*Article*

# Genome-Resolved Metagenomics of Nitrogen Transformations in the Switchgrass Rhizosphere Microbiome on Marginal Lands

**Richard Allen White III[1-2], Aaron Garoutte[5], Emily E. Mclachlan[3-5], Lisa K. Tiemann[7], Sarah Evans[6,8] and Maren L. Friesen[3-5,*]**

[1]Department of Bioinformatics and Genomics, The University of North Carolina at Charlotte, 9201 University City Boulevard, Charlotte, NC 28223, USA
[2]Department of Bioinformatics and Genomics, The University of North Carolina at Charlotte, 150 Research Campus Drive, Kannapolis, NC 28081, USA
[3]Department of Plant Pathology, Washington State University, Pullman, Washington USA
[4]Department of Crop and Soil Sciences, Washington State University, Pullman, Washington USA
[5]Department of Plant Biology, Michigan State University, East Lansing, Michigan, USA
[6]Department of Integrative Biology, Michigan State University, East Lansing, Michigan, USA
[7]Department of Plant, Soil and Microbial Sciences, Michigan State University, East Lansing, Michigan, USA
[8]W. K. Kellogg Biological Station, Michigan State University, East Lansing, Michigan, USA
***Correspondence:** Maren L. Friesen, Email: m.friesen@wsu.edu

**Abstract:** Switchgrass (*Panicum virgatum* L.) remains the preeminent American perennial (C4) bioenergy crop for cellulosic ethanol that could help displace over a quarter of the US current petroleum consumption. Intriguingly, there is often little response to nitrogen fertilizer once stands are established. The rhizosphere microbiome plays a critical role in nitrogen cycling and overall plant nutrient uptake. We used high-throughput metagenomic sequencing to characterize the switchgrass rhizosphere microbial community before and after a nitrogen fertilization event for established stands on marginal land. We examined community structure, bulk metabolic potential, and resolved 29 individual bacteria genomes via metagenomic *de novo* assembly. Community structure and diversity were not significantly different before and after fertilization; however, the bulk metabolic potential of carbohydrate-active enzymes was depleted after fertilization. We resolved 29 metagenomic assembled genomes including some from the 'most wanted' soil taxa such as Verrucomicrobia, Candidate phyla UBA10199, Acidobacteria (rare subgroup 23), Dormibacterota, and the very rare Candidatus Eisenbacteria. The Dormibacterota (formally candidate division AD3) we identified have the potential for autotrophic CO utilization, which may impact carbon partitioning and storage. Our study also suggests that the rhizosphere microbiome may be involved in providing associative nitrogen fixation (ANF) via the novel diazotroph *Janthinobacterium*, which may partially explain why switchgrass growth is insensitive to fertilizer.

**Keywords:** rhizosphere; phyllosphere; metagenomics; microbiome; nutrient cycling; metagenomic assembled-genomes (MAGs); nitrogen-fixation; nitrogen

## 1. Introduction

"Plants wear their guts on the outside" wrote Janzen (1985), since the rhizosphere of terrestrial plants–the ~millimeter interface between plant roots and surrounding soil–plays critical roles in nutrient uptake, absorption, degradation via the diverse microbes it contains (Berendsen et al. 2012; Ramírez-Puebla et al. 2012; White III et al. 2017ab). The rhizosphere connects plants to ecosystem processes including the cycling and sequestration of water, nitrogen (N), carbon and other nutrients (Ahkami et al. 2017). The rhizosphere represents one of the most dynamic and diverse interfaces on the planet, containing up to 1011 microbial cells per gram root, potentially representing over 30,000 bacterial species that also interact with fungi, picoeukaryotes, bacteriophage, and viruses

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 March 2023 doi:10.20944/preprints202303.0168.v1

(Berendsen et al. 2012; White III et al. 2017ab). The rhizosphere microbiome can alter the physical and chemical environment of plants by directly promoting plant growth via nutrient fixation (e.g., N fixation), increase bioavailability of soil nutrients (e.g., phosphorus, iron, zinc, and copper), and alter plant hormones and signaling (Ahkami et al. 2017; Friesen et al. 2012). Hence, the rhizosphere represents a critical interface of plant-microbial interactions that directly impacts plant and soil health and harnessing it requires knowledge about the identities and functionality of rhizosphere microbial communities

Metagenomics of soil and rhizosphere can provide a direct measure of its metabolic capabilities, functional potential, and the genomes of individual members through 'metagenome assembled genomes' (MAGs) (Bowers et al., 2017; White III et al. 2017ab). Soil and rhizosphere ecosystems have been considered the 'grand challenge' of metagenomics due to the low coverage of individual organisms, uneven sampling, high genetic diversity, and large amounts of sequence data required (Howe et al. 2014; White III et al. 2016). In general, metagenomics in these environments yield poor de novo assemblies, because often up to 80% of the data cannot be assembled, there is typically low read map coverage (< 20%) and few contigs >8 kbp (< 10 contigs) (Howe et al. 2014; White III et al. 2016). Until recently, obtaining MAGs from soils was deemed impossible; however, three studies have been able to resolve MAGs directly without amendment in soil ecosystems (Butterfield et al. 2016; White III et al. 2016; Kroeger et al. 2018). MAGs and genome-resolved metagenomics have provided a wealth of knowledge on the vast candidate bacterial phyla and their metabolic potential and functions that have never before been described due to their unculturability (Wrighton et al. 2012; Kantor et al. 2013).

The activity of the rhizosphere microbiome, including community assembly, recruitment, uptake and degradation of nutrients are driven by plant root exudates (Philippot et al. 2013; Zhalnina et al. 2018). Greater than a century of research into the rhizosphere has revealed "the rhizosphere effect", by which plants enhance the growth of soil microbes via the exudation of organic molecules, particularly the carbon compounds exuded (Hartmann et al. 2008). This carbon fuels the metabolic processes of the rhizosphere microbiome, of which the nitrogen cycle is absolutely critical to plant growth. Carbon inputs directly impact nitrogen fixation, which is an energetically expensive process (Smercina et al. 2019). Carbon can also stimulate denitrification in parallel with nitrogen fixation, resulting in a net loss of N (deCatanzaro and Beauchamp, 1985). Denitrification stimulation occurs more strongly with simple substrates (e.g., glucose) versus complex substrates (e.g., cellulose and lignin) (deCatanzaro and Beauchamp, 1985). Understanding how the rhizosphere microbiome utilizes carbohydrates via carbohydrate-active enzymes (CAZy) can further elucidate the balance and interaction between C and N cycling.

The rhizosphere microbiome is primarily responsible for nitrogen transformations in soil. The nitrogen cycle consists of assimilation, fixation, denitrification, nitrate reduction, nitrification, anaerobic ammonium oxidation (ANAMMOX), dissimilatory nitrate reduction to ammonium (DNRA), and complete ammonium oxidation (COMMAMOX) (Pajares and Bohannan, 2016). Biological nitrogen fixation (BNF) is the reduction of atmospheric molecular nitrogen [N2] to ammonia [NH3] via nitrogenase (encoded by the nifHDK gene cluster); this reaction accounts for approximately two-thirds of the fixed nitrogen available to biology on the planet (Masson-Boivin et al. 2009). Soil denitrification occurs via three mechanisms: 1) nitrite [NO2-] to molecular nitrogen [N2] via dissimilatory nitrite reductase nirKS gene cluster, 2) nitric oxide [NO] reduction to molecular nitrogen [N2] via the norB nitric oxide reductase gene, and 3) nitrous [N2O] reduction to molecular nitrogen [N2] via the nitrous oxide reductase nosZ gene (Demanèche et al. 2009). Soil nitrate reduction is the conversion of nitrate [NO3-] to nitrite [NO2-] via the napA/narG nitrate reductase genes. DNRA occurs via the transformation of nitrite [NO2-] to ammonia [NH3] via the nrfA nitrite reductase gene (Giles et al. 2018). ANAMMOX reaction converts nitrate [NO3-] to ammonia [NH3] which occurs via the hzo hydrazine oxidoreductase gene (Koch et al. 2012). COMMAMOX converts to am-

 doi:10.20944/preprints202303.0168.v1

monia [NH3] to nitrate [NO3-] requires the amoABC ammonium oxidase gene cluster, the hao hydroxylamine oxidoreductase gene and nxr nitrite oxidoreductase gene (Koch et al. 2018). The abundance and diversity of nitrogen cycling genes and their connection to carbohydrate utilization genes provides a window into the coupling of the C and N cycles within the rhizosphere.

Switchgrass (*Panicum virgatum* L.) is the principal United States bioenergy model C4 perennial crop for use in cellulosic ethanol production, biogas, and combustion (McLaughlin and Kzos, 2005), which could displace up to 30% of the current petroleum consumption (Schmer et al. 2008). Its high biomass productivity in low-nutrient soils common to marginal lands (Wright, 2007; Wright and Turhollow, 2010; Ruan et al. 2016; Emery et al. 2017) is key because growth on marginal land avoids competition with food crops on arable lands (Wright, 2007; Ruan et al. 2016; Emery et al. 2017). Switchgrass and other cellulosic ethanol sources could displace up to 30% of the current petroleum consumption (Schmer et al. 2008).

Contrary to many cropping systems which show N limitation, N fertilization of switchgrass often has limited to no impact on productivity (Duran et al. 2016; Ruan et al. 2018) and there is a resulting gap in the N budget (Ruan et al. 2018). This suggests the rhizosphere microbiome could contribute to the N requirement (Singer et al. 2019), through free-living diazotrophic bacteria (Roley et al. 2018; Roley et al. 2019; Smercina et al. 2019).

Studies to date have investigated the effect of N fertilizer on the switchgrass rhizosphere microbiome using 16S rRNA amplicon sequencing (Chen et al. 2019; Roley et al. 2018; Singer et al. 2019). 16S rRNA amplicon studies typically resolve fail to resolve novel 'candidate' phyla representing vast portions of the tree of life (Rinke et al. 2013; Hug et al. 2016). These 16S rRNA amplicon approaches fail to measure metabolic capabilities directly, functional potential, gene-gene assortment on operons via long-range sequence contiguity, gene transfer via horizontal gene transfer (HGT), or provide MAGs. Shotgun metagenomics thus provides a powerful approach to characterize both diversity and functionality simultaneously (Milanese et al. 2019).

Here, we use high-throughput metagenomic shotgun sequencing of the switchgrass rhizosphere on marginal lands to resolve its taxonomic composition, functional metabolic potential, and resolve individual genomes (MAGs). We further compare plots pre- and post-fertilization to determine how the microbiome responds to N addition and give insight into its responsiveness over a two week interval. We first assess metagenome quality and overall diversity across our study. We next describe overall metabolic potential and shifts between timepoints, delving particularly into nitrogen cycling and CAZy. Finally, we describe MAGs of abundant bacteria and use these to elucidate their potential roles in nitrogen cycling and coupling the N and C cycles in the rhizosphere.

## 2. Materials and Methods

*Study site description and management*

Switchgrass rhizosphere soil was sampled (April to May 2016) from the Lux Arbor reserve (42.48 N -85.44 W) as part of the Great Lake Bioenergy Research Center's (GLBRC) marginal land experiments. The Lux Arbor reserve marginal land soil is a sandy clay loam, mesic Hapludalf (Crum and Collins, 1995). The mean annual temperature (MAT) is 10.1 ºC with a mean annual precipitation of 1.005 m. Four blocks of switchgrass rhizosphere soil from the Cave-in-rock variety were sampled in a randomized block design after a two-week pulse treatment of N amendment. The assigned blocks were 64 ft x 40 ft in size. Amended soil is approximately half a plot with the other half receiving no amendment. The soil is classified as a Kalamazoo loam with a 2-6% slope. Blocking diagram and map of the location is provided in supplemental materials (**Supplemental Fig. S1**).

The plots were amended with pelleted lime (454 kg/A), urea (53 kg/A or 24kg/A N) on April 5th, 2016 and then again on May 13th, 2016. The urea fertilizer was SUPERU®

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 March 2023 doi:10.20944/preprints202303.0168.v1

(Koch Agronomic Services - KAS, Wichita, KS) brand which is 45.5% urea nitrogen, contains 0.06% (600ppm) N-(n-butyl) thiophosphoric triamide (NBPT) a urease inhibitor, and 0.85 % (8500 ppm) Dicyandiamide (DCD) a nitrification inhibitor.

We collected soil cores (2 cm diameter × 15 cm deep) from near the centers of each subplot, ~50 g of soil per core, one sample was taken per subplot. Cores from each subplot on each date were pooled, sieved through a 4 mm sieve to remove rocks and large roots, and then flash frozen liquid nitrogen until processed. Rhizosphere is typically considered to be the zone of soil that is influenced by roots, so given that these soil cores were sampled at the base of a well-established perennial plant and that the cores themselves contained a sizeable amount of root material, we consider these samples to be indicative of the switchgrass rhizosphere. The "tightly bound" rhizosphere and rhizoplane are typically defined operationally by the particles of soil that are stuck to a root after shaking and the surface of a root after all soil particles are removed, respectively. We did not seek to partition these compartments, but note that the tightly bound and rhizoplane microbes are captured in our samples

*DNA extraction and sequencing*

Total DNA was extracted from ~2 g switchgrass rhizosphere from field flash frozen samples using the MoBio PowerSoil DNA (Carlsbad, CA, USA) according to the manufacturer's instructions. Samples were quantified using the Qubit Fluorometer 2.0 (Invitrogen, Carlsbad, CA, USA), quality checked using a Nanodrop-1000 (Thermo Fisher, Waltham, MA, USA). Michigan State University Research Technology Support Facility (RTSF) sequencing core completed Illumina library preparation, library quantification, and sequenced on HiSeq 4000 150 bp paired-end read format.

We analyzed eight metagenomes from switchgrass rhizosphere within marginal lands of southern Michigan Lux Arbor Reserve, comprising 4 plots sampled pre- and post-fertilizer application. Samples that were taken pre-fertilization (i.e., without N fertilizer) will be called I1 to I4 (or pre-fertilized plots 1 to 4). Post-fertilized samples that received N amendment in the form of urea fertilizer will be called P1 to P4 (or post-fertilized plots 1 to 4).

*Metagenomic assembly, annotation, differential abundance statistical analysis and genome reconstruction*

Paired-end shotgun reads were quality filtered, assembled, and decontaminated using ATLAS (White III et al. 2017c). In short, the bbduk module quality filtered, trimmed, and decontaminated for φX174 phage DNA, a common Illumina sequencing spike-in, and for all Illumina adapters. Metagenomic *de novo* assembly was performed using Megahit (k-mer 21 - 121, version 1.1.3) (Li et al. 2015). We used only >5 kbp contigs only for all downstream analysis; this includes taxonomic and functional annotation, and metagenomic binning. Protein-coding open reading frames (ORFs) and RNA prediction were completed using Prokka (Seemann, 2014). Eggnog-mapper (--diamond mode, version 0.8.22.84) was used to obtain updated KEGG (KO) numbers for MAGs and contigs (Huerta-Cepas et al. 2017). CAZy predictions were completed in diamond (version 0.8.22.84) (Buchfink et al. 2015) for MAGs and contigs (July 31, 2018, database update) from dbcan2 (Zhang et al. 2018).

DESeq2 R package (version 1.18.1) (Love et al. 2014) was used to obtain differential statistics on taxonomic composition and functional annotations from predicted ORFs from contigs using COG, CAZy and KO annotation abundances. The contig orf counts abundances within DESeq2 followed a paired analysis which blocked by sampling plot pre/post-fertilization, then normalized with variance stabilizing normalization.

We pooled contigs from pre/post-fertilization then used the differential read abundances across all samples to obtain MAGs. Contigs were binned using Concoct (Alneberg et al. 2014), Maxbin2 (Wu et al. 2016), and Metabat2 (Kang et al. 2019) including the refinement program within metawrap (Uritskiy et al. 2018). Metawrap was used for re-

 doi:10.20944/preprints202303.0168.v1

finement of MAGs, blobiology prediction, and quantification of bin (MAG) abundance via quant_bin module (Uritskiy et al. 2018). CheckM was used to evaluate completeness, contamination, redundancy, and genome properties of the MAGs (Parks et al. 2015). All MAG qualities were reported as per was reported according to the MIMAG standards (Bowers et al. 2017). We tested a variety of methods to resolve the taxonomy of the MAGs including metawrap's classifier, classify genomes (https://github.com/AlessioMilanese/classify-genomes) which uses the metagenomic operational taxonomic units (mOTU) v2 taxonomy, JSpeciesWS Tetra Correlation Search (Richter et al. 2016), GTDB-Tk (https://github.com/Ecogenomics/GTDBTk) and blastp of ribosomal protein S9 gene (obtained from prokka annotation). Only GTDB-Tk provided the taxonomic identifications that were supported by ribosomal protein S9 gene blastp results. GTDB-Tk provided all further taxonomy for the MAGs downstream.

*Read based mOTU picking and statistical analysis*

mOTUs analysis of the quality filtered and decontaminated reads (not contigs) used in the *de novo* metagenomic assembly were used as input mOTUs v2 (Milanese et al. 2019), then parsed and further analyzed with the phyloseq (version 1.22.3) R package (McMurdie et al. 2013). Alpha-diversity measurements with statistical testing for mOTUs (including t-test, Wilcox, Kruskal, and anova) were completed in the phyloseq (version 1.22.3) R package. Beta-diversity metrics were obtained for mOTUs using the UniFrac (weighted/unweighted) and Brays-Curtis distances in phyloseq (version 1.22.3) R package. DESeq2 R package (version 1.18.1) was using within phyloseq (version 1.22.3) R package for mOTU differential statistics using negative binomial distribution corrected with variance stabilizing normalization.

*Data and analysis code availability*

Raw sequence data, assembled contigs, supplemental data, are all available on https://osf.io/mzrvj/. All code for this study is available on www.github.com/friesenlab/MMPRNT_panicum_metagenome_mags/.

## 3. Results

*Assessment of assembly and metagenomic assembled genomes within Lux Arbor*

The raw data represents 5.37 billion Illumina reads with 805 Gbp in 490 gigabytes of compressed data with ~100 Gigabytes of uncompressed data per sample (Table 1). On average, 34.5% of the data was removed per sample due to quality, length, adapters, or were phiX174 bacteriophage Illumina spike-in DNA library (Table 1). Upon metagenomic *de novo* assembly with MEGAHIT, each sample averaged 4.6 million total contigs (>200 bp) contained, on average, 3.2 Gbp with an average N50 of 737 bp (Table 1). However, the best assessment of a soil and rhizosphere metagenome *de novo* assembly is how many contigs are >1 kbp and are longer than 5 kbp (Howe et al. 2014; White III et al. 2016). On average, ~60,000 contigs per sample were contained on contigs >1 kbp with an average of 2.21 Gbp assembly size and an average N50 value of 1,982 bp (Table 1). Across all samples we obtained 190,172 contigs >5 kbp with an average of 23,771 contigs of 5 kbp per sample on a 1.19 Gbp assembly size with an average N50 value of 9,254 bp (Table 1). Of those 190,172 >5 kbp contigs, 44,171 of them were >10 kbp in length and 237 were >100 Kbps. Max contig length was 697,599 across all samples.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 March 2023 doi:10.20944/preprints202303.0168.v1

**Table 1.** Metagenome assembly statistics with pre- and post-fertilization processing read counts.

| sample_ID | Sample_dis | total reads | number of raw bases | total trim-decon reads | No_contigs | Size | N50 |
|---|---|---|---|---|---|---|---|
| P1 | Post-fertilization | 678370656 | 101755598400 | 459843606 | 4362794 | 2620101511 | 607 |
| P2 | Post-fertilization | 749126640 | 112368996000 | 498478482 | 5542436 | 3877662650 | 743 |
| P3 | Post-fertilization | 663645284 | 99546792600 | 447352834 | 4946258 | 3305489932 | 701 |
| P4 | Post-fertilization | 547683338 | 82152500700 | 341287068 | 3839925 | 2659609828 | 736 |
| I1 | Pre-fertilization | 677015916 | 101552387400 | 448049362 | 4829012 | 3418287128 | 754 |
| I2 | Pre-fertilization | 644655868 | 96698380200 | 443420266 | 4892751 | 3349942551 | 722 |
| I3 | Pre-fertilization | 717346388 | 107601958200 | 459104210 | 4772799 | 3543944556 | 825 |
| I4 | Pre-fertilization | 693199720 | 103979958000 | 421553744 | 4097967 | 3013287898 | 810 |
| Average | | 5371043810 | 805656571500 | 3519089572 | 4660492 | 3223540756 | 737 |

| sample_ID | No_contigs_1K | size_1K | N50_1K | No_contigs_5K | size_5k | N50_5k | GC% |
|---|---|---|---|---|---|---|---|
| P1 | 387207 | 99770640 | 1741 | 9864 | 701049285 | 10169 | 62.3 |
| P2 | 721633 | 293099267 | 2034 | 31195 | 1453398921 | 9402 | 62.76 |
| P3 | 595079 | 188806023 | 1914 | 20404 | 1144783484 | 9058 | 62.22 |
| P4 | 504757 | 168539602 | 1951 | 18173 | 983824479 | 9139 | 62.55 |
| I1 | 649392 | 256716261 | 2027 | 28056 | 1302389212 | 9123 | 61.98 |
| I2 | 618357 | 214086402 | 1969 | 23648 | 1212045193 | 8891 | 60.92 |
| I3 | 726107 | 304500857 | 2140 | 32660 | 1503212627 | 9013 | 61.35 |
| I4 | 613447 | 246163115 | 2081 | 26172 | 1250578446 | 9243 | 61.37 |
| Average | 601997 | 221460270 | 1982 | 23771 | 1193910205 | 9254 | 61.93 |

We pooled all contigs based then used concoct, maxbin2, and metabat2, including the refinement program within metawrap. Concoct yielded no usable bins. Metabat2 produced more raw MAGs than Maxbin2 (435 vs. 319), the completeness was lower (39.7% vs. 48.0%), as well as higher contamination was present amongst the metabat2 bins (28.0% vs. 16.4%) (**Supplemental Fig. S2**). Pooling results in metawraps bin refinement yielded 29 MAGs in total (14 pre- and 15 for post-fertilization) (**Supplemental Fig. S2**).

*Microbiome diversity and composition of Lux Arbor switchgrass rhizosphere*

Only 571 mOTUs were identified across all eight samples combined. Comparing alpha diversity between pre- and post-fertilization samples showed no statistically significant difference in diversity metrics, which include observed mOTUs, ACE richness, Shannon diversity, or Simpson evenness (**Fig. 1A**). Qualitatively the variances were more substantial and more variable for alpha diversity metrics within post-fertilized plots, then pre-fertilized (**Fig. 1A**). There was no statistically significant difference in alpha diversity metrics (**Fig. 1A**), due to high variance observed in post-fertilization. Using UniFrac (weighted and unweighted) (**Fig. 1B, Fig. 1C**) and Bray-Curtis (**Fig. 1D**) distance samples using a paired analysis block effects with treatment are highly variable with slight clustering by pre- and post-fertilization treatment. Adonis testing (permanova) suggested no statistically significant difference by treatment using either the UniFrac (weighted and unweighted) or Bray-Curtis difference (**Supplemental Table S2**).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 March 2023 doi:10.20944/preprints202303.0168.v1

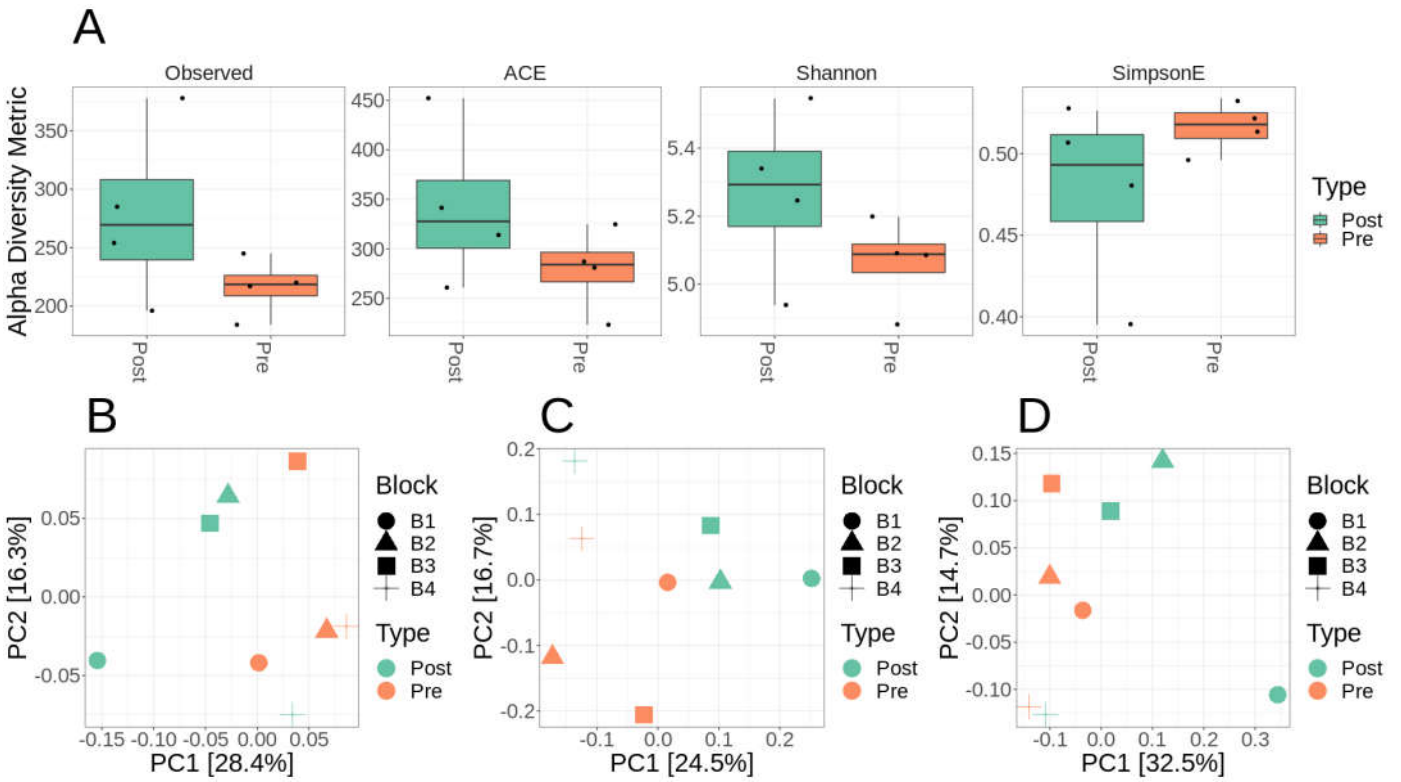


**Figure 1.** Alpha and Beta-diversity metrics for mOTUs (metagenomic OTUs) analysis. A) mOTU alpha diversity statistics observed, ACE, Shannon diversity, and Simpson eveness completed in phyloseq R without rarefaction. B-D) mOTU Beta-diversity ordinations using MDS/PcoA in phyloseq R without rarefaction using unweighted UniFrac (B), weighted UniFrac (C), and Bray-Curtis distance (D).

The microbial taxonomic composition of the Lux arbor switchgrass rhizosphere plots based on mOTUs were numerically dominated by Proteobacteria (>70% whether they were pre- or post-fertilizer), followed by Actinobacteria (>10% pre- or post-fertilizer), then the other phyla <5% each which include Acidobacteria, Bacteriodetes, Chloroflexi, Cyanobacteria/Melainabacteria, Firmicutes, Gemmatimonadetes, and Verrucomicrobia (**Fig. 2**). A single taxon, OTU158, represented >50% of the mOTU abundance in all samples (**Supplemental Table S2**). OTI158 closest reference genome is the N-fixing Alphaproteobacteria *Bradyrhizobium japonicum* based on mOTU taxonomy. OTU603 is the next most abundant OTU representing >20% the bacterial composition; it is a Betaproteobacteria *Paraburkholderia* sp. [C caribensis/terrae] and found in all samples (**Supplemental Table S2**). The most numerically dominate non-proteobacteria was OTI3128, which is *Blastococcus* sp. URHD0036, an Actinobacteria in the family Geodermatophilaceae at >10% abundance in all samples (**Supplemental Table S2**).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 March 2023 doi:10.20944/preprints202303.0168.v1

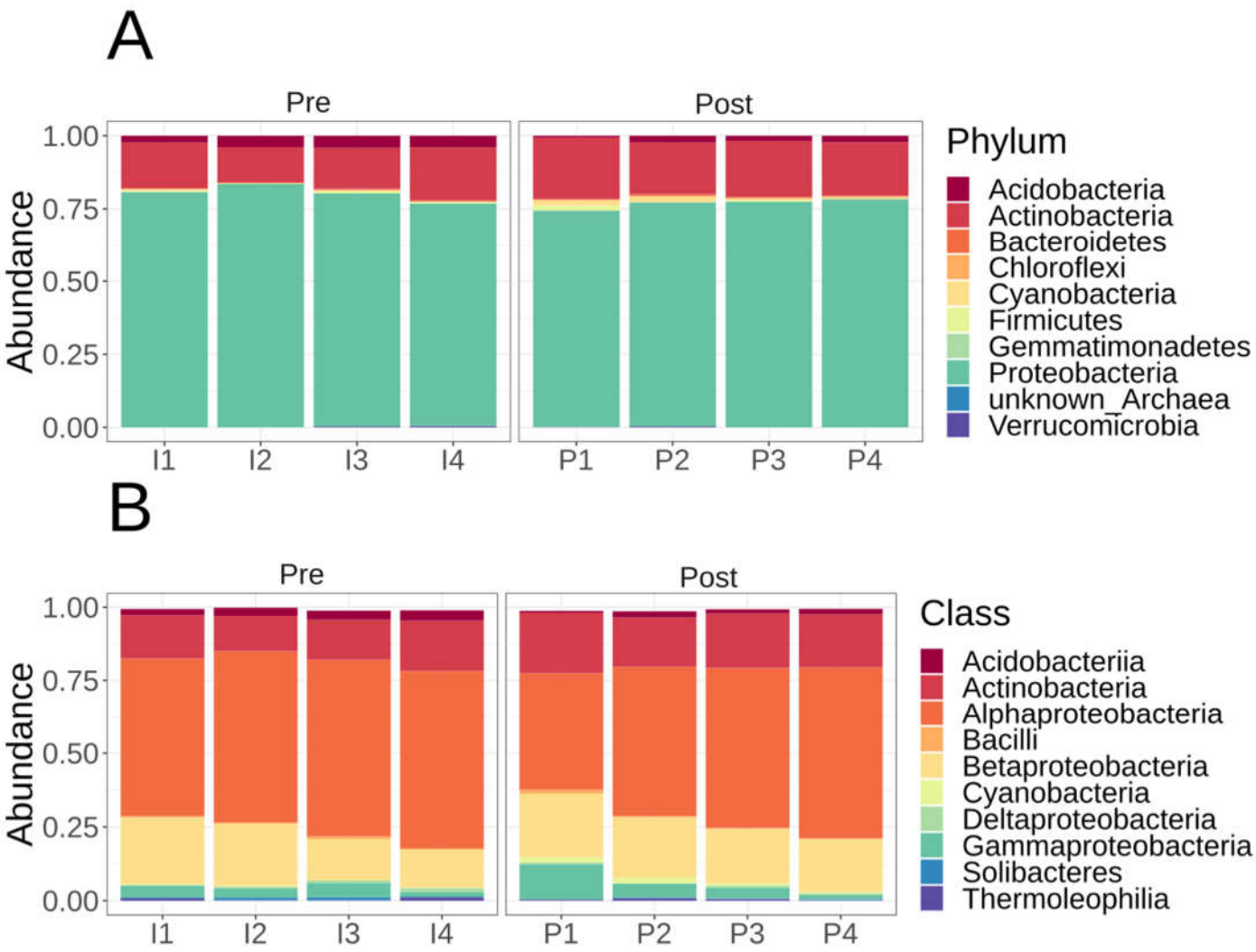


**Figure 2.** mOTU taxonomic affiliation relative abundances. A) Phyla level mOTU taxonomic relative abundance using phyloseq R without rarefaction. B) Class level mOTU taxonomic relative abundance using phyloseq R without rarefaction. Samples are labeled P1-4 for post-fertilization whereas pre-fertilization are labeled I1-4.

*Overall metabolic potential and differential metabolic genes of Lux Arbor switchgrass rhizosphere*

Our primary annotation enlisted KEGG for pathway and gene level metabolic potential and functions. Amongst our KEGG annotations (KO), >75% of the metabolic potential was metabolism based (KO level 1) which was followed by Environmental Information Processing (KO level 1) at ~11 % (**Supplemental Table S3**). Amongst the total metabolic potential of the Lux Arbor switchgrass rhizosphere >15% of the metabolism based (KO level 2) annotations were for Amino Acid and Carbohydrate metabolism (**Supplemental Table S3**).

We compared contig protein coding orf functionality using DESeq2 via paired analysis post-fertilization using KEGG KO annotations. Using a MDS/PCoA ordination of the DESeq2 KO annotations, fertilization had minimal effect on blocks 2-3, but resulted in large shifts for block 1 and 4, though we lack replicated samples to test this statistically (**Fig. 3A**). Out of 3204, nonzero ORFs counts, only 19 were significantly different in the paired DESeq2 analysis (**Fig. 3B, Supplemental Table S3**). Of those 19 differentially significant KO annotated ORFs, 15 were decreased whereas 4 were increased post-fertilization (**Fig. 3B, Supplemental Table S3**). The four that were differentially increased post-fertilization were K00171 (pyruvate ferredoxin oxidoreductase), K07691 (two-component system NarL family - ComA), KO11624 (two-component system, NarL family, response regulator YdfI) and K07694 (two-component system, NarL family, van-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 March 2023 doi:10.20944/preprints202303.0168.v1

comycin resistance associated response regulator VraR) (**Fig. 3B, Supplemental Table S3**). The NarL is a two-component system involved in signal transduction and environmental information processing. While 15 annotated ORFs were differentially decreased post-fertilization, only four showed ~2 $log_2$ fold change (**Fig. 3B, Supplemental Table S3**). Most of the KO ORFs that were significant by DESeq2 were depleted post-fertilization including four that were >2 $log_2$ fold change (**Fig. 3B, Supplemental Table S3**). These four most depleted KO's post-fertilization were K03763 (DNA polymerase III subunit alpha), K00141 (benzaldehyde dehydrogenase (NAD) [EC:1.2.1.28]), K03943 (NADH dehydrogenase (ubiquinone) flavoprotein 2), and K07406 (alpha-galactosidase) (**Fig. 3B, Supplemental Table S3**). The K07406 is involved in carbohydrate metabolism and is linked to sphingo, glycerol and glycoshingolipid metabolism and biosynthesis. The K00141 (also known as *xylC*) is involved with hydrocarbon degradation which include xylene, toluene, aminobenzoate and steroids.

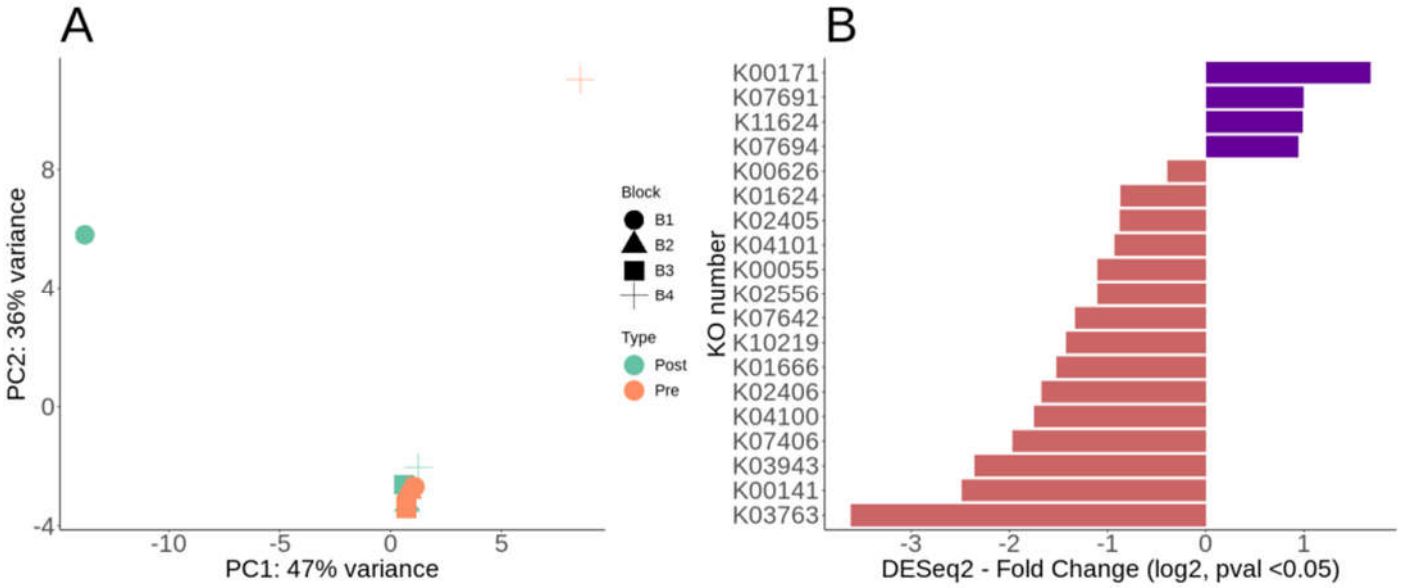


**Figure 3.** KEGG KO DESeq2 paired analysis for functional gene annotation. A) DESeq2 MDS/PcoA ordination of the KO abundances paired by sample block (B1-4) and post-fertilization. B) Divergent barplot of the KO DESeq2 $log_2$ fold changes post-fertilization all with p-value <0.05. If enriched (+ value) means more represented post-fertilization, whereas depleted (- value) less represented post-fertilization.

*Nitrogen cycle metabolic potential within the switchgrass rhizosphere microbiome*

Nitrogen cycling metabolic potential in switchgrass is critical to understanding how switchgrass gains little benefit from N fertilizer addition. We compared the abundances of genes related to various steps in the N cycle including N-fixation (*nifDHK*), denitrification (*nirSK, norB* and *nosZ*), ammonification (respiration/assimilation, *nrfA, napA, narG, nasA*), urea catabolism (*ureABC*) and anammox (*hzo*). The urea fertilization treatment also contained urease N-(n-butyl) thiophosphoric triamide (NBPT) and nitrification dicyandiamide (DCD) inhibitors. Ammonia monooxygenase enzyme encoded by *amoA* was not found in any samples. Anammox (*hzo*) were not detected in any of the samples.

We further compared the N cycling functional ORFs impact on the whole functional metabolic potential profile using DESeq2 paired analysis followed by MDS/PCoA ordination. Similar to the KO MDS/PCoA ordination samples don't cluster by pre- and post-fertilization (**Fig. 4A**). However, as with the KO analysis block 1 and 4 had greatly separated linearly post-fertilization for N specific functional ORFs (**Fig. 4A**). Functionally it appears that certain blocks were more differential by post-fertilization based on meta-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 March 2023 doi:10.20944/preprints202303.0168.v1

bolic potential. Comparing relative abundances of N cycling genes suggests similar abundances for *ureABC, nosZ, nasA,* and *napA* (**Fig. 4B**).

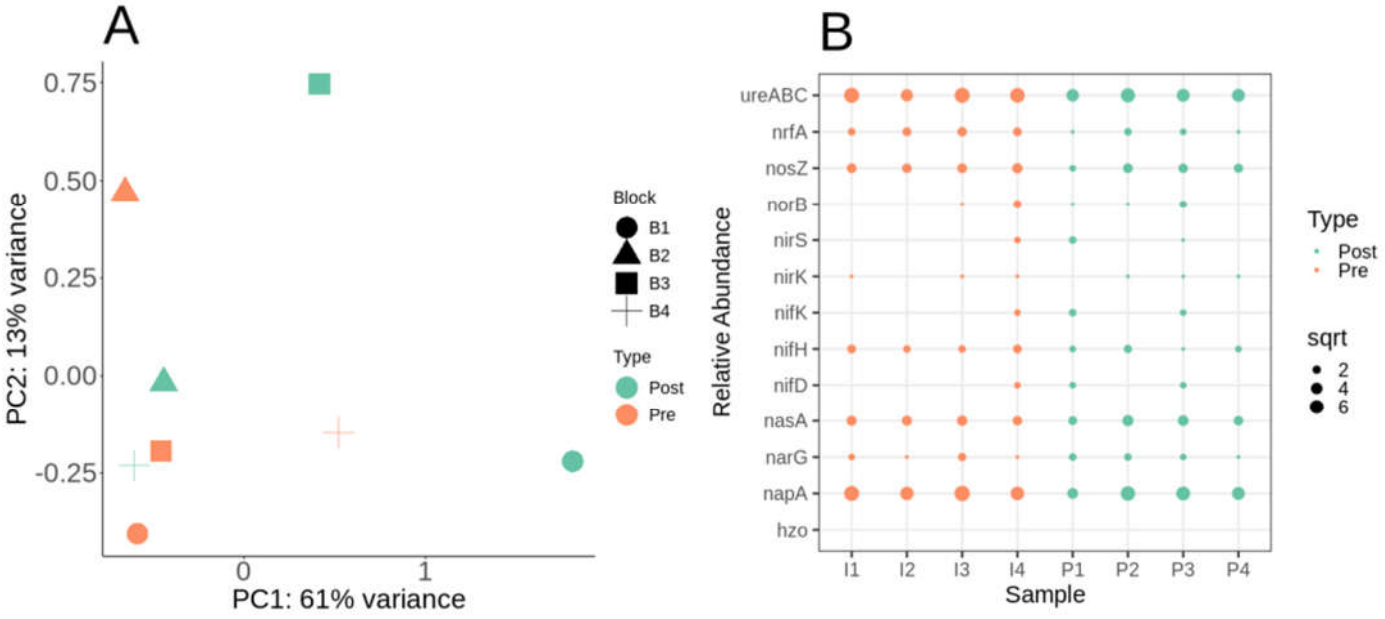


**Figure 4.** Nitrogen cycling functional genes (Prokka - COG) DESeq2 paired analysis for functional gene annotation. A) DESeq2 MDS/PcoA ordination of the nitrogen functional abundances paired by sample block (B1-4) and tested for post-fertilization. B) Dotplot of square root (sqrt normalized) relative abundances of nitrogen cycling functional genes. Samples are labeled P1-4 for post-fertilization whereas pre-fertilization are labeled I1-4.

Comparing N fixation, *nifHDK* gene cluster was detected in more post-fertilized samples over pre-fertilized (**Fig. 4B**). Blast analysis of *nifD,* the molybdenum-iron nitrogenase alpha chain, found that all full-length sequences were Betaproteobacterial in origin (Supplemental Table S4). Of the five *nifD* sequences found amongst the assembled contigs, two belong to unclassified Betaproteobacteria, two are from *Dechloromonas* sp., one from *Sulfuriferula* sp, and lastly one from *Herbaspirillum* sp (Supplemental Table S4). All *nifD* sequences belong to Betaproteobacteria, with none detected from Alphaproteobacteria even amongst the high abundance of Bradyrhizobium detected in mOTU analysis. The N fixation gene *nifH* on average had 60% higher abundance in plots pre-fertilization (**Fig. 4B**). Diverse members of proteobacteria contained *nifHK* genes including: *Azonexus hydrophilus, Herbaspirillum frisingense, Herbaspirillum rubrisubalbicans, Herbaspirillum* sp. HC18, *Dechloromonas aromatica, Dechloromonas* sp. HYN0024, *Dechloromonas* sp. Dech2017, and *Rhodocyclaceae* bacterium (Supplemental Table S4). The alternative nitrogenase gene clusters *anf* (iron containing nitrogenase) or *vnf* (vanadium containing nitrogenase) were not detected. No rhizobium (e.g., bradyrhizobium) nitrogenase genes were detected.

Nitric oxide reductase (*norB*) was detected in 3 out of 4 plots post-fertilization (Fig. 4B). NBPT had little inhibition of the gene level counts of *ureABC,* which encode all the major subunits of the urease enzyme as all samples had high levels of *ureABC* abundance (**Fig. 4B**). Both *ureABC* and *napA* were the most abundant N cycle-related gene regardless of the plot or timepoint (**Fig. 4B**). DCD also inhibits nitrous oxide ($N_2O$) production when applied to the soil (Lan et al. 2013), but *nosZ,* which encodes nitrous oxide reductase, had similar abundances across all plots (**Fig. 4B**).

*Differential CAZy potential within the switchgrass rhizosphere microbiome*

We further examined CAZy to characterize carbon utilization, uptake and degradation within pre- and post-fertilization in our switchgrass rhizospheres. Comparing fertilization effects paired by field plot, we identified 21 CAZy enzyme genes differential

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 March 2023 doi:10.20944/preprints202303.0168.v1

pre- versus post-fertilization (**Table 2**). An MDS/PcoA plot of the DESeq2 paired analysis results again shows sample block 1 and 4 having the largest effect post-fertilization (**Fig. 5A**). Of the 21 differential CAZy predictions, 71% were depleted with only five that were enriched ~2 $\log_2$ fold change post-fertilization (**Fig. 5B, Table 2**). Complete CAZy families were not enriched but individual CAZy enzymes from various taxa were (**Fig. 5B, Table 2**). Those five CAZy that were enriched include a GT41 (*Geobacter sulfurreducens*), GT28 (*Singulisphaera acidiphila*), GH9 (Uncultured bacterium BLR10), GT2 (*Acidobacterium capsulatum*), GT51 (*Brevibacterium linens*), and GH33 (*Cyclobacterium amurskyense*) (**Fig. 5B, Table 2**). The GT2, GT28, GT41 are families of glycosyltransferases that function on substrates such as N-acetyl-α-D-glucosamine, glycerol, galactose, cellulose, chitin, and glucans by an inverting mechanisms. GH9 are glycoside hydrolases that function to catabolize cellulose, lichenin, cellobiose, and other plant components. GH33 is a specialized for glycogen, dextrin, and other aminosugars. Three CAZy predicted enzymes had > -3 $\log_2$ fold change post-fertilization which include a Carbohydrate-binding module (CBM54), glycosyltransferase (GT4) a multi-domain auxiliary activity (AA3_1/AA8) (**Fig. 5B, Table 2**). CBM54 binds to xylan, yeast cell wall glucan and chitin (Dvortsov et al. 2009), but this family's function is relatively unknown. The AA3_1/AA8 has a heme binding site, a cytochrome domain, a cellobiose dehydrogenase, and a choline dehydrogenase or flavoprotein and is of Basidiomycota fungal origin (Table 2). The GT4 is another major transferases family for simple sugars via a retaining mechanism which includes sucrose synthase (EC 2.4.1.13), sucrose-phosphate synthase (EC 2.4.1.14) and α-glucosyltransferase (EC 2.4.1.52). Of the most significant CAZy enzymes, 52% are associated with plants, plant-associated zones (phyllosphere or rhizosphere) or are from soil directly (**Table 2**). CAZy enzymes that were significant represented three kingdoms (fungi, archaea, bacteria), seven bacterial phyla, and two uncultivated organisms (**Table 2**).

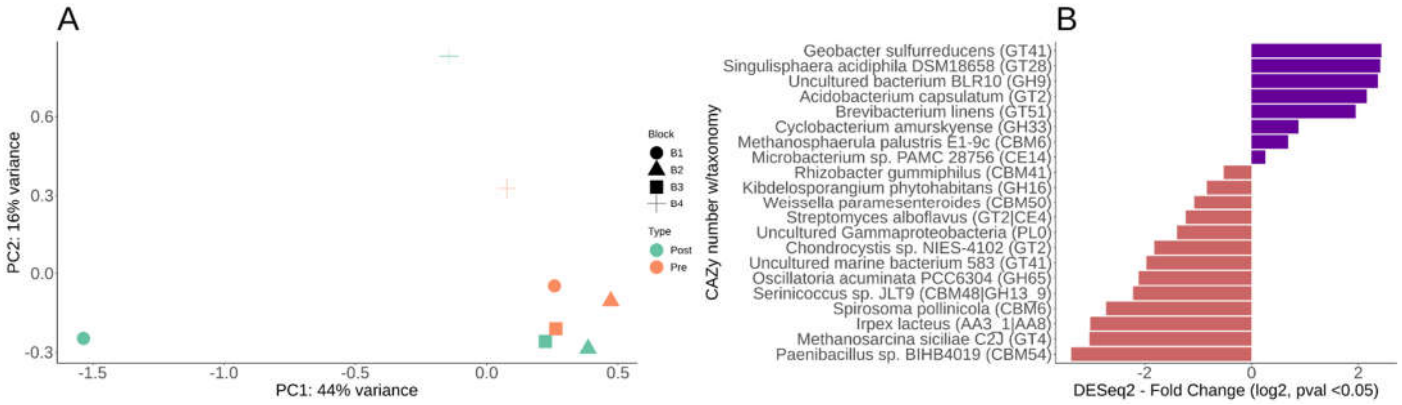


**Figure 5.** DESeq2 paired analysis of CAZy for carbohydrate active genes gene annotation. A) DESeq2 MDS/PcoA ordination of the CAZy functional abundances paired by sample block (B1-4) and post-fertilization. B) Divergent barplot of the CAZy DESeq2 $\log_2$ fold changes post-fertilization all with p-value <0.05. If enriched (+ value) means more represented post-fertilization, whereas depleted (- value) less represented post-fertilization.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 March 2023 doi:10.20944/preprints202303.0168.v1

**Table 2.** CAZy DESeq2 paired analysis statistical table. This includes the taxonomic accession from CAZy database, the genbank taxonomy with CAZy family (funtaxa), the genbank phyla and class taxonomy, the location of isolation from GenBank, the DESeq2 $\log_2$ fold change which is paired by sample location block (B1-4) then tested for post-fertilization, and p-value from DESeq2. If enriched (+ value) means more represented post-fertilization, whereas depleted (- value) less represented post-fertilization.

| accession | funtaxa | phyla | class | isolation | log2FoldChange | pvalue |
|---|---|---|---|---|---|---|
| ANY66681.1 | Paenibacillus sp. BIHB4019 (CBM54) | Firmicutes | Bacilli | Rhizosphere | -3.38 | 0.02 |
| AKB38096.1 | Methanosarcina siciliae C2J (GT4) | Euryarchaeota | Methanomicrobia | Unknown | -3.03 | 0.04 |
| ALJ82902.1 | Irpex lacteus (AA3_1\|AA8) | Basidiomycota | Agaricomycetes | Wood | -3.02 | 0.04 |
| AUD02463.1 | Spirosoma pollinicola (CBM6) | Bacteroidetes | Cytophagia | Pollen | -2.72 | 0.02 |
| ANS78621.1 | Serinicoccus sp. JLT9 (CBM48\|GH13_9) | Actinobacteria | Actinobacteria | Hydrothermal sea | -2.22 | 0.05 |
| AFY81829.1 | Oscillatoria acuminata PCC6304 (GH65) | Cyanobacteria | Cyanophyceae | Soil | -2.11 | 0.02 |
| AAR38497.1 | Uncultured marine bacterium 583 (GT41) | Uncultured | Uncultured | Ocean | -1.97 | 0.02 |
| BAZ44095.1 | Chondrocystis sp. NIES-4102 (GT2) | Cyanobacteria | Cyanophyceae | Unknown | -1.82 | 0.04 |
| BAL56682.1 | Uncultured Gammaproteobacteria (PL0) | Proteobacteria | Gammaproteobacteria | Microbial mat | -1.40 | 0.05 |
| ARX88346.1 | Streptomyces alboflavus (GT2\|CE4) | Actinobacteria | Actinomycetes | Rhizosphere | -1.23 | 0.02 |
| ATF41409.1 | Weissella paramesenteroides (CBM50) | Firmicutes | Bacilli | Unknown | -1.07 | 0.02 |
| ALG08540.1 | Kibdelosporangium phytohabitans (GH16) | Actinobacteria | Actinobacteria | Phyllosphere | -0.84 | 0.04 |
| ATU64527.1 | Rhizobacter gummiphilus (CBM41) | Proteobacteria | Gammaproteobacteria | Soil | -0.52 | 0.05 |
| AMG83817.1 | Microbacterium sp. PAMC 28756 (CE14) | Actinobacteria | Actinobacteria | Lichen | 0.26 | 0.04 |
| ACL17090.1 | Methanosphaerula palustris E1-9c (CBM6) | Euryarchaeota | Methanomicrobia | Peatland Soil | 0.69 | 0.03 |
| AKP50194.1 | Cyclobacterium amurskyense (GH33) | Bacteroidetes | Flavobacteria | Ocean | 0.88 | 0.05 |
| AMT93207.1 | Brevibacterium linens (GT51) | Actinobacteria | Actinomycetes | Sediment | 1.95 | 0.02 |
| ACO33523.1 | Acidobacterium capsulatum (GT2) | Acidobacteria | Acidobacteria | Soil | 2.16 | 0.02 |
| ACN58963.1 | Uncultured bacterium BLR10 (GH9) | Uncultured | Uncultured | Soil | 2.37 | 0.05 |
| AGA24658.1 | Singulisphaera acidiphila DSM18658 (GT28) | Planctomycetes | Planctomycetacia | Peat bog wetland | 2.41 | 0.04 |
| BBA71022.1 | Geobacter sulfurreducens (GT41) | Proteobacteria | Deltaproteobacteria | Sediment | 2.43 | 0.04 |

*Genome-resolved metagenomics elucidates resolves members of the rare biosphere*

We obtained 190,172 >5 kbp contigs in total from our eight metagenomes and used the differential read abundance pre- and post-fertilization, which resulted in 29 MAGs from many phyla (e.g., Actinobacteria, Acidobacteria, Dormibacterota (formally candidate division AD3), *Nitrospira*, Gemmatimonadetes, Proteobacteria, and Verrucomicrobia). These MAGs represent high abundance (Actinobacteria, Acidobacteria, Proteobacteria, and Verrucomicrobia) and low abundance members (*Nitrospira* and Gemmatimonadetes) of common soil phyla as well as members of the rare biosphere Dormibacterota, Candidatus Eisenbacteria, Candidate phyla UBA10199 formerly Deltaproteobacteria) (**Fig. 6A, Table 3**). Acidobacteriota (Acidobacteria) phyla had the most representative MAGs with eight (**Fig. 6A, Table 3**). The MAG genome sizes ranged from 2.5 to 11 Mbp, with a G+C content of 55 to 71.7%, with a total contig range from 38 to 527 (Table 3). Amongst the 14 MAGs within the pre-fertilized samples, 2 are close to being high-quality drafts. High-quality drafts are defined by the minimum information about a metagenome-assembled genome (MIMAG) reporting guidelines at >90% complete, <5% contamination, and with the presence of one entire rRNA operon (5S, 16S, 23S) and 18 tRNAs (Bowers et al. 2017). The rest of the pre-fertilized MAGs are medium quality drafts, and no low-quality drafts were used in downstream analysis. For the 15 MAGs within the post-fertilized plot, eight are near high-quality drafts with the rest being medium draft quality (Table 3).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 March 2023 doi:10.20944/preprints202303.0168.v1

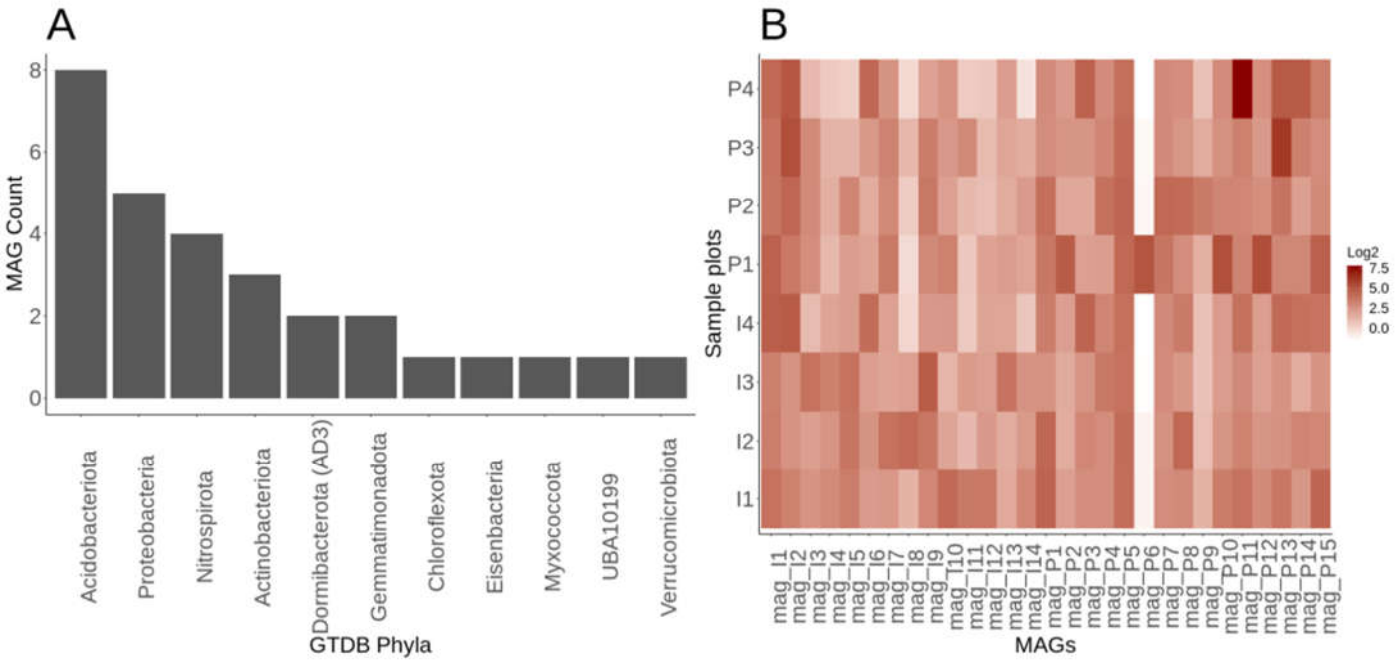


**Figure 6.** Metagenomic assembled genome (MAG) statistics and sample relative abundances. A) Barplot of the number of MAGs per phyla using the GTDB taxonomy. B) Quantification of relative abundances of the MAGs using Metawraps tool (quant_bin) with values expressed in the heatmap as genome copies per million reads.

**Table 3.** Metagenomic assembled genome (MAG) assembly statistics with GTDB taxonomy. This includes the predicted taxonomy by the GTDB-Tk tool using the GTDB database. The assembly statistics with checkM MAG completeness and contamination is included. The metagenome-assembled genome (MIMAG) of bacteria and archaea quality ranking (Bowers et al. 2017) is included. MAGs are labeled by rhizosphere soil type for pre-fertilization (I1-I14) and post-fertilization (P1-P15).

| | GTDB-Tk taxonomy | Size | Contigs | N50 | GC | completeness | contamination | MIMAG |
|---|---|---|---|---|---|---|---|---|
| magI1 | Actinobacteriota;Thermoleophilia;20CM-4-69-9;20CM-4-69-9 | 3377677 | 169 | 24250 | 69.9% | 81 | 1.293 | Medium |
| magI2 | Acidobacteriota;Thermoanaerobaculia | 5321068 | 131 | 66773 | 66.0% | 95.96 | 2.849 | Medium |
| magI3 | Nitrospirota;Nitrospiria;Nitrospirales;Nitrospiraceae;Nitrospira_C | 3293157 | 304 | 11734 | 58.7% | 87.87 | 7.929 | Medium |
| magI4 | Eisenbacteria;RBG-16-71-46 | 2561625 | 211 | 13768 | 67.9% | 84.84 | 1.098 | Medium |
| magI5 | Gemmatimonadota;Gemmatimonadetes;Gemmatimonadales;GWC2-71-9 | 3461988 | 62 | 148829 | 67.7% | 94.18 | 2.197 | Medium |
| magI6 | Acidobacteriota;Acidobacteriae;Acidobacteriales;Koribacteraceae | 4052648 | 57 | 120172 | 56.0% | 91.05 | 0.854 | Medium |
| magI7 | Acidobacteriota;Thermoanaerobaculia | 7820903 | 527 | 18142 | 67.6% | 80.65 | 3.703 | Medium |
| magI8 | Nitrospirota;Nitrospiria;Nitrospirales;Nitrospiraceae;GCA-2737345 | 4109394 | 341 | 12463 | 56.4% | 84.66 | 4.545 | Medium |
| magI9 | Myxococcota;Polyangia;Kofleriales;Kofleriaceae | 11007611 | 231 | 77779 | 69.0% | 81.93 | 1.474 | Medium |
| magI10 | Dormibacterota;Dormibacteria | 3788213 | 181 | 27454 | 71.7% | 88 | 1.851 | Medium |
| magI11 | Nitrospirota;Nitrospiria;Nitrospirales;Nitrospiraceae;Nitrospira_C | 4541993 | 243 | 24855 | 55.3% | 96.36 | 3.989 | Medium |
| magI12 | Gemmatimonadetes;Gemmatimonadales;GWC2-71-9;40CM-2-70-7 | 2649947 | 193 | 16824 | 67.7% | 82 | 1.098 | Medium |
| magI13 | Dormibacterota;Dormibacteria | 4029220 | 85 | 69215 | 70.7% | 98.61 | 0.925 | Medium |
| magI14 | Nitrospirota;Nitrospiria;Nitrospirales;Nitrospiraceae;Nitrospira_C | 3065303 | 170 | 26094 | 56.7% | 82.01 | 6.363 | Medium |

We compared multiple methods to obtain taxonomic identity for MAGs, finding GTDB-Tk to be the most reliable. We classified MAG taxonomy by metawrap's classifier, classify genomes which use the mOTU v2 taxonomy, JSpeciesWS Tetra Correlation Search (TCS) (Richter et al. 2016; White III et al. 2016), GTDB-Tk, and blastp of ribosomal protein S9 gene. Only GTDB-Tk provided the taxonomic identifications that were sup-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 March 2023 doi:10.20944/preprints202303.0168.v1

ported by ribosomal protein S9 gene blastp results. Classify genomes supplied no identifications beyond "Bacteria." (Supplemental Table S5). JspeciesWS TCS misclassified candidate phyla such as Dormibacterota, which it classified incorrectly as “Mycobacterium.” (Supplemental Table S5). Metawraps classifier also provided no identifications for candidate phyla like Dormibacterota (Supplemental Table S5). GTDB-Tk based taxonomy was therefore used for all downstream MAG taxonomy.

While the presence of all 29 MAGs was present in all samples, whether pre- or post-fertilization (Fig. 6B), the abundances differed across the MAGs resolved. MAG P9 had the lowest average abundance across samples (Fig. 6B), whereas MAG P11 had the highest average overall abundance (Fig. 6B). While we had limited replication (4 replicates), Acidobacteria and Actinobacteria phyla were also found amongst the top 10 mOTUs phyla based on composition. The samples pre-fertilization resolved MAGs from Gemmatimonadetes, Candidatus Eisenbacteria, *Nitrospira*, and Dormibacterota, but these were not as highly resolved or as abundant in the post-fertilization samples (Table 3). We had one MAG in the pre-fertilization samples belonging to the "Myxococcota," formally Deltaproteobacteria but now its own phyla based on the GTDB. The post-fertilization samples yielded MAGs from Verrucomicrobia, Chlorflexota, Candidate phyla UBA10199 that were not resolved well in the pre-fertilization samples (**Table 3**).

Gammaproteobacterial MAGs from the *Lelliottia* and *Janthinobacterium* genus (Table 3) were only found in the post-fertilization samples. MAG P6 a *Lelliottia* (Enterobacteriaceae) poorly represented across the samples all but one post-fertilization (P1) (Fig. 6B). The high abundance of P6 in a single sample could suggest an infection of plant roots within the switchgrass rhizosphere. Lelliottia are opportunistic pathogens of roots and implicated in post harvest onion rot (Liu et al. 2016).

*Betaproteobacterial MAG with molybdenum based nitrogen-fixation gene cluster*

We screened all our MAGs for N-fixation genes such as the dominant molybdenum based (*nif* gene cluster) and the alternative N-fixation clusters based on vanadium (*vnf* gene cluster) and iron (*anf* gene cluster). As mentioned above, the only *nifD* genes have similarity to Proteobacteria with no other phyla represented. We resolved four proteobacterial MAGs in the post-fertilization treatment only. MAG P10, which is classified as *Janthinobacterium*, has two copies of *nifD* and *nifK*, one *nifH*, and a one *nifW* arranged on a single gene cluster for molybdenum-based N-fixation. The *nifHDK* was previously detected as HGW-Betaproteobacteria-11 and HGW-Betaproteobacteria-7 (Supplemental Table S4), which are located on the same contig within our MAG P10 genome.

*Acidobacteria related to rare subdivision 23 with utilization nitrate*

Acidobacteria represented eight of the total twenty-nine MAGs within our Lux Arbor switchgrass rhizosphere with relations to Thermoanaerobaculia (subdivision 23), Koribacteraceae (subdivision 1), and unclassified Acidobacteriales (subdivision 1). Five MAGs were related to Thermoanaerobaculia (subdivision 23), with 2 in pre-fertilized and 3 in post-fertilized (Table 3). The Thermoanaerobaculia genomes resolved ranged from 4 to 5.4 Mbp, 56 to 66% G+C, completeness from 91 to 96%, and contamination 0.8 to 5.5% (Table 3).

Our Thermoanaerobaculia appears to utilize nitrate but not ammonium, urea, or have the ability to fix N. Ammonia monooxygenase (*amoA* or *amoB*), urease (alpha or gamma, *ureA* or *ureC*), nitrification genes (*nxrAB*), nitrous oxide reductase (*nosZ*), or nitrite reductases (*nirK* or *nirS*) were not detected in the Thermoanaerobaculia MAGs. The anaerobic nitric oxide reductase transcription regulator (NorR) was present in all Thermoanaerobaculia genomes with up to nine copies in I2. Nitrate reductase (1.7.99.4; *napA* and *nasC*), formate dehydrogenase nitrate-inducible (*fdnH*), and nitrate transporter (*narT*) was present amongst the genomes. The *nasA* nitrate reductase wasn't present in any of the MAGs. MAG P4, while not classified past Acidobacteriales, had nitrate utili-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 March 2023 doi:10.20944/preprints202303.0168.v1

zation genes including *narT* transporter, *nasC* nitrate reductase, and the formate dehydrogenase nitrate-inducible gene (*fdnH*).

We further characterized the carbohydrate utilization in the Thermoanaerobaculia MAGs for potential carbon source utilization. The most abundant genes included major glycosyltransferases (GT2 and GT4) families that synthesize diverse substrates including cellulose, chitin, sucrose, sucrose-phosphates, and glucose-glycerol phosphates. The top glycoside hydrolases (GH) encoded by the Thermoanaerobaculia MAGs included GH23 and GH0 families. GH0 is the uncharacterized family of GH, which comprises completely novel and unknown enzymes. The GH23 is specific enzyme family is a rather specific substrate family which contains lysozyme type G (EC 3.2.1.17); peptidoglycan lyase (EC 4.2.2.n1) and chitinase (EC 3.2.1.14). GH3 and GH18 were also numerically abundant amongst the Thermoanaerobaculia MAGs, which encode GH3 (β-glucosidase (EC 3.2.1.21); xylan 1,4-β-xylosidase (EC 3.2.1.37); β-glucosylceramidase (EC 3.2.1.45) and GH18 (chitinase (EC 3.2.1.14); lysozyme (EC 3.2.1.17); endo-β-N-acetylglucosaminidase (EC 3.2.1.96); peptidoglycan hydrolase). Carbohydrate-binding module 50 (CBM50) was the most abundant CBM enzyme amongst the Thermoanaerobaculia MAGs, which contains 50 residues has a LysM domain and works synergistically with GH23 or other enzymes that cleave chitin or peptidoglycan. CBM2 was the second most abundant CBM enzyme present in the Thermoanaerobaculia MAGs, which is 100 residues with modules that bind cellulose, chitin, and xylan.

*Nitrospira hydrolysis of urea, nitrate reduction with limited nitrite reduction*

*Nitrospira* MAGs were only well resolved in the pre-fertilized plots. Four MAGs (I3, I8, I11, I14) in the pre-fertilized plots were classified as *Nitrospira*, the genomes resolved ranged from 3 to 4.5 Mbp, 55 to 58% G+C, completeness from 82 to 96%, and contamination 4 to 8% (Table 4). *Nitrospira* MAG I11 was the most complete at 96% with the lowest contamination at ~4% (Table 4). *Nitrospira* MAG I11 MAG had up to 0.11% of all the reads in an pre-fertilized plot map directly to the genome, representing relatively high abundance.

We further examined the N metabolism of the *Nitrospira* related MAGs within Lux Arbor to identify how N metabolism functions these recovered MAGs. None of the MAGs had genes related to ammonia monooxygenase (*amoA* or *amoB*), so no gene annotations support the presence of ammonia oxidation metabolism. No nitrification genes (*nxrAB),* nitrous oxide reductase (*nosZ*), or nitrite reductases (*nirK* or *nirS*) genes were detected amongst the annotations for these MAGs. All the *Nitrospira* MAGs had urease subunits (alpha or gamma, *ureA* or *ureC*) plus accessory proteins (*ureDGF*) present. As for denitrification, all *Nitrospira* MAGs had denitrification regulatory protein (*nirQ*), but that was the only *nir* gene found amongst the MAGs. Nitric oxide reduction pathway, which encodes an anaerobic nitric oxide reductase (*norV*/*norW*), were not found; however, the anaerobic nitric oxide reductase transcription regulator (*norR*) was found amongst the MAGs. All MAGs have assimilatory nitrite reductase (*nasE*), and MAG I14 has a copper-containing nitrite reductase, but no other genes for nitrite metabolism were found. MAG I14 had genes relating to nitrate influx and reduction to nitrite, and genes include the nitrate transporters (*nasA*, 3 copies), and nitrate reductase (*napA*, 1.7.99.4). MAG I11 had *napA* but didn't have the *nasA* transporters. N fixation has never been found amongst the *Nitrospira* and was not found in any of our resolved MAGs.

Using CAZy, we compared the carbohydrate-active enzymes present within our *Nitrospira* MAGs relating to carbon source metabolism. Glycosyltransferases (GT2 and GT4) were the most prevalent CAZy enzymes present in the *Nitrospira* MAGs which metabolism cellulose, chitin, or simple sugars like sucrose. As with the Thermoanaerobaculia MAGs, the most numerically abundant GHs were GH23 and GH0 in the *Nitrospira* MAGs. Carbohydrate esterases with the most numerical abundance included CE11, CE1, and CE14. CE1 contains acetyl xylan esterase (EC 3.1.1.72), cinnamoyl esterase (EC 3.1.1.-) and feruloyl esterase (EC 3.1.1.73). The CE1 family also contains intracellular

 doi:10.20944/preprints202303.0168.v1

poly(3-hydroxybutyrate) (PHB) depolymerases. CE14 family contains N-acetyl-1-D-myo-inosityl-2-amino-2-deoxy-α-D-glucopyranoside deacetylase (EC 3.5.1.89) and diacetylchitobiose deacetylase (EC 3.5.1.-). Diacetylchitobiose deacetylase is involved in chitin degradation and metabolism.

*Dormibacterota MAGs metabolic potential within the switchgrass rhizosphere*

Dormibacterota MAGs were resolved in the pre-fertilization samples only (MAGs I10 and I13). The two MAGs ranged from 3.7 to 4 Mbp, 71 to 72% G+C content, completeness 88 to 98%, contamination 0.9 to 2% (Table 4). Dormibacterota I13 is a near highly-quality genome at 98.6% complete with 0.92% contamination and was the most resolved MAG in the pre-fertilized treatments (Table 4).

The two Dormibacterota MAGs contains no complete gene clusters for N-fixation, urea, or nitrite utilization. Urea and nitrite utilization genes were also not found. I10 has a *nifH* nitrogenase but is missing the rest of the genes required for N-fixation such as *nifDK*. MAG I13 has no N fixation genes. Both have a nitrate-inducible formate dehydrogenase (fdnH) and the *napA* nitrate reductase. The *nasC* nitrate reductase and *narT* nitrate transporter were not found in I10 or I13 MAG. The I13 MAG lacked the *nasA* nitrate transporter, whereas I10 has the *nasA* transporter gene.

Carbon monoxide and dioxide utilization have been previously associated with Dormibacterota in soils (Ji et al. 2017), there was examined whether our MAGs had similar gene clusters. Carbon monoxide dehydrogenases were found amongst the I10 and I13 MAGs, which catalyze the oxidation of carbon monoxide to carbon dioxide using a quinone donor (EC:1.2.5.3). Carbon monoxide aerobic dehydrogenases have named either *cox* or *cut* gene clusters under the same EC 1.2.5.3. I10 and I13 had *cutL* (large chain), *cutM* (medium chain) and *cutS* (small chain) genes present. I10 had one *coxS* (small chain) gene, but I13 had zero *cox* genes related to CO dehydrogenases. No ribulose-1,5-bisphosphate carboxylase (RuBisCO) (*rbcL*).

Our Dormibacterota MAGs had similar high abundances of GT2, GT4, CBM2, CBM50, GH0, GH18, GH23 and CE14 in the top ten of their CAZy repertoire as did our Acidobacteria and *Nitrospira* MAGs. This suggests that our Dormibacterota MAGs I10 and I13 can utilize cellulose, chitin, or simple sugars like sucrose.

## 4. Discussion

Metagenomic assembly of soil and rhizosphere ecosystems has remained an enormous challenge due to the difficulty in obtaining long contigs to reconstruct high-quality MAGs. While metagenomics has improved since the original prairie soil assembly, which contained only a few contigs >5 kbp (Howe et al. 2014), due to the development of better software (Li et al. 2015; White III et al. 2017c), short reads still provide significant challenges. The Lux Arbor marginal land switchgrass rhizosphere metagenomes we report represent an excellent model system with very long contigs from short reads (150 bp paired-end), due to lower microbial community diversity and complexity. In prairie soils (e.g., Kansas or Iowa) (Howe et al. 2014; White III et al. 2016), it is not possible to obtain long contigs with just short reads and computation alone (White III et al. 2016) and only long read technologies (i.e., moleculo) have yielded similar results to our Lux Arbor short read assembly. To compare, on average a single sample from Lux Arbor had 23,771 contigs >5 kbp, while a Kansas native prairie soil with a similar amount of data had 4,683 (100 bp paired-end) and 8,532 (250 bp paired-end) (White III et al. 2016). A single moleculo sequence library from a pooled Kansas prairie sample yielded 10,198 contigs >10 kbp in length, and our Lux Arbor side yielded 44,171 >10 kbp in length using only short reads. The max contig length obtained from a hybrid assembly of Kansas prairie was <63 kbp, whereas Lux Arbor had 237 contigs >100 kbp in length with a max contig of 697,599 bp (White III et al. 2016). Recently, a closed bacterial genome has been

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 March 2023 doi:10.20944/preprints202303.0168.v1

obtained from the Saccharibacteria formally candidate phyla TM7 from stable isotope labeled rhizosphere metagenome suggesting binning complete genomes directly from soil is possible (Starr et al. 2018). Our data suggest that Lux Arbor soils have lower microbial community complexity than Kansas prairie soil based on the quality of metagenomic de novo assembly obtained. This nominates Lux Arbor and possibly other marginal soils may provide a testbed for soil and rhizosphere metagenomics.

Obtaining metagenomic bins to resolve individual microbial genomes within the soil and rhizosphere has remained problematic as the common assumption is that the higher the microbial complexity, the harder it is to resolve genomes directly from a sample. Low-complexity permafrost soil has had great success in resolving genomes, with over 1,500 individual MAGs resolved with expressed metabolisms using transcriptomics and proteomics (Woodcroft et al. 2018), but we have yet to obtain this order of magnitude with ease in non-permafrost soil. The first genome-centric view of a soil ecosystem was in the Kansas native prairie where 129 MAGs were obtained, but on average the genome completeness was quite low at ~40% (White III et al. 2016). The second of a grassland soil resolved 372 total genomic bins with 181 partial to near complete (Butterfield et al. 2016). A recent study of Amazon soil (using MIMAG guidelines) had 29 MAGs that were medium quality representing over ten phyla including members of the Candidate phyla radiation (Kroeger et al. 2018). Mediterranean grassland soil MAG study obtained 793 MAGs that were near complete (Diamond et al. 2019). We compared concoct, maxbin2, and metabat2 within metawrap and found that on average maxbin2 provided high-quality with lower contamination MAGs than metabat2, resulting in the 29 MAGs that we describe (**Supplemental Fig. S2**).

In this study, we resolved genomes in the Lux Arbor switchgrass rhizosphere that represent uncultivated phyla including the Acidobacteria group (rare subgroup 23), Candidate phyla UBA10199, Candidatus Eisenbacteria and Dormibacterota (AD3). Acidobacteria is dominant soil phyla representing upwards of 20% of all soil bacteria, highly diverse, and are physiologically active (Naether et al. 2012). Acidobacteria MAGs from Kansas prairie soil was highly transcriptionally active (White III et al. 2016), and genomes have been resolved from grassland (Butterfield et al. 2016) and Amazonian soil (Kroeger et al. 2018). Lux Arbor Thermoanaerobaculia MAGs have previously never been described in soil, only in wastewater, sediments, and hot springs (Losey et al. 2013; Parks et al. 2017). The Thermoanaerobaculia MAGs we describe are the first representatives of Acidobacteria subgroup 23 from a soil or rhizosphere environment. The Acidobacteria and Dormibacterota phyla are on the 'most wanted list,' of organisms from the soil and rhizosphere ecosystem for cultivation (Carini, 2019) and genome references via single-cell genomes or MAGs (Choi et al. 2017). Thermoanaerobaculia and Dormibacterota MAGs have the potential to utilize nitrate, but not molecular N, urea, or ammonia. Their carbohydrate metabolism is similar to the other Lux Arbor MAGs in terms of utilization of cellulose, chitin, and simple sugars like sucrose. The Dormibacterota have been previously implicated in carbon gas exchange (CO and CO2), in Antarctic soils including RuBisCO and carbon monoxide dehydrogenases (Ji et al. 2017). Recently, Dormibacterota MAGs have been resolved in subsurface soil horizons which have had genes identified to aid survival in low-nutrient environments (Brewer et al. 2019). We find the Lux Arbor MAGs lack RuBisCo for CO2 capture and utilization, but do have CO-dehydrogenases which may allow CO metabolism; CO metabolism may be thus be conserved in Dormibacterota found in soil, rhizosphere and permafrost ecosystems. Brewer et al. 2019 found that their Dormibacterota lack both RuBisCo and dehydrogenases lacking autotrophic metabolism (Brewer et al. 2019). Dormibacterota may lose autotrophy under more stressful environmental conditions.

We resolved the only the third representative from the elusive candidate phyla Eisenbacteria. Our MAG is the first to be found amongst soil or rhizosphere ecosystems. The previous two where found via genome resolved metagenomics in the Atlantic Ocean deep vent sample (BioSample: SAMN09287800) named Candidatus Eisenbacteria bacterium SZUA-252 and Rifle, Colorado USA background sediment (BioSample:

 doi:10.20944/preprints202303.0168.v1

SAMN04313721) named Candidatus Eisenbacteria bacterium RBG_16_71_46 (Anantharaman et al. 2016). This phyla appears to be extremely rare as in 8,000 MAG study did not find any representatives across thousands of samples (Parks et al. 2017). Here, we add another representative of this rare phyla for further comparative genome analysis.

In contrast to previous studies, the metagenomes we report nominate the betaproteobacterium Janthinobacterium as a candidate organism for association nitrogen fixation in the switchgrass rhizosphere. Nitrogen-fixation (nif gene cluster) was present within the bulk metagenome amongst diverse betaproteobacterial members: Azonexus, hydrophilus, Herbaspirillum, Dechloromonas, Rhodocyclaceae, and Sulfuriferula. No other nitrogenase (nif/anf/vnf) genes outside of betaproteobacterial class were discovered. Our reconstruction of the Janthinobacterium P10 MAG demonstrated that this genome contains a complete nif gene cluster. A related Janthinobacterium lividum V30-G6 isolated from permafrost showed low levels of N-fixation via the acetylene reduction assay (Hara et al. 2014). Bradyrhizobium spp. are highly represented in our Lux Arbor mOTU data and previous nifH data from switchgrass rhizosphere soils (Roley et al. 2019). However, we were unable to find Bradyrhizobium nifDKH genes or resolve a Bradyrhizobium MAG in our study. Longer read sequencing or further depth seems to be required to address the Bradyrhizobium in Lux Arbor. Many Bradyrhizobium lack nif genes, as previously described in soils from North America to England (VanInsberghe et al. 2015; Jones et al. 2016).

Two inhibitors are included in the N fertilizer (SuperU) by the manufacturer inhibit nitrate reduction and urease: dicyandiamide (DCD) and N-(n-butyl) thiophosphoric triamide (NBPT), respectively. We measured the potential effects of these inhibitors at the community, individual, and gene level. Ammonium oxidation (amo) genes were not present in the bulk metagenome or associated with a resolved MAG. Dicyandiamide (DCD) limits the conversion of ammonium to nitrite via the ammonium monooxygenase (amo) but has no effect on urea hydrolysis (Ning et al. 2018; Yang et al. 2018). DCD has been shown to sharply reduce amo gene copy numbers in ammonium oxidizing bacteria (AOB) but has relatively little effect on ammonium oxidizing archaea (AOA) amo gene copy number (Yang et al. 2018). The AOB community is strongly shifted by DCD with impacts on function, namely nitrification, regardless of whether it's an AOB or AOA amo (Yang et al. 2018). This could be the reason why we were unable to detect any amo genes in post-fertilized plots, but cannot explain the lack of amo in our untreated plots. The AOB and AOA communities may be limited in abundance in these marginal lands due to lack of available ammonium that is rapidly fixed then utilized by plant roots. This could be due to rapid utilization of ammonium or its loss by volatilization, leaving little ammonium available for microbes with ammonium oxidation capabilities. In addition, lack of ammonia oxidation would lead to lower levels of nitrate available for denitrification and thus reduce potential N2O emissions. Indeed, DCD has been shown to mitigate $N_2O$ emissions (Lan et al. 2013) and other studies without DCD have observed lower $N_2O$ emission under switchgrass compared to other bioenergy crops. We find that the abundance of nitrous oxide reductase (*nosZ*) genes to be similar pre- and post- fertilization, suggesting that inhibition occurs beyond the gene level, either at an enzymatic level or due to substrate (nitrate) limitation. This is supported by previous work that found DCD did not affect *nosZ* gene presence or abundance (Di et al. 2014).

Urease genes were numerically abundant both pre- and post-fertilization and were prevalent within the *Nitrospira* MAGs that we assembled. This points to the importance of *Nitrospira* in transforming urea to ammonium and these organisms showed high abundance both pre- and post- fertilization. In another system, *Nitrospira* was enriched five-fold after N fertilizer treatment in agricultural soils under corn and soybean rotation, and *Nitrospira* MAGs were resolved with complete ammonia oxidization (comammox) (Orellana et al. 2018). Our *Nitrospira* MAGs lack the genes required for comammox, including the amo gene. The *Nitrospira* of Lux Arbor have the metabolic potential for urease and the urease inhibitor N-(n-butyl) thiophosphoric triamide (NBPT) didn't alter the ure gene copy number present pre- versus post-fertilization. However, it is unknown

 doi:10.20944/preprints202303.0168.v1

from our data whether urease enzyme function was altered in the post-fertilized plots, but NBPT is most successful urease inhibitor on the market reducing ammonium volatilization loss by 53% (Cantarella et al. 2018). Further analysis of urease gene expression and urease enzyme function is needed to validate that NBPT inhibition is on the expression or functional level.

Carbon substrate metabolism predicted using CAZy in the switchgrass roots of Lux Arbor appears to be limited in terms of both the bulk metabolic potential and the individual genome level. CAZy abundances were differential under fertilizer treatment, and most were depleted in post-fertilized plots, including genes related to cellobiose, cellulose, xylan, wood-degradation, chitin, and N-acetylglucosamine. This may be due to increases in exudation of simpler carbon sources by switchgrass roots, either stimulated by fertilization or through other shifts in the system between sampling timepoints. Fertilizer treatment has been previously reported to impact CAZy enzyme function (Zhang et al. 2015). Enzyme assays could be used to validate CAZy function shifts in Lux Arbor. The MAG metabolic potentials based on CAZy had similar genes in very high abundance with limited diversity relating to cellulose, chitin, or simple sugars.

The soils sampled here were from the same sample block, sampled just two weeks apart, immediately before, and two weeks after fertilization. We found strong fertilizer effects in some blocks and little to no effect in others suggesting resilience of the rhizosphere microbiome or variation in the timescales of these responses. Even with this variation, we were able to better resolve individual MAG communities present within the switchgrass rhizosphere. We are also not able to definitively conclude that the shifts we observe are due to N fertilization since it is confounded with the time of sampling, but the functional relationships we characterize suggest that they are due to varying inputs of nitrogen and carbon in this system.

## 5. Conclusions

The Lux Arbor marginal land switchgrass plots provide an excellent model system to study a lower-complexity and diversity rhizosphere soil ecosystem. We have described a snap-shot of how a N fertilization event impacts the bulk metabolic potential of carbon and N metabolism and resolved MAGs relating to N-fixation (Janthinobacterium) and nitrate utilization (*Nitrospira*). We have also characterized the potential roles of several 'most wanted taxa' in the soil, resolving genomes from Thermoanaerobaculia and Dormibacterota. Dormibacterota have the potential for autotrophic CO utilization, which may impact carbon partitioning and storage. Further culture-dependent and multi-omics studies are needed to evaluate the use of Janthinobacterium diazotrophs for ANF in switchgrass grown in marginal lands.

**Supplemental Materials: Figure S1:** Comparison of Metabat2 and Maxbin2 binning statistics; **Figure S2:** Comparison of Metabat2 and Maxbin2 binning statistics**; Table S1**: mOTU adonis statistical testing within phyloseq R results. Paired analysis using sample block (B1-4) "Block," and post-fertilization; **Table S2**: Top 10 mOTU abundance table from phyloseq R. This includes taxonomy of mOTUs and sample metadata; **Table S3:** KEGG KO contig protein-coding orf count table. This includes level 1 to level II raw counts for KEGG KO. The DESeq2 KO paired analysis significant table out of 3204 KO's with nonzero total orf count (p-value < 0.05); **Table S4:** Nitrogenase molybdenum-iron protein alpha chain gene (nifD) blast table. This was blastp analysis against genbank/refseq with score, accession and taxonomy; **Table S5:** Comparison of MAG taxonomy annotation. Comparing metawraps MAG taxonomy tool, Classify genomes tool, Jspecies and ribosomal gene S9 with genbank/refseq with score, accession and taxonomy.

**Author Contributions:** A.G and M.L.F conceived and designed the experiments. A.R and E.E.M completed measurements and experiments of metagenomics. R.A.W.III conducted data analysis, metagenomic assemblies, metagenomic annotation, and cowrote the paper with M.L.F, L.K.T and S.E. All authors contributed to editing and read and approved the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** The U.S. Department of Energy provided support for this research, Office of Science, Office of Biological and Environmental Research (Award DE-SC0018409 to SEE, LKT, MLF), the Great Lakes Bioenergy Research Center, U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research (Award DE-FC02-07ER64494), the National Science Foundation Long-term Ecological Research Program (DEB 1637653) at the Kellogg Biological Station, and Michigan State University AgBioResearch.

**Data Availability Statement:** Raw sequence data, assembled contigs, supplemental data, are all available on https://osf.io/mzrvj/. All code for this study is available on www.github.com/friesenlab/MMPRNT_panicum_metagenome_mags/.

**Acknowledgments:** We would also like to thank the Kellogg Biological Station and Michigan State University AgBioResearch. Special thanks to H. Vander Stel, L. Bell-Dereske, G. Davis, M. Rabbitt, D. Marinas, J. Priebe, C. Landis, and Z. Ye for experimental assistance.

**Conflicts of Interest:** The authors declare that there are no conflict of interest. RAWIII is the CEO of RAW Molecular Systems (RAW), INC, but no financial, IP, or others from RAW INC were used or contributed to the study.

## References

Alneberg, J., Bjarnason, B. S., de Bruijn. I., Schirmer, M., Quick, J., Ijaz, U. Z., Lahti, L., Loman, N. J., Andersson, A. F, and Quince, C. 2014. Binning metagenomic contigs by coverage and composition. Nat. Methods. 11:1144-1146.

Anantharaman, K., Brown, C. T., Burstein, D., Castelle, C. J., Probst, A. J., Thomas, B. C., Williams, K. H., & Banfield, J. F. (2016). Analysis of five complete genome sequences for members of the class Peribacteria in the recently recognized Peregrinibacteria bacterial phylum. PeerJ, 4, e1607. https://doi.org/10.7717/peerj.1607

Berendsen, R. L., Pieterse, C. M., and Bakker PA. 2012. The rhizosphere microbiome and plant health. Trends Plant Sci., 17:478-486.

Bowers, R. M., Kyrpides, N.C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T. B. K., Schulz, F., Jarett, J., Rivers, A. R., Eloe-Fadrosh, E. A., Tringe, S.G., Ivanova, N. N., Copeland, A., Clum, A., Becraft, E. D., Malmstrom, R. R., Birren, B., Podar, M., Bork, P., Weinstock, G. M., Garrity, G. M., Dodsworth, J. A., Yooseph, S., Sutton, G., Glöckner, F. O., Gilbert, J. A., Nelson, W. C., Hallam, S. J., Jungbluth, S. P., Ettema, T. J. G., Tighe, S., Konstantinidis, K. T., Liu, W. T., Baker, B. J., Rattei, T., Eisen, J. A., Hedlund, B., McMahon, K. D., Fierer, N., Knight, R., Finn, R., Cochrane, G., Karsch-Mizrachi, I., Tyson, G. W., Rinke, C., Genome Standards Consortium, Lapidus, A., Meyer, F., Yilmaz, P., Parks, D. H., Eren, A. M., Schriml, L., Banfield, J. F., Hugenholtz, P., and Woyke, T. 2017. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat. Biotechnol. 35:725-731.

Brewer, T. E., Aronson, E. L., Arogyaswamy, K., Billings, S. A., Botthoff, J. K., Campbell, A. N., Dove, N. C., Fairbanks, D., Gallery, R. E., Hart, S. C., Kaye, J., King, G., Logan, G., Lohse, K. A., Maltz, M.R., Mayorga, E., O'Neill, C., Owens, S. M., Packman, A., Pett-Ridge, J., Plante, A. F., Richter, D. D., Silver, W. L., Yang, W. H., and Fierer, N. 2019. Ecological and genomic attributes of novel bacterial taxa that thrive in subsurface soil horizons. Mbio 10:e01318-19.

Buchfink, B., Xie, C., and Huson, D. H. 2015. Fast and sensitive protein alignment using DIAMOND. Nat. Methods 12:59-60.

Butterfield, C. N., Li, Z., Andeer, P. F., Spaulding, S., Thomas, B. C., Singh, A., Hettich, R. L., Suttle, K. B., Probst, A. J., Tringe, S. G., Northen, T., Pan, C., and Banfield, J. F. 2016. Proteogenomic analyses indicate bacterial methylotrophy and archaeal heterotrophy are prevalent below the grass root zone. PeerJ 4:e2687.

Cantarella, H., Otto, R., Soares, J. R., and Silva, A. G. B. 2018. Agronomic efficiency of NBPT as a urease inhibitor: A review. J. Adv. Res. 13:19-27.

Carini, P. 2019. A "cultural" renaissance: genomics breathes new life into an old craft. Msystems 4:e00092-19.

Choi, J., Yang, F., Stepanauskas, R., Cardenas, E., Garoutte, A., Williams, R., Flater, J., Tiedje, J. M., Hofmockel, K. S., Gelder, B., Howe, A. 2017. Strategies to improve reference databases for soil microbiomes. ISMEJ. 11:829-834.

Chen, H., Yang, Z. K., Yip, D., Morris, R. H., Lebreux, S. J., Cregger, M. A., Klingeman, D. M., Hui, D., Hettich, R. L., Wilhelm, S. W., Wang, G., Löffler, F. E., & Schadt, C. W. 2019. One-time nitrogen fertilization shifts switchgrass soil microbiomes within a context of larger spatial and temporal variation. *PloS one*, *14*(6), e0211310. https://doi.org/10.1371/journal.pone.0211310

Demanèche, S., Philippot, L., David, M. M., Navarro, E., Vogel, T. M., and Simonet, P. 2009. Characterization of denitrification gene clusters of soil bacteria via a metagenomic approach. Appl. Environ. Microbiol. 75:534-537.

Di, H. J., Cameron, K. C., Podolyan, A, and Robinson, A. 2014. Effect of soil moisture status and a nitrification inhibitor, dicyandiamide, on ammonia oxidizer and denitrifier growth and nitrous oxide emissions in a grassland soil. Soil Biology and Biochemistry. 73:59-68.

Diamond, S., Andeer, P. F., Li, Z., Crits-Christoph, A., Burstein, D., Anantharaman, K., Lane, K. R., Thomas, B. C., Pan, C., Northen, T. R., and Banfield JF. 2019. Mediterranean grassland soil C-N compound turnover is dependent on rainfall and depth, and is mediated by genomically divergent microorganisms. Nat. Microbiol. 4:1356-1367.

Dvortsov, I. A., Lunina, N. A., Chekanovskaya, L. A., Schwarz, W. H., Zverlov, V. V., and Velikodvorskaya, G. A. Carbohydrate-binding properties of a separately folding protein module from beta-1,3-glucanase Lic16A of *Clostridium thermocellum*. Microbiology. 155:2442-2449.

Emery I, Mueller S, Qin Z, Dunn JB. Evaluating the Potential of Marginal Land for Cellulosic Feedstock Production and Carbon Sequestration in the United States. Environ Sci Technol. 2017 51(1):733-741. doi: 10.1021/acs.est.6b04189.

Friesen ML, Porter SS, Stark SC, von Wettberg EJ, Sachs JL, Martinez-Romero E. Microbially Mediated Plant Functional Traits. 2011. Annu. Rev. Ecol. Evol. Syst. 42, 23-46.

Giles, M. E., Morley, N. J., Baggs, E. M., and Daniell, T. J. (2012). Soil nitrate reducing processes – drivers, mechanisms for spatial variation and significance for nitrous oxide production. Front. Microbiol. 3:407. doi: 10.3389/fmicb.2012.00407

Hara., S., Desyatkin, R. V., and Hashidoko, Y. 2014. Investigation of the mechanisms underlying the high acetylene-reducing activity exhibited by the soil bacterial community from BC2 horizon in the permafrost zone of the East Siberian larch forest bed. J. Appl. Microbiol. 116:865-876.

Hartmann, A., and Rothballer, M. 2008. Schmid Lorenz Hiltner, a pioneer in rhizosphere microbial ecology and soil bacteriology research. Plant Soil, 312:7-14.

Howe, A. C., Jansson, J. K., Malfatti, S. A., Tringe, S. G., Tiedje, J. M., and Brown, C. T. 2014. Tackling soil diversity with the assembly of large, complex metagenomes. Proc. Natl. Acad. Sci. USA 111:4904–4909.

Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., von Mering, C., and Bork, P. 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper. Mol. Biol. Evol. 2017 34:2115-2122.

Hug, L. A., Baker, B. J., Anantharaman, K., Brown, C. T., Probst, A. J., Castelle, C. J., Butterfield, C. N., Hernsdorf, A. W., Amano, Y., Ise, K., Suzuki, Y., Dudek, N., Relman, D. A., Finstad, K. M., Amundson, R., Thomas, B. C., and Banfield, J. F. 2016. A new view of the tree of life. Nat. Microbiol. 1:16048.

Janzen, D. H., 1985. The natural history of mutualisms. In The Biology of Mutualism, ed. DH Boucher, 3:40-99. London: Croom Helm.

Ji, M., Greening, C., Vanwonterghem, I., Carere, C. R., Bay, S. K., Steen, J. A., Montgomery, K., Lines, T., Beardall, J., van Dorst, J., Snape, I., Stott, M. B., Hugenholtz, P., and Ferrari, B. C. 2017. Atmospheric trace gases support primary production in Antarctic desert surface soil. Nature. 552:400-403.

Jones, F. P., Clark, I. M., King, R., Shaw, L. J., Woodward, M. J., and Hirsch, P. R. 2016. Novel european free-living, non-diazotrophic *Bradyrhizobium* isolates from contrasting soils that lack nodulation and nitrogen fixation genes - a genome comparison. Sci. Rep. 6:25858.

Kang, D., Li, F., Kirton, E. S., Thomas, A., Egan, R. S., An, H., and Wang, Z. 2019. MetaBAT2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ Preprints 7:e27522v1

Kantor, R. S., Wrighton, K. C., Handley, K. M., Sharon, I., Hug, L. A., Castelle, C. J., Thomas, B. C., and Banfield, J. F. 2013. Small genomes and sparse metabolisms of sediment-associated bacteria from four candidate phyla. MBio. 4:e00708-13.

Koch, H., van Kessel, M.A.H.J., and Lücker, S. 2018. Complete nitrification: insights into the ecophysiology of comammox *Nitrospira*. Appl. Microbiol. Biotechnol. 103:177-189.

Kroeger, M. E., Delmont, T. O., Eren, A. M., Meyer, K. M., Guo, J., Khan, K., Rodrigues, J. L. M., Bohannan, B. J. M., Tringe, S. G., Borges, C. D., Tiedje, J. M., Tsai, S. M., and Nüsslein, K. 2018. New biological insights Into how deforestation in Amazonia affects soil microbial communities Using metagenomics and metagenome-assembled genomes. Front Microbiol. 9:1635.

Lan, T., Han, Y., Roelcke, M., Nieder, R., and Cai, Z. 2013. Effects of the nitrification inhibitor dicyandiamide (DCD) on gross N transformation rates and mitigating $N_2O$ emission in paddy soils. Soil Biology and Biochemistry. 67:174-182.

Li, D., Liu, C. M., Luo, R., Sadakane, K., and Lam, T. W. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 31:1674-1676.

Liu, S. Y., Tang, Y. X., Wang, D. C., Lin, N. Q., and Zhou, J. N. 2016. Identification and characterization of a new Enterobacter onion bulb decay caused by *Lelliottia amnigena* in China. App. Micro. Open Access. 2:114.

Losey, N. A., Stevenson, B. S., Busse, H. J., Sinninghe Damste, J. S., Rijpstra, W. I., Rudd, S., and Lawson, P. A. *Thermoanaerobaculum aquaticum* gen. nov., sp. nov., the first cultivated member of Acidobacteria subdivision 23, isolated from a hot spring. 2013. Int. J. Syst. Evol. Microbiol. 63:4149-4157.

Love, M. I., Huber, W., and Anders, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15:550.

McLaughlin, S. B., and Kzos, L. A. 2005. Development of switchgrass (*Panicum virgatum*) as a bioenergy feedstock in the United States. Biomass and Bioenergy. 28:515–535.

McMurdie, P. J., and Holmes, S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One. 2013 8(4):e61217. doi: 10.1371/journal.pone.0061217.

Milanese, A., Mende, D. R., Paoli, L., Salazar, G., Ruscheweyh, H. J., Cuenca, M., Hingamp, P., Alves, R., Costea, P. I., Coelho, L. P., Schmidt, T. S. B., Almeida, A., Mitchell, A. L., Finn, R. D., Huerta-Cepas, J., Bork, P., Zeller, G., and Sunagawa, S. 2019. Microbial abundance, activity and population genomic profiling with mOTUs2. Nat. Commun. 10:1014.

Monti, M., Barbanti, L., Zatta, A., and Zegada-Lizarazu, W. 2012. The contribution of switchgrass in reducing GHG emissions. Global Change Biology Bioenergy 4:420-434

Ning, J., Ai, S., and Cui, L. 2018. Dicyandiamide has more inhibitory activities on nitrification than thiosulfate. PLoS One 13:e0200598.

Naether, A., Foesel, B. U., Naegele, V., Wüst, P. K., Weinert, J., Bonkowski, M., Alt, F., Oelmann, Y., Polle, A., Lohaus, G., Gockel, S., Hemp, A., Kalko, E. K., Linsenmair, K. E., Pfeiffer, S., Renner, S., Schöning, I., Weisser, W. W., Wells, K., Fischer, M., Overmann, J., and Friedrich, M. W. 2012. Environmental factors affect Acidobacterial communities below the subgroup level in grassland and forest soils. Appl. Environ. Microbiol. 78:7398-7406.

Orellana, L. H., Chee-Sanford, J. C., Sanford, R. A., Löffler, F. E., and Konstantinidis, K. T. 2018. Year-round shotgun metagenomes reveal stable microbial communities in agricultural soils and novel ammonia oxidizers responding to fertilization. Appl. Environ. Microbiol. 84:e01646-17.

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.

Parks, D. H., Rinke, C., Chuvochina, M., Chaumeil, P. A., Woodcroft, B. J., Evans, P. N., Hugenholtz, P., Tyson, G. W. 2017. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. Nat. Microbiol. 2:1533-1542.

Philippot, L., Raaijmakers, J. M., Lemanceau, P., and van der Putten, W. H. 2013. Going back to the roots: the microbial ecology of the rhizosphere. Nat. Rev. Microbiol. 11:789-799.

Ramírez-Puebla, S. T., Servín-Garcidueñas, L. E., Jiménez-Marín, B., Bolaños, L. M., Rosenblueth, M., Martínez, J., Rogel, M. A., Ormeño-Orrillo, E., and Martínez-Romero, E. 2013. Gut and root microbiota commonalities. Appl. Environ. Microbiol. 2013. 79:2-9.

Richter, M., Rosselló-Móra, R., Oliver Glöckner, F., and Peplies, J. 2016. JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. Bioinformatics 32:929-931.

Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N. N., Anderson, I. J., Cheng, J. F., Darling, A., Malfatti, S., Swan, B. K., Gies, E. A., Dodsworth, J. A., Hedlund, B. P., Tsiamis, G., Sievert, S. M., Liu, W. T., Eisen, J. A., Hallam, S. J., Kyrpides, N. C., Stepanauskas, R., Rubin, E. M., Hugenholtz, P., and Woyke, T. 2013. Insights into the phylogeny and coding potential of microbial dark matter. Nature 499:431-437.

Roley, S. S., Duncan, D. S., Liang, D., Garoutte, A., Jackson, R. D., Tiedje, J. M., and Robertson, G. P. 2018. Associative nitrogen fixation (ANF) in switchgrass (*Panicum virgatum L.*) across a nitrogen input gradient. PLoS One 13:e0197320.

Roley, S. S., Xue, C., Hamilton, S. K., Tiedje, J. M., and Robertson, G. P. 2019. Isotopic evidence for episodic nitrogen fixation in switchgrass (*Panicum virgatum L.*). Soil Biology and Biochemistry. 129:90-98.

Ruan, L., Bhardwaj, A. K., Hamilton, S. K., and Robertson, G. P. 2016. Nitrogen fertilization challenges the climate benefit of cellulosic biofuels. Environ. Res. Lett. 11:064007

Schmer, M. R., Vogel, K. P., Mitchell, R. B., and Perrin, R. K. 2008. Net energy of cellulosic ethanol from switchgrass. Proc. Natl. Acad. Sci. USA. 105:464-469.

Seemann, T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 30:2068-2069.

Singer, E., Bonnette, J., Kenaley, S. C., Woyke, T., & Juenger, T. E. (2019). Plant compartment and genetic variation drive microbiome composition in switchgrass roots. *Environmental microbiology reports*, *11*(2), 185–195. https://doi.org/10.1111/1758-2229.12727

Starr, E. P., Shi, S., Blazewicz, S. J., Probst, A. J., Herman, D. J., Firestone, M. K., and Banfield, J. F. 2018. Stable isotope informed genome-resolved metagenomics reveals that Saccharibacteria utilize microbially-processed plant-derived carbon. Microbiome 6:122.

Uritskiy, G. V., DiRuggiero, J., Taylor J. MetaWRAP-a flexible pipeline for genome-resolved metagenomic data analysis. Microbiome. 2018 6(1):158. doi: 10.1186/s40168-018-0541-1.

VanInsberghe, D., Maas, K. R., Cardenas, E., Strachan, C. R., Hallam, S. J., and Mohn, W. W. Non-symbiotic *Bradyrhizobium* ecotypes dominate North American forest soils. ISME J. 2015 9(11):2435-2441.

Woodcroft, B. J., Singleton, C. M., Boyd, J. A., Evans, P. N., Emerson, J. B., Zayed, A. A. F., Hoelzle, R. D., Lamberton, T. O., McCalley, C. K., Hodgkins, S. B., Wilson, R. M., Purvine, S. O., Nicora, C. D., Li, C., Frolking, S., Chanton, J. P., Crill, P. M., Saleska, S. R., Rich, V. I., Tyson, G. W. 2018. Genome-centric view of carbon processing in thawing permafrost. Nature. 560:49-54.

White III, R. A., Bottos, E. M., Roy Chowdhury, T., Zucker, J. D., Brislawn, C. J., Nicora, C. D., Fansler, S. J., Glaesemann, K. R., Glass, K., Jansson, J. K. 2016. Moleculo long-read sequencing facilitates assembly and genomic binning from complex soil metagenomes. Msystems 1:e00045-16.

White III, R. A., Rivas-Ubach, A., Borkum, M. I., Köberl, M., Bilbao, A., Colby, S. M., Hoyt, D. W., Bingol, K., Kim, Y. M., Wendler, J. P., Hixson, K. K., and Jansson, C. 2017 a. The state of rhizospheric science in the era of multi-omics: a practical guide to omics technologies. Rhizosphere 3:212-221.

White III, R. A., Borkum, M. I., Rivas-Ubach, A., Bilbao, A., Wendler, J. P., Colby, S. M., Köberl, M., and Jansson, C. 2017 b. From data to knowledge: the future of multi-omics data analysis for the rhizosphere. Rhizosphere 3:222-229.

White III, R. A., Brown, J., Colby, S., Overall, C. C., Lee, J., Zucker, J. D., Glaesemann, K. R., Jansson, C., and Jansson, J. K. 2017 c. ATLAS (Automatic Tool for Local Assembly Structures) – a comprehensive infrastructure for assembly, annotation, and genomic binning of metagenomic and metatranscriptomic data. PeerJ Preprints 5:e2843v1.

Wright, L. 2007. Historical perspective on how and why switchgrass was selected as a "Model" high-potential energy crop. Oak Ridge National Laboratory, Oak Ridge, TN. ORNL/TM-2007/109.

Wright, L., and Turhollow, A. 2010. Switchgrass selection as a “model” bioenergy crop: a history of the process. Biomass and Bioenergy 34:851-868

 doi:10.20944/preprints202303.0168.v1

Wrighton, K. C., Thomas, B. C., Sharon, I., Miller, C. S., Castelle, C. J., VerBerkmoes, N. C., Wilkins, M. J., Hettich, R. L., Lipton, M. S., Williams, K. H., Long, P. E., Banfield, J. F. 2012. Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. Science 337:1661-1665.

Yang, W., Wang, Y., Tago, K., Tokuda, S., and Hayatsu, M. 2017. Comparison of the effects of phenylhydrazine hydrochloride and dicyandiamide on ammonia-oxidizing bacteria and archaea in andosols. Front Microbiol. 8:2226.

Wu, Y. W., Simmons, B. A., and Singer, S. W. 2016. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics 32:605-607.

Zhang, L., Chen, W., Burger, M., Yang, L., Gong, P., and Wu, Z. 2015. Changes in soil carbon and enzyme activity as a result of different long-term fertilization regimes in a greenhouse field. PLoS One. 10:e0118371.

Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., Busk, P. K., Xu, Y., and Yin, Y. 2018. dbCAN2: a meta-server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 46:W95-W101.