universität wien

DISSERTATION / DOCTORAL THESIS

Titel der Dissertation /Title of the Doctoral Thesis

„Identification and quantification of soil necromass biomarkers“

verfasst von / submitted by

Erika Kristel Salas Hernández

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Doctor of Philosophy (PhD)

Wien, 2024 / Vienna 2024

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 794 685 437
Dissertationsgebiet lt. Studienblatt / field of study as it appears on the student record sheet:	Biologie
Betreut von / Supervisor:	Uni.-Prof. Dr. Wolfgang Wanek
Betreut von / Supervisor:	Assoz.-Prof. Dr. Christina Kaiser

Table of contents

Acknowledgements

First, I would like to thank my supervisors, Wolfgang Wanek and Christina Kaiser, who gave me the opportunity to come to Vienna and carry out my PhD on both their work groups. Thank you for your support and guidance during every stage of my PhD. I have learned so much from both of you and without your help I would not be where I am today. You motivate me to become a better scientist every day.

I want to thank everyone at TER for making me always feel welcomed into our very special group. We have an exceptional work group and we are always there for each other. I will always appreciate the laughs and memories I made with all of you along the way. A special thanks to Ludwig, who always took his time to teach me and help me in the lab when I most needed it. Thank you to Rahul and Xiaofei as well for always being there and assisting me when I needed help. Also, to all my office companions because laughs were always present, especially when we were supposed to be focusing.

I would like to give special thanks to Vicky and Julia, who have been there for me since my first day. The support you have given on the good days but specially on the hard days is incomparable. I will always treasure all our moments together, especially the conversations and the really nice trips. It seems incredible that this crazy adventure we are doing together brought me such close friends.

A mi familia, quienes siempre me han estado apoyando en todo momento. Tantos años de mucho trabajo, esfuerzo, dedicación y sacrificios no hubieran sido posibles sin tenerlos siempre a mi lado. Y en especial a Sebas, quien ha estado ahí todos los días, en los buenos y los malos momentos. Eres mi apoyo y sin tu ayuda no hubiera podido llegar hasta donde estoy.

And finally, to everyone who has helped me achieve this incredible life goal, I have no words to express how thankful I am. This work is dedicated to all of us.

Abstract

Microbial necromass is a key contributor to soil organic matter (SOM, >50%) as microbes assimilate plant carbon (C) into their biomass and microbial residues later become stabilized in soils. Microbial necromass biomarkers, particularly amino sugars (AS), are commonly used to trace microbial C contributions to SOM. Most studies have focused on using glucosamine (GlcN) to trace fungal C and muramic acid (MurA) for bacterial C; however, the knowledge of other potential necromass biomarkers such as neutral sugars (NS) and amino acids (AA), as well as the contribution of other microbial groups such as archaea is still lacking. In this PhD thesis, I identified and quantified necromass biomarkers in archaeal, bacterial, fungal and plant species. I developed and used derivatization protocols, ultra-high performance liquid chromatography (UHPLC) and Orbitrap mass spectrometry to identify and quantify AS, NS and AA necromass biomarkers in these taxonomic groups. Multivariate approaches alongside machine learning methods were used to determine the biomarker potential of these compounds. Our results showed that AS were the best discriminant biomarkers to separate all taxa. Alongside the use of GlcN and MurA as biomarkers specific for fungi and bacteria respectively, we identified new specific biomarkers: talosaminuronic acid (TalA) for Euryarchaeota, hydroxyproline for plants, and L,L-diaminopimelic acid for bacteria. By combining AS, NS and AA biomarkers we were able to extend the biomarker approach from separating bacterial and fungal necromass to separate microbes from plants. Moreover, the simultaneous measurement of MurA, TalA, hydroxyproline and hexosamines would allow to measure plant and microbial contributions to the continuum from litter decomposition to SOM formation through microbial processes. Overall, this PhD thesis expanded the repertoire of necromass biomarkers by including new specific biomarkers as well as combining AS, NS and AA biomarkers found in a wide variety of taxa.

Zusammenfassung

Mikrobielle Nekromasse trägt wesentlich zu dem bodeneigenen organischen Material (SOM, >50 %) bei, da Mikroben pflanzlichen Kohlenstoff (C) in ihre Biomasse aufnehmen und mikrobielle Rückstände später im Boden stabilisiert werden. Mikrobielle Nekromasse-Biomarker, insbesondere Aminozucker (AS), werden häufig verwendet, um mikrobielle C-Beiträge zum SOM zu verfolgen. Die meisten Studien konzentrierten sich auf die Verwendung von Glucosamin (GlcN) zur Verfolgung von Pilz-C und Muraminsäure (MurA) für Bakterien-C. Es mangelt jedoch noch an Kenntnissen über andere potenzielle Nekromasse-Biomarker wie Neutralzucker (NS) und Aminosäuren (AA) sowie über den Beitrag anderer mikrobieller Gruppen zur Nekromassebildung wie Archaeen. In diesem Dissertationsprojekt habe ich Nekromasse-Biomarker in Archaeen-, Bakterien-, Pilz- und Pflanzenarten identifiziert und quantifiziert. Ich habe Derivatisierungsprotokolle entwickelt und angewandt, basierend auf Ultrahochleistungsflüssigkeitschromatographie (UHPLC) und Orbitrap-Massenspektrometrie, um AS-, NS- und AA-Nekromasse-Biomarker in diesen taxonomischen Gruppen zu identifizieren und zu quantifizieren. Zur Bestimmung des Biomarkerpotenzials dieser Verbindungen wurden multivariate Ansätze sowie Methoden des maschinellen Lernens eingesetzt. Unsere Ergebnisse zeigten, dass AS die besten diskriminierenden Biomarker zur Trennung aller Taxa waren. Neben der Verwendung von GlcN und MurA als spezifische Biomarker für Pilze bzw. Bakterien identifizierten wir neue spezifische Biomarker: Talosaminuronsäure (TalA) für Euryarchaeota, Hydroxyprolin für Pflanzen und L,L-Diaminopimelinsäure für Bakterien. Durch die Kombination von AS-, NS- und AA-Biomarkern konnten wir den Biomarker-Ansatz von der Trennung von Bakterien- und Pilznekromasse auf die Trennung von Mikroben von Pflanzen erweitern. Darüber hinaus würde die gleichzeitige Messung von MurA, TalA, Hydroxyprolin und Hexosaminen die Messung pflanzlicher und mikrobieller Beiträge zum Kontinuum von der Streuzersetzung bis zur SOM-Bildung durch mikrobielle Prozesse ermöglichen. Insgesamt erweiterte dieses Dissertationsprojekt das Repertoire an Nekromasse-Biomarkern durch die Entwicklung neuer spezifischer Biomarker sowie die kombinierte Messung von AS-, NS- und AA-Biomarkern, die in einer Vielzahl von Taxa vorkommen.

Overview of manuscripts

Chapter II

Manuscript title: A rapid and sensitive assay to quantify amino sugars, neutral sugars and uronic acid necromass biomarkers using pre-column derivatization, ultra-high-performance liquid chromatography and high-resolution mass spectrometry

Authors: Erika Salas, Markus Gorfer, Dragana Bandian, Baorong Wang, Christina Kaiser, and Wolfgang Wanek.

Reference: Salas, E., Gorfer, M., Bandian, D., Wang, B., Kaiser, C., & Wanek, W. (2023). A rapid and sensitive assay to quantify amino sugars, neutral sugars and uronic acid necromass biomarkers using pre-column derivatization, ultra-high-performance liquid chromatography and high-resolution mass spectrometry. *Soil Biology and Biochemistry*, *177*, 108927. https://doi.org/10.1016/j.soilbio.2022.108927

Author contributions: E.S.: Conceptualization, Formal analysis, Writing – original draft. M.G.: Resources, Writing – review & editing. D.B.: Resources, Writing – review & editing. B.W.: Resources, Writing – review & editing. C.K.: Conceptualization, Writing – review & editing. W.W.: Conceptualization, Writing – review & editing.

Chapter III

Manuscript title: Reevaluation and novel insights into amino sugar and neutral sugar necromass biomarkers in archaea, bacteria, fungi, and plants

Authors: Erika Salas, Markus Gorfer, Dragana Bandian, Stephanie A. Eichorst, Hannes Schmidt, Julia Horak, Simon K.-M. R. Rittmann, Christa Schleper, Barbara Reischl, Thomas Pribasnig, Jan Jansa, Christina Kaiser, and Wolfgang Wanek.

Reference: Salas, E., Gorfer, M., Bandian, D., Eichorst, S. A., Schmidt, H., Horak, J., Rittman, S.K.-M.R., Schleper, C., Reischl, B., Pribasnig, T., Jansa, J., Kaiser, C., & Wanek, W. (2024). Reevaluation and novel insights into amino sugar and neutral sugar necromass biomarkers in archaea, bacteria, fungi, and plants. *Science of the Total Environment*, *906*, 167463. https://doi.org/10.1016/j.scitotenv.2023.167463

Author contributions: E.S.: Conceptualization, Formal analysis, Writing – original draft. M.G.: Resources, Writing – review & editing. D.B.: Resources, Writing – review & editing. S.A.E.: Resources, Writing – review & editing. H.S.: Resources, Writing – review & editing. J.H.: Resources, Writing – review & editing. S.K.-M.R.R.: Resources, Writing – review & editing. C.S.: Resources, Writing – review & editing. B.R.: Resources, Writing – review & editing. T.P.: Resources, Writing – review & editing. J.J.: Resources, Writing – review & editing. C.K.: Conceptualization, Writing – review & editing. W.W.: Conceptualization, Writing – review & editing.

Chapter IV

Manuscript title: Novel method for estimation of necromass deriving from soil archaea, bacteria, fungi, and plants by quantifying amino acid and amino sugar biomarkers in a single approach

Authors: Erika Salas, Markus Gorfer, Dragana Bandian, Stephanie A. Eichorst, Hannes Schmidt, Julia Horak, Simon K.-M. R. Rittmann, Christa Schleper, Barbara Reischl, Thomas Pribasnig, Jan Jansa, Christina Kaiser, and Wolfgang Wanek.

Reference: Draft manuscript

Author contributions: E.S.: Conceptualization, Formal analysis, Writing – original draft. M.G.: Resources, Writing – review & editing. D.B.: Resources, Writing – review & editing. S.A.E.: Resources, Writing – review & editing. H.S.: Resources, Writing – review & editing. J.H.: Resources, Writing – review & editing. S.K.-M.R.R.: Resources, Writing – review & editing. C.S.: Resources, Writing – review & editing. B.R.: Resources, Writing – review & editing. T.P.: Resources, Writing – review & editing. J.J.: Resources, Writing – review & editing. C.K.: Conceptualization, Writing – review & editing. W.W.: Conceptualization, Writing – review & editing.

Chapter I

General introduction, thesis outline and main goals

General introduction

1. Microbial necromass contribution to soil organic matter

Soil organic matter (SOM) is the largest pool of terrestrial carbon (C) and a key component of terrestrial biogeochemical cycles. The composition of SOM is both chemically and physically heterogeneous and complex, with its primary source being plant inputs at various stages of decomposition (Kallenbach et al., 2016). Soil microbial communities also contribute greatly to the SOM pool. The soil microbial carbon pump is an ecological framework that describes how microorganisms contribute to SOM. Accordingly, microbes assimilate plant C into new microbial biomass and, following death and lysis, part of their necromass and other microbial residues become stabilized in soils as part of the SOM pool (Liang et al., 2017; Zhu et al., 2020), particularly in the more stable mineral-associated organic matter (MAOM). This ultimately leads to the buildup of soil organic C stocks as a mix of plant and microbial derived residues (Chang et al., 2024).

Microbial necromass is largely composed of fragmented cell wall residues, extracellular polymeric substances (EPS), and cytoplasmic components from dead fungi and bacteria (Liang et al., 2019). Microbial metabolites, necromass and other residues can be stabilized in soils, specifically in the MAOM fraction, due to the physicochemical interactions between SOM and the mineral matrix that limits microbial decomposition (Buckeridge et al., 2020). Some of these interactions include sorption to mineral surfaces, incorporation into organo-mineral complexes, and occlusion into aggregate pore spaces at the microscale (Liang et al., 2017). Necromass components can also be stabilized via organo-organic (necromass-necromass) interactions regulated by cell chemistry (Buckeridge et al., 2020). These compounds thereby become inaccessible to microorganisms and contribute to the long-term storage of soil organic C (SOC) (Zhu et al., 2020).

The microbial buildup mechanism represents an important part of the ecosystem C cycle, since fungal and bacterial necromass are the main C-sources contributing to stable SOM pools (Liang et al., 2017, 2019). While living microbial biomass only contributes ~1% to the SOC (Xu et al., 2013), many studies have suggested that microbial necromass on average comprises 50% or more of the SOC (Creamer et al., 2019; Liang et al., 2019; B. Wang et al., 2021). Although other studies have pro-

duced lower estimates of microbial-derived SOM (Angst et al., 2021; Chang et al., 2024; Z. Zhang et al., 2022), microbial contribution to C storage in soils is significant.

Several qualitative and quantitative methods have been used to study the contribution of microbial necromass to SOM. Imaging techniques have shown microbial necromass associations to soil minerals (Buckeridge et al., 2022; Creamer et al., 2019; Miltner et al., 2012). A stoichiometric approach using isotope mixing models estimated the proportional contribution of plant residue versus microbial residues to MAOM (Chang et al., 2024). Molecular fingerprinting as well as mathematical models have proposed that microbial residues strongly dominate SOM (Fan & Liang, 2015; Liang et al., 2011; Miltner et al., 2012; Simpson et al., 2007). Nonetheless, the most common method used to quantify the microbial necromass contribution to SOM involves the use of microbial necromass biomarkers.

2. Microbial necromass biomarkers

After microbial cell death, lysis and decomposition, some organic cell components such as cytosolic proteins, ribosomes and other biopolymers are rapidly and efficiently reused by living microbes in close vicinity, while other residual compounds are retained in the soil matrix (Liang et al., 2019). In soils, cell wall residues from microbial necromass are commonly found in their polysaccharide form such as chitin or peptidoglycan (Joergensen, 2018). Mixtures of oligosaccharides, monosaccharides, amino acids, and oligopeptides are also found, which makes it challenging to trace their origin. Biomarkers are organic compounds with known phylogenetic origin that can be used as indicators for the source and transformation of other molecules or organisms (Amelung, 2003). Biomarkers of soil microbial necromass allow to distinguish between C found in microbial residues from non-microbial organic C (Liang et al., 2019). Cell wall fragments have a slow turnover and accumulate in soils, therefore biomarkers in microbial cell wall fragments are the most widely used ones.

Specific compounds such as amino sugars and amino acids have been used as biomarkers for the identification of bacterial and fungal communities in soils (Joergensen, 2018; Zhu et al., 2020). Some of these biomarkers have also been used for the quantification of the contribution of microbial derived C to SOM (Liang et al., 2019; Zhu et al., 2020), and as indicators of microbial necromass versus living microbial biomass contributions to SOM (Liang et al., 2017).

There are different classes of compounds that are used as microbial and plant necromass biomarkers:

i. Amino sugars

Amino sugars have been used to trace the contribution of microorganisms to SOM since plants do not produce them in significant quantities (Parsons, 2021). Amino sugars represent 5-12% of soil total nitrogen (N) and 2-5% of soil organic C (Joergensen, 2018; X. Zhang & Amelung, 1996). The amino sugars most used as microbial biomarkers are glucosamine (GlcN), galactosamine (GalN), mannosamine (ManN) and muramic acid (MurA) (Amelung, 2003; Joergensen, 2018).

GlcN is the most common amino sugar found in soils, contributing ~68% of total amino sugars in soil (Joergensen, 2018). It is the main component of fungal chitin where it is found in β-(1,4) links of N-acetylglucosamine (Parsons, 2021). It is also the major structural component of chitosan, where it is found in its deacetylated form. Additionally, GlcN is one of the main components of bacterial peptidoglycan, where it forms β-(1,4) links with N-acetylmuramic acid (Subedi et al., 2021), as well as archaeal pseudopeptidoglycan (a structural analog of bacterial peptidoglycan), where it is linked to N-acetyltalosaminuronic acid (Albers & Meyer, 2011).

MurA is the most specific amino sugar biomarker because it is found exclusively in the murein layers of peptidoglycan in bacterial cell walls (Vollmer et al., 2008). It can contribute up to 30% of murein mass in gram-positive and gram-negative bacteria (Joergensen, 2018). Using the assumption that MurA and GlcN occur at a molar ratio of 1:2 in bacteria, fungal GlcN can be estimated based on the contents of bacterial MurA (Engelking et al., 2007; Joergensen, 2018).

GalN is found in plant material (Liang et al., 2007), fungal cell walls (Beauvais & Latgé, 2018) and cell walls of rare archaea species (Albers & Meyer, 2011). ManN is found as integral part of sialic acids (Brigham et al., 2009), as part of capsular components in gram-negative bacteria (Lewis et al., 2016) and in EPS of gram-positive bacteria (Degeest et al., 2001). Both GalN and ManN are produced by fungal and bacterial groups, and therefore are not considered specific biomarkers (Joergensen, 2018). While the GalN content in soils can represent between 17% and 42% of total soil amino sugars, the ManN contents are usually minor (~4%) (Joergensen, 2018).

Other amino sugars have rarely been evaluated for their biomarker potential. For example, talosaminuronic acid (TalA) is found linked in its acetylated form to N-

acetylmuramic acid in the pseudopeptidoglycan of some Euryarchaeota species (Subedi et al., 2021).

ii. Neutral sugars

Neutral sugars are the most abundant organic monomers in the biosphere and the main energy sources for soil microbes as well as the primary input of organic C in soil. Neutral sugars mostly enter soils as plant polysaccharides that are decomposed by exoenzymes into oligo- and monosaccharides (Gunina & Kuzyakov, 2015). These compounds are soluble and easily available for microbes for maintenance metabolism, growth, and C storage (Bore et al., 2019).

Plants are considered the primary sources of neutral sugars in soils due to the decomposition of plant litter and rhizodeposits. Plant cellulose consists mainly of glucose (Glc, $\beta 1 \rightarrow 4$ linked), while hemicellulose contains various polymer-bound hexoses and pentoses, such as glucans, xylans, mannans, galactans and arabinogalactans (Angst et al., 2021). Soil microorganisms are considered a secondary source of soil sugars as they synthesize their carbohydrates from other C sources, such as plant litter (Gunina & Kuzyakov, 2015). Fungal cell walls also contain a large fraction of pure β–1,3-linked glucans and some bacteria also produce cellulose (Fesel & Zuccaro, 2016). Starch (Glc, $\alpha 1 \rightarrow 4/1 \rightarrow 6$ linked) serves as an energy storage compound in plants and contributes to plant-derived polymeric Glc-based sugars in soils. Glycogen (Glc, $\alpha 1 \rightarrow 4$ linked) is also form of energy storage compound in fungi, bacteria, and animals which can also contribute to Glc-based polymers in soils.

Hexoses account for most of the bound polymeric sugars in soil, with Glc being the most abundant one, followed by mannose (Man) and galactose (Gal) (Gunina & Kuzyakov, 2015). Due to the complex soil matrix, it is difficult to determine whether the origin of sugars is microbial or plant derived. Pentoses, particularly arabinose (Ara) and xylose (Xyl), mostly originate from plant residues, given that microorganisms preferentially use and incorporate hexoses into their biomass (Apostel et al., 2015; Gunina & Kuzyakov, 2015). Microbial polysaccharides therefore almost exclusively consist of hexoses, though plants also produce them (Parsons, 2021). The use of a hexose/pentose ratio of (galactose + mannose)/(arabinose + xylose) (GM/AX) has been proposed to determine the plant versus microbial origin of polysaccharides in SOM (Gunina & Kuzyakov, 2015; Oades, 1984).

Less common sugars and sugar derivatives such as fucose, rhamnose and uronic acids could potentially be used as biomarkers. For example, glucuronic acid (GluA) and galacturonic acid (GalA) are sugar acids and are unique to the exopolymers found outside the cell membrane (Fazio et al., 1982). GluA has been found in teichuronic acids from gram-positive microbes grown under phosphorus limitation (Bhavsar et al., 2004) and in some gram-negative microbial lipopolysaccharides (Fazio et al., 1982). GalA is the main component of pectin in higher plant cell walls (Xiao & Anderson, 2013).

iii. Amino acids

Bound amino acids represent an important part of soil N, comprising ~50% of soil total N (Hu et al., 2017; Schulten & Schnitzer, 1997). Free amino acid pools in soil are very small and highly dynamic, with turnover times that range from minutes to a few hours (Hu et al., 2018; Wanek et al., 2010). Amino acids in their polymeric state can enter soils through a variety of sources such as root turnover and litterfall, microbial exudation and turnover, and animal inputs, with plants supplying the largest inputs (Jones et al., 2002; Jones & Kielland, 2002).

L-amino acids are ubiquitous in nature while D-amino acids are rare or absent in eukaryotes. However, D-amino acids are important components of bacterial peptidoglycan, and they have been used as biomarkers for microbial residues in soil and sediments (Dworzanski et al., 2005; Hu et al., 2018; Radkov & Moe, 2014; Veuger et al., 2005). Peptide chains in bacterial peptidoglycan are composed of four or five amino acids consisting of L-alanine (L-Ala), D-alanine (D-Ala), and D-glutamine (Gln), and either lysine (Lys) in gram-positive bacterial or *meso* (D,L)-diaminopimelic acid (mDAP) in gram-negative bacteria (Vollmer et al., 2008). Archaeal pseudopeptidoglycan does not contain D-amino acid enantiomers and its peptide chain is composed of L-Glutamate, L-Ala and Lys attached to N-acetyltalosaminuronic acid (Subedi et al., 2021).

Other less common non-proteinogenic amino acids have the potential to be used as biomarkers. L,L-diaminopimelic acid (LLDAP) has been found instead of Lys in the peptidoglycan of Actinobacteria, such as Actinomyces and Streptomyces (Busse & Schumann, 1999; Leyh-Bouille et al., 1970). Similarly, hydroxyproline (Hyp) has been used as a tracer for plant residue N in soils, since its two major natural sources

are Hyp-rich glycoproteins in plant cell walls and animal collagen (Philben & Benner, 2013), collagen contributing little to bound amino acids in soils.

3. Upscaling biomarker data to microbial necromass

In order to quantify the contribution of microbial necromass to SOM using biomarkers, Appuhn and Joergensen (2006) developed conversion factors based on the GlcN and MurA content in biomass of selected cultured fungal and bacterial species. By using the conversion factors of 9 for fungal GlcN and 45 for bacterial MurA, soil GlcN and MurA contents can be converted into fungal and bacterial necromass C, respectively. Similarly, Liang et al. (2019) proposed conversion factors for N in bacterial residues to bacterial biomass in soils using the C/N ratio of the living soil microbial biomass. Using these conversion factors, the contribution of microbial necromass C to SOC has been estimated to vary approximately between 40-70% in cropland soils, 30-62% in grasslands and 15-35% in forest soils (Angst et al., 2021; Khan et al., 2016; Liang et al., 2019; B. Wang et al., 2021; Whalen et al., 2022).

Even though many studies have used these conversion factors, there are uncertainties regarding these based on assumptions that could over- or underestimate the real contribution of microbial necromass C to SOM. For example, the conversion factor for MurA to bacterial C is calculated based on the assumption that the gram-positive to gram-negative bacterial ratio in soils is 65:35 on a percentage basis (Appuhn & Joergensen, 2006; Fanin et al., 2019). Additionally, necromass conversion factors are derived from living cultured microbes (biomass) which is considered as the sum of cell envelope and cytoplasm; however, microbial necromass comprises not only cell wall and cytoplasmic residues, but also exo-enzymes and EPS (Joergensen, 2018; Joergensen & Wichern, 2018). It is also assumed that all amino sugars in soils have the same stabilization and recycling rates, as well as the same ratios in microbial biomass compared to microbial necromass (Whalen et al., 2022).

Even though necromass biomarkers have increasingly been used to study soil microbial communities and their necromass contribution to SOM (Joergensen, 2018; Joergensen & Wichern, 2018), there is much to learn about their presence in specific soil microbial groups and in plant litter, their transformation and (de)stabilization pathways and their pool sizes. Likewise, more accurate information on specific organisms would be useful to generate more precise conversion factors to determine microbial- and plant-derived C from necromass biomarker data.

4. Current methods in necromass biomarker research

The compounds targeted generally as necromass biomarkers are relatively common compounds that can occur in their free low-molecular weight (LMW) form and in the bound polymeric form in soils. Most of the LMW metabolites in soils are found at very low concentrations (Jansson & Hofmockel, 2018; Oburger & Jones, 2018), which means that their detection can be challenging. Multiple analytical techniques have been developed to study these microbial metabolites. These techniques must allow the quantification of trace level compounds in complex matrices; therefore, they require high sensitivity and selectivity (J. Wang et al., 2021). Although a multitude of compounds have been discovered, the identification and quantification of compounds can be strongly affected by the constant uptake by microbial communities, the loss due to formation of organo-mineral associations, the concurrent SOM breakdown, and the method used for sampling (Oburger & Jones, 2018). In contrast to LMW metabolites, the same compounds can be bound in microbial necromass compounds in polymeric form and accumulate in soils. In polymeric form they are not amenable to direct quantification by chromatography and mass spectrometry. Therefore, they need to be released first from bound to free form before analysis, which is performed by acid (6 M HCl or 4 M TFA) hydrolysis before becoming accessible for quantitative analysis. By acid hydrolysis the biomarkers are also released in greater amounts omitting the need for ultra-trace analysis.

Many studies have used derivatization methods to trace amino sugar, neutral sugar and amino acid biomarkers in soils (Dai et al., 2010; Gonzalez et al., 2020; Salazar et al., 2012). Derivatization methods consist of a chemical reaction where a target compound is modified by adding a derivatization reagent, creating a newly derivatized compound with novel properties (Qi et al., 2014). Derivatization has been used in combination with high throughput methods, such as liquid chromatography (LC) or gas chromatography (GC) coupled to mass spectrometry (MS), to improve the selectivity and separation of compounds, to enhance ionization efficiency, and to facilitate the separation of structural isomers (Qi et al., 2014).

Finally, multivariate machine learning (ML) approaches have been used to detect biomarkers in clinical metabolomics and biomonitoring studies (Cambon et al., 2023; Dias-Audibert et al., 2020; Hirakawa et al., 2022; J. D. Zhang et al., 2023). Similar statistical approaches have not yet been used to investigate the potential of necro-

mass biomarkers in soils. Multivariate ML methods provide several advantages for the detection and analysis of biomarkers since they can handle large, non-linear, and heterogenous datasets (Pichler & Hartig, 2023). Therefore, combining high-resolution analytical techniques with multivariate unsupervised and supervised ML approaches would help advance our understanding of soil necromass biomarkers as well as improve the detection of these compounds in complex matrixes such as soil environments.

Thesis outline and main objectives

The use of necromass biomarker data can help elucidate changes in microbial community composition over time and the contribution of microbial necromass to soil organic matter. There still are many uncertainties in the interpretation of necromass biomarkers as well as limited biomarker data for many representative soil groups and plants.

The main goal of this PhD thesis was to identify and quantify necromass biomarkers in representative groups of soil archaea, bacteria, fungi, and plants. For this purpose, I analysed (potential) amino sugar, neutral sugar, and amino acid necromass biomarkers found in the biomass of the previously mentioned taxonomic groups and divided this PhD thesis into three subprojects.

In the first subproject (Chapter II), I developed a method to separate, identify and quantify amino sugar and neutral sugar compounds in a single assay. I used 1-phenyl-3-methyl-5-pyrazolone (PMP) derivatization coupled to reversed-phase ultra-high performance liquid chromatography (UHPLC) and high-resolution Orbitrap mass spectrometry (MS). The method was validated and successfully applied in both microbial and plant biomass as well as in soil samples.

In the second subproject (Chapter III), I identified and quantified amino sugar and neutral sugar necromass biomarkers in the biomass of representative soil groups of archaea, bacteria, fungi, and plants. In this subproject, I utilized the PMP derivatization method developed in the first subproject. I evaluated the biomarker potential of both amino sugars and neutral sugars, as well as proposed new conversion factors to allow the estimation of microbial necromass residue C from amino sugar biomass contents.

Finally, in the third subproject (Chapter IV), I analysed amino acid and amino sugar biomarkers in the biomass of the same representative soil groups used in the second subproject. In this subproject, I used 6-aminoquinolyl-N-hydroxysuccinimidyl carbamate (AQC) derivatization and reversed-phase UHPLC coupled to Orbitrap MS. Additionally, I used multivariate unsupervised and supervised machine learning approaches to evaluate the biomarker potential of amino acids and amino sugars. By including both microbial and plant necromass biomarkers, this approach will further help understand the plant versus microbial contribution to the organic matter continuum from litter decomposition to soil organic matter formation via microbial processing.

References

Albers, S.-V., & Meyer, B. H. (2011). The archaeal cell envelope. *Nature Reviews Microbiology*, *9*(6), 414–426. https://doi.org/10.1038/nrmicro2576

Amelung, W. (2003). Nitrogen biomarkers and their fate in soil. *Journal of Plant Nutrition and Soil Science*, *166*(6), 677–686. https://doi.org/10.1002/jpln.200321274

Angst, G., Mueller, K. E., Nierop, K. G. J., & 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. https://doi.org/10.1016/j.soilbio.2021.108189

Apostel, C., Dippold, M., & Kuzyakov, Y. (2015). Biochemistry of hexose and pentose transformations in soil analyzed by position-specific labeling and 13C-PLFA. *Soil Biology and Biochemistry*, *80*, 199–208. https://doi.org/10.1016/j.soilbio.2014.09.005

Appuhn, A., & Joergensen, R. (2006). Microbial colonisation of roots as a function of plant species. *Soil Biology and Biochemistry*, *38*(5), 1040–1051. https://doi.org/10.1016/j.soilbio.2005.09.002

Beauvais, A., & Latgé, J.-P. (2018). Special Issue: Fungal Cell Wall. *Journal of Fungi*, *4*(3), 91. https://doi.org/10.3390/jof4030091

Bhavsar, A. P., Erdman, L. K., Schertzer, J. W., & Brown, E. D. (2004). Teichoic Acid Is an Essential Polymer in Bacillus subtilis That Is Functionally Distinct from Teichuronic Acid. *Journal of Bacteriology*, *186*(23), 7865–7873. https://doi.org/10.1128/JB.186.23.7865-7873.2004

Bore, E. K., Kuzyakov, Y., & Dippold, M. A. (2019). Glucose and ribose stabilization in soil: Convergence and divergence of carbon pathways assessed by position-specific labeling. *Soil Biology and Biochemistry*, *131*, 54–61. https://doi.org/10.1016/j.soilbio.2018.12.027

Brigham, C., Caughlan, R., Gallegos, R., Dallas, M. B., Godoy, V. G., & Malamy, M. H. (2009). Sialic acid (N-acetyl neuraminic acid) utilization by bacteroides fragilis requires a novel N-acetyl mannosamine epimerase. *Journal of Bacteriology*, *191*(11), 3629–3638. https://doi.org/10.1128/JB.00811-08

Buckeridge, K. M., Creamer, C., & Whitaker, J. (2022). Deconstructing the microbial necromass continuum to inform soil carbon sequestration. *Functional Ecology*, 1–15. https://doi.org/10.1111/1365-2435.14014

Buckeridge, K. M., La Rosa, A. F., Mason, K. E., Whitaker, J., McNamara, N. P., Grant, H. K., & Ostle, N. J. (2020). Sticky dead microbes: Rapid abiotic retention of microbial necromass in soil. *Soil Biology and Biochemistry*, *149*, 107929. https://doi.org/10.1016/j.soilbio.2020.107929

Busse, H.-J., & Schumann, P. (1999). Polyamine profiles within genera of the class Actinobacteria with LL-diaminopimelic acid in the peptidoglycan. *International Journal of Systematic and Evolutionary Microbiology*, *49*(1), 179–184. https://doi.org/10.1099/00207713-49-1-179

Cambon, M. C., Trillat, M., Lesur-Kupin, I., Burlett, R., Chancerel, E., Guichoux, E., Piouceau, L., Castagneyrol, B., Le Provost, G., Robin, S., Ritter, Y., Van Halder, I., Delzon, S., Bohan, D. A., & Vacher, C. (2023). Microbial biomarkers of tree water status for next-generation biomonitoring of forest ecosystems. *Molecular Ecology*, *32*(22), 5944–5958. https://doi.org/10.1111/mec.17149

Chang, Y., Sokol, N. W., Van Groenigen, K. J., Bradford, M. A., Ji, D., Crowther, T. W., Liang, C., Luo, Y., Kuzyakov, Y., Wang, J., & Ding, F. (2024). A stoichiometric approach to estimate sources of mineral-associated soil organic matter. *Global Change Biology*, *30*(1), e17092. https://doi.org/10.1111/gcb.17092

Creamer, C. A., Foster, A. L., Lawrence, C., McFarland, J., Schulz, M., & Waldrop, M. P. (2019). Mineralogy dictates the initial mechanism of microbial necromass association. *Geochimica et Cosmochimica Acta*, *260*, 161–176. https://doi.org/10.1016/j.gca.2019.06.028

Dai, J., Wu, Y., Chen, S., Zhu, S., Yin, H., Wang, M., & Tang, J. (2010). Sugar compositional determination of polysaccharides from Dunaliella salina by modified RP-HPLC method of precolumn derivatization with 1-phenyl-3-methyl-5-pyrazolone. *Carbohydrate Polymers*, *82*(3), 629–635. https://doi.org/10.1016/j.carbpol.2010.05.029

Degeest, B., Vaningelgem, F., Laws, A. P., & De Vuyst, L. (2001). UDP-N-acetylglucosamine 4-epimerase activity indicates the presence of N-acetylgalactosamine in exopolysaccharides of Streptococcus thermophilus

strains. *Applied and Environmental Microbiology*, *67*(9), 3976–3984. https://doi.org/10.1128/AEM.67.9.3976-3984.2001

Dias-Audibert, F. L., Navarro, L. C., De Oliveira, D. N., Delafiori, J., Melo, C. F. O. R., Guerreiro, T. M., Rosa, F. T., Petenuci, D. L., Watanabe, M. A. E., Velloso, L. A., Rocha, A. R., & Catharino, R. R. (2020). Combining Machine Learning and Metabolomics to Identify Weight Gain Biomarkers. *Frontiers in Bioengineering and Biotechnology*, *8*, 6. https://doi.org/10.3389/fbioe.2020.00006

Dworzanski, J. P., Tripathi, A., Snyder, A. P., Maswdeh, W. M., & Wick, C. H. (2005). Novel biomarkers for Gram-type differentiation of bacteria by pyrolysis–gas chromatography–mass spectrometry. *Journal of Analytical and Applied Pyrolysis*, *73*(1), 29–38. https://doi.org/10.1016/j.jaap.2004.09.003

Engelking, B., Flessa, H., & Joergensen, R. G. (2007). Shifts in amino sugar and ergosterol contents after addition of sucrose and cellulose to soil. *Soil Biology and Biochemistry*, *39*(8), 2111–2118. https://doi.org/10.1016/j.soilbio.2007.03.020

Fan, Z., & Liang, C. (2015). Significance of microbial asynchronous anabolism to soil carbon dynamics driven by litter inputs. *Scientific Reports*, *5*(1), 9575. https://doi.org/10.1038/srep09575

Fanin, N., Kardol, P., Farrell, M., Nilsson, M.-C., Gundale, M. J., & Wardle, D. A. (2019). The ratio of Gram-positive to Gram-negative bacterial PLFA markers as an indicator of carbon availability in organic soils. *Soil Biology and Biochemistry*, *128*, 111–114. https://doi.org/10.1016/j.soilbio.2018.10.010

Fazio, S. A., Uhlinger, D. J., Parker, J. H., & White, D. C. (1982). Estimations of Uronic Acids as Quantitative Measures of Extracellular and Cell Wall Polysaccharide Polymers from Environmental Samples. *Applied and Environmental Microbiology*, *43*(5), 1151–1159. https://doi.org/10.1128/aem.43.5.1151-1159.1982

Fesel, P. H., & Zuccaro, A. (2016). β-glucan: Crucial component of the fungal cell wall and elusive MAMP in plants. *Fungal Genetics and Biology*, 90, 53-60. https://doi.org/10.1016/j.fgb.2015.12.004

Gonzalez, N. M., Fitch, A., & Al-Bazi, J. (2020). Development of a RP-HPLC method for determination of glucose in Shewanella oneidensis cultures utilizing 1-phenyl-3-methyl-5-pyrazolone derivatization. *PLOS ONE*, *15*(3), e0229990. https://doi.org/10.1371/journal.pone.0229990

Gunina, A., & Kuzyakov, Y. (2015). Sugars in soil and sweets for microorganisms: Review of origin, content, composition and fate. *Soil Biology and Biochemistry*, *90*, 87–100. https://doi.org/10.1016/j.soilbio.2015.07.021

Hirakawa, Y., Yoshioka, K., Kojima, K., Yamashita, Y., Shibahara, T., Wada, T., Nangaku, M., & Inagi, R. (2022). Potential progression biomarkers of diabetic kidney disease determined using comprehensive machine learning analysis of non-targeted metabolomics. *Scientific Reports*, *12*(1), 16287. https://doi.org/10.1038/s41598-022-20638-1

Hu, Y., Zheng, Q., & Wanek, W. (2017). Flux Analysis of Free Amino Sugars and Amino Acids in Soils by Isotope Tracing with a Novel Liquid Chromatography/High Resolution Mass Spectrometry Platform. *Analytical Chemistry*, *89*(17), 9192–9200. https://doi.org/10.1021/acs.analchem.7b01938

Hu, Y., Zheng, Q., Zhang, S., Noll, L., & Wanek, W. (2018). Significant release and microbial utilization of amino sugars and D-amino acid enantiomers from microbial cell wall decomposition in soils. *Soil Biology and Biochemistry*, *123*, 115–125. https://doi.org/10.1016/j.soilbio.2018.04.024

Jansson, J. K., & Hofmockel, K. S. (2018). The soil microbiome—From metagenomics to metaphenomics. *Current Opinion in Microbiology*, *43*, 162–168. https://doi.org/10.1016/j.mib.2018.01.013

Joergensen, R. G. (2018). Amino sugars as specific indices for fungal and bacterial residues in soil. *Biology and Fertility of Soils*, *54*(5), 559–568. https://doi.org/10.1007/s00374-018-1288-3

Joergensen, R. G., & Wichern, F. (2018). Alive and kicking: Why dormant soil microorganisms matter. *Soil Biology and Biochemistry*, *116*, 419–430. https://doi.org/10.1016/j.soilbio.2017.10.022

Jones, D. L., & Kielland, K. (2002). Soil amino acid turnover dominates the nitrogen flux in permafrost-dominated taiga forest soils. *Soil Biology*, *34*, 209–219. https://doi.org/10.1016/S0038-0717(01)00175-4

Jones, D. L., Owen, A. G., & Farrar, J. F. (2002). Simple method to enable the high resolution determination of total free amino acids in soil solutions and soil extracts. *Soil Biology and Biochemistry*, *34*(12), 1893–1902. https://doi.org/10.1016/S0038-0717(02)00203-1

Kallenbach, C. M., Frey, S. D., & Grandy, A. S. (2016). Direct evidence for microbial-derived soil organic matter formation and its ecophysiological controls. *Nature Communications*, *7*(1), 13630. https://doi.org/10.1038/ncomms13630

Khan, K. S., Mack, R., Castillo, X., Kaiser, M., & Joergensen, R. G. (2016). Microbial biomass, fungal and bacterial residues, and their relationships to the soil organic matter C/N/P/S ratios. *Geoderma*, *271*, 115–123. https://doi.org/10.1016/j.geoderma.2016.02.019

Lewis, A. L., Robinson, L. S., Agarwal, K., & Lewis, W. G. (2016). Discovery and characterization of de novo sialic acid biosynthesis in the phylum Fusobacterium. *Glycobiology*, *26*(10), 1107–1119. https://doi.org/10.1093/glycob/cww068

Leyh-Bouille, M., Bonaly, R., Ghuysen, J.-M., Tinelli, R., & Tipper, D. (1970). LL-Diaminopimelic acid containing peptidoglycans in walls of Streptomyces sp. And of Clostridium perfringens (Type A). *Biochemistry*, *9*(15), 2944–2952. https://doi.org/10.1021/bi00817a002

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. https://doi.org/10.1111/gcb.14781

Liang, C., Cheng, G., Wixon, D. L., & Balser, T. C. (2011). An Absorbing Markov Chain approach to understanding the microbial role in soil carbon stabilization. *Biogeochemistry*, *106*(3), 303–309. https://doi.org/10.1007/s10533-010-9525-3

Liang, C., Schimel, J. P., & Jastrow, J. D. (2017). The importance of anabolism in microbial control over soil carbon storage. *Nature Microbiology*, *2*(8), 17105. https://doi.org/10.1038/nmicrobiol.2017.105

Liang, C., Zhang, X., Rubert, K. F., & Balser, T. C. (2007). Effect of plant materials on microbial transformation of amino sugars in three soil microcosms. *Biology and Fertility of Soils*, *43*(6), 631–639. https://doi.org/10.1007/s00374-006-0142-1

Miltner, A., Bombach, P., Schmidt-Brücken, B., & Kästner, M. (2012). SOM genesis: Microbial biomass as a significant source. *Biogeochemistry*, *111*(1–3), 41–55. https://doi.org/10.1007/s10533-011-9658-z

Oades, J. M. (1984). Soil organic matter and structural stability: Mechanisms and implications for management. *Plant and Soil*, *76*, 319–337. https://doi.org/10.1007/BF02205590

Oburger, E., & Jones, D. L. (2018). Sampling root exudates – Mission impossible? *Rhizosphere*, *6*, 116–133. https://doi.org/10.1016/j.rhisph.2018.06.004

Parsons, J. W. (2021). Chemistry and Distribution of Amino Sugars in Soil and Soil Organisms. In *Soil Biochemistry* (pp. 197–228). CRC Press.

Philben, M., & Benner, R. (2013). Reactivity of hydroxyproline-rich glycoproteins and their potential as biochemical tracers of plant-derived nitrogen. *Organic Geochemistry*, *57*, 11–22. https://doi.org/10.1016/j.orggeochem.2013.01.003

Pichler, M., & Hartig, F. (2023). Machine learning and deep learning—A review for ecologists. *Methods in Ecology and Evolution*, *14*(4), 994–1016. https://doi.org/10.1111/2041-210X.14061

Qi, B.-L., Liu, P., Wang, Q.-Y., Cai, W.-J., Yuan, B.-F., & Feng, Y.-Q. (2014). Derivatization for liquid chromatography-mass spectrometry. *Trends in Analytical Chemistry*, *59*, 121–132. https://doi.org/10.1016/j.trac.2014.03.013

Radkov, A. D., & Moe, L. A. (2014). Bacterial synthesis of d-amino acids. *Applied Microbiology and Biotechnology*, *98*(12), 5363–5374. https://doi.org/10.1007/s00253-014-5726-3

Salazar, C., Armenta, J. M., & Shulaev, V. (2012). An UPLC-ESI-MS/MS assay using 6-aminoquinolyl-N-hydroxysuccinimidyl carbamate derivatization for targeted amino acid analysis: Application to screening of arabidopsis thaliana mutants. *Metabolites*, *2*(3), 398–428. https://doi.org/10.3390/metabo2030398

Schulten, H.-R., & Schnitzer, M. (1997). The chemistry of soil organic nitrogen: A review. *Biology and Fertility of Soils*, *26*(1), 1–15. https://doi.org/10.1007/s003740050335

Simpson, A. J., Simpson, M. J., Smith, E., & Kelleher, B. P. (2007). Microbially Derived Inputs to Soil Organic Matter: Are Current Estimates Too Low? *Environmental Science & Technology*, *41*(23), 8070–8076. https://doi.org/10.1021/es071217x

Subedi, B. P., Martin, W. F., Carbone, V., Duin, E. C., Cronin, B., Sauter, J., Schofield, L. R., Sutherland-Smith, A. J., & Ronimus, R. S. (2021). Archaeal pseudomurein and bacterial murein cell wall biosynthesis share a common evolutionary ancestry. *FEMS Microbes*, *2*, xtab012. https://doi.org/10.1093/femsmc/xtab012

Veuger, B., Middelburg, J. J., Boschker, H. T. S., & Houtekamer, M. (2005). Analysis of 15 N incorporation into D-alanine: A new method for tracing nitrogen uptake

by bacteria. *Limnology and Oceanography: Methods*, *3*(5), 230–240. https://doi.org/10.4319/lom.2005.3.230

Vollmer, W., Blanot, D., & De Pedro, M. A. (2008). Peptidoglycan structure and architecture. *FEMS Microbiology Reviews*, *32*(2), 149–167. https://doi.org/10.1111/j.1574-6976.2007.00094.x

Wanek, W., Mooshammer, M., Blöchl, A., Hanreich, A., & Richter, A. (2010). Determination of gross rates of amino acid production and immobilization in decomposing leaf litter by a novel 15N isotope pool dilution technique. *Soil Biology and Biochemistry*, *42*(8), 1293–1302. https://doi.org/10.1016/j.soilbio.2010.04.001

Wang, B., An, S., Liang, C., Liu, Y., & Kuzyakov, Y. (2021). Microbial necromass as the source of soil organic carbon in global ecosystems. *Soil Biology and Biochemistry*, *162*, 108422. https://doi.org/10.1016/j.soilbio.2021.108422

Wang, J., Zhao, J., Nie, S., Xie, M., & Li, S. (2021). Mass spectrometry for structural elucidation and sequencing of carbohydrates. *TrAC Trends in Analytical Chemistry*, *144*, 116436. https://doi.org/10.1016/j.trac.2021.116436

Whalen, E. D., Grandy, A. S., Sokol, N. W., Keiluweit, M., Ernakovich, J., Smith, R. G., & Frey, S. D. (2022). Clarifying the evidence for microbial- and plant-derived soil organic matter, and the path toward a more quantitative understanding. *Global Change Biology*, *00*, 1–19. https://doi.org/10.1111/gcb.16413

Xiao, C., & Anderson, C. T. (2013). Roles of pectin in biomass yield and processing for biofuels. *Frontiers in Plant Science*, *4*, 67. https://doi.org/10.3389/fpls.2013.00067

Xu, X., Thornton, P. E., & Post, W. M. (2013). A global analysis of soil microbial biomass carbon, nitrogen and phosphorus in terrestrial ecosystems. *Global Ecology and Biogeography*, *22*(6), 737–749. https://doi.org/10.1111/geb.12029

Zhang, J. D., Xue, C., Kolachalama, V. B., & Donald, W. A. (2023). Interpretable machine learning on metabolomics data reveals biomarkers for Parkinson's disease. *ACS Central Science*, *9*(5), 1035–1045. https://doi.org/10.1021/acscentsci.2c01468

Zhang, X., & Amelung, W. (1996). Gas chromatographic determination of muramic acid, glucosamine, mannosamine, and galactosamine in soils. *Soil Biology and Biochemistry*, *28*(9), 1201–1206. https://doi.org/10.1016/0038-0717(96)00117-4

Zhang, Z., Kaye, J. P., Bradley, B. A., Amsili, J. P., & Suseela, V. (2022). Cover crop functional types differentially alter the content and composition of soil organic carbon in particulate and mineral-associated fractions. *Global Change Biology*, *28*(19), 5831–5848. https://doi.org/10.1111/gcb.16296

Zhu, X., Jackson, R. D., DeLucia, E. H., Tiedje, J. M., & Liang, C. (2020). The soil microbial carbon pump: From conceptual insights to empirical assessments. *Global Change Biology*, *26*(11), 6032–6039. https://doi.org/10.1111/gcb.15319

Chapter II

A rapid and sensitive assay to quantify amino sugars, neutral sugars and uronic acid necromass biomarkers using pre-column derivatization, ultra-high-performance liquid chromatography and high-resolution mass spectrometry

A rapid and sensitive assay to quantify amino sugars, neutral sugars and uronic acid necromass biomarkers using pre-column derivatization, ultra-high-performance liquid chromatography and high-resolution mass spectrometry

Erika Salas[a,b*], Markus Gorfer[c], Dragana Bandian[c], Baorong Wang[d], Christina Kaiser[a], and Wolfgang Wanek[a]

[a]Division of Terrestrial Ecosystem Research, Centre for Microbiology and Environmental Systems Science, University of Vienna, Vienna, Austria
[b]Doctoral School in Microbiology and Environmental Science, University of Vienna, Vienna, Austria
[c]AIT Austrian Institute of Technology GmbH, Bioresources, Tulln, Austria
[d]State Key Laboratory of Soil Erosion and Dryland Farming on the Loess Plateau, Institute of Soil and Water Conservation, Chinese Academy of Science and Ministry of Water Resources, Yangling, China

Abstract

Microbial necromass comprises a large fraction of soil organic matter (SOM) due to the accumulation and stabilization of microbial residues from dead archaea, bacteria and fungi. Amino sugars, neutral sugars and uronic acids have been used as microbial necromass biomarkers to trace the origin and composition of microbial residues in the SOM pool. Due to the structural complexity of sugars, derivatization reactions and high-throughput analytical methods are required to separate and quantify these sugar-related compounds. Our aim was to develop a rapid and sensitive assay to measure amino sugar, neutral sugar and uronic acid compounds using pre-column 1-phenyl-3-methyl-5-pyrazolone (PMP) derivatization. PMP-derivatives were separated and quantified via reversed phase (RP) ultra-high-performance liquid chromatography (UPLC) coupled to high-resolution Orbitrap mass spectrometry (MS). The method was validated and applied on hydrolyzed peptidoglycans and the biomass of archaeal, bacterial, fungal and plant species, as well as with soils. This developed PMP method allowed the separation and quantification of 18 sugar-related compounds, including four amino sugars, three N-acetyl amino sugars, eight neutral sugars, and three uronic acids within 20 min. This PMP method showed a precision for isotope enrichment detection of 0.03-0.05 atom % ^{13}C for D-glucose and D-glucosamine. This is the first time talosaminuronic acid (deriving from archaeal pseudopeptidoglycan) was identified and quantified using PMP derivatization. The application of this novel PMP method on pure hydrolyzed biomass and soils, showed the successful chromatographic and mass spectrometric separation and quantification of amino sugar, neutral sugar and uronic acid compounds. A multivariate analysis using these sugar-related PMP derivatives showed a clustering of the species according to their respective taxonomic group (archaea, gram-positive bacteria, gram-negative bacteria, fungi and plants). The modified PMP method can be applied to identify and quantify soil microbial necromass biomarkers, as well as their contribution to SOM. The sensitive isotope tracer detection allows tracing isotopically labeled materials into necromass biomarkers in SOM pools.

Key words

Microbial necromass biomarkers, amino sugars, neutral sugars, uronic acids, PMP derivatization, UPLC-Orbitrap MS, isotopic enrichment precision

1. Introduction

Soil organic matter (SOM) is largely made up of microbial necromass, including fragmented cell walls, cytoplasmic components, and extracellular polymeric substances from dead archaea, bacteria and fungi (Liang et al., 2019). Fungal and bacterial necromass comprises on average 50% of soil organic carbon (SOC) (Wang B. et al., 2021a), while living microbial biomass only contributes approximately 1% to SOC (Xu et al., 2013). The stabilization of microbial necromass residues contributes to the buildup of SOC stocks through the interaction with mineral surfaces, the incorporation into organo-mineral complexes, and occlusion into soil pore spaces and in aggregates (Liang et al., 2017). Biomarkers are used to trace microbial necromass concentrations and dynamics in soils since they allow to distinguish carbon found in microbial residues from non-microbial organic C (Liang et al., 2019; Whalen et al., 2022), and to distinguish between bacterial and fungal contributions to soil organic matter.

Different classes of organic compounds have been used as microbial necromass biomarkers, such as amino sugars and neutral sugars. Amino sugars are important structural components of archaeal, bacterial and fungal cell walls, in the form of (pseudo)peptidoglycan and chitin, respectively (Engelking et al., 2007; Joergensen, 2018). Once part of the soil microbial necromass, bound amino sugars from cell wall fragments are stabilized and contribute significantly to the soil organic matter (SOM) accrual (Liang et al., 2017). Glucosamine (GlcN) is the most common amino sugar found in soils, contributing up to 68% of total amino sugars in soils (Joergensen, 2018). Muramic acid (MurN), an ether of lactic acid and glucosamine, is the most specific amino sugar biomarker, which is found exclusively in the murein layers of bacterial peptidoglycan (Vollmer et al., 2008). Galactosamine (GalN) and mannosamine (ManN) are produced by both fungal and bacterial groups, and therefore they are not considered specific biomarkers (Joergensen, 2018).

Neutral sugars are the dominant low molecular weight (LMW) compounds found in soils and they mainly derive from plants (Gunina & Kuzyakov, 2015). Cellulose consists primarily of glucose (Glc) monomers, while hemicellulose contains various polymer-bound hexoses and pentoses. Starch and glycogen from plants, animals, fungi and bacteria also contribute to Glc-based polymers in soil (Bore et al., 2019; Drigo et al., 2012; German et al., 2011). While both, pentoses and hexoses, are

produced and utilized by plants, microorganisms preferentially use and incorporate hexoses into their biomass (Guggenberger et al., 1994). Consequently, the pentose:hexose ratio is indicative of the plant versus microbial origin of polysaccharides in SOM (Gunina & Kuzyakov, 2015; Ludwig et al., 2015).

Uronic acids, such as glucuronic acid (GluA) and galacturonic acid (GalA), are sugar acids that can potentially be used as biomarkers. GluA has been found in some lipopolysaccharides of Gram-negative microbes (Fazio et al., 1982) and in teichuronic acids of Gram-positive microbes grown under phosphorus limitation (Bhavsar et al., 2004). GalA is also a main structural component of plant pectins (Xiao & Anderson, 2013).

Due to the structural complexity of sugar-containing compounds, high-throughput identification and characterization approaches of pure microbial isolates and of complex soil communities have remained a challenge. The combination of ultra-high performance separation techniques, such as reversed-phase high-performance liquid chromatography (RP-HPLC), with high-resolution mass spectrometry (HRMS) have allowed the analysis of compound mixtures with isobaric structures (Wang B. et al., 2021b). Derivatization using 1-phenyl-3-methyl-5-pyrazolone (PMP) has been widely used for the analysis of polymeric carbohydrates, because it rapidly reacts with the aldehyde moiety of reducing sugars under basic conditions (Honda et al., 1989; Strydom, 1994). This reaction forms a sugar-derivative with two PMP molecules (bis-PMP derivatives) and no stereoisomers, which can be readily separated using LC-MS techniques.

While PMP derivatization has proven useful in the analysis of sugars (Dai et al., 2010; Gonzalez et al., 2020; Xia et al., 2017), there have been many variations in the reaction duration and retention characteristic of the different sugar derivatives, and this technique was not adopted for concurrent analysis of the multitude of neutral sugars, amino sugars and uronic acids. Our aim was to develop a modified PMP derivatization method that allows the concurrent separation and mass spectrometric detection of amino sugars, neutral sugars and uronic acids in a single assay. We tested the applicability of this method in one archaea species [*Methanobacterium* sp., also seeking for the biomarker talosaminuronic acid (TalA)], six bacterial species (*Escherichia coli*, *Bacillus subtilis*, *Micrococcus luteus*, *Streptomyces* sp., *Staphylococcus aureus*, and *Synechococcus* sp.), two fungal species (*Clonostachys rosea*, *Fusarium fujikuroi*), and two plant species (*Anthriscus cerefolium*, *Malva*

silvestris). We also tested the applicability of the method on forest, grassland and cropland soils with contrasting edaphic properties. To the best of our knowledge, this is the first time such a wide range of necromass biomarkers, including four amino sugars, three N-acetyl amino sugars, eight neutral sugars, and three uronic acids, were separated simultaneously in a single run within 20 min, including the ability of HRMS for sensitive isotope label quantification.

2. Materials and methods

2.1 Standards and reagents

Amino sugar, neutral sugar and uronic acid standards [arabinose (Ara), fucose (Fuc), galactosamine (GalN), galactose (Gal), galacturonic acid (GalA), glucosamine (GlcN), glucose (Glc), glucuronic acid (GluA), mannosamine (ManN), mannose (Man), muramic acid (MurN), *N*-acetylgalactosamine (NAcGal), *N*-acetylglucosamine (NAcGlc), *N*-acetylmannosamine (NAcMan), rhamnose (Rha), ribose (Rib), and xylose (Xyl)], 1-methyl-3-phenyl-2-pyralozone (PMP), ammonium formate solution, ammonia solution, formic acid, hydrochloric acid and acetonitrile (hypergrade for LC-MS) solution were purchased from Sigma Aldrich (St. Louis, MO). Peptidoglycan (from *Methanobacterium* sp., *B. subtilis*, *M. luteus*, *Streptomyces* sp., and *S. aureus*), and lyophilized cells (from *M. luteus*, *Synechococcus* sp., and *E. coli*) were also obtained from Sigma.

Clonostachys rosea EBL1201_S1 was isolated from strawberry roots. Identification is based on the markers ITS/LSU (GenBank: OP476509) and β-tubulin (GenBank: OP485759) as outlined in Gorfer et al. (2011). *Fusarium fujikuroi* DSM 893 was obtained from the DSMZ German Collection of Microorganisms and Cell Cultures GmbH. Biomass of fungi for biomarker analysis was produced in 50 ml malt extract broth. Liquid cultures were incubated for two days at room temperature (~22 °C) in an orbital shaker at 180 rpm. Mycelium was harvested by vacuum filtration over GN-6 Metricel filters (Pall). The mycelium was rinsed with sterile distilled water and lyophilized.

2.2 Standard solutions

Stock solutions of all standards were prepared at 20 mM and mixed to give one combined standard (1 mM for each analyte). A serial dilution of the calibration standards that ranged from 20 nM to 25 µM was used for method validation. Dilute

single standards were used prior for retention time identification of isobaric compounds.

2.3 Derivatization with PMP

PMP solution was prepared in methanol to a final concentration of 0.5 M. For the derivatization reaction, 100 µL 0.5 M PMP and 200 µL ammonia were added to aliquots of 100 µL of saccharide standards or of hydrolyzed samples. The reactions were carried out in a 70 °C water bath for 20 min, and then cooled to room temperature. 200 µL of formic acid were then added to neutralize the reaction.

Excess PMP was removed via liquid-liquid extraction, by adding 1 mL chloroform and 1 mL water to the derivatized samples. The solution was mixed thoroughly for 1 min and the organic layer was discarded. This process was repeated twice. The aqueous layer was then filtered through a 25 mm syringe filter with a 0.2 µm acetate membrane (VWR International^{TM}) and injected into the UPLC-Orbitrap MS system.

2.4 UPLC-Orbitrap MS conditions

Derivatized samples were measured using a UPLC Ultimate 3000 system (Thermo Fisher Scientific, Bremen, Germany) coupled to an Obitrap Q Exactive HRMS system (Thermo Fisher Scientific) with an electrospray ionization (H-ESI) source. Chromatographic separation was carried out on a Waters AccQ.Tag Ultra C18 column (2.1 mm x 100 mm, 1.7 µm particles) (Milford, MA, USA). An ACQUITY column in-line guard filter (2.1 mm, 0.2 µm) (Milford, MA, USA) was used. The column temperature was 35 °C. The mobile phase was composed of 20 mM ammonium formate as eluent A and acetonitrile as eluent B. The following solvent gradient was used: 0-0.5 min kept constant at 13% B, 0.5-2 min increased to 16% B, 2-15 min increased to 21% B, 15-16 min increased to 60% B, 16-17 min increased to 90% B, and 17-20 min decreased to 13% B for column re-equilibration. The flow rate was set to 0.3 mL min^{-1} and the injection volume was 2 µL.

The Orbitrap Q Exactive HRMS system was operated using full-mass scan mode (m/z 150-1000) in the ESI positive mode. The automatic gain control (AGC) target values were set to $3x10^6$ and the resolution was set to 70,000. Lock mass correction was set to 214.0896 (contaminant from plasticizer). Spray voltage was 3.5

kV, capillary temperature 300 °C, sheath gas was 35 arbitrary units and auxiliary gas was 15 arbitrary units.

2.5 Method validation

Relative standard deviations (RSD) of the retention times, mass errors, linearities, limits of detection (LOD), limits of quantification (LOQ), repeatabilities (RSD of peak areas), and precision for ^{13}C isotope enrichment detection were tested during method validation. A serial dilution of the calibration standards ranging from 20 nM to 25 µM was injected to test the RSD of the retention time, linearity and repeatability. Mass errors were determined based on the relative difference between the expected *m/z* values and the observed *m/z* values. LOD and LOQ for the method were determined as 3σ/*s* or 10σ/*s* respectively, where σ refers to the standard deviation of the *y* intercept of the regression line and *s* refers to the slope of the calibration curve (Shrivastava & Gupta, 2011).

The precision of the measurement of ^{13}C isotope enrichment was investigated using U-^{13}C D-Glucose (99 atom % ^{13}C, Sigma Aldrich) and D-glucosamine-1-^{13}C (99 atom % ^{13}C, Sigma Aldrich). A mixed serial isotopic dilution was prepared for both isotopically enriched standards with natural abundance compounds and ranged from 99 atom % ^{13}C to natural ^{13}C abundance (1.1 atom % ^{13}C). Each 100 µl standard dilutions were derivatized with PMP and injected into the UPLC-ESI-Orbitrap HRMS system. The isotopic dilution series were carried out in triplicates at a concentration of 50 µM.

In LC-HRMS applications, the atom % ^{13}C is calculated accounting for multiple isotopologues by using the relative abundance of the signal (S_k) weighted by the number of isotopically labelled carbons (k) (Equation 1):

$$at\%\ ^{13}C = \sum_{k=1}^{n}(k * S_k)/(n * \sum_{k=0}^{n} S_k) * 100 \qquad \text{(Eq. 1)}$$

where n is the total number of carbons in the molecule (Abadie & Tcherkez, 2021). Since both glucose and glucosamine contain 6 C atoms, in this study we account only for the isotopologues of M0 (all carbon atoms in the molecule are in their light isotopic variant, ^{12}C) to M6 (all carbon atoms in the molecule are in their heavy isotopic variant, ^{13}C).

However, during PMP-derivatization 20 C atoms are added to the molecules, creating bis-PMP derivatives for both glucose and glucosamine with 26 C atoms in

total. In order to properly estimate the atom % ^{13}C in these large molecules, a correction for natural isotope occurrence from the derivatization reagents must be carried out to account for the ^{13}C isotopologues from M0 to M26. This correction considers the difference between the theoretical and the experimental mass isotopomer distributions of the samples, including differences derived from tracer impurities (Heinrich et al., 2018; Selivanov et al., 2017).

2.6 Applicability of the PMP method

In order to test the applicability of this method, pure biomass from isolates of different taxonomic groups, such as archaeal, bacterial, fungal and plant species, were hydrolyzed and derivatized using the proposed PMP method. Peptidoglycan and/or lyophilized cells from *Methanobacterium* sp., *E. coli*, *B. subtilis*, *M. luteus*, *Streptomyces* sp., *S. aureus*, and *Synechococcus* sp. were used. Fungal biomass from *Clonostachys rosea* and plant biomass from *Malva silvestris* were also hydrolyzed. Four replicates of each species were tested.

Aliquots of 2 mg of peptidoglycan, lyophilized cells or freeze-dried isolates were hydrolyzed using 1.5 mL 6 M HCl. The solution was heated to 105 °C for 16 h and then cooled to room temperature. Vials were then centrifuged at 10,000 g for 20 min at room temperature and the supernatant was transferred into new vials and dried under a nitrogen stream. 1 mL water was then added to the samples and dried again under a nitrogen stream to remove residual HCl. In order to remove any salts, the dried samples were dissolved in 1 mL of MS-grade methanol, which does not extract salts. The samples were then centrifuged at 10,000 g for 10 min at room temperature and the supernatant was transferred and finally dried under a nitrogen stream. The dried hydrolyzed samples were dissolved in 1 mL water for the following PMP derivatization. Derivatization was carried out using 100 µL of dissolved hydrolyzed sample and following the same steps as previously depicted in the PMP-derivatization section.

The applicability of the method was also tested on soil samples collected from Andalucía, Spain and Styria, Austria. At both sites, soils under forest, grassland and cropland management were sampled using a soil corer (8 cm diameter, n=5, soils mixed after sieving by 2 mm) at 0-15 cm depth. Soil properties from both locations can be found in Table S.1. For acid hydrolysis, 0.1-0.4 g dry finely ground soil (equivalent to ~0.3 mg N) were mixed with 10 ml 6 M HCl. The mixture was heated to 105 °C for 8 h and then cooled to room temperature. The soil extract was filtered using qualitative

filter papers 150 mm diameter (Whatman®) and dried under nitrogen stream. The dried extract was re-dissolved with 12 mL of deionized water and the pH was adjusted to 6.6 – 6.8 using 0.4 M KOH and 0.2 M HCl. The extracts were centrifuged at 1200 g for 10 min and the supernatant was transferred and freeze-dried. In order to remove any salts, the freeze-dried extracts were dissolved in 8 mL MS-grade methanol and centrifuged at 1600 g for 15 min. The supernatant was transferred and finally dried under a nitrogen stream. The dried extracts were dissolved in 1 mL water for PMP derivatization.

2.7 Data analysis

In order to distinguish between taxonomic groups (archaea, bacteria, fungi and plants), a Principal Component Analysis (PCA) was carried out using all PMP-derivatives obtained from the different species. A PCA biplot was constructed using only the first and second principal components (PC1 and PC2), in order to visualize the clustering of the taxonomical groups. A permutational multivariate analysis of variance (perMANOVA) was carried out to test the effect of taxonomic group on the content of amino sugars, neutral sugars and uronic acids. All statistical analyses were performed in R version 4.2.0 (R Core Team, 2021).

3. Results

3.1 Separation of PMP derivatives

The developed PMP-derivatization protocol allowed the successful separation of three *N*-acetyl amino sugars, four amino sugars, eight neutral sugars and two uronic acids in a 20 min run using RP-UPLC-ESI-Orbitrap HRMS, including the separation of all isobars (Fig. 1). All compounds were identified based on their retention time and their PMP-derivative molecular ions $[M + H]^+$ with *m/z* values 481.2079 (Ara, Rib, Xyl), 495.2235 (Fuc, Rha), 510.2349 (GlcN, GalN, ManN), 511.2186 (Glu, Gal, Man), 525.1978 (GalA, GluA), 552.2453 (NAcGlc, NAcGal, NAcMan) and 582.2554 (MurN).

3.2 PMP method validation

Table 1 shows the method validation parameters. Retention time stability was very high (mean RSD < 0.4%). PMP-derivative molecular ions for all monosaccharides were found at the expected *m/z* values and the highest mass error detected was only 0.8 ppm (Table 1). All PMP-derivatives showed very good linearity (R^2 > 0.9991) and

repeatability (RSDs < 2.6%). The LOD and LOQ values ranged from 0.03 – 1.73 µM and 0.10 – 5.76 µM, respectively.

The precision of isotopic enrichment was calculated using U-^{13}C D-Glucose and ^{13}C for D-glucosamine-1-^{13}C (Fig. S1). For D-Glucose the following isotopologue m/z values were used: 511.2193 (M0), 512.2227 (M1), 513.2260 (M2), 514.2294 (M3), 515.2328 (M4), 516.2361 (M5) and 517.2395 (M6). For D-glucosamine the following isotopologue m/z values were used: 510.2353 (M0), 511.2387 (M1), 512.2420 (M2), 513.2454 (M3), 514.2487 (M4), 515.2521 (M5) and 516.2555 (M6). Figure 2 shows the regression lines corresponding to the ^{13}C isotopic calibrations. The U-^{13}C D-Glucose regression had a slope of 1.06, a *y*-intercept of -3.27 and an R^2 = 0.9978. The D-glucosamine-1-^{13}C regression had a slope of 0.97, a *y*-intercept of -3.07 and an R^2 = 0.9991. The precision of the measurement of ^{13}C isotopic enrichment was 0.03 atom % ^{13}C for U-^{13}C D-Glucose and 0.05 atom % ^{13}C D-glucosamine-1-^{13}C, based on the atom % ^{13}C standard deviation of repeated analyses of low-enrichment samples close to natural ^{13}C abundance.

3.3 Application of the PMP method

The developed PMP protocol was tested in archaeal, bacterial, fungal and plant samples for the quantification of amino sugars, neutral sugars, and uronic acids (Fig. 3). Since strong acid hydrolysis (> 1.3 M and > 50 °C) causes deacetylation of bound amino sugars, *N*-acetylamino sugars were not detectable in the hydrolyzed samples.

The results of sugar-PMP derivatives quantified in the archaeal, bacterial, fungal and plant species are shown in Table 2. Both fungal species showed the highest content of glucosamine (GlcN), followed by Gram-positive bacteria. *Methanobacterium* sp. had the highest galactosamine (GalN) content compared to the other species. Both plant species showed the highest content of glucose (Glu). Muramic acid (MurN) was only found in bacterial samples, where it had a higher content in Gram-positive bacteria compared to Gram-negative bacteria.

Based on its PMP-derivative molecular ion of *m/z* value 524.2128, we were able to identify talosaminuronic acid in archaea. This uronic acid was only found in *Methanobacterium* sp.. Mannosamine (ManN), xylose (Xyl) and galacturonic acid (GalA) were not detected above LOQ in any of the species tested.

The PCA showed the clustering of the different taxonomic groups to be driven by the PMP derivatives content in each of the species (Fig. 4). In the PCA biplot, the

principal component 1 (PC1) accounted for 32.2% of the variance and principal component 2 (PC2) for 22.9% of the variance. There was a significant effect of taxonomic group on the amino sugar, neutral sugar and uronic acid contents, as shown by perMANOVA analysis ($P<0.001$).

The PMP-derivatization method was also tested on soil samples (Fig. S.2 and S.3). The results of sugar-PMP derivatives quantified in forest, grassland and cropland soils are shown in Table 3. GlcN was the most abundant amino sugar found in all soil samples, while Xyl was the most abundant monosaccharide quantified (Table 3). GlcN contents varied from 1.2 – 5.1 mg g^{-1} of dry weight (dw), while MurN varied from 0.25 – 2.5 mg g^{-1} dw. Talosaminuronic acid (TalA) could be quantified in almost all soils tested. Mannosamine (ManN), galactose (Gal) arabinose (Ara), ribose (Rib), galacturonic acid (GalA) and glucuronic acid (GluA) were not detected in any of the soils tested.

4. Discussion

4.1 Separation of PMP-derivatives

Our proposed PMP-derivatization method allowed the successful separation and quantification of four amino sugars, three *N*-acetyl amino sugars, eight neutral sugars and three uronic acids in a single run. While Orbitrap HRMS could distinguish between non-isomeric PMP-derivatives, UPLC separation was required to separate the different isomers. Both dimensions were necessary to fully resolve the 18 sugar-related compounds.

Method validation parameters (Table 1) showed the successful application of the PMP-derivatization method. Retention time stability (RSD < 0.4%) was very high and within range of those observed by Hu et al. (2017). All PMP-derivative molecular ions were found to be similar to the expected m/z values and the highest mass error detected was only 0.8 ppm (Table 1). This not only shows the high precision and sensitivity of the proposed PMP-method as analyzed by an UPLC-Orbitrap HRMS system, but also the mass error was markedly lower than that observed for most monosaccharides in another study using similar PMP-derivatization and ion-trap time-of-flight (IT-TOF) MS detection (Guo et al., 2019). A very good linearity of the MS response was obtained (R^2 > 0.9991), better than that observed by using triple quadrupole ion trap MS (TQ-MS/MS) detection (Xia et al., 2017). The repeatability of

the method showed high precision for the PMP-derivatization methods, with all RSDs being below 2.6% for peak areas.

The detection and quantification limits established in the current method were lower than those published for PMP-derivatization followed by diode array detection (DAD) (Wang W. et al., 2021), evaporative light scattering detection (ELSD) (Wang W. et al., 2021), IT-TOF MS detection (Guo et al., 2019), or capillary electrophoresis coupled to UV detection (Hu et al., 2014; Liu et al., 2022). Likewise, this PMP-method has lower limits of detection for most monosaccharides when compared to *o*-phthalaldehyde (OPA) and *p*-aminobenzoic acid (*p*-AMBA) derivatization methods (Indorf et al., 2011; Meyer et al., 2001). The coupling of RP-UPLC separation to Q Exactive Orbitrap HRMS therefore provided highest specificity and sensitivity for the detection of the targeted biomarkers.

The precision of isotopic enrichment carried out for the PMP-derivatization method using UPLC-Orbitrap HRMS showed the unique sensitivity for the separation of ^{13}C isotopologues for both glucose and glucosamine, with a precision of ^{13}C enrichment detection around 0.03 atom % ^{13}C enrichment for glucose and 0.05 atom % ^{13}C enrichment for glucosamine, which is much more sensitive than what is achieved by GC/MS (Patterson et al., 1997; Preston & Slater, 1994). This highlights the potential use of this PMP-derivatization approach for the sensitive measurement of isotope incorporation in microbial and plant necromass biomarkers, without the use of more specialized instrumentation such as an isotope ratio mass spectrometer (IRMS) (Sessions, 2006). Though IRMS has a 10-fold higher sensitivity in terms of isotope incorporation, this comes at the cost of markedly decreased concentration sensitivity.

In this study, the atom % ^{13}C was calculated using only the isotopologues M0 to M6 (Equation 1) since both glucose and glucosamine contain 6 C atoms. While these calculations provided useful proxies for the determination of the precision of ^{13}C isotopic enrichment, a more precise estimation of the atom % ^{13}C would require the consideration of all possible M0 to M26 ^{13}C isotopologues since the PMP-derivatization reaction adds 20 C to the molecules and each of these C atoms come with an atom % ^{13}C at natural abundance (1.1 atom % ^{13}C). To generate this more precise atom % ^{13}C estimation, a correction for the natural abundance of all possible isotopologues introduced by the derivatization reaction would be required alongside a correction for possible tracer impurities (Heinrich et al., 2018; Selivanov et al., 2017).

4.2 Application of the method

We were able to identify and quantify most of the PMP derivatives in all prokaryotic and eukaryotic species tested. The amino sugar content quantified using our modified PMP-method is within range of what has been measured in other microorganisms. The measured GlcN content ranged from 60.3 – 76.6 mg g^{-1} dw for fungal species, and the MurN content from 32.4 – 65.1 mg g^{-1} dw for Gram-positive bacteria and 0.9 – 4.5 mg g^{-1} dw for Gram-negative bacteria (Table 2). Appuhn and Joergensen (2006) showed that GlcN content in different fungal species ranged between 9 – 112 mg g^{-1} dw, while MurN ranged from 1.7 – 42.2 mg g^{-1} dw for Gram-positive bacteria and 2.0 – 7.2 mg g^{-1} dw for Gram negative bacteria. Similarly, Glaser et al. (2004) observed a GlcN content of 27.1 mg C g^{-1} of total organic carbon (TOC) for fungi and 34.7 mg C g^{-1} TOC for bacteria (contents on a dry mass basis ~2-fold higher at C content of 50%), while the MurN content varied from 6.6 mg C g^{-1} TOC in bacteria to <0.1 mg C g^{-1} TOC in fungi.

Differences between the amino sugar content presented in this paper and previous papers could be explained by the different species used. Species-specific variations in the cell wall thickness in bacteria could explain differences observed in MurN and GlcN content (Vollmer et al., 2008), which was also observed in the wide range of MurN contents of Gram-positive bacteria presented in Appuhn and Joergensen (2006).

Fungal chitin is mainly composed of interlinked *N*-acetylglucosamine, while bacterial peptidoglycan contains both, *N*-acetylglucosamine and *N*-acetylmuramic acid (Joergensen, 2018). This explains not only why the highest GlcN content was found in both fungal species, but also why Gram-positive bacteria had a high GlcN content. Gram-positive bacteria also had a higher MurN content compared to Gram-negative bacteria species, due to the greater thickness of the peptidoglycan layer in Gram-positive bacteria (Vollmer et al., 2008). Also, muramic acid was only found in the bacterial species, which coincides with it being a very specific bacterial necromass biomarker that is only found in bacterial peptidoglycan.

Talosaminuronic acid was only found in *Methanobacterium* sp., but not in any other taxonomic group. This is due to archaeal pseudopeptidolglycan containing both *N*-acetylglucosamine and *N*-talosaminuronic acid (Subedi et al., 2021). König et al. (1993) also proposed that some methanobacterial pseudomurein glycan strands can

contain *N*-acetylgalactosamine along with or instead of *N*-acetylglucosamine, which could account for the high galactosamine content observed in this archaeal species.

Regarding the neutral sugars, all microbial species had lower contents of hexoses, pentoses and uronic acids, when compared to their amino sugar contents, in contrast to higher plants (*A. cerefolium* and *M. sylvestris*), which showed the highest glucose contents and the lowest amino sugar contents out of all the other species tested. Since cellulose is an important structural polymer of plant cell walls and consists of interlinked glucose units, it explained the high glucose content in plants. Similarly, amino sugars have been found only in trace amounts in higher plants (Parsons, 2021).

PCA results showed that the different taxonomic groups clustered based on their PMP-derivative distributions (Fig. 4). This indicates the successful application of the PMP-method for the measurement of amino sugar, neutral sugar and uronic acid contents and its application for the quantification of soil necromass biomarkers.

When applied to soils, we were able to identify and quantify amino sugars, neutral sugars and uronic acids using the adopted PMP-derivatization protocol. GlcN was the most abundant amino sugar measured in all soils (Table 3), ranging from 1.2 – 5.1 mg g^{-1} dw. According to Joergensen (2018), GlcN contributes between 47-68 % to total amino sugars in soils (54 ± 8% in this study) and its major source are fungal cell walls. However, GlcN from bacterial origin (peptidoglycan) can also contribute up to 30% in saline soils with high bacterial residues (Khan et al., 2016). High bacterial necromass was also found here in the Calcisols (~pH 8.6), with GlcN:MurN ratios of 1.9-2.0 being close to those expected from bacteria alone (~2), with no GlcN contribution by fungi. The highest MurN content was observed in alkaline forest and grassland soils (Table 3), indicating a high abundance of bacterial structural residues. Both GlcN and MurN contents in all soils tested with the PMP-derivatization method were within the higher ranges of those presented in other studies using aldonitrile or OPA derivatization methods along with gas or liquid chromatography (Khan et al., 2016; Wang B. et al., 2021a; Yang et al., 2020). High clay contents in soils might also explain the higher content of GlcN and MurN in Calcisols, since amino sugar polymers can be stabilized on clay particles and thereby resist microbial degradation (Khan et al., 2016).

Furthermore, strong acid hydrolysis, which is required for the successful release of amino sugars from their polymeric form (Einbu & Vårum, 2007), can affect

the composition and yield of neutral sugars (Lavarack et al., 2002) and uronic acids. The hydrolysis of samples using 6 M HCl could account for the lower concentrations of neutral sugars and of uronic acids compared to amino sugars in all species. Guo et al. (2019) found that optimal hydrolysis conditions vary between amino sugars, neutral sugars and uronic acids, indicating that differential hydrolysis protocols should be applied to the same sample, when aiming to determine all three biomarker fractions. However, this is not only time-consuming, but it might also not be feasible if only low quantities of sample are available such as from slow-growing archaeal and bacterial isolates. While using 6 M HCl for hydrolysis is preferential for the release of amino sugars, we observed that neutral sugars and uronic acids, although low in concentration, are still quantifiable, but that they are likely underestimated (Table 2).

This PMP derivatization method was successfully applied for the identification and quantification of microbial necromass biomarkers in soils. Moreover, this method can be used for tracing isotopically labeled substrates (e.g. in ^{15}N and/or ^{13}C, separately or combined, the latter due to the separation of ^{13}C and ^{15}N-isotopologues with almost same mass by Orbitrap MS) (Hu et al., 2017) into necromass biomarkers, and these into different SOM pools, such as occluded and free particulate organic matter and mineral-associated organic matter. However, in order to accurately determine the contribution of microbial necromass C to SOM, it is important to consider that the quantification of these biomarkers do not reflect the entirety of the microbial necromass pool since they do not account for the presence of microbial extracellular polymeric substances and rhizodeposits (Whalen et al., 2022).

Conversion factors are used to estimate bacterial and fungal necromass C from MurN and GlcN content in soils. These conversion factors are derived from microbial biomarkers measured for a few isolates of bacteria and fungi and vary strongly, e.g. between Gram-positive and Gram-negative bacteria and among fungi (Appuhn & Joergensen, 2006). This bears a large uncertainty for the estimations of soil microbial necromass and further assumes that: 1) all necromass biomarkers are stabilized at the same rate in soils, and 2) their ratios found in microbial necromass are the same as in microbial biomass (Whalen et al., 2022). While more studies are required to investigate these assumptions, determining the ratio of Gram-positive to Gram-negative bacteria, as well as the bacteria to fungi ratio in soils, would improve the quantification of microbial necromass C when using conversion factors.

5. Conclusions

We developed a modified PMP-derivatization method coupled to RP-UPLC-ESI-Orbitrap HRMS, which allowed the successful separation and quantification of 18 sugar-containing compounds, including four amino sugars, three *N*-acetylamino sugars, eight neutral sugars and three uronic acids. When applied to acid hydrolyzed samples from archaea, bacteria, fungi and plant species, a multivariate statistical analysis showed the separation between these taxonomic groups based on their amino sugar, neutral sugar and uronic acid contents, supporting the use of these compounds as necromass biomarkers. Similarly, our developed PMP-derivatization method was successfully applied to forest, grassland and cropland soils with contrasting edaphic properties. Our proposed PMP-derivatization method can be applied to identify and quantify soil microbial necromass biomarkers and their contribution to the stabilization of SOM, as well as for tracing isotopically labeled compounds into these necromass biomarkers and further into different SOM pools.

Acknowledgements

The authors would like to thank Ludwig Seidl for his technical support with the UPLC-Orbitrap measurements. Likewise, we would like to thank Sabrina Pober and Sabine Maringer for their assistance in the laboratory work.

References

Abadie, C., & Tcherkez, G. (2021). 13C isotope labelling to follow the flux of photorespiratory intermediates. *Plants*, *10*(3), 427. https://doi.org/10.3390/plants10030427

Appuhn, A., & Joergensen, R. (2006). Microbial colonisation of roots as a function of plant species. *Soil Biology and Biochemistry, 38*, 1040–1051. https://doi.org/10.1016/j.soilbio.2005.09.002

Bhavsar, A. P., Erdman, L. K., Schertzer, J. W., & Brown, E. D. (2004). Teichoic acid is an essential polymer in Bacillus subtilis that is functionally distinct from teichuronic acid. *Journal of Bacteriology, 186*, 7865–7873. https://doi.org/10.1128/jb.186.23.7865-7873.2004

Bore, E. K., Kuzyakov, Y., & Dippold, M. A. (2019). Glucose and ribose stabilization in soil: Convergence and divergence of carbon pathways assessed by position-specific labeling. *Soil Biology and Biochemistry, 131*, 54–61. https://doi.org/10.1016/j.soilbio.2018.12.027

Dai, J., Wu, Y., Chen, S. W., Zhu, S., Yin, H. P., Wang, M., & Tang, J. (2010). Sugar compositional determination of polysaccharides from Dunaliella salina by modified RP-HPLC method of precolumn derivatization with 1-phenyl-3-methyl-5-pyrazolone. *Carbohydrate Polymers, 82*, 629–635. https://doi.org/10.1016/j.carbpol.2010.05.029

Drigo, B., Anderson, I. C., Kannangara, G. S. K., Cairney, J. W. G., & Johnson, D. (2012). Rapid incorporation of carbon from ectomycorrhizal mycelial necromass into soil fungal communities. *Soil Biology and Biochemistry, 49*, 4–10. https://doi.org/10.1016/j.soilbio.2012.02.003

Einbu, A., & Vårum, K. M. (2007). Depolymerization and de-N-acetylation of chitin oligomers in hydrochloric acid. *Biomacromolecules, 8*, 309–314. https://doi.org/10.1021/bm0608535

Engelking, B., Flessa, H., & Joergensen, R. G. (2007). Shifts in amino sugar and ergosterol contents after addition of sucrose and cellulose to soil. *Soil Biology and Biochemistry, 39*, 2111–2118. https://doi.org/10.1016/j.soilbio.2007.03.020

Fazio, S. A., Uhlinger, D. J., Parker, J. H., & White, D. C. (1982). Estimations of uronic acids as quantitative measures of extracellular and cell wall polysaccharide polymers from environmental samples. *Applied and*

Environmental Microbiology, 43, 1151–1159. https://doi.org/10.1128/aem.43.5.1151-1159.1982

German, D. P., Chacon, S. S., & Allison, S. D. (2011). Substrate concentration and enzyme allocation can affect rates of microbial decomposition. *Ecology, 92*, 1471–1480. https://doi.org/10.1890/10-2028.1

Glaser, B., Turrión, M.-B., & Alef, K. (2004). Amino sugars and muramic acid—Biomarkers for soil microbial community structure analysis. *Soil Biology and Biochemistry, 36*, 399–407. https://doi.org/10.1016/j.soilbio.2003.10.013

Gonzalez, N. M., Fitch, A., & Al-Bazi, J. (2020). Development of a RP-HPLC method for determination of glucose in Shewanella oneidensis cultures utilizing 1-phenyl-3-methyl-5-pyrazolone derivatization. *PLOS ONE, 15*, e0229990. https://doi.org/10.1371/journal.pone.0229990

Gorfer, M., Blumhoff, M., Klaubauf, S., Urban, A., Inselsbacher, E., Bandian, D., Mitter, B., Sessitsch, A., Wanek, W., & Strauss, J. (2011). Community profiling and gene expression of fungal assimilatory nitrate reductases in agricultural soil. *The ISME Journal, 5*, 1771–1783. https://doi.org/10.1038/ismej.2011.53

Guggenberger, G., Christensen, B. T., & Zech, W. (1994). Land-use effects on the composition of organic matter in particle-size separates of soil: I. Lignin and carbohydrate signature. *European Journal of Soil Science, 45*, 449–458. https://doi.org/10.1111/j.1365-2389.1994.tb00530.x

Gunina, A., & Kuzyakov, Y. (2015). Sugars in soil and sweets for microorganisms: Review of origin, content, composition and fate. *Soil Biology and Biochemistry, 90*, 87–100. https://doi.org/10.1016/j.soilbio.2015.07.021

Guo, N., Bai, Z., Jia, W., Sun, J., Wang, W., Chen, S., & Wang, H. (2019). Quantitative analysis of polysaccharide composition in Polyporus umbellatus by HPLC–ESI–TOF–MS. *Molecules, 24*, 2526. https://doi.org/10.3390/molecules24142526

Heinrich, P., Kohler, C., Ellmann, L., Kuerner, P., Spang, R., Oefner, P. J., & Dettmer, K. (2018). Correcting for natural isotope abundance and tracer impurity in MS-, MS/MS-and high-resolution-multiple-tracer-data from stable isotope labeling experiments with IsoCorrectoR. *Scientific reports, 8*, 1-10. https://doi.org/10.1038/s41598-018-36293-4

Honda, S., Akao, E., Suzuki, S., Okuda, M., Kakehi, K., & Nakamura, J. (1989). High-performance liquid chromatography of reducing carbohydrates as strongly

ultraviolet-absorbing and electrochemically sensitive 1-phenyl-3-methyl5-pyrazolone derivatives. *Analytical Biochemistry, 180*, 351–357. https://doi.org/10.1016/0003-2697(89)90444-2

Hu, Y., Wang, T., Yang, X., & Zhao, Y. (2014). Analysis of compositional monosaccharides in fungus polysaccharides by capillary zone electrophoresis. *Carbohydrate Polymers, 102*, 481–488. https://doi.org/10.1016/j.carbpol.2013.11.054

Hu, Y., Zheng, Q., & Wanek, W. (2017). Flux analysis of free amino sugars and amino acids in soils by isotope tracing with a novel liquid chromatography/high resolution mass spectrometry platform. *Analytical Chemistry, 89*, 9192–9200. https://doi.org/10.1021/acs.analchem.7b01938

Indorf, C., Dyckmans, J., Khan, K. S., & Joergensen, R. G. (2011). Optimisation of amino sugar quantification by HPLC in soil and plant hydrolysates. *Biology and Fertility of Soils, 47*, 387–396. https://doi.org/10.1007/s00374-011-0545-5

Joergensen, R. G. (2018). Amino sugars as specific indices for fungal and bacterial residues in soil. *Biology and Fertility of Soils, 54*, 559–568. https://doi.org/10.1007/s00374-018-1288-3

Khan, K.S., Mack, R., Castillo, X., Kaiser, M., & Joergensen, R.G. (2016). Microbial biomass, fungal and bacterial residues, and their relationships to the soil organic matter C/N/P/S ratios. *Geoderma, 271*, 115–123. https://doi.org/10.1016/j.geoderma.2016.02.019

König, H., Hartmann, E., & Kärcher, U. (1993). Pathways and principles of the biosynthesis of Methanobacterial cell wall polymers. *Systematic and Applied Microbiology, 16*, 510–517. https://doi.org/10.1016/S0723-2020(11)80320-6

Lavarack, B. P., Griffin, G. J., & Rodman, D. (2002). The acid hydrolysis of sugarcane bagasse hemicellulose to produce xylose, arabinose, glucose and other products. *Biomass and Bioenergy, 23*, 367–380. https://doi.org/10.1016/S0961-9534(02)00066-1

Liang, C., Amelung, W., Lehmann, J., & Kästner, M. (2019). Quantitative assessment of microbial necromass contribution to soil organic matter. *Global Change Biology, 25*, 3578–3590. https://doi.org/10.1111/gcb.14781

Liang, C., Schimel, J. P., & Jastrow, J. D. (2017). The importance of anabolism in microbial control over soil carbon storage. *Nature Microbiology, 2*, 17105. https://doi.org/10.1038/nmicrobiol.2017.105

Liu, S., Li, X., Yang, X., Zhou, L., Liang, X., Qiu, R., & Fa, Y. (2022). A capillary electrophoresis method for the determination of soluble monosaccharides in Ginkgo biloba leaves. *Journal of Separation Science, 45*, 623–630. https://doi.org/10.1002/jssc.202100749

Ludwig, M., Achtenhagen, J., Miltner, A., Eckhardt, K.-U., Leinweber, P., Emmerling, C., & Thiele-Bruhn, S. (2015). Microbial contribution to SOM quantity and quality in density fractions of temperate arable soils. *Soil Biology and Biochemistry, 81*, 311–322. https://doi.org/10.1016/j.soilbio.2014.12.002

Meyer, A., Raba, C., & Fischer, K. (2001). Ion-Pair RP-HPLC determination of sugars, amino sugars, and uronic acids after derivatization with p-aminobenzoic acid. *Analytical Chemistry, 73*, 2377–2382. https://doi.org/10.1021/ac001402s

Parsons, J. W. (2021). Chemistry and Distribution of Amino Sugars in Soil and Soil Organisms. In: Paul, E.A., Ladd, J.N. (Eds.), Soil Biochemistry, CRC Press.

Patterson, B. W., Zhang, X.-J., Chen, Y., Klein, S., & Wolfe, R. R. (1997). Measurement of very low stable isotope enrichments by gas chromatography/mass spectrometry: Application to measurement of muscle protein synthesis. *Metabolism, 46*, 943–948. https://doi.org/10.1016/S0026-0495(97)90084-6

Preston, T., & Slater, C. (1994). Mass spectrometric analysis of stable-isotope-labelled amino acid tracers. *Proceedings of the Nutrition Society, 53*, 363–372. https://doi.org/10.1079/PNS19940042

R Core Team., 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Selivanov, V. A., Benito, A., Miranda, A., Aguilar, E., Polat, I. H., Centelles, J. J, Jayaraman, A., Lee, P.W.N., Marin, S., & Cascante, M. (2017). MIDcor, an R-program for deciphering mass interferences in mass spectra of metabolites enriched in stable isotopes. *BMC bioinformatics, 18*, 1-9. https://doi.org/10.1186/s12859-017-1513-3

Sessions, A. L. (2006). Isotope-ratio detection for gas chromatography. Journal of *Separation Science, 29*, 1946–1961. https://doi.org/10.1002/jssc.200600002

Shrivastava, A., & Gupta, V. B., 2011. Methods for the determination of limit of detection and limit of quantitation of the analytical methods. *Chronicles of Young Scientists, 2*, 21–25. https://doi.org/10.4103/2229-5186.79345

Strydom, D. J. (1994). Chromatographic separation of 1-phenyl-3-methyl-5-pyrazolone-derivatized neutral, acidic and basic aldoses. *Journal of Chromatography A, 678*, 17–23. https://doi.org/10.1016/0021-9673(94)87069-1

Subedi, B. P., Martin, W. F., Carbone, V., Duin, E. C., Cronin, B., Sauter, J., Schofield, L. R., Sutherland-Smith, A. J., & Ronimus, R. S. (2021). Archaeal pseudomurein and bacterial murein cell wall biosynthesis share a common evolutionary ancestry. *FEMS Microbes, 2*, xtab012. https://doi.org/10.1093/femsmc/xtab012

Vollmer, W., Blanot, D., & De Pedro, M. A. (2008). Peptidoglycan structure and architecture. *FEMS Microbiology Reviews, 32*, 149–167. https://doi.org/10.1111/j.1574-6976.2007.00094.x

Wang, B., An, S., Liang, C., Liu, Y., & Kuzyakov, Y. (2021a). Microbial necromass as the source of soil organic carbon in global ecosystems. *Soil Biology and Biochemistry, 162*, 108422. https://doi.org/10.1016/j.soilbio.2021.108422

Wang, B., Liang, C., Yao, H., Yang, E., & An, S. (2021b). The accumulation of microbial necromass carbon from litter to mineral soil and its contribution to soil organic carbon sequestration. *Catena, 207*, 105622. https://doi.org/10.1016/j.catena.2021.105622

Wang, W., Wang, Y., Chen, F., & Zheng, F. (2021). Comparison of determination of sugar-PMP derivatives by two different stationary phases and two HPLC detectors: C18 vs. amide columns and DAD vs. ELSD. *Journal of Food Composition and Analysis, 96*, 103715. https://doi.org/10.1016/j.jfca.2020.103715

Whalen, E. D., Grandy, A. S., Sokol, N. W., Keiluweit, M., Ernakovich, J., Smith, R. G., & Frey, S. D. (2022). Clarifying the evidence for microbial- and plant-derived soil organic matter, and the path toward a more quantitative understanding. *Global Change Biology, 28*, 7167–7185. https://doi.org/10.1111/gcb.16413

Xia, Y.-G., Wang, T.-L., Sun, L.-M., Liang, J., Yang, B.-Y., & Kuang, H.-X. (2017). A new UPLC-MS/MS method for the characterization and discrimination of polysaccharides from genus Ephedra based on enzymatic digestions. *Molecules, 22*, 1992. https://doi.org/10.3390/molecules22111992

Xiao, C., & Anderson, C. T. (2013). Roles of pectin in biomass yield and processing for biofuels. *Frontiers in Plant Science, 4*, 67. https://doi.org/10.3389/fpls.2013.00067

Xu, X., Thornton, P. E., & Post, W. M. (2013). A global analysis of soil microbial biomass carbon, nitrogen and phosphorus in terrestrial ecosystems. *Global Ecology and Biogeography, 22*, 737–749. https://doi.org/10.1111/geb.12029

Yang, L., Lyu, M., Li, X., Xiong, X., Lin, W., Yang, Y., & Xie, J. (2020). Decline in the contribution of microbial residues to soil organic carbon along a subtropical elevation gradient. *Science of The Total Environment, 749*, 141583. https://doi.org/10.1016/j.scitotenv.2020.141583

Table 1. Retention times, detected masses, mass accuracies, linearities, limits of detection (LOD), limits of quantification (LOQ) and repeatabilities (RSD, relative standard deviation) of 17 PMP-labeled amino sugars, neutral sugars and uronic acids measured by RP-UPLC-ESI-Orbitrap HRMS.

Compound	Chemical formula	Retention time (min)	Retention time RSD (%)	m/z $[M+H]^+$	Expected PMP derivatives $[M+H]^+$	Observed PMP derivatives $[M+H]^+$	Mass error (ppm)	Linearity (R^2)	LOD (μM)	LOQ (μM)	Repeatability RSD (%)
Ribose	$C_5H_{10}O_5$	6.75	0.2	151.0601	481.2082	481.2079	0.6	0.9999	0.03	0.10	0.3
Arabinose	$C_5H_{10}O_5$	7.89	0.1	151.0601	481.2082	481.2080	0.5	0.9993	0.17	0.57	0.7
Xylose	$C_5H_{10}O_5$	7.97	0.2	151.0601	481.2082	481.2079	0.6	0.9991	0.19	0.62	0.7
Rhamnose	$C_6H_{12}O_5$	6.76	0.2	165.0758	495.2239	495.2235	0.8	0.9999	0.03	0.10	0.7
Fucose	$C_6H_{12}O_5$	8.31	0.1	165.0758	495.2239	495.2236	0.6	0.9993	0.23	0.77	1.8
Glucosamine	$C_6H_{13}NO_5$	6.71	0.3	180.0866	510.2347	510.2349	-0.4	0.9996	0.26	0.85	1.6
Galactosamine	$C_6H_{13}NO_5$	7.22	0.3	180.0866	510.2347	510.2347	-0.1	0.9997	0.16	0.52	1.5
Mannosamine	$C_6H_{13}NO_5$	6.90	0.1	180.0866	510.2347	510.2347	0.0	0.9997	0.84	2.78	0.5
Glucose	$C_6H_{12}O_6$	7.61	0.2	181.0707	511.2188	511.2186	0.4	0.9997	0.08	0.26	0.2
Galactose	$C_6H_{12}O_6$	7.97	0.2	181.0707	511.2188	511.2187	0.2	0.9996	0.23	0.77	1.6
Mannose	$C_6H_{12}O_6$	6.16	0.3	181.0707	511.2188	511.2185	0.6	0.9993	0.13	0.44	1.2
Galacturonic acid	$C_6H_{10}O_7$	7.30	0.1	195.0499	525.1980	525.1978	0.5	0.9992	0.25	0.84	2.2
Glucuronic acid	$C_6H_{10}O_7$	7.14	0.4	195.0499	525.1980	525.1978	0.4	0.9998	0.13	0.44	1.6
N-acetylglucosamine	$C_8H_{15}NO_6$	7.23	0.3	222.0972	552.2453	552.2452	0.2	0.9994	0.22	0.74	1.1
N-acetylgalactosamine	$C_8H_{15}NO_6$	7.55	0.2	222.0972	552.2453	552.2451	0.3	0.9997	0.11	0.37	0.1
N-acetylmannosamine	$C_8H_{15}NO_6$	8.01	0.4	222.0972	552.2453	552.2451	0.3	0.9996	0.36	1.12	1.6
Muramic acid	$C_9H_{17}NO_7$	6.24	0.3	252.1078	582.2559	582.2554	0.8	0.9993	1.73	5.76	2.6

Table 2. Concentrations (± standard deviation) of amino sugars, neutral sugars and uronic acids measured in pure biomass from archaea, bacteria, fungi and plant species using PMP-derivatization and RP-UPLC-Orbitrap HRMS.

Taxonomical group	Species	GlcN (mg g^{-1} dw)	GalN (mg g^{-1} dw)	MurN (mg g^{-1} dw)	Glc (mg g^{-1} dw)	Gal (mg g^{-1} dw)	Man (mg g^{-1} dw)	Rib (mg g^{-1} dw)	Ara (mg g^{-1} dw)	Fuc (mg g^{-1} dw)	GluA (mg g^{-1} dw)	TalA (mg g^{-1} dw)
Archaea	*Methanobacterium* sp.	20.3 (± 5.3)	26.7 (±6.7)	-	0.8 (±0.3)	0.3 (±0.1)	0.1 (±0.0)	0.2 (±0.1)	0.2 (±0.0)	-	-	11.8 (±4.9)
Gram-negative bacteria	*Escherichia coli*	1.8 (± 0.6)	0.2 (±0.1)	0.9 (±0.3)	0.2 (±0.1)	-	-	0.1 (±0.1)	0.3 (±0.2)	-	-	-
	Synechococcus sp.	7.3 (±2.3)	0.9 (±0.3)	4.5 (±1.4)	2.2 (±0.4)	0.9 (±0.2)	0.2 (±0.1)	0.3 (±0.1)	0.3 (±0.1)	0.1 (±0.0)	-	-
Gram-positive bacteria	*Bacillus subtilis*	32.3 (±16.0)	-	53.8 (±27.7)	0.2 (±0.1)	0.2 (±0.0)	-	0.1 (±0.0)	0.3 (±0.0)	-	-	-
	Micrococcus luteus (peptidoglycan)	75.7 (±6.4)	20.1 (±1.9)	32.4 (±2.7)	0.5 (±0.2)	0.3 (±0.1)	0.2 (±0.0)	0.3 (±0.1)	0.4 (±0.1)	-	-	-
	Micrococcus luteus (lyophilized cells)	38.3 (±4.8)	-	65.1 (±11.1)	2.7 (±0.6)	-	0.2 (±0.0)	0.5 (±0.1)	0.4 (±0.0)	-	-	-
	Staphylococcus aureus	23.4 (±16.2)	0.8 (±0.6)	43.5 (±33.1)	0.2 (±0.1)	-	-	0.1 (±0.0)	0.5 (±0.0)	-	-	-
	Streptomyces sp.	23.8 (±1.9)	1.5 (±0.1)	39.6 (±3.9)	0.2 (±0.0)	0.2 (±0.0)	0.1 (±0.0)	0.1 (±0.0)	0.2 (±0.0)	-	-	-
Fungi	*Clonostachys rosea*	76.6 (±7.1)	6.1 (±0.5)	-	4.7 (±0.9)	1.9 (±0.2)	0.1 (±0.0)	0.1 (±0.0)	0.2 (±0.0)	0.2 (±0.0)	0.1 (±0.0)	-
	Fusarium fujikurio	60.3 (±5.9)	-	-	1.8 (±0.4)	0.7 (±0.1)	0.1 (±0.0)	0.1 (±0.0)	0.1 (±0.0)	-	-	-
Plants	*Anthriscus cerefolium*	0.7 (±0.0)	0.1 (±0.0)	-	5.9 (±0.6)	1.9 (±0.3)	0.1 (±0.0)	0.1 (±0.0)	0.1 (±0.0)	0.1 (±0.0)	-	-
	Malva silvestris	2.7 (±0.2)	0.2 (±0.0)	-	14.0 (±4.1)	0.8 (±0.2)	0.1 (±0.0)	0.1 (±0.0)	0.2 (±0.1)	0.1 (±0.0)	-	-

Table 3. Concentrations of amino sugar and neutral sugar biomarkers measured in Calcisols and Cambisols under forest, grassland and cropland using PMP-derivatization and UPLC-Orbitrap MS.

Site	Management	GlcN (µg mg^{-1} dw)	GalN (µg mg^{-1} dw)	MurN (µg mg^{-1} dw)	Glc (µg mg^{-1} dw)	Man (µg mg^{-1} dw)	Xyl (µg mg^{-1} dw)	Fuc (µg mg^{-1} dw)	Rha (µg mg^{-1} dw)	TalA (µg mg^{-1} dw)
1	Forest	5.05	2.94	2.53	4.90	1.60	20.90	0.42	0.08	0.15
2	Grassland	4.69	2.39	2.41	4.03	1.30	14.80	0.35	0.07	0.18
3	Cropland	1.15	0.75	0.62	1.20	0.34	4.40	0.11	0.02	0.05
4	Cropland	2.97	1.91	0.47	2.27	1.01	17.83	0.36	0.07	0.22
5	Grassland	4.92	3.14	0.45	3.24	1.25	24.64	0.47	0.09	0.48
6	Forest	2.03	0.79	0.25	1.40	0.78	10.80	0.31	0.06	-

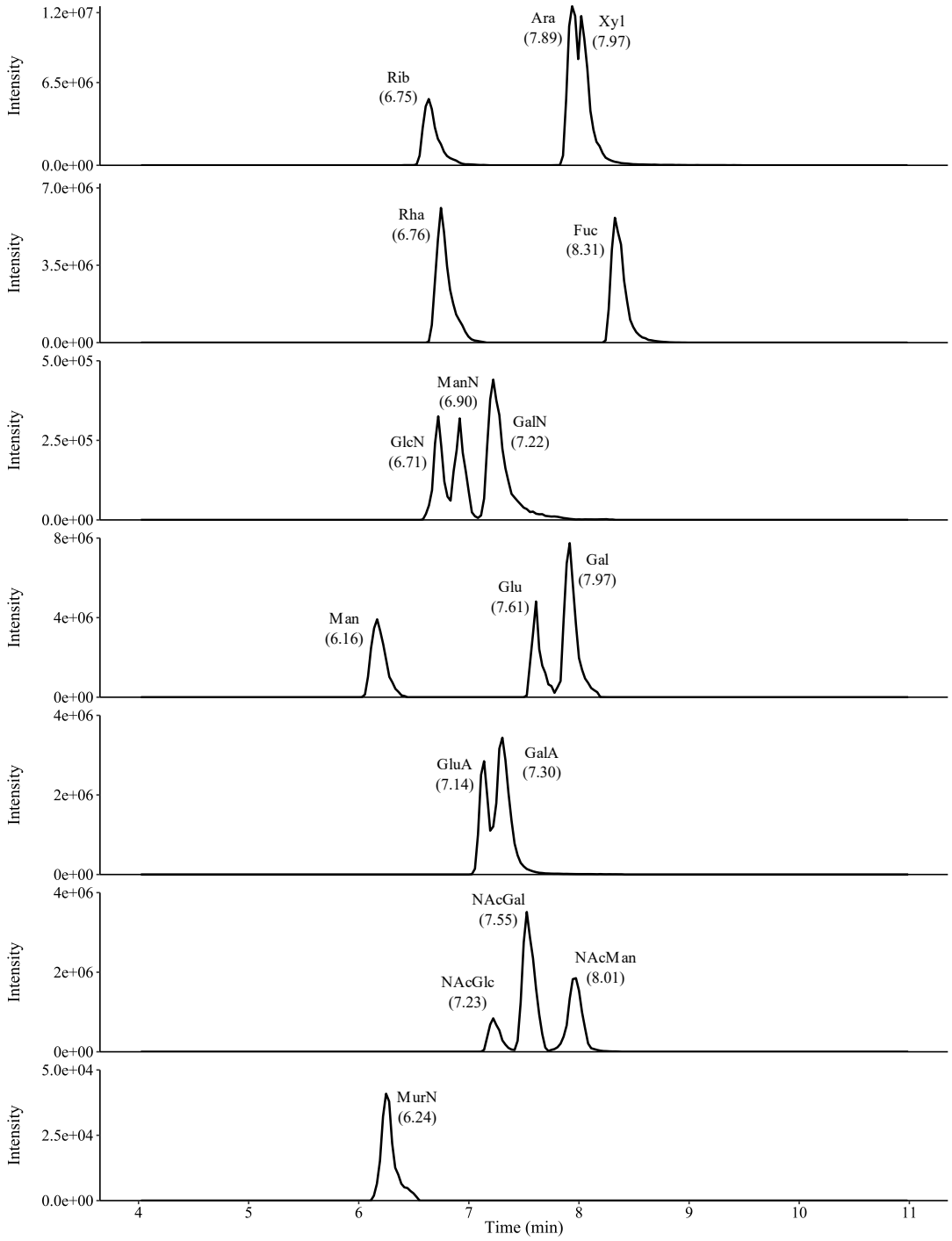

Figure 1. Chromatograms of amino sugar, neutral sugar and uronic acid standards derivatized with PMP and injected into the RP-UPLC-ESI-Orbitrap HRMS system.

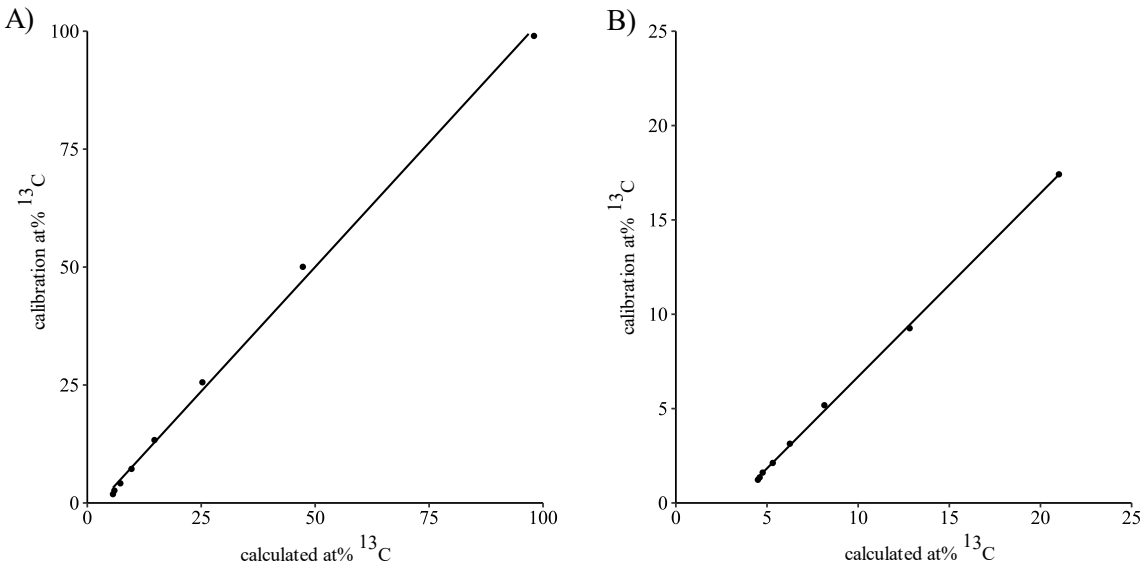

Figure 2. ^{13}C isotopic enrichment precision of U-^{13}C D-glucose (A) and D-glucosamine-1-^{13}C (B). The calculated atom % ^{13}C (x-axis) is determined using the isotopologues M0 to M6, where the sum of the signal of all isotopologues weighted by the number of ^{13}C substitutions is divided by the sum of the signals of all isotopologues times the total number of C atoms in the molecule (Equation 1). The calibrated atom % ^{13}C (y-axis) is based on the estimation of atom % ^{13}C from the mixed serial isotopic dilutions that ranged from 99 atom % ^{13}C to natural ^{13}C abundance (1.1 atom % ^{13}C) for U-^{13}C D-glucose, and from 17.4 atom % ^{13}C to natural ^{13}C abundance for D-glucosamine-1-^{13}C.

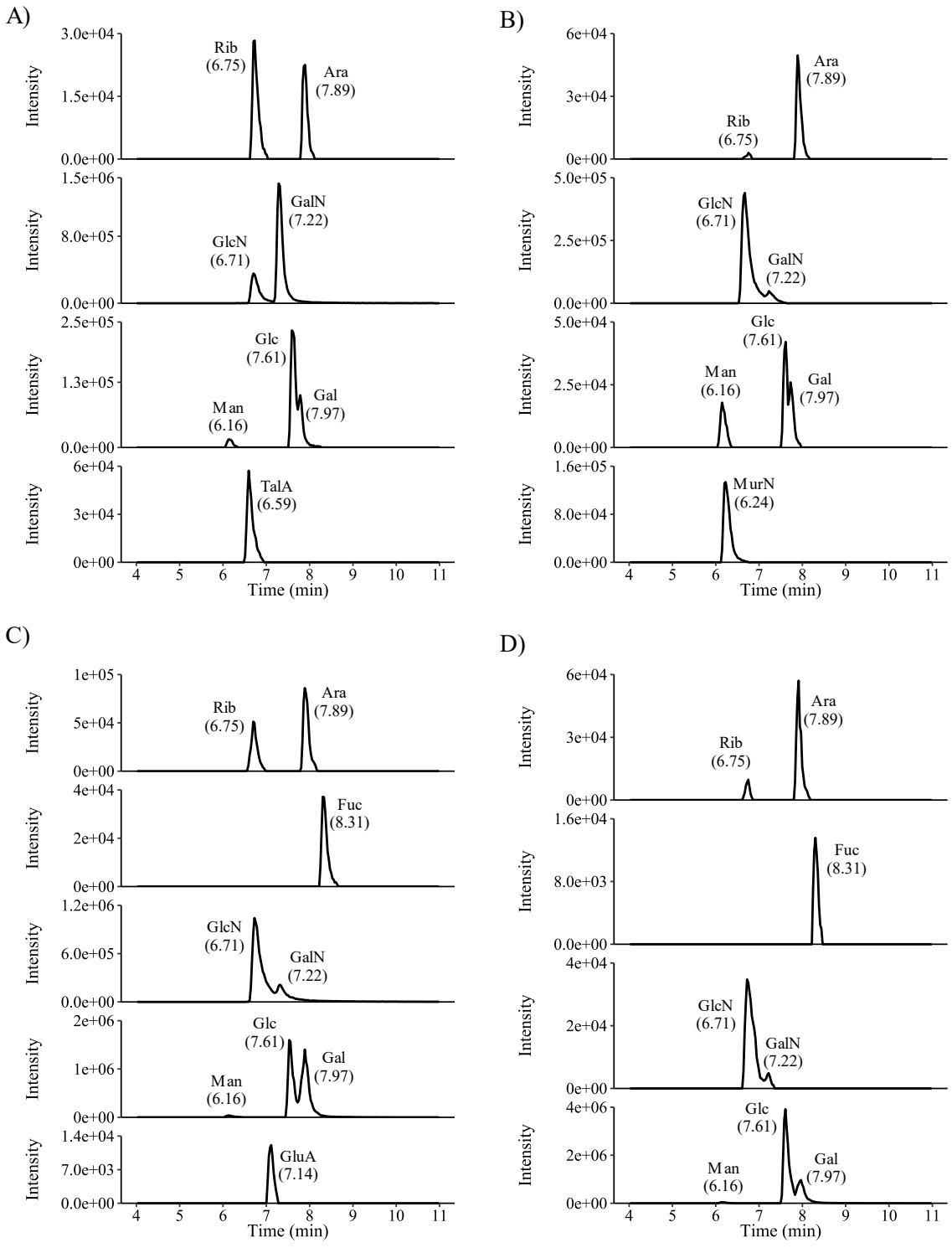

Figure 3. Chromatograms of PMP-derivatives from acid-hydrolyzed *Methanobacterium* sp. (A), *Streptomyces* sp. (B), *Clonostachys rosea* (C) and *Malva silvestris* (D).

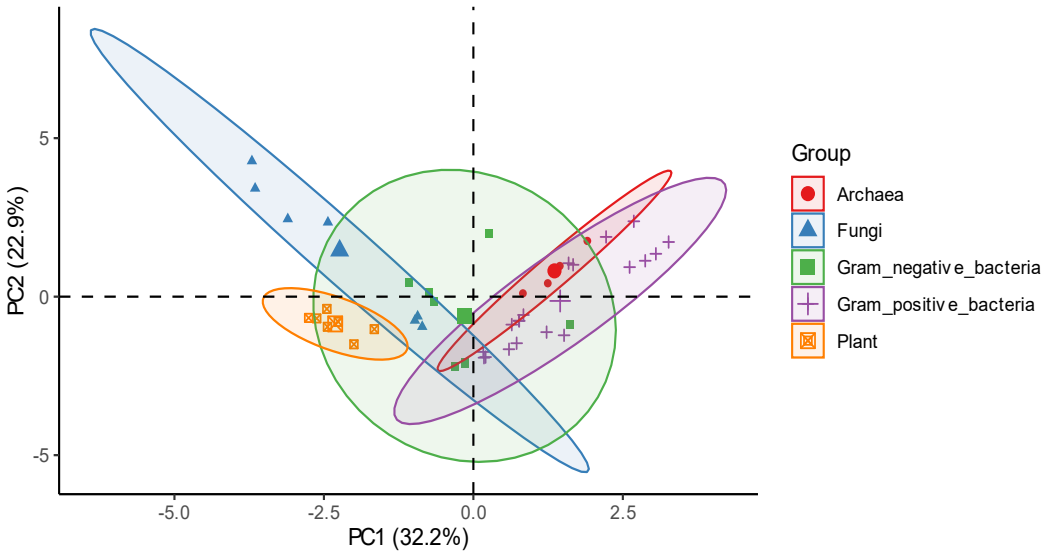

Figure 4. Biplot from principal component analysis (PCA) using PMP-derivatives from acid hydrolyzed samples of species belonging to different taxonomic groups. The small symbols represent all the individual replicates and the big symbols their centroid.

Supplements

Table S.1 Edaphic properties of forest, grassland and cropland soils used for testing the applicability of the PMP-derivatization method via UPLC-Orbitrap MS.

Site	Location	Latitude	Longitud	Management	Bedrock	Soil Type	pH*	WHC	Total N (mg/g)	Total C (mg/g)	Sand %	Silt %	Clay %	CEC
1	Andalucía, Spain	36°26'44"N	6°2'10"W	Forest	Carbonate	Calcisol	8.60	0.93	2.6	31.8	31	26.7	42.4	38.0
2	Andalucía, Spain	36°24'45"N	6°6'21"W	Grassland	Carbonate	Calcisol	8.50	0.74	2.4	26.4	23.8	43.8	32.4	26.6
3	Andalucía, Spain	36°45'43"N	5°11'42"W	Cropland	Carbonate	Calcisol	8.70	0.44	0.7	7.8	33.6	28.6	37.8	17.1
4	Styria, Austria	47°29'46"N	14°6'3"E	Cropland	Silicate	Cambisol	6.40	0.68	2.4	24.8	49.5	44.4	6.2	9.8
5	Styria, Austria	47°29'47"N	14°6'23"E	Grassland	Silicate	Cambisol	5.50	0.75	3.3	32.9	49.9	42.4	7.7	6.1
6	Styria, Austria	47°28'57"N	14°5'37"E	Forest	Silicate	Cambisol	4.40	0.89	2.1	44.1	49.3	34.1	16.6	6.4

*pH was measured using MQ 1:5 ratio.

WHC: Water holding capacity (g H_2O/g dw).

CEC: Cation exchange capacity (cmol/kg).

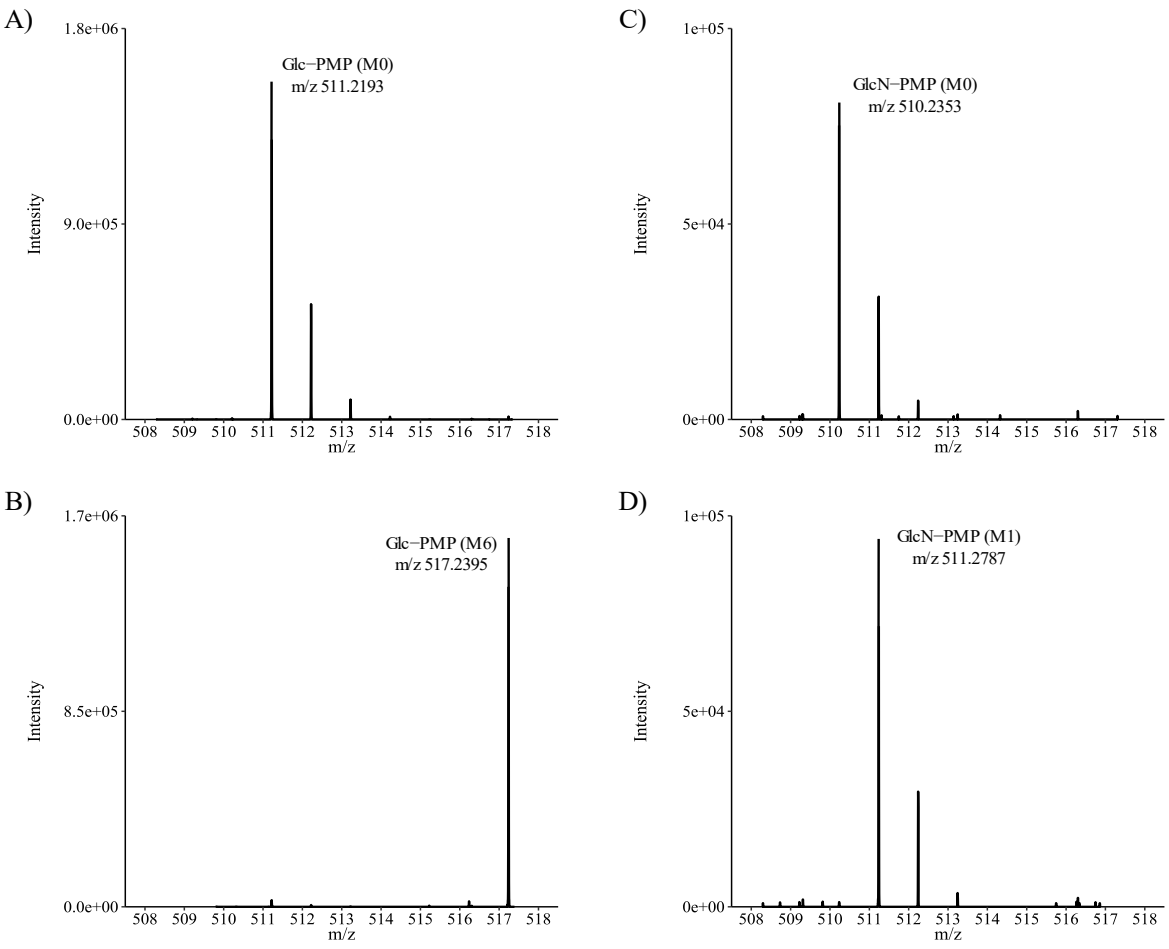

Figure S.1 Mass spectra of Glc-PMP at natural abundance ^{13}C (M0 = 511.2193 m/z, A) and the isotopically enriched variant U-^{13}C-Glc-PMP (M6 = 517.2395 m/z, B). Mass spectra of GlcN-PMP at natural abundance (M0 = 510.2353 m/z, C) and the isotopically enriched variant 1-^{13}C-GlcN-PMP (M1= 511.2387 m/z, D). In M0, all carbon atoms in the molecule are in their light isotopic variant (^{12}C). In M6, 6 carbon atoms in the molecule are in their heavy isotopic variant (^{13}C).

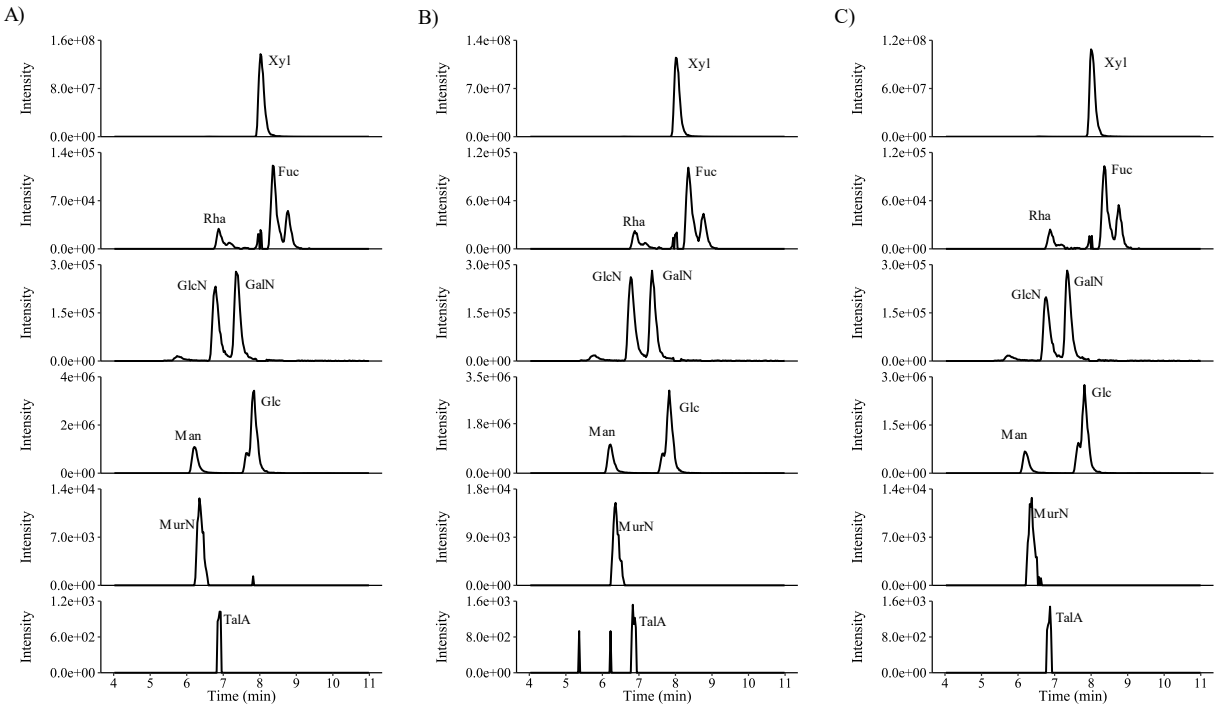

Figure S.2 Chromatograms of PMP-derivatives from acid-hydrolyzed Calcisol soils under forest (A), grassland (B) and cropland (C) management located in Andalucía, Spain.

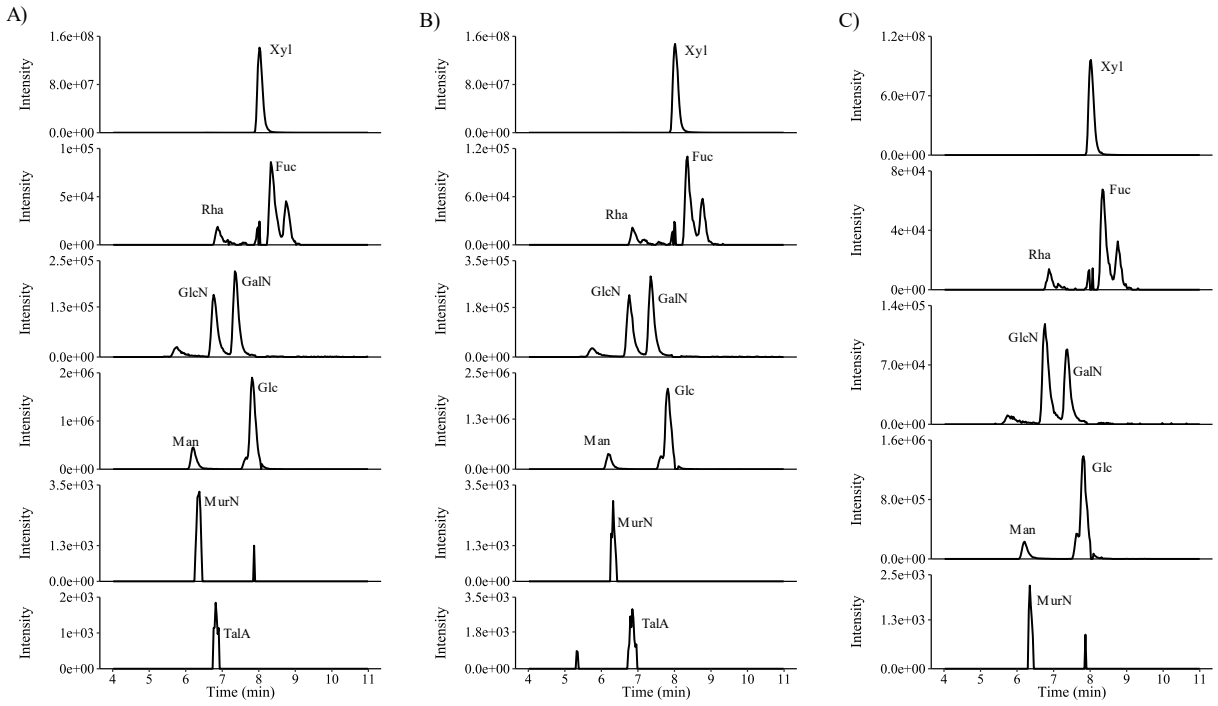

Figure S.3 Chromatograms of PMP-derivatives from acid-hydrolyzed Cambisol soils under cropland (A), grassland (B) and forest (C) management located in Styria, Austria.

Chapter III

Reevaluation and novel insights into amino sugar and neutral sugar necromass biomarkers in archaea, bacteria, fungi, and plants

Reevaluation and novel insights into amino sugar and neutral sugar necromass biomarkers in archaea, bacteria, fungi, and plants

Erika Salas[a,b*], Markus Gorfer[c], Dragana Bandian[c], Stephanie A. Eichorst[d], Hannes Schmidt[a], Julia Horak, Simon K.-M. R. Rittmann[e], Christa Schleper[f], Barbara Reischl[e], Thomas Pribasnig[f], Jan Jansa[g], Christina Kaiser[a], and Wolfgang Wanek[a]

[a] Division of Terrestrial Ecosystem Research, Centre for Microbiology and Environmental Systems Science, University of Vienna, Vienna, Austria
[b] Doctoral School in Microbiology and Environmental Science, University of Vienna, Vienna, Austria
[c] AIT Austrian Institute of Technology GmbH, Bioresources, Tulln, Austria
[d] Department of Microbiology and Ecosystem Science, Centre for Microbiology and Environmental Systems Science, University of Vienna, Vienna, Austria
[e] Archaea Physiology & Biotechnology Group, Department of Functional and Evolutionary Ecology, University of Vienna, Vienna, Austria
[f] Archaea Biology and Ecogenomics Unit, Department of Functional and Evolutionary Ecology, University of Vienna, Vienna, Austria
[g] Institute of Microbiology, Czech Academy of Sciences, Prague, Czech Republic

Abstract

Soil microbial necromass is an important contributor to soil organic matter (>50%) and it is largely composed of microbial residues. In soils, fragmented cell wall residues are mostly found in their polysaccharide forms of fungal chitin and bacterial peptidoglycan. Microbial necromass biomarkers, particularly amino sugars (AS) such as glucosamine (GlcN) and muramic acid (MurA) have been used to trace fungal and bacterial residues in soils, and to distinguish carbon (C) found in microbial residues from non-microbial organic C. Neutral sugars (NS), particularly the hexose/pentose ratio, have also been proposed as tracers of plant polysaccharides in soils. In our study, we extended the range of biomarkers to include AS and NS compounds in the biomass of 120 species belonging to archaea, bacteria, fungi, or plants. GlcN was the most common AS found in all taxa, contributing 42-91% to total AS content, while glucose was the most common NS found, contributing 56-79% to total NS. We identified talosaminuronic acid, found in archaeal pseudopeptidoglycan, as a new potential biomarker specific for Euryarchaeota. We compared the variability of these compounds between the different taxonomic groups using multivariate approaches, such as non-metric multidimensional scaling (NMDS) and partial least squares discriminant analysis (PLS-DA) and statistically evaluated their biomarker potential via indicator species analysis. Both NMDS and PLS-DA showcased the variability in the AS and NS contents between the different taxonomic groups, highlighting their potential as necromass residue biomarkers and allowing their extension from separating bacterial and fungal necromass to separating microbes from plants. Finally, we estimated new conversion factors where fungal GlcN is converted to fungal C by multiplying by 10 and MurA is converted to bacterial C by multiplying by 54. Conversion factors for talosaminuronic acid and galactosamine are also proposed to allow estimation of archaeal or all-microbial necromass residue C, respectively.

Key words

Soil necromass biomarkers, chitin, peptidoglycan, PLS-DA, necromass conversion factors.

1. Introduction

Soil organic matter (SOM) is the largest reservoir of terrestrial organic carbon (C), and soil microbes play a key role in the formation and long-term storage in this pool. While living microbial biomass is assumed to contribute approximately 1% to soil organic carbon (SOC) (Xu et al., 2013), dead microbes, also called microbial necromass, comprise on average 50% or more of the SOC (Liang et al., 2019; Wang et al., 2021). The accumulation and stabilization of microbial necromass, and consequently of SOC, is caused by organo-mineral interactions between necromass and mineral surfaces, organic-organic interactions between necromass and other organic matter, and necromass-necromass associations (Buckeridge et al., 2020; Creamer et al., 2019; Liang et al., 2019).

Soil microbial necromass is largely made up of microbial residues and products such as fragmented cell walls, cytoplasmic components, and extracellular polymeric substances (EPS) from fungi, bacteria, and archaea. Microbial residues, particularly cell wall fragments of fungi and bacteria, are mostly found in polysaccharide form such as chitin and peptidoglycan (Kästner et al., 2021). In soils, cell wall residues are decomposed and recycled by extracellular hydrolytic enzymes, resulting in a mixture of oligo- and monosaccharides, as well as amino acids and oligopeptides, which makes it difficult to distinguish their origin (Hu et al., 2018; Schimel & Bennett, 2004). Biomarkers of soil microbial necromass have been used to differentiate between C found in microbial residues from non-microbial organic C (Liang et al., 2019).

Amino sugars are the most commonly used biomarkers to trace bacterial and fungal necromass residues in soils, since plants only produce them in trace amounts (Parsons, 2021). Glucosamine (GlcN) is the most common amino sugar found in soils, contributing up to 68% of total amino sugars in soils (Joergensen, 2018). GlcN is the main component of fungal chitin, and one of the main components of bacterial peptidoglycan and archaeal pseudopeptidoglycan (Albers & Meyer, 2011; Joergensen, 2018; Parsons, 2021). Muramic acid (MurA) is considered the most specific microbial necromass biomarker because it is found exclusively in the murein layers of the bacterial peptidoglycan (Vollmer et al., 2008). Galactosamine (GalN) and mannosamine (ManN) are nonspecific amino sugars which can be found in plants (Indorf et al., 2011), fungal cell walls (Beauvais & Latgé, 2018), EPS of bacteria (Caruso et al., 2018; Cheah & Chan, 2022; Vaningelgem et al., 2004), and in the cell envelope of some archaeal species (Albers & Meyer, 2011). Talosaminuronic acid

(TalA) can be found as part of the archaeal pseudopeptidoglycan (Albers & Meyer, 2011); however, its biomarker potential has yet to be demonstrated.

Beside amino sugars, neutral sugars are also potential biomarkers for necromass in soils. They can be found in their polymeric form as building blocks of cellulose, hemicellulose, starch, and glycogen. Due to the decomposition of plant litter and rhizodeposits, plants are considered the primary sources of sugars in soils, while microbes are a secondary source since they synthesize their carbohydrates from other C sources (e.g., from plant litter) (Gunina & Kuzyakov, 2015). Hexoses account for most of the bound sugars found in soils, with glucose (Glc) being the most abundant one, followed by mannose (Man) and galactose (Gal) (Gunina & Kuzyakov, 2015). Pentoses, particularly arabinose (Ara) and xylose (Xyl), are thought to mostly originate from plant residues since microbes preferentially synthesize hexoses compared to pentoses (Apostel et al., 2015; Gunina & Kuzyakov, 2015). A hexose/pentose ratio GM/AX (galactose + mannose/arabinose + xylose) has thus been proposed to determine the plant vs. microbial origin of polysaccharides in soil, where a GM/AX > 2 is indicative of microbial polysaccharides and a GM/AX < 0.5 is indicative of plant polysaccharides (Gunina & Kuzyakov, 2015; Oades, 1984). Other “rare” sugars and sugar derivatives, such as fucose, rhamnose, and uronic acids have also been found in EPS and other polymers attached to the cell walls of bacteria and archaea (Albers & Meyer, 2011; Bhavsar et al., 2004; Fazio et al., 1982; Roca et al., 2015) and in plant hemicelluloses (Scheller & Ulvskov, 2010).

In order to estimate the contribution of microbial necromass to SOM, Appuhn and Joergensen (2006) developed conversion factors based on the GlcN and MurA content in the biomass of selected fungal and bacterial species. These conversion factors convert soil GlcN or MurA content into fungal or bacterial necromass C, respectively. However, the extrapolation using these conversion factors is based on critical assumptions. For example, the MurA to bacterial C conversion factor of 45 was calculated assuming a ratio of gram-positive to gram-negative bacteria of 65% to 35% in soils (Appuhn & Joergensen, 2006; Joergensen & Potthoff, 2005), which could potentially under- or overestimate the contribution of bacterial C depending on the ecosystem (Cao et al., 2023; Fanin et al., 2019). Other assumptions include the same stabilization and recycling rates in soils for all amino sugars, and the same amino sugar ratios in microbial biomass and necromass.

Microbial necromass contribution to SOM and long-term C storage has gained important recognition, and with it the use of microbial necromass biomarkers. Thus far, necromass biomarker studies have focused mainly on GlcN and MurA as biomarkers for fungi and bacteria (Angst et al., 2021; Hu et al., 2023; Wang et al., 2021). There is a major research gap regarding other potential amino sugar biomarkers, as well as the potential of neutral sugars as biomarkers. Additionally, the knowledge of the content and variation of these biomarkers in representative soil microbial groups, including archaea, as well as plants is still lacking. In our study, we aimed to expand the repertoire of biomarkers by: (1) identifying and quantifying amino sugar and neutral sugar compounds in the biomass of archaea, bacteria, fungi, and plants, (2) determining the variability of these compounds amongst the taxonomic groups in order to evaluate their biomarker potential, and (3) deriving new conversion factors that estimate microbial residue C based on specific amino sugar biomarkers. We used a single and robust approach to measure for the first-time amino sugar and neutral sugar biomarkers in a wide variety of taxa using the same hydrolysis method, derivatization protocol and analytical conditions via ultra-high performance liquid chromatography (UPLC) coupled to Orbitrap high-resolution mass spectrometry (HRMS).

2. Materials and methods

2.1 Standards and reagents

Amino sugar and neutral sugar standards [arabinose (Ara), deoxyribose (DRib), fucose (Fuc), galactosamine (GalN), galactose (Gal), galacturonic acid (GalA), glucosamine (GlcN), glucose (Glc), glucuronic acid (GluA), mannosamine (ManN), mannose (Man), muramic acid (MurA), rhamnose (Rha), ribose (Rib), and xylose (Xyl)], 1-methyl-3-phenyl-2-pyralozone (PMP), ammonium formate solution BioUltra 10 M in H_2O, ammonia solution 25%, formic acid solution BioUltra 1 M in H_2O, hydrochloric acid and acetonitrile (hypergrade for LC-MS) solutions were purchased from Sigma Aldrich (St. Louis, MO).

2.2 Sample biomass preparation

We analyzed biomass from 120 species belonging to archaea (13), bacteria (23), fungi (56), or plants (28) with the aim to capture a large diversity of soil microbial groups, as well as plants. These microbial species belonged to representative clades

of soil microorganisms (Fierer, 2017), such as Crenarchaeota and Euryarchaeota for archaea; Acidobacteria, Bacteroidetes, Proteobacteria, Verrucomicrobia, Actinobacteria and Firmicutes for bacteria; and Ascomycota, Basidiomycota and Mucuromycota for fungi. Microbial cultivation methods can be found in the supplementary methods. The biomass from bacterial, fungal, and archaeal species was freeze-dried prior to acid hydrolysis. Plant materials (leaves and/or roots obtained during the peak growing season in and around Vienna, Austria) were dried in paper bags at 70 °C for 24 h or purchased from Herbathek® (Berlin, Germany). Acid hydrolysis and PMP derivatization was carried out as previously described in Salas et al. (2023). Briefly, 2 mg of pure freeze-dried isolate biomass or dried plant material were hydrolyzed using 1.5 mL of 6 M HCl and heated to 105 °C for 16 h. After cooling, the solution was centrifuged at 10,000 g for 20 min and dried under a nitrogen stream. To remove any salts, 1 mL of MS-grade methanol was added which dissolves amino sugars and neutral sugars released from cell mass but not other precipitates (Pinho & Macedo, 2005). The solution was centrifuged at 10,000 g for 10 min and dried under a nitrogen stream. The dried samples were dissolved in 1 mL MQ water prior to derivatization using 1-phenyl-3-methyl-5-pyrazolone (PMP).

Aliquots of 100 µL of hydrolyzed samples were derivatized by adding 100 µL 0.5 M PMP and 200 µL ammonia 25%. The samples were incubated in a 70 °C water bath for 20 min. After cooling, 200 µL of 1 M formic acid were added to neutralize the reaction. Excess PMP was removed via liquid:liquid extraction by adding 1 mL chloroform and 1 mL water to the derivatized samples. The samples were mixed thoroughly, and the organic layer (bottom) was discarded. This process was repeated twice. Finally, the aqueous layer was filtered through a 25 mm syringe filter with a 0.2 µm acetate membrane (VWR International™).

Derivatized samples were injected into a UPLC Ultimate 3000 system (Thermo Fisher Scientific, Bremen, Germany) coupled to an Orbitrap Q Exactive HRMS system (Thermo Fisher Scientific) with an electrospray ionization (H-ESI) source. A Waters AccQ.Tag Ultra C18 column (2.1 mm x 100 mm, 1.7 µm particles) (Milford, MA, USA) coupled to an ACQUITY column in-line guard filter (2.1 mm, 0.2 µm) (Milford, MA, USA) was used. The UPLC and ESI-Orbitrap HRMS settings used were the same as outlined in Salas et al. (2023).

2.3 Data analysis

The following mass to charge ratio (*m/z*) values were used to identify and quantify amino sugar and neutral sugar biomarkers: 510.2349 for GlcN, GalN and ManN; 582.2554 for MurA; 524.2133 for TalA; 511.2186 for Glc, Gal and Man; 481.2079 for Ara, Rib and Xyl, 495.2235 for Fuc and Rha; 525.1978 for GluA and GalA; and 465.2135 for DRib. Stereoisomers were identified based on their different retention times (Salas et al., 2023). Data was extracted and compounds were identified and quantified by their mass spectrometric signals using the FreeStyle™ 1.7 software (Thermo Scientific).

All statistical analyses were carried out using R version 4.2.3 (R Core Team, 2021). Amino sugar and neutral sugar content data were checked for outliers and log-transformed when necessary to obtain homoscedasticity and normality. Differences in amino sugar and neutral sugar content between taxonomic groups were tested using Kruskal-Wallis, Dunn and Wilcoxon tests. Spearman correlations were used to test correlation between MurA and GlcN in all bacterial groups.

We carried out non-metric multidimensional scaling (NMDS) analyses to visualize patterns of clustering among the different taxa based on their amino sugar and/or neutral sugar content. We performed NMDS analyses using the metaMDS function of the vegan package (Oksanen et al., 2008) and vegdist to create the Bray-Curtis dissimilarity matrix. To test whether the groups differed based on their amino sugar and neutral sugar biomass content, we performed analysis of similarity (ANOSIM) using multiple permutations (n = 9999). Additionally, to discriminate between taxonomic groups based on their amino sugar and neutral sugar content, we used partial least squares discriminant analysis (PLS-DA), a supervised method based on partial least squares regression (PLS) analysis where the response variable is categorical (in this case, the taxonomic groups). PLS-DA was carried out using leave-one-out cross-validation in the mixOmics package (Rohart et al., 2017). PLS-DA loadings from components 1 and 2 were used to detect which amino sugar and neutral sugar biomarkers were the most discriminatory between the different taxonomic groups.

To identify amino sugar and neutral sugar biomarkers associated to specific taxonomic groups, we ran an indicator species analysis using the multipatt function from the indcspecies package (Caceres & Legendre, 2009) using multiple permutations (n = 9999).

We calculated conversion factors from specific amino sugar biomarkers into biomass C for each taxon. These conversion factors were calculated based on: (1) the weighted averages of amino sugar content found in the hydrolyzed biomass of each group, and (2) the assumption that fungal, bacterial, archaeal and plant dry biomass contain approximately 46% organic C (Jenkinson, 1988; Joergensen & Potthoff, 2005; Ma et al., 2018).

3. Results

We analyzed amino sugar and neutral sugar contents in pure isolates of 120 species belonging to archaea, bacteria, fungi, and plants using the same acid hydrolysis and PMP derivatization method via UPLC-Orbitrap MS (Table S1). In most cases, bacterial samples were separated into two major groups: gram-positive bacteria and gram-negative bacteria, to showcase the different biomarker content between these two groups.

3.1 Amino sugar and neutral sugar content in pure hydrolyzed biomass

Glucosamine was the most common amino sugar found in the hydrolyzed biomass of all taxonomic groups and represented 42 - 91% of total amino sugar content (Table 1). Both fungi and gram-positive bacteria had the highest GlcN content with an average of 45.9 μg GlcN mg^{-1} dry weight (dw) and 37.3 μg GlcN mg^{-1} dw, respectively. Plants had the lowest GlcN content with 1.4 μg mg^{-1} dw.

Muramic acid was found only in bacterial samples, where it was higher in gram-positive bacteria than in gram-negative bacteria, with an average of 24.7 μg MurA mg^{-1} dw and 3.9 μg MurA mg^{-1} dw respectively (Table 1). We calculated a MurA:GlcN molar ratio of 1:2 for gram-positive bacteria and 1:4 for gram-negative bacteria. We observed a positive significant correlation between GlcN and MurA only in gram-positive bacteria, but not in gram-negative bacteria (Fig. 1A). When carried out within the phyla of gram-positive bacteria, we found a strong positive correlation between GlcN and MurA in Firmicutes, but not in Actinobacteria (Fig. 1B). There was no correlation between GlcN and MurA in any of the phyla of gram-negative bacteria (Fig. S1).

Galactosamine content was highest in archaea with 15.2 μg GalN mg^{-1} dw. In bacteria, fungi and plants GalN content only ranged from 1.1 to 4.0 μg GalN mg^{-1} dw

(Table 1). We calculated GalN:GlcN molar ratios of 1:1 for archaea and plants, and 1:10 for fungi and bacteria.

Talosaminuronic acid was only found in archaea (Table 1), specifically in species belonging to Euryarchaeota, with an average of 6.4 µg TalA mg^{-1} dw (Table S2). A TalA:GlcN molar ratio of 1:2 was calculated for Euryarchaeota.

Hexoses (glucose, galactose, and mannose) were the most abundant neutral sugars found in the biomass of all groups, with glucose representing 56-79% of the total neutral sugars measured in this study (Table 2). We observed higher hexose contents in fungi and plants compared to bacteria and archaea. Gram-positive bacteria had negligible contents of Fuc, Xyl and GluA. GluA was not detected in any plant samples (Table S3). DRib content was similar between all groups, where it ranged from 36.1 ng mg^{-1} dw in plants to 50.0 ng mg^{-1} dw in fungi.

Table 2 showcased the GM/AX ratios for all taxonomic groups. Bacterial groups had the lowest GM/AX ratios of 3.2 and 9.0 for gram-positive bacteria and gram-negative bacteria, respectively. The highest GM/AX ratios were found in fungi and plants with 24.5 and 22.6, respectively. Archaea had an intermediate GM/AX value of 13.2.

ManN and GalA could not be quantified for almost all species and are therefore not included in the data analyses.

3.2 Amino sugars and neutral sugars as necromass biomarkers

The NMDS ordination using only amino sugar biomarkers showed the separation of gram-positive and gram-negative bacteria from the clusters of archaea, fungi, and plants, which overlapped (Fig. 2A, ANOSIM's $R = 0.59$, $P < 0.001$). This bacterial separation was driven by MurA (Fig. 2D). When considering only neutral sugar biomarkers, the NMDS showed a tight clustering of fungi and plants (eukaryotes) compared to archaea and bacteria (prokaryotes; Fig. 2B, ANOSIM's $R = 0.34$, $P < 0.001$). Glc and Man content drove the tight clustering of the eukaryotes (Fig. 2E). Finally, NMDS analysis using both amino sugar and neutral sugar biomarkers showed the separation of plants from all other microbial groups (Fig. 2C and 2F, ANOSIM's $R = 0.63$, $P < 0.001$).

The results of the PLS-DA model using amino sugar and neutral sugar biomarkers showed the separation of bacteria, fungi and plant groups, with components 1 and 2 explaining 29% of the variance (Fig. 3A). MurA was the major

contributor to the separation of bacteria, while GlcN was the major contributor to the separation of fungi (Fig. 3B and 3C). Plants were mainly separated due to their Glc and Gal content. The receiver operating characteristic (ROC) curves showcasing the performance of the PLS-DA model for the classification of each taxonomic group are shown in Figure S2.

The indicator species analysis showed the association of different taxonomic groups to specific biomarkers (Table 3). GlcN was indicative of archaea, bacteria, and fungi, while MurA was a specific indicator for bacteria. Both GalN and TalA were specific indicators for archaea. Glc and Gal were indicators for fungi and plants. Fuc was indicative of plants, Rib of archaea, and both Xyl and GluA were indicators for fungi. DRib was the only sugar associated to all groups.

We estimated conversion factors that convert amino sugar content (e.g., µg MurA/g dw) to microbial necromass C (e.g., bacterial necromass C/g dw). We calculated a GlcN conversion factor of 10 for fungal C, 12 for gram-positive bacterial C, 41 for gram-negative bacterial C, 29 for archaeal C and 329 for plant C (Table 1). Using the averages of MurA content in bacteria, we estimated conversion factors from MurA to bacterial C of 19 for gram-positive bacteria and 118 for gram-negative bacteria. If a proportion of 35 % to 65% for gram-negative to gram-positive bacteria in soils is assumed (Appuhn and Joergensen, 2006), we calculated a conversion factor from MurA to bacterial C of 54. Based on the TalA content found in the pure biomass of euryarchaeota (Table S2), we estimated a TalA conversion factor of 72 for euryarchaeota C.

4. Discussion

4.1 Amino sugar and neutral sugar contents in hydrolyzed biomass

Our data showed that GlcN was the most abundant amino sugar, with fungi and gram-positive bacteria presenting the highest GlcN contents in their biomass (Table 1). Fungal GlcN contents were similar to those presented in other studies (Appuhn & Joergensen, 2006; Engelking et al., 2007; Glaser et al., 2004). Most fungal GlcN originates from the *β*(1-4) linked *N*-acetylglucosamine in chitin and *β*(1-4) linked glucosamine in chitosan (naturally occurring deacetylated chitin), where it can represent up to 15% of the fungal cell wall mass (Arroyo et al., 2016). Moreover, our data showed the largest variation in the GlcN content between fungal species (Table

S1), which is often explained by phylogenetic differences in cell wall composition (Gow et al., 2017).

The MurA content in gram-negative bacteria was similar to that presented by previous authors (Appuhn & Joergensen, 2006; Engelking et al., 2007). In contrast, we measured a higher MurA content in gram-positive bacteria, which could be due to the different species included in this study as compared to the previous ones. The large range of MurA contents between the upper and lower confidence limits in gram-positive bacteria has also been previously observed (Appuhn & Joergensen, 2006). Additionally, since bacterial murein contains alternating *N*-acetylglucosamine and *N*-acetylmuramic acid linked by *β*(1-4) bonds (Subedi et al., 2021), the thicker peptidoglycan cell wall of gram-positive bacteria resulted in the higher GlcN and MurA content when compared to gram-negative bacteria (Table 1).

We observed a positive correlation between the GlcN and MurA content only in gram-positive bacteria, specifically only in Firmicutes (Fig. 1A, 1B). This suggests that most of the GlcN in Firmicutes can be found in their peptidoglycan structure. This also indicates that MurA and GlcN co-variations in Firmicutes are mostly deriving from species differences in cell wall thickness (Megrian et al., 2020), while cell wall thickness in the species of Actinobacteria was similar (as observed from relatively constant MurA contents, Fig 1B). On the contrary, the higher GlcN content relative to MurA in Actinobacteria and gram-negative bacteria can be due to GlcN originating from other structural components of bacterial cell walls aside from peptidoglycan. For example, GlcN is a component of lipid A in lipopolysaccharides of gram-negative bacteria (Raetz et al., 2007), and it can be found alongside ManN in teichoic acids (TAs), where TAs can constitute up to 60% of the cell wall mass of gram-positive bacteria (Brown et al., 2013).

In archaea, GlcN probably originated from the glycoproteins on the surface layer (S-layer) or from pseudopeptidoglycan in the cell wall. Pseudopeptidoglycan contains a backbone of *β*(1-3) linked *N*-acetylglucosamine and *N*-acetyltalosaminuronic acid (Subedi et al., 2021), and this has only been found in a few methanogens (Albers & Meyer, 2011). This would also explain why TalA was only found in euryarchaeota species (Table S2).

Similarly, we found high concentrations of GalN in archaea (Table 1). Some Methanosarcinales produce methanochondroitin during cell wall formation (Albers & Meyer, 2011). Methanochondroitin is composed of two *N*-acetylgalactosamine

molecules per GluA, which could explain the high GalN content in this archaeal group. In fungi, GalN could originate from galactosaminogalactan (GAG), a polysaccharide composed of GalN and Gal that is associated to cell walls and plays an important role in fungal pathogenesis (Bardalaye & Nordin, 1976; Latgé & Beauvais, 2014). GalN is also a common component in the bacterial cell wall, where it can be found in lipopolysaccharides (Leyn et al., 2012) and teichoic acids (Freymond et al., 2006), which explains the presence of this amino sugar in bacterial species.

Plants have been known to have trace amounts of amino sugars, which coincided with the low GlcN and GalN content we measured from their hydrolyzed biomass (Table 1). The GlcN and GalN content we detected in plants most likely originated from glycoproteins in the plant cell wall (Nguema-Ona et al., 2014) and to a small extent from microbial endophytes (Appuhn & Joergensen, 2006).

Neutral sugar data showed Glc as the most common neutral sugar found in all taxonomic groups, particularly in fungi and plants (Table 2). Fungal glucan contains high amounts of Glc and can represent 50-60% of the cell wall dry weight (Garcia-Rubio et al., 2020), which coincides with the high Glc content we found in fungi. In plants, cellulose and hemicellulose are mainly composed of linked Glc, but plant hemicelluloses and glycoproteins can contain substantial amounts of Gal, Man, Ara and Xyl (Nguema-Ona et al., 2014; Scheller & Ulvskov, 2010).

We also detected more rare neutral sugars, such as Fuc, Rha and GluA, in almost all taxonomic groups. These rare neutral sugars have been found in cell wall glycoproteins (Messner, 1997) and in microbial EPS (Roca et al., 2015). Fuc, Rha and GluA are important constituents of glycoproteins of all prokaryotes and eukaryotes (Eichler, 2020; Nguema-Ona et al., 2014; Tra & Dube, 2014) and are complex and diverse macromolecules that can be found in almost all organisms (Messner, 1997). Planktonic isolates as analyzed as biomarker source in this study do not produce EPS in large quantities. However, in environmental samples common and rare sugars also derive from EPS to a substantial proportion (Flemming & Wingender, 2010) but there is still quite an extensive research gap regarding the identity and stability of compounds found in EPS. The use of EPS extraction methods (Redmile-Gordon et al., 2014) could improve the current understanding on the composition of EPS is soils, the stability of EPS compounds and their role on soil aggregation.

The GM/AX ratio assumes that values > 2 are indicative of microbial polysaccharides, while values < 0.5 are indicative of plants (Gunina & Kuzyakov, 2015;

Oades, 1984). These values were based on the incorporation of ^{14}C into sugars in soils amended with ^{14}C-glucose or ^{14}C-dextran (Oades & Wagner, 1971). Our measurements based on pure isolates and plant biomass revealed a wider range in the GM/AX ratio, where it varied from 2.1 in gram-positive bacteria to 22.9 in plants (Table 2). This shows that not only plants have considerable amounts of hexoses (Gal and Man), but also microbial polysaccharides contain significant amounts of pentoses (Ara and Xyl). Therefore, our data reveals that the GM/AX ratio is an inadequate indicator for the origin of neutral sugars as it behaved inverse to the pattern proposed for plants versus microbes (GM/AX proposed values should indicate microbes > plants, but our data shows plants ≥ microbes). The current analyses highlight a different biomarker potential of this GM/AX ratio, differentiating between a plants/fungi group (GM/AX > 22) and a bacterial group (GM/AX < 9), however this requires further confirmation. Archaea had intermediate values but contribute very little to soil bound sugars (Bates et al., 2011).

Strong acid hydrolysis (6 M HCl) is required to depolymerize amino sugars; however, such acidic conditions can degrade the released monosaccharides, particularly neutral sugars (Liu et al., 2021). While other hydrolysis methods, such as 4 M TFA, are used as a mild alternative to depolymerize neutral sugars from polysaccharides, it does not completely depolymerize the amino sugar bonds (Garna et al., 2004; Liu et al., 2021). Therefore, the choice of acid hydrolysis method must be a compromise between the best depolymerization of polysaccharides for amino sugars with the least degradation of released monosaccharides. Since amino sugars are the most used necromass biomarkers for soil microbial groups, 6 M HCl hydrolysis was chosen to ensure the efficient depolymerization of amino sugar polymers at the expense of possibly degrading neutral sugars. To ensure the total depolymerization of both amino sugars and neutral sugars from microbial and plant residues, we recommend using both 6 M HCl and 4 M TFA hydrolysis methods in parallel.

Some biomarkers could not be detected in certain species (Table S1). This could be due to: (1) biomarkers not being present in a specific species, (2) biomarkers being depolymerized during the acid hydrolysis, or (3) biomarker content in the hydrolyzed biomass is lower than the limit of detection of the UPLC-Orbitrap MS system (Salas et al., 2023).

4.2 Amino sugars and neutral sugars as necromass biomarkers

By focusing only on the amino sugar content, the NMDS showed the separation of bacterial groups, from archaea, fungi, and plants (Fig. 2A). MurA content drove this separation (Fig. 2D), therefore showing the importance of this amino sugar to trace necromass residues from bacteria. This was also corroborated by the PLS-DA (Fig. 3). When only neutral sugar content is considered, the high hexose content in fungi and plants is most likely the cause of the tighter clustering of these groups compared to archaea and bacteria (Fig. 2B & 2E). Even though there was a significant separation of the taxonomic groups based on their neutral sugar content, these sugars are ubiquitous and therefore should not be considered alone as specific necromass biomarkers.

When both amino sugar and neutral sugar contents were included, the NMDS showed the separation of plants from all other microbial groups (Fig. 2C). Both the low amino sugar content and the high neutral sugar content, particularly hexoses, in plants produced this separation (Fig. 2F). This suggests that the combination of both amino sugar and neutral sugar profiles can be used to separate plants from microbial groups, which holds great potential to use the same set of sugar-related molecular compounds to separate plants from fungi and bacteria when using multivariate methods instead of relying only on single compounds. PLS-DA demonstrated the clear separation of plants, fungi and bacteria based on amino sugars and neutral sugars, driven by contributions from MurA, GlcN, Glc, Gal, and Xyl (Fig. 3). Separation between organic matter derived from microbes and plants is otherwise performed by plant-compounds absent in microbes, such as lignin monomers and specific alkanes which requires separate analysis approaches (Angst et al., 2021).

Multivariate necromass attribution methods have not yet been applied to microbial and plant residues in soils and would involve a two-step process: (1) group definition and separation by mixed biomarker signals in a multivariate space, and (2) source mixing or source attribution modeling based on multivariate group separators. Group separation by multicomponent analysis is based on discriminant analysis (DA) such as linear discriminant analysis (LDA) and PLS-DA (Oliveri et al., 2021; Ruiz-Perez et al., 2020). Source attribution models are based on the multivariate definition of the sources including error terms and Bayesian source mixing models, which was developed for stable isotope in animal nutrition and related research but can be extended to any biomarkers (Hopke, 2016; Stock et al., 2018). Such an approach

holds great potential to bring biomarker analysis to a new level using machine learning approaches.

The indicator species analysis showed that GlcN was an indicator of archaea, bacteria and fungi showcasing its function as a commonly used microbial necromass biomarker (Table 3). MurA was only associated to bacteria due to its sole presence in the peptidoglycan structure. This showcases MurA as a robust bacterial necromass biomarker. Similarly, both GalN and TalA were only associated with archaea, likely due to pseudo-murein and methanochondroitin occurrence. Since TalA has only been found in methanogens (Subedi et al., 2021), we highlight the potential of TalA as a new necromass biomarkers specific for these organisms.

In soils, amino sugars from microbial necromass are stabilized in soils as polymers due to organo-mineral interactions, organic-organic interactions and necromass-necromass associations (Buckeridge et al., 2020; Creamer et al., 2019). Due to the decomposition of the homopolysaccharide cellulose, the most abundant biopolymer on Earth, Glc is the most abundant neutral sugar released in soils (Gunina & Kuzyakov, 2015). Once polysaccharides are decomposed, oligo- and monosaccharides become easily available for microbes and are therefore taken up by microbes and used for growth, maintenance metabolism and C storage. Additionally, sugar polymers can be adsorbed to other soil components such as organic matter or clay particles and thereby play an important role in aggregate formation, or they are leached to deeper soil horizons within dissolved organic matter (DOM) and/or more rarely taken up by plants (Gunina & Kuzyakov, 2015). Consequently, amino sugars and neutral sugars can only be stabilized in soils long-term if they are found in recalcitrant polymeric forms such as cellulose or microbial residues (Dungait et al., 2012). The decomposition, recycling, and stabilization of amino sugars and neutral sugars in soils should therefore be considered when using these compounds as necromass residue biomarkers.

4.3 Conversion factors derived from the amino sugar content in biomass

Most of the necromass biomarker studies use conversion factors to extrapolate amino sugar content measured in soils into microbial C. However, these conversion factors were derived only from the GlcN and MurA biomass content in selected fungal and bacterial species. In our study, we estimated conversion factors that allow to convert amino sugar content in soils into microbial necromass residue C for all

taxonomic group (Table 1). These conversion factors were derived from the amino sugar content found in the biomass of each taxon. This is the first-time GlcN conversion factors were calculated for archaea, bacteria, and plants. It is also the first-time conversion factors specific for GalN and TalA have been proposed for archaea.

Appuhn and Joergensen (2006) estimated a conversion factor of 9 for fungal GlcN with 95% confidence limits ranging from 8 to 10. While the conversion factor estimated in this study is similar (10), we calculated a much larger 95% confidence interval (9 to 55). Overall, our study not only included a larger diversity of fungal species than those presented by the former authors, but also included more species belonging to Ascomycota, which generally present a wide range of fungal cell wall structures and cell wall chitin contents (Erwig & Gow, 2016). Given that we also measured more than twice the amount of GlcN in Ascomycota and Mucuromycota than in Basidiomycota (Table S2), this could explain the large variability in the average fungal GlcN content.

We estimated similar MurA conversion factors than those presented by Appuhn and Joergensen (2006); however, we again calculated a larger 95% confidence interval. We estimated two conversion factors for MurA: 19 for gram-positive bacteria and 118 for gram-negative bacteria (Table 1). Appuhn and Joergensen (2006) assumed a representative ratio of 65% to 35% gram-positive to gram negative bacteria in soils to calculate a single conversion factor of 45 for MurA into bacterial C. When calculated using the same bacterial ratio, we obtained a MurA to bacterial necromass C conversion factor of 54. The difference between both conversion factors, as well as the larger variation observed (Table 1), is likely caused by the inclusion of a larger diversity of bacterial species, as well as a higher MurA content measured in gram-positive bacteria than that presented by Appuhn and Joergensen, (2006). Fanin et al. (2019) showed that the gram-positive to gram-negative bacteria ratios differed between soil ecosystem conditions using phospholipid-derived fatty acid analysis (PLFAs). A global analysis also showed that gram-positive to gram-negative bacteria ratios predictably vary with soil moisture and pH (Cao et al., 2023). Therefore, we suggest determining the proportion of gram-positive to gram-negative bacteria in the corresponding soils in order to estimate accurately the contribution of bacterial residue C to SOM using a dynamic MurA conversion factor tailored to the present gram-positive to gram-negative bacteria ratio of the respective soil.

We propose TalA as a new biomarker for archaeal species, specifically for methanogens within Euryarchaeota. The archaeal cell wall and envelope are quite diverse between species and across phylogeny (Albers & Meyer, 2011). To extrapolate a conversion factor of TalA into archaea-derived C in soils, we propose that it is necessary to know the ratio of archaeal groups that do not contain TalA in their cell envelope (such as Crenarchaeota and Thaumarchaeota) to TalA-containing archaea that occur in soils, e.g., using 16S-gene amplicon information (Bates et al., 2011; Starke et al., 2021). Salas et al. (2023) were able to quantify TalA from soil samples, showcasing the potential use of this amino sugar as a soil necromass biomarker.

While the GlcN and MurA conversion factors calculated in this study were similar to those proposed previously (Appuhn & Joergensen, 2006; Engelking et al., 2007), the larger 95% confidence intervals can have important implications when using conversion factors to estimate necromass residue C contribution to SOM. This means that the interspecies variation of amino sugar contents should be considered via error propagation when extrapolating amino sugar measurements into necromass residue C. This is particularly important for the GlcN conversion factor since GlcN content in fungi is highly variable (Table 1). This could be done by including the 95% confidence intervals in the extrapolations using amino sugar conversion factors. Similarly, the representative nature of the organisms used for the quantification of amino sugar contents should also be considered since assessing a higher diversity of species would lead to the derivation of more accurate conversion factors.

4.4 Considerations for the application of conversion factors in soils

To accurately apply the conversion factors calculated in this study as a proxy for the contribution of necromass residues to SOM, certain aspects need to be considered. First, the amino sugar content in soils should be ideally quantified using the same analytical method to achieve full comparability. Second, the ratio of gram-positive to gram-negative bacteria should be determined for each soil by an independent method to generate a correct MurA conversion factor to bacterial necromass C. Third, the contribution of fungal GlcN to total GlcN is currently estimated assuming a 1:2 molar ratio of MurA:GlcN for bacteria, and assuming negligible contribution of GlcN from archaea and plants. Finally, the TalA conversion factor

should be considered as a proxy for methanogens, unless the ratio of methanogens to other soil archaea groups is known.

Even though necromass biomarker studies are increasingly used to trace the contribution of microbial necromass C to the SOM pool (Liang et al., 2019; Wang et al., 2021), certain limitations must be addressed:

(1) The use of necromass biomarkers does not reflect the entire contribution of microbes and/or plants to SOM (Whalen et al., 2022). The necromass biomarkers quantified in this paper only account for the presence of necromass residues and not for extracellular residues such as plant rhizodeposits or microbial EPS.

(2) The use of biomarkers as a proxy for microbial or plant residue C assumes that the specific biomarkers and their ratios behave similarly during death, decomposition, and recycling. The use of conversion factors derived from microbial residues creates direct conclusions from biomass biomarkers to necromass biomarkers. Post-mortem modification of compounds, as well as different degradation rates once these biomarkers reach the soil matrix could affect the ratios in which biomarkers are found. Camenzind et al. (2023) proposed that different microbial death pathways might determine the quantity and quality of microbial necromass in soils due to the chemical transformations microbial biomass suffer before and when becoming necromass. For example, predation by bacteria could lead to the uptake of cytoplasmic components leaving behind mainly cell wall fragments, predation by fauna would cause ingestion and assimilation only of digestible compounds and leave behind non-digested necromass, and fungal compartmental senescence would facilitate internal recycling of nutrients while leaving only modified cell walls behind (Camenzind et al., 2023). All these post-mortem modifications lead to fragmented cell walls residues in soils, and to a variable loss of non-cell wall mass such as cytoplasm. The loss of cytoplasm by predation and solute leaching can eventually increase the relative proportion of cell wall biomarkers. Also, growth stage and physiological state can alter the biochemical composition of microbes and thereby eventually affect the biomarker contents. Therefore, cultured biomass can only be regarded as a proximate yet close match for necromass biomarker composition. As such, while post-mortem modifications should be considered and the use of conversion factors based on “living” cultured strains should be done cautiously, conversion factors estimated from microbial residues are good estimators of microbial C contribution.

(3) Biomarker approaches assume that all necromass components are stabilized at the same rate in soils. Differential turnover dynamics of amino acids and amino sugars from bacterial necromass were recently shown by Zheng et al. (2023). Similarly, Craig et al. (2022) observed that necromass retention could not be predicted just by necromass production. Physical and chemical protection of cell wall fragments due to necromass-necromass interactions (Buckeridge et al., 2020), stabilization mechanisms due to the formation of aggregates (Angst et al., 2021), and the binding to minerals (Creamer et al., 2019) could not only alter the chemistry but also the stability of specific biomarkers. Different stabilization rates of biomarkers can also alter the ratios in which compounds are found in soil microbial necromass compared to the ratios found in microbial biomass.

These limitations need to be considered when using conversion factors derived from microbial cultures. Similarly, interspecies variation in the biomarker content within a taxonomic group should be addressed, which could be done via the inclusion of error propagation calculations. While there exist still substantial research gaps regarding the long-term transformation, recycling, turnover and stabilization of microbial necromass compounds in soils, the use of necromass biomarkers and the derived conversion factors do represent the golden standard to estimate necromass residue-C contribution to SOC.

5. Conclusions

We were able to identify and quantify amino sugars and neutral sugars in the pure biomass of a wide range of archaea, bacteria, fungi, and plant species. Our data showcased the variability in the amino sugar and neutral sugar content amongst the different taxonomic groups highlighting their potential as necromass residue biomarkers and allowing their extension from separating bacterial and fungal necromass to separating microbes from plants using multicomponent multivariate approaches. Moreover, we developed more accurate GlcN and MurA conversion factors, as well as new conversion factors for GalN and TalA that can be used to extrapolate amino sugar content into soil necromass C. Similarly, these necromass biomarkers and their derived conversion factors can help elucidate necromass contribution to the SOM pool.

Acknowledgements

The authors would like to thank Ludwig Seidl for his technical support with the UPLC-Orbitrap measurements. Likewise, we would like to thank Sabrina Pober, Sabine Maringer, Rahul Samrat and Xiaofei Liu for their assistance in the laboratory work.

References

Albers, S.-V., & Meyer, B.H. (2011). The archaeal cell envelope. *Nature Reviews Microbiology, 9*, 414–426. https://doi.org/10.1038/nrmicro2576

Angst, G., Mueller, K.E., Nierop, K.G.J., & 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. https://doi.org/10.1016/j.soilbio.2021.108189

Apostel, C., Dippold, M., & Kuzyakov, Y. (2015). Biochemistry of hexose and pentose transformations in soil analyzed by position-specific labeling and 13C-PLFA. *Soil Biology and Biochemistry, 80*, 199–208. https://doi.org/10.1016/j.soilbio.2014.09.005

Appuhn, A., & Joergensen, R. (2006). Microbial colonisation of roots as a function of plant species. *Soil Biology and Biochemistry, 38*, 1040–1051. https://doi.org/10.1016/j.soilbio.2005.09.002

Arroyo, J., Farkaš, V., Sanz, A.B., & Cabib, E. (2016). Strengthening the fungal cell wall through chitin-glucan cross-links: effects on morphogenesis and cell integrity. *Cellular Microbiology, 18*, 1239–1250. https://doi.org/10.1111/cmi.12615

Bardalaye, P.C., & Nordin, J.H. (1976). Galactosaminogalactan from Cell Walls of Aspergillus niger. *Journal of Bacteriology, 125*, 655–669. https://doi.org/10.1128/jb.125.2.655-669.1976

Bates, S.T., Berg-Lyons, D., Caporaso, J.G., Walters, W.A., Knight, R., & Fierer, N. (2011). Examining the global distribution of dominant archaeal populations in soil. *The ISME Journal, 5*, 908–917. https://doi.org/10.1038/ismej.2010.171

Beauvais, A., & Latgé, J.-P. (2018). Special Issue: Fungal Cell Wall. *Journal of Fungi, 4*, 91. https://doi.org/10.3390/jof4030091

Bhavsar, A.P., Erdman, L.K., Schertzer, J.W., & Brown, E.D. (2004). Teichoic acid is an essential polymer in Bacillus subtilis that is functionally distinct from teichuronic acid. *Journal of Bacteriology, 186*, 7865–7873. https://doi.org/10.1128/JB.186.23.7865-7873.2004

Brown, S., Santa Maria, J.P., & Walker, S. (2013). Wall teichoic acids of gram-positive bacteria. *Annual Review of Microbiology, 67*, 313–336. https://doi.org/10.1146/annurev-micro-092412-155620

Buckeridge, K.M., La Rosa, A.F., Mason, K.E., Whitaker, J., McNamara, N.P., Grant, H.K., & Ostle, N.J. (2020). Sticky dead microbes: Rapid abiotic retention of microbial necromass in soil. *Soil Biology and Biochemistry, 149*, 107929. https://doi.org/10.1016/j.soilbio.2020.107929

Caceres, M.D., & Legendre, P. (2009). Associations between species and groups of sites: indices and statistical inference. *Ecology, 90*, 3566–3574. https://doi.org/10.1890/08-1823.1

Camenzind, T., Mason-Jones, K., Mansour, I., Rillig, M.C., & Lehmann, J. (2023). Formation of necromass-derived soil organic carbon determined by microbial death pathways. *Nature Geoscience, 16*, 115–122. https://doi.org/10.1038/s41561-022-01100-3

Cao, Y., Ding, J., Li, J., Xin, Z., Ren, S., & Wang, T. (2023). Necromass-derived soil organic carbon and its drivers at the global scale. *Soil Biology and Biochemistry, 181*, 109025. https://doi.org/10.1016/j.soilbio.2023.109025

Caruso, C., Rizzo, C., Mangano, S., Poli, A., Di Donato, P., Finore, I., Nicolaus, B., Di Marco, G., Michaud, L., Lo & Giudice, A. (2018). Production and biotechnological potential of extracellular polymeric substances from sponge-associated Antarctic bacteria. *Applied and Environmental Microbiology, 84*, e01624-17. https://doi.org/10.1128/AEM.01624-17

Cheah, Y.T., & Chan, D.J.C. (2022). A methodological review on the characterization of microalgal biofilm and its extracellular polymeric substances. *Journal of Applied Microbiology, 132*, 3490–3514. https://doi.org/10.1111/jam.15455

Craig, M.E., Geyer, K.M., Beidler, K.V., Brzostek, E.R., Frey, S.D., Stuart Grandy, A., Liang, C., & Phillips, R.P. (2022). Fast-decaying plant litter enhances soil carbon in temperate forests but not through microbial physiological traits. *Nature Communications, 13*, 1229. https://doi.org/10.1038/s41467-022-28715-9

Creamer, C.A., Foster, A.L., Lawrence, C., McFarland, J., Schulz, M., & Waldrop, M.P. (2019). Mineralogy dictates the initial mechanism of microbial necromass association. *Geochimica et Cosmochimica Acta, 260*, 161–176. https://doi.org/10.1016/j.gca.2019.06.028

Dungait, J.A.J., Hopkins, D.W., Gregory, A.S., & Whitmore, A.P. (2012). Soil organic matter turnover is governed by accessibility not recalcitrance. *Global Change Biology, 18*, 1781–1796. https://doi.org/10.1111/j.1365-2486.2012.02665.x

Eichler, J. (2020). N-glycosylation in Archaea—New roles for an ancient posttranslational modification. *Molecular Microbiology, 114*, 735–741. https://doi.org/10.1111/mmi.14569

Engelking, B., Flessa, H., & Joergensen, R.G. (2007). Shifts in amino sugar and ergosterol contents after addition of sucrose and cellulose to soil. *Soil Biology and Biochemistry, 39*, 2111–2118. https://doi.org/10.1016/j.soilbio.2007.03.020

Erwig, L.P., & Gow, N.A.R. (2016). Interactions of fungal pathogens with phagocytes. *Nature Reviews Microbiology, 14*, 163–176. https://doi.org/10.1038/nrmicro.2015.21

Fanin, N., Kardol, P., Farrell, M., Nilsson, M.-C., Gundale, M.J., & Wardle, D.A. (2019). The ratio of Gram-positive to Gram-negative bacterial PLFA markers as an indicator of carbon availability in organic soils. *Soil Biology and Biochemistry, 128*, 111–114. https://doi.org/10.1016/j.soilbio.2018.10.010

Fazio, S.A., Uhlinger, D.J., Parker, J.H., & White, D.C. (1982). Estimations of uronic acids as quantitative measures of extracellular and cell wall polysaccharide polymers from environmental samples. *Applied and Environmental Microbiology, 43*, 1151–1159.

Fierer, N. (2017). Embracing the unknown: disentangling the complexities of the soil microbiome. *Nature Reviews Microbiology, 15*, 579–590. https://doi.org/10.1038/nrmicro.2017.87

Flemming, H.-C. & Wingender, J. (2010). The biofilm matrix. *Nature Reviews Microbiology, 8*, 623–633. https://doi.org/10.1038/nrmicro2415

Freymond, P.-P., Lazarevic, V., Soldo, B., & Karamata, D. (2006). Poly (glucosyl-N-acetylgalactosamine 1-phosphate), a wall teichoic acid of Bacillus subtilis 168: its biosynthetic pathway and mode of attachment to peptidoglycan. *Microbiology, 152*, 1709–1718. https://doi.org/10.1099/mic.0.28814-0

Garcia-Rubio, R., de Oliveira, H.C., Rivera, J., & Trevijano-Contador, N. (2020). The fungal cell wall: Candida, Cryptococcus, and Aspergillus species. *Frontiers in Microbiology, 10*, 2993. https://doi.org/10.3389/fmicb.2019.02993

Garna, H., Mabon, N., Wathelet, B., & Paquot, M. (2004). New method for a two-step hydrolysis and chromatographic analysis of pectin neutral sugar chains. *Journal of Agricultural and Food Chemistry, 52*, 4652–4659. https://doi.org/10.1021/jf049647j

Glaser, B., Turrión, M.-B., & Alef, K. (2004). Amino sugars and muramic acid—biomarkers for soil microbial community structure analysis. *Soil Biology and Biochemistry, 36*, 399–407. https://doi.org/10.1016/j.soilbio.2003.10.013

Gow, N.A.R., Latge, J.-P., & Munro, C.A. (2017). The fungal cell wall: Structure, biosynthesis, and function. *Microbiology Spectrum, 5*, FUNK-0035-2016. https://doi.org/10.1128/microbiolspec.FUNK-0035-2016

Gunina, A., & Kuzyakov, Y. (2015). Sugars in soil and sweets for microorganisms: Review of origin, content, composition and fate. *Soil Biology and Biochemistry, 90*, 87–100. https://doi.org/10.1016/j.soilbio.2015.07.021

Hopke, P.K. (2016). Review of receptor modeling methods for source apportionment. *Journal of the Air and Waste Management Association, 66*, 237–259. https://doi.org/10.1080/10962247.2016.1140693

Hu, J., Du, M., Chen, J., Tie, L., Zhou, S., Buckeridge, K.M., Cornelissen, J.H.C., Huang, C., & Kuzyakov, Y. (2023). Microbial necromass under global change and implications for soil organic matter. *Global Change Biology, 29*, 1–13. https://doi.org/10.1111/gcb.16676

Hu, Y., Zheng, Q., Zhang, S., Noll, L., & Wanek, W. (2018). Significant release and microbial utilization of amino sugars and D-amino acid enantiomers from microbial cell wall decomposition in soils. *Soil Biology and Biochemistry, 123*, 115–125. https://doi.org/10.1016/j.soilbio.2018.04.024

Indorf, C., Dyckmans, J., Khan, K.S., & Joergensen, R.G. (2011). Optimisation of amino sugar quantification by HPLC in soil and plant hydrolysates. *Biology and Fertility of Soils, 47*, 387–396. https://doi.org/10.1007/s00374-011-0545-5

Jenkinson, D.S. (1988). The determination of microbial biomass carbon and nitrogen in soils, in: Advances in Nitrogen Cycling in Agricultural Ecosystems. CAB International Press.

Joergensen, R.G. (2018). Amino sugars as specific indices for fungal and bacterial residues in soil. *Biology and Fertility of Soils, 54*, 559–568. https://doi.org/10.1007/s00374-018-1288-3

Joergensen, R.G., & Potthoff, M. (2005). Microbial reaction in activity, biomass, and community structure after long-term continuous mixing of a grassland soil. *Soil Biology and Biochemistry, 37*, 1249–1258. https://doi.org/10.1016/j.soilbio.2004.11.021

Kästner, M., Miltner, A., Thiele-Bruhn, S., & Liang, C. (2021). Microbial necromass in soils—Linking microbes to soil processes and carbon turnover. *Frontiers in Environmental Science, 9*, 756378. https://doi.org/10.3389/fenvs.2021.756378

Latgé, J.-P., & Beauvais, A. (2014). Functional duality of the cell wall. *Current Opinion in Microbiology, 20*, 111–117. https://doi.org/10.1016/j.mib.2014.05.009

Leyn, S.A., Gao, F., Yang, C., & Rodionov, D.A. (2012). N-acetylgalactosamine utilization pathway and regulon in Proteobacteria: Genomic reconstruction and experimental characterization in Shewanella. *Journal of Biological Chemistry, 287*, 28047–28056. https://doi.org/10.1074/jbc.M112.382333

Liang, C., Amelung, W., Lehmann, J., & Kästner, M. (2019). Quantitative assessment of microbial necromass contribution to soil organic matter. *Global Change Biology, 25*, 3578–3590. https://doi.org/10.1111/gcb.14781

Liu, D., Tang, W., Yin, J.-Y., Nie, S.-P., & Xie, M.-Y. (2021). Monosaccharide composition analysis of polysaccharides from natural sources: Hydrolysis condition and detection method development. *Food Hydrocolloids, 116*, 106641. https://doi.org/10.1016/j.foodhyd.2021.106641

Ma, S., He, F., Tian, D., Zou, D., Yan, Z., Yang, Y., Zhou, T., Huang, K., Shen, H., & Fang, J. (2018). Variations and determinants of carbon content in plants: a global synthesis. *Biogeosciences, 15*, 693–702. https://doi.org/10.5194/bg-15-693-2018

Megrian, D., Taib, N., Witwinowski, J., Beloin, C., & Gribaldo, S. (2020). One or two membranes? Diderm Firmicutes challenge the Gram-positive/Gram-negative divide. *Molecular Microbiology, 113*, 659–671. https://doi.org/10.1111/mmi.14469

Messner, P. (1997). Bacterial glycoproteins. *Glycoconjugate Journal, 14*, 3–11. https://doi.org/10.1023/A:1018551228663

Nguema-Ona, E., Vicre-Gibouin, M., Gotte, M., Plancot, B., Lerouge, P., Bardor, M., & Driouich, A. (2014). Cell wall O-glycoproteins and N-glycoproteins: aspects of biosynthesis and function. *Frontiers in Plant Science, 5*, 499. https://doi.org/10.3389/fpls.2014.00499

Oades, J.M. (1984). Soil organic matter and structural stability: mechanisms and implications for management. *Plant and Soil, 76*, 319–337. https://doi.org/10.1007/BF02205590

Oades, J.M., & Wagner, G.H. (1971). Biosynthesis of sugars in soils incubated with 14C glucose and 14C dextran. *Soil Science Society of America Journal, 35*, 914–917. https://doi.org/10.2136/sssaj1971.03615995003500060021x

Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Simpson, G.L., Solymos, P., Stevens, M.H.H., & Wagner, H. (2008). The vegan Package. *Community ecology package, 10*, 719.

Oliveri, P., Malegori, C., Mustorgi, E., & Casale, M. (2021). Qualitative pattern recognition in chemistry: Theoretical background and practical guidelines. *Microchemical Journal, 162*, 105725. https://doi.org/10.1016/j.microc.2020.105725

Parsons, J.W. (2021). Chemistry and distribution of amino sugars in soil and soil organisms, in: Soil Biochemistry. CRC Press.

Pinho, S.P., & Macedo, E.A. (2005). Solubility of NaCl, NaBr, and KCl in water, methanol, ethanol, and their mixed solvents. *Journal of Chemical and Engineering Data, 50*, 29–32. https://doi.org/10.1021/je049922y

R Core Team. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Raetz, C.R.H., Reynolds, C.M., Trent, M.S., & Bishop, R.E. (2007). Lipid A modification systems in gram-negative bacteria. *Annual Review of Biochemistry, 76*, 295–329. https://doi.org/10.1146/annurev.biochem.76.010307.145803

Redmile-Gordon, M.A., Brookes, P.C., Evershed, R.P., Goulding, K.W.T., & Hirsch, P.R. (2014). Measuring the soil-microbial interface: Extraction of extracellular polymeric substances (EPS) from soil biofilms. *Soil Biology and Biochemistry, 72*, 163–171. https://doi.org/10.1016/j.soilbio.2014.01.025

Roca, C., Alves, V.D., Freitas, F., & Reis, M.A.M. (2015). Exopolysaccharides enriched in rare sugars: bacterial sources, production, and applications. *Frontiers in Microbiology, 6*, 288. https://doi.org/10.3389/fmicb.2015.00288

Rohart, F., Gautier, B., Singh, A., & Lê Cao, K.-A. (2017). mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS Computational Biology, 13*, e1005752. https://doi.org/10.1371/journal.pcbi.1005752

Ruiz-Perez, D., Guan, H., Madhivanan, P., Mathee, K., & Narasimhan, G. (2020). So you think you can PLS-DA? *BMC Bioinformatics, 21*, 2. https://doi.org/10.1186/s12859-019-3310-7

Salas, E., Gorfer, M., Bandian, D., Wang, B., Kaiser, C., & Wanek, W. (2023). A rapid and sensitive assay to quantify amino sugars, neutral sugars and uronic acid necromass biomarkers using pre-column derivatization, ultra-high-performance liquid chromatography and high-resolution mass spectrometry. *Soil Biology and Biochemistry, 177*, 108927. https://doi.org/10.1016/j.soilbio.2022.108927

Scheller, H.V., & Ulvskov, P. (2010). Hemicelluloses. *Annual Review of Plant Biology, 61*, 263–289. https://doi.org/10.1146/annurev-arplant-042809-112315

Schimel, J.P., & Bennett, J. (2004). Nitrogen mineralization: Challenges of a changing paradigm. *Ecology, 85*, 591–602. https://doi.org/10.1890/03-8002

Starke, R., Siles, J.A., Fernandes, M.L.P., Schallert, K., Benndorf, D., Plaza, C., Jehmlich, N., Delgado-Baquerizo, M., & Bastida, F. (2021). The structure and function of soil archaea across biomes. *Journal of Proteome Research, 237*, 104147. https://doi.org/10.1016/j.jprot.2021.104147

Stock, B.C., Jackson, A.L., Ward, E.J., Parnell, A.C., Phillips, D.L., & Semmens, B.X. (2018). Analyzing mixing systems using a new generation of Bayesian tracer mixing models. *PeerJ, 6*, e5096. https://doi.org/10.7717/peerj.5096

Subedi, B.P., Martin, W.F., Carbone, V., Duin, E.C., Cronin, B., Sauter, J., Schofield, L.R., Sutherland-Smith, A.J., & Ronimus, R.S. (2021). Archaeal pseudomurein and bacterial murein cell wall biosynthesis share a common evolutionary ancestry. *FEMS Microbes, 2*, xtab012. https://doi.org/10.1093/femsmc/xtab012

Tra, V.N., & Dube, D.H. (2014). Glycans in pathogenic bacteria – potential for targeted covalent therapeutics and imaging agents. *Chemical Communications, 50*, 4659–4673. https://doi.org/10.1039/C4CC00660G

Vaningelgem, F., Zamfir, M., Mozzi, F., Adriany, T., Vancanneyt, M., Swings, J., & De Vuyst, L. (2004). Biodiversity of exopolysaccharides produced by Streptococcus thermophilus strains is reflected in their production and their molecular and functional characteristics. *Applied and Environmental Microbiology, 70*, 900–912. https://doi.org/10.1128/AEM.70.2.900-912.2004

Vollmer, W., Blanot, D., & De Pedro, M.A. (2008). Peptidoglycan structure and architecture. *FEMS Microbiology Reviews, 32*, 149–167. https://doi.org/10.1111/j.1574-6976.2007.00094.x

Wang, B., An, S., Liang, C., Liu, Y., & Kuzyakov, Y. (2021). Microbial necromass as the source of soil organic carbon in global ecosystems. *Soil Biology and Biochemistry, 162*, 108422. https://doi.org/10.1016/j.soilbio.2021.108422

Whalen, E.D., Grandy, A.S., Sokol, N.W., Keiluweit, M., Ernakovich, J., Smith, R.G., & Frey, S.D. (2022). Clarifying the evidence for microbial- and plant-derived soil organic matter, and the path toward a more quantitative understanding. *Global Change Biology, 28*, 1–19. https://doi.org/10.1111/gcb.16413

Xu, X., Thornton, P.E., & Post, W.M. (2013). A global analysis of soil microbial biomass carbon, nitrogen and phosphorus in terrestrial ecosystems. *Global Ecology and Biogeography, 22*, 737–749. https://doi.org/10.1111/geb.12029

Zheng, T., Miltner, A., Liang, C., Nowak, K.M., & Kästner, M. (2023). Turnover of bacterial biomass to soil organic matter via fungal biomass and its metabolic implications. *Soil Biology and Biochemistry, 180*, 108995. https://doi.org/10.1016/j.soilbio.2023.108995

Table 1. Average amino sugar contents (95% confidence intervals in brackets) and conversion factors (C.F.) derived from hydrolyzed biomass of archaea, bacteria, fungi, and plant species.

	GlcN		MurA		GalN		TalA		N
	μg mg^{-1} dw	C.F.	μg mg^{-1} dw	C.F.	μg mg^{-1} dw	C.F.	μg mg^{-1} dw	C.F.	
Archaea	15.8 a	29	ND		15.2 a	30	6.4*	72*	13
	(10.1-21.5)	(21-46)			(7.9-22.4)	(21-58)	(1.8-8.0)	(58-256)	
Gram-negative bacteria	11.3 a	41	3.9 a	118	1.1 bc	418	ND		18
	(4.1-18.5)	(25-112)	(1.5-6.2)	(74-307)	(0.3-1.9)	(242-1533)			
Gram-positive bacteria	37.3 b	12	24.7 b	19	3.8 bc	121	ND		29
	(28.3-46.3)	(10-16)	(18.4-30.9)	(15-25)	(1.3-6.3)	(73-354)			
Fungi	45.9 b	10	ND		4.0 b	115	ND		66
	(8.3-53.5)	(9-55)			(2.5-5.6)	(82-184)			
Plants	1.4 c	329	ND		1.2 c	383	ND		38
	(1.1-1.7)	(271-418)			(0.6-1.7)	(271-767)			

N: number of replicates. ND: not detectable. *Only detected in Euryarchaeota.
Different letters indicate significant differences ($P<0.01$) between taxonomy groups. Numbers in brackets indicate 95% confidence interval.

Table 2. Average neutral sugar contents (95% confidence intervals in brackets) in the hydrolyzed biomass of archaea, bacteria, fungi, and plant species.

	Glc µg mg-1 dw	Gal µg mg^{-1} dw	Man µg mg^{-1} dw	Fuc ng mg-1 dw	Rha ng mg^{-1} dw	Rib ng mg^{-1} dw	Xyl ng mg^{-1} dw	Ara ng mg^{-1} dw	GluA ng mg^{-1} dw	DRib ng mg^{-1} dw	GM/AX ratio	N
Archaea	3.1 a (0.4-5.8)	0.1 a (0.0-0.2)	1.0 a (0.5-1.5)	50.0 ab (0.0-147.7)	99.9 (28.1-171.6)	175.6 a (144.6-206.6)	5.4 ab (2.5-8.3)	120.4 a (76.0-164.9)	5.5 (0.0-16.1)	43.5 (41.1-45.9)	13.2 ab (6.4-19.9)	13
Gram-negative bacteria	2.1 a (1.0-3.2)	0.7 a (0.1-1.3)	0.3 b (0.0-0.6)	297.4 a (0.0-600.5)	1.9 (0.0-4.1)	136.6 bc (84.9-188.3)	5.3 b (1.1-9.5)	179.7 ab (107.4-252.0)	25.8 (0.0-55.6)	46.7 (32.6-60.8)	9.0 a (3.2-14.7)	18
Gram-positive bacteria	2.4 a (0.8-4.0)	0.3 a (0.1-0.6)	0.2 b (0.1-0.2)	73.3 b (0.0-177.4)	0.6 (0.0-1.5)	160.0 bc (100.0-220.0)	1.5 b (0.0-3.7)	232.8 b (183.6-281.7)	9.8 (0.0-24.4)	45.5 (41.2-49.8)	3.2 a (1.9-4.5)	29
Fungi	15.2 b (11.5-19.0)	2.2 b (1.3-3.1)	1.3 a (1.0-1.6)	317.4 ab (145.1-484.4)	56.8 (25.2-92.5)	109.6 b (91.6-123.6)	110.5 a (55.5-170.9)	104.2 a (84.5-120.9)	46.9 (23.6-69.4)	50.0 (46.0-54.0)	24.5 b (18.4-30.2)	66
Plants	15.7 b (13.2-18.2)	2.2 b (1.6-2.9)	1.1 a (0.9-1.3)	474.6 c (403.5-545.7)	57.4 (26.3-88.5)	88.0 c (73.1-102.9)	10.6 a (8.4-12.7)	132.9 a (117.8-147.9)	ND	36.1 (31.8-40.5)	22.6 b (18.8-26.4)	38

N: number of replicates. ND: not detectable. GM/AX: galactose + mannose / arabinose + xylose.
Different letters indicate significant differences ($P<0.05$) between taxonomy groups. Numbers in brackets indicate 95% confidence interval.

Table 3. Indicator species analysis of amino sugar and neutral sugar biomarkers found in the hydrolyzed biomass of archaea, bacteria, fungi, and plant species.

Biomarker	Indicator value	p value	Taxa associated
GlcN	0.77	<0.001	Archaea, Bacteria, Fungi
MurA	0.90	<0.001	Bacteria
GalN	0.39	<0.001	Archaea
TalA	0.72	<0.001	Archaea
Glc	0.66	<0.001	Fungi, Plants
Gal	0.63	<0.001	Fungi, Plants
Man	0.45	<0.001	Archaea, Fungi, Plants
Fuc	0.64	<0.001	Plants
Rib	0.33	0.003	Archaea
Ara	0.31	0.003	Archaea, Bacteria, Plants
Xyl	0.42	<0.001	Fungi
GluA	0.33	0.006	Fungi

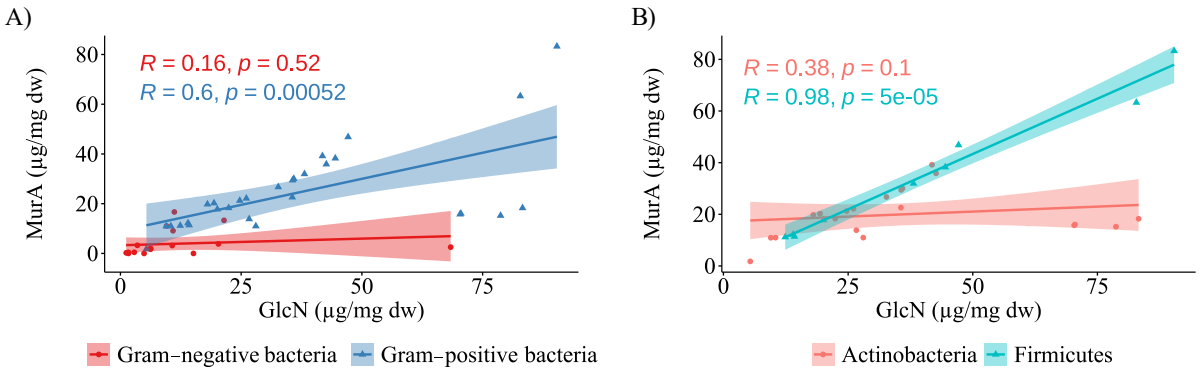

Figure 1. Correlation between glucosamine (GlcN) and muramic acid (MurA) contents in gram-negative and gram-positive bacteria (A), and in Actinobacteria and Firmicutes (B). Shaded background indicates 95% confidence intervals. Spearman correlations were used to test the correlation between GlcN and MurA in all bacterial groups.

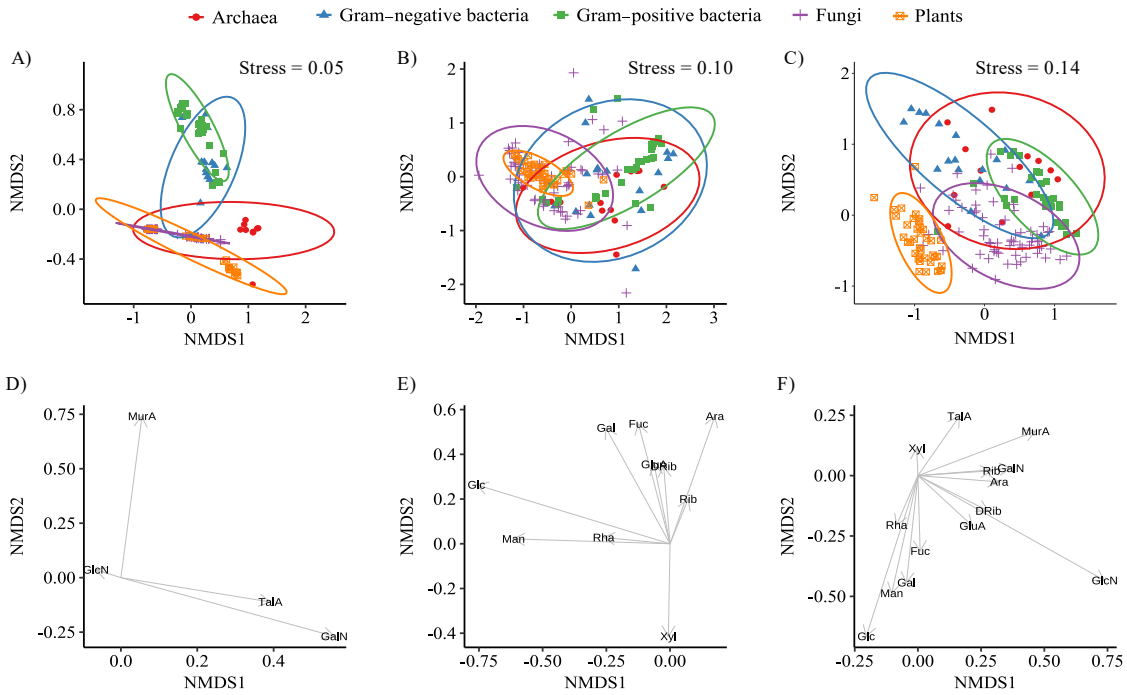

Figure 2. Non-metric multidimensional scaling (NMDS) analysis (A,B,C) and the sugars driving the separation (D,E,F) based only on amino sugar contents (A,D), only on neutral sugar contents (B,E) and on amino and neutral sugar biomarker contents (C,F) measured in the biomass of archaea, bacteria, fungi and plant species. Ellipses represent the 95% confidence interval.

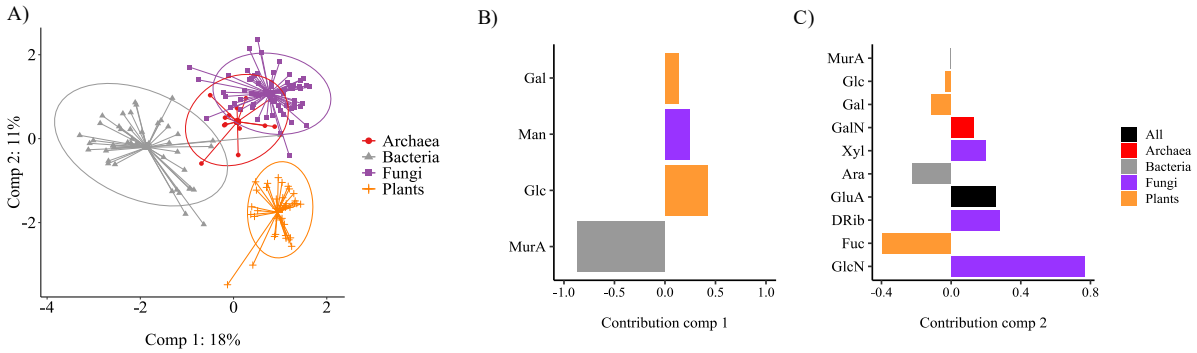

Figure 3. Partial least squares discriminant analysis (PLS-DA) score of amino sugar and neutral sugar contents of archaea, bacteria, fungi, and plants (A), and the contribution plots of the most discriminatory sugars for component 1 (B) and component 2 (C). Ellipses represent the 95% confidence interval. The colors indicate the taxonomic group in which the sugar compound has maximal importance based on the median. Y-axes in plots B and C represent the loading weight of each sugar in components 1 and 2.

Supplements

Table S1. Amino sugar and neutral sugar biomarker contents (± standard deviation) in the hydrolyzed biomass of archaea, bacteria, fungi, and plant species.

Taxon	Phylum* or Division	Species	GlcN (μg mg^{-1} dw)	MurA (μg mg^{-1} dw)	GalN (μg mg^{-1} dw)	TalA (μg mg^{-1} dw)	Glc (μg mg^{-1} dw)	Gal (μg mg^{-1} dw)	Man (μg mg^{-1} dw)	Rib (ng mg^{-1} dw)	Xyl (ng mg^{-1} dw)	Ara (ng mg^{-1} dw)	Fuc (μg mg^{-1} dw)	Rha (μg mg^{-1} dw)	GluA (ng mg^{-1} dw)	DRib (ng mg^{-1} dw)	N
Archaea	Crenarchaeota	*Saccharolobus solfataricus*	9.6	ND	ND	ND	17.4	ND	1.5	137.9	8.7	91.4	ND	ND	ND	43.1	1
	Crenarchaeota	*Sulfolobus acidocaldarius*	36.7	ND	ND	ND	8.1	ND	2.8	141.5	8.5	73.5	ND	0.2	ND	43.0	1
	Crenarchaeota	*Sulfolobus islandicus*	10.4	ND	ND	ND	6.2	ND	1.8	162.2	ND	65.8	ND	0.3	ND	45.7	1
	Euryarchaeota	*Haloferax volcanii*	ND	ND	ND	ND	1.7	ND	1.3	143.1	ND	55.0	ND	ND	ND	34.8	1
	Euryarchaeota	*Methanobacterium* sp.	20.3 (± 5.3)	ND	26.7 (±6.7)	11.8 (±4.9)	0.8 (±0.3)	0.3 (±0.1)	0.1 (±0.0)	230.4 (±68.0)	ND	236.5 (±24.3)	ND	ND	ND	45.5 (± 6.6)	4
	Euryarchaeota	*Methanothermobacter marburgensis*	16.8	ND	22.2	7.0	1.2	ND	1.4	148.3	11.0	65.8	ND	ND	ND	47.5	1
	Euryarchaeota	*Methanothermus fervidus*	28.7	ND	17.5	4.2	0.0	ND	1.9	125.1	9.8	60.2	ND	0.3	ND	42.6	1
	Euryarchaeota	*Methanocaldococcus jannaschii*	11.9	ND	6.2	2.8	2.5	ND	ND	132.8	11.7	64.4	ND	ND	ND	38.1	1
	Euryarchaeota	*Methanothermococcus okinawensis*	2.0	ND	3.9	ND	0.3	ND	ND	155.7	ND	67.4	ND	0.3	ND	38.0	1
	Euryarchaeota	*Methanotorris igneus*	5.6	ND	5.0	1.5	0.6	ND	1.7	129.3	9.9	65.3	ND	0.3	ND	41.2	1
	Euryarchaeota	*Methanospirillum stamsii*	ND	ND	ND	ND	5.4	0.0	1.5	199.1	16.1	78.5	0.9	ND	ND	44.1	1
	Euryarchaeota	*Methanopyrus kandleri*	ND	ND	ND	ND	0.2	0.3	ND	ND	ND	ND	ND	ND	ND	35.1	1
	Euryarchaeota	*Methanosarcina spelaei*	2.5	ND	35.4	1.4	0.6	ND	1.5	228.4	11.0	65.7	0.6	ND	70.9	44.7	1
Gram-negative bacteria	Acidobacteria	*Koribacter versatilis*	1.7	ND	ND	ND	0.0	ND	0.6	ND	ND	12.8	0.0	ND	ND	ND	1
	Acidobacteria	*Terriglobus* sp.	34.2	1.3	3.4	ND	5.6	6.6	2.2	336.8	34.5	443.5	2.1	ND	105.1	102.0	2
	Acidobacteria	*Solibacter usitatus*	3.5	3.3	2.0	ND	3.7	ND	0.4	42.5	ND	22.4	0.0	0.0	ND	ND	1
	Bacteroidetes	*Spirosoma spitsbergense*	9.6	11.0	1.0	ND	6.6	ND	0.2	ND	ND	49.9	0.0	0.0	ND	36.6	1
	Cyanobacteria	*Synechococcus* sp.	7.3 (±2.3)	4.5 (±1.4)	0.9 (±0.3)	ND	2.2 (±0.4)	0.9 (±0.2)	0.2 (±0.0)	259.6 (±82.1)	ND	276.8 (±109.1)	0.1 (±0.0)	ND	ND	44.0 (± 16.8)	4
	Proteobacteria	*Chelativorans multitrophicus*	11.2	16.7	0.6	ND	2.7	ND	ND	ND	17.7	20.0	ND	ND	ND	40.1	1

	Proteobacteria	*Neorhizobium* sp.	21.5	13.3	ND	ND	0.3	4.2	ND	253.6	20.4	222.3	2.4	ND	182.0	88.5	1
	Proteobacteria	*Labilithrix luteola*	10.9	9.1	0.6	ND	0.6	ND	0.2	9.9	ND	12.6	ND	0.0	ND	38.5	1
	Proteobacteria	*Klebsiella aerogenes*	4.9	ND	ND	ND	0.7	ND	ND	156.0	17.3	70.9	ND	ND	ND	38.1	1
	Proteobacteria	*Enterobacter* sp.	ND	67.5	53.4	ND	9.7	10.3	ND	1.3	0.1	0.9	ND	2.8	ND	487.0	1
	Proteobacteria	*Escherichia coli*	1.8 (± 0.6)	0.9 (±0.3)	0.2 (±0.1)	ND	0.2 (±0.1)	ND	ND	138.7 (±91.2)	ND	284.8 (±186.1)	ND	ND	ND	39.6 (± 8.1)	5
	Verrucomicrobia	*Chthoniobacter flavus*	20.3	3.8	3.8	ND	3.0	ND	0.3	63.8	ND	67.4	0.1	ND	ND	51.6	1
Gram-positive bacteria	Actinobacteria	*Arthrobacter globiformis*	28.0	11.0	ND	ND	8.5	ND	0.3	48.0	ND	39.7	ND	0.0	ND	37.2	1
	Actinobacteria	*Arthrobacter tecti*	18.0	19.8	7.0	ND	10.5	ND	0.3	49.5	ND	36.2	ND	ND	ND	39.9	1
	Actinobacteria	*Micrococcus luteus* (lyophylized cells)	37.7 (±4.3)	30.9 (±6.7)	ND	ND	2.7 (±0.6)	ND	0.3 (±0.0)	488.1 (±131.4)	ND	409.9 (±39.3)	ND	ND	ND	42.9 (± 4.4)	5
	Actinobacteria	*Micrococcus luteus* (peptidoglycan)	75.7 (±6.4)	16.32 (±1.4)	20.1 (±1.9)	ND	0.5 (±0.2)	0.3 (±0.1)	0.2 (±0.0)	283.5 (±55.6)	ND	366.8 (±77.3)	ND	ND	ND	59.1 (± 11.9)	4
	Actinobacteria	*Pseudarthrobacter oxydans*	26.7	13.8	0.7	ND	6.9	ND	0.3	41.9	ND	37.9	ND	0.0	ND	39.6	1
	Actinobacteria	*Rhodococcus* sp.	10.4	10.9	ND	ND	0.8	0.5	0.3	46.1	ND	63.2	0.0	ND	ND	43.8	1
	Actinobacteria	*Solirubrobacter soli*	5.4	1.8	1.3	ND	19.8	ND	0.3	ND	ND	86.8	ND	ND	ND	52.6	1
	Actinobacteria	*Streptomyces* sp.	27.2 (±5.8)	22.8 (±4.7)	1.7 (±0.5)	ND	0.2 (±0.1)	0.3 (±0.1)	0.2 (±0.1)	127.0 (±45.0)	ND	267.3 (±95.0)	ND	ND	ND	50.7 (± 19.8)	5
	Actinobacteria	*Streptosporangium roseum*	19.4	20.2	1.6	ND	0.4	ND	0.2	19.4	ND	27.1	ND	ND	ND	37.4	1
	Firmicutes	*Bacillus subtilis*	32.3 (±16.0)	27.2 (±14.0)	ND	ND	0.2 (±0.1)	0.2 (±0.0)	ND	110.3 (±15.6)	ND	249.7 (±40.8)	ND	ND	ND	43.5 (± 8.4)	4
	Firmicutes	*Paenibacillus alginolyticus*	90.4	83.3	ND	ND	0.6	ND	0.3	14.6	ND	98.1	ND	ND	ND	35.4	1
	Firmicutes	*Staphylococcus aureus*	23.4 (±16.2)	43.5 (±33.1)	0.8 (±0.6)	ND	0.2 (±0.1)	0.3 (±0.0)	ND	107.7 (±12.8)	ND	230.9 (±33.8)	ND	ND	ND	38.8 (± 5.6)	4
Fungi	Ascomycota	*Alternaria alternata*	58.9	ND	ND	ND	13.5	1.7	1.7	134.9	14.3	88.8	ND	ND	ND	51.9	1
	Ascomycota	*Aspergillus brasiliensis*	81.8	ND	7.3	ND	5.9	1.3	1.4	125.6	16.7	76.3	ND	ND	ND	46.9	1
	Ascomycota	*Aspergillus nidulans*	40.0	ND	26.7	ND	21.8	2.5	1.6	152.3	ND	133.9	ND	ND	ND	69.9	1
	Ascomycota	*Blastobotrys* sp.	50.2	ND	4.2	ND	37.7	2.4	2.6	114.0	15.2	130.6	ND	ND	ND	46.9	1
	Ascomycota	*Cadophora finlandica*	33.7	ND	4.3	ND	30.4	2.4	1.6	175.0	ND	115.5	ND	ND	ND	37.5	1
	Ascomycota	*Cephalotrichum microsporum*	93.9	ND	20.6	ND	15.3	0.8	1.7	119.2	12.0	82.8	0.3	0.1	ND	41.7	2

Ascomycota	*Clonostachys rosea*	76.6 (±6.2)	ND	6.1 (±0.4)	ND	4.7 (±0.8)	1.9 (±0.2)	0.1 (±0.0)	135.15 (±22.2)	ND	182.0 (±35.0)	0.2 (±0.0)	ND	83.8 (±26.7)	51.5 (± 5.2)	5
Ascomycota	*Coniochaeta hoffmannii*	30.1	ND	5.7	ND	11.2	1.7	1.8	119.0	10.4	64.0	ND	0.2	ND	50.2	1
Ascomycota	*Epicoccum nigrum*	36.9	ND	ND	ND	36.6	1.5	1.8	149.4	ND	150.9	ND	ND	ND	56.7	1
Ascomycota	*Exophiala* sp.	13.2	ND	ND	ND	20.1	2.3	2.2	137.2	ND	111.3	ND	ND	ND	61.6	1
Ascomycota	*Fusarium fujikurio*	60.3 (±5.9)	ND	4.6	ND	1.8 (±0.4)	0.7 (±0.1)	0.1 (±0.0)	68.4 (±5.1)	ND	92.0 (±7.9)	ND	ND	ND	49.4 (± 13.1)	4
Ascomycota	*Geomyces* sp.	35.3	ND	4.4	ND	30.7	ND	2.2	137.1	ND	128.8	ND	0.6	ND	46.9	1
Ascomycota	*Gibellulopsis nigrescens*	23.1	ND	4.2	ND	9.1	2.6	1.5	129.7	8.4	88.7	ND	0.3	ND	38.9	1
Ascomycota	*Gonytrichum macrocladum*	27.2	ND	6.9	ND	21.8	2.7	2.8	232.1	ND	209.2	ND	ND	ND	66.5	1
Ascomycota	*Heydenia* sp.	55.0	ND	4.9	ND	14.4	4.6	2.8	140.3	11.8	151.6	ND	ND	ND	48.1	1
Ascomycota	*Microdochium bolleyi*	54.9	ND	ND	ND	3.3	2.5	1.6	130.1	13.0	66.3	ND	ND	ND	38.1	1
Ascomycota	*Morchella elata*	14.6	ND	ND	ND	4.2	ND	0.3	15.0	652.6	ND	ND	ND	ND	42.4	1
Ascomycota	*Penicillium chrysogenum*	85.8	ND	ND	ND	5.1	1.5	1.5	144.1	ND	72.8	ND	ND	ND	44.6	1
Ascomycota	*Periconia macrospinosa*	9.6	ND	0.6	ND	8.5	ND	0.3	ND	441.8	ND	ND	ND	ND	40.7	1
Ascomycota	*Peziza ostracoderma*	82.8	ND	33.7	ND	24.5	ND	1.9	133.2	41.2	81.0	ND	0.2	ND	36.6	1
Ascomycota	*Phialocephala helvetica*	25.9	ND	8.2	ND	37.1	11.5	2.9	212.9	38.5	250.1	ND	0.5	ND	82.3	1
Ascomycota	*Pichia fermentans*	5.7	ND	ND	ND	27.5	ND	1.6	130.1	22.5	115.8	ND	ND	ND	41.1	1
Ascomycota	*Pyrenochaeta acicola*	100.0	ND	ND	ND	5.3	1.0	1.5	136.0	10.3	113.7	ND	ND	ND	45.2	1
Ascomycota	*Pyrenochaeta cava*	72.6	ND	4.2	ND	15.0	1.2	1.5	116.3	ND	93.0	0.6	ND	ND	52.6	1
Ascomycota	*Schizosaccharomyces pombe*	2.1	ND	ND	ND	16.3	3.7	1.7	121.2	14.0	81.0	ND	ND	ND	52.4	1
Ascomycota	*Schwanniomyces capriottii*	30.1	ND	4.2	ND	87.6	ND	7.9	120.3	90.9	161.8	ND	ND	ND	42.8	1
Ascomycota	*Tetracladium furcatum*	7.3	ND	ND	ND	2.8	ND	2.1	180.0	18.6	86.5	ND	0.4	ND	59.6	1
Ascomycota	*Tetracladium marchalianum*	53.9	ND	ND	ND	82.4	28.2	32.5	ND	225.4	1471.8	ND	5.5	ND	848.0	1
Ascomycota	*Tetracladium maxilliforme*	13.1	ND	ND	ND	41.5	16.8	1.9	159.0	ND	148.7	ND	0.3	ND	70.8	1
Ascomycota	*Trichocladium asperum*	61.0	ND	13.5	ND	15.3	2.8	1.6	153.2	12.5	125.3	ND	ND	ND	47.2	1

Ascomycota	*Trichoderma flavofuscum*	5.1	ND	ND	ND	18.4	ND	0.4	12.7	315.0	ND	ND	ND	ND	47.1	1
Ascomycota	*Tuber* sp.	33.4	ND	ND	ND	4.3	ND	0.3	14.3	1168.1	ND	ND	ND	ND	44.3	1
Ascomycota	*Warcupia* sp.	281.9	ND	6.1	ND	9.5	1.3	2.0	144.3	ND	178.9	ND	ND	ND	61.1	1
Ascomycota	Boliniaceae	30.3	ND	ND	ND	26.9	22.3	5.2	339.7	24.0	208.6	ND	ND	ND	121.9	1
Ascomycota	Chaetosphaeriaceae	26.7	ND	3.7	ND	17.0	1.7	1.6	114.8	ND	62.1	ND	0.2	ND	39.6	1
Ascomycota	Helotiales	44.0	ND	4.5	ND	35.9	4.0	3.0	138.4	ND	149.7	ND	0.5	ND	38.0	1
Basidiomycota	*Agaricus* sp.	11.7	ND	ND	ND	0.8	ND	0.2	14.0	672.1	ND	ND	ND	ND	45.7	1
Basidiomycota	*Amanita muscaria*	16.1	ND	ND	ND	8.9	ND	0.2	ND	ND	ND	ND	ND	ND	35.8	1
Basidiomycota	*Armillaria bulbosa*	21.9	ND	ND	ND	6.1	ND	0.3	12.4	485.6	ND	0.0	ND	ND	41.4	1
Basidiomycota	*Auricularia polytricha*	7.5	ND	0.7	ND	6.8	ND	0.9	19.2	333.0	ND	0.0	ND	271.4	45.9	1
Basidiomycota	*Craterellus cornucopioides*	19.0	ND	ND	ND	0.0	ND	0.3	ND	1041.3	ND	ND	ND	ND	45.8	1
Basidiomycota	*Grifola frondosa*	13.9	ND	ND	ND	5.2	ND	0.3	19.2	357.4	ND	ND	ND	ND	43.4	1
Basidiomycota	*Hericium* sp.	8.6	ND	ND	ND	11.4	ND	0.3	15.2	510.9	ND	0.0	ND	ND	46.8	1
Basidiomycota	*Lentinula edodes*	30.1	ND	ND	ND	3.4	ND	0.3	ND	ND	ND	ND	ND	ND	39.8	1
Basidiomycota	*Psilocybe cyanescens*	14.6	ND	ND	ND	10.0	ND	0.3	159.2	204.8	ND	ND	ND	ND	48.1	1
Basidiomycota	*Ramaria* sp.	50.9	ND	ND	ND	16.0	ND	2.8	122.3	11.4	153.2	0.6	ND	83.7	50.7	1
Basidiomycota	*Volvariella volvacea*	18.3	ND	ND	ND	11.3	ND	0.3	34.9	267.7	ND	ND	ND	215.5	35.2	1
Mucuromycota	*Mortierella alpina*	18.7	ND	5.0	ND	8.7	ND	1.4	148.1	9.6	89.4	0.9	ND	69.2	53.8	1
Mucuromycota	*Mortierella gamsii*	68.9	ND	ND	ND	43.1	5.5	1.6	166.0	83.3	182.5	2.4	ND	174.6	62.5	1
Mucuromycota	*Mucor hiemalis*	62.8	ND	4.4	ND	35.1	1.0	1.8	140.1	10.2	117.2	1.2	ND	234.7	45.7	1
Mucuromycota	*Rhizophagus irregularis*	53.3	ND	4.3	ND	3.1	5.6	1.4	133.1	17.2	133.7	2.1	0.3	79.9	44.8	1
Mucuromycota	*Rhizophagus irregularis* (hyphae and spores)	108.5 (±41.0)	ND	3.9 (±0.2)	ND	0.5 (±0.3)	2.0 (±1.1)	ND	119.0 (±11.0)	ND	233.7 (±82.8)	0.8 (±0.2)	ND	107.1 (±41.8)	45.2 (± 5.7)	3
Mucuromycota	*Rhizophagus irregularis* (hyphae)	85.7	ND	ND	ND	3.1	5.6	1.4	ND	45.6	243.2	3.9	ND	313.1	129.9	1
Mucuromycota	*Rhizopus oligosporus*	4.0	ND	0.6	ND	5.8	ND	ND	13.2	181.3	ND	ND	ND	ND	39.4	1
Mucuromycota	*Rhizopus oryzae*	55.6	ND	3.8	ND	8.4	0.9	1.1	79.2	7.8	69.0	0.7	ND	150.2	29.9	1
Mucuromycota	*Thamnidium elegans*	92.4	ND	6.6	ND	32.4	5.6	1.9	149.7	ND	217.4	2.0	ND	532.1	36.1	1

Plants	Mucuromycota	*Umbelopsis ramanniana*	37.7	ND	5.5	ND	27.4	1.3	1.8	140.7	16.3	117.2	1.4	ND	125.7	38.1	1
	Mucuromycota	*Umbelopsis vinacea*	31.0	ND	5.7	ND	43.3	3.3	1.9	149.1	12.3	150.1	1.1	ND	107.7	59.6	1
	Bryophyta	*Artichum undulatum*	1.0	ND	3.0	ND	20.9	3.0	1.4	95.0	16.2	128.0	0.5	0.2	ND	34.6	1
	Bryophyta	*Brachythecium rutabulum*	2.0	ND	3.9	ND	30.7	2.9	1.4	102.8	15.5	164.6	0.5	0.2	ND	36.3	1
	Bryophyta	*Hypnum cupressiforme*	1.4	ND	3.4	ND	30.2	6.0	1.4	116.7	12.4	181.4	0.5	0.2	ND	39.3	1
	Bryophyta	*Sphagnum magellanicum*	1.1	ND	3.6	ND	29.7	4.2	1.4	120.5	22.1	134.8	0.6	0.2	ND	39.8	1
	Bryophyta	*Tortella tortuosa*	1.9	ND	3.9	ND	30.2	7.6	2.1	173.0	ND	227.4	0.6	0.2	ND	41.9	1
	Marchantiophyta	*Bazzania trilobata*	1.8	ND	3.9	ND	9.6	5.2	1.2	101.7	15.0	94.1	0.5	ND	ND	32.1	1
	Marchantiophyta	*Marchantia polymorpha*	1.7	ND	3.4	ND	21.2	4.9	1.3	95.2	17.2	121.1	0.6	0.2	ND	ND	1
	Marchantiophyta	*Metzgeria conjugata*	1.5	ND	3.8	ND	20.9	8.7	1.4	154.6	13.0	286.1	0.6	0.2	ND	44.3	1
	Marchantiophyta	*Plagiochila asplenioides*	3.0	ND	4.0	ND	15.0	4.6	1.4	129.4	13.7	170.8	0.6	ND	ND	40.7	1
	Anthocerophyta	*Anthoceros* sp.	1.2	ND	3.1	ND	1.3	0.6	1.1	86.3	7.6	57.6	0.5	ND	ND	28.3	1
	Polypodiophyta	*Equisetum arvense*	ND	ND	ND	ND	24.4	1.4	1.5	104.1	8.3	131.6	0.5	ND	ND	35.8	1
	Conipherophyta	*Picea abies*	0.8	ND	ND	ND	20.1	1.2	1.1	87.1	12.6	114.0	0.5	ND	ND	28.3	1
	Gynkgophyta	*Ginko biloba*	1.4	ND	ND	ND	11.0	0.5	1.3	ND	8.6	93.1	0.6	ND	ND	37.1	1
	Monocot	*Allium ursinum*	2.9	ND	ND	ND	10.2	0.9	1.2	97.5	13.7	137.7	0.5	ND	ND	30.5	1
	Eudicot	*Alchemilla vulgaris*	ND	ND	ND	ND	8.5	1.0	1.1	91.5	9.2	51.6	0.5	ND	ND	30.3	1
	Eudicot	*Anthriscus cerefolium*	0.8 (±0.1)	ND	0.1 (±0.0)	ND	6.5 (±1.3)	1.8 (±0.4)	0.3 (±0.6)	81.0 (±19.1)	ND	98.7 (±18.6)	0.1 (±0.0)	ND	ND	35.9 (± 0.6)	4
	Eudicot	*Betula pendula*	ND	ND	ND	ND	15.3	2.7	1.4	118.7	13.7	201.1	0.6	0.2	ND	43.2	1
	Eudicot	*Capsella bursa-pastoris*	1.2	ND	ND	ND	17.1	1.8	ND	114.4	14.9	148.7	0.6	ND	ND	50.8	1
	Eudicot	*Dictamnus albus*	1.2	ND	ND	ND	19.2	1.2	1.4	ND	13.8	128.7	0.6	ND	ND	43.5	1
	Eudicot	*Fumaria officinalis*	2.1	ND	3.1	ND	12.6	0.9	1.2	100.4	8.8	136.2	0.5	ND	ND	67.8	1
	Eudicot	*Galium verum*	ND	ND	ND	ND	23.6	2.5	1.4	113.8	10.4	199.6	0.6	0.2	ND	45.3	1

Eudicot	*Malva silvestris*	2.6 (±0.3)	ND	0.1 (±0.1)	ND	14.0 (±4.1)	0.8 (±0.2)	0.1 (±0.0)	46.7 (±8.0)	ND	148.5 (±45.5)	0.1 (±0.0)	ND	ND	37.2 (± 5.7)	5
Eudicot	*Petroselinum crispum* (leaves)	ND	ND	ND	ND	10.4	0.8	1.1	94.8	9.8	96.3	0.5	ND	ND	37.8	1
Eudicot	*Petroselinum crispum* (roots)	ND	ND	ND	ND	17.0	1.6	1.5	ND	10.0	91.3	0.6	ND	ND	39.2	1
Eudicot	*Plantago major*	1.3	ND	ND	ND	16.8	1.5	1.5	128.1	26.2	139.2	0.7	ND	ND	39.5	1
Eudicot	*Polygonum aviculare*	1.6	ND	ND	ND	12.1	1.5	1.3	105.5	10.5	131.2	0.6	ND	ND	42.6	1
Eudicot	*Taraxacum officinale (leaves)*	2.5	ND	ND	ND	10.1	1.1	2.5	200.6	23.5	141.4	1.1	ND	ND	67.4	1
Eudicot	*Taraxacum officinale* (roots)	ND	ND	ND	ND	3.1	ND	ND	ND	14.0	82.7	0.6	ND	ND	ND	1
Eudicot	*Urtica dioica* (leaves)	1.2	ND	ND	ND	12.8	2.1	1.2	106.0	11.1	109.8	0.5	0.2	ND	29.9	1
Eudicot	*Urtica dioica* (roots)	1.9	ND	4.3	ND	23.7	1.5	1.5	ND	15.7	121.0	0.7	ND	ND	41.1	1
Eudicot	*Vaccinium myrtillus*	1.0	ND	ND	ND	13.7	1.2	1.2	96.5	11.1	102.2	0.5	ND	ND	31.1	1

N: number of replicates. ND: not detectable. * Phylum classification of fungi was based on the National Center for Biotechnology Information (NCBI) taxonomy.

Table S2. Average amino sugar contents in the hydrolyzed biomass of archaea, bacteria, fungi, and plant species.

		GlcN		MurA		GalN		TalA		
		μg mg^{-1} dw		μg mg^{-1} dw		μg mg^{-1} dw		μg mg^{-1} dw		N
		mean	95% C.I.	mean	95% C.I.	mean	95% C.I.	mean	95% C.I.	
Archaea	Crenarchaeota	18.9	1.5-36.3	ND		ND		ND		3
	Euryarchaeota	14.8	9.0-20.7	ND		19.7	12.5-27.0	6.4	2.9-10.0	10
Gram-negative bacteria	Acidobacteria	24.5	1.7-67.5	1.9	0.0-3.9	2.9	0.0-6.9	ND		3
	Bacteroidetes	9.5		11.0		1.0		ND		1
	Cyanobacteria	7.3	5.1-9.6	2.2	1.5-2.8	0.9	0.6-1.2	ND		4
	Proteobacteria	7.9	3.1-12.6	4.5	0.1-8.9	0.2	0.1-0.4	ND		9
	Verrucomicrobia	20.3		3.8		3.8		ND		1
Gram-positive bacteria	Actinobacteria	35.9	25.8-45.9	20.0	16.0-23.9	5.1	1.6-8.6	ND		20
	Firmicutes	40.4	21.2-59.7	35.1	18.4-51.8	0.9	0.0-2.0	ND		9
Fungi	Ascomycota	47.1	38.6-55.7	ND		5.2	2.9-7.4	ND		42
	Basidiomycota	19.3	12.1-26.5	ND		0.1	0.0-0.2	ND		11
	Mucuromycota	64.3	43.1-85.5	ND		3.7	2.5-4.8	ND		13
Plants	Bryophyta	1.5	1.1-1.9	ND		3.6	3.2-3.9	ND		5
	Marchantiophyta	2.0	1.0-3.0	ND		3.8	2.3-5.3	ND		4
	Anthocerotophyta	1.2		ND		3.1		ND		1
	Polypodiophyta	ND		ND		ND		ND		1
	Coniferophyta	0.8		ND		ND		ND		1
	Ginkgophyta	1.4		ND		ND		ND		1
	Monocot	2.9		ND		ND		ND		1
	Eudicot	1.3	0.9-1.7	ND		0.4	0.0-0.8	ND		24

N: number of replicates. ND: not detectable. C.I.: Confidence interval.

Table S3. Average neutral sugar contents (95% confidence intervals in brackets) in the hydrolyzed biomass of archaea, bacteria, fungi, and plant species.

		Glc μg mg^{-1} dw	Gal μg mg^{-1} dw	Man μg mg^{-1} dw	Fuc ng mg^{-1} dw	Rha ng mg^{-1} dw	Rib ng mg^{-1} dw	Xyl ng mg^{-1} dw	Ara ng mg^{-1} dw	GluA ng mg^{-1} dw	DRib ng mg^{-1} dw	N
Archaea	Crenarchaeota	10.6 (3.8-17.4)	ND	2.0 (1.3-2.8)	ND	163.2 (1.8-324-7)	147.2 (132.4-162.0)	5.7 (0.1-11.3)	76.9 (62.1-91.7)	ND	43.9 (42.2-45.7)	3
	Euryarchaeota	0.8 (0.4-1.3)	0.1 (0.0-0.2)	0.7 (0.2-1.2)	65.0 (0.0-192.0)	80.9 (0.2-161.6)	184.1 (145.1-223.1)	5.3 (1.8-8.9)	133.5 (77.9-189.2)	7.1 (0.0-21.0)	43.4 (40.3-46.5)	10
Gram-negative bacteria	Acidobacteria	1.6 (0.0-3.8)	1.2 (0.0-3.6)	0.4 (0.0-0.7)	553.9 (0.0-1579.0)	4.3 (0.0-12.7)	87.4 (0.0-219.3)	9.0 (0.0-26.8)	140.4 (0.0-381.0)	70.1 (0.0-207.4)	28.3 (0.0-83.7)	3
	Bacteroidetes	6.6	ND	0.2	33.1	7.5	ND	ND	49.9	ND	36.6	1
	Cyanobacteria	2.2 (1.7-2.6)	0.9 (0.7-1.1)	0.2 (0.1-0.2)	133.4 (103.6-163.2)	ND	259.6 (179.2-340.0)	ND	276.8 (169.9-383.7)	ND	44.0 (27.5-60.5)	4
	Proteobacteria	1.6 (0.1-3.6)	0.6 (0.0-1.4)	0.3 (0.0-0.9)	340.0 (0.0-868.2)	1.6 (0.0-4.8)	121.6 (59.7-183.6)	7.6 (1.6-13.6)	176.6 (68.7-284.5)	28.3 (0.0-69.1)	54.6 (33.4-75.8)	9
	Verrucomicrobia	3.0	ND	0.3	65.1	ND	63.8	ND	67.4	ND	51.6	1
Gram-positive bacteria	Actinobacteria	3.4 (0.2-5.6)	0.2 (0.1-0.3)	0.2 (0.2-0.3)	37.5 (0.0-110.1)	0.9 (0.0-2.1)	193.2 (110.7-275.7)	0.7 (0.0-2.1)	235.9 (166.1-305.7)	3.9 (0.0-11.6)	47.8 (42.1-53.4)	20
	Firmicutes	0.3 (0.1-0.4)	0.6 (0.0-1.3)	0.0 (0.0-0.1)	153.0 (0.0-453.0)	ND	86.3 (56.0-116.5)	3.3 (0.0-9.6)	225.4 (118.3-262.7)	22.8 (0.0-67.5)	40.5 (36.0-45.0)	9
Fungi	Ascomycota	17.0 (12.1-22.0)	2.6 (1.2-3.9)	1.6 (1.2-2.1)	48.3 (4.7-82.9)	82.9 (36.3-135.9)	127.7 (105.8-143.2)	70.8 (10.1-140.2)	111.9 (91.9-127.3)	10.0 (1.1-17.5)	51.1 (46.5-55.8)	42
	Basidiomycota	7.3 (4.4-10.1)	ND	0.6 (0.1-1.0)	65.3 (0.0-178.0)	ND	36.0 (4.5-67.6)	353.1 (165.0-541.3)	13.9 (0.0-41.2)	51.9 (0.0-110.2)	43.5 (40.6-46.4)	11
	Mucuromycota	16.1 (6.7-25.5)	2.7 (1.5-3.9)	1.0 (0.5-1.4)	1401.3 (858.4-1944.3)	20.6 (0.0-60.7)	113.6 (85.1-142.0)	33.2 (6.3-60.1)	155.4 (110.0-200.9)	162.2 (88.0-236.4)	52.0 (38.3-65.7)	13
Plants	Bryophyta	28.3 (24.7-32.0)	4.8 (3.0-6.5)	1.5 (1.2-1.8)	545.9 (474.8-616.9)	216.7 (197.0-236.4)	121.6 (94.8-148.3)	13.3 (6.1-20.5)	167.2 (132.1-202.4)	ND	38.4 (35.80-40.9)	5
	Marchantiophyta	16.7 (10.6-22.8)	5.9 (4.2-7.6)	1.3 (0.7-2.0)	572.7 (340.7-804.7)	108.5 (29.8-187.2)	120.2 (74.2-166.3)	14.7 (8.3-21.2)	168.1 (122.4-213.7)	ND	29.3 (14.9-43.6)	4
	Anthocerotophyta	1.3	0.6	1.1	470.5	ND	86.3	7.6	57.6	ND	28.3	1
	Polypodiophyta	24.4	1.4	1.5	520.0	ND	104.1	8.3	131.6	ND	35.8	1
	Coniferophyta	20.1	1.2	1.1	456.6	ND	87.1	12.6	114.0	ND	28.3	1
	Ginkgophyta	11.0	0.5	1.3	576.8	ND	ND	8.6	93.1	ND	37.1	1
	Monocot	10.2	0.9	1.2	507.2	ND	97.5	13.7	137.7	ND	30.5	1
	Eudicot	13.4 (11.1-15.7)	1.4 (1.1-1.6)	0.9 (0.6-1.2)	436.8 (328.0-545.6)	27.7 (0.0-57.8)	78.3 (58.8-97.8)	9.4 (6.4-12.3)	125.3 (109.7-140.9)	ND	37.6 (31.6-43.7)	24

N: number of replicates. ND: not detectable.

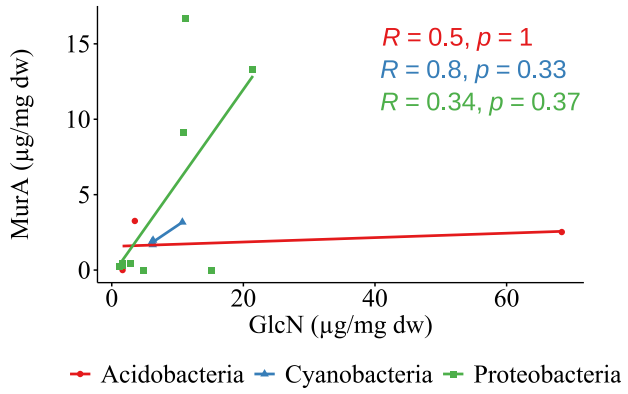

Figure S1. Correlation between glucosamine (GlcN) and muramic acid (MurA) contents in phyla of gram-negative bacteria. Only phyla with 3 or more replicates were included in the analysis. Spearman correlations was used to test the correlation between GlcN and MurA in all bacterial groups.

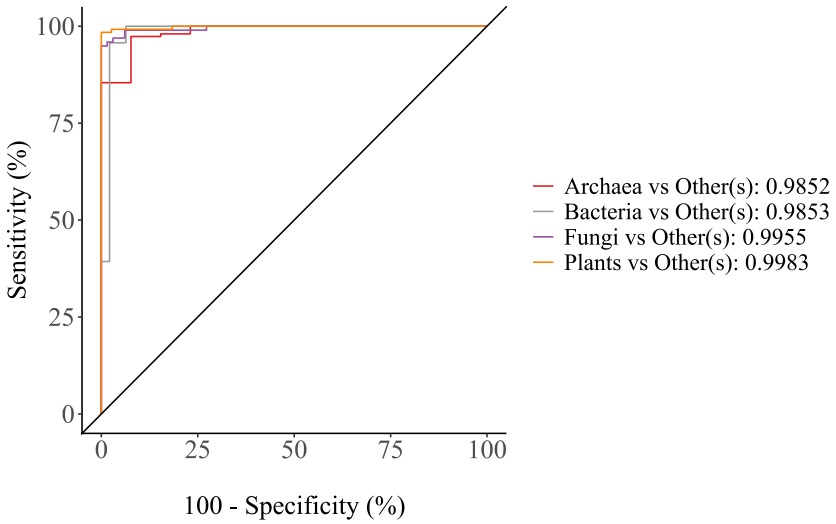

Figure S2. Receiver operating characteristic (ROC) curves using components 1,2 and 3 showcasing the performance evaluation of the PLS-DA model for each taxonomy group. Specificity indicates the rates of true negatives. Sensitivity indicates the rates of true positives. Numbers on the right indicate the accuracy of the classification performed by the PLS-DA model.

Chapter IV

Novel method for estimation of necromass deriving from soil archaea, bacteria, fungi, and plants by quantifying amino acid and amino sugar biomarkers in a single approach

Novel method for estimation of necromass deriving from soil archaea, bacteria, fungi and plants by quantifying amino acid and amino sugar biomarkers in a single approach

Erika Salas[a,b*], Markus Gorfer[c], Dragana Bandian[c], Stephanie A. Eichorst[d], Hannes Schmidt[a], Julia Horak, Simon K.-M. R. Rittmann[e], Christa Schleper[f], Barbara Reischl[e], Thomas Pribasnig[f], Jan Jansa[g], Christina Kaiser[a], and Wolfgang Wanek[a]

[a] Division of Terrestrial Ecosystem Research, Centre for Microbiology and Environmental Systems Science, University of Vienna, Vienna, Austria

[b] Doctoral School in Microbiology and Environmental Science, University of Vienna, Vienna, Austria

[c] AIT Austrian Institute of Technology GmbH, Bioresources, Tulln, Austria

[d] Department of Microbiology and Ecosystem Science, Centre for Microbiology and Environmental Systems Science, University of Vienna, Vienna, Austria

[e] Archaea Physiology & Biotechnology Group, Department of Functional and Evolutionary Ecology, University of Vienna, Vienna, Austria

[f] Archaea Biology and Ecogenomics Unit, Department of Functional and Evolutionary Ecology, University of Vienna, Vienna, Austria

[g] Institute of Microbiology, Czech Academy of Sciences, Prague, Czech Republic

Abstract

Soil organic matter is the largest carbon (C) pool in terrestrial ecosystems, and it is largely composed of microbial necromass. Microbes contribute to the long-term C storage in soils by incorporating C from plants into their biomass and, consequently, microbial necromass becomes stabilized mostly in mineral associated organic matter. So far, most studies have focused on tracing microbial necromass using amino sugar biomarkers, while plant contributions to soil organic matter are predominantly traced using lignin or long-chain alkanes. Glucosamine and muramic acid are amino sugars commonly used as biomarkers of fungal and bacterial necromass, respectively. Amino acids, such as D-enantiomers and non-proteinogenic amino acids have also been used though rarely as microbial and/or plant necromass tracers. For instance, meso (D,L)-diaminopimelic acid can be found in the peptidoglycan peptide chain of gram-negative bacteria, while hydroxyproline is commonly found in glycoproteins of plant cell walls. Currently, only very few studies have measured amino sugars alongside primary and secondary amino acids as biomarkers of plant and microbial necromass. In this study, we propose a new method that allows the simultaneous exploration of microbial and plant residue biomarkers using a single run via 6-aminoquinolyl-N-hydroxysuccinimidyl carbamate (AQC) derivatization, followed by ultra-high performance liquid chromatography (UHPLC) and Orbitrap high resolution mass spectrometry. For this, we analyzed 121 species of archaea, bacteria, fungi, and plants. We were able to quantify amino acids and amino sugar biomarkers in the biomass of all taxonomic groups, as well as compare how these biomarker contents varied between broad taxonomic groups. We confirmed the biomarker potential of non-proteinogenic amino acids and amino sugars using indicator species analysis as well as supervised multivariate approaches, such as random forest and partial least squares discriminant analysis (PLS-DA). Our results showed that hydroxyproline is a biomarker specific for plants, while L,L-diaminopimelic acid can be used alongside muramic acid as biomarkers specific for bacteria. Talosaminuronic acid represents a biomarker specific for archaea, while glucosamine was a biomarker indicative of archaea, bacteria, and fungi, being absent in plants. Our results showcase an unparalleled approach to trace both plant and microbial contributions to soil organic matter which will help improve our understanding of how different organic matter sources contribute to soil carbon formation and stabilization. This approach also allows

the quantitation of plant versus microbial contributions to the continuum from litter decomposition to soil organic matter formation through microbial processing, the contribution of plant, fungal and bacterial organic matter to mineral-associated organic matter (MaOM) versus particulate organic matter (POM), and to soil macro- and microaggregate formation.

Key words

Soil necromass, machine-learning, random forest, AQC chemistry, Orbitrap

1. Introduction

Microbes play a critical role in the conversion of plant organic matter and the buildup of soil organic matter (SOM), as well as in concurrent ecosystem C and nitrogen (N) cycling. The soil microbial carbon pump is an ecological framework that describes how microorganisms assimilate plant C into new microbial biomass and how, subsequently, part of their necromass and other microbial residues become stabilized in soil as part of SOM (Liang et al., 2017; Zhu et al., 2020). Necromass biomarkers, mainly amino sugars, are used to trace this microbial contribution to the soil organic C stocks (Joergensen, 2018; Wang et al., 2021).

Specific amino acids can be used as biomarkers for soil microorganisms. While L-amino acids are ubiquitous in nature, the less abundant D-amino acids are important components of bacterial peptidoglycan but are rare to absent in eukaryotes, and therefore have been used to trace bacterial residues in soils and sediments (Dworzanski et al., 2005; Hu et al., 2018; Radkov & Moe, 2014; Veuger et al., 2005). The bacterial peptidoglycan peptide chain is composed of four or five amino acids consisting of L-alanine (Ala), D-glutamine (Gln) as the first two, and D-alanine as the last one or two (Radkov & Moe, 2014). The third amino acid is usually a diamino acid, being lysine (Lys) in Gram-positive bacterial and *meso* (D,L)-diaminopimelic acid (mDAP) in Gram-negative bacteria (Vollmer et al., 2008). In some Gram-positive organisms a 5-glycine chain interconnects the peptide chains. Pseudopeptidoglycan, a structural analog of peptidoglycan found only in archaea, does not contain D-amino acid enantiomers and is composed of L-glutamate (Glu), L-Ala and Lys (Subedi et al., 2021).

Isotope tracing in mDAP and D-Ala has been used to estimate the decomposition of bacterial peptidoglycan (Hu et al., 2018). L,L-diaminopimelic acid (LLDAP) has been found to replace Lys in the peptidoglycan of some species of Actinobacteria (Gram-positive bacteria), such as *Actinomyces* and *Streptomyces* (Busse & Schumann, 1999; Leyh-Bouille et al., 1970), though this has not yet been explored in biomarker research. Hydroxyproline (Hyp) has been used to trace plant N in soils since its two major natural sources are Hyp-rich glycoproteins in plant cell walls and animal collagen (Nguema-Ona et al., 2014; Philben & Benner, 2013). Even though the origin and mineralization of amino acid biomarkers have been studied to some extent (Miltner et al., 2009; Vranova et al., 2012), our knowledge regarding amino acid residues from other soil microbial groups and their contribution to soil environments is highly fragmentary.

In contrast, amino sugars have been regularly used as biomarkers of soil microbial necromass to trace the contribution of microbial-derived C to SOM (Liang et al., 2019; Wang et al., 2021), however excluding plants which are largely devoid of amino sugars. Muramic acid (MurA), a major component of bacterial peptidoglycan, is used as a specific biomarker for bacterial residue C (Joergensen, 2018). Glucosamine (GlcN), the other amino sugar component of bacterial peptidoglycan and the major constituent of fungal chitin, is used as a biomarker of fungal C once the contribution of bacterial glucosamine has been corrected using the bacterial GlcN:MurA molar ratio (Appuhn & Joergensen, 2006; Joergensen, 2018). Galactosamine (GalN) and mannosamine (ManN) are less specific microbial necromass biomarkers since they can be found in bacteria, fungi and archaea (Joergensen, 2018; Salas et al., 2024). Similar to bacteria, some archaea contain a pseudopeptidoglycan cell wall composed of alternating N-acetylated GlcN and talosaminuronic acid (TalA); however, only recently has TalA been proposed as biomarker specific for Euryarchaeota in soils (Salas et al., 2024).

Multivariate supervised machine learning (ML) methods have been applied for classification, regression and biomarker detection in clinical metabolomics studies (Dias-Audibert et al., 2020; Hirakawa et al., 2022; Zhang et al., 2023) and in biomonitoring investigations (Cambon et al., 2023). ML algorithms have been widely used recently since they can handle nonlinear, large and heterogeneous datasets (Pichler & Hartig, 2023). Although some ML models present an overfitting risk, others such as partial least squares discriminant analysis (PLS-DA) and random forest

algorithms have low overfitting risk while maintaining high interpretability (Ghannam & Techtmann, 2021; Liebal et al., 2020). While supervised ML approaches have many advantages to analyze multivariate data, they have not yet been used to investigate the potential of amino acids together with amino sugars as soil residue/necromass biomarkers and to distinguish amongst microbial and/or plant contributions.

Soil necromass biomarker studies that quantify both microbial and plant contribution to SOM are of great value to better understand their dynamics and role in long-term soil C storage (Angst et al., 2021). So far, there have been neither approaches nor studies that have quantified both microbial and plant biomarkers using the same analytical approach or compound class. This derives from the problem that in current biogeochemical research amino sugars are applied as unique tracers of microbes while lignin or long-chain alkanes are used as specific tracers for plants (Angst et al., 2021; Whalen et al., 2022). Yet different compound classes such as lignins and proteins have different turnover characteristics and therefore cannot be directly compared (Gleixner et al., 1999). 6-aminoquinolyl-N-hydroxysuccinimidyl carbamate (AQC) derivatization provides numerous advantages for the analysis of amino compounds (Cohen & Michaud, 1993; Salazar et al., 2012); however, most studies focused on the analysis of amino acids only and very few have extended this to primary and secondary amines (Boughton et al., 2011) or to the concurrent measurement of amino acids and amino sugars (Díaz et al., 1996). In our study, we aimed to: (1) identify and quantify amino acid and amino sugar biomarkers in representative groups of soil archaea, bacteria, fungi, and plants using AQC derivatization, (2) determine the variability of amino acids and amino sugars between the different taxonomic groups, and (3) evaluate the biomarker potential of amino acids and amino sugars using multivariate machine learning approaches.

2. Methods

2.1 Chemicals and reagents

Amino acid and amino sugar standards [L-alanine (Ala), L-arginine (Arg), L-asparagine (Asn), L-aspartic acid (Asp), L-cysteine (Cys), L-glutamine (Gln), L-glutamic acid (Glu), glycine (Gly), L-histidine (His), L-hydroxyproline (Hyp), L-isoleucine (Ile), L-leucine (Leu), L-lysine (Lys), L-methionine (Met), *meso* (D,L)-diaminopimelic acid (mDAP), L,L- diaminopimelic acid (LLDAP), L-phenylalanine

(Phe), L-proline (Pro), L-serine (Ser), L-threonine (Thr), L-tryptophan (Trp), L-tyrosine (Tyr), L-valine (Val), glucosamine (GlcN), galactosamine (GalN), mannosamine (ManN), muramic acid (MurA)], formic acid solution (BioUltra 1 M in H_2O), hydrochloric acid, and acetonitrile (hypergrade for LC-MS) solutions were purchased from Sigma-Aldrich (St. Louis, MO). The AccQ•Tag Ultra derivatization kit (AccQ•Tag Ultra borate buffer, AccQ•Tag Ultra reagent powder, and AccQ•Tag Ultra reagent diluent) was purchased from Waters Corporation (Milford, MA, USA).

2.2 Sample preparation

With the aim to cover a large diversity of soil microbial groups, as well as plants, we analyzed the biomass from 121 species belonging to representative soil taxa of archaea, bacteria, fungi, and plants. Biomass samples from these taxonomic groups were hydrolyzed in a previous study (Salas et al., 2024). Briefly, 2 mg of pure freeze-dried cultured biomass or dried plant material were hydrolyzed by addition of 1.5 mL 6 M HCl and heating to 105 °C for 16 h. After cooling, the solution was centrifuged at 10,000 g for 20 min and dried under a nitrogen stream. To remove salts, 1 mL MS-grade methanol was added. The solution was centrifuged at 10,000 g for 10 min and dried under a nitrogen stream. The dried samples were dissolved in 1 mL Milli-Q water prior to derivatization using AQC.

AQC derivatization was carried out using the AccQ•Tag Ultra derivatization kit (Waters™, USA) following the protocol from the manufacturer. Briefly, 70 µL borate buffer were mixed with 10 µL sample aliquot or amino acid standard, followed by addition of 20 µL AccQ•Tag Ultra reagent (3 mg/mL AQC in acetonitrile). The mixture was briefly vortexed and kept at room temperature for 1 min. Samples were then incubated at 55 °C for 10 min, allowed to cool to room temperature and injected into the UHPLC-Orbitrap MS system.

2.3 UHPLC-Orbitrap MS system

Derivatized samples were injected into a UHPLC Ultimate 3000 system (Thermo Fisher Scientific, Bremen, Germany) coupled to an Orbitrap Q Exactive HRMS system (Thermo Fisher Scientific) with an electrospray ionization (H-ESI) source. Reversed phase chromatographic separations were performed on a Waters AccQ.Tag Ultra C18 column (2.1 mm x 100 mm, 1.7 µm particles) (Milford, MA, USA) with ACQUITY column in-line guard filter (2.1 mm, 0.2 µm) (Milford, MA, USA). The

column temperature was set to 55 °C. The mobile phase was composed of 0.001% formic acid in Milli-Q as eluent A and 0.001% formic acid in acetonitrile as eluent B. The following solvent gradient was used: 0-0.5 min constant at 0.1% B, 0.5-2.5 min increase to 5% B, 2.5-8 min increase to 20% B, 8-8.25 min increase to 90% B, 8.25-11 min constant at 90% B, 11-11.2 min decrease to 0.1% B, and 11.2-14.5 min constant at 0.1% B for column re-equilibration. The flow rate was set to 0.4 mL min^{-1} and the injection volume was 1 µL.

The Orbitrap Q Exactive HRMS system was operated using full-mass scan mode (m/z 150-1000) in the ESI positive mode. The automatic gain control (AGC) target value was set to $3x10^6$ and the resolution to 70,000. Lock mass correction was set to 214.0896 (contaminant from plasticizer). Spray voltage was 3.5 kV, capillary temperature 300 °C, sheath gas 35 arbitrary units and auxiliary gas 15 arbitrary units.

2.4 Data analysis

The mass to charge ratios (*m/z*) and retention times of all AQC-amino acid and amino sugar derivatives analyzed in this study can be found in Table S1. Structural isomers (e.g. leucine and isoleucine) were identified based on their distinct retention times. Data were extracted and compounds were identified and quantified by their mass spectrometric signal using the FreeStyle™ 1.7 software (Thermo Scientific) and Skyline software (Adams et al., 2020).

All statistical analyses were carried out using R version 4.2.3 (R Core Team, 2021). Amino acid and amino sugar data were checked for outliers and log-transformed when necessary to obtain normality and homoscedasticity. Kruskal-Wallis, Dunn, and Wilcoxon tests were performed to detect differences in amino acid and amino sugar content between taxonomic groups. To visualize the clustering of the taxonomic groups using amino acid and amino sugar contents we carried out principal component analyses (PCA). To test whether the taxonomic groups differed based on their amino acid and amino sugar biomass content in multivariate space, we used analyses of similarity (ANOSIM) with multiple permutations (n = 9999).

Indicator species analysis was used to identify amino acid and amino sugar biomarkers associated to specific taxonomic groups. Indicator species analysis uses the relative abundance and occurrence of species to estimate the strength of its association with a priori groups of interest and evaluates the probability of association based on randomization. We used log-transformed data and the multipatt function

from the indcspecies package (Caceres & Legendre, 2009) with multiple permutations (n = 9999).

We used two supervised multivariate approaches: random forest and partial least squares discriminant analysis (PLS-DA) to verify the use of amino compounds as biomarkers for different soil microbial groups and for plant necromass. The random forest algorithm trained on amino acid and amino sugar data was used to classify samples according to their taxonomic group (archaea, bacteria, fungi, or plants) using the randomForest package (Breiman, 2001). Data partitioning was set to 70% for model training and 30% for model testing. We set the number of trees to 500 and *mtry*, a parameter that corresponds to the number of features (in our study, number of amino compounds) randomly selected by the algorithm for each split of the decision trees, to 3 using five-fold cross-validation. The Gini index was used to determine the most important features during the decision tree classification process, where high values indicate more informative features (Cambon et al., 2023). A multidimensional scaling (MDS) plot was created using the proximity matrix calculated by the random forest. This proximity matrix is based on the frequency that two data points end up in the same leaf nodes.

To discriminate between taxonomic groups based on their amino acid and amino sugar content, we ran a PLS-DA using leave-one-out cross-validation in the mixOmics package (Rohart et al., 2017). PLS-DA loadings from components 1 and 2 were used to determine which amino acids and amino sugars were the most discriminatory between the different taxonomic groups. The area under the receiver operating characteristic (AUROC) curves showcased the performance evaluation of the PLS-DA model for each group.

3. Results

We analyzed amino acid and amino sugar contents in 12 archaeal species, 24 bacterial species, 57 fungal species, and 28 species of plants, using the same acid hydrolysis and AQC derivatization method via UHPLC-Orbitrap MS (Table S2).

3.1 Amino acid and amino sugar content in hydrolyzed biomass

We quantified 16 proteinogenic amino acids in the biomass of all taxonomic groups (Fig. S1). The highest contents of proteinogenic amino acids were found in both archaeal and bacterial samples, while the lowest contents were found in plants.

Gln, Asn, Trp and Cys were degraded by 6 M HCl hydrolysis, the former two deamidated (to Glu and Asp) and the latter two oxidized (to cystine and oxindolylalanine), and therefore were not analyzed here.

In our study, we focused on three non-proteinogenic amino acids with biomarker potential: mDAP, LLDAP and Hyp. Archaea had the highest mDAP content, followed by bacteria and fungi and plants the lowest (Fig. 1A). LLDAP was found at highest level in bacteria, but only in one archaeon and not in fungi and plants (Fig. 1B, Table S2). Hyp was consistently found in plants and at low levels in fungi and bacteria (Fig. 1C).

We were able to simultaneously quantify amino sugar compounds using AQC derivatization. We measured the MurA, TalA, and the sum of hexosamines (HexN = GlcN + GalN + ManN), the latter since the individual AQC-derivatives of HexN could not be chromatographically resolved. Fungal samples had the highest HexN content, followed by bacteria and archaea and lowest contents were measured in plants (Fig. 1D). MurA was only detected in bacterial samples (Fig. 1E), while TalA was detected only in archaeal samples (Fig. 1F).

When compared between bacterial groups, LLDAP, HexN, and MurA contents were higher in gram-positive than in gram-negative bacteria, while mDAP was higher in gram-negative than in gram-positive bacteria (Fig. 2).

3.2 Biomarker potential of amino acids and amino sugars

All proteinogenic amino acids were strongly correlated to each other (Fig. S2), with a PCA biplot of proteinogenic amino acids explaining 89.2% of the total variance (Fig. 3). Since the PC1 explained 83.8% of the variance in proteinogenic amino acids, we used these PC1 scores (termed PC1*) as a proxy of the content of proteinogenic amino acids to remove excessive co-variance from distorting all subsequent statistical analyses.

The full biomarker PCA based on non-proteinogenic amino acids, amino sugars and the PC1* proxy of proteinogenic amino acids showed a clear separation of prokaryotes (archaea + bacteria) from eukaryotes (fungi + plants) (Fig. 4). Their separation was explained by the inverse effects of PC1* proxy versus TalA, MurA and both DAP isomers on PC1. Along PC2 MurA and LLDAP contents drove the separation of bacteria, while mDAP and TalA were positively associated with archaea. Moreover,

the separation of plants and fungi was driven by PC3 (Fig. S3), following differences in Hyp and HexN content between non-plants and plants.

The indicator species analysis showed that TalA was a specific indicator of archaea while Hyp was specific to plants (Table 1). MurA and LLDAP were indicators for bacteria. HexN and mDAP were indicators of archaea, bacteria, and fungi. The PC1* proxy was indicative of eukaryotes (plants + fungi).

Our random forest model had an out-of-bag error of 8.3%, an accuracy of 92.3%, and a classification error of 25.0% for archaea, 8.6% for bacteria, 6.8% for fungi and 3.4% for plants. Random forest results also showed that HexN and MurA, followed by the PC1* proxy, are the most important biomarkers during the decision tree classification process, shown by their greatest negative impact on the Gini index (Fig. 5A). An MDS plot using the proximity matrix calculated from the random forest model showed a clear separation of all taxa (Fig. 5B).

The results of the PLS-DA model also showed the separation of bacteria, fungi, and plants, where components 1 and 2 explained 49% of the variance (Fig. 6A). PC1* proxy was a major contributor to the definition of plants, followed by MurA and LLDAP for bacteria (Fig. 6 B). HexN drove the separation of fungi, and TalA that of archaea (Fig. 6C). The receiver operating characteristic (ROC) curves of the performance of the PLS-DA model showed a classification error of 1.8% for archaea, 7.5% for bacteria, 2.9% for fungi and 1.6% for plants (Fig. 6D).

Finally, we assessed whether the relative composition of proteinogenic amino acids improves the separation of the taxonomic groups. We ran a PCA using the relative molar percentages of proteinogenic amino acids, in contrast to using their absolute contents (Fig. S4). The PC1 (PC1_ m) and PC2 (PC2_m) scores of the relative molar percentages of proteinogenic amino acids were included along with the PC1* proxy of proteinogenic amino acid contents and the non-proteinogenic amino acids and amino sugars into the multivariate analyses. However, the indicator species analysis, random forest model and PLS-DA showed very similar results to those mentioned above (Table S3, Fig. S5, Fig. S6).

4. Discussion

In this study we demonstrate the excellent separation of the major organismal groups contributing to SOM formation by their residues and necromass, based on

canonical biomarkers (amino sugars including HexN and MurA) and non-canonical or novel biomarkers (TalA, LLDAP, mDAP, and proteinogenic PC1*). We measured higher contents of proteinogenic amino acids in archaea and bacteria compared to fungi and plants (Fig. S1), which drove the separation of prokaryotes from eukaryotes (Fig. 3). Proteinogenic amino acids, especially L-enantiomers, are ubiquitous in nature and therefore, are commonly found in all taxa since they are the building blocks of peptides and key intermediates in biochemical processes. Therefore, proteinogenic amino acids have not yet been used as biomarkers.

Non-proteinogenic amino acids and amino sugars (e.g. GlcN as HexN member, and MurA) have been proposed and used as biomarkers for soil microorganisms (Appuhn & Joergensen, 2006; Dworzanski et al., 2005). Yet the presence and absence of potential non-proteinogenic amino acid biomarkers has not been systematically tested thus far. Indicator species analysis showed the association of specific amino compounds as biomarkers for specific taxonomic groups (Table 1). While mDAP is known as major constituent of the peptidoglycan peptide chain of gram-negative bacteria (Vollmer et al., 2008), we also found significant quantities in gram-positive bacteria, archaea and fungi (Fig. 1A, Fig 2A). Furthermore, our results showed that mDAP is not specific to bacteria but instead is indicative of archaea, bacteria, and fungi.

On the contrary, we found that LLDAP was an excellent biomarker specific for bacteria (Table 1, Fig. 1B), with gram-positive bacteria having a higher content compared to gram-negative bacteria (Fig. 2B). Leyh-Bouille et al. (1970) proposed that LLDAP could replace Lys in the peptidoglycan peptide chain of some Actinobacteria. A systematic survey later showed the prevalence of the LLDAP in Actinobacteria (Busse & Schumann, 1999). While both LLDAP and mDAP can be also be found as intermediates of the Lys biosynthesis pathways in prokaryotes and plants (Hudson et al., 2006), in our study most of the measured LLDAP and mDAP in bacteria likely derived from peptidoglycan peptides since we detected negligible contents of LLDAP in archaea and of mDAP in plants (Fig. 2).

We observed that Hyp was prevalent in plant samples and our indicator species analysis confirmed that Hyp can be used as a biomarker specific for plants (Fig. 1C, Table 1). Hyp is produced by hydroxylation of proline by the enzyme prolyl hydroxylase following protein synthesis, via post-translational modification. Hyp is a major component of the animal protein collagen, comprising ~13.5% of protein. In plants, the

major source of Hyp are the Hyp-rich glycoproteins (HRPGs) in their cell walls (Nguema-Ona et al., 2014). HRGPs comprise a superfamily of plant cell wall proteins that function in diverse aspects of plant growth and development. Given the expectedly small contribution of fauna (and collagen) to soil biota and to soil organic matter (Bar-On et al., 2018), Philben & Benner (2013) proposed Hyp as a tracer of plant N in soils. This demonstrates that Hyp is a suitable biomarker that can be used to trace plant cell wall residues in soil environments. The current study therefore circumvents one cardinal problem of current soil biomarker analysis: adding an amino compound (Hyp) as biomarker specific of plants to the known set of amino biomarkers for soil microbes. This relaxes problems of relying on chemical disparate classes of biomarkers (lignins or alkanes versus amino compounds) with very different turnover times and recycling kinetics.

Amino sugars, particularly GlcN and MurA, have been used as fungal and bacterial biomarkers, respectively, to trace microbial necromass C dynamics and contributions to soil organic matter (Joergensen, 2018; Liang et al., 2019; Wang et al., 2021). In our study, we found that HexN (lumping GlcN, GalN and ManN together) were indicative of archaea, bacteria, and fungi, while MurA was a biomarker specific for bacteria and TalA was specific for archaea (Table 1, Fig. 1 D-F). Though it be favorable to separate HexN into its constituents – which does not work with reversed phase chromatography of AQC-derivatized HexN components – it was shown that GlcN is dominant in HexN and on average constitutes 64%, followed by GalN (32%) and ManN (4%), with little differences across soils (Joergensen, 2018). The main components of archaeal pseudopeptidoglycan are alternating N-acetylated GlcN and TalA (Albers & Meyer, 2011), while bacterial peptidoglycan is composed of alternating N-acetylated GlcN and MurA (Subedi et al., 2021). In fungi, linked N-acetylated (or non-acetylated) GlcN is the major component of chitin (or chitosan) in the cell wall (Arroyo et al., 2016). Our results coincide with those from a previous study, in which a different derivatization protocol specific for neutral and amino sugars was used (Salas et al., 2024). This confirms that AQC derivatization can be successfully applied for the quantification of amino sugar biomarkers of archaea, bacteria, and fungi.

Our random forest model had a high accuracy percentage (92.3%), as well as low classification error for bacteria, fungi, and plants (<8.6%). The higher classification error for archaea (25.0%) was likely due to the small number of samples and the extreme diversity of archaeal cell wall structures almost rivaling that of all other

organisms together (Albers & Meyer, 2011). The Gini index results showed that HexN and MurA were the most important biomarkers for the classification of the different taxa, followed by the PC1* proxy of proteinogenic amino acids, LLDAP and Hyp (Fig. 5A). The MDS of the proximity matrix constructed by the random forest model clearly showed the separation of the different taxonomic groups (Fig. 5B), highlighting the importance of these amino compounds as biomarkers. Overall, our random forest model results verify that these amino compounds can be used as biomarkers and highlight their importance as tracers for the classification and for quantitative partitioning of soil organic matter into the different taxonomic groups.

Similarly, the PLS-DA results also showcased the biomarker value of these amino compounds to discriminate between phylogenetic broad groups, along with very low classification errors for all groups (<8%, Fig. 6D). The PLS-DA model showed the separation of all groups (Fig. 6), where the PC1* proxy of proteinogenic amino acids along with Hyp drove the separation of plants. HexN defined the separation of fungi, LLDAP and MurA caused the grouping of bacteria, and archaea were separated due to the presence of TalA. This further suggests that the amino compounds analyzed in this study can be used as biomarkers to discriminate between the main taxonomic groups contributing to soil organic matter in terms of their necromass and residues.

Both random forest and PLS-DA approaches demonstrated the importance of using a combined set of specific biomarkers for the classification of the different taxonomic groups. This multivariate biomarker approach differs from the use of single biomarkers specific to one group only. When compared to traditional statistics, ML algorithms provide higher flexibility and automatic data complexity optimization (Pichler & Hartig, 2023). Within ML approaches Menze et al. (2009) showed that random forest algorithms outperformed PLS-DA in variable selection but PLS-DA outperformed random forest algorithms in classification tasks. We also observed a similar trend in the classification error of both ML models; therefore, exploring a combination of different ML models can improve the robustness of the results in biomarker identification, particularly if more biomarker analyses are added to this dataset.

Multivariate ML models can also be applied in soil ecology to identify the contributions and dynamics of microbial and plant residues in soil environments. This requires not only the definition and separation of biomarker signals for all groups targeted, but also the application of a multivariate source mixing or attribution

modelling approach. Bayesian mixing models are used to estimate the contribution of several sources to a mixture, where tracer data (i.e., stable isotopes or biomarkers) are required to characterize both the sources and the mixture (Stock et al., 2018; Tackmann et al., 2018). This holds great potential for biomarker tracing analysis in complex mixtures such as soils in the future.

Furthermore, AQC chemistry provided an unparalleled approach to necromass biomarker analyses, where biomarkers for both microbes and plants can be measured and analyzed simultaneously. While previous biomarker studies have focused on amino sugars as tracers for soil microbes and lignin for plants (Angst et al., 2021; Whalen et al., 2022), the analysis of these different compounds required different analytical methods due to their chemical differences. By confirming Hyp as a biomarker specific for plants, our results showed that we were able to successfully measure Hyp alongside amino sugars and other amino acids using the same AQC derivatization protocol. To the best of our knowledge, this is the first time a full set of plant and microbial necromass biomarkers can be quantified using the same extraction and analytical methods.

5. Conclusions

In our study we describe a novel method to quantify both microbial and plant necromass biomarkers simultaneously using AQC derivatization via UHPLC and Orbitrap MS. We showed that amino acid and amino sugar contents varied in the hydrolyzed biomass of archaea, bacteria, fungi, and plants. We confirmed that non-proteinogenic amino acids, specifically LLDAP and Hyp, can be used as necromass biomarkers specific for bacteria and plants, respectively. The biomarker potential of both amino acids and amino sugars were evaluated using multivariate ML models. The amino biomarkers presented in this study can be used to trace both plant biomass and archaeal, bacterial, and fungal necromass in soils using a single analytical platform, and therefore can help improve our current knowledge on the plant and microbial contributions to soil organic matter.

Acknowledgements

The authors would like to thank Ludwig Seidl for his technical support with the UPLC-Orbitrap measurements. Likewise, we would like to thank Sabrina Pober,

Sabine Maringer, Rahul Samrat and Xiaofei Liu for their assistance in the laboratory work.

References

Adams, K. J., Pratt, B., Bose, N., Dubois, L. G., St. John-Williams, L., Perrott, K. M., Ky, K., Kapahi, P., Sharma, V., MacCoss, M. J., Moseley, M. A., Colton, C. A., MacLean, B. X., Schilling, B., Thompson, J. W., & Alzheimer's Disease Metabolomics Consortium. (2020). Skyline for small molecules: A unifying software package for quantitative metabolomics. *Journal of Proteome Research*, *19*(4), 1447–1458. https://doi.org/10.1021/acs.jproteome.9b00640

Albers, S.-V., & Meyer, B. H. (2011). The archaeal cell envelope. *Nature Reviews Microbiology*, *9*(6), 414–426. https://doi.org/10.1038/nrmicro2576

Angst, G., Mueller, K. E., Nierop, K. G. J., & 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. https://doi.org/10.1016/j.soilbio.2021.108189

Appuhn, A., & Joergensen, R. (2006). Microbial colonisation of roots as a function of plant species. *Soil Biology and Biochemistry*, *38*(5), 1040–1051. https://doi.org/10.1016/j.soilbio.2005.09.002

Arroyo, J., Farkaš, V., Sanz, A. B., & Cabib, E. (2016). Strengthening the fungal cell wall through chitin-glucan cross-links: Effects on morphogenesis and cell integrity. *Cellular Microbiology*, *18*(9), 1239–1250. https://doi.org/10.1111/cmi.12615

Bar-On, Y. M., Phillips, R., & Milo, R. (2018). The biomass distribution on Earth. *Proceedings of the National Academy of Sciences*, *115*(25), 6506–6511. https://doi.org/10.1073/pnas.1711842115

Boughton, B. A., Callahan, D. L., Silva, C., Bowne, J., Nahid, A., Rupasinghe, T., Tull, D. L., McConville, M. J., Bacic, A., & Roessner, U. (2011). Comprehensive profiling and quantitation of amine group containing metabolites. *Analytical Chemistry*, *83*(19), 7523–7530. https://doi.org/10.1021/ac201610x

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*, 5–32. https://doi.org/10.1023/A:1010933404324

Busse, H.-J., & Schumann, P. (1999). Polyamine profiles within genera of the class Actinobacteria with LL-diaminopimelic acid in the peptidoglycan. *International*

Journal of Systematic and Evolutionary Microbiology, *49*(1), 179–184. https://doi.org/10.1099/00207713-49-1-179

Caceres, M. D., & Legendre, P. (2009). Associations between species and groups of sites: Indices and statistical inference. *Ecology*, *90*(12), 3566–3574. https://doi.org/10.1890/08-1823.1

Cambon, M. C., Trillat, M., Lesur-Kupin, I., Burlett, R., Chancerel, E., Guichoux, E., Piouceau, L., Castagneyrol, B., Le Provost, G., Robin, S., Ritter, Y., Van Halder, I., Delzon, S., Bohan, D. A., & Vacher, C. (2023). Microbial biomarkers of tree water status for next-generation biomonitoring of forest ecosystems. *Molecular Ecology*, *32*(22), 5944–5958. https://doi.org/10.1111/mec.17149

Cohen, S. A., & Michaud, D. P. (1993). Synthesis of a fluorescent derivatizing reagent, 6-aminoquinolyl-N- hydroxysuccinimidyl carbamate, and its application for the analysis of hydrolysate amino acids via high-performance liquid chromatography. In *Analytical Biochemistry* (Vol. 211, Issue 2, pp. 279–287). https://doi.org/10.1006/abio.1993.1270

Dias-Audibert, F. L., Navarro, L. C., De Oliveira, D. N., Delafiori, J., Melo, C. F. O. R., Guerreiro, T. M., Rosa, F. T., Petenuci, D. L., Watanabe, M. A. E., Velloso, L. A., Rocha, A. R., & Catharino, R. R. (2020). Combining Machine Learning and Metabolomics to Identify Weight Gain Biomarkers. *Frontiers in Bioengineering and Biotechnology*, *8*, 6. https://doi.org/10.3389/fbioe.2020.00006

Díaz, J., Lliberia, J. Ll., Comellas, L., & Broto-Puig, F. (1996). Amino acid and amino sugar determination by derivatization with 6-aminoquinolyl-N-hydroxysuccinimidyl carbamate followed by high-performance liquid chromatography and fluorescence detection. *Journal of Chromatography A*, *719*(1), 171–179. https://doi.org/10.1016/0021-9673(95)00372-X

Dworzanski, J. P., Tripathi, A., Snyder, A. P., Maswdeh, W. M., & Wick, C. H. (2005). Novel biomarkers for Gram-type differentiation of bacteria by pyrolysis–gas chromatography–mass spectrometry. *Journal of Analytical and Applied Pyrolysis*, *73*(1), 29–38. https://doi.org/10.1016/j.jaap.2004.09.003

Ghannam, R. B., & Techtmann, S. M. (2021). Machine learning applications in microbial ecology, human microbiome studies, and environmental monitoring. *Computational and Structural Biotechnology Journal*, *19*, 1092–1107. https://doi.org/10.1016/j.csbj.2021.01.028

Gleixner, G., Bol, R., & Balesdent, J. (1999). Molecular insight into soil carbon turnover. *Rapid Communications in Mass Spectrometry*, *13*(13), 1278–1283. https://doi.org/10.1002/(SICI)1097-0231(19990715)13:13<1278::AID-RCM649>3.0.CO;2-N

Hirakawa, Y., Yoshioka, K., Kojima, K., Yamashita, Y., Shibahara, T., Wada, T., Nangaku, M., & Inagi, R. (2022). Potential progression biomarkers of diabetic kidney disease determined using comprehensive machine learning analysis of non-targeted metabolomics. *Scientific Reports*, *12*(1), 16287. https://doi.org/10.1038/s41598-022-20638-1

Hu, Y., Zheng, Q., Zhang, S., Noll, L., & Wanek, W. (2018). Significant release and microbial utilization of amino sugars and D-amino acid enantiomers from microbial cell wall decomposition in soils. *Soil Biology and Biochemistry*, *123*, 115–125. https://doi.org/10.1016/j.soilbio.2018.04.024

Hudson, A. O., Singh, B. K., Leustek, T., & Gilvarg, C. (2006). An LL-diaminopimelate aminotransferase defines a novel variant of the lysine biosynthesis pathway in plants. *Plant Physiology*, *140*(1), 292–301. https://doi.org/10.1104/pp.105.072629

Joergensen, R. G. (2018). Amino sugars as specific indices for fungal and bacterial residues in soil. *Biology and Fertility of Soils*, *54*(5), 559–568. https://doi.org/10.1007/s00374-018-1288-3

Leyh-Bouille, M., Bonaly, R., Ghuysen, J. M., Tinelli, R., & Tipper, D. (1970). LL-Diaminopimelic Acid Containing Peptidoglycans in Walls of Streptomyces sp. And of Clostridium perfringens (Type A). *Biochemistry*, *9*(15), 2944–2952. https://doi.org/10.1021/bi00817a002

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. https://doi.org/10.1111/gcb.14781

Liang, C., Schimel, J. P., & Jastrow, J. D. (2017). The importance of anabolism in microbial control over soil carbon storage. *Nature Microbiology*, *2*(8), 1–6. https://doi.org/10.1038/nmicrobiol.2017.105

Liebal, U. W., Phan, A. N. T., Sudhakar, M., Raman, K., & Blank, L. M. (2020). Machine learning applications for mass spectrometry-based metabolomics. *Metabolites*, *10*(6), 243. https://doi.org/10.3390/metabo10060243

Menze, B. H., Kelm, B. M., Masuch, R., Himmelreich, U., Bachert, P., Petrich, W., & Hamprecht, F. A. (2009). A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC Bioinformatics*, *10*(1), 213. https://doi.org/10.1186/1471-2105-10-213

Miltner, A., Kindler, R., Knicker, H., Richnow, H. H., & Kästner, M. (2009). Fate of microbial biomass-derived amino acids in soil and their contribution to soil organic matter. *Organic Geochemistry*, *40*(9), 978–985. https://doi.org/10.1016/j.orggeochem.2009.06.008

Nguema-Ona, E., Vicre-Gibouin, M., Gotte, M., Plancot, B., Lerouge, P., Bardor, M., & Driouich, A. (2014). Cell wall O-glycoproteins and N-glycoproteins: Aspects of biosynthesis and function. *Frontiers in Plant Science*, *5*, 499. https://doi.org/10.3389/fpls.2014.00499

Philben, M., & Benner, R. (2013). Reactivity of hydroxyproline-rich glycoproteins and their potential as biochemical tracers of plant-derived nitrogen. *Organic Geochemistry*, *57*, 11–22. https://doi.org/10.1016/j.orggeochem.2013.01.003

Pichler, M., & Hartig, F. (2023). Machine learning and deep learning—A review for ecologists. *Methods in Ecology and Evolution*, *14*(4), 994–1016. https://doi.org/10.1111/2041-210X.14061

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Radkov, A. D., & Moe, L. A. (2014). Bacterial synthesis of d-amino acids. *Applied Microbiology and Biotechnology*, *98*(12), 5363–5374. https://doi.org/10.1007/s00253-014-5726-3

Rohart, F., Gautier, B., Singh, A., & Lê Cao, K.-A. (2017). mixOmics: An R package for 'omics feature selection and multiple data integration. *PLOS Computational Biology*, *13*(11), e1005752. https://doi.org/10.1371/journal.pcbi.1005752

Salas, E., Gorfer, M., Bandian, D., Eichorst, S. A., Schmidt, H., Horak, J., Rittmann, S. K.-M. R., Schleper, C., Reischl, B., Pribasnig, T., Jansa, J., Kaiser, C., & Wanek, W. (2024). Reevaluation and novel insights into amino sugar and neutral sugar necromass biomarkers in archaea, bacteria, fungi, and plants.

Science of The Total Environment, *906*, 167463. https://doi.org/10.1016/j.scitotenv.2023.167463

Salazar, C., Armenta, J. M., & Shulaev, V. (2012). An UPLC-ESI-MS/MS assay using 6-aminoquinolyl-N-hydroxysuccinimidyl carbamate derivatization for targeted amino acid analysis: Application to screening of arabidopsis thaliana mutants. *Metabolites*, *2*(3), 398–428. https://doi.org/10.3390/metabo2030398

Stock, B. C., Jackson, A. L., Ward, E. J., Parnell, A. C., Phillips, D. L., & Semmens, B. X. (2018). Analyzing mixing systems using a new generation of Bayesian tracer mixing models. *PeerJ*, *6*, e5096. https://doi.org/10.7717/peerj.5096

Subedi, B. P., Martin, W. F., Carbone, V., Duin, E. C., Cronin, B., Sauter, J., Schofield, L. R., Sutherland-Smith, A. J., & Ronimus, R. S. (2021). Archaeal pseudomurein and bacterial murein cell wall biosynthesis share a common evolutionary ancestry. *FEMS Microbes*, *2*, 1–13. https://doi.org/10.1093/femsmc/xtab012

Tackmann, J., Arora, N., Schmidt, T. S. B., Rodrigues, J. F. M., & Von Mering, C. (2018). Ecologically informed microbial biomarkers and accurate classification of mixed and unmixed samples in an extensive cross-study of human body sites. *Microbiome*, *6*(1), 192. https://doi.org/10.1186/s40168-018-0565-6

Veuger, B., Middelburg, J. J., Boschker, H. T. S., & Houtekamer, M. (2005). Analysis of 15 N incorporation into D-alanine: A new method for tracing nitrogen uptake by bacteria. *Limnology and Oceanography: Methods*, *3*(5), 230–240. https://doi.org/10.4319/lom.2005.3.230

Vollmer, W., Blanot, D., & De Pedro, M. A. (2008). Peptidoglycan structure and architecture. *FEMS Microbiology Reviews*, *32*(2), 149–167. https://doi.org/10.1111/j.1574-6976.2007.00094.x

Vranova, V., Zahradnickova, H., Janous, D., Skene, K. R., Matharu, A. S., Rejsek, K., & Formanek, P. (2012). The significance of D-amino acids in soil, fate and utilization by microbes and plants: Review and identification of knowledge gaps. *Plant and Soil*, *354*, 21–39. https://doi.org/10.1007/s11104-011-1059-5

Wang, B., An, S., Liang, C., Liu, Y., & Kuzyakov, Y. (2021). Microbial necromass as the source of soil organic carbon in global ecosystems. *Soil Biology and Biochemistry*, *162*, 108422. https://doi.org/10.1016/j.soilbio.2021.108422

Whalen, E. D., Grandy, A. S., Sokol, N. W., Keiluweit, M., Ernakovich, J., Smith, R. G., & Frey, S. D. (2022). Clarifying the evidence for microbial- and plant-

derived soil organic matter, and the path toward a more quantitative understanding. *Global Change Biology*, *00*, 1–19. https://doi.org/10.1111/gcb.16413

Zhang, J. D., Xue, C., Kolachalama, V. B., & Donald, W. A. (2023). Interpretable machine learning on metabolomics data reveals biomarkers for Parkinson's disease. *ACS Central Science*, *9*(5), 1035–1045. https://doi.org/10.1021/acscentsci.2c01468

Zhu, X., Jackson, R. D., DeLucia, E. H., Tiedje, J. M., & Liang, C. (2020). The soil microbial carbon pump: From conceptual insights to empirical assessments. *Global Change Biology*, *26*(11), 6032–6039. https://doi.org/10.1111/gcb.15319

Table 1. Indicator species analysis of amino acid and amino sugar contents measured in the hydrolyzed biomass of archaea, bacteria, fungi, and plants.

Taxa	Amino acid	Indicator value
Archaea	TalA	0.771
Bacteria	MurA	0.808
	LLDAP	0.637
Plants	Hyp	0.561
Fungi + Plants	PC1*	0.642
Archaea + Bacteria + Fungi	HexN	0.708
	mDAP	0.385

PC1* refers to the PC1 scores of the proteinogenic amino acids content.

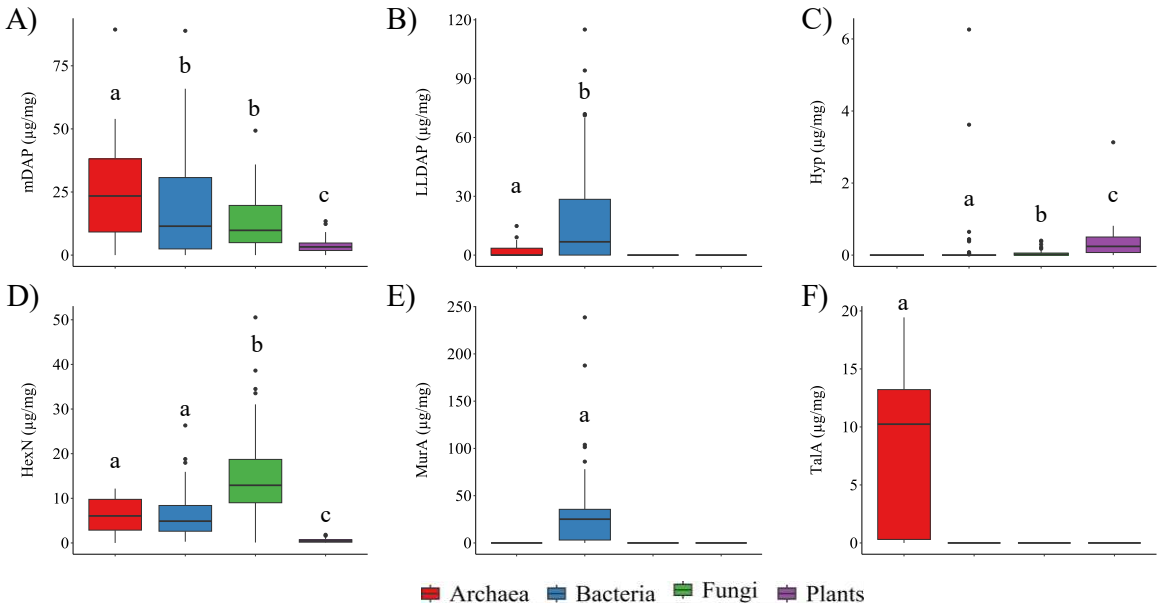

Figure 1. Non-proteinogenic amino acid (A, B, C) and amino sugar (D, E, F) contents measured in the hydrolyzed biomass of archaea, bacteria, fungi, and plants using AQC derivatization via UHPLC-Orbitrap MS. Different letters indicate significant differences ($P < 0.05$) between taxonomic groups.

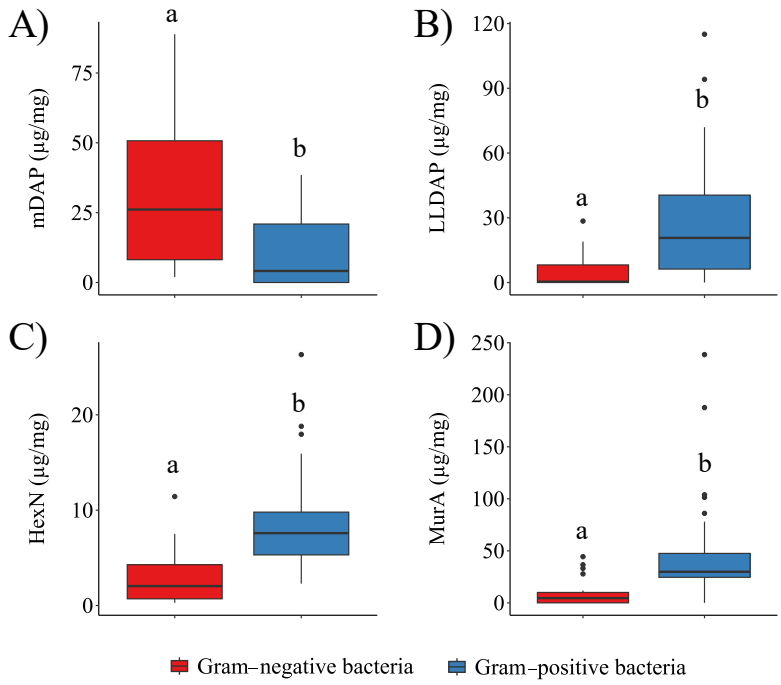

Figure 2. Non-proteinogenic amino acid (A & B) and amino sugar (C & D) contents in the hydrolyzed biomass of gram-negative and gram-positive bacteria. Different letters indicate significant differences ($P < 0.05$) between bacterial groups.

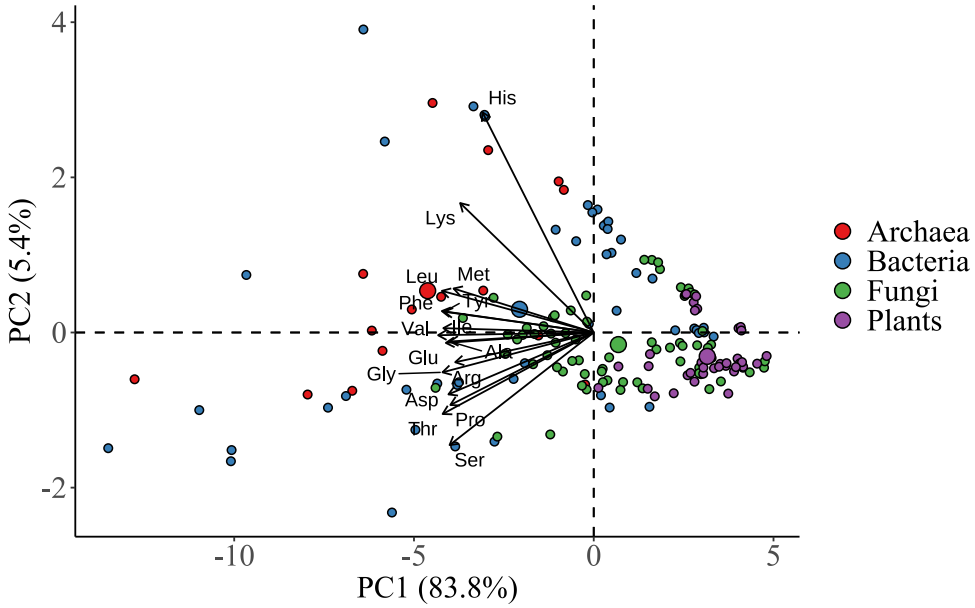

Figure 3. Principal component analysis (PCA) biplot of proteinogenic amino acid contents in the hydrolyzed biomass of archaea, bacteria, fungi, and plants. Large circles represent the data centroids.

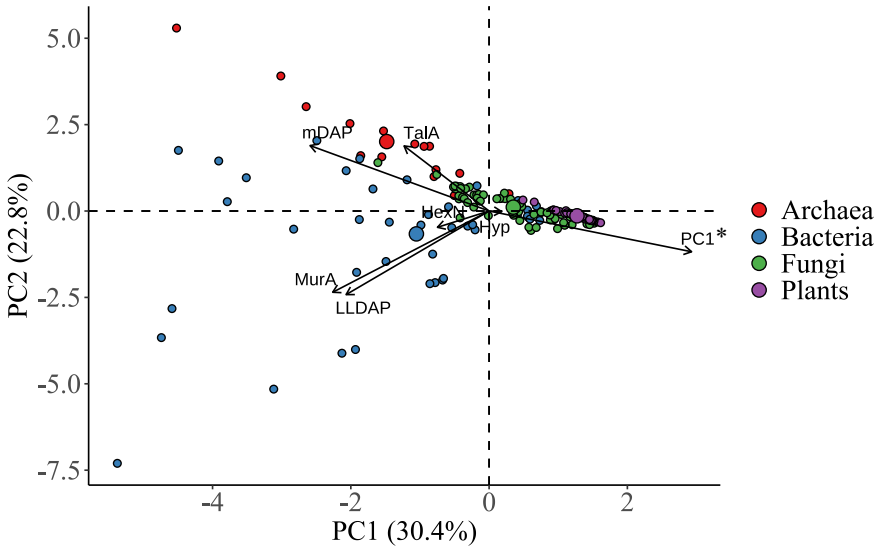

Figure 4. Principal component analysis (PCA) biplot using non-proteinogenic amino acid contents and they PC1 proxy of proteinogenic amino acid contents (PC1*) in samples of archaea, bacteria, fungi, and plants. Large circles represent the data centroids.

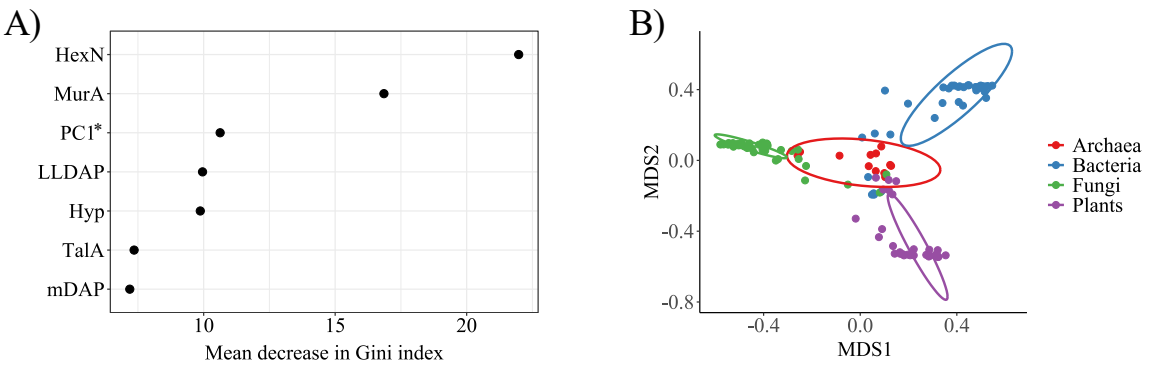

Figure 5. Random forest model results using amino sugars, non-proteinogenic amino acid contents and the PC1 proxy of proteinogenic amino acid contents (PC1*) in samples of archaea, bacteria, fungi, and plants. (A) Feature importance for decision tree classification process in the random forest model as detected by the mean decrease in Gini index. (B) Multidimensional scaling (MDS) plot of the random forest model proximity matrix.

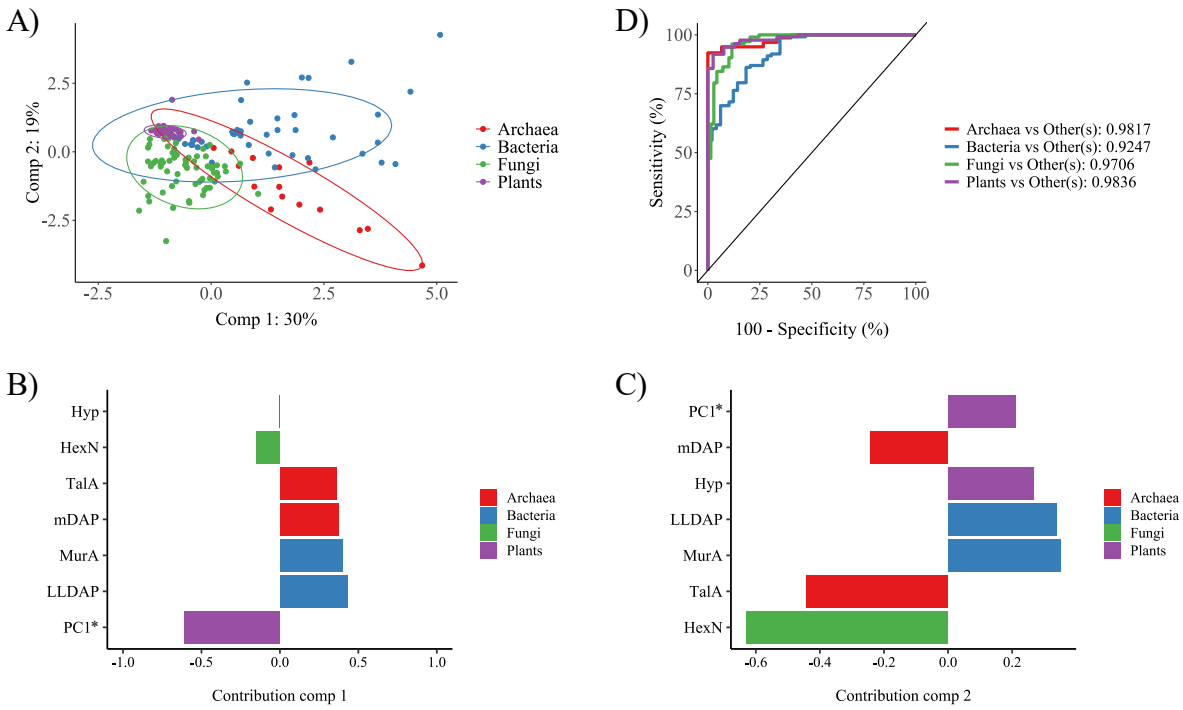

Figure 6. (A) Partial least squares discriminant analysis (PLS-DA) score results using amino sugars, non-proteinogenic amino acid contents and the PC1 proxy of proteinogenic amino acid contents (PC1*) in samples of archaea, bacteria, fungi, and plants. Ellipses represent the 95% confidence interval. (B & C) Contribution plots of the most discriminatory amino compounds for component 1 and component 2. Y-axes in plots B and C represent the loading weight of each amino compound in components 1 and 2, respectively. (D) Receiver operating characteristic (ROC) curves using components 1,2 and 3 showcasing the performance evaluation of the PLS-DA model for each taxonomic group. Specificity indicates the rates of true negatives. Sensitivity indicates the rates of true positives. Numbers on the right indicate the accuracy of the classification performed by the PLS-DA model.

Supplements

Table S1. Detected masses and retention times of AQC-labeled amino acids and amino sugars measured by RP-UPLC-ESI-Orbitrap HRMS.

Compound	Chemical formula	*m/z* $[M+H]^+$	AQC derivatives $[M+H]^+$	Retention time (min)
Glycine	$C_2H_5NO_2$	75.0320	246.0872	5.05
Alanine	$C_3H_7NO_2$	89.0477	260.1026	6.55
Serine	$C_3H_7NO_3$	105.0426	276.0977	4.46
Proline	$C_5H_9NO_2$	115.0633	286.1184	7.30
Valine	$C_5H_{11}NO_2$	117.0790	288.1340	9.00
Threonine	$C_4H_9NO_3$	119.0582	290.1133	6.44
Cysteine	$C_3H_7NO_2S$	121.0198	292.0745	7.30
Hydroxyproline	$C_5H_9NO_3$	131.0582	302.1134	1.94
Isoleucine	$C_6H_{13}NO_2$	131.0946	302.1497	10.34
Leucine	$C_6H_{13}NO_2$	131.0946	302.1497	10.47
Asparagine	$C_4H_8N_2O_3$	132.0535	303.1091	2.43
Aspartic acid	$C_4H_7NO_4$	133.0375	304.0926	5.54
Glutamine	$C_5H_{10}N_2O_3$	146.0691	317.1248	4.78
Lysine	$C_6H_{14}N_2O_2$	146.1055	317.1606	8.08
Glutamic acid	$C_5H_9NO_4$	147.0532	318.1090	5.93
Methionine	$C_5H_{11}NO_2S$	149.0511	320.1061	8.77
Histidine	$C_6H_9N_3O_2$	155.0695	326.1244	1.05
Phenylalanine	$C_9H_{11}NO_2$	165.0790	336.1339	10.55
Arginine	$C_6H_{14}N_4O_2$	174.1117	345.1670	3.20
Hexosamines	$C_6H_{13}NO_5$	179.0794	350.1339	0.96
Tyrosine	$C_9H_{11}NO_3$	181.0739	352.1288	8.47
meso-diaminopimelic acid	$C_7H_{14}N_2O_4$	190.0954	361.1508	6.96 / 7.09
L,L-diaminopimelic acid	$C_7H_{14}N_2O_4$	190.0954	361.1502	7.08
Talosaminuronic acid	$C_6H_{11}NO_6$	193.0586	364.1136	1.10
Tryptophan	$C_{11}H_{12}N_2O_2$	204.0899	375.1447	10.59
Muramic acid	$C_9H_{17}NO_7$	251.1005	422.1552	5.45 / 6.21

Table S2. Average (± standard deviation) amino acid and amino sugar contents (μg mg^{-1} dw) in pure hydrolyzed biomass of archaea, bacteria, fungi, and plants using AQC derivatization and UPLC-Orbitrap MS.

Taxon	Species	Ala	Ser	Glu	Asp	Arg	Gly	His	Phe	Met	Lys	Pro	Leu	Ile	Thr	Tyr	Val	Hyp	LLDAP	mDAP	HexN	TalA	MurA	N
Archaea	Average	62.6 (±20.7)	53.2 (±31.6)	12.6 (±3.8)	120.3 (±51.0)	41.1 (±12.6)	50.7 (±20.5)	7.2 (±3.1)	28.7 (±10.2)	15.7 (±6.8)	70.9 (±20.6)	32.7 (±12.5)	52.4 (±15.1)	33.7 (±11.3)	49.3 (±23.2)	32.9 (±14.9)	35.7 (±12.4)	n.d	2.6 (± 4.7)	28.3 (± 22.8)	5.9 (±4.1)	8.2 (±7.4)	n.d	15
	Saccharolobus solfataricus	54.6	69.7	11.9	139.6	41.5	57.2	5.4	34.7	16.1	69.8	39.7	59.4	40.6	60.0	50.0	40.2	n.d.	n.d.	23.4	3.1	n.d.	n.d.	1
	Sulfolobus acidocaldarius	59.8	97.4	13.8	154.7	51.9	68.1	5.5	41.5	18.2	77.1	47.3	73.1	46.7	77.0	57.1	51.1	n.d.	n.d.	n.d.	10.0	n.d.	n.d.	1
	Sulfolobus islandicus	60.7	82.9	13.6	143.9	42.9	64.9	4.7	35.8	16.3	72.1	44.7	68.2	42.0	70.3	51.5	42.0	n.d.	n.d.	43.1	3.8	n.d.	n.d.	1
	Haloferax vulcanii	36.2	33.5	11.0	93.8	25.9	33.6	3.4	14.1	6.8	16.8	16.6	27.2	15.2	31.9	13.8	21.5	n.d.	n.d.	41.0	n.d.	n.d.	n.d.	1
	Methanobacterium sp.	52.8 (±12.0)	23.9 (±5.9)	8.8 (±2.9)	61.9 (±13.8)	36.6 (±9.3)	29.6 (±6.8)	10.9 (±2.3)	25.2 (±4.7)	13.4 (±3.2)	66.1 (±16.2)	23.2 (±5.7)	46.6 (±9.8)	24.1 (±5.2)	27.8 (±6.5)	22.2 (±6.1)	27.2 (±6.3)	n.d.	9.7 (± 3.6)	28.8 (± 6.3)	7.5 (±1.5)	14.5 (±3.1)	n.d.	4
	Methanothermobacter marburgensis	91.2	52.9	19.9	137.9	59.7	63.8	7.8	28.4	23.7	88.9	38.4	56.5	36.5	49.3	28.7	40.7	n.d.	n.d.	89.4	9.9	19.4	n.d.	1
	Methanothermus fervidus	71.5	33.6	14.1	100.9	31.0	40.3	4.7	20.3	11.9	80.7	26.6	37.7	30.0	43.8	24.0	29.2	n.d.	n.d.	8.7	12.2	11.7	n.d.	1
	Methanocaldococcus jannaschii	74.0	58.9	14.4	162.4	44.5	66.8	6.2	31.9	9.2	86.6	39.8	61.7	46.2	54.6	38.7	46.8	n.d.	n.d.	54.0	4.7	15.7	n.d.	1
	Methanothermococcus okinawensis	40.8	36.6	10.1	107.8	23.7	38.4	3.6	16.1	9.7	55.2	22.0	33.2	25.1	34.1	22.1	22.1	n.d.	n.d.	7.7	0.7	1.5	n.d.	1
	Methanotorris igneus	56.8	42.3	13.0	134.9	35.1	53.0	5.7	25.3	14.8	75.9	30.6	50.5	37.9	45.2	30.9	32.0	n.d.	n.d.	8.8	2.7	10.2	n.d.	1
	Methanospirillum stamsii	115.9	131.1	19.7	248.4	67.6	99.5	10.8	53.6	33.6	104.9	61.6	79.4	54.8	109.6	59.1	64.8	n.d.	n.d.	23.2	0.2	0.6	n.d.	1
	Methanosarcina spelaei	67.2	64.1	12.4	132.3	45.9	57.0	6.9	26.9	22.0	70.4	31.0	52.1	33.6	52.3	29.5	36.3	n.d.	n.d.	9.5	10.9	5.3	n.d.	1
Bacteria	Average	68.5 (±36.8)	37.8 (±29.9)	9.9 (±5.2)	77.7 (±57.5)	35.3 (±26.1)	45.9 (±29.1)	6.9 (±4.0)	23.3 (±13.9)	9.8 (±7.0)	38.9 (±23.3)	26.2 (±16.2)	41.5 (±21.6)	17.7 (±11.4)	38.1 (±27.8)	18.4 (±13.7)	26.5 (±15.5)	0.2 (±1.0)	19.1 (±26.5)	18.6 (±21.1)	6.3 (±5.6)	n.d.	32.2 (±46.2)	49
	Koribacter versatilis	41.1	37.7	6.2	72.4	23.8	33.5	2.9	17.4	4.1	20.9	20.4	27.7	11.2	35.7	10.8	18.0	n.d.	0.8	14.5	0.6	n.d.	1.4	1
	Terriglobus roseus	102.3	84.9	15.1	156.1	66.9	76.8	8.5	37.6	15.0	55.9	52.6	63.2	34.4	81.9	35.5	44.5	0.1	2.9	65.9	2.6	n.d.	5.2	1
	Terriglobus sp.	91.7	66.7	12.7	135.2	54.4	65.0	6.6	32.7	12.3	54.5	41.9	54.1	25.0	66.7	27.7	35.4	0.4	4.7	55.4	1.7	n.d.	7.3	1
	Solibacter usitatus	96.2	78.8	12.4	137.0	44.2	80.6	4.2	36.2	9.9	31.5	49.8	57.0	27.3	91.0	20.3	45.3	n.d.	4.4	30.7	2.4	n.d.	7.0	1
	Spirosoma spitsbergense	138.1	95.5	16.5	201.7	82.1	102.2	8.4	53.3	13.9	63.1	61.0	76.8	39.9	96.7	53.7	50.5	n.d.	19.0	58.0	4.3	n.d.	33.1	1
	Synechococcus sp.	23.7 (±9.2)	13.0 (±5.4)	4.1 (±1.7)	30.5 (±12.1)	14.2 (±5.9)	13.6 (±5.8)	3.1 (±1.2)	12.7 (±4.7)	2.8 (±1.6)	12.3 (±5.1)	9.5 (±3.8)	21.1 (±7.9)	7.6 (±3.2)	15.0 (±6.1)	4.6 (±2.6)	11.0 (±4.3)	n.d.	n.d.	12.8 (±6.0)	1.7 (±0.7)	n.d.	4.2 (±1.6)	1
	Chelativorans multitrophicus	140.1	89.6	19.7	194.2	102.0	114.8	9.9	55.0	25.3	54.0	58.5	84.7	39.0	88.1	48.1	58.2	n.d.	28.5	49.2	4.9	n.d.	44.4	4
	Neorhizobium sp.	135.0	95.5	20.9	207.9	80.3	110.5	7.0	52.2	19.8	56.7	54.5	78.2	41.2	83.3	42.4	55.9	n.d.	17.6	88.9	4.3	n.d.	36.6	1

	Labilithrix luteola	156.1	112.9	20.2	228.3	119.4	119.2	10.7	65.1	21.6	68.2	75.4	106.4	49.4	103.3	56.6	71.0	n.d.	15.4	59.1	4.5	n.d.	27.8	1
	Klebsiella aerogenes	104.5	86.3	18.0	193.4	80.7	79.8	13.0	42.7	27.1	81.1	46.0	83.4	41.5	76.2	43.4	54.8	0.0	7.3	35.2	3.1	n.d.	9.4	1
	Enterobacter sp.	44.1	33.3	6.4	66.8	26.4	34.0	1.5	15.0	3.3	25.0	21.2	26.6	11.4	32.5	9.5	15.2	n.d.	n.d.	32.5	7.5	n.d.	n.d.	1
	Eschericia coli	21.8 (±10.9)	11.8 (±5.6)	2.7 (±1.5)	26.2 (±12.2)	11.2 (±5.2)	34.1 (±17.5)	3.8 (±1.7)	8.6 (±3.6)	2.6 (±1.2)	15.5 (±6.5)	21.6 (±11.7)	16.3 (±8.1)	5.4 (±1.8)	10.3 (±4.6)	6.4 (±3.1)	9.5 (±4.4)	n.d.	n.d.	3.0 (±1.5)	0.4 (±0.2)	n.d.	n.d.	5
	Chthoniobacter flavus	83.9	69.7	11.4	128.0	43.3	72.4	6.3	35.2	12.2	36.8	44.3	54.1	25.1	74.7	21.1	38.7	n.d.	10.8	37.7	11.4	n.d.	11.9	1
	Arthrobacter globiformis	88.1	62.4	14.8	102.2	43.3	53.0	5.0	20.2	19.5	47.7	30.9	43.1	18.9	59.1	16.2	31.2	6.3	n.d.	20.9	9.8	n.d.	29.1	1
	Arthrobacter tecti	128.5	34.5	12.5	82.2	33.3	42.7	3.3	15.0	18.0	35.9	24.5	31.2	13.5	52.7	12.1	24.1	3.6	n.d.	13.7	8.4	n.d.	47.5	1
	Micrococcus luteus (lyophylized cells)	56.7 (±3.0)	11.7 (±0.6)	7.9 (±0.5)	32.2 (±1.8)	22.9 (±1.4)	27.4 (±1.2)	6.5 (±0.5)	11.2 (±0.5)	13.0 (±2.4)	51.4 (±3.6)	17.1 (±1.7)	24.6 (±1.4)	8.0 (±1.0)	16.6 (±1.3)	7.3 (±0.6)	16.4 (±0.8)	n.d.	6.2 (±0.4)	4.1 (±0.2)	5.7 (±0.3)	n.d.	n.d.	5
	Micrococcus luteus (peptidoglycan)	50.4 (±1.9)	16.7 (±0.7)	8.1 (±0.5)	42.4 (±1.1)	19.1 (±0.6)	18.0 (±0.7)	8.7 (±0.5)	18.6 (±1.0)	4.1 (±0.7)	34.3 (±1.1)	13.8 (±0.6)	32.2 (±3.0)	16.0 (±1.1)	19.3 (±1.0)	14.9 (±0.7)	19.4 (±0.9)	n.d.	40.6 (±2.3)	n.d.	16.0 (±1.5)	n.d.	27.1 (±1.5)	4
	Pseudarthrobacter oxydans	100.1	63.1	21.1	101.9	41.7	50.9	4.7	18.6	25.8	48.9	29.8	41.6	18.4	63.9	15.4	32.1	0.6	0.0	13.8	9.7	n.d.	35.6	1
	Rhodococcus sp.	89.7	80.9	15.2	115.8	43.6	62.2	5.1	20.8	11.9	29.2	29.7	47.5	21.8	60.7	15.7	33.6	0.4	22.4	34.1	4.4	n.d.	34.1	1
	Solirubrobacter soli	37.7	25.5	5.0	42.4	17.0	30.2	0.8	11.3	3.9	9.7	15.4	21.6	9.5	24.8	6.1	15.4	n.d.	2.6	12.9	2.3	n.d.	4.7	1
	Streptomyces sp.	74.4 (±26.6)	33.3 (±22.3)	9.2 (±4.1)	66.6 (±28.7)	38.8 (±14.2)	45.1 (±16.4)	10.4 (±5.5)	22.5 (±8.1)	7.0 (±2.3)	29.0 (±11.0)	26.9 (±9.6)	47.0 (±16.1)	14.5 (±7.2)	33.1 (±14.3)	17.0 (±6.8)	28.6 (±9.7)	0.1 (±0.2)	38.6 (±19.1)	25.3 (±8.8)	6.3 (±1.6)	n.d.	53.7 (±28.5)	5
	Streptosporangium roseum	109.1	68.7	15.3	144.2	83.4	79.5	8.2	35.0	12.1	51.4	44.0	58.5	25.8	75.1	36.4	53.2	n.d.	30.5	34.8	8.1	n.d.	45.5	1
	Bacillus subtilis	69.9 (±33.8)	23.5 (±13.4)	10.2 (±4.3)	46.6 (±30.4)	17.0 (±9.6)	29.3 (±20.5)	6.6 (±4.4)	23.6 (±6.4)	5.5 (±4.6)	34.8 (±32.2)	13.3 (±4.8)	39.6 (±8.1)	17.0 (±7.1)	21.5 (±10.3)	17.2 (±7.5)	19.3 (±6.8)	n.d.	66.5 (±44.3)	n.d.	9.9 (±6.6)	n.d.	94.4 (±67.7)	4
	Paenibacillus alginolyticus	93.6	51.9	11.0	90.2	60.1	70.1	5.6	24.9	11.3	37.3	35.1	55.3	18.4	50.6	19.5	35.7	0.0	0.8	30.2	26.3	n.d.	238.6	1
	Staphylococcus aureus	59.2 (±9.4)	29.5 (±11.0)	10.0 (±1.4)	76.0 (±35.3)	29.8 (±12.8)	49.2 (±23.5)	12.8 (±4.9)	31.3 (±8.6)	12.3 (±6.1)	73.2 (±37.5)	18.7 (±5.9)	52.4 (±12.7)	23.0 (±7.3)	31.1 (±11.8)	27.6 (±9.8)	28.0 (±8.8)	n.d.	29.3 (±43.3)	2.1 (±1.5)	5.1 (±4.3)	n.d.	46.5 (±38.5)	4
Fungi	Average	31.9 (±17.8)	32.2 (±17.5)	6.5 (±3.4)	60.9 (±30.4)	29.3 (±18.9)	25.1 (±12.4)	4.2 (±2.3)	13.0 (±6.5)	4.3 (±2.8)	28.5 (±14.5)	18.9 (±10.2)	20.7 (±10.5)	10.4 (±5.6)	24.6 (±13.0)	11.6 (±6.0)	13.8 (±7.1)	0.0 (±0.1)	n.d.	n.d.	15.0 (±9.5)	n.d.	n.d.	69
	Alternaria alternata	50.2	49.0	10.6	78.7	54.4	38.5	6.3	18.6	8.3	44.9	27.0	28.7	14.4	32.8	19.4	19.5	0.1	n.d.	15.4	18.7	n.d.	n.d.	1
	Aspergillus brasiliensis	60.7	53.4	11.5	84.5	52.3	39.7	7.7	20.8	8.9	38.0	30.5	35.2	18.1	40.7	21.3	24.2	0.0	n.d.	19.9	20.9	n.d.	n.d.	1
	Aspergillus nidulans	32.5	32.0	9.5	65.0	32.3	24.9	4.1	13.5	4.1	24.0	20.7	21.8	11.3	24.4	11.7	14.9	0.1	n.d.	8.5	21.8	n.d.	n.d.	1
	Blastobotrys sp.	27.2	28.6	8.0	57.6	24.0	22.8	4.2	11.9	4.5	30.0	50.1	18.2	9.1	28.9	11.5	12.8	0.0	n.d.	19.9	12.3	n.d.	n.d.	1
	Cadophora finlandica	63.9	51.7	10.9	86.1	47.3	59.1	6.1	19.7	8.1	50.2	25.3	31.3	16.4	33.0	18.9	21.5	0.0	n.d.	17.6	12.0	n.d.	n.d.	1
	Cephalotrichum microsporum	31.9	31.3	5.7	50.9	18.5	23.8	2.4	10.7	3.3	22.0	22.5	18.4	9.4	24.1	10.7	12.9	0.0	n.d.	5.7	28.2	n.d.	n.d.	1

Cephalotrichum sp.	36.8	34.3	7.2	55.9	23.3	24.9	2.5	11.7	4.0	23.7	25.5	19.4	10.7	26.9	10.7	14.2	0.0	n.d.	8.4	24.1	n.d.	n.d.	1
Clonostachys rosea	14.5 (±0.9)	16.3 (±3.5)	3.1 (±0.2)	25.2 (±1.4)	14.7 (±0.6)	9.7 (±0.6)	4.4 (±0.2)	9.3 (±0.7)	1.7 (±0.4)	21.6 (±1.6)	8.6 (±0.6)	14.9 (±1.4)	6.3 (±0.4)	10.5 (±0.8)	7.3 (±0.4)	8.2 (±0.7)	n.d.	n.d.	4.9 (±0.6)	14.1 (±1.1)	n.d.	n.d.	5
Coniochaeta hoffmannii	49.7	39.4	8.0	79.8	36.1	35.9	5.1	19.1	8.9	38.5	26.8	30.1	14.1	34.6	15.8	20.1	0.1	n.d.	27.8	9.2	n.d.	n.d.	1
Epicoccum nigrum	51.9	43.2	8.4	72.9	57.1	35.5	6.8	17.0	8.5	40.7	23.0	25.7	13.0	30.2	16.0	18.1	0.1	n.d.	19.9	11.2	n.d.	n.d.	1
Exophiala sp.	34.5	40.7	8.8	61.3	90.8	29.8	5.5	13.4	4.4	53.3	19.1	21.8	11.0	34.6	11.8	15.5	0.1	n.d.	12.9	3.9	n.d.	n.d.	1
Fusarium fujikuroi	22.2 (±4.6)	12.0 (±1.8)	3.8 (±0.6)	33.9 (±5.2)	16.0 (±2.8)	13.8 (±1.3)	6.2 (±1.1)	12.8 (±0.8)	3.3 (±0.1)	22.2 (±4.6)	14.2 (±2.2)	21.9 (±1.9)	7.2 (±1.4)	16.6 (±3.4)	10.1 (±1.4)	11.6 (±1.5)	n.d.	n.d.	5.5 (±1.3)	6.3 (±3.6)	n.d.	n.d.	5
Geomyces sp.	44.7	48.9	10.4	80.4	34.9	35.8	5.0	16.2	6.5	37.0	28.2	28.9	13.3	36.5	16.1	17.6	0.0	n.d.	31.4	16.8	n.d.	n.d.	1
Gibellulopsis nigrescens	47.8	47.3	11.3	80.2	41.3	37.1	8.0	18.9	6.4	42.4	32.8	29.4	15.9	47.3	18.8	21.7	0.1	n.d.	18.4	11.2	n.d.	n.d.	1
Gonytrichum macrocladum	41.6	45.9	10.5	78.3	35.3	34.3	4.4	18.0	6.0	33.0	23.8	28.6	14.6	35.1	14.3	21.2	0.1	n.d.	8.6	10.9	n.d.	n.d.	1
Heydenia sp.	52.3	39.8	7.0	61.9	32.6	28.1	3.8	12.8	5.2	32.1	34.9	22.3	10.8	35.0	16.5	18.0	0.3	n.d.	31.9	16.9	n.d.	n.d.	1
Microdochium bolleyi	61.0	59.8	13.7	95.4	69.3	45.5	8.6	22.4	7.8	66.7	34.4	34.3	20.1	53.9	26.9	26.3	0.0	n.d.	20.0	12.1	n.d.	n.d.	1
Morchella elata	35.6	35.2	6.6	64.2	28.2	28.9	3.4	14.6	6.3	24.0	20.4	23.6	14.2	31.0	13.3	16.8	n.d.	n.d.	9.8	9.0	n.d.	n.d.	1
Penicillium chrysogenum	49.4	54.7	9.0	95.7	40.5	40.8	5.5	24.7	11.8	43.3	30.4	40.0	20.2	44.2	21.5	26.6	n.d.	n.d.	16.7	21.5	n.d.	n.d.	1
Periconia macrospinosa	3.6	6.3	0.4	3.7	1.1	3.1	0.0	1.1	0.0	0.6	4.5	1.9	0.8	5.2	0.9	2.0	0.0	n.d.	2.5	8.1	n.d.	n.d.	1
Peziza ostracoderma	34.4	37.3	7.0	63.5	25.7	28.7	3.7	15.3	5.1	32.3	40.9	22.7	11.9	36.1	13.1	16.2	0.1	n.d.	10.1	34.5	n.d.	n.d.	1
Phialocephala helvetica	43.2	35.4	7.9	56.4	44.2	28.9	3.2	12.4	5.0	24.2	20.5	19.6	10.1	27.2	11.1	12.9	0.4	n.d.	19.0	9.2	n.d.	n.d.	1
Pichia fermentans	13.7	19.8	5.5	28.9	7.8	10.8	2.2	7.6	2.2	14.8	7.7	11.0	5.8	13.3	5.0	7.2	0.1	n.d.	2.6	1.5	n.d.	n.d.	1
Pyrenochaeta acicola	61.9	49.4	10.5	81.3	59.0	35.7	6.7	20.9	8.2	45.1	30.0	29.8	15.6	36.3	18.2	20.4	0.1	n.d.	19.7	25.5	n.d.	n.d.	1
Pyrenochaeta cava	44.1	45.8	8.4	64.0	74.8	29.5	4.7	14.1	7.4	34.1	20.2	22.4	11.5	26.9	13.1	15.0	0.1	n.d.	19.0	21.0	n.d.	n.d.	1
Schizosaccharomyces pombe	30.0	41.0	7.8	71.0	53.7	33.6	5.3	15.6	7.1	34.8	18.4	24.0	13.6	29.3	15.2	18.3	n.d.	n.d.	18.7	0.5	n.d.	n.d.	1
Schwanniomyces capriottii	21.7	27.4	9.5	46.8	21.8	17.3	3.3	9.0	2.1	27.0	11.5	14.2	8.0	19.6	9.8	9.3	0.0	n.d.	17.2	9.6	n.d.	n.d.	1
Tetracladium furcatum	29.4	17.5	3.5	24.4	9.6	24.2	0.9	3.8	0.4	28.6	5.2	6.8	3.7	7.0	2.4	4.0	n.d.	n.d.	2.9	2.6	n.d.	n.d.	1
Tetracladium marchalianum	57.3	65.0	11.3	113.4	61.1	46.8	2.3	24.2	6.2	57.7	36.4	43.0	16.9	46.4	10.6	23.4	0.4	n.d.	25.3	11.1	n.d.	n.d.	1
Tetracladium maxilliforme	16.0	20.2	4.3	36.7	26.0	14.5	2.6	7.6	2.7	18.2	10.1	12.1	5.7	13.6	6.4	7.4	0.0	n.d.	9.5	3.3	n.d.	n.d.	1
Trichocladium asperum	49.6	54.7	11.5	87.4	42.0	37.6	5.8	18.5	5.7	37.3	25.8	32.6	17.2	39.6	18.2	23.9	0.0	n.d.	20.9	22.5	n.d.	n.d.	1
Trichoderma flavofuscum	3.3	4.2	0.7	5.9	1.8	3.2	0.0	1.5	0.0	1.3	2.7	2.1	1.0	3.5	1.2	1.5	n.d.	n.d.	1.4	3.8	n.d.	n.d.	1

	Tuber sp.	68.2	74.4	8.2	69.5	25.0	42.1	3.7	15.0	5.0	21.7	24.0	26.0	24.9	39.2	13.8	23.8	n.d.	n.d.	14.4	29.3	n.d.	n.d.	1
	Warcupia sp.	36.7	39.5	7.0	75.0	24.4	33.8	8.5	16.5	4.8	27.6	21.5	25.3	13.0	27.6	13.1	17.4	0.1	n.d.	14.8	50.5	n.d.	n.d.	1
	Boliniaceae	28.2	31.2	6.7	50.6	26.5	24.0	1.9	11.0	3.5	19.2	15.0	18.0	8.2	25.5	8.8	11.7	0.0	n.d.	17.1	12.4	n.d.	n.d.	1
	Chaetosphaeriaceae	43.9	42.9	11.8	74.5	60.5	32.2	6.1	17.7	6.7	48.7	23.7	28.0	13.0	36.5	17.0	18.6	0.0	n.d.	13.2	7.0	n.d.	n.d.	1
	Helotiales	52.0	56.8	10.1	80.6	42.5	36.6	7.0	15.7	4.2	38.1	27.6	25.8	15.0	40.0	15.8	18.0	0.1	n.d.	30.7	14.3	n.d.	n.d.	1
	Agaricus sp.	41.9	42.9	7.0	68.8	27.2	31.4	3.9	17.9	4.7	26.7	22.6	28.8	17.3	31.6	10.9	20.4	n.d.	n.d.	8.5	14.2	n.d.	n.d.	1
	Amanita muscaria	26.8	26.7	5.3	43.5	17.0	19.7	3.5	11.1	3.4	16.2	13.9	17.1	9.0	20.1	7.6	10.8	n.d.	n.d.	7.8	10.3	n.d.	n.d.	1
	Armillaria bulbosa	5.7	8.6	1.4	16.3	4.1	8.3	0.6	3.2	0.0	2.9	6.2	4.8	3.3	8.4	1.7	3.9	0.0	n.d.	3.0	14.6	n.d.	n.d.	1
	Auricularia polytricha	19.5	18.8	2.8	29.8	10.5	13.0	1.3	7.5	1.5	7.8	9.2	11.3	4.8	16.9	6.6	7.4	n.d.	n.d.	7.0	5.5	n.d.	n.d.	1
	Craterellus cornucopioides	12.6	19.2	2.6	28.8	10.6	13.4	1.5	7.6	1.6	9.8	10.0	11.3	7.5	16.2	6.2	7.9	n.d.	n.d.	0.0	15.1	n.d.	n.d.	1
	Grifola frondosa	31.8	38.0	7.1	65.0	24.4	30.2	3.6	15.8	5.8	23.8	19.9	25.2	13.1	29.2	12.7	17.0	n.d.	n.d.	9.5	11.7	n.d.	n.d.	1
	Hericium sp.	6.6	8.3	1.4	14.5	3.6	6.0	0.0	2.9	0.0	1.8	4.5	4.9	2.4	5.9	2.1	3.1	0.0	n.d.	0.0	6.7	n.d.	n.d.	1
	Lentinula edodes	27.9	32.7	8.3	57.2	23.8	23.3	2.2	13.6	3.8	22.9	17.2	20.8	11.7	27.3	11.6	15.9	0.0	n.d.	9.7	17.8	n.d.	n.d.	1
	Psilocybe cyanescens	14.6	20.1	2.1	26.5	6.7	14.8	0.6	7.4	1.2	5.9	10.8	8.9	4.9	14.4	5.5	7.0	0.0	n.d.	5.0	38.6	n.d.	n.d.	1
	Ramaria sp.	17.4	18.5	3.3	29.2	9.6	16.8	1.3	7.3	1.4	8.8	10.2	10.8	5.7	15.2	6.5	7.3	0.0	n.d.	12.7	11.8	n.d.	n.d.	1
	Volvariella volvacea	57.5	79.0	11.2	126.0	49.4	50.2	6.4	36.4	8.6	51.2	39.8	54.2	25.0	55.8	27.7	32.9	n.d.	n.d.	49.3	17.5	n.d.	n.d.	1
	Mortierella alpina	50.7	56.6	9.3	110.4	47.0	44.3	8.6	24.4	10.5	57.1	29.3	37.9	18.3	41.5	21.8	23.9	0.2	n.d.	35.9	6.8	n.d.	n.d.	1
	Mortierella gamsii	54.2	39.2	8.3	80.0	26.4	34.4	4.1	17.1	4.1	34.4	22.1	28.0	13.0	29.8	14.8	18.7	0.2	n.d.	24.7	17.7	n.d.	n.d.	1
	Mucor hiemalis	37.0	39.8	10.2	95.2	33.6	30.4	7.0	15.7	5.3	40.0	20.9	25.3	13.6	28.6	15.0	16.5	0.1	n.d.	21.2	19.1	n.d.	n.d.	1
	Rhizophagus irregularis	6.4	9.8	1.7	78.5	32.4	13.9	2.8	3.6	0.5	27.0	5.9	4.4	2.7	6.4	5.0	3.0	n.d.	n.d.	1.8	13.8	n.d.	n.d.	1
	Rhizophagus irregularis (hyphae)	7.8	13.1	1.9	57.3	15.5	14.3	0.7	3.9	0.0	10.6	8.2	5.1	2.5	7.2	3.3	2.9	n.d.	n.d.	1.5	24.7	n.d.	n.d.	1
	Rhizophagus irregularis (hyphae and spores)	7.2 (±2.9)	14.4 (±3.7)	1.9 (±0.4)	123.1 (±24.3)	21.9 (±14.3)	19.0 (±3.9)	3.9 (±1.5)	3.2 (±0.9)	0.8 (±0.4)	34.0 (±12.5)	8.3 (±2.1)	4.0 (±1.6)	2.7 (±1.1)	6.9 (±2.7)	4.9 (±2.1)	3.1 (±0.9)	n.d.	n.d.	2.2 (±1.2)	27.1 (±9.0)	n.d.	n.d.	3
	Rhizopus oligosporus	17.9	19.9	5.1	33.3	16.1	16.3	1.8	9.3	3.5	9.5	10.4	14.9	7.1	12.9	8.7	10.2	0.1	n.d.	1.8	2.9	n.d.	n.d.	1
	Rhizopus oryzae	16.2	21.3	4.0	44.9	15.1	15.5	2.7	8.3	2.6	19.1	10.1	12.5	6.9	15.0	8.3	8.3	0.0	n.d.	10.7	17.3	n.d.	n.d.	1
	Thamnidium elegans	51.3	38.4	8.4	75.6	22.6	26.2	6.5	15.9	5.1	30.8	23.4	23.5	11.6	28.4	15.4	14.7	0.0	n.d.	22.3	29.6	n.d.	n.d.	1
	Umbelopsis ramanniana	43.3	40.4	9.2	84.4	38.7	33.0	6.6	17.6	7.5	42.2	23.6	28.0	14.5	33.8	19.1	18.6	0.0	n.d.	27.5	12.9	n.d.	n.d.	1
	Umbelopsis vinacea	40.7	45.4	9.7	86.3	44.7	32.9	6.0	17.9	6.9	41.1	24.8	28.5	15.4	31.7	19.6	18.2	0.1	n.d.	29.6	12.7	n.d.	n.d.	1
Plants	Average	12.9 (±6.7)	13.2 (±7.6)	2.7 (±1.4)	29.7 (±13.7)	7.5 (±7.0)	12.8 (±6.9)	1.5 (±1.1)	7.5 (±4.2)	2.3 (±1.5)	9.1 (±5.0)	10.8 (±7.9)	12.0 (±7.0)	4.3 (±3.4)	10.5 (±5.4)	5.1 (±2.8)	7.4 (±4.0)	0.4 (±0.5)	n.d.	n.d.	0.5 (±0.5)	n.d.	n.d.	39

Artichum undulatum	9.5	7.7	1.6	20.4	4.3	9.9	0.7	4.1	1.0	5.5	5.0	6.3	2.7	7.1	3.0	4.0	0.2	n.d.	3.9	0.4	n.d.	n.d.	1
Brachythecium rutabulum	13.6	13.7	2.7	28.6	8.6	14.5	1.0	6.1	2.0	7.2	7.9	9.2	4.3	10.8	4.0	6.3	0.5	n.d.	6.3	1.5	n.d.	n.d.	1
Hypnum cupressiforme	14.3	16.1	3.0	42.7	13.3	17.6	1.1	7.0	1.8	8.8	8.2	10.2	4.2	11.8	5.2	6.6	0.2	n.d.	6.6	0.7	n.d.	n.d.	1
Sphagnum magellanicum	5.0	5.3	1.0	18.1	2.4	4.7	0.2	1.8	0.3	2.1	2.8	3.1	1.3	4.1	1.5	2.2	0.2	n.d.	0.9	0.6	n.d.	n.d.	1
Tortella tortuosa	8.9	9.3	1.5	21.1	5.4	10.2	0.5	3.8	0.9	4.4	4.8	5.4	2.3	7.1	2.8	3.9	0.5	n.d.	4.7	1.3	n.d.	n.d.	1
Bazzania trilobata	8.2	8.6	1.5	18.4	6.4	8.6	0.6	3.6	0.9	4.4	5.1	5.8	2.5	7.2	2.6	3.7	0.1	n.d.	3.2	1.0	n.d.	n.d.	1
Marchantia polymorpha	10.4	11.2	2.2	33.9	8.0	9.9	0.8	4.8	1.6	6.7	5.9	7.1	3.1	8.7	3.5	4.7	0.2	n.d.	4.3	0.8	n.d.	n.d.	1
Metzgeria conjugata	7.7	8.5	1.5	16.8	3.7	8.1	0.3	3.6	0.6	3.9	4.9	5.2	2.4	7.2	2.6	3.6	0.2	n.d.	3.1	0.7	n.d.	n.d.	1
Plagiochila asplenioides	9.4	10.4	2.1	27.0	5.1	9.7	0.5	4.0	0.7	4.6	5.5	6.4	3.1	8.5	2.5	4.3	0.2	n.d.	2.8	1.8	n.d.	n.d.	1
Anthoceros sp.	3.4	3.2	0.5	4.9	2.8	3.7	0.1	1.3	n.d.	1.7	1.9	2.1	0.9	2.4	0.3	1.5	0.1	n.d.	0.7	0.5	n.d.	n.d.	1
Equisetum arvense	15.5	15.3	2.9	28.3	9.6	14.7	1.1	8.3	2.9	9.8	9.2	12.7	5.7	11.5	6.0	7.7	0.2	n.d.	5.2	0.2	n.d.	n.d.	1
Picea abies	7.2	8.1	1.4	14.3	5.0	7.7	0.3	4.0	1.1	4.6	5.1	6.1	3.0	6.0	2.4	3.9	0.3	n.d.	1.5	0.2	n.d.	n.d.	1
Ginko biloba	15.9	16.8	3.2	33.3	9.5	16.6	1.4	8.8	2.3	10.8	17.8	12.2	6.9	12.3	5.9	8.7	0.6	n.d.	2.9	0.6	n.d.	n.d.	1
Allium ursinum	26.9	23.4	5.0	47.7	15.1	23.6	2.2	13.9	4.7	13.8	15.0	20.3	10.5	19.8	10.7	12.9	0.2	n.d.	6.1	1.1	n.d.	n.d.	1
Alchemilla vulgaris	12.8	13.2	2.3	24.6	8.2	13.0	1.1	7.0	2.6	8.3	8.0	10.7	4.9	10.4	5.7	6.8	0.2	n.d.	2.8	0.1	n.d.	n.d.	1
Anthriscus cerefolium	14.5 (±0.7)	11.0 (±0.5)	2.8 (±0.1)	24.9 (±0.8)	n.d.	12.0 (±0.5)	3.4 (±0.4)	12.0 (±1.0)	3.4 (±0.4)	14.1 (±1.0)	10.3 (±0.5)	23.0 (±1.0)	2.3 (±0.3)	11.0 (±0.5)	7.9 (±0.5)	9.5 (±0.5)	n.d.	n.d.	4.8 (±0.3)	0.1 (±0.0)	n.d.	n.d.	5
Betula pendula	14.5	15.7	2.9	28.3	9.5	15.6	1.1	8.0	2.3	8.2	9.7	12.1	5.6	11.5	5.7	7.3	0.7	n.d.	9.1	0.3	n.d.	n.d.	1
Capsella bursa-pastoris	18.8	20.9	4.5	38.2	12.6	19.7	1.6	10.7	4.2	13.2	35.5	16.0	9.7	16.7	7.6	13.3	0.5	n.d.	2.7	0.4	n.d.	n.d.	1
Dictamnus albus	13.2	15.7	2.7	36.7	8.4	14.6	1.0	8.5	2.7	8.0	18.8	12.2	6.3	12.3	6.3	8.2	0.4	n.d.	2.4	0.4	n.d.	n.d.	1
Fumaria officinalis	15.9	18.0	3.6	38.3	11.4	16.3	1.4	8.6	2.9	9.2	14.8	13.3	7.1	13.8	6.1	9.5	0.6	n.d.	3.4	0.9	n.d.	n.d.	1
Galium verum	10.0	11.1	1.8	20.3	5.5	10.1	0.8	5.3	1.7	6.1	8.6	8.2	3.7	8.1	3.8	5.4	0.5	n.d.	4.4	0.3	n.d.	n.d.	1
Malva silvestris	6.5 (±0.2)	5.6 (±0.2)	1.3 (±0.1)	22.3 (±0.7)	n.d.	5.3 (±0.1)	1.9 (±0.1)	5.1 (±0.2)	1.3 (±0.1)	6.7 (±0.3)	5.2 (±0.2)	8.2 (±0.4)	n.d.	4.9 (±0.3)	3.0 (±0.2)	4.5 (±0.1)	0.0	n.d.	1.4 (±0.5)	0.4 (±0.0)	n.d.	n.d.	5
Petroselinum crispum (leaves)	23.2	23.5	4.0	47.3	15.1	23.5	2.5	13.5	5.1	17.1	14.3	20.4	9.7	20.2	10.4	14.0	0.3	n.d.	3.9	0.1	n.d.	n.d.	1
Petroselinum crispum (roots)	7.5	7.4	5.8	26.3	4.0	6.2	0.3	2.9	0.9	4.6	3.6	5.1	2.6	6.0	1.8	3.8	0.3	n.d.	n.d.	0.1	n.d.	n.d.	1
Plantago major	10.8	12.3	2.3	21.7	7.2	11.8	0.7	6.4	2.2	5.7	7.9	9.2	5.1	9.5	4.6	6.3	0.4	n.d.	n.d.	0.4	n.d.	n.d.	1
Polygonum aviculare	22.1	24.2	4.9	40.9	13.6	22.0	2.0	12.3	4.6	13.8	32.6	17.8	9.2	17.4	8.6	12.8	0.5	n.d.	5.5	0.9	n.d.	n.d.	1
Taraxacum officinale (leaves)	31.1	34.3	6.8	76.3	19.9	33.0	2.9	18.4	5.5	20.8	28.1	27.6	14.0	25.5	10.6	19.0	0.4	n.d.	13.4	1.6	n.d.	n.d.	1

Taraxacum officinale (roots)	4.9	5.4	2.6	27.4	20.6	4.6	0.3	2.4	0.8	3.9	22.5	3.4	2.4	4.7	1.6	3.0	0.5	n.d.	n.d.	0.1	n.d.	n.d.	1
Urtica dioica (leaves)	31.0	33.7	5.5	63.9	21.6	30.1	3.0	15.8	5.5	23.2	19.2	25.2	11.1	23.1	9.0	16.3	0.8	n.d.	12.3	0.4	n.d.	n.d.	1
Urtica dioica (roots)	15.3	25.4	3.8	52.3	27.8	18.7	1.5	7.6	2.2	13.1	15.6	10.2	6.9	15.8	5.8	11.4	3.1	n.d.	2.8	1.0	n.d.	n.d.	1
Vaccinium myrtillus	10.9	13.4	1.9	22.7	6.5	12.1	0.8	6.2	2.0	6.7	7.8	9.8	4.2	9.0	4.4	6.0	0.7	n.d.	3.2	0.3	n.d.	n.d.	1

N: number of replicates. n.d.: not detectable

Table S3. Indicator species analysis of amino acid and amino sugar contents measured in the hydrolyzed biomass of archaea, bacteria, fungi, and plants.

Taxa	Amino acid	Indicator value
Archaea	TalA	0.771
Bacteria	MurA	0.808
	LLDAP	0.637
Plants	Hyp	0.561
Archaea + Bacteria	PC1.m[a]	0.445
Fungi + Plants	PC1*[b]	0.642
Archaea + Bacteria + Fungi	HexN	0.708
	mDAP	0.385

[a] PC1.m refers to the PC1 scores of the relative molar percentage of proteinogenic amino acids content. [b] PC1* refers to the PC1 scores of the proteinogenic amino acids content.

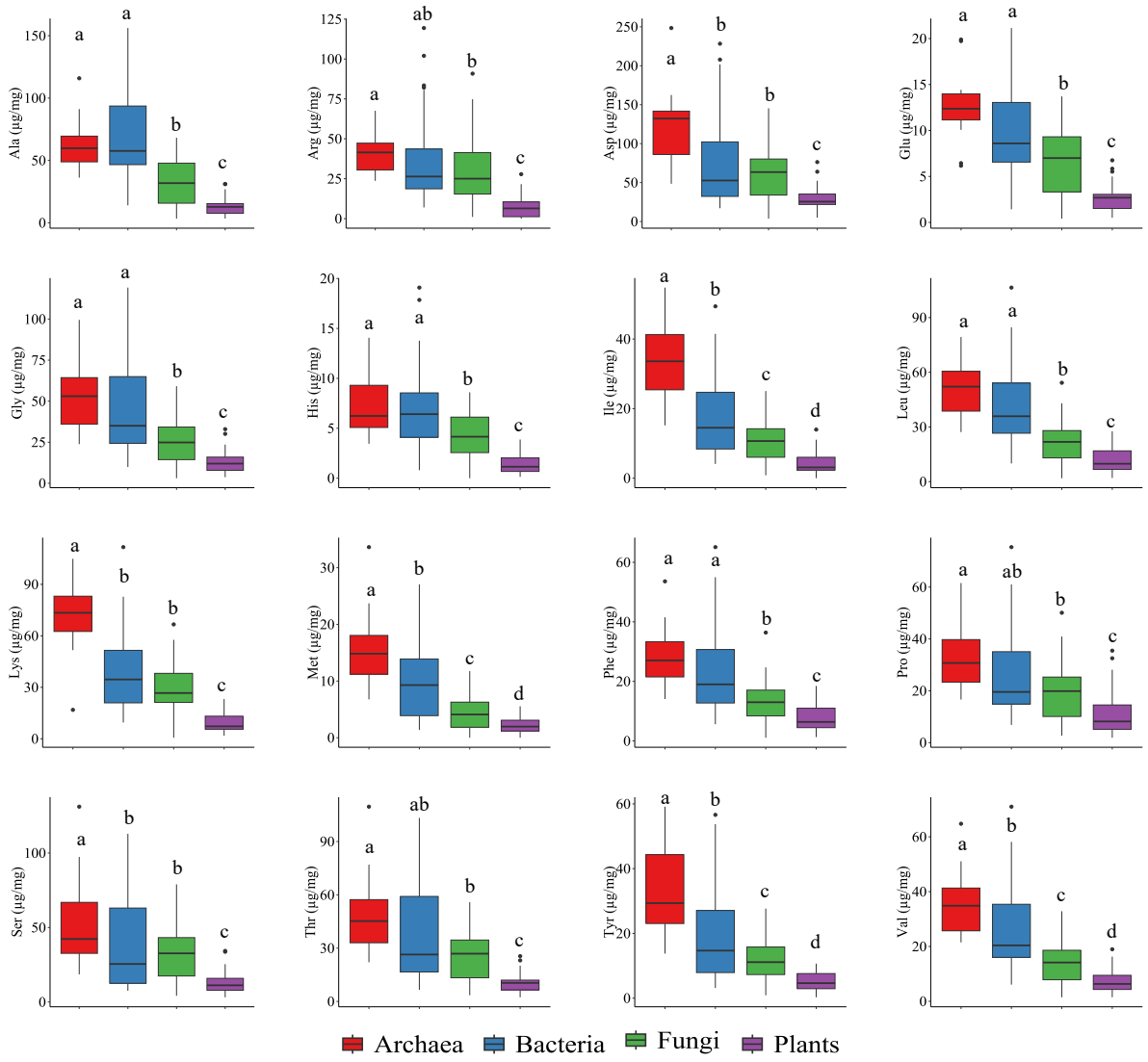

Figure S1. Proteinogenic amino acid contents in the hydrolyzed biomass of archaea, bacteria, fungi, and plants using AQC derivatization via UHPLC-Orbitrap MS. Different letters indicate significant differences ($P < 0.05$) between taxonomic groups.

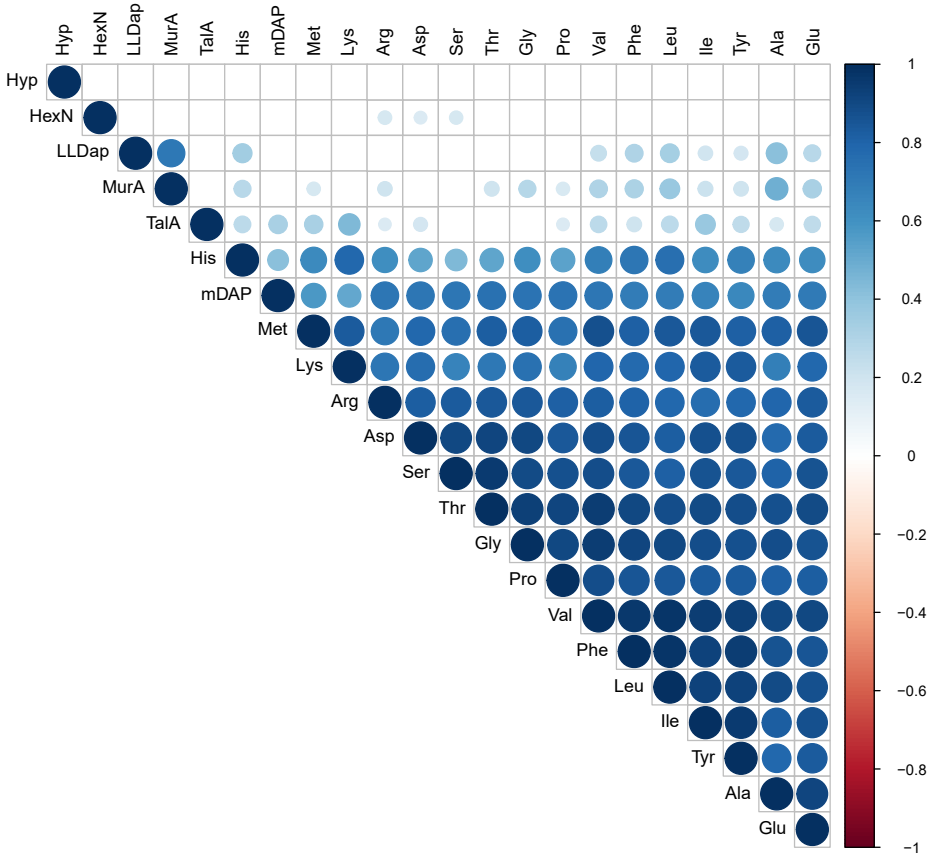

Figure S2. Correlation plot of proteinogenic amino acid contents in the hydrolyzed biomass of archaea, bacteria, fungi, and plants. Circles indicate a significant correlation ($p<0.01$), where bigger circles indicate a stronger correlation.

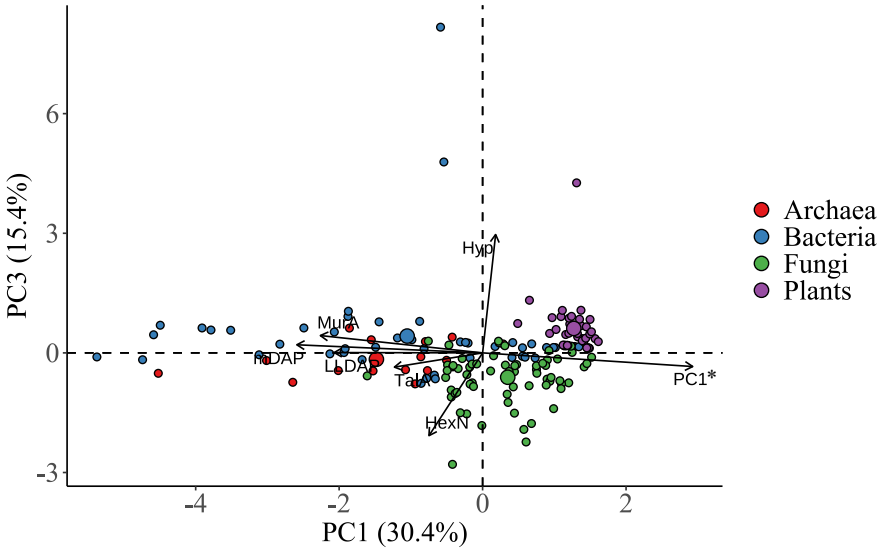

Figure S3. Principal component analysis (PCA) biplot using components 1 and 3 of non-proteinogenic amino acid contents and the PC1 proxy of proteinogenic amino acid contents (PC1*) in samples of archaea, bacteria, fungi, and plants. Large circles represent the data centroids.

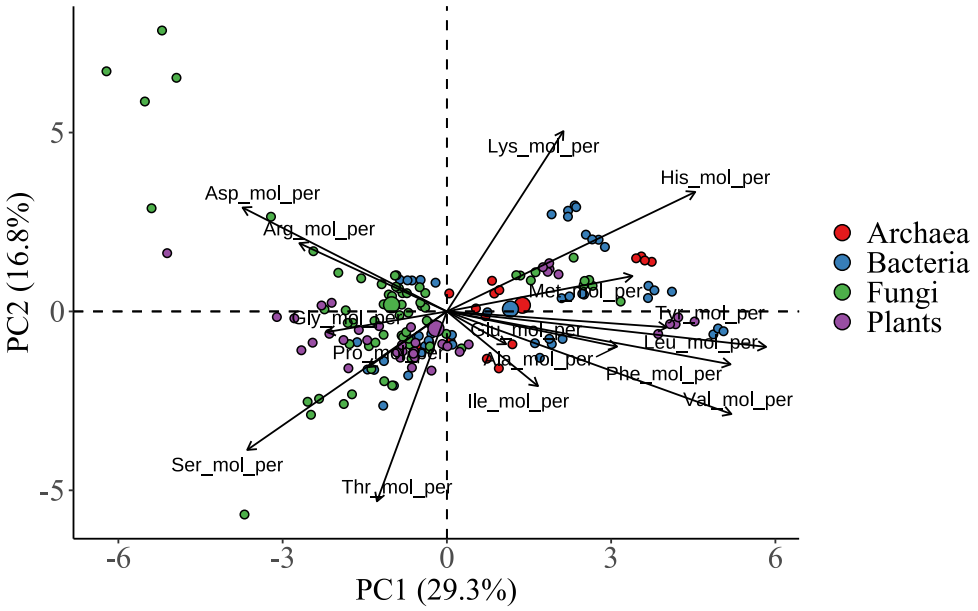

Figure S4. Principal component analysis (PCA) biplot using the relative molar percentages of proteinogenic amino acid contents in samples of archaea, bacteria, fungi, and plants. Large circles represent the data centroids.

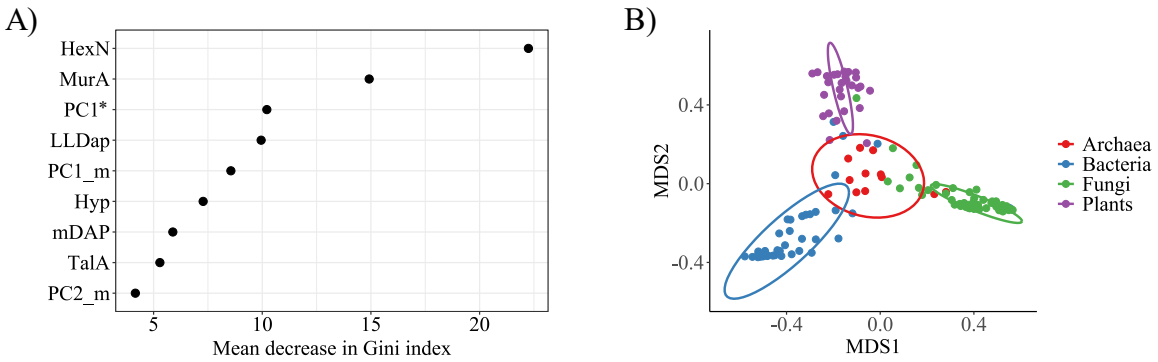

Figure S5. Random forest model results using amino sugars, non-proteinogenic amino acid contents, PC1 scores of proteinogenic amino acid contents (PC1*), and PC1 (PC1_m) and PC2 (PC2_m) scores of the relative contents (molar percentages) of proteinogenic amino acid contents in samples of archaea, bacteria, fungi, and plants. (A) Feature importance for decision tree classification process in the random forest model as detected by the mean decrease in Gini index. (B) Multidimensional scaling (MDS) plot of the random forest model proximity matrix.

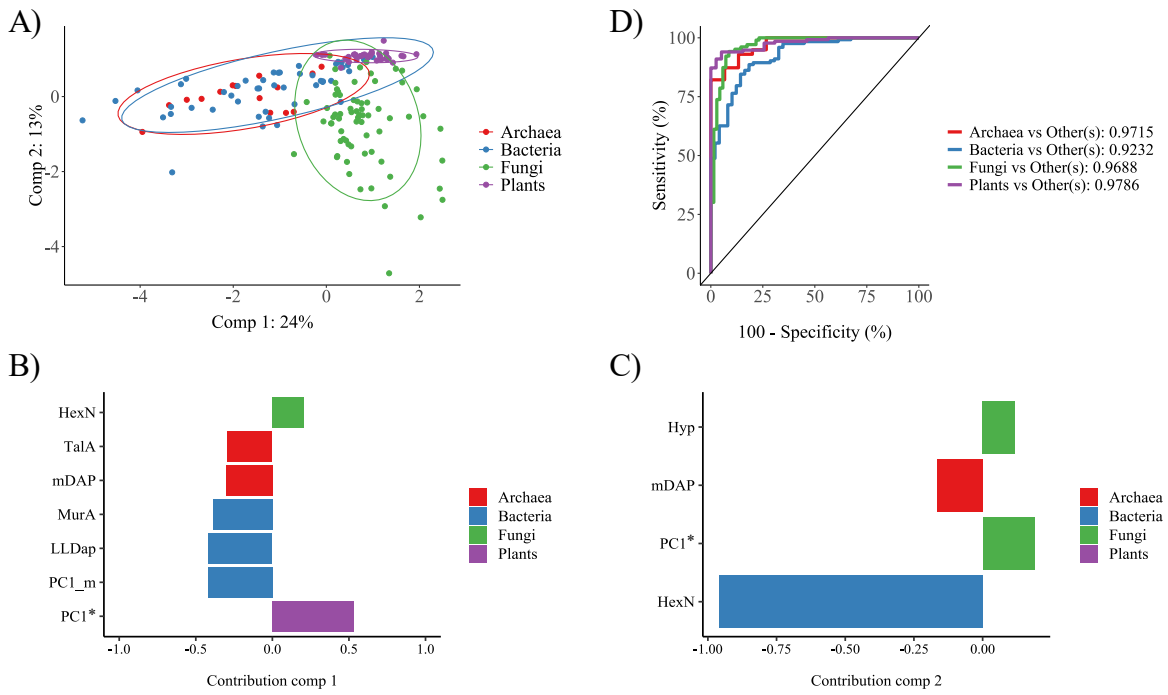

Figure S6. (A) Partial least squares discriminant analysis (PLS-DA) score results using amino sugars, non-proteinogenic amino acid contents, PC1 scores of proteinogenic amino acid contents (PC1*), and PC1 (PC1_m) and PC2 (PC2_m) scores of the relative contents (molar percentages) of proteinogenic amino acid contents in samples of archaea, bacteria, fungi, and plants. Ellipses represent the 95% confidence interval. (B & C) Contribution plots of the most discriminatory amino compounds for component 1 and component 2. Y-axes in plots B and C represent the loading weight of each amino compound in components 1 and 2, respectively. (D) Receiver operating characteristic (ROC) curves using components 1,2 and 3 showcasing the performance evaluation of the PLS-DA model for each taxonomic group. Specificity indicates the rates of true negatives. Sensitivity indicates the rates of true positives. Numbers on the right indicate the accuracy of the classification performed by the PLS-DA model.

Chapter V

Summary

Summary

Soil necromass biomarkers are commonly used to estimate the microbial contribution to soil organic matter. However, there are still many uncertainties regarding the interpretation of microbial necromass biomarkers as well as limited information on biomarkers from many representative soil organism groups and plants. In this PhD project I identified and quantified amino sugar, neutral sugar, and amino acid necromass biomarkers in the biomass of representative soil groups of archaea, bacteria, fungi, and plants. In chapter II, I focused on developing a method to quantify amino sugar and neutral sugar biomarkers simultaneously. Chapters III and IV were dedicated to gaining deeper insights into amino sugar, neutral sugar, and amino acid necromass biomarkers. This provided a comprehensive analysis of current and novel potential necromass biomarkers encompassing a large range of compounds in a wide variety of soil taxonomic groups.

In chapter II, I developed a method to identify and quantify amino sugars and neutral sugars as necromass biomarkers using 1-phenyl-3-methyl-5-pyrazolone (PMP) derivatization coupled to ultra-high performance liquid chromatography (UHPLC) and high-resolution Orbitrap mass spectrometry. The method successfully separated and allowed the quantification of 18 sugar-related compounds in a single 20 minute assay. Method validation showed very low relative standard deviations in retention time (<0.4%) and repeatability (<2.6%), very low mass error (<0.8 ppm), very high linearity (>0.9991) and low limits of detection (0.03 – 1.73 µM) and quantification (0.10 – 5.76 µM). The precision of the measurement of ^{13}C isotopic enrichment was calculated to be 0.03 atom % for U-^{13}C D-glucose and 0.05 atom % for ^{13}C D-glucosamine-1-^{13}C. This meant that this PMP protocol is more sensitive to ^{13}C enrichment detection than gas chromatography/mass spectrometry methods (Patterson et al., 1997; Preston & Slater, 1994), while at the same time it allows a lower concentration sensitivity than what is achieved by isotope ratio mass spectrometry (Sessions, 2006).

This PMP method was successfully applied on acid-hydrolysed (6 M HCl) biomass of archaea, bacteria, fungi and plants, as well as on samples from forest, grassland, and cropland soils. This was the first time such a wide range of compounds were separated simultaneously using PMP derivatization in a single assay, including the capability of using Orbitrap mass spectrometry for isotope enrichment quantification.

It was also the first time talosaminuronic acid was quantified using PMP derivatization alongside other amino sugars and neutral sugars.

In chapter III, I analysed amino sugar and neutral sugar biomarkers in the biomass of archaeal, bacterial, fungal, and plant species using the PMP method developed in chapter II. The results showed that glucosamine (GlcN) was the most common amino sugar found across taxonomic groups (representing 42-91% of total amino sugar contents), being indicative of archaea, bacteria, and fungi. Muramic acid was associated only to bacteria, and it was positively correlated to GlcN content only in Firmicutes. In Actinobacteria and gram-negative bacteria, the MurA contents remained constant even at high GlcN contents, which suggested that GlcN in these organisms likely originated from other structural components and not only from peptidoglycan. We also found that both galactosamine and talosaminuronic acid were specific indicators for archaea. Galactosamine likely originated from methanochondroitin, a cell wall polymer produced by some Methanosarcinales (Albers & Meyer, 2011), while talosaminuronic acid originated from archaeal pseudopeptidoglycan (Subedi et al., 2021). We found that hexoses were more abundant than pentoses in the biomass of all taxonomic groups. We also calculated a hexose/pentose ratio of 9 or lower for bacterial groups and 22 or higher for plants and fungi. Our results showed that a combination of amino sugars and neutral sugar biomarkers improved the separation of plants from other microbial groups.

We calculated new conversion factors for the extrapolation of amino sugar contents into microbial or plant C. We estimated a GlcN conversion factor of 10 for fungal carbon (C) and two MurA conversion factors: 19 for gram-positive bacterial C and 118 for gram-negative bacterial C. The conversion factors estimated in this chapter had wider confidence intervals than those presented by Appuhn & Joergensen (2006). This likely arose from a larger diversity of fungal species and bacterial groups analysed in this subproject. Therefore, a measurement of error propagation should be accounted for when extrapolating amino sugar contents into soil microbial necromass.

When using these conversion factors, several limitations should be considered. For example, the gram-positive to gram-negative bacterial ratio in the soil should be estimated before using only one conversion factor from MurA to bacterial C since the MurA conversion factor varied significantly between these two bacterial groups (19 versus 118). Additionally, conversion factors were derived from the biomass of organisms which included cell wall and cytoplasmic components. This means that the

conversion factors were representative of cell wall residues and not necessarily total microbial necromass since extracellular polymeric substances (EPS) were not analysed in this subproject. Lastly, biomarker approaches assume that amino sugars and their ratios remain the same after cell death, decomposition and recycling. However, amino sugar stabilization and turnover in complex matrixes such as soils still require further studies.

In chapter IV I analysed amino acid and amino sugar biomarkers in the biomass of the same species utilized in chapter III using 6-aminoquinolyl-N-hydroxysuccinimidyl carbamate (AQC) derivatization coupled to UHPLC and Orbitrap mass spectrometry. By using a combination of indicator species analysis and multivariate machine learning methods (random forest and discriminant analysis), we confirmed the biomarker potential of non-proteinogenic amino acids and amino sugars. Our results showed the separation of prokaryotes (archaea + bacteria) from eukaryotes (fungi + plants) based on the proteinogenic amino acid content in the biomass of these taxonomic groups. Both non-proteinogenic amino acids and amino sugars (the sum of hexosamines, muramic acid and talosaminuronic acid) can be used as necromass biomarkers. Hydroxyproline (Hyp) was shown to be an amino acid biomarker specific for plants, since the most common sources in soils are Hyp-rich glycoproteins of plant cell walls (HRPGs) (Nguema-Ona et al., 2014). Similarly, we found that L,L-diaminopimelic acid could be used alongside muramic acid as a biomarker specific for bacteria, where gram-positive bacteria had higher contents compared to gram-negative bacteria.

Both machine learning models used had very high accuracy, with the random forest model having a 92.3% accuracy and the discriminant analysis having less than 8% classification error. The combined use of machine learning algorithms improved the robustness of our results and showed that the concurrent use of both amino acid and amino sugar biomarkers can successfully discriminate between broad taxonomic groups.

Finally, we showed that AQC derivatization successfully separated amino acid and amino sugar biomarkers in a single run, showcasing an unparallel method to trace both microbial and plant necromass biomarkers simultaneously. This approach would allow to quantify the microbial versus plant contributions to soil organic matter along the continuum from litter decomposition to soil C stabilization via microbial processing.

Overall, this PhD study provided novel insights into the identification and quantification of soil necromass biomarkers from archaea, bacteria, fungi, and plants. By using a method specifically developed for the analysis of reducing sugars (PMP derivatization, Chapter II), I could analyse amino sugars and neutral sugar biomarkers from prominent soil microbial groups as well as plants (Chapter III). By using a method specific for amino groups (AQC derivatization), I could extend the biomarker analysis to include not only amino sugars, but also less commonly used non-proteinogenic amino acids (Chapter IV).

With our results, we expanded on the current knowledge of GlcN as a tracer for fungal and MurA as a tracer for bacterial residues (Engelking et al., 2007; Joergensen, 2018; Liang et al., 2019; Wang et al., 2021). First, we identified new necromass biomarkers specific for archaea (talosaminuronic acid), plants (Hyp), and bacteria (LLDAP). This not only significantly expanded the number and scope of compounds that are currently used as soil necromass biomarkers (Appuhn & Joergensen, 2006; Joergensen, 2018), but also it will allow to trace necromass residues from archaea and plants together with those from bacteria and fungi in soils using the same analytical protocols. Second, we proposed new conversion factors to extrapolate amino sugar contents into microbial or plant carbon which we derived from the biomass analysis of a broad range of taxonomic groups. This brought new insights into the necromass biomarker composition of other representative soil groups including archaea and plants. Third, we pointed to the limitations of using conversion factors from necromass biomarkers to extrapolate microbial contributions to soil organic carbon. Specifically, by generating conversion factors with broader confidence intervals than previously reported (Appuhn & Joergensen, 2006; Engelking et al., 2007), we highlight the importance of adding error propagation measures to avoid over- or underestimating the quantification of microbial necromass carbon in soils. Finally, we simultaneously quantified microbial and plant necromass using amino sugar and amino acid biomarkers. Different analytical protocols are currently needed when comparing microbial versus plant-derived organic matter in soils since amino sugars are used to trace microbes and lignin and long-chain alkanes are applied for plants (Angst et al., 2021; Craig et al., 2022; Whalen et al., 2022). By combining the measurements of microbial and plant residues into one protocol of amino-compound biomarkers, we could eliminate the use of different compound classes with varying turnover kinetics as biomarkers (lignin /alkanes versus amino sugars). At the same

time, it allows to follow carbon stabilization from litter to soil organic matter via microbial processing namely the microbial carbon pump.

This PhD has provided important new perspectives into the use of amino sugars, neutral sugars, and amino acids as biomarkers to trace both microbial and plant necromass into soil ecosystems and how they contribute to soil organic C accumulation. While there is still a large gap regarding how these necromass biomarkers are transformed after cell death, as well as on their turnover and (de)stabilization rates in soils, our findings pave the way to the next exploratory steps to increase our understanding of soil organic matter formation and the role microbes and plants play in it.

References

Albers, S.-V., & Meyer, B. H. (2011). The archaeal cell envelope. *Nature Reviews Microbiology*, *9*(6), 414–426. https://doi.org/10.1038/nrmicro2576

Angst, G., Mueller, K. E., Nierop, K. G. J., & 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. https://doi.org/10.1016/j.soilbio.2021.108189

Appuhn, A., & Joergensen, R. (2006). Microbial colonisation of roots as a function of plant species. *Soil Biology and Biochemistry*, *38*(5), 1040–1051. https://doi.org/10.1016/j.soilbio.2005.09.002

Craig, M. E., Geyer, K. M., Beidler, K. V., Brzostek, E. R., Frey, S. D., Stuart Grandy, A., Liang, C., & Phillips, R. P. (2022). Fast-decaying plant litter enhances soil carbon in temperate forests but not through microbial physiological traits. *Nature Communications*, *13*(1), 1229. https://doi.org/10.1038/s41467-022-28715-9

Engelking, B., Flessa, H., & Joergensen, R. G. (2007). Shifts in amino sugar and ergosterol contents after addition of sucrose and cellulose to soil. *Soil Biology and Biochemistry*, *39*(8), 2111–2118. https://doi.org/10.1016/j.soilbio.2007.03.020

Joergensen, R. G. (2018). Amino sugars as specific indices for fungal and bacterial residues in soil. *Biology and Fertility of Soils*, *54*(5), 559–568. https://doi.org/10.1007/s00374-018-1288-3

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. https://doi.org/10.1111/gcb.14781

Nguema-Ona, E., Vicre-Gibouin, M., Gotte, M., Plancot, B., Lerouge, P., Bardor, M., & Driouich, A. (2014). Cell wall O-glycoproteins and N-glycoproteins: Aspects of biosynthesis and function. *Frontiers in Plant Science*, *5*, 499. https://doi.org/10.3389/fpls.2014.00499

Patterson, B. W., Zhang, X.-J., Chen, Y., Klein, S., & Wolfe, R. R. (1997). Measurement of very low stable isotope enrichments by gas chromatography/mass spectrometry: Application to measurement of muscle protein synthesis. *Metabolism*, *46*(8), 943–948. https://doi.org/10.1016/S0026-0495(97)90084-6

Preston, T., & Slater, C. (1994). Mass spectrometric analysis of stable-isotope-labelled amino acid tracers. *Proceedings of the Nutrition Society*, *53*(2), 363–372. https://doi.org/10.1079/PNS19940042

Sessions, A. L. (2006). Isotope-ratio detection for gas chromatography. *Journal of Separation Science*, *29*(12), 1946–1961. https://doi.org/10.1002/jssc.200600002

Subedi, B. P., Martin, W. F., Carbone, V., Duin, E. C., Cronin, B., Sauter, J., Schofield, L. R., Sutherland-Smith, A. J., & Ronimus, R. S. (2021). Archaeal pseudomurein and bacterial murein cell wall biosynthesis share a common evolutionary ancestry. *FEMS Microbes*, *2*, xtab012. https://doi.org/10.1093/femsmc/xtab012

Wang, B., An, S., Liang, C., Liu, Y., & Kuzyakov, Y. (2021). Microbial necromass as the source of soil organic carbon in global ecosystems. *Soil Biology and Biochemistry*, *162*, 108422. https://doi.org/10.1016/j.soilbio.2021.108422

Whalen, E. D., Grandy, A. S., Sokol, N. W., Keiluweit, M., Ernakovich, J., Smith, R. G., & Frey, S. D. (2022). Clarifying the evidence for microbial- and plant-derived soil organic matter, and the path toward a more quantitative understanding. *Global Change Biology*, *00*, 1–19. https://doi.org/10.1111/gcb.16413