–1310.

Mesuere, B., Debyser, G., Aerts, M., Devreese, B., Vandamme, P., & Dawyndt, P. (2015). The Unipept metaproteomics analysis pipeline. *Proteomics*, *15*(8), 1437–1442.

Montgomery, D. R. (2007). Soil erosion and agricultural sustainability. *Proceedings of the National Academy of Sciences*, *104*(33), 13268–13272.

Moore, E. K., Nunn, B. L., Faux, J. F., Goodlett, D. R., & Harvey, H. R. (2012). Evaluation of electrophoretic protein extraction and database-driven protein identification from marine sediments. *Limnology and Oceanography: Methods*, *10*(5), 353–366.

Muth, T., Behne, A., Heyer, R., Kohrs, F., Benndorf, D., Hoffmann, M., Lehtevä, M., Reichl, U., Martens, L., & Rapp, E. (2015). The MetaProteomeAnalyzer: A powerful open-source software suite for metaproteomics data analysis and interpretation. *Journal of Proteome Research*, *14*(3), 1557–1565.

Muth, T., Kolmeder, C. A., Salojärvi, J., Keskitalo, S., Varjosalo, M., Verdam, F. J., Rensen, S. S., Reichl, U., De Vos, W. M., & Rapp, E. (2015). Navigating through metaproteomics data: A logbook of database searching. *Proteomics*, *15*(20), 3439–3453.

Nannipieri, P. (2006). Role of stabilised enzymes in microbial ecology and enzyme extraction from soil with potential applications in soil proteomics. In P. Nannipieri & K. Smalla (Eds.), *Nucleic acids and proteins in soil* (pp. 75–94). Springer Science & Business Media.

Nichols, K. A., & Wright, S. F. (2006). Carbon and nitrogen in operationally defined soil organic matter pools. *Biology and Fertility of Soils*, *43*, 215–220.

Novotny, N. R., Capley, E. N., & Stenson, A. C. (2014). Fact or artifact: The representativeness of ESI-MS for complex natural organic mixtures. *Journal of Mass Spectrometry*, *49*(4), 316–326.

Ogunseitan, O. A. (2006). Soil proteomics: Extraction and analysis of proteins from soils. In P. Nannipieri & K. Smalla (Eds.), *Nucleic acids and proteins in soil*. Springer Science & Business Media.

Paterson, E. (2009). Comments on the regulatory gate hypothesis and implications for C-cycling in soil. *Soil Biology and Biochemistry*, *41*(6), 1352–1354.

Paul, E. (2014). *Soil microbiology, ecology, and biochemistry*. Academic Press.

Paul, E. A. (2016). The nature and dynamics of soil organic matter: Plant inputs, microbial transformations, and organic matter stabilization. *Soil Biology and Biochemistry*, *98*, 109–126.

Peralta, G., Di Paolo, L., Luotto, I., Omuto, C., Mainka, M., Viatkin, K., & Yigini, Y. (2022). *Global soil organic carbon sequestration potential map (GSOCseq v1.1) – Technical manual*. FAO. https://doi.org/10.4060/cb2642en

Piccolo, A. (2001). The supramolecular structure of humic substances. *Soil Science*, *166*(11), 810–832.

Pot, V., Portell, X., Otten, W., Garnier, P., Monga, O., & Baveye, P. C. (2022). Understanding the joint impacts of soil architecture and microbial dynamics on soil functions: Insights derived from microscale models. *European Journal of Soil Science*, *73*(3), e13256.

Qian, C., & Hettich, R. L. (2017). Optimized extraction method to remove humic acid interferences from soil samples prior to microbial proteome measurements. *Journal of Proteome Research*, *16*(7), 2537–2546.

Quinn, G. A., Abdelhameed, A., Banat, I. M., Berrar, D., Doerr, S. H., Dudley, E., Francis, L. W., Gazze, S. A., Hallin, I., Matthews, G. P., Swain, M. T., Whalley, W. R., & van Keulen, G. (2022). Complimentary protein extraction methods increase the identification of the park grass experiment metaproteome. *Applied Soil Ecology*, *173*, 104388.

Redmile-Gordon, M., Armenise, E., White, R., Hirsch, P., & Goulding, K. (2013). A comparison of two colorimetric assays, based upon Lowry and Bradford techniques, to estimate total protein in soil extracts. *Soil Biology and Biochemistry*, *67*, 166–173.

Reemtsma, T., & These, A. (2003). On-line coupling of size exclusion chromatography with electrospray ionization-tandem mass spectrometry for the analysis of aquatic fulvic and humic acids. *Analytical Chemistry*, *75*(6), 1500–1507.

Renella, G., Ogunseitan, O., Giagnoni, L., & Arenella, M. (2014). Environmental proteomics: A long march in the pedosphere. *Soil Biology and Biochemistry*, *69*, 34–37.

Rennert, T., Gockel, K. F., & Mansfeldt, T. (2007). Extraction of water-soluble organic matter from mineral horizons of forest soils. *Journal of Plant Nutrition and Soil Science*, *170*(4), 514–521.

Rillig, M. C., Caldwell, B. A., Wösten, H. A., & Sollins, P. (2007). Role of proteins in soil carbon and nitrogen storage: Controls on persistence. *Biogeochemistry*, *85*(1), 25–44.

Rojas, E. R., Billings, G., Odermatt, P. D., Auer, G. K., Zhu, L., Miguel, A., Chang, F., Weibel, D. B., Theriot, J. A., & Huang, K. C. (2018). The outer membrane is an essential load-bearing element in gram-negative bacteria. *Nature*, *559*(7715), 617–621.

Schlüter, S., Leuther, F., Albrecht, L., Hoeschen, C., Kilian, R., Surey, R., Mikutta, R., Kaiser, K., Mueller, C. W., & Vogel, H. (2022). Microscale carbon distribution around pores and particulate organic matter varies with soil moisture regime. *Nature Communications*, *13*(1), 2098.

Schmitt-Kopplin, P., & Kettrup, A. (2003). Capillary electrophoresis–electrospray ionization–mass spectrometry for the characterization of natural organic matter: An evaluation with free flow electrophoresis-off-line flow injection electrospray ionization-mass spectrometry. *Electrophoresis*, *24*(17), 3057–3066.

Schulze, W. X., Gleixner, G., Kaiser, K., Guggenberger, G., Mann, M., & Schulze, E. (2005). A proteomic fingerprint of dissolved organic carbon and of soil particles. *Oecologia*, *142*(3), 335–343.

Spohn, M., & Kuzyakov, Y. (2014). Spatial and temporal dynamics of hotspots of enzyme activity in soil as affected by living and dead roots—A soil zymography analysis. *Plant and Soil*, *379*(1), 67–77.

Stahr, K., Kandeler, E., Herrmann, L., & Streck, T. (2016). *Bodenkunde und Standortlehre* (Vol. 2967). Ulmer UTB.

Stenson, A. C. (2008). Reversed-phase chromatography fractionation tailored to mass spectral characterization of humic substances. *Environmental Science & Technology*, *42*(6), 2060–2065.

Sutton, R., & Sposito, G. (2005). Molecular structure in soil humic substances: The new view. *Environmental Science & Technology*, *39*(23), 9009–9015.

Tan, K. H. (2014). *Humic matter in soil and the environment: Principles and controversies*. CRC Press.

Tanca, A., Palomba, A., Deligios, M., Cubeddu, T., Fraumene, C., Biosa, G., Pagnozzi, D., Addis, M. F., & Uzzau, S. (2013). Evaluating the impact of different sequence databases on metaproteome analysis: Insights from a lab-assembled microbial mixture. *PLoS One*, *8*(12), e82981.

Team R. (2019). *R: A language and environment for statistical computing, version 4.05*. R Foundation for Statistical Computing. https://www.r-project.org/

Thorn, C. E., Bergesch, C., Joyce, A., Sambrano, G., Mcdonnell, K., Brennan, F., Heyer, R., Benndorf, D., & Abram, F. (2019). A robust, cost-effective method for DNA, RNA and protein co-extraction from soil, other complex microbiomes and pure cultures. *Molecular Ecology Resources*, *19*(2), 439–455.

Trubetskaya, O. E., Richard, C., & Trubetskoj, O. A. (2015). Evaluation of Suwannee River NOM electrophoretic fractions by RP-HPLC with online absorbance and fluorescence detection. *Environmental Science and Pollution Research*, *22*(13), 9989–9998.

Trubetskoi, O., & Trubetskaya, O. (2015). Reversed-phase high-performance liquid chromatography of the stable electrophoretic fractions of soil humic acids. *Eurasian Soil Science*, *48*(2), 148–156.

Verberkmoes, N. C., Russell, A. L., Shah, M., Godzik, A., Rosenquist, M., Halfvarson, J., Lefsrud, M. G., Apajalahti, J., Tysk, C., & Hettich, R. L. (2009). Shotgun metaproteomics of the human distal gut microbiota. *The ISME Journal*, *3*(2), 179–189.

Von Lützow, M., Kögel-Knabner, I., Ekschmitt, K., Flessa, H., Guggenberger, G., Matzner, E., & Marschner, B. (2007). SOM fractionation methods: Relevance to functional pools and to stabilization mechanisms. *Soil Biology and Biochemistry*, *39*(9), 2183–2207.

Wallenstein, M. D., & Burns, R. G. (2011). Ecology of extracellular enzyme activities and organic matter degradation in soil: A complex community-driven process. In *Methods of soil enzymology* (pp. 35–55). John Wiley & Sons, Ltd.

Wickham, H., Averick, M., Bryan, J., Chang, W., Mcgowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., & Hester, J. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686.

Wickner, W., & Schekman, R. (2005). Protein translocation across biological membranes. *Science*, *310*(5753), 1452–1456.

Xia, Y., Chrisman, P. A., Erickson, D. E., Liu, J., Liang, X., Londry, F. A., Yang, M. J., & Mcluckey, S. A. (2006). Implementation of ion/ion reactions in a quadrupole/time-of-flight tandem mass spectrometer. *Analytical Chemistry*, *78*(12), 4146–4154.

Zimmermann, M., Leifeld, J., Schmidt, M., Smith, P., & Fuhrer, J. (2007). Measured soil organic matter fractions can be related to pools in the RothC model. *European Journal of Soil Science*, *58*(3), 658–667.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Waibel, M., McDonnell, K., Tuohy, M., Shirran, S., Synowsky, S., Thornton, B., Paterson, E., Brennan, F., & Abram, F. (2023). Assessing the impact of interfering organic matter on soil metaproteomic workflow. *European Journal of Soil Science*, *74*(3), e13392. https://doi.org/10.1111/ejss.13392