



## Article

# Soil Metabolomics Predict Microbial Taxa as Biomarkers of Moisture Status in Soils from a Tidal Wetland

Taniya RoyChowdhury <sup>1,\*</sup>, Lisa M. Bramer <sup>1</sup>, Joseph Brown <sup>1,†</sup>, Young-Mo Kim <sup>1</sup>, Erika Zink <sup>1</sup>, Thomas O. Metz <sup>1</sup>, Lee Ann McCue <sup>1</sup>, Heida L. Diefenderfer <sup>2</sup> and Vanessa Bailey <sup>1</sup>

<sup>1</sup> Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA 99352, USA

<sup>2</sup> Energy and Environment Directorate, Pacific Northwest National Laboratory, Sequim, WA 98382, USA

\* Correspondence: troychowdhury@woodwellclimate.org; Tel.: +1-614-619-2131

† Present address: Woodwell Climate Research Center, Falmouth, MA 02540, USA.

‡ Present address: Department of Human Genetics, University of Utah, Salt Lake City, UT 84112, USA.

**Abstract:** We present observations from a laboratory-controlled study on the impacts of extreme wetting and drying on a wetland soil microbiome. Our approach was to experimentally challenge the soil microbiome to understand impacts on anaerobic carbon cycling processes as the system transitions from dryness to saturation and vice-versa. Specifically, we tested for impacts on stress responses related to shifts from wet to drought conditions. We used a combination of high-resolution data for small organic chemical compounds (metabolites) and biological (community structure based on 16S rRNA gene sequencing) features. Using a robust correlation-independent data approach, we further tested the predictive power of soil metabolites for the presence or absence of taxa. Here, we demonstrate that taking an untargeted, multidimensional data approach to the interpretation of metabolomics has the potential to indicate the causative pathways selecting for the observed bacterial community structure in soils.

**Keywords:** 16S rRNA gene; soil metabolomics; random forest classification; wetland microbiome



**Citation:** RoyChowdhury, T.; Bramer, L.M.; Brown, J.; Kim, Y.-M.; Zink, E.; Metz, T.O.; McCue, L.A.; Diefenderfer, H.L.; Bailey, V. Soil Metabolomics Predict Microbial Taxa as Biomarkers of Moisture Status in Soils from a Tidal Wetland.

*Microorganisms* **2022**, *10*, 1653.

<https://doi.org/10.3390/microorganisms10081653>

Academic Editor: Matthias Noll

Received: 21 April 2022

Accepted: 12 August 2022

Published: 16 August 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Hydrologic fluctuations are a driving factor regulating wetland biogeochemical processes, but controlled studies investigating the effect of rapid and extreme moisture regime shifts on the soil microbiome, specifically on anaerobic carbon (C) cycling pathways, are limited. The few studies on the impact of fluctuating precipitation regimes (extreme wetting and drying) on the soil microbiome have been carried out on upland systems either by characterizing microbial community composition shifts [1–5], or by identifying correlative associations between the microbial community and metabolite data [6,7]. The latter correlations are useful for describing, predicting, and managing functional traits of the microbiome [8] and can inform new hypotheses regarding microbe-dependent ecosystem processes [9,10]. For example, we know that there are direct links between soil redox conditions and the soil microbiome that impact biogeochemical functions including greenhouse gas production, iron reduction, and C-sequestration potentials [11–14]. Still, metrics such as correlations are limited in detecting linear associations. To uncover causal relationships, statistical approaches free of the assumption of linear relationships, as is commonly seen in multidimensional microbiome datasets from soil, must be utilized. Such approaches that link community membership information (e.g., 16S rRNA gene surveys) with microbial metabolic processes remain unexplored in soil systems.

Untargeted mass spectrometry approaches to characterize soil/sediment metabolomes have only recently been developed and applied [3,15,16]. From these techniques, metabolic pathways have been inferred using network analysis methods [7] or through integration of gene expression (metatranscriptomes) with metabolite profiles using association networks [3]. The rationale for the use of association networks to analyze metabolomics data

is that differences or commonalities in biological processes are reflected in the characteristics of the inferred correlation networks [17]. Generally, the correlations observed in metabolomics data are the result of the combination of all reactions and regulatory processes in the network; there may be no correlations between metabolites that are close in a certain metabolic pathway [18]. Pairwise correlations are also limited because only pairs of features can be assessed at once and correlation quantifies the linear relationship between two features. Such approaches allow us to visualize and examine correlations/relationships between pairs of features simultaneously but may not work for soil metabolomic profiling that are typically characterized by complex multivariate relationships with the microbial community structure. Inferring soil functions from correlations between observed features for microbial taxa and small organic molecules (metabolite and/or substrate) is inadequate and provides limited insight into relationships between these features.

An area of soil multi-omics research needing major advances is in the application of metabolome data from untargeted approaches, i.e., mass spectrometry, to infer microbial community composition and predict community metabolism. Derivatives of the relative abundance of a set of Operational Taxonomic Units (OTUs) from soil microbiomes have been used to infer ecological modules or clusters [19,20]. However, microbial taxa composition alone is a poor predictor of soil function because it does not link ecology to phenotype, i.e., the plethora of physiological adaptations expressed in prokaryotes, such as the use of different C functional groups, and terminal electron acceptors as a soil environment change. Here, we adopted the definition of the metabolome as an effective phenotype modulator [21] and used it to infer soil microbiome biomarkers under altered soil moisture conditions as described below.

We used controlled laboratory incubations to simulate the following three soil moisture conditions: constant saturation, fluctuating, and drying as previously described [22]. Our objective was to force shifts in the soil microbial community composition and metabolite profiles. We hypothesized that the accumulation of metabolites central to fermentative pathways, and anaerobic respiration under wet/saturated conditions would provide evidence for the presence of microbial taxa specializing in these pathways. Similarly, microbial taxa resilient to rapid drying would be predicted by metabolites indicative of drought/stress tolerance. We applied a data-integration approach to determine if and how the composition and abundance of metabolites predict the presence/absence of microbial taxa in soils. We used a random forest classifier [23] to integrate data from 16S rRNA sequencing and metabolite profiles from gas chromatography–mass spectrometry (GC–MS). Our findings indicated that the use of metabolites as predictors of presence/absences of taxa using random forest models can be a meaningful approach to link community structure to function.

## 2. Materials and Methods

### 2.1. Soil Collection and Incubation Conditions

#### 2.1.1. Site

The study site is located within the tidally influenced, freshwater Secret River wetland on the Columbia River estuary floodplain in southwestern Washington State, USA. In a recent synoptic study of wetlands in the Pacific Northwest USA, the tidal forested wetland at Secret River was shown to have one of the highest total ecosystem C stocks in the world, exceeding mangroves, and more C was sequestered in the soil (to 3-m depth) than in aboveground biomass (Kauffman et al., 2020 [24]). Further details about the study site and soil core collection have been previously described [22].

Fifteen soil cores were sampled in opaque polyvinyl chloride columns (5 cm height  $\times$  2.5-cm i.d.) in 2016 using a grid sampling technique [22] and transported on ice to the Pacific Northwest National Laboratory, Richland, WA, USA, where they were stored at 4 °C until further use. Cores were pre-incubated at 20 °C for 18 days under field-moist conditions (~50% water *w/w*) and monitored daily for greenhouse gas production, i.e., carbon dioxide (CO<sub>2</sub>) and methane (CH<sub>4</sub>), by gas chromatography. At the end of the

pre-incubation period, cores were randomly assigned to each of the three treatments: Wet-Dry (W-D,  $n = 5$ ), Dry-Wet (D-W,  $n = 4$ ), and continuous saturation (Sat,  $n = 3$ ). At least 3 replicates per treatment were deemed minimum for the experiment; additional replicates for the W-D and D-W treatments were included to perform statistical power analyses to determine the minimum replication to confidently attribute sources of variance observed for bulk process measurements (soil respiration, previously published [22]) and molecular analysis, i.e., 16S rRNA sequencing and metabolomics, presented in this study. Three of the 15 cores collected were used for additional soil property analysis [22].

### 2.1.2. Soil Incubation

Samples were maintained in a humidity-controlled (60% relative humidity) chamber for 14 days following treatment initiation and thereafter transferred to a biosafety cabinet at 25 °C for the duration of the incubation (120 days). It took up to 16 h for soil cores to be saturated by placing them in standing water after which an additional 1–2 mL of water was added to the soil column surface to limit evaporative losses. The continuously saturated (Sat) samples were saturated as described above, maintained at constant weights by addition of 1–2 mL of water and incubated anoxically (headspace purged with 99.9% N<sub>2</sub>). The W-D treatment was saturated and initially incubated under a microaerophilic headspace (purged with 70% N<sub>2</sub>) for 14 days and subsequently maintained anoxically for the next 25 days (wet phase). For the D-W treatment, pre-incubated field-moist samples were allowed to dry to a constant weight by evaporation under a microaerophilic headspace for 14 days and then maintained under aerobic conditions for the next 25 days of the dry phase at 20 °C.

After 39 days from treatment onset, the moisture and headspace conditions for the W-D and D-W treatments were reversed: the W-D samples initially under wet/anoxic conditions were subjected to drying by evaporation under aerobic conditions at 20 °C (moisture data published in [22]). The D-W samples from the initially dry treatments were wetted by addition of de-gassed sterile de-ionized water until the rate of evaporative loss was lower than the volume of water added and incubated under an anoxic headspace. Wetting was conducted in the same manner as the initial wetting treatments, described above. Gas measurements for headspace CO<sub>2</sub> and CH<sub>4</sub> were measured throughout the incubation and have been previously reported [22].

### 2.1.3. Soil Core Sampling and Analyses

All sample handling and post-incubation processing for the D-W and Sat samples were conducted inside an anoxic glove bag (Coy Laboratories, O<sub>2</sub> < 5 ppm, 2.1% H<sub>2</sub>): soil cores were extruded from the PVC columns, homogenized, and subsampled for deoxyribonucleic acid (DNA) and extracellular metabolite extraction analyses (described below). DNA was extracted immediately on the last day of incubation. DNA was stored at –80 °C until further analysis. Soil subsamples were preserved at –20 °C until metabolite extraction. The experiment generated a total of 60 samples for sequence and metabolite analysis.

### 2.2. DNA Extraction and 16S rRNA Sequence Analysis

DNA was extracted from 0.25 g portions (five technical replicates per core) of homogenized subsamples using the PowerSoil DNA isolation kit (MoBio Laboratories Inc., Carlsbad, CA, USA) according to the manufacturer's instructions. DNA extracts were quantified using the Qubit Fluorometer 2.0 (Thermo Fisher Scientific Inc., Waltham, MA, USA) and checked for quality (A260/A280 > 1.9) using the NanoDrop 2000 (Thermo Fisher Scientific Inc.). Additional DNA purification was performed using the ZR-96 Genomic DNA Clean and Concentrator-5 kit (Zymo Research Corporation, Irvine, CA, USA). The V4 region of the 16S rRNA gene was amplified using the PCR protocol developed by the Earth Microbiome Project [25]; this protocol was modified to include a 12-base barcode sequence in the forward primer. Amplicons were sequenced on an Illumina MiSeq using

the 500 cycle MiSeq Reagent Kit v2 (<http://www.illumina.com/> (accessed on 16 September 2016)) according to manufacturer's instructions.

Raw sequence reads were de-multiplexed using *ea-utils* [26] with zero mismatches allowed in the barcode sequence. Reads were quality-filtered with BBDuk2 [27] to remove adapter sequences and PhiX with a matching kmer length of 31 bp at a hamming distance of 1. Reads shorter than 51 bp were discarded. The forward and reverse reads were merged using USEARCH [28] with a minimum length threshold of 175 bp and maximum mismatch of 1%. Sequences were de-replicated (minimum sequence abundance of 2) and clustered using the distance-based, "greedy clustering" method (i.e., maximize similarity within clusters, minimize similarity between clusters) of USEARCH at 97% pairwise sequence identity among OTU member sequences. De novo prediction of chimeric sequences was performed using USEARCH during clustering. Taxonomy was assigned to OTU sequences at a minimum identity cutoff fraction of 0.8 using the global alignment method implemented in USEARCH across RDP Trainset database version 15 trained with the UTAX 250 bp configuration [29]. OTU seed sequences were filtered against RDP Gold reference database version 9 to identify chimeric OTUs using USEARCH.

### 2.3. Metabolome Extraction and GC-MS Analysis

Extracellular metabolites were extracted from 2 g of wet soil from each core (approx. 1–1.5 g dry weight equivalent) (with five technical replicates) using 10 mL of de-ionized water. Samples were shaken on a mechanical shaker for 12 h at 4 °C, centrifuged at 2500× g for 15 min at 4 °C, and filtered through 0.2 µm polyethersulfone membrane filters (Pall Corporation, New York, NY, USA). The polar metabolites were completely dried under a speed vacuum concentrator, then chemically derivatized and analyzed by gas chromatography–mass spectrometry (GC–MS). An Agilent 7890A gas chromatograph coupled with a quadrupole 5975C mass spectrometer (Agilent Technologies, Inc., Santa Clara, CA, USA) was used for all analyses. Metabolites were derivatized as previously described [3,30] by adding 20 µL of methoxyamine solution (30 mg mL<sup>-1</sup> in pyridine) and were incubated at 37 °C for 90 min to protect the carbonyl groups and reduce carbohydrate isoforms. Then, 80 µL of *N*-methyl-*N*-(trimethylsilyl)-trifluoroacetamide with 1% trimethylchlorosilane were added to each sample to trimethylsilylate the hydroxyl and amine groups for 30 min. The samples were cooled to room temperature prior to GC–MS analysis. Data collected by GC–MS were processed using the MetaboliteDetector software, version 2.5 beta [31]. Retention indices of detected metabolites were calculated based on analysis of the fatty acid methyl esters mixture (C8–C28), followed by chromatographic alignment across all analyses after deconvolution. Metabolites were initially identified by matching experimental spectra to a PNNL-augmented version of the Agilent Fiehn Metabolomics Library, containing spectra and validated retention indices for over 900 metabolites [32]. Additionally, metabolites were cross-checked by matching with NIST17 GC-MS Spectral Library. All metabolite identifications were manually validated to minimize deconvolution and identification errors during the automated data processing.

## 3. Data Analyses

### 3.1. 16S rRNA Sequence Data Filtering

The number of samples that successfully sequenced for each treatment were: 10 (out of 20) for D-W, 20 (out of 25) for W-D, and 13 (out of 15) for Sat (uneven replicates were the result of limits on soil core collection and sequencing failures). A total of 14,999 OTUs were observed in at least one of the 43 samples. An outlier detected using a randomization test based on pairwise Jaccard distances was removed as was a sample without corresponding metabolite data. Additionally, any OTUs that were observed in less than 2 samples were removed. A total of 4191 such OTUs were removed, leaving 12,649 OTUs in the dataset of 41 samples.

Rarefaction curves for 16S rRNA data for each sample were generated using the R package *vegan* [33], using R programming language (R Core Team, 2018), to assess the sufficiency of depth of sequencing to calculate species richness and inform outlier

removal (Supplementary Figure S1). OTU tables were generated using the *phyloseq* R package [34]. OTUs present were counted after removal of singletons for all downstream analyses. Count data were screened for changes in relative abundance based on upper 75th-quartile normalized library counts (Bullard et al., 2010 [35]). A significant change in relative abundance was considered at  $p\text{-adj.} < 0.05$ . Significance values were corrected for multiple tests using the Benjamini–Hochberg procedure (Dalmasso et al., 2005 [36]). A nonmetric multidimensional scaling (NMDS) plot with all OTUs was used to visualize the data based on Bray–Curtis distance metrics. NMDS was performed using *vegan* package v.2.5.4 [33]. Figures were rendered using *ggplot2* v.3.1.1.

### 3.2. Metabolomics Data Univariate Analysis

A total of 125 metabolites were observed in at least one of the 41 samples. Any metabolites that did not have enough data to conduct quantitative statistical testing (observed in 2 samples for at least 2 of the 3 treatments) were removed and missing observations were replaced with NAs. The data were  $\log_2$  transformed and then normalized via median centering as previously reported [3]. For each metabolite, a mixed-effects linear model was fit to the data assuming a conditional normal distribution. The fixed effect was “moisture”, and a random effect for the soil core was included using *lme4* v1.1-19. All pairwise group comparisons were conducted, and a one-step multiple test adjustment was made to the resulting  $p$ -values to obtain the adjusted  $p$ -values using *multcomp* v1.4-10.

### 3.3. Metabolite Data Processing for Random Forest Classification Approach

A total of 125 metabolites were observed in at least one of the 41 samples. Because random forest classifiers require complete observations, all missing values were imputed by replacing the missing value with half of the lowest observed  $\log_2$  metabolite abundance across all the data (3.2992124). The data were then normalized via median centering. The number of samples for each treatment used for subsequent analysis were: 10 for D-W, 18 for W-D, and 13 for Sat. Filtering down to samples successfully analyzed for metabolites and 16S resulted in both data sources having the same set of 41 samples.

### 3.4. Predicting OTU Presence/Absence with Metabolomic Profiles

To test if metabolite abundance profiles could be used to predict whether an OTU would be observed (i.e., have a nonzero count) or not within a given moisture treatment, we used the following approach: for each OTU, data were converted to binary presence/absence data for each sample, i.e., if an OTU was observed with a non-zero count, it was denoted as present and otherwise absent. A random forest model [23] was fit to the data for each OTU using the *randomForest* R package [37]. The binary outcome of interest was presence/absence of the OTU for each sample, and the normalized  $\log_2$  metabolite abundance values were used as the explanatory variables. The random forest model uses bootstrap aggregating as a validation strategy to protect against overfitting and to fairly evaluate the model’s predictive ability.

OTUs that were present for all or nearly all samples, or were absent for nearly all samples, are not amenable to classification. Moreover, they are not biologically interesting for the purposes of understanding how metabolites might predict the presence/absence of OTUs for individual samples. Thus, we required an OTU to have a minimum of 3 samples where the OTU was observed and a minimum of 3 samples for which the OTU was not observed. The number of OTUs meeting the above requirements were: 4916 for the Sat, 2987 for the D-W, and 5531 for the W-D treatments.

For each random forest model, i.e., each OTU and treatment, the following information was obtained: (i) variable importance metric for each variable, i.e., a metabolite; (ii) predictions (present/absent) for each sample based on the metabolomic abundance profiles; (iii) overall model predictive efficacy. Variable importance metrics were calculated as the mean decrease in the Gini index [22], a metric that can be used across models utilizing the same set of predictors.

We used the balanced accuracy metric (Powers, 2011), the mean of the true positive and true negative prediction rates, to summarize model efficacy for the different moisture treatments (Supplementary Figure S2). Balanced accuracy is used when there is an imbalance in the number of positive and negative labels and is equivalent to accuracy when labels are balanced between the two groups [38]. We set a balanced accuracy threshold greater than or equal to 80% to predict the presence/absence of OTUs to identify models to further investigate.

#### 4. Results and Discussion

We aimed to test whether microbial taxa enrichment under a given soil moisture treatment could be a predictable trait as a function of the observed metabolites, particularly those belonging to classes of complex polysaccharides, long-chain fatty acids, and nitrogenous compounds. Therefore, in this study, we tested for the predictive power of observed metabolites in the ecological context of bacterial population shifts as result of moisture perturbations in wetland soils. We used a random forest classifier model to assess the effectiveness of metabolite data in the prediction of the presence/absence of patterns of microbial taxa obtained from 16S rRNA sequencing. The advantages of this approach include the ability to examine and detect linear and nonlinear relationships between observed metabolites and taxa in a multivariate manner; random forest analyses provide both a metric of how well the presence/absence of an OTU can be predicted and how useful a metabolite is for such a prediction in the presence of other metabolites. It also allowed us to examine patterns across similar taxa, i.e., assuming functional traits, and to make comparisons across treatments. In addition to high predictive performance, random forest classifiers can reveal feature importance [39]. Here, we applied this approach in the context of how the soil bacterial community responds to fluctuating moisture regimes.

##### 4.1. Shifts in Relative Abundances of Bacterial Taxa

The 16S rRNA gene sequences retrieved from all samples corresponded to 32 bacterial phyla-, 129 family-, and 218 genus-level OTUs. Ordination analyses performed using all (Figure 1a) and the top 100 OTUs (Supplementary Figure S2) demonstrated separation of samples by moisture treatments. The relative abundances (>0.5%) of phyla are shown in Figure 1b. Among the dominant phyla based on the mean relative abundance across the three treatments were Acidobacteria, Actinobacteria, Bacteroidetes, Gemmatimonadetes, Latescibacteria, Planctomycetes, Proteobacteria, Verrucomicrobia, and Firmicutes. In the Sat treatment, Acidobacteria, Actinobacteria, Planctomycetes, and Firmicutes were significantly lower in relative abundance while Latescibacteria was higher ( $p\text{-adj} < 0.05$ ) relative to both the W-D and D-W treatments. Organisms in the phylum Latescibacteria are capable of fermentative pathways as well as of oxygen-dependent metabolic reactions and have been recovered from oxygenated and partially/seasonally oxygenated aquatic habitats [40]. Their higher abundance in the Sat treatment is potentially due to this metabolic specialization compared to the other phyla, which are more physiologically diverse and ubiquitous in soils.

Relative shifts in class abundances in response to the moisture fluctuations may be reflective of ecological selection based on their primary respiratory pathways, i.e., aerobic or anaerobic. Acidobacteria Gp2, Acidobacteria Gp7, and Betaproteobacteria were higher in the Sat treatment compared to those in W-D and D-W (Figure 1c).

Acidobacteria\_Gp1, Clostridia (obligate anaerobes), Deltaproteobacteria (aerobic), Gemmatimonadetes (both aerobic and anaerobic), and Sphingobacteria decreased in abundance in response to drying in the W-D samples while correspondingly increasing in response to wetting in the D-W samples. Spartobacteria and Alphaproteobacteria were higher in the D-W samples compared to the Sat samples. The Gemmatimonadetes are found in a wide range of environments and can adapt to low-moisture conditions [41]. Spartobacteria have been observed in diverse ecological niches and display aerobic, facultative anaerobic, or obligate heterotrophic lifestyles [42]. They are capable of utilizing

various C compounds, e.g., cellulose, xylan [43,44], and sugars [45] or methane [46]. We interpret the relative shifts in class abundances in response to the moisture fluctuations to be reflective of ecological selection based on their primary respiratory pathways, i.e., aerobic or anaerobic.

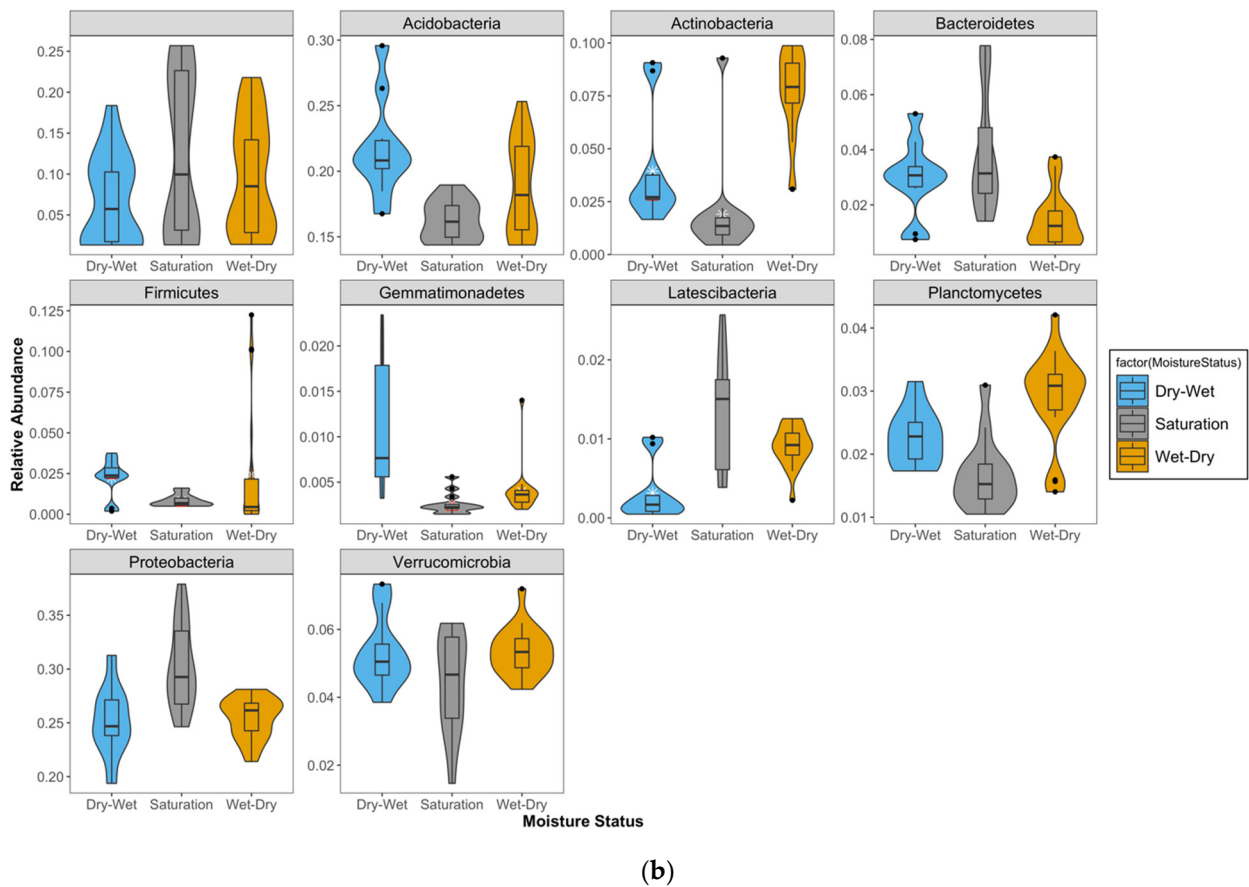
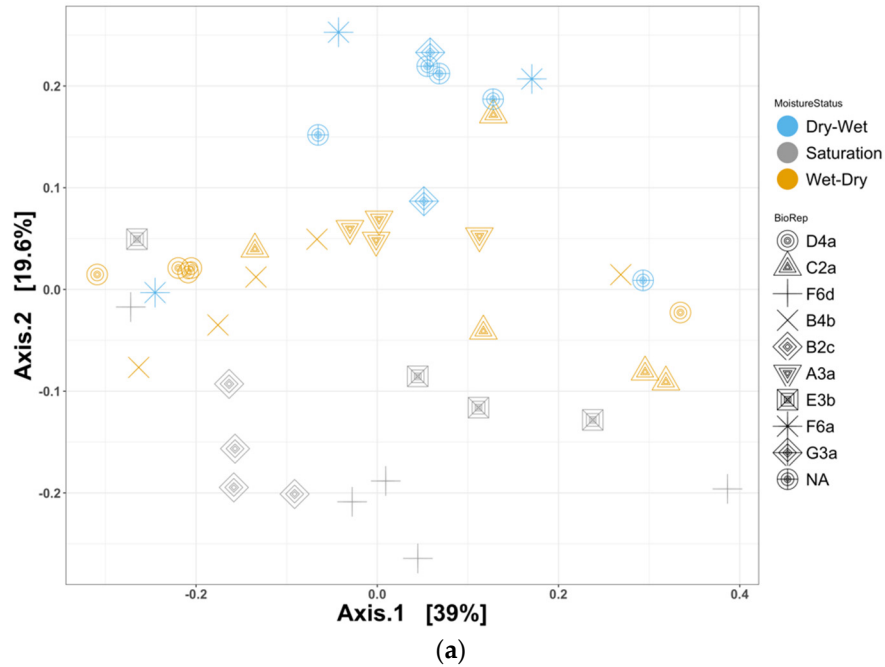
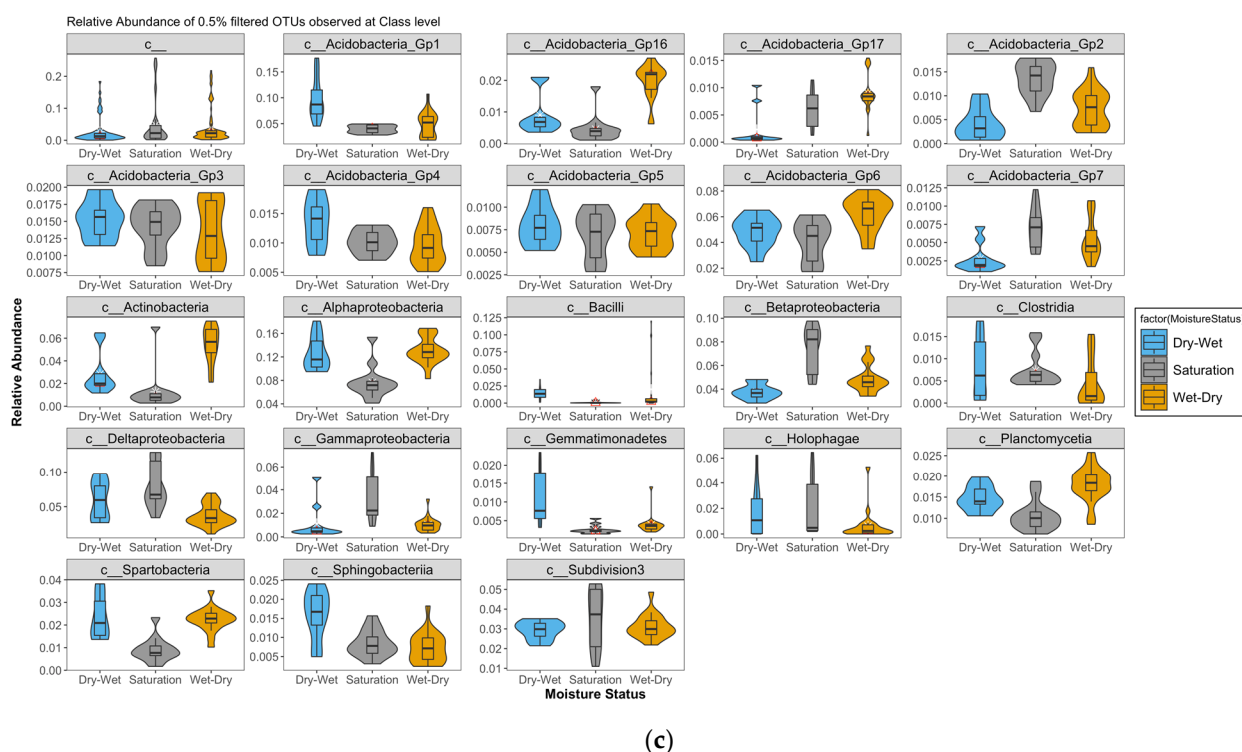


Figure 1. Cont.

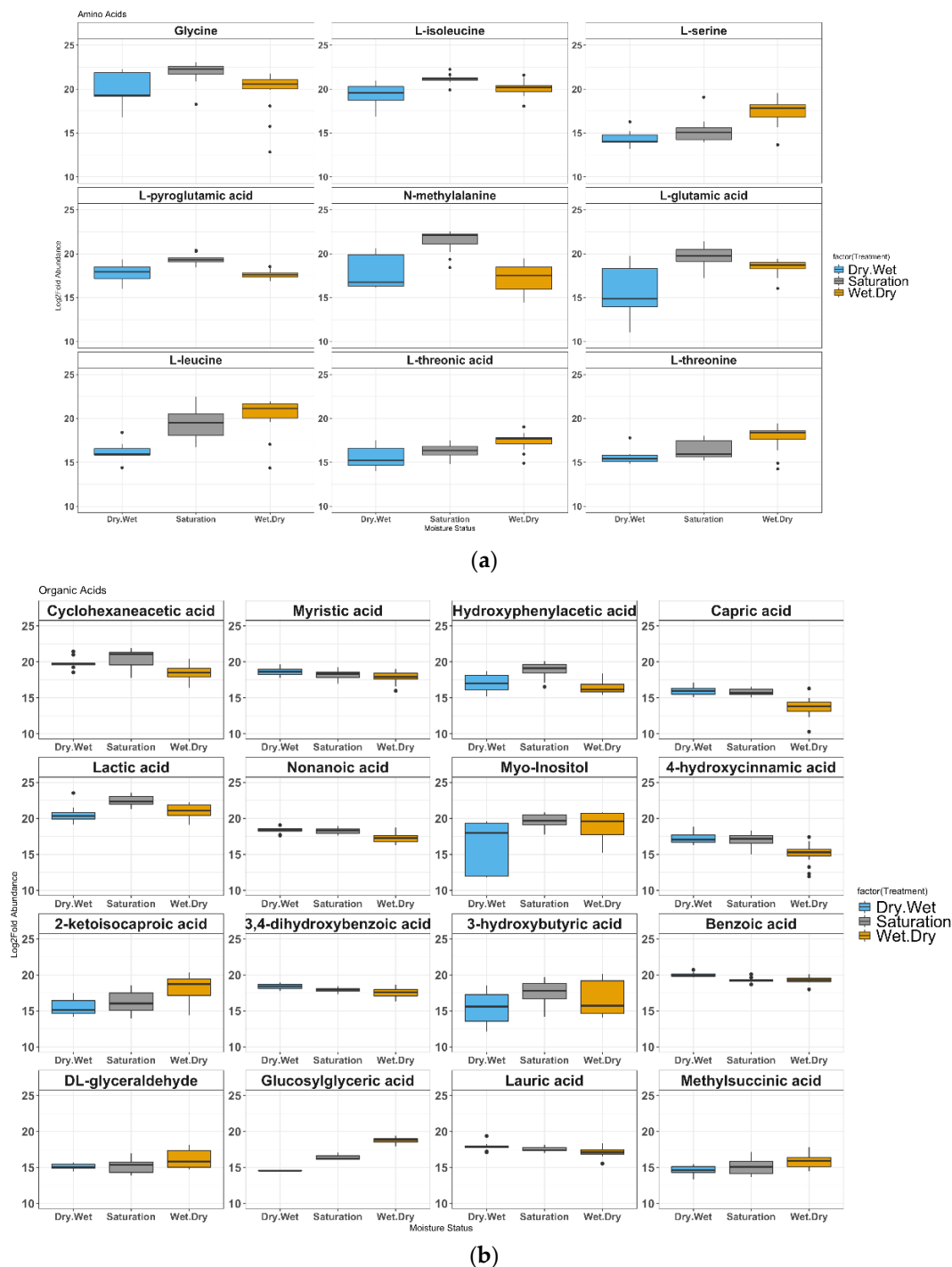


**Figure 1.** (a) Ordination plots showing separation of all OTUs by moisture treatments visualized by nonmetric multidimensional scaling (NMDS) of Bray–Curtis distance metrics. The three moisture treatments are shown in blue (Dry-Wet), grey (Saturation), and orange (Wet-Dry). The biological replicates and their technical replicates are shown by distinct symbols for each moisture treatment and listed under BioRep for their unique experimental IDs. (b) Relative abundance (>0.5%) of OTUs at the phylum level measured for each moisture treatment (Dry-Wet, Wet-Dry, and Saturation). Data are shown as violin plots where whiskers indicate the most extreme values within 1.5 multiplied by the interquartile region. Box, 25% quartile; median, 75% quartile. Pairwise comparisons of means to test treatment effects were performed after outlier removal. (c) Relative abundance (>0.5%) of OTUs at the class level measured for each moisture treatment (Wet-Dry, Dry-Wet, and Saturation). Data are shown as violin plots where whiskers indicate the most extreme values within 1.5 multiplied by the interquartile region. Box, 25% quartile; median, 75% quartile. Pairwise comparisons of means to test treatment effects were performed after outlier removal.

#### 4.2. Metabolite Shifts in Response to Hydrologic Shifts

Drying generally increased abundances of amino acids compared to wetting (Figure 2a). The abundances of nine amino acids shifted in both the W-D and D-W treatments relative to the Sat condition. For example, L-glutamic acid ( $p = 0.015$ ) and L-isoleucine ( $p = 0.038$ ) decreased significantly in the D-W compared to the Sat treatment (Figure 2a). Glycine ( $p = 0.024$ ) decreased while L-serine ( $p = 0.004$ ) and L-threonine ( $p = 0.000$ ) increased in W-D compared to Sat.





**Figure 2.** (a) Impact of soil moisture perturbations on metabolites shown as log<sub>2</sub>-fold abundances of amino acid that shifted significantly in response to a moisture treatment (Dry-Wet, Saturation, or Wet-Dry). Nine amino acids changed significantly under the wet or dry treatments compared to control ( $p < 0.05$ , one-way ANOVA). Whiskers indicate the most extreme values within 1.5 multiplied by the interquartile region. Box, 25% quartile; median, 75% quartile. Pairwise comparisons of means to test treatment effects were performed after outlier removal. (b) Impact of soil moisture perturbations on the soil metabolome shown as log<sub>2</sub>-fold abundances of metabolites (other than amino acids) that shifted significantly in response to a moisture treatment (Dry-Wet, Saturation, or Wet-Dry). Fifteen of 70 detected metabolites changed significantly under the wet or dry treatments compared to control ( $p < 0.05$ , one-way ANOVA). Whiskers indicate the most extreme values within 1.5 multiplied by the interquartile region. Box, 25% quartile; median, 75% quartile. Pairwise comparisons of means to test treatment effects were performed after outlier removal.

*N*-methylalanine (or 2-aminoisobutyric acid) decreased in both W-D ( $p = 0.009$ ) and D-W ( $p = 0.038$ ) relative to Sat. Under conditions of a high ratio of organic C to nitrogen, heterotrophic microorganisms compete with primary producers for the available nitrogen [47]. If much of the nitrogen is present in organic form, it is likely that the ability to use organic nitrogenous compounds as nitrogen sources would confer some advantages to organisms with this ability over those lacking it. Such utilization of ethylamine, beta-alanine, and L-serine as nitrogen but not as sole C sources have been well established. The ability to use methylamines as nitrogen sources by bacteria but not as sources of C has been widely shown in natural habitats [48]. In addition, Methylotrophic bacteria able to utilize methylamines as sole C and energy sources first convert the methyl groups to formaldehyde, which is subsequently either assimilated or oxidized to CO<sub>2</sub> [49,50].

L-pyroglutamic acid decreased in both D-W ( $p = 0.022$ ) and W-D ( $p = 0.007$ ) relative to Sat. Among the amino acids that significantly increased in abundance in W-D compared to D-W were L-glutamic acid ( $p = 0.045$ ), L-leucine ( $p = 0.000$ ), L-serine ( $p = 0.000$ ), L-threonic acid ( $p = 0.037$ ), and L-threonine ( $p = 0.000$ ). Protein amino acids such as glutamic acid, serine, and glycine can play significant roles as constitutive osmolytes or precursors thereof under soil drought [51,52]. The higher abundances of L-serine, L-leucine, L-threonic acid/threonine, and glutamic acid in the W-D treatment could be a stress response or simply accumulation of these small molecules in the soil exometabolome.

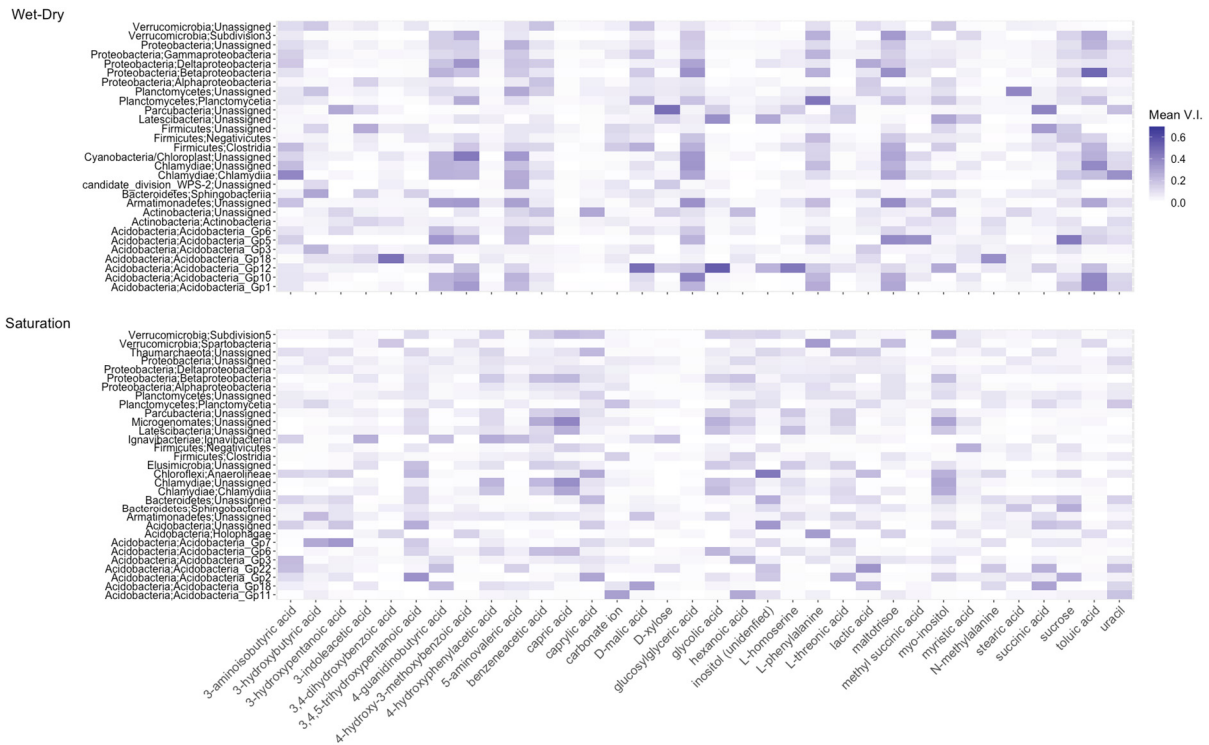
Higher relative abundances of methylalanine and glycine in the Sat treatment could indicate stimulation of pathways related to methylamine synthesis or degradation. Methylamines are produced from glycine and other amino acids (from protein or peptidoglycan), compatible solutes, and lipids. Expression of glycine-cleavage transcripts in methanogenic environments [53,54] and methanogenesis from methylamines by methanogens [55] and methylotrophs [56] have only been recently reported. Therefore, the higher abundance of methylalanine indicates conditions conducive to methanogenesis in the Sat treatment.

Organic acids such as benzoic acid, phenylacetate, and capric and lauric acid have been observed in anaerobic systems and have also been implicated in the methanogenic degradation of lignin derivatives [57]. Thirteen of the measured organic acids decreased in abundance in both D-W and W-D when compared to Sat (Figure 2b). Several organic acids increased in W-D compared to D-W, including those of 2-ketoisocaproic acid ( $p = 0.05$ ), 3-hydroxybutyric acid ( $p = 0.048$ ), DL-glyceraldehyde ( $p = 0.020$ ), glycosylglyceric acid ( $p = 0.006$ ), and methylsuccinic acid ( $p = 0.035$ ). The organic acids that decreased in W-D compared to D-W included 3,4-hydroxybenzoic acid ( $p = 0.019$ ), 4-hydroxycinnamic acid ( $p = 0.000$ ), benzoic acid ( $p = 0.006$ ), lauric acid ( $p = 0.001$ ), myristic acid ( $p = 0.005$ ), and nonanoic acid ( $p = 0.018$ ). A previously reported accumulation of short-chain fatty acids such as formate, isobutyrate, isovalerate, propionate, and pyruvate in the W-D samples occurred during the antecedent wet phase when anaerobic processes were dominant [22]. Therefore, the redox-active metabolites from fermentation and homoacetogenesis could be the precursors for active methanogenesis.

#### 4.3. Metabolites Predict Taxa

We rank metabolites based on their variable importance metric and present those found to be the strongest predictors of OTUs under each moisture condition (Figure 3). These were metabolites that had a variable importance score in the top 20% of variable importance scores across all models. Among these 34 top ranked metabolites, 24 were organic acids, 4 were amino acids, 3 were sugars, 1 was a carbonate ion, 1 was inositol, and 1 was uracil. Collectively, these metabolites were among the strongest predictors of 31 OTUs for the Sat treatment and 29 OTUs for the W-D treatment. However, we observed strong variabilities in the strength of metabolite–OTU predictability as a function of the moisture condition in question. For example, 3,4,5-trihydroxypentanoic acid, capric acid, and glycolic acid were among the strong predictors of the observed OTUs under the Sat condition. In contrast, 4-guanidinobutyric acid, 4-hydroxy-3-methoxybenzoic acid, 5-aminovaleric acid, D-xylose, glycosylglyceric acid, L-phenylalanine, maltotriose, sucrose,

and toluic acid were strong predictors of OTUs under the W-D condition. Surprisingly, none of the metabolites had sufficient predictive power for the OTUs observed under the D-W condition (Supplementary Figure S4).



**Figure 3.** Metabolites ranked based on their mean variable importance for predicting OTUs under Wet-Dry (**top**) and Saturation (**bottom**) treatments.

In the W-D treatment, 4-guanidinobutyric acid, 4-hydroxy-3-methoxybenzoic acid, and glucosylglyceric acid strongly predicted the presence of Acidobacteria (Gp1, Gp10, Gp12, Gp5, and Gp6), Armatimonadetes, Planctomycetes, Deltaproteobacteria, and Subdivision3. D-xylose was a strong predictor of Parcubacteria, while sucrose was the strongest predictor of Acidobacteria\_Gp5 among others from this phylum. Succinic acid also strongly predicted the presence of Parcubacteria and Firmicutes under the W-D condition and that of Acidobacteria\_Gp22 and Gp18 under the Sat condition. Toluic acid ranked among the strongest predictors of Proteobacteria (Betaproteobacteria), Chlamydiae, Armatimonadetes, and Acidobacteria (Gp12, Gp10, and Gp1).

Our results show the high predictive power of D-xylose and sucrose for several members of the phylum Acidobacteria under the W-D treatment. Acidobacteria is a highly diverse phylum and is capable of metabolizing a wide range of carbohydrates as C sources, ranging from simple sugars to complex substrates such as D-glucose, D-xylose, and other polysaccharides [58]. Maltotriose, another carbohydrate, was also among the strong predictors of Acidobacteria although its metabolism is not well documented. These observations are consistent with previous studies that showed positive correlations between Acidobacteria subgroups and gene families related to C degradation, specifically those involved in hemicellulose degradation. For example, subgroup Gp4 has been described as the most versatile in the use of C from sucrose, maltose, and chitins [59,60]. It is important to note that much of our understanding of C degradation by Acidobacteria in soil is limited to correlative studies. The lack of isolates of most Acidobacteria subgroups is a challenge due to discrepancies between information from genomes of cultured representatives and that from marker gene assays.

In the Sat treatment, capric acid was consistently a strong predictor of Verrucomicrobia (Subdivision 5), Proteobacteria (betaproteobacteria), Firmicutes (Negativicutes), Chlamy-

diae (Chlamydia), and Acidobacteria Gp6. Myo-inositol (a polyol) also had sufficient predictive power for many taxa including Verrucomicrobia Subdivision 5, Betaproteobacteria, Chloroflexi (Anaerolineae), Chlamydiae, and Acidobacteria Gp22. 3-aminoisobutyric acid and N-methylalanine, the methylamines that occurred in higher abundances in the Sat treatment, did not rank very high in predictive power for most OTUs, but in the W-D treatment were strong predictors of Firmicutes (Clostridia), Chlamydiae, and Acidobacteria\_Gp18. None of the other amino acids ranked high enough based on our criterion for metabolite of predictive importance, although their relative abundances shifted significantly with wetting and drying. Glucosylglyceric acid, significantly increased in abundance in the W-D treatment, has characteristics of compatible solutes under combined salt stress and nitrogen-limiting conditions and has been observed in prokaryotes across a wide taxonomic distribution [61]. Enrichment of the Latescibacteria in Sat treatment and their high predictability by capric acid, benzenoacetic acid, glycolic acid, polyols such as inositol and myo-inositols, and myristic acid support the diverse metabolic capacities of this phylum as also corroborated by recent reports of their biogeography [62].

The conversion of CO<sub>2</sub> back to acetate by homoacetogenesis is a strictly anaerobic process; homoacetogenesis was shown to be preferred over methanogenesis in profundal sediments of Lake Constance [63], and anaerobic peatland soils [64,65] notwithstanding the predominance in low-temperature ecosystems. The higher relative abundances of Holophagae in the D-W and Sat treatments (Figure 1c) might be an indicator of homoacetogenesis as an active process that could potentially outcompete methanogenesis, as previously reported in swamp peats at pH 6.5 [64]. One member of the Holophagales order, *Holophoga foetida*, can also degrade aromatic compounds [66], an observation supported by our measurements of higher abundances of lactic acid, hydroxyphenylacetic, and 3-hydroxybutyric and methylsuccinic acids under saturated conditions (Figure 2b). All these acids have been implicated in anaerobic lignin degradation pathways [57]. Based on these observations, we hypothesize that homoacetogenesis is an active pathway in these soil systems.

The presence of *Clostridia* under the wet and saturated conditions may explain the high levels of isopropanol and acetone accumulation reported previously [22]. The metabolic pathway of *Clostridia* remains poorly understood but recent studies suggest they may use pyruvate for biosynthetic reactions [67]. Elevated concentrations of acetate or butyrate, or both as we previously observed [22], have been suggested to trigger acetone production in *C. acetobutylicum* [68]. Close interactions between *Methanosarcina* and cellulolytic *Clostridia* have also been reported in freshwater methanogenic environments [69]. The combined observations of high concentrations of acetate and butyrate [22] and the strong predictive power of 3-aminoisobutyric acid from this study may support the co-occurrence of methanogenesis and fermentation processes upon rewetting soils.

The spatiotemporal variability in hydrological fluctuations in wetlands causes the development of temporal and spatial redox gradients [70–73], such that the subsequent manifestation of microbial processes is a function of the availability of terminal electron acceptors and metabolic substrates and products. Here, we provide evidence for shifts in microbial taxa and metabolite abundance in response to moisture shifts as a proxy for redox. Specifically, we find enrichment of drought-tolerant microbial taxa and sugars as metabolites upon drying a previously saturated soil; in addition, an increasing abundance of fatty acids with high predictive power for anaerobes such as *Clostridia*, *Holophagales*, and *Latescibacteria* under saturated conditions suggest a predominance of anaerobic C cycling. We, therefore, infer that present soil redox status may enrich for metabolite classes linked to anaerobic or aerobic metabolism, which, in turn, may predict the dominant microbial community structure in tidal wetlands under fluctuating moisture conditions.

## 5. Conclusions

Soil moisture legacy and redox metabolism are intimately linked, yet it is difficult to discern how these soil conditions drive microbiome shifts within a predictive framework. Our results illustrate the potential of untargeted statistical approaches such as Random

Forest models to interrogate complex soil microbiomes such as wetland sediments using multidimensional data such as metabolomics and the 16S rRNA gene-based community structure. The predictive power to consider discrete datasets for microbial population and/or substrate characteristics for wetland biogeochemistry, collectively and in relation, has previously been limited to correlation analyses. Elucidation of the chemical composition of microbial substrates such as fatty acids, polysaccharides and sugars, and amino acids and their dynamics in response to environmental factors is a critical step toward a predictive understanding of the complex relationship between soil chemical environment and microbial communities. Here, we demonstrate that taking a multidimensional data analysis approach to the interpretation of metabolomics has the potential to unravel the causative pathways selecting for the observed community structure. For example, the dominance of sugars and osmolytes in response to drought stress and that of organic acids under saturation conditions provide evidence of the dynamic nature of metabolome shifts with moisture while enriching for the specialized or facultative microbial taxa. The advantage of our approach is that the random forest model does not make any assumptions about the distribution (e.g., normal distribution) of variable values and can detect nonlinear relationships between predictors and response variables, unlike stepwise regression or correlation-based models that can only detect linear relationships. Such a linear relationship is not relevant to a complex microbial community structure.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/microorganisms10081653/s1>, Figure S1: Rarefaction curves of bacterial assemblages from individual samples subjected to (a) Dry-Wet, (b) Wet-Dry, and (c) Saturation treatments. OTUs were determined by a cutoff value of 97% sequence similarity; Figure S2. Ordination plots showing separation of top 100 OTUs by moisture treatments visualized by nonmetric multidimensional scaling (NMDS) of Bray–Curtis distance metrics. The three moisture treatments are shown in blue (Dry-Wet), grey (Saturation), and orange (Wet-Dry). Figure S3: Histogram showing the balanced accuracy metrics corresponding to the number of OTUs observed in the Dry-Wet, Saturation or Wet-Dry treatments. The red bars and fonts call out the number of OTUs with a balanced accuracy of >0.8 for a given moisture condition; Figure S4: Metabolites ranked based on their mean variable importance for predicting OTUs under Dry-Wet treatment.

**Author Contributions:** Conceptualization, T.R. and V.B.; bioinformatic analysis, J.B. and L.A.M.; 16S rRNA data analysis and visualization, T.R.; metabolomics data generation and analysis, E.Z., Y.-M.K. and T.O.M.; writing—original draft preparation, T.R. and V.B.; writing—review and editing, H.L.D.; random forest analysis and visualization, L.M.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research is part of the Laboratory Directed Research and Development Program at Pacific Northwest National Laboratory (PNNL), a multi-program national laboratory operated by Battelle for the U.S. Department of Energy (U.S. DOE) under Contract DE-AC05-76RL01830. Some data interpretation and writing were supported by COMPASS-FME, a multi-institutional project supported by the U.S. DOE, Office of Science, Biological and Environmental Research as part of the Environmental System Science Program. A portion of this research was performed using EMSL, a DOE Office of Science user facility sponsored by the Department of Energy’s Office of Biological and Environmental Research and located at PNNL.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.

**Acknowledgments:** We are grateful to J. Stegen and A. Borde for their assistance in planning the soil sampling trip and in obtaining the soil cores. We are very thankful to S. Fansler for sequencing the samples.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Barnard, R.L.; Osborne, C.A.; Firestone, M.K. Changing precipitation pattern alters soil microbial community response to wet-up under a Mediterranean-type climate. *Isme J.* **2015**, *9*, 946–957. [[CrossRef](#)] [[PubMed](#)]
2. Kieft, T.; Soroker, E.; Firestone, M. Microbial biomass response to a rapid increase in water potential when dry soil is wetted. *Soil Biol. Biochem.* **1987**, *19*, 119–126. [[CrossRef](#)]
3. RoyChowdhury, T.; Lee, J.-Y.; Bottos, E.M.; Brislawn, C.M.; White, R.A., III; Bramer, L.M.; Brown, J.N.; Zucker, J.D.; Kim, Y.-M.; Jumpponen, A.; et al. Metaphenomic response of a native prairie soil microbiome to moisture perturbations. *mSystems* **2019**, *4*, e00061-19.
4. Xiang, S.; Doyle, A.; Holden, P.; Schimel, J. Drying and rewetting effects on C and N mineralization and microbial activity in surface and subsurface California grassland soils. *Soil Biol. Biochem.* **2008**, *40*, 2281–2289. [[CrossRef](#)]
5. Evans, S.; Wallenstein, M. Soil microbial community response to drying and rewetting stress: Does historical precipitation regime matter? *Biogeochemistry* **2012**, *109*, 101–116. [[CrossRef](#)]
6. Cong, J.; Yang, Y.; Liu, X.; Lu, H.; Liu, X.; Zhou, J.; Li, D.; Yin, H.; Ding, J.; Zhang, Y. Analyses of soil microbial community compositions and functional genes reveal potential consequences of natural forest succession. *Sci. Rep.* **2015**, *5*, 10007. [[CrossRef](#)]
7. Swenson, T.; Karaoz, U.; Swenson, J.; Bowen, B.; Northen, T. Linking soil biology and chemistry in biological soil crust using isolate exometabolomics. *Nat. Commun.* **2018**, *9*, 19. [[CrossRef](#)]
8. Widder, S.; Allen, R.J.; Pfeiffer, T.; Curtis, T.P.; Wiuf, C.; Sloan, W.T.; Cordero, O.X.; Brown, S.P.; Momeni, B.; Shou, W.Y.; et al. Challenges in microbial ecology: Building predictive understanding of community function and dynamics. *Isme J.* **2016**, *10*, 2557–2568. [[CrossRef](#)]
9. Krause, S.; Le Roux, X.; Niklaus, P.A.; Van Bodegom, P.M.; Lennon, J.T.; Bertilsson, S.; Grossart, H.P.; Philippot, L.; Bodelier, P.L.E. Trait-based approaches for understanding microbial biodiversity and ecosystem functioning. *Front. Microbiol.* **2014**, *5*, 251. [[CrossRef](#)]
10. Treseder, K.K.; Balsler, T.C.; Bradford, M.A.; Brodie, E.L.; Dubinsky, E.A.; Eviner, V.T.; Hofmockel, K.S.; Lennon, J.T.; Levine, U.Y.; MacGregor, B.J.; et al. Integrating microbial ecology into ecosystem models: Challenges and priorities. *Biogeochemistry* **2012**, *109*, 7–18. [[CrossRef](#)]
11. DeAngelis, K.; Silver, W.; Thompson, A.; Firestone, M. Microbial communities acclimate to recurring changes in soil redox potential status. *Environ. Microbiol.* **2010**, *12*, 3137–3149. [[CrossRef](#)] [[PubMed](#)]
12. Weber, K.; Urrutia, M.; Churchill, P.; Kukkadapu, R.; Roden, E. Anaerobic redox cycling of iron by freshwater sediment microorganisms. *Environ. Microbiol.* **2006**, *8*, 100–113. [[CrossRef](#)] [[PubMed](#)]
13. Pett-Ridge, J.; Silver, W.; Firestone, M. Redox fluctuations frame microbial community impacts on N-cycling rates in a humid tropical forest soil. *Biogeochemistry* **2006**, *81*, 95–110. [[CrossRef](#)]
14. Pett-Ridge, J.; Firestone, M. Redox fluctuation structures microbial communities in a wet tropical soil. *Appl. Environ. Microbiol.* **2005**, *71*, 6998–7007. [[CrossRef](#)]
15. Swenson, T.L.; Karaoz, U.; Swenson, J.M.; Bowen, B.P.; Northen, T.R. Linking soil biology and chemistry using bacterial isolate exometabolite profiles. *bioRxiv* **2017**, 109330. [[CrossRef](#)]
16. Swenson, T.L.; Jenkins, S.; Bowen, B.P.; Northen, T.R. Untargeted soil metabolomics methods for analysis of extractable organic matter. *Soil Biol. Biochem.* **2015**, *80*, 189–198. [[CrossRef](#)]
17. Szymanski, J.; Jozefczuk, S.; Nikoloski, Z.; Selbig, J.; Nikiforova, V.; Catchpole, G.; Willmitzer, L. Stability of Metabolic Correlations under Changing Environmental Conditions in *Escherichia coli*—A Systems Approach. *PLoS ONE* **2009**, *4*, e7441. [[CrossRef](#)] [[PubMed](#)]
18. Rosato, A.; Tenori, L.; Cascante, M.; Carulla, P.; dos Santos, V.; Saccenti, E. From correlation to causation: Analysis of metabolomics data using systems biology approaches. *Metabolomics* **2018**, *14*, 37. [[CrossRef](#)]
19. van der Heijden, M.; Bardgett, R.; van Straalen, N. The unseen majority: Soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. *Ecol. Lett.* **2008**, *11*, 296–310. [[CrossRef](#)]
20. Guijas, C.; Montenegro-Burke, J.; Warth, B.; Spilker, M.; Siuzdak, G. Metabolomics activity screening for identifying metabolites that modulate phenotype. *Nat. Biotechnol.* **2018**, *36*, 316–320. [[CrossRef](#)]
21. RoyChowdhury, T.; Bramer, L.; Hoyt, D.; Kim, Y.; Metz, T.; Mccue, L.; Diefenderfer, H.; Jansson, J.; Bailey, V. Temporal dynamics of CO<sub>2</sub> and CH<sub>4</sub> loss potentials in response to rapid hydrological shifts in tidal freshwater wetland soils. *Ecol. Eng.* **2018**, *114*, 104–114. [[CrossRef](#)]
22. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
23. Delgado-Baquerizo, M.; Reith, F.; Dennis, P.; Hamonts, K.; Powell, J.; Young, A.; Singh, B.; Bissett, A. Ecological drivers of soil microbial diversity and soil biological networks in the Southern Hemisphere. *Ecology* **2018**, *99*, 583–596. [[CrossRef](#)] [[PubMed](#)]
24. Kauffman, J.B.; Adame, M.F.; Arifanti, V.B.; Schile-Beers, L.M.; Bernardino, A.F.; Bhomia, R.K.; Donato, D.C.; Feller, I.C.; Ferreira, T.O.; Garcia, M.d.J.; et al. Total ecosystem carbon stocks of mangroves across broad global environmental and physical gradients. *Ecol. Monogr.* **2020**, *90*, e01405. [[CrossRef](#)]
25. Thompson, L.R.; Sanders, J.; McDonald, D.; Amir, A.; Ladau, J.; KJ, L.; Prill, R.; Tripathi, A.; Gibbons, S.; Ackermann, G.; et al. The Earth Microbiome Project Consortium- A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* **2017**, *551*, 457–463. [[CrossRef](#)] [[PubMed](#)]
26. Aronesty, E. Comparison of sequencing utility programs. *Open Bioinform. J.* **2013**, *7*, 1–8. [[CrossRef](#)]

27. Bushnell, B. *BBMap: A Fast, Accurate, Splice-Aware Aligner*; Lawrence Berkeley National Laboratory (LBNL): Berkeley, CA, USA, 2014.
28. Edgar, R.C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **2010**, *26*, 2460–2461. [[CrossRef](#)]
29. Cole, J.R.; Wang, Q.; Fish, J.A.; Chai, B.L.; McGarrell, D.M.; Sun, Y.N.; Brown, C.T.; Porras-Alfaro, A.; Kuske, C.R.; Tiedje, J.M. Ribosomal Database Project: Data and tools for high throughput rRNA analysis. *Nucleic Acids Res.* **2014**, *42*, D633–D642. [[CrossRef](#)]
30. Kim, Y.M.; Nowack, S.; Olsen, M.T.; Becraft, E.D.; Wood, J.M.; Thiel, V.; Klapper, I.; Kuhl, M.; Fredrickson, J.K.; Bryant, D.A.; et al. Diel metabolomics analysis of a hot spring chlorophototrophic microbial mat leads to new hypotheses of community member metabolisms. *Front. Microbiol.* **2015**, *6*, 209. [[CrossRef](#)]
31. Hiller, K.; Hangebrauk, J.; Jager, C.; Spura, J.; Schreiber, K.; Schomburg, D. MetaboliteDetector: Comprehensive Analysis Tool for Targeted and Nontargeted GC/MS Based Metabolome Analysis. *Anal. Chem.* **2009**, *81*, 3429–3439. [[CrossRef](#)]
32. Kind, T.; Wohlgemuth, G.; Lee, D.Y.; Lu, Y.; Palazoglu, M.; Shahbaz, S.; Fiehn, O. FiehnLib: Mass Spectral and Retention Index Libraries for Metabolomics Based on Quadrupole and Time-of-Flight Gas Chromatography/Mass Spectrometry. *Anal. Chem.* **2009**, *81*, 10038–10048. [[CrossRef](#)] [[PubMed](#)]
33. Oksanen, J.; Blanchet, F.G.; Kindt, R. *Vegan: Community Ecology Package*; R Package Version 2.3-0; 2015. Available online: [https://www.researchgate.net/publication/311493644\\_Vegan\\_commity\\_ecology\\_package\\_R\\_package\\_version\\_23-0](https://www.researchgate.net/publication/311493644_Vegan_commity_ecology_package_R_package_version_23-0) (accessed on 20 April 2022).
34. McMurdie, P.J.; Holmes, S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE* **2013**, *8*, e61217. [[CrossRef](#)] [[PubMed](#)]
35. Bullard, J.H.; Purdom, E.; Hansen, K.D.; Dudoit, S. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinform.* **2010**, *11*, 94. [[CrossRef](#)]
36. Dalmasso, C.; Broet, P.; Moreau, T. A simple procedure for estimating the false discovery rate. *Bioinformatics* **2005**, *21*, 660–668. [[CrossRef](#)] [[PubMed](#)]
37. Liaw, A.; Wiener, M. Classification and Regression by randomForest. *R News* **2002**, *2*, 18–22.
38. Powers, D.M.W. Evaluation: From Precision, Recall and F-measure to ROC, Informedness, Markedness and Correlation. *J. Mach. Learn. Technol.* **2011**, *2*, 37–63.
39. Basu, S.; Kumbier, K.; Brown, J.; Yu, B. Iterative random forests to discover predictive and stable high-order interactions. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 1943–1948. [[CrossRef](#)]
40. Farag, I.; Youssef, N.; Elshahed, M. Global Distribution Patterns and Pangenomic Diversity of the Candidate Phylum “Latescibacteria” (WS3). *Appl. Environ. Microbiol.* **2017**, *83*, e00521-17. [[CrossRef](#)]
41. DeBruyn, J.; Nixon, L.; Fawaz, M.; Johnson, A.; Radosevich, M. Global Biogeography and Quantitative Seasonal Dynamics of Gemmatimonadetes in Soil. *Appl. Environ. Microbiol.* **2011**, *77*, 6295–6300. [[CrossRef](#)]
42. Herlemann, D.; Lundin, D.; Labrenz, M.; Jurgens, K.; Zheng, Z.; Aspeborg, H.; Andersson, A. Metagenomic De Novo Assembly of an Aquatic Representative of the Verrucomicrobial Class Spartobacteria. *Mbio* **2013**, *4*, e00569-12. [[CrossRef](#)]
43. Chin, K.; Hahn, D.; Hengstmann, U.; Liesack, W.; Janssen, P. Characterization and identification of numerically abundant culturable bacteria from the anoxic bulk soil of rice paddy microcosms. *Appl. Environ. Microbiol.* **1999**, *65*, 5042–5049. [[CrossRef](#)] [[PubMed](#)]
44. Janssen, P.; Schuhmann, A.; Morschel, E.; Rainey, F. Novel anaerobic ultramicrobacteria belonging to the Verrucomicrobiales lineage of bacterial descent isolated by dilution culture from anoxic rice paddy soil. *Appl. Environ. Microbiol.* **1997**, *63*, 1382–1388. [[CrossRef](#)] [[PubMed](#)]
45. Pol, A.; Heijmans, K.; Harhangi, H.; Tedesco, D.; Jetten, M.; den Camp, H. Methanotrophy below pH1 by a new Verrucomicrobia species. *Nature* **2007**, *450*, 874–878. [[CrossRef](#)] [[PubMed](#)]
46. Dunfield, P.; Yuryev, A.; Senin, P.; Smirnova, A.; Stott, M.; Hou, S.; Ly, B.; Saw, J.; Zhou, Z.; Ren, Y.; et al. Methane oxidation by an extremely acidophilic bacterium of the phylum Verrucomicrobia. *Nature* **2007**, *450*, 879–882. [[CrossRef](#)] [[PubMed](#)]
47. Alexander, M. *Introduction to Soil Microbiology*; John Wiley: New York, NY, USA, 1977.
48. Bicknell, B.; Owens, J. Utilization of methyl amines as nitrogen-sources by non-methylotrophs. *J. Gen. Microbiol.* **1980**, *117*, 89–96. [[CrossRef](#)]
49. Anthony, C. Biochemistry of methylotrophic microorganisms. *Sci. Prog.* **1975**, *62*, 167–206.
50. Lin, M.; Wagner, C. Purification and characterization of n-methylalanine dehydrogenase. *J. Biol. Chem.* **1975**, *250*, 3746–3751. [[CrossRef](#)]
51. Warren, C. Do microbial osmolytes or extracellular depolymerisation products accumulate as soil dries? *Soil Biol. Biochem.* **2016**, *98*, 54–63. [[CrossRef](#)]
52. Warren, C.R. Response of osmolytes in soil to drying and rewetting. *Soil Biol. Biochem.* **2014**, *70*, 22–32. [[CrossRef](#)]
53. Tveit, A.; Urich, T.; Frenzel, P.; Svenning, M. Metabolic and trophic interactions modulate methane production by Arctic peat microbiota in response to warming. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E2507–E2516. [[CrossRef](#)]
54. Angle, J.; Morin, T.; Solden, L.; Narrowe, A.; Smith, G.; Borton, M.; Rey-Sanchez, C.; Daly, R.; Mirfenderesgi, G.; Hoyt, D.; et al. Methanogenesis in oxygenated soils is a substantial fraction of wetland methane emissions. *Nat. Commun.* **2017**, *8*, 1567. [[CrossRef](#)] [[PubMed](#)]

55. Nayak, D.; Metcalf, W. Methylamine-specific methyltransferase paralogs in *Methanosarcina* are functionally distinct despite frequent gene conversion. *Isme J.* **2019**, *13*, 2173–2182. [[CrossRef](#)] [[PubMed](#)]
56. Narrowe, A.B.; Borton, M.A.; Hoyt, D.W.; Smith, G.J.; Daly, R.A.; Angle, J.C.; Eder, E.K.; Wong, A.R.; Wolfe, R.A.; Pappas, A.; et al. Uncovering the diversity and activity of methylotrophic methanogens in freshwater wetland soils. *mSystems* **2019**, *4*, e00320-19. [[CrossRef](#)] [[PubMed](#)]
57. Healy, J.; Young, L.; Reinhard, M. Methanogenic decomposition of ferulic acid, a model lignin derivative. *Appl. Environ. Microbiol.* **1980**, *39*, 436–444. [[CrossRef](#)]
58. Kielak, A.; Barreto, C.; Kowalchuk, G.; van Veen, J.; Kuramae, E. The Ecology of Acidobacteria: Moving beyond Genes and Genomes. *Front. Microbiol.* **2016**, *7*, 744. [[CrossRef](#)]
59. Huber, K.; Wust, P.; Rohde, M.; Overmann, J.; Foesel, B. *Aridibacter famidurans* gen. nov., sp nov and *Aridibacter kavangonensis* sp nov., two novel members of subdivision 4 of the *Acidobacteria* isolated from semiarid savannah soil. *Int. J. Syst. Evol. Microbiol.* **2014**, *64*, 1866–1875. [[CrossRef](#)]
60. Lee, S.; Ka, J.; Cho, J. Members of the phylum Acidobacteria are dominant and metabolically active in rhizosphere soil. *Fems Microbiol. Lett.* **2008**, *285*, 263–269. [[CrossRef](#)]
61. Empadinhas, N.; da Costa, M. Osmoadaptation mechanisms in prokaryotes: Distribution of compatible solutes. *Int. Microbiol.* **2008**, *11*, 151–161. [[CrossRef](#)]
62. Song, M.; Peng, W.; Zeng, F.; Du, H.; Peng, Q.; Xu, Q.; Chen, L.; Zhang, F. Spatial Patterns and Drivers of Microbial Taxa in a Karst Broadleaf Forest. *Front. Microbiol.* **2018**, *9*, 1691. [[CrossRef](#)]
63. Schulz, S.; Conrad, R. Influence of temperature on pathways to methane production in the permanently cold profundal sediment of Lake Constance. *Fems Microbiol. Ecol.* **1996**, *20*, 1–14. [[CrossRef](#)]
64. Ye, R.; Jin, Q.; Bohannon, B.; Keller, J.; Bridgham, S. Homoacetogenesis: A potentially underappreciated carbon pathway in peatlands. *Soil Biol. Biochem.* **2014**, *68*, 385–391. [[CrossRef](#)]
65. Zalman, C.; Meade, N.; Chanton, J.; Kostka, J.; Bridgham, S.; Keller, J. Methylotrophic methanogenesis in Sphagnum-dominated peatland soils. *Soil Biol. Biochem.* **2018**, *118*, 156–160. [[CrossRef](#)]
66. Liesack, W.; Bak, F.; Kreft, J.; Stackebrandt, E. *Holophaga foetida* gen-nov, sp-nov, a new, homoacetogenic bacterium degrading methoxylated aromatic-compounds. *Arch. Microbiol.* **1994**, *162*, 85–90. [[CrossRef](#)] [[PubMed](#)]
67. Amador-Noguez, D.; Feng, X.; Fan, J.; Roquet, N.; Rabitz, H.; Rabinowitz, J. Systems-Level Metabolic Flux Profiling Elucidates a Complete, Bifurcated Tricarboxylic Acid Cycle in *Clostridium acetobutylicum*. *J. Bacteriol.* **2010**, *192*, 4452–4461. [[CrossRef](#)] [[PubMed](#)]
68. George, H.; Johnson, J.; Moore, W.; Holdeman, L.; Chen, J. Acetone, isopropanol, and butanol production by *Clostridium-beijerinckii* (syn *Clostridium-butylicum*) and *Clostridium-aurantibutyricum*. *Appl. Environ. Microbiol.* **1983**, *45*, 1160–1163. [[CrossRef](#)]
69. Stams, A. Metabolic interactions between anaerobic-bacteria in methanogenic environments. *Antonie van Leeuwenhoek Int. J. Gen. Mol. Microbiol.* **1994**, *66*, 271–294. [[CrossRef](#)]
70. Conrad, R.; Rothfuss, F. Methane oxidation in the soil surface-layer of a flooded rice field and the effect of ammonium. *Biol. Fertil. Soils* **1991**, *12*, 28–32. [[CrossRef](#)]
71. Frenzel, P.; Rothfuss, F.; Conrad, R. Oxygen profiles and methane turnover in a flooded rice microcosm. *Biol. Fertil. Soils* **1992**, *14*, 84–89. [[CrossRef](#)]
72. Wang, J.; Krause, S.; Muiyzer, G.; Meima-Franke, M.; Laanbroek, H.; Bodelier, P. Spatial patterns of iron- and methane-oxidizing bacterial communities in an irregularly flooded, riparian wetland. *Front. Microbiol.* **2012**, *3*, 64. [[CrossRef](#)]
73. Peralta, A.; Ludmer, S.; Matthews, J.; Kent, A. Bacterial community response to changes in soil redox potential along a moisture gradient in restored wetlands. *Ecol. Eng.* **2014**, *73*, 246–253. [[CrossRef](#)]