Soil Biology and Biochemistry 125 (2018) 290–299

Contents lists available at ScienceDirect

Soil Biology and Biochemistry

journal homepage: www.elsevier.com/locate/soilbio

Addressing the challenge of soil metaproteome complexity by improving metaproteome depth of coverage through two-dimensional liquid chromatography

Stephen J. Callister[a], Thomas L. Fillmore[b], Carrie D. Nicora[a], Jared B. Shaw[b], Samuel O. Purvine[b], Daniel J. Orton[a], Richard Allen White III[a], Ronald J. Moore[a], Meagan C. Burnet[a], Ernesto S. Nakayasu[a], Samuel H. Payne[a], Janet K. Jansson[a,*], Ljiljana Paša-Tolić[b,**]

[a] Biological Sciences Division, Pacific Northwest National Laboratory, Richland WA 99352, USA
[b] Environmental Molecular Sciences Division, Pacific Northwest National Laboratory, Richland WA 99352, USA

ARTICLE INFO

Keywords:
Metaproteomics
Proteomics
Mass spectrometry
Soil
Prairie
Liquid chromatography

ABSTRACT

Metaproteomics conducted on soil is challenged by a low depth of protein coverage that can potentially result in an underrepresentation of the functional underpinnings of important biological processes and interactions. Typically, the utilization of an on-line two-dimensional chromatographic separation approach (2D LC-MS/MS) can significantly improve depth of coverage. Herein, we evaluate different fractionation modalities to determine the optimal approach for LC MS based soil metaproteomics. The first approach fractionates the digested soil proteome in 2 dimensions while coupled directly to the MS instrument ("online" approach). The second approach performs the first dimension of fractionation "offline" prior to injection to the MS ("offline" approach). While both approaches are commonly utilized for proteomic research, they have not been directly compared for soils. We rigorously compared these approaches applied to: 1) a mock community consisting of 47 different microorganisms, and 2) to natural soil. The results provide insight into protein dynamic range, the presence of mass spectrometry interfering substances, and other factors that may contribute to an observed low metaproteome depth of coverage from complex and highly diverse samples, such as soil. We observed that the "offline" approach generally resulted in the highest metaproteome coverage; however, there are advantages to using the "online" approach when dealing with limited biomass. Both approaches resulted in a larger number of protein identifications from the synthetic metaproteome rather than from the soil metaproteome, although the soil metaproteome had a significantly larger number of predicted proteins. A large dynamic range in abundances of proteins resulting from metabolically active and inactive populations within the soil metaproteome explains this observation.

1. Introduction

Mass spectrometry (MS) based metaproteomics conducted on soil represents an important analytical tool for understanding biological function within the soil environment. The generation of metaproteomics data begins with isolating proteins from the soil matrix and preparing them for the MS analysis. The generalized workflow for a soil metaproteomic study involves protein extraction using direct or indirect extraction methods, followed by protein digestion, sample clean-up, and gel based or liquid chromatography based separation of peptides prior to ionization and analysis by the tandem mass spectrometer (MS/MS) (Bastida et al., 2009). In some cases, following extraction, proteins are separated by polyacrylamide gel electrophoresis (1D-PAGE and 2D-PAGE) then digested (Maron et al., 2007) and analyzed using LCMS. The workflow is completed with application of bioinformatics tools for processing the acquired mass spectra, assignment of amino acid sequences to measured peptide tandem mass spectra (MS/MS), and protein identification. While this workflow is conceptually straightforward it is often challenging to carry out because of edaphic factors, the presence of interfering humic substances (Qian and Hettich, 2017) other contaminants, low microbial biomass, sample complexity, high microbial diversity and high evenness. In addition, the majority of soil

* Corresponding author. 902 Battelle Boulevard, P.O. Box 999, MSIN K8-98 Richland, WA 99352, USA.
** Corresponding author. 902 Battelle Boulevard, P.O. Box 999, MSIN K8-98, Richland, WA 99352, USA.
E-mail addresses: janet.jansson@pnnl.gov (J.K. Jansson), ljijana.pasatolic@pnnl.gov (L. Paša-Tolić).

https://doi.org/10.1016/j.soilbio.2018.07.018
Received 15 February 2018; Received in revised form 30 May 2018; Accepted 19 July 2018

microbes have not yet been cultivated and their functional genes have not been sequenced and annotated. The extraction of proteins from the soil also remains a continual challenge, where low yields and different extraction methods can influence the observed metaproteome (Taylor and Williams, 2010; Keiblinger et al., 2012; Becher et al., 2013). Taken together, these factors can place a severe limitation on the number of proteins that can be identified by commonly used database searches against annotated soil genomes and metagenomes (Hultman et al., 2015), and has often resulted in metaproteomics being performed on amended soils (Canizares et al., 2011; Bastida et al., 2016a; Starke et al., 2016, 2017; Liu et al., 2017). However, for soils where amendments have not been undertaken, these factors frequently result in the suboptimal recovery of proteins and a relatively low number of protein identifications (i.e. low depth of metaproteome coverage) as expected given the predicted genetic potential within a gram of soil equates to approximately 10^{12} genes (Prosser, 2015).

The degree of metaproteome complexity within the soil matrix is empirically difficult to determine, and is largely described in qualitative terms relative to a soil metagenomic database (when it exists). Generally, the quality and coverage of a given metaproteome is scored on the basis of several factors including the dynamic range and abundances of proteins; the number of identified proteins and their sequence similarities; and the identities of specific members of the microbiome that contribute to the soil metaproteome at any given time (the active microbiome (Artursson et al., 2005; Singer et al., 2017). Because most soil metaproteomic workflows are peptide centric, proteome complexity increases significantly following protein digestion.

To mitigate metaproteome complexity and increase metaproteome depth of coverage, the use of high-performance, or ultra-performance liquid chromatographic separation of peptides (HPLC or UPLC) is commonly employed, reviewed in (Kota and Stolowitz, 2016). For soil proteomics research, LC separation has traditionally been carried out using a one-dimensional (1D) C18 reversed-phase (RP) separation coupled to the MS, yielding a highly variable number of proteins, from the hundreds to the thousands (for examples see (Bastida et al., 2014; Lunsmann et al., 2016; Sidibe et al., 2016; Mattarozzi et al., 2017). However, the utilization of a two-dimensional (2D) separation approach can improve depth of coverage by employing orthogonal separation methods. In general, peptides separated in the first dimension separation are eluted onto the second dimension separation as “online fractions” by modifying the mobile phase gradient in an iterative or step-wise manner. Originally described as the Multidimensional Protein Identification Technology (MuDPIT) approach (Washburn et al., 2001; Wolters et al., 2001), a strong cation exchange (SCX) separation is integrated with a RP separation in a biphasic, or triphasic (RP, SCX, RP) micro-capillary column (Florens and Washburn, 2006). Other 2D online orthogonal combinations are also possible such as RP and size exclusion (SEC) (Gilar et al., 2005). Initial soil metaproteomics demonstrations of 2D liquid chromatography with tandem mass spectrometry (2D LC-MS/MS) were used to successfully evaluate different protein extraction methods (Chourey et al., 2010; Keiblinger et al., 2012), while a more relatively recent example used this approach to elucidate the metaproteomes associated with different stages of thawing permafrost soil (Hultman et al., 2015). These examples illustrate initial demonstrations of the 2D online approach for conducting soil metaproteomics, but to our knowledge it has not been successfully evaluated against other LC separation approaches applied to the soil.

The LC separation of peptides “offline” using SEC, SCX or C18 high-pH RP separation prior to low-pH RPLC-MS/MS analysis has also been utilized to reduce proteome complexity predominantly from non-soil environments, as described by Wang et al. (2011) (Wang et al., 2011). This approach uses LC to separate peptides that are collected as fractions and their concentrations adjusted prior to LC-MS/MS analysis. Technically, this approach also represents a 2D LC-MS/MS approach, but both dimensions are not directly coupled to the mass spectrometer and thus is referred to as offline fractionation. Major advantages of this approach include the ability to adjust the number of fractions generated to suite the anticipated complexity of the metaproteome and the ability to work with separate fractions requiring additional clean-up, or concentration/dilution to achieve an optimal on-column mass for the second dimension of separation. While several studies utilizing an offline fractionation approach have been published for microbial communities associated with groundwater (Wilkins et al., 2009; Callister et al., 2010), insect gut (Burnum et al., 2011), ant fungal gardens (Aylward et al., 2012), and biofilms (Stuart et al., 2016), to our knowledge no demonstration utilizing this approach has yet been published for the soil microbiome.

As such, the objective our study was to evaluate how 1D, 2D offline and 2D online (modified from the original MuDPIT set-up) separation approaches, coupled with MS/MS, perform with regard to soil metaproteome depth of coverage given the same initial metaproteome complexity. To accomplish this, we compared these separation approaches for a soil metaproteome, and from a mock community consisting of 47 different microorganisms; where, proteins extracted from each member in pure culture were combined to generate mass ratios ranging from 1:10 to 1:100. The organisms selected include known members of the same genus and those of more distant taxon, and in whole represent a genetic database composed of 199,560 predicted proteins. For the, bulk soil samples were collected from the Konza Prairie Biological Station (KPBS), a long–term ecological research (LTER) site located in Eastern Kansas, USA. The soil microbiome has previously been genomically characterized for organism and functional diversity (White et al., 2016).

2. Materials and methods

2.1. Soil sample collection

Three subsamples (0–15 cm depth; minimum of 10 m apart) from a silty-loam soil (C:N 12, pH 6.5, 2% clay) at the KPBS were excavated with a shovel, roots and large rocks were removed and the soil was homogenized by manually mixing first in a 5 gallon bucket and then transferring the mixed soil to a 1 gallon sealable plastic bag. The subsamples were composited into one sample and immediately frozen under liquid nitrogen, then shipped to the Pacific Northwest National Laboratory where they were quickly thawed, sieved (4 mm) and aliquoted into 18–40 g portions in Falcon tubes. The samples were re-frozen under liquid nitrogen and stored at − 80 °C until processing.

2.2. Protein extraction from soil

The extraction of proteins from soil can often result in a low protein yield. In order to obtain a sufficient mass for our LC MS/MS approach evaluation, we extracted proteins from 60 g of collected Kansas prairie soil. Soil was quickly thawed and weighed into six 10 g aliquots (60 g total) in 50 mL methanol/chloroform compatible tubes (Genesee Scientific, San Diego, CA) along with 10 mL of 0.9–2.0 mm stainless steel beads, 0.1 mm zirconia beads and 0.1 mm garnet beads. All beads had previously been washed with chloroform and methanol and dried in a fume hood. Protein extraction occurred using a modified method of the Folch extraction (Folch et al., 1951; Nakayasu et al., 2016). Here, 4 mL of ice-cold ultrapure “Type 1” water (Millipore, Billerica, MA) was added to each sample and transferred to an ice bucket in a fume hood. Using a 25 mL glass serological pipette, ice-cold (− 20 °C) 2:1 chloroform:methanol (v/v) (Sigma-Aldrich, St. Louis, MO), was added to the sample in a 5:1 ratio over sample volume (20 mL) and vigorously mixed (by vortexing). The tubes were attached to a 50 mL tube vortex-attachment and horizontally mixed for 10 min at 4 °C and placed inside a − 80 °C freezer for 5 min. Using a probe sonicator (model FB505, Thermo Fisher Scientific, Waltham, MA) inside a fume hood, each sample was sonicated with a 6 mm probe (20 kHz fixed ultrasonic frequency) at 60% of the maximum amplitude for 30 s on ice, allowed to