

UC Berkeley

UC Berkeley Previously Published Works

Title

Genome-Resolved Proteomic Stable Isotope Probing of Soil Microbial Communities Using  $^{13}\text{CO}_2$  and  $^{13}\text{C}$ -Methanol

Permalink

<https://escholarship.org/uc/item/2d29z026>

Authors

Li, Zhou

Yao, Qiuming

Guo, Xuan

et al.

Publication Date

2019

DOI

10.3389/fmicb.2019.02706

Peer reviewed



# Genome-Resolved Proteomic Stable Isotope Probing of Soil Microbial Communities Using $^{13}\text{C}\text{O}_2$ and $^{13}\text{C}$ -Methanol

OPEN ACCESS

**Edited by:**

Phil B. Pope,  
Norwegian University of Life Sciences,  
Norway

**Reviewed by:**

Jeppe Lund Nielsen,  
Aalborg University, Denmark  
Robert Heyer,  
Otto von Guericke University  
Magdeburg, Germany

**\*Correspondence:**

Chongle Pan  
cpan@ou.edu

**† Present address:**

Qiuming Yao,  
Department of Pathology, Center  
for Cancer Research, Massachusetts  
General Hospital and Harvard Medical  
School, Boston, MA, United States  
Xuan Guo,  
Department of Computer Science  
and Engineering, University of North  
Texas, Denton, TX, United States

**Specialty section:**

This article was submitted to  
Systems Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 16 July 2019

**Accepted:** 08 November 2019

**Published:** 06 December 2019

**Citation:**

Li Z, Yao Q, Guo X,  
Crits-Christoph A, Mayes MA,  
Hervey WJ IV, Lebeis SL, Banfield JF,  
Hurst GB, Hettich RL and Pan C  
(2019) Genome-Resolved Proteomic  
Stable Isotope Probing of Soil  
Microbial Communities Using  $^{13}\text{C}\text{O}_2$   
and  $^{13}\text{C}$ -Methanol.  
*Front. Microbiol.* 10:2706.  
doi: 10.3389/fmicb.2019.02706

Zhou Li<sup>1,2,3</sup>, Qiuming Yao<sup>1†</sup>, Xuan Guo<sup>1†</sup>, Alexander Crits-Christoph<sup>4</sup>, Melanie A. Mayes<sup>5</sup>, William Judson Hervey IV<sup>6</sup>, Sarah L. Lebeis<sup>7</sup>, Jillian F. Banfield<sup>8,9</sup>, Gregory B. Hurst<sup>3</sup>, Robert L. Hettich<sup>2,3</sup> and Chongle Pan<sup>1,2,10\*</sup>

<sup>1</sup> Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN, United States, <sup>2</sup> Graduate School of Genome Science and Technology, The University of Tennessee, Knoxville, TN, United States, <sup>3</sup> Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, United States, <sup>4</sup> Department of Plant and Microbial Biology, University of California, Berkeley, Berkeley, CA, United States, <sup>5</sup> Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, United States, <sup>6</sup> Naval Research Laboratory, Center for Biomolecular Science and Engineering, Washington, DC, United States, <sup>7</sup> Department of Microbiology, The University of Tennessee, Knoxville, TN, United States, <sup>8</sup> Department of Earth and Planetary Science, University of California, Berkeley, Berkeley, CA, United States, <sup>9</sup> Department of Environmental Science, Policy, and Management, University of California, Berkeley, Berkeley, CA, United States, <sup>10</sup> School of Computer Science and Department of Microbiology and Plant Biology, University of Oklahoma, Norman, OK, United States

Stable isotope probing (SIP) enables tracking the nutrient flows from isotopically labeled substrates to specific microorganisms in microbial communities. In proteomic SIP, labeled proteins synthesized by the microbial consumers of labeled substrates are identified with a shotgun proteomics approach. Here, proteomic SIP was combined with targeted metagenomic binning to reconstruct metagenome-assembled genomes (MAGs) of the microorganisms producing labeled proteins. This approach was used to track carbon flows from  $^{13}\text{C}\text{O}_2$  to the rhizosphere communities of *Zea mays*, *Triticum aestivum*, and *Arabidopsis thaliana*. Rhizosphere microorganisms that assimilated plant-derived  $^{13}\text{C}$  were capable of metabolic and signaling interactions with their plant hosts, as shown by their MAGs containing genes for phytohormone modulation, quorum sensing, and transport and metabolism of nutrients typical of those found in root exudates. XoxF-type methanol dehydrogenases were among the most abundant proteins identified in the rhizosphere metaproteomes.  $^{13}\text{C}$ -methanol proteomic SIP was used to test the hypothesis that XoxF was used to metabolize and assimilate methanol in the rhizosphere. We detected 7  $^{13}\text{C}$ -labeled XoxF proteins and identified methylotrophic pathways in the MAGs of 8  $^{13}\text{C}$ -labeled microorganisms, which supported the hypothesis. These two studies demonstrated the capability of proteomic SIP for functional characterization of active microorganisms in complex microbial communities.

**Keywords:** stable isotope probing, metagenomic analyses, metaproteomic analysis, microbial ecology, rhizosphere

## INTRODUCTION

Plants release a substantial fraction of their photosynthetically fixed carbon from roots into rhizosphere as carbohydrates, amino acids, organic acids, and other compounds (Zhalnina et al., 2018). These carbon substrates can be consumed by microorganisms in the rhizosphere communities. In return, rhizosphere communities help plants extract nutrients from soil (Bulgarelli et al., 2013), modulate plant growth with phytohormones (Costacurta and Vanderleyden, 1995), and defend plants against soil-borne pathogens (Mendes et al., 2011). Studies of rhizosphere communities have shed light on the cycling of carbon and nutrients between plants and soil in many natural terrestrial ecosystems (Jones et al., 2009). However, microbe-plant interactions in the rhizosphere are still poorly understood owing to the immense complexity of rhizosphere communities and the difficulty of comprehensively cultivating rhizosphere microorganisms from diverse taxa (Bulgarelli et al., 2012; Lundberg et al., 2012; Lebeis et al., 2015).

Cultivation-independent stable isotope probing (SIP) has been used to track carbon flows from plants to specific members of rhizosphere communities (el Zahar Haichar et al., 2016). Growing under a  $^{13}\text{C}$  atmosphere, plants fix  $^{13}\text{C}$  and release  $^{13}\text{C}$ -labeled compounds into the rhizosphere. Microorganisms that directly assimilate these  $^{13}\text{C}$ -labeled compounds should produce  $^{13}\text{C}$ -labeled biomass, including DNA, RNA, and proteins. In previous studies (Lu and Conrad, 2005; Vandenkoornhuysen et al., 2007; el Zahar Haichar et al., 2008; Drigo et al., 2010),  $^{13}\text{C}$ -labeled DNAs and RNAs were isolated by density gradient ultracentrifugation and then amplicon-sequenced to identify the microbial taxa that assimilated  $^{13}\text{C}$  in rhizosphere communities. For example, a DNA SIP study showed that bacteria in the order *Sphingobacteriales* and the genus *Myxococcus* assimilated plant-derived carbon in the rhizosphere of wheat, maize, rapeseed, and medicago (el Zahar Haichar et al., 2008). RNA SIP was used to uncover the carbon transfer from rice to a rhizosphere methanogen belonging to Rice Cluster I Archaea (Lu and Conrad, 2005) and from white clover to arbuscular mycorrhizal fungi (Vandenkoornhuysen et al., 2007). Recently, a genome-resolved metagenomic approach was implemented in a DNA SIP study which recovered a complete genome for *Saccharibacteria* (TM7) from the rhizosphere of wild oats (Starr et al., 2018).

In comparison with DNA SIP and RNA SIP, proteomic SIP has the advantages of accurate atom% estimation and a low detection limit for the labeling levels (as low as 2% for  $^{13}\text{C}$  labeling) (Pan et al., 2011; Justice et al., 2014; Bryson et al., 2016, 2017; Marlow et al., 2016), because it is based on precise measurement of the mass shifts of labeled peptides. However, proteomic SIP provides less direct information than DNA SIP or RNA SIP for taxonomy identification of the labeled microorganisms. Binning of metagenome-assembled genomes (MAGs) has been shown in previous studies to be a powerful cultivation-independent method to recover novel genomes from complex metagenomics (Crits-Christoph et al., 2018; Delmont et al., 2018). Genome-resolved metagenomics and proteomics have been combined to survey the protein expression profiles of microbial communities

(Ram et al., 2005). Protein-based SIP was used with a binned metagenome for characterization of  $^{13}\text{C}$ -acetate turnover in anaerobic digestion communities (Mosbaek et al., 2016). In this study, proteomic SIP was coupled with targeted metagenomic binning to identify  $^{13}\text{C}$ -labeled proteins and reconstruct MAGs of  $^{13}\text{C}$ -labeled microorganisms in a  $^{13}\text{CO}_2$  rhizosphere SIP experiment. Proteomic SIP and metagenomics also provided a cultivation-independent method to validating the substrates of enzymes used by microbial communities for carbon uptake. This method was used in a  $^{13}\text{C}$ -methanol SIP experiment to support methanol as a putative substrate of the abundant XoxF-type methanol dehydrogenase in the rhizosphere.

## RESULTS

### $^{13}\text{CO}_2$ Stable Isotope Probing, Metagenomic Sequencing, and Metaproteomics Analyses of Rhizosphere Communities

Three model plants, including *Z. mays*, *T. aestivum*, and *A. thaliana*, were cultivated in pots filled with the same initial soil collected from the field (4 pots for each plant species for a total of 12 pots) (Supplementary Figure 1). After 29 days of growth in the normal atmosphere, the plants were moved to grow in a labeling chamber with a  $^{13}\text{CO}_2$  atmosphere (99% atom%  $^{13}\text{C}$ ). Duplicate pots from each plant species were harvested after 3 days of  $^{13}\text{CO}_2$  labeling to generate timepoint-1 (T1) rhizosphere samples. The remaining duplicate pots from each plant species were harvested after 8 days of  $^{13}\text{CO}_2$  labeling to generate timepoint-2 (T2) rhizosphere samples. The 12 rhizosphere soil samples (3 plant species X 2 timepoints X 2 replicates) and 2 initial soil samples as the controls were measured with metagenomics and metaproteomics (Butterfield et al., 2016; Yao et al., 2018).

Because the 14 samples originated from the same initial soil, the 203 Gbp of metagenomic sequencing data (a total of 1.4 billion  $2 \times 150$ -bp reads) from all samples were merged and co-assembled into a composite metagenome (Supplementary Table 1). The composite metagenome contained 6.5 Gbp of scaffolds longer than 1 kbp with an N50 score of 2,331 bp. It encoded a total of 8.4 million predicted protein-coding genes. The metaproteomics measurements identified an average of 5,287 unlabeled microbial proteins or protein groups per soil sample with protein false discovery rates (FDRs) all less than 2% (Supplementary Table 1).

Proteomic SIP searches identified 19  $^{13}\text{C}$ -labeled microbial proteins or protein groups with  $>10\%$   $^{13}\text{C}$  enrichment and no decoy  $^{13}\text{C}$ -labeled protein (Supplementary Table 2), which indicated a low FDR in the identification of  $^{13}\text{C}$ -labeled proteins. Furthermore, no  $^{13}\text{C}$ -labeled protein was identified from the 2 initial soil samples, which also supported the low FDR of the  $^{13}\text{C}$ -labeled protein identification in this study. This was consistent with the low FDR of proteomic SIP searches in our previous studies (Pan et al., 2011; Justice et al., 2014; Bryson et al., 2016, 2017; Marlow et al., 2016). The average  $^{13}\text{C}$  atom% of  $^{13}\text{C}$ -labeled

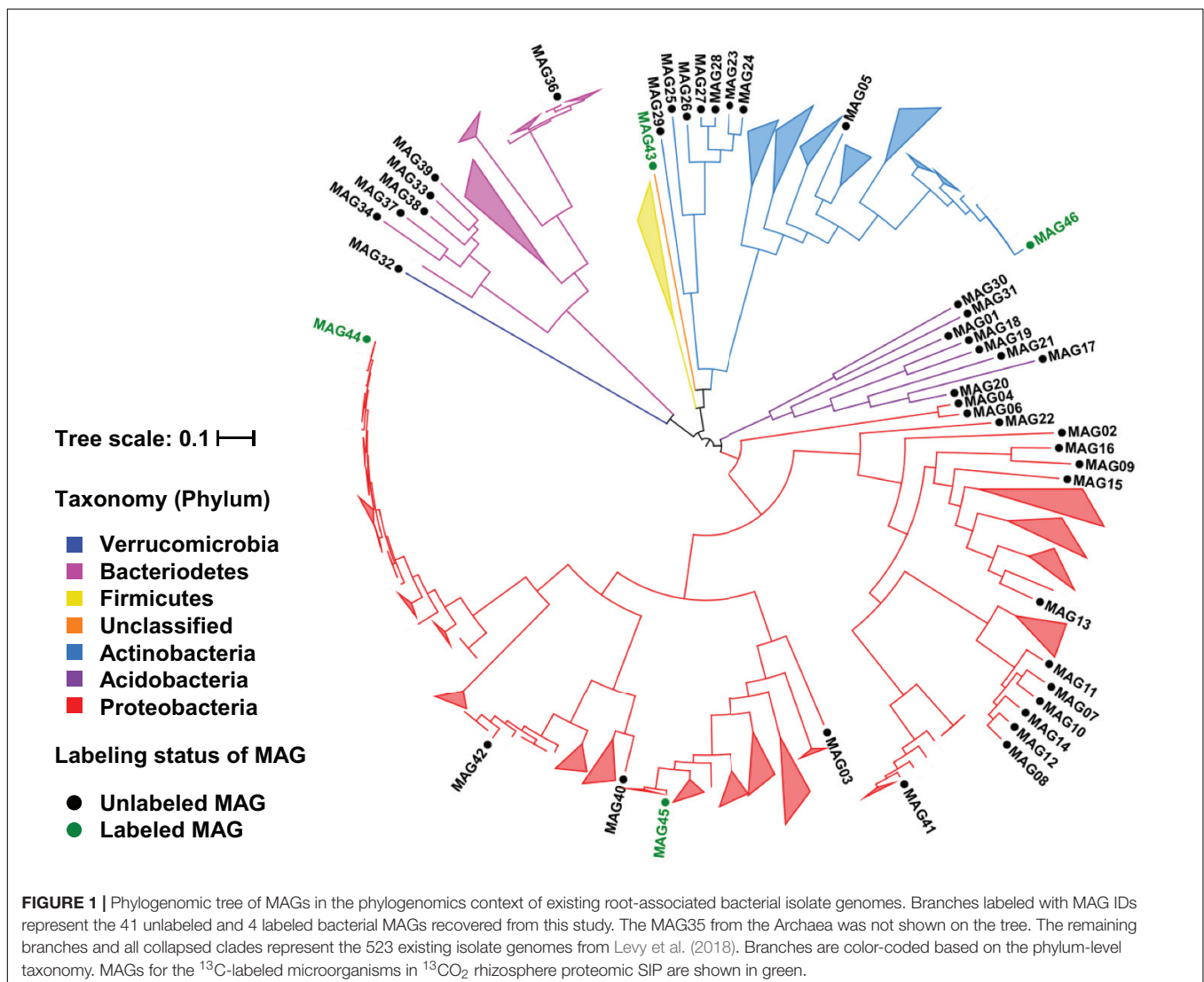
proteins from the T2 rhizosphere samples was higher than that from the T1 rhizosphere samples ( $p$  value = 0.00262, Mann-Whitney  $U$  test). Most of the identified  $^{13}\text{C}$ -labeled proteins were abundant housekeeping proteins, such as ribosomal proteins, chaperones, and glycolysis enzymes.

## Reconstruction of MAGs for the Unlabeled and $^{13}\text{C}$ -Labeled Rhizosphere Microorganisms

The abundance series of scaffolds in the composite metagenome assembly across the 14 soil samples were used to bin scaffolds into MAGs. The untargeted binning produced 41 medium-quality bacterial MAGs and 1 medium-quality archaeal MAG. All these MAGs had more than 70% genome completeness and less than 10% genome contamination (Supplementary Table 3). The 41 bacterial MAGs were analyzed in the phylogenomics context of 523 isolate genomes of root-associated bacteria from a previous study (Levy et al., 2018) (Figure 1). The 523

existing isolate genomes all belonged to four major bacterial phyla, including *Proteobacteria*, *Actinobacteria*, *Bacteroidetes*, and *Firmicutes*. Consistently, 32 of the 41 bacterial MAGs belonged to *Proteobacteria* ( $n = 18$ ), *Actinobacteria* ( $n = 8$ ), and *Bacteroidetes* ( $n = 6$ ). Most of the MAGs formed distinct clades separate from the existing isolate genomes, including a clade of 7 MAGs in the *Actinobacteria* phylum, a clade of 5 MAGs in the *Sphingobacteriales* order of the *Bacteroidetes* phylum, and a clade of 6 MAGs in the *Sphingomonadales* order of the *Proteobacteria* phylum. Furthermore, 10 MAGs were recovered from phyla under-represented by the existing root-associated isolate genomes, including 8 bacterial MAGs from the *Acidobacteria* phylum, 1 bacterial MAG from the *Verrucomicrobia* phylum, and 1 archaeal MAG from the *Thaumarchaeota* phylum. Thus, the MAGs from our untargeted binning expanded the phylogenomic coverage and genomic diversity of root-associated bacteria and archaea.

Metagenome-assembled genomes of the rhizosphere microorganisms that produced  $^{13}\text{C}$ -labeled proteins were



recovered using a targeted binning approach. The scaffolds containing the genes for  $^{13}\text{C}$ -labeled unique proteins were used as the binning targets to cluster other scaffolds around. Each obtained MAG was required to have a minimum of two independent identifications of  $^{13}\text{C}$ -labeled unique proteins in order to increase the confidence on the  $^{13}\text{C}$ -labeling status of the microorganisms. The contamination levels of the MAGs were controlled below 5%. Out of the 28  $^{13}\text{C}$ -labeled protein identifications, 14 were binned into four MAGs, including the medium-quality MAG45 with genome completeness at 54% and three low-quality MAGs (MAG43, MAG44, and MAG46) with genome completeness between 21% and 42% (Table 1). The  $^{13}\text{C}$  atom% enrichment of the binned protein identifications ranged between 15% and 46%. MAG44 belonged to the *Pseudomonas* genus, MAG45 to the Oxalobacteraceae family, MAG46 to the *Arthrobacter* genus, and MAG43 to an unclassified taxon related to the Chloroflexi-related group (CHLX; Anantharaman et al., 2016). MAG44, MAG45, and MAG46 were closely related to other root-associated bacterial isolates (Figure 1).

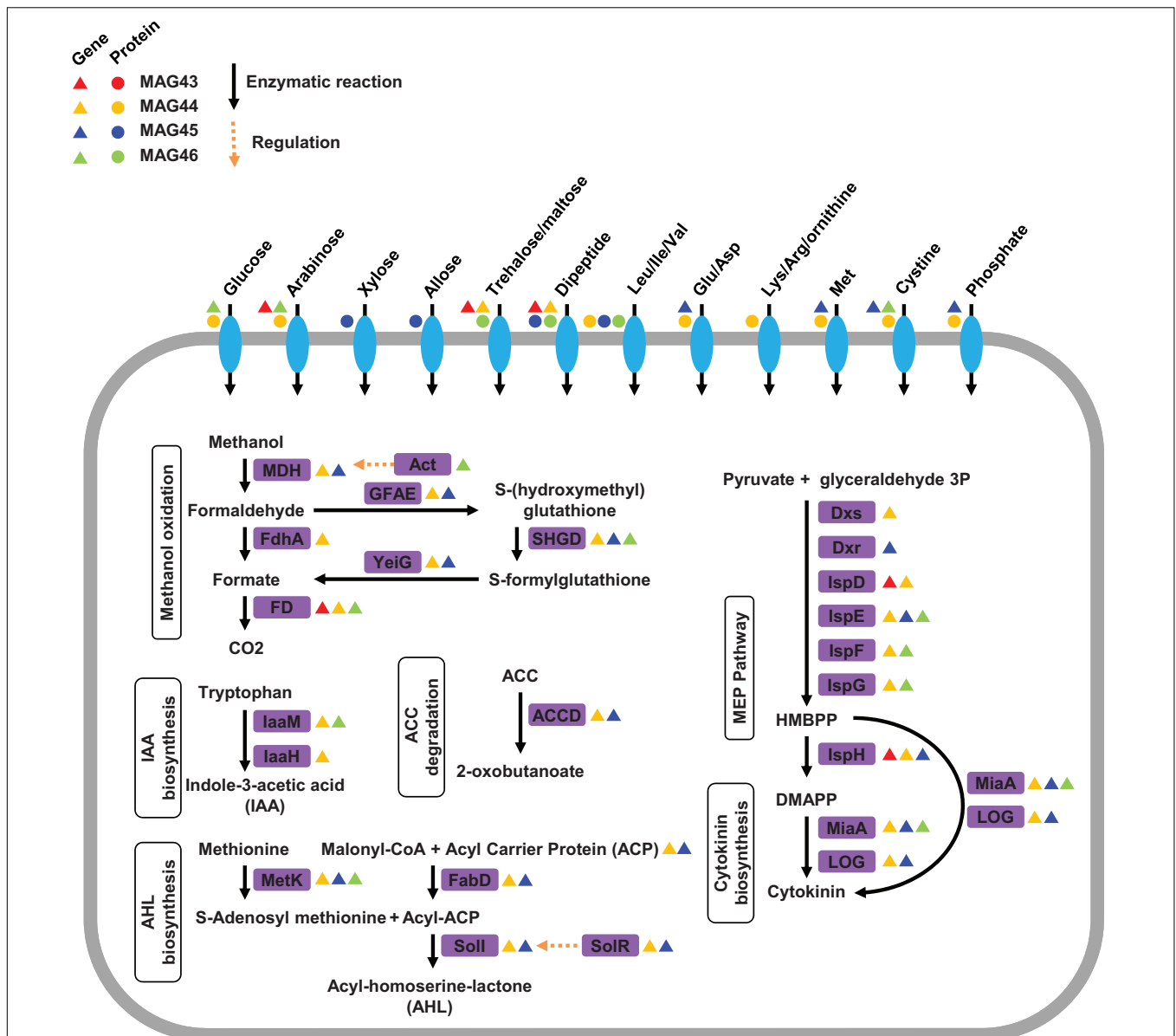
## Genetic Potential and Expressed Functions in the $^{13}\text{C}$ -Labeled Rhizosphere Microorganisms

The genetic potential and expressed functions of the four  $^{13}\text{C}$ -labeled rhizosphere microorganisms were inferred based on their genes contained in the MAGs and their unlabeled proteins identified by the regular label-free proteomics searches. These microorganisms encoded genes and had the capability to produce proteins for a variety of transporters of sugars and amino acids (Figure 2). We identified proteins for transporters of glucose and arabinose in MAG44, xylose and allose transporters in MAG45, and a trehalose/maltose transporter in MAG46. We also identified proteins for transporters of dipeptides and branched-chain amino acids (Leu/Ile/Val) in MAG45 and MAG46 and transporters of a variety of amino acids in MAG44. Identification of these transporter proteins in the  $^{13}\text{C}$ -labeled microorganisms suggested their extraction of sugars and amino acids from the rhizosphere soils. These MAGs also encoded genes for methanol oxidation through formaldehyde and formate (Chistoserdova et al., 2009) (Figure 2). The sugars, amino acids, and/or methanol may carry plant-derived  $^{13}\text{C}$  to these microorganisms.

Rhizosphere microorganisms can modulate the physiology of their plant hosts by producing phytohormones, such as cytokinin and auxin. MAG44 and MAG45 encoded the genes for tRNA: isopentenyltransferase (MiaA) and cytokinin-specific phosphoribohydrolase 'Lonely guy' (LOG) in the cytokinin synthesis pathway (Frébert et al., 2011) (Figure 2). MAG46 also encoded the gene for MiaA. This was consistent with a previous rhizosphere study (Starr et al., 2018). The substrates for cytokinin synthesis can be produced through the 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway (Frébert et al., 2011). Out of the seven genes in the MEP pathway, 6 were found in MAG44, 3 in MAG45 and MAG46, and 2 in MAG43. The absence of the other genes in this pathway may be attributed to the low completeness of the MAGs.

**TABLE 1** |  $^{13}\text{C}$ -labeled proteins and microorganisms from  $^{13}\text{CO}_2$  rhizosphere proteomic SIP.

Labeled protein (Protein ID)	Plant host	Time point	$^{13}\text{C}$ atom%	MAG ID	Completeness	Contamination	Taxonomy	Genome size (Mbp)	# of genes
Glycerialdehyde-3-phosphate dehydrogenase (scaff_0000095019_7)	<i>Z. mays</i>	T1	24%	MAG43	21%	3.6%	Unclassified	0.7	692
	<i>Z. mays</i>	T2	15%						
	<i>T. aestivum</i>	T2	30%						
Cold shock protein CapA (scaff_0000016092_2)	<i>Z. mays</i>	T1	18%	MAG44	42%	1.8%	<i>Pseudomonas</i>	5.5	5207
Chaperone DnaK (scaff_0000160764_4)	<i>Z. mays</i>	T1	16%						
Protein with unknown function (scaff_0000030713_6)	<i>Z. mays</i>	T2	18%	MAG45	54%	0.3%	Oxalobacteraceae	2.9	2853
Colicin I receptor (scaff_0000049197_5)	<i>Z. mays</i>	T2	17%						
Enolase (scaff_0000017170_6)	<i>T. aestivum</i>	T2	20%	MAG46	29%	0%	<i>Arthrobacter</i>	2.3	2252
	<i>T. aestivum</i>	T2	28%						
	<i>T. aestivum</i>	T1	16%						
	<i>T. aestivum</i>	T2	26%						
	<i>T. aestivum</i>	T2	46%						
	<i>T. aestivum</i>	T2	34%						
Protein with unknown function (scaff_0000159039_6)	<i>Z. mays</i>	T1	29%						
Chaperonin GroEL (scaff_0000063839_3)									



**FIGURE 2 |** Phytohormone modulation, nutrient transportation, quorum sensing, and methanol metabolism in the four  $^{13}\text{C}$ -labeled microorganisms. Enzymes in purple boxes are labeled with triangles for encoded genes and circles for expressed proteins in these microorganisms. The triangles and circles are color-coded for each microorganism. MDH: NAD-dependent methanol dehydrogenase, Act: methanol dehydrogenase activator, FdhA: glutathione-independent formaldehyde dehydrogenase, FD: formate dehydrogenase, GFAE: glutathione-dependent formaldehyde-activating enzyme, SHGD: S-(hydroxymethyl)glutathione dehydrogenase, YeiG: S-formylglutathione hydrolase, IaaM: tryptophan 2-monooxygenase, IaaH: indole acetamide hydrolase, MetK: methionine adenosyltransferase, ACCD: 1-aminocyclopropane-1-carboxylate deaminase, FabD: malonyl CoA-acyl carrier protein transacylase, SoI: acyl-homoserine-lactone synthase, SoIR: transcriptional activator/quorum-sensing receptor, Dxs: 1-deoxy-D-xylulose 5-phosphate synthase, Dxr: 2-C-methyl-D-erythritol 4-phosphate synthase, IspD: 2-C-methyl-D-erythritol 4-phosphate cytidyltransferase, IspE: 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase, IspF: 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase, IspG: 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase, IspH: 4-hydroxy-3-methylbut-2-enyl diphosphate reductase, MiaA: tRNA isopentenyltransferase, LOG: cytokinin-specific phosphoribohydrolase 'Lonely guy', HMBPP: (E)-4-Hydroxy-3-methyl-but-2-enyl pyrophosphate, DMAPP: dimethylallyl pyrophosphate.

Indole-3-acetic acid (IAA) is a major type of auxin and can be produced from tryptophan using a tryptophan monooxygenase (IaaM) and an indole-3-acetamide hydrolase (IaaH) (Patten and Glick, 1996). A gene encoding IaaM was found in MAG44 and MAG46 (Figure 2). In MAG44, the *iaaM* gene was adjacent to a

carbon-nitrogen hydrolyase gene (E.C. 3.5.-.-), which may carry out the function of IaaH (E.C. 3.5.1.-). MAG44 also encoded genes for a two-component system GacS/GacA and a sigma factor RpoS, which can regulate the transcription of IaaM and IaaH (Bulgarelli et al., 2013).

Plants can synthesize ethylene to regulate their growth (Morgan and Drew, 1997). 1-aminocyclopropane-1-carboxylate (ACC) is a precursor for ethylene synthesis (Glick et al., 2007). ACC can be exuded by plant roots and degraded into  $\alpha$ -ketobutyrate and ammonia by the ACC deaminase (ACCD) in rhizosphere microorganisms (Bulgarelli et al., 2013). MAG44 and MAG45 both possessed a gene for the ACCD (**Figure 2**). The degradation of ACC by MAG44 and MAG45 may lower the ethylene level inside their plant hosts to attenuate the ethylene-dependent growth inhibition (Bulgarelli et al., 2013).

Rhizosphere microorganisms can use acyl-homoserine-lactone (AHL) as a quorum-sensing autoinducer (DeAngelis et al., 2008). Both MAG44 and MAG45 contained the gene for an AHL synthase, *SolI*. The *solI* gene in both MAGs was adjacent to a *solR* gene which encodes a AHL receptor that can activate the transcription of *solI* (Lerat and Moran, 2004). Both MAG44 and MAG45 also encoded the genes for S-adenosylmethionine synthetase (MetK), an acyl-carrier protein (ACP), and malonyl CoA-acyl carrier protein transacylase (FabD) for AHL synthesis (**Figure 2**). The presence of the AHL synthesis genes and the AHL receptor gene in MAG44 and MAG45 indicated the quorum sensing capability in these two microorganisms.

## Protein Expression Profiles of the Rhizosphere Communities for Carbon Metabolism

Besides the  $^{13}\text{C}$ -labeled proteins, metaproteomics identified between 4,165 and 7,258 unlabeled proteins or protein groups from the 12 rhizosphere communities, which were used to characterize the expressed molecular functions at the community level. Protein identifications were aggregated into biological functions represented by gene ontology (GO) terms and EC numbers (Yao et al., 2018). Out of 5,200 GO terms and 2,569 EC numbers encoded in the metagenome, 1,352 GO terms and 509 EC numbers had protein expression detected and quantified by metaproteomics (**Supplementary Figure 2**).

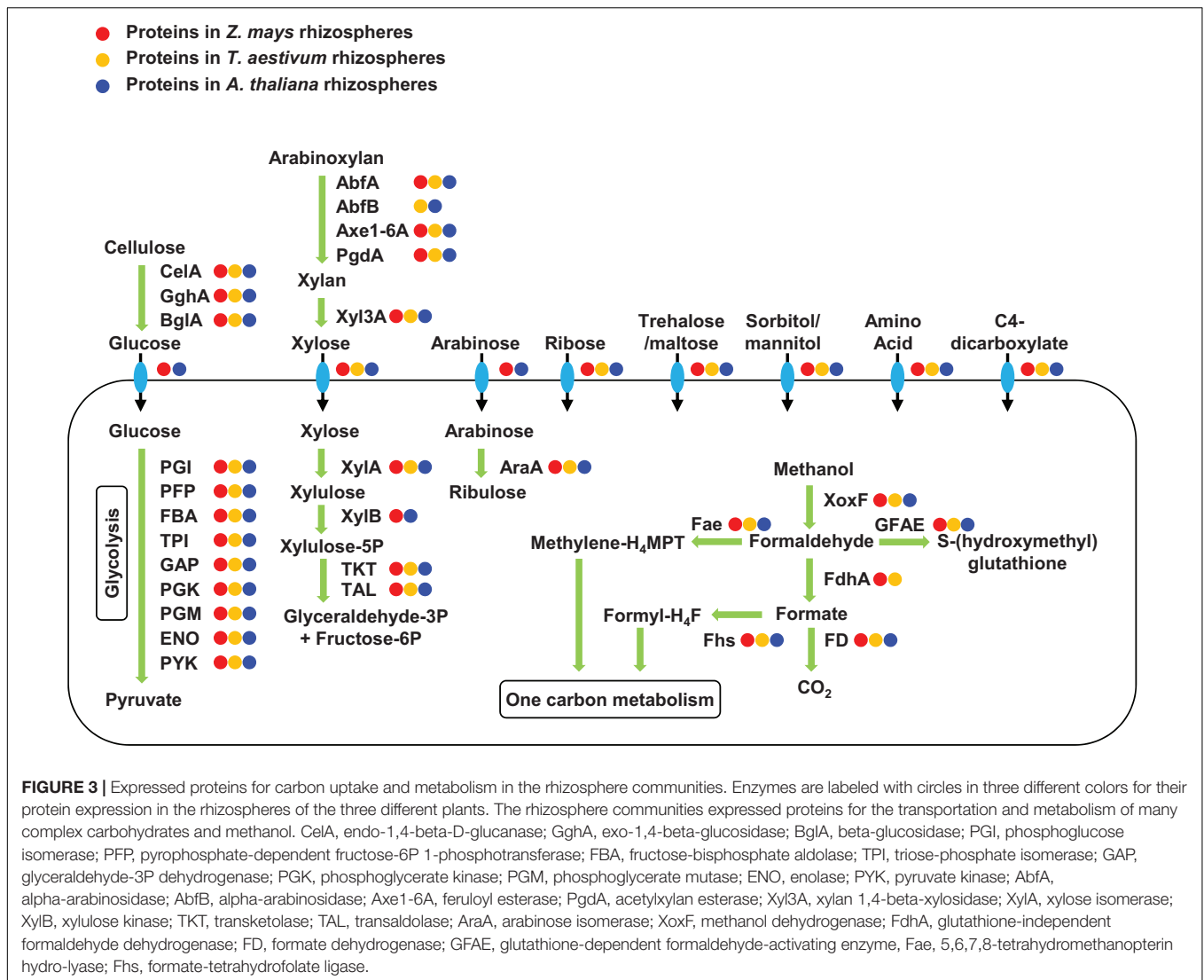
The rhizosphere communities expressed proteins for degrading cellulose, xylan, and arabinoxylan (**Figure 3**). Cellulose can be deconstructed to glucose by the identified endo-1,4-beta-D-glucanase (CelA), exo-1,4-beta-glucosidase (GghA), and beta-glucosidase (BglA) in the rhizosphere metaproteomes. Glucose can be transported into cells by the expressed glucose porin and glucose ABC transporter and be degraded through an active glycolysis pathway. Arabinoxylan can be debranched to xylan by the identified alpha-arabinosidase (AbfA and AbfB), feruloyl esterase (Axe1-6A), and acetylxylan esterase (PdfA). Xylan can be depolymerized to xylose by the expressed xylan 1,4-beta-xylosidase (Xyl3A). Xylose can be transported into cell by the expressed transporters, XylT and XylF, and be degraded by the identified xylose isomerase (XylA), xylulose kinase (XylB), and transketolase and transaldolase in the pentose phosphate pathway. Metaproteomics also identified transporters of other compounds, including arabinose, ribose, trehalose/maltose, sorbitol/mannitol, C4-dicarboxylate, and many amino acids. These compounds extracted from the rhizosphere soil may be important carbon substrates for the rhizosphere communities.

Metaproteomics identified 68 XoxF-type methanol dehydrogenases out of a total of 567 XoxF genes encoded in the metagenomes. The aggregate abundance of the identified XoxF proteins accounted for an average of 11% of the total abundance of all identified proteins in the metaproteomes. The phylogenetics of the 68 identified XoxF proteins was analyzed using 146 reference XoxF sequences (Taubert et al., 2015), which categorized the identified XoxF proteins into 4 groups, namely XoxF1, XoxF2, XoxF3, and XoxF5 (**Supplementary Figure 3**). Metaproteomics also identified glutathione-independent formaldehyde dehydrogenase (FdhA) and formate dehydrogenase (FD) for oxidation of formaldehyde and formate, respectively, in the methanol oxidation pathway to  $\text{CO}_2$  (**Figure 3**). Formaldehyde and formate can also be converted to methylene-tetrahydromethanopterin (methylene- $\text{H}_4\text{MPT}$ ) and formyl-tetrahydrofolate (formyl- $\text{H}_4\text{F}$ ) for the one-carbon metabolism by the identified  $\text{H}_4\text{MPT}$  hydro-lyase (Fae) and formate- $\text{H}_4\text{F}$  ligase (Fhs), respectively (**Figure 3**).

## Testing of the Methylotrophic Metabolism With $^{13}\text{C}$ -Methanol Proteomic SIP

Because of the existence of diverse XoxF proteins with high abundance across the rhizosphere and initial soils, we hypothesized that methanol was an important carbon source for many of these soil microorganisms via XoxF-based methylotrophic pathways. We tested this hypothesis using a  $^{13}\text{C}$ -methanol proteomic SIP experiment. A fresh sample of the initial soil was collected and  $\sim 6.1$  micromoles of  $^{13}\text{C}$ -methanol per gram of soil was added daily into six replicates of the fresh initial soil. A triplicate of samples was collected after 3 days of  $^{13}\text{C}$ -methanol addition and incubation and another triplicate was collected after 8 days (**Supplementary Figure 4**). Proteomic SIP searches detected 22  $^{13}\text{C}$ -labeled microbial proteins or protein groups from 48 protein identifications in the soil samples amended with  $^{13}\text{C}$ -methanol (**Supplementary Table 4**). No decoy  $^{13}\text{C}$ -labeled protein was identified in any sample. This indicated that soil microorganisms incorporated carbon from methanol into their proteomes. Eighteen of these 48  $^{13}\text{C}$ -labeled protein identifications were XoxF proteins, suggesting that the addition of methanol in the soils stimulated the biosynthesis of new XoxF proteins in the microorganisms that assimilated methanol. In addition, we identified a  $^{13}\text{C}$ -labeled membrane-bound aldehyde dehydrogenase which can oxidize formaldehyde to formate and a  $^{13}\text{C}$ -labeled  $\text{H}_4\text{MPT}$  hydro-lyase for converting formaldehyde to methylene- $\text{H}_4\text{MPT}$  following the methanol oxidation.

Scaffolds containing these  $^{13}\text{C}$ -labeled proteins were used as targets for metagenomic binning, which generated one medium-quality MAGs and seven low-quality MAGs with genome contamination below 5% (**Table 2**). Each of these MAGs had  $^{13}\text{C}$ -labeled protein identifications from at least two replicates in a  $^{13}\text{C}$ -methanol SIP incubation triplicate. Four of the eight MAGs contained a *xoxF* gene (**Figure 4**). Among these four MAGs, three produced  $^{13}\text{C}$ -labeled XoxF proteins and the remaining one expressed an unlabeled XoxF protein.



**FIGURE 3 |** Expressed proteins for carbon uptake and metabolism in the rhizosphere communities. Enzymes are labeled with circles in three different colors for their protein expression in the rhizospheres of the three different plants. The rhizosphere communities expressed proteins for the transportation and metabolism of many complex carbohydrates and methanol. CelA, endo-1,4-beta-D-glucanase; GghA, exo-1,4-beta-glucosidase; BglA, beta-glucosidase; PGI, phosphoglucose isomerase; PFP, pyrophosphate-dependent fructose-6P 1-phosphotransferase; FBA, fructose-bisphosphate aldolase; TPI, triose-phosphate isomerase; GAP, glyceraldehyde-3P dehydrogenase; PGK, phosphoglycerate kinase; PGM, phosphoglycerate mutase; ENO, enolase; PYK, pyruvate kinase; AbfA, alpha-arabinosidase; AbfB, alpha-arabinosidase; Axe1-6A, feruloyl esterase; PgdA, acetylxylan esterase; Xyl3A, xylan 1,4-beta-xylosidase; XylA, xylose isomerase; XylB, xylulose kinase; TKT, transketolase; TAL, transaldolase; AraA, arabinose isomerase; XoxF, methanol dehydrogenase; FdhA, glutathione-independent formaldehyde dehydrogenase; FD, formate dehydrogenase; GFAE, glutathione-dependent formaldehyde-activating enzyme, Fae, 5,6,7,8-tetrahydromethanopterin hydro-lyase; Fhs, formate-tetrahydrofolate ligase.

All 8  $^{13}\text{C}$ -labeled MAGs encoded many genes involved in the oxidation of methanol to formaldehyde and formate and the carbon assimilation via the serine cycle (Figure 4). For example, the MAG53 encoded genes for methanol oxidation (*xoxF*), a two-component system involved in regulating expression of methanol dehydrogenase (*moxX* and *moxY*) (Harms et al., 1993),  $\text{H}_4\text{MPT}$ -dependent formaldehyde oxidation, formate oxidation, and serine cycle for carbon assimilation (serine-glyoxylate transaminase, hydroxypyruvate reductase, glycerate 3-kinase, malate dehydrogenase, malyl-CoA synthetase, and malyl-CoA lyase) (Chistoserdova et al., 2009; Chistoserdova, 2011). Thus, the MAGs of these  $^{13}\text{C}$ -labeled microorganisms indicated their genetic potential for *XoxF*-based methylotrophic metabolism.

## DISCUSSION

In this study, proteomic SIP was combined with targeted metagenomic binning in two  $^{13}\text{C}$  SIP experiments on complex

soil communities. Proteomic SIP was first used to identify  $^{13}\text{C}$ -labeled proteins with a low FDR. Then, targeted binning based on the abundance series of scaffolds across 14 samples generated MAGs for the microorganisms that produced the identified  $^{13}\text{C}$ -labeled proteins. Finally, the genetic potential and expressed functions of these microorganisms were inferred from their MAGs and identified proteins. In comparison with other SIP approaches, proteomic SIP was able to identify specific  $^{13}\text{C}$ -labeled microorganisms in a community and simultaneously reveal their partial genomes and proteomes.

This approach revealed four microorganisms that synthesized  $^{13}\text{C}$ -labeled proteins in the rhizosphere of plants growing under a  $^{13}\text{CO}_2$  atmosphere. The  $^{13}\text{C}$ -labeled microbial proteins indicated the transfer of photosynthetically fixed  $^{13}\text{C}$  from the plant hosts to these rhizosphere microorganisms. The  $^{13}\text{C}$ -labeled microorganisms expressed proteins for transport of carbohydrates and amino acids, which are known components of plant root exudates (Zhalnina et al., 2018). These compounds may be the  $^{13}\text{C}$  intermediates between the plant hosts and these



**TABLE 2** |  $^{13}\text{C}$ -labeled proteins and microorganisms from  $^{13}\text{C}$ -methanol proteomic SIP.

Labeled protein (Protein ID)	Day-3 ( $^{13}\text{C}$ Atom%)			Day-8 ( $^{13}\text{C}$ Atom%)			MAG ID	Completeness	Contamination	Taxonomy	Genome size (Mbp)	# of Genes
	Rep 1	Rep 2	Rep 3	Rep 1	Rep 2	Rep 3						
	Threonine-tRNA ligase (scaff_000000132_40)	97	98	98	98	97						
XoxF-type methanol dehydrogenase (scaff_0000127167_4)	55	55	59	72	68	59	MAG48	37%	0%	Rhizobiales	2.3	2418
XoxF-type methanol dehydrogenase (scaff_0000032941_13)	95	98	94	-	-	96	MAG49	17%	2%	Alphaproteobacteria	1.4	1477
SOS ribosomal protein L5 (scaff_0000001175_15)	93	-	93	-	-	-	MAG50	48%	5%	Proteobacteria	3.1	2996
XoxF-type methanol dehydrogenase (scaff_0000138535_1)	-	-	-	96	95	95	MAG51	8%	0%	Proteobacteria	0.3	300
Transaldolase (scaff_0000073426_7)	-	-	-	95	-	94	MAG52	23%	3%	Unclassified	1.4	1374
Protein with unknown function (scaff_0000002428_1)	-	-	-	84	-	84	MAG53	66%	4%	Rhizobiales	2.3	2354
Malate dehydrogenase (scaff_0000012663_3)	-	-	-	-	58	67	MAG54	31%	3%	Bradyrhizobiaceae	1.8	1781

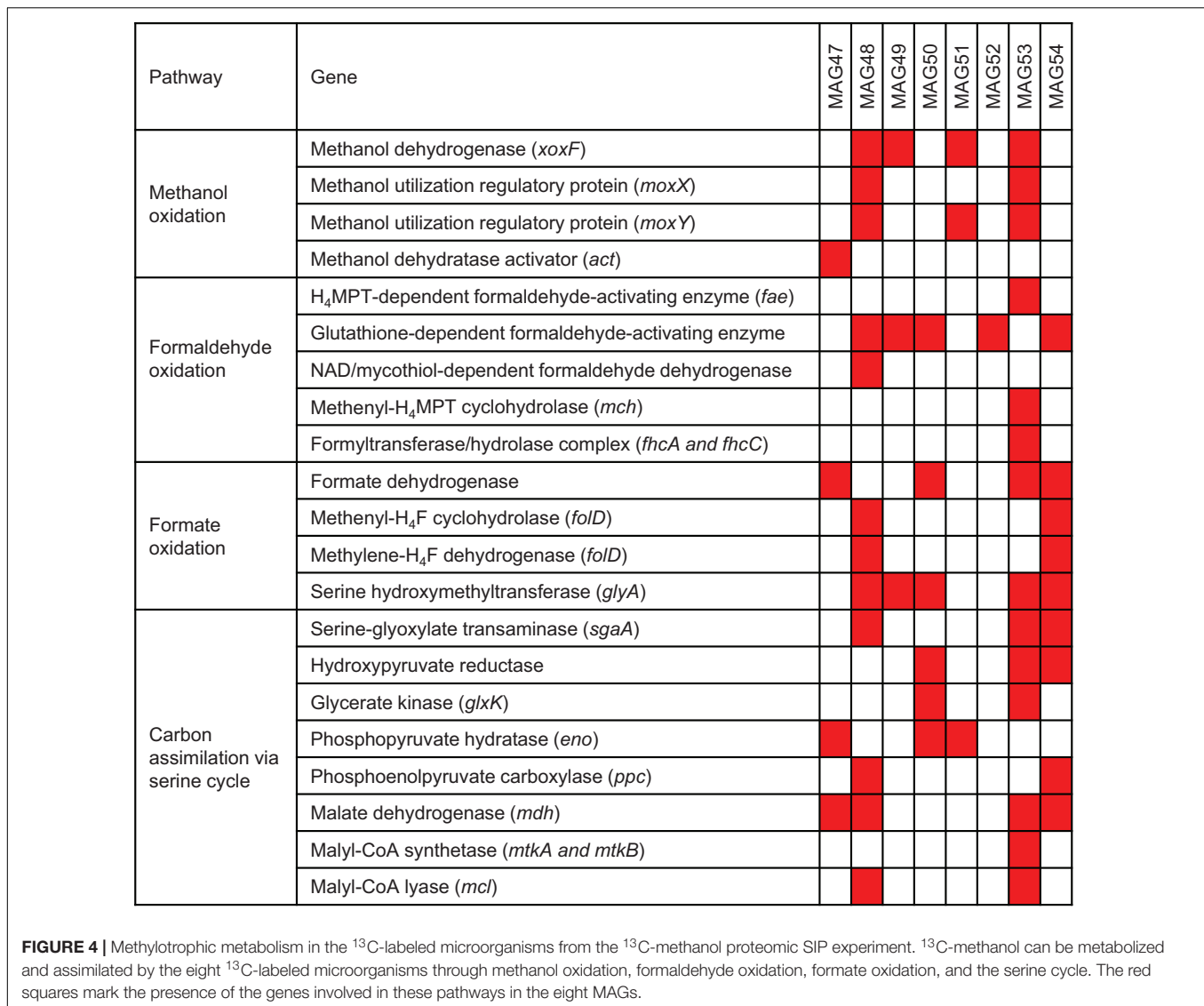
$^{13}\text{C}$ -labeled microorganisms. Besides the metabolic coupling, these  $^{13}\text{C}$ -labeled microorganisms had the genetic potential for extensive signaling interaction with the plant hosts by modulating plant hormones. In particular, the MAG44 encoded the genes involved in synthesizing cytokinin and auxin and suppressing ethylene via ACCD. This suggested that the  $^{13}\text{C}$ -labeled microorganism for MAG 44 can potentially modulate all these three plant hormones together to fine-tune the plant metabolic activities. Overall, our results supported many metabolic activities of rhizosphere microorganisms described in previous studies (Tien et al., 1979; Costacurta and Vanderleyden, 1995; Starr et al., 2018).

XoxF-type methanol dehydrogenase was found to be one of the most abundant proteins expressed by microbial communities in sea water (Sowell et al., 2011), grassland soil (Butterfield et al., 2016; Diamond et al., 2018), and phyllosphere of soybean, clover, and *Arabidopsis thaliana* (Delmotte et al., 2009), and in the rhizosphere soils of this study. The methanol oxidation activity of XoxF from pure cultures has been confirmed in laboratory (Schmidt et al., 2010). In rhizosphere soil, methanol may be released from plant roots as a by-product of pectin demethylation and lignin degradation (Galbally and Kirstine, 2002). In this study, proteomic SIP and metagenomics was used to establish a direct connection between the methanol addition and the biosynthesis of new XoxF proteins in methanol-utilizing microorganisms. The MAGs of these  $^{13}\text{C}$ -labeled methanol-utilizing microorganisms suggested that they can assimilate  $^{13}\text{C}$ -methanol via methanol oxidation, formaldehyde oxidation, formate oxidation, and the serine cycle. Proteomic SIP provided much stronger support for the methanol oxidation activity of XoxF than a potential comparative metaproteomics measurement that can only associate the substrate addition with protein abundance changes with or without new protein synthesis in organisms that may or may not directly utilize the added substrate. The  $^{13}\text{C}$ -methanol SIP experiment showed that proteomic SIP and metagenomics can link the consumption of a certain substrate with new biosynthesis of specific enzymes in organisms that have assimilated the substrate.

## MATERIALS AND METHODS

### Plant Growth and $^{13}\text{CO}_2$ Labeling

Twelve 0.25-L pots were filled with 330 g of a moistened loamy sand soil collected from a biological *Lolium perenne* grass field (51°59'28.0"N 5°39'43.1"E; upper 0–10 cm). The soil was sieved (5 mm) and subsequently adjusted to about 60% of field capacity (16%, w/w) by addition of deionized water. At day 0, seeds of *Zea mays* L. cv 'Yukon Chief' and *Triticum aestivum* L. cv 'Baldus' were germinated on wet tissue paper in petri dishes and transferred to the pots at Day 6 with two seedlings per pot. Ca. 30 seeds of *Arabidopsis thaliana* (L.) Heynh. cv 'Columbia-O' were directly sown in the pots. The plants were cultured in IsoLife's (Wageningen, The Netherlands) stable isotope labeling facility under the following conditions: temperatures: 18/14°C (day/night), relative humidity: 75% (continuously), irradiance: 600  $\mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$  (PAR) with a day/night rhythm of 12/12 h,



and CO<sub>2</sub> level: 400 ppm. Nutrients were given as a half-strength Hoagland's solution with corresponding N content of 10, 20, and 30 mg N per pot for *A. thaliana*, *T. aestivum*, and *Z. mays*, respectively. This fertilizer solution was given twice during the experiment, on Day 20, and on Day 33. From Day 30 till Day 38, the plants were labeled with 99 <sup>13</sup>C atom% <sup>13</sup>CO<sub>2</sub>. During the <sup>13</sup>C-labeling period, the enrichment level decreased to ~90 <sup>13</sup>C atom% due to <sup>12</sup>CO<sub>2</sub> respiration from the plants. Three and eight days after the start of the <sup>13</sup>C-labeling, half of the pots (6) were harvested. All shoots were cut at soil level. These shoots and the undisturbed soils were directly frozen and stored at -80°C prior to further processing.

## Rhizosphere Soil Harvesting

Rhizosphere samples were harvested from roots as previously described (Lundberg et al., 2012; Lebeis et al., 2015). Briefly, roots were removed from each pot and separated from loose soil. The roots with rhizosphere soil were transferred to a sterile 50-mL

Falcon tube with 25 mL of sterile harvesting buffer (6.33 g of NaH<sub>2</sub>PO<sub>4</sub>·H<sub>2</sub>O, 16.5 g of Na<sub>2</sub>HPO<sub>4</sub>·H<sub>2</sub>O, and 200 μL Silwet L-77 in 1 L of water). The Falcon tubes were then vortexed for 15 s to separate the rhizosphere soil from the roots. Large roots were removed from the Falcon tubes using sterile tweezers. The turbid solutions were filtered through a sterile 100 μm nylon mesh cell strainer (BD Biosciences) into a new sterile 50-mL Falcon tube to remove sand and fine roots. The resulting turbid filtrates were centrifuged to generate pellets of rhizosphere soil. The initial soil samples collected from the sieved soil were processed in the same way as the rhizosphere soil samples as described above.

## <sup>13</sup>C-Methanol Stable Isotope Probing

Fresh soil was collected identically from the same site as the soil used above in the plant growth experiment. Six microcosms were set up with 4 g of soil in glass jars. The <sup>13</sup>C-methanol labeling solution was 5% v/v of <sup>13</sup>C-methanol (99 atom% <sup>13</sup>C, Sigma-Aldrich) in a half-strength Hoagland's solution. Every day,

20  $\mu\text{L}$  of the  $^{13}\text{C}$ -methanol labeling solution was added into each microcosm's soil, which was then homogenized and incubated at the room temperature. The amount of  $^{13}\text{C}$ -methanol added to the soil communities in this study was substantially lower than a previous study of  $^{13}\text{C}$ -methanol SIP of soil communities (Radajewski et al., 2000). A triplicate of the microcosms was harvested after 3 days of  $^{13}\text{C}$ -methanol addition and incubation. Another triplicate was harvested after 8 days.

## DNA Extraction and Metagenomic Sequencing

Total DNAs in the collected soil samples were extracted using the PowerSoil DNA Isolation kit (Mo Bio Laboratories) in technical triplicates, which were pooled and purified using PowerClean Pro DNA Clean-Up kit (Mo Bio Laboratories). DNA concentrations were quantified with the Qubit fluorometric assay (Invitrogen) and 260/280 and 260/230 ratios were quantified using NanoDrop Spectrophotometer (Thermo Scientific). After TruSeq PCR-free library preparation, the 14 DNA samples (12 from the rhizosphere and 2 from the initial soil) were multiplexed together and sequenced by two  $2 \times 150\text{-bp}$  sequencing lanes on Illumina HiSeq 3000 at the Center for Genome Research and Biocomputing of Oregon State University.

## Metagenomic Assembly, Gene Prediction, and Functional Annotation

Metagenomic reads were pre-processed using BBTools<sup>1</sup> for adapter sequence removal, read trimming and filtering, and sequencing error correction. The pre-processed reads from 14 samples were co-assembled to a composite metagenome with Omega 3.0 (Haider et al., 2014). Genes and proteins were predicted from scaffolds  $>1\text{ kbp}$  using the Prodigal (Hyatt et al., 2010). Protein functions were annotated using the UniFam protein family database (Chai et al., 2014).

## Untargeted Binning

Untargeted binning of scaffolds  $>5\text{ kbp}$  were performed with anvi'o v3 following the workflow described previously (Eren et al., 2015; Delmont et al., 2018). We used the anvi-cluster-with-concoct program which implemented the binning algorithm CONCOCT (Alneberg et al., 2014) to automatically cluster the scaffolds to 10 clusters. Then, we used the anvi-refine program to manually bin each CONCOCT cluster to generate MAGs with more than 70% completeness and less than 10% contamination estimated by anvi'o. We further estimated the quality of each MAG by checkM (Parks et al., 2015) and required each MAG to have more than 70% completeness and less than 10% contamination estimated by lineage-specific marker genes in checkM. The completeness and contamination of MAG43 was estimated by anvi'o because the MAG43 does not contain lineage-specific marker genes.

## Targeted Binning

Targeted binning followed the same procedures as the untargeted binning, except that the anvi-refine program was executed

with the `-additional-layers` parameter to highlight the scaffolds which contain genes encoding  $^{13}\text{C}$ -labeled proteins in each CONCOCT cluster. We carried out manual binning around these highlighted scaffolds and extracted MAGs that had less than 5% contamination estimated by checkM and contained at least two  $^{13}\text{C}$ -labeled proteins identified in the rhizosphere or one  $^{13}\text{C}$ -labeled protein identified from at least two replicates in a SIP incubation triplicate in the  $^{13}\text{C}$ -methanol SIP experiment.

## Protein Extraction

Initial soils, rhizosphere soils, and methanol SIP microcosm soils were prepared for protein extraction in the same manner. One gram of soil from each sample was ground to fine powder in liquid nitrogen using a mortar and pestle. The ground soil sample was transferred to a 2-mL Eppendorf tube. One mL of the Solution SP1 of the Novipure Soil Protein Extraction Kit (Mo Bio Laboratories) was added to the 2-mL tube which was then vortexed to completely mix the soil with the buffer, incubated at  $4^\circ\text{C}$  for 10 min, sonicated (20% amplitude, 10 s pulse with 10 s rest, 2 min total pulse time). After the sonication, 0.3 mL of the Solution SP2 was added into the tube which was then vortexed, incubated at  $4^\circ\text{C}$  for 30 min, and sonicated again using the same setting as described above. The tube was centrifuged at  $10,000 \times g$  for 5 min and the supernatant containing the crude protein extract was transferred into a new 2-mL Eppendorf tube. The crude protein extract in the supernatant was precipitated by trichloroacetic acid (Sigma-Aldrich) overnight at  $4^\circ\text{C}$ , pelleted by centrifugation, and washed with ice-cold acetone. The pelleted protein was re-suspended in 200  $\mu\text{L}$  of 6 M guanidine (Sigma-Aldrich). In order to extract sufficient protein for triplicate metaproteomic measurements of each sample, duplicate protein extractions were conducted for each sample. The extracted proteins from the duplicate extractions of the same sample were pooled and quantified by BCA assay (Pierce Biotechnology). After the BCA assay, dithiothreitol (Sigma-Aldrich) was added to the tube for reducing the disulfide bonds of proteins. Fifty  $\mu\text{g}$  of proteins from each soil sample were further processed using the filter-aided sample preparation as described previously (Wiśniewski et al., 2009; Li et al., 2014). Proteins were first digested with trypsin (Promega) using an enzyme:substrate ratio of 1:100 (w:w) for overnight at room temperature, followed by an additional digestion with trypsin using the same enzyme:substrate ratio for 4 h. The peptide samples were stored at  $-80^\circ\text{C}$ .

## Liquid Chromatography-Mass Spectrometry (LC-MS)

Metaproteomic measurements were carried out by using the 11-step multidimensional protein identification technology (MudPIT) (Washburn et al., 2001) on an LTQ-Orbitrap Elite mass spectrometer (Thermo Scientific), as described previously (Li et al., 2014). Ten microgram of proteins were loaded onto a MudPIT run. The MudPIT runs were configured with 11 salt pulses at 5, 7, 10, 12, 15, 17, 20, 25, 35, 50, and 100% of Solvent D which contains 500 mM ammonium acetate dissolved in the Solvent A (95% water, 5% acetonitrile and 0.1% formic acid). Each

<sup>1</sup><http://jgi.doe.gov/data-and-tools/bbtools/>

salt pulse was followed by a 110-min gradient from 100% of the Solvent A to 60% of the Solvent B (30% water, 70% acetonitrile, and 0.1% formic acid). The LC eluent was directly nanosprayed (Proxeon) into the LTQ Orbitrap Elite mass spectrometer. MS scans were acquired in Orbitrap with a resolution of 30,000. The top 8 most abundant ions were selected for collision-induced dissociation and fragment ions were measured in Orbitrap with a resolution of 15,000. Each soil sample had technical triplicate LC-MS measurements.

## Peptide and Protein Identification by Database Searching

The regular database searching for identification of unlabeled proteins was conducted using Sipros Ensemble (Guo et al., 2017). The soil microbial protein database comprised of >1.8 million proteins predicted from scaffolds >5 kbp in the composite metagenome assembly. The initial soil samples were searched against the soil microbial protein database. An *A. thaliana* proteome database downloaded from [https://phytozome.jgi.doe.gov/pz/portal.html#!bulk?org=Org\\_Athaliana](https://phytozome.jgi.doe.gov/pz/portal.html#!bulk?org=Org_Athaliana) (Araport11) was added to the soil microbial protein database to search the *A. thaliana* rhizosphere soil samples. A *Z. mays* proteome database from [https://phytozome.jgi.doe.gov/pz/portal.html#!bulk?org=Org\\_Zmays](https://phytozome.jgi.doe.gov/pz/portal.html#!bulk?org=Org_Zmays) (Ensembl-18) was added to the soil microbial protein database to search the *Z. mays* rhizosphere soil samples. A *T. aestivum* proteome database from [https://phytozome.jgi.doe.gov/pz/portal.html#!bulk?org=Org\\_Taestivum\\_er](https://phytozome.jgi.doe.gov/pz/portal.html#!bulk?org=Org_Taestivum_er) (v2.2) was added to the soil microbial protein database to search the *T. aestivum* rhizosphere soil samples. Plant proteins identified in the rhizosphere samples were shown in **Supplementary Table 5**. The methanol incubation soil samples were searched against the soil microbial database. All these protein databases were constructed with concatenated forward and reverse proteins for FDR estimation. The searches used the following parameters: parent mass offsets of -1, 0, 1, 2, 3 Da; mass tolerance of 0.03 and 0.01 Da for parent ions and fragment ions, respectively; up to three missed cleavages; dynamic modification of methionine residue by oxidation; and full enzyme specificity required. The database searching results were filtered to achieve 1% FDR at the peptide level. Proteins were inferred from the identified peptides using the parsimony rule (Nesvizhskii and Aebersold, 2005). Indistinguishable proteins were combined as a protein group. A minimum of one peptide unique to each inferred protein or protein groups was required. The FDR at the protein level was estimated to be less than 2% for all metaproteome samples. The abundance of identified proteins was quantified based on normalized peak areas using ProRata (Pan et al., 2006).

The SIP database searching for identification of <sup>13</sup>C-labeled proteins was conducted using Sipros with the weighted dot-production scoring function as described in previous studies (Pan et al., 2011; Justice et al., 2014; Bryson et al., 2016, 2017; Marlow et al., 2016). The SIP database for the rhizosphere samples of each plant species included all the proteins identified from these samples in the regular searches. Similarly, the SIP

database for the methanol incubation samples included all the proteins identified in these samples in the regular searches. SIP searches were also conducted for the initial soil samples as a negative control using a protein database containing all the unlabeled proteins identified in the initial soil samples. All SIP searches used the following parameters: parent mass offsets of -4, -3, -2, -1, 0, 1, 2, 3, 4 Da; 0.03 Da of mass tolerances for parent ions; 0.01 Da of mass tolerances for fragment ions; up to three missed cleavages; full enzyme specificity required. SIP searching results were filtered to achieve 1% FDR at the peptide level using a target-decoy approach (Elias and Gygi, 2007). Proteins were inferred from the peptide identification results by requiring a minimum of one unique peptide per protein or protein group. Each <sup>13</sup>C-labeled protein was required to have a minimum of >10% <sup>13</sup>C enrichment and it also must have a minimum of two peptides identified in at least one of the four rhizosphere samples of the same plant species with which it was associated. <sup>13</sup>C-labeled peptides were clustered into protein isotopologues until the average atom% difference between clusters was >10%.

## Taxonomy Assignment

The taxonomy of MAGs were initially inferred using three different tools, namely, checkM (Parks et al., 2015), AMPHORA (Wu and Eisen, 2008), and best USEARCH (Edgar, 2010) hits of predicted proteins on MAG scaffolds to the Uniprot, Uniref90, and KEGG databases. The AMPHORA-based taxonomy assignment of these MAGs was conducted on the AmphoraNet webserver (Kerepesi et al., 2014) and each MAG was assigned with a taxon if at least 75% of the identified marker genes in that MAG generated a concordant taxonomy or with a putative taxon if more than 50% and less than 75% of the identified marker genes generated a concordant taxonomy (Muller et al., 2014). If the taxonomy assignments of a MAG from at least two of the three approaches agreed at the phylum level, the lowest common ancestor of these taxonomy assignments were selected as the taxon of that MAG.

## Phylogenomic Analysis

The root-associated isolate genomes were downloaded from the **Supplementary Material** of Levy et al. (2018). We used Phylosift (Darling et al., 2014) to extract marker genes from all the MAGs and the root-associated isolated genomes. A phylogenomic tree of these genomes was constructed based on the concatenated protein sequence alignment of marker genes using the program anvi-gen-phylogenomic-tree in anvi'o and the FastTree algorithm (Price et al., 2010). The tree was uploaded into the Interactive Tree Of Life server for visualization.

## Phylogenetic Analysis of XoxF

The protein sequences of 68 XoxF identified by metaproteomics in this study were aligned to 143 reference XoxF sequences obtained from a previously study (Taubert et al., 2015) using Clustal Omega at <https://www.ebi.ac.uk/Tools/msa/clustalo/>. A newick-formatted tree was generated after the sequence alignment. The tree was uploaded into the Interactive Tree Of Life server for visualization.

## DATA AVAILABILITY STATEMENT

Metagenomic sequencing reads associated with this study have been deposited in BioProject under accession PRJNA488251. Mass spectrometry raw files for this study have been deposited in ProteomeXchange under accessions PXD011737, PXD011738, PXD011739, PXD011891, and PXD011892. MAGs recovered in this study can be downloaded from <https://figshare.com/s/2a812c513ab14e6c8161>.

## AUTHOR CONTRIBUTIONS

CP, ZL, and GH designed the study. ZL, QY, MM, RH, and SL performed the experiments. ZL, QY, XG, WH, and CP implemented data analysis algorithms. ZL, QY, A-CC, and JB analyzed metagenomics binning results. ZL and CP analyzed the results and prepared the manuscript. All authors discussed the results and edited the manuscript.

## FUNDING

This work was supported by the U.S. Department of Energy (DOE), Office of Science, Office of Biological and Environmental Research, the ORNL Plant-Microbe Interfaces Scientific Focus Area project, and the DOE Grants DOE-SC10010566 and DE-SC0020356. WH acknowledges support

## REFERENCES

- Alneberg, J., Bjarnason, B. S., De Bruijn, I., Schirmer, M., Quick, J., Ijaz, U. Z., et al. (2014). Binning metagenomic contigs by coverage and composition. *Nat. Methods* 11:1144. doi: 10.1038/nmeth.3103
- Anantharaman, K., Brown, C. T., Hug, L. A., Sharon, I., Castelle, C. J., Probst, A. J., et al. (2016). Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat. Commun.* 7:13219. doi: 10.1038/ncomms13219
- Bryson, S., Li, Z., Chavez, F., Weber, P. K., Pett-Ridge, J., Hettich, R. L., et al. (2017). Phylogenetically conserved resource partitioning in the coastal microbial loop. *ISME J.* 11:2781. doi: 10.1038/ismej.2017.128
- Bryson, S., Li, Z., Pett-Ridge, J., Hettich, R. L., Mayali, X., Pan, C., et al. (2016). Proteomic stable isotope probing reveals taxonomically distinct patterns in amino acid assimilation by coastal marine bacterioplankton. *MSystems* 1:e00027-15.
- Bulgarelli, D., Rott, M., Schlaeppi, K., van Themaat, E. V. L., Ahmadinejad, N., Assenza, F., et al. (2012). Revealing structure and assembly cues for Arabidopsis root-inhabiting bacterial microbiota. *Nature* 488:91. doi: 10.1038/nature11336
- Bulgarelli, D., Schlaeppi, K., Spaepen, S., van Themaat, E. V. L., and Schulze-Lefert, P. (2013). Structure and functions of the bacterial microbiota of plants. *Annu. Rev. Plant Biol.* 64, 807–838. doi: 10.1146/annurev-arplant-050312-120106
- Butterfield, C. N., Li, Z., Andeer, P. F., Spaulding, S., Thomas, B. C., Singh, A., et al. (2016). Proteogenomic analyses indicate bacterial methylotrophy and archaeal heterotrophy are prevalent below the grass root zone. *PeerJ* 4:e2687. doi: 10.7717/peerj.2687
- Chai, J., Kora, G., Ahn, T.-H., Hyatt, D., and Pan, C. (2014). Functional phylogenomics analysis of bacteria and archaea using consistent genome annotation with UniFam. *BMC Evol. Biol.* 14:207. doi: 10.1186/s12862-014-0207-y
- Chistoserdova, L. (2011). Modularity of methylotrophy, revisited. *Environ. Microbiol.* 13, 2603–2622. doi: 10.1111/j.1462-2920.2011.02464.x

from the HASI program and Naval Research Laboratory core funding. MM acknowledges funding support from U.S. Department of Energy (DOE), Office of Science Early Career Award. ORNL is managed by UT-Battelle, LLC, under contract DE-AC05-00OR22725 for the U.S. Department of Energy. The opinions and assertions contained herein are those of the authors and are not to be construed as those of the U.S. Navy, Military Service at large, and/or the U.S. Government.

## ACKNOWLEDGMENTS

We thank Ton Gorissen at the Isolife for assistance with the plant growth and  $^{13}\text{CO}_2$  labeling. This research used resources of the Oak Ridge Leadership Computing Facility and Department of Defense's High Performance Computing Modernization Program Application Software Initiative (HASI) under the U.S. Army Corps of Engineers, Engineer Research and Development Center (ERDC) by the Department of the Army.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2019.02706/full#supplementary-material>

- Chistoserdova, L., Kalyuzhnaya, M. G., and Lidstrom, M. E. (2009). The expanding world of methylotrophic metabolism. *Annu. Rev. Microbiol.* 63, 477–499. doi: 10.1146/annurev.micro.091208.073600
- Costacurta, A., and Vanderleyden, J. (1995). Synthesis of phytohormones by plant-associated bacteria. *Crit. Rev. Microbiol.* 21, 1–18. doi: 10.3109/10408419509113531
- Crits-Christoph, A., Diamond, S., Butterfield, C. N., Thomas, B. C., and Banfield, J. F. (2018). Novel soil bacteria possess diverse genes for secondary metabolite biosynthesis. *Nature* 558, 440–444. doi: 10.1038/s41586-018-0207-y
- Darling, A. E., Jospin, G., Lowe, E., Matsen, F. A. I. V., Bik, H. M., et al. (2014). PhyloSift: phylogenetic analysis of genomes and metagenomes. *PeerJ* 2:e243. doi: 10.7717/peerj.243
- DeAngelis, K. M., Lindow, S. E., and Firestone, M. K. (2008). Bacterial quorum sensing and nitrogen cycling in rhizosphere soil. *FEMS Microbiol. Ecol.* 66, 197–207. doi: 10.1111/j.1574-6941.2008.00550.x
- Delmont, T. O., Quince, C., Shaiber, A., Esen, Ö.C., Lee, S. T., Rappé, M. S., et al. (2018). Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. *Nat. Microbiol.* 3, 804–813. doi: 10.1038/s41564-018-0176-9
- Delmotte, N., Knief, C., Chaffron, S., Innerebner, G., Roschitzki, B., Schlapbach, R., et al. (2009). Community proteogenomics reveals insights into the physiology of phyllosphere bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 106, 16428–16433. doi: 10.1073/pnas.0905240106
- Diamond, S., Andeer, P., Li, Z., Crits-Christoph, A., Burstein, D., Anantharaman, K., et al. (2018). Processing of grassland soil CN compounds into soluble and volatile molecules is depth stratified and mediated by genomically novel bacteria and archaea. *bioRxiv[preprint]* doi: 10.1101/445817
- Drigo, B., Pijl, A. S., Duyts, H., Kielak, A. M., Gamper, H. A., Houtekamer, M. J., et al. (2010). Shifting carbon flow from roots into associated microbial communities in response to elevated atmospheric CO<sub>2</sub>. *Proc. Natl. Acad. Sci. U.S.A.* 107, 10938–10942. doi: 10.1073/pnas.0912421107

- Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi: 10.1093/bioinformatics/btq461
- el Zahar Haichar, F., Heulin, T., Guyonnet, J. P., and Achouak, W. (2016). Stable isotope probing of carbon flow in the plant holobiont. *Curr. Opin. Biotechnol.* 41, 9–13. doi: 10.1016/j.copbio.2016.02.023
- el Zahar Haichar, F., Marol, C., Berge, O., Rangel-Castro, J. L., Prosser, J. I., Balesdent, J., et al. (2008). Plant host habitat and root exudates shape soil bacterial community structure. *ISME J.* 2:1221. doi: 10.1038/ismej.2008.80
- Elias, J. E., and Gygi, S. P. (2007). Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* 4:207. doi: 10.1038/nmeth1019
- Eren, A. M., Esen, Ö.C., Quince, C., Vineis, J. H., Morrison, H. G., Sogin, M. L., et al. (2015). Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* 3:e1319. doi: 10.7717/peerj.1319
- Frébort, I., Kowalska, M., Hluska, T., Frébortová, J., and Galuszka, P. (2011). Evolution of cytokinin biosynthesis and degradation. *J. Exp. Bot.* 62, 2431–2452. doi: 10.1093/jxb/err004
- Galbally, I., and Kirstine, W. (2002). The production of methanol by flowering plants and the global cycle of methanol. *J. Atmos. Chem.* 43, 195–229.
- Glick, B. R., Cheng, Z., Czarny, J., and Duan, J. (2007). "Promotion of plant growth by ACC deaminase-producing soil bacteria," in *New Perspectives and Approaches in Plant Growth-Promoting Rhizobacteria Research*, eds P. A. H. M. Bakker, J. M. Raaijmakers, G. Bloembergen, M. Höfte, P. Lemanceau, and B. M. Cooke, (Dordrecht: Springer), 329–339. doi: 10.1007/978-1-4020-6776-1\_8
- Guo, X., Li, Z., Yao, Q., Mueller, R. S., Eng, J. K., Tabb, D. L., et al. (2017). Sipros ensemble improves database searching and filtering for complex metaproteomics. *Bioinformatics* 34, 795–802. doi: 10.1093/bioinformatics/btx601
- Haider, B., Ahn, T.-H., Bushnell, B., Chai, J., Copeland, A., and Pan, C. (2014). Omega: an overlap-graph de novo assembler for metagenomics. *Bioinformatics* 30, 2717–2722. doi: 10.1093/bioinformatics/btu395
- Harms, N., Reijnders, W. N., Anazawa, H., van der Palen, C. J., van Spanning, R. J., Oltmann, L. F., et al. (1993). Identification of a two-component regulatory system controlling methanol dehydrogenase synthesis in *Paracoccus denitrificans*. *Mol. Microbiol.* 8, 457–470. doi: 10.1111/j.1365-2958.1993.tb01590.x
- Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* 11:119. doi: 10.1186/1471-2105-11-119
- Jones, D. L., Nguyen, C., and Finlay, R. D. (2009). Carbon flow in the rhizosphere: carbon trading at the soil–root interface. *Plant Soil* 321, 5–33. doi: 10.1007/s11104-009-9925-0
- Justice, N. B., Li, Z., Wang, Y., Spaulding, S. E., Mosier, A. C., Hettich, R. L., et al. (2014). 15 N-and 2 H proteomic stable isotope probing links nitrogen flow to archaeal heterotrophic activity. *Environ. Microbiol.* 16, 3224–3237. doi: 10.1111/1462-2920.12488
- Kerepesi, C., Banky, D., and Grolmusz, V. (2014). AmphoraNet: the webserver implementation of the AMPHORA2 metagenomic workflow suite. *Gene* 533, 538–540. doi: 10.1016/j.gene.2013.10.015
- Lebeis, S. L., Paredes, S. H., Lundberg, D. S., Breakfield, N., Gehring, J., McDonald, M., et al. (2015). Salicylic acid modulates colonization of the root microbiome by specific bacterial taxa. *Science* 349, 860–864. doi: 10.1126/science.aaa8764
- Lerat, E., and Moran, N. A. (2004). The evolutionary history of quorum-sensing systems in bacteria. *Mol. Biol. Evol.* 21, 903–913. doi: 10.1093/molbev/msh097
- Levy, A., Gonzalez, I. S., Mittelviehhaus, M., Clingenpeel, S., Paredes, S. H., Miao, J., et al. (2018). Genomic features of bacterial adaptation to plants. *Nat. Genet.* 50:138.
- Li, Z., Wang, Y., Yao, Q., Justice, N. B., Ahn, T.-H., Xu, D., et al. (2014). Diverse and divergent protein post-translational modifications in two growth stages of a natural microbial community. *Nat. Commun.* 5:4405. doi: 10.1038/ncomms5405
- Lu, Y., and Conrad, R. (2005). In situ stable isotope probing of methanogenic archaea in the rice rhizosphere. *Science* 309, 1088–1090. doi: 10.1126/science.1113435
- Lundberg, D. S., Lebeis, S. L., Paredes, S. H., Yourstone, S., Gehring, J., Malfatti, S., et al. (2012). Defining the core *Arabidopsis thaliana* root microbiome. *Nature* 488:86. doi: 10.1038/nature11237
- Marlow, J. J., Skennerton, C. T., Li, Z., Chourey, K., Hettich, R. L., Pan, C., et al. (2016). Proteomic stable isotope probing reveals biosynthesis dynamics of slow growing methane based microbial communities. *Front. Microbiol.* 7:563. doi: 10.3389/fmicb.2016.00563
- Mendes, R., Kruijt, M., De Bruijn, I., Dekkers, E., van der Voort, M., Schneider, J. H., et al. (2011). Deciphering the rhizosphere microbiome for disease-suppressive bacteria. *Science* 332, 1097–1100. doi: 10.1126/science.1203980
- Morgan, P. W., and Drew, M. C. (1997). Ethylene and plant responses to stress. *Physiol. Plant* 100, 620–630. doi: 10.1111/j.1399-3054.1997.tb03068.x
- Mosbaek, F., Kjeldal, H., Mulat, D. G., Albertsen, M., Ward, A. J., Feilberg, A., et al. (2016). Identification of syntrophic acetate-oxidizing bacteria in anaerobic digesters by combined protein-based stable isotope probing and metagenomics. *ISME J.* 10, 2405–2418. doi: 10.1038/ismej.2016.39
- Muller, E. E., Pinel, N., Laczny, C. C., Hoopmann, M. R., Narayanasamy, S., Lebrun, L. A., et al. (2014). Community-integrated omics links dominance of a microbial generalist to fine-tuned resource usage. *Nat. Commun.* 5:5603. doi: 10.1038/ncomms6603
- Nesvizhskii, A. I., and Aebersold, R. (2005). Interpretation of shotgun proteomic data: the protein inference problem. *Mol. Cell. Proteomics* 4, 1419–1440. doi: 10.1074/mcp.r500012-mcp200
- Pan, C., Fischer, C. R., Hyatt, D., Bowen, B. P., Hettich, R. L., and Banfield, J. F. (2011). Quantitative tracking of isotope flows in proteomes of microbial communities. *Mol. Cell. Proteomics* 10:M110.006049. doi: 10.1074/mcp.M110.006049
- Pan, C., Kora, G., McDonald, W. H., Tabb, D. L., VerBerkmoes, N. C., Hurst, G. B., et al. (2006). ProRata: a quantitative proteomics program for accurate protein abundance ratio estimation with confidence interval evaluation. *Anal. Chem.* 78, 7121–7131. doi: 10.1021/ac060654b
- Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055. doi: 10.1101/gr.186072.114
- Patten, C. L., and Glick, B. R. (1996). Bacterial biosynthesis of indole-3-acetic acid. *Can. J. Microbiol.* 42, 207–220. doi: 10.1139/m96-032
- Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490. doi: 10.1371/journal.pone.0009490
- Radajewski, S., Ineson, P., Parekh, N. R., and Murrell, J. C. (2000). Stable-isotope probing as a tool in microbial ecology. *Nature* 403, 646–649. doi: 10.1038/35001054
- Ram, R. J., VerBerkmoes, N. C., Thelen, M. P., Tyson, G. W., Baker, B. J., Blake, R. C., et al. (2005). Community proteomics of a natural microbial biofilm. *Science* 308, 1915–1920.
- Schmidt, S., Christen, P., Kiefer, P., and Vorholt, J. A. (2010). Functional investigation of methanol dehydrogenase-like protein XoxF in *Methylobacterium extorquens* AM1. *Microbiology* 156, 2575–2586. doi: 10.1099/mic.0.038570-0
- Sowell, S. M., Abraham, P. E., Shah, M., Verberkmoes, N. C., Smith, D. P., Barofsky, D. F., et al. (2011). Environmental proteomics of microbial plankton in a highly productive coastal upwelling system. *ISME J.* 5:856. doi: 10.1038/ismej.2010.168
- Starr, E. P., Shi, S., Blazewicz, S. J., Probst, A. J., Herman, D. J., Firestone, M. K., et al. (2018). Stable isotope informed genome-resolved metagenomics reveals that Saccharibacteria utilize microbially-processed plant-derived carbon. *Microbiome* 6:122. doi: 10.1186/s40168-018-0499-z
- Taubert, M., Grob, C., Howat, A. M., Burns, O. J., Dixon, J. L., Chen, Y., et al. (2015). XoxF encoding an alternative methanol dehydrogenase is widespread in coastal marine environments. *Environ. Microbiol.* 17, 3937–3948. doi: 10.1111/1462-2920.12896
- Tien, T., Gaskins, M., and Hubbell, D. (1979). Plant growth substances produced by *Azospirillum brasilense* and their effect on the growth of pearl millet (*Pennisetum americanum* L.). *Appl. Environ. Microbiol.* 37, 1016–1024.

- Vandenkoornhuyse, P., Mahé, S., Ineson, P., Staddon, P., Ostle, N., Cliquet, J.-B., et al. (2007). Active root-inhabiting microbes identified by rapid incorporation of plant-derived carbon into RNA. *Proc. Natl. Acad. Sci. U.S.A.* 104, 16970–16975. doi: 10.1073/pnas.0705902104
- Washburn, M. P., Wolters, D., and Yates, J. R. I. I. I. (2001). Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* 19:242. doi: 10.1038/85686
- Wiśniewski, J. R., Zougman, A., Nagaraj, N., and Mann, M. (2009). Universal sample preparation method for proteome analysis. *Nat. Methods* 6:359. doi: 10.1038/nmeth.1322
- Wu, M., and Eisen, J. A. (2008). A simple, fast, and accurate method of phylogenomic inference. *Genome Biol.* 9:R151. doi: 10.1186/gb-2008-9-10-r151
- Yao, Q., Li, Z., Song, Y., Wright, S. J., Guo, X., Tringe, S. G., et al. (2018). Community proteogenomics reveals the systemic impact of phosphorus availability on microbial functions in tropical soil. *Nat. Ecol. Evol.* 2:499. doi: 10.1038/s41559-017-0463-5
- Zhalnina, K., Louie, K. B., Hao, Z., Mansoori, N., da Rocha, U. N., Shi, S., et al. (2018). Dynamic root exudate chemistry and microbial substrate preferences drive patterns in rhizosphere microbial community assembly. *Nat. Microbiol.* 3:470. doi: 10.1038/s41564-018-0129-3

**Disclaimer:** This manuscript has been authored in part by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the U.S. Department of Energy (DOE). The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a non-exclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for U.S. Government purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (<http://energy.gov/downloads/doe-public-access-plan>). The opinions and assertions contained herein are those of the authors and are not to be construed as those of the U.S. Government.

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Li, Yao, Guo, Crits-Christoph, Mayes, Hervey, Lebeis, Banfield, Hurst, Hettich and Pan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms. The U.S. Government retains a nonexclusive license to this work for noncommercial purposes.