PNAS

Supporting Information for

Nutrient and moisture limitation reveal keystone metabolites that link switchgrass rhizosphere metabolome and microbiome dynamics

Nameer R. Baker[1], Kateryna Zhalnina[1,2], Mengting Yuan, Don Herman, Javier A. Ceja-Navarro, Joelle Sasse, Jacob S. Jordan, Benjamin P. Bowen, Liyou Wu, Christina Fossum, Aaron Chew, Ying Fu, Malay Saha, Jizhong Zhou, Jennifer Pett-Ridge, Trent R. Northen, Mary K. Firestone[2]

[1]N.R.B. and K.Z. contributed equally to this work.
[2]To whom correspondence may be addressed. Email: mkfstone@berkeley.edu or kzhalnina@lbl.gov.

This PDF file includes:

Other supporting materials for this manuscript include the following:

Datasets S1 to S5 (separate files)

Supporting Information Text

Materials and Methods:

Experimental Design

One meter deep soil mesocosms were constructed in 19.7 cm diameter impact-resistance polycarbonate tubes with the ”A”, “B”, and “C” horizon soils to recreate the marginal soil environment. Six replicate mesocosms were created per treatment, for a total of 30 mesocosms. Prior to packing the mesocosms to field bulk density (1.41 ± 0.04, 1.53 ± 0.18, 1.64 ± 0.07 g dry soil cm^{-3} for the “A” “B” and “C” horizons, respectively), we added slow-release coated urea (ESN Smart Nitrogen, 44-0-0, Agrium) to the top horizon of the +N treatment mesocosms (0.13 g kg^{-1} soil), and rock phosphate (0-3-0, Espoma) to the +P treatment mesocosms (0.48 g kg^{-1} soil). Coated urea and rock phosphate were both added (at concentrations as above) to a third set for the +NP treatment. Two further sets of mesocosms were created - a control treatment with no nutrient amendments, and a low water (–W) treatment which received half the water of all the other treatments once plants became established. Mesocosms were watered with 50% (-W) or 100% (control, +N, +P, +NP) of the mean monthly rainfall (2012-2017, NOAA) at the field site in Oklahoma in the summer months (roughly 100 mL each day). Soil moisture sensors (EC-20; METER Group, Pullman, WA) were installed in the “A” horizon of the control and –W treatment to confirm differences in soil moisture. Individual clonal ramets of the Alamo switchgrass genotype, NFSG 18-01, from the Nested Association Mapping population generated at the Nobel Research Institute were planted in each mesocosm in May 2017 and grown in the greenhouse, under a natural light regime and a 32 °C daytime and 21°C nighttime temperature cycle. Following a period of 18 weeks, the mesocosms were subject to destructive harvesting, which involved longitudinally cutting open the polycarbonate tube and processing the soil by horizon.

Sample Collection and Processing

Roots and associated rhizosphere soil from the top “A” horizon were immediately collected for DNA extractions in 15 mL tubes containing 5 mL Lifeguard Soil Preservation Solution (QIAGEN) using chilled soil-processing trays. For metabolite extractions, roots and rhizosphere soil from all three horizons were collected and immediately placed on dry ice and then stored at -80 °C. Bulk soils from the “A” horizon were collected and stored at 4 °C for gravimetric soil moisture measurements and all remaining roots were collected, dried and weighed.

DNA extractions

Rhizosphere soil in Lifeguard solution was pelleted for DNA extractions by vortexing samples for 2 min to fully resuspend soil particles before centrifuging for 5 min at 5,000 g and 4 °C, and then re-centrifuging for 2 min after picking out remaining roots, if necessary. The resulting supernatant was discarded and tubes with soil pellets were stored at -80 °C.

Soil pellets were collected by cracking surrounding tubes with a foil-wrapped, DNAse-free hammer on a foil-wrapped, DNAse-free ring stand underlain by dry ice. Pellets were collected in Whirl Paks (Nasco) and stored on dry ice. 0.5 g of each soil sample was aliquoted into a sterile, pre-weighed 2 mL microcentrifuge tube (screw-top, self-standing) after fragmenting the soil pellets in Whirl Paks with the hammer.

Soil DNA was extracted using a modified RNA/DNA phenol chloroform co-extraction protocol via bead-beating (1, 2). Notably, the 5% hexadecyl-trimethyammonium bromide/0.7 M NaCl/240 mM K-PO4 buffer (pH 8) was modified to include 1% β-mercaptoethanol, PEG 8000 was used in place of PEG 6000, and GlycoBlue was used to stain nucleic acid pellets. The combined RNA and DNA in resulting pellets were purified using a modified lithium chloride extraction protocol (3).

Purified DNA pellets were stored at -20 °C. DNA was quantified within a week of extraction via PicoGreen fluorescence (4).

Extracted DNA was then fractionated via isopynic centrifugation in a cesium chloride density gradient, using the same method detailed by Blazewicz et al. (5). Immediately following centrifugation, the sample was fractionated into ~40 fractions using a syringe pump. DNA concentrations in each fraction were quantified via PicoGreen fluorescence, and fractions containing DNA were recombined with neighboring fractions to create seven "bins" of equal DNA content (not concentration). Microbial community composition of the DNA in each bin was characterized with a sequencing library prepared at the University of Oklahoma via a phasing amplification technique targeting the V4 region of the 16s ribosomal RNA gene with the 515F and 806R primer set (6, 7). The Illumina MiSeq platform was used to sequence the samples in a 2x250 bp format. We note that the process of fractionating the samples before sequencing likely altered the community that we observed - however, this DNA should also be cleaner as a result having been run through the cesium chloride gradient and the results of the sequencing run should contain fewer artifacts.

Analysis of 16S rRNA gene sequences pipeline

A total of 10,516,421 raw reads were imported (using the Earth Microbiome Project protocol) into Qiime2, and were demultiplexed, trimmed of non-biological primer sequences, and denoised using DADA2 (225 bp forward reads, 223 bdp reverse) (8). This resulted in 7856 denoised ASVs, which accounted for 3,942,210 (37%) of the total reads. These ASVs were aligned to a rooted tree and assigned taxonomy by a feature classifier. The feature classifier was trained on reads extracted using the aforementioned primer sequences from the SILVA_132 16S reference database for 99% identity. The resulting ASVs were grouped by initial soil sample, such that ASVs from the seven bins the extracted DNA had been fractionated into were re-combined to assess the nature of the switchgrass rhizosphere 16S community and its response to our treatments. These reads and the rooted tree were exported from Qiime2 for further analysis via *phyloseq* in R (9).

In *phyloseq*, all chloroplast or mitochondrial sequences were removed from the dataset, as were bacterial and archaeal sequences that lacked a designation at the phylum level. 7481 ASVs accounting for 3,792,761 reads passed this filter, and were used to calculate α-diversity metrics via the *estimate_richness* function. We tested for significant α-diversity treatment effects using the *lm* and *anova* functions in R, and used the *emmeans* function to test the significance of pairwise comparisons between treatments. For all analyses other than α-diversity, all sequences that only occurred once or twice across the entire dataset were removed in phyloseq. This resulted in 7093 ASVs accounting for 3,792,087 reads, with a range of 104,776 - 194,526 reads per sample. The *distance* function was used to develop a Unifrac dissimilarity matrix by sample that was then assessed for significant treatment effects using the *adonis* function of the *vegan* package in R, with 99999 permutations. We assessed significance of pairwise comparisons between treatments with the *pairwise.adonis* function of the *pairwiseAdonis* package, with 99999 permutations and used the Hochberg correction for multiple comparisons.

Differentially abundant taxa between each of the individual treatments and the control samples were assessed with the *DESeq* function of the *DESeq2* package, employing a local fit for dispersion estimates and optimizing for a significance cutoff of $P < 0.01$ after visually assessing the distribution of P-values. Effect sizes for the magnitude of differences in ASV abundance between the relevant treatment and control samples were calculated as Cohen's *d* in R, using pooled standard deviations between the control samples and those from the relevant treatment.

Cluster analysis of ASVs and metabolite abundances were performed using the *vegan* package in R (10). The most significant associations were selected by filtering (i) differentially abundant ASVs with more than three significant positive or negative correlations (Spearman's rank

correlation, $r > 0.7$ or $r < -0.7$, $P < 0.05$) with metabolites; and (ii) metabolites with more than one significant positive or negative correlation (Spearman's rank correlation, $r > 0.7$, or $r < -0.7$, $P<0.05$) with ASVs.

Soil metabolite extraction, analysis and identification

To extract soil metabolites, 10 ml of ice-cold LC-MS-grade water (pH 7.4) was added to 15 ml polypropylene Falcon tubes filled with roots and rhizosphere soil and held on ice. Samples were vigorously vortexed for 10 sec to detach rhizosphere soil from the root. Roots were removed with sterile tweezers and samples were shaken on an orbital shaker (Orbital-Genie, Scientific Industries, Bohemia, NY) at 200 rpm for 1 h at 4 °C then centrifuged at 3220 g for 15 min at 4 °C. For each sample, the supernatant was filtered through a 0.45 µm syringe filter (Pall Acrodisc Supor membrane) into a 15 mL tube. Two ml of supernatant was aliquoted for total organic carbon (TOC) analysis, using a Shimadzu TOC-L Analyzer. Five additional ml of supernatant transferred into a 15 ml Falcon tube, lyophilized using a Labconoco FreeZone 2.5 lyophilizer and stored at -80 °C.

Dried metabolite extracts were resuspended in 1 ml LC-MS grade methanol (Honeywell Burdick & Jackson, Morristown, NJ, USA), vortexed, sonicated on ice for 30 min and incubated at 4 °C overnight. Samples were centrifuged at 3220 g for 15 min at 4 °C, after which the supernatant was dried in a Savant SpeedVac SPD111V (Thermo Scientific, Waltham, MA) for 3 h. Dried samples were resuspended in 100% ice-cold methanol to achieve a final concentration of 3000 ppm/L TOC; these were then sonicated for 15 min using an ultrasonic bath (VWR). Internal standards (1 µg per ml 2-amino-3-bromo-5-methylbenzoic acid, 5 µg per ml ^{13}C-^{15}N-l-phenylalanine and 2 µg per ml 9-anthracenecarboxylic acid) were spiked into the methanol used to resuspend samples (Supplementary Table S4). The resulting extracts were filtered with 0.22 µm microcentrifuge PVDF filters (Merck Millipore), and 150 µl aliquots were transferred to LC-MS vials for metabolite analysis.

All chromatography was performed using an Agilent 1290 LC stack, with MS and tandem mass spectrometry (MS/MS) fragmentation data collected in both positive and negative ion mode using a Thermo Q Exactive mass spectrometer (Thermo Fisher Scientific) in the Northen Lab at Lawrence Berkeley National Laboratory (Table S4). For each 3 µl sample injection, full MS spectra were acquired for m/z 70–1,050 at 70,000 FWHM (full-width at half-maximum) resolution. MS/MS fragmentation data were acquired using collision energies of 10-40 eV at 17,500 resolution. Sample injection order was randomized and an injection blank of only methanol was run between samples. Normal-phase chromatography was performed using a HILIC column (Agilent InfinityLab Poroshell 120 HILIC-Z, 150 mm × 2.1 mm, 2.7 µm) warmed to 40 °C with a flow rate of 0.45 ml min^{-1} equilibrated with 100% buffer B (95:5 acetonitrile:water w/ 5 mM ammonium acetate) for 1 min, followed by a linear gradient diluting buffer B down to 89% with buffer A (100% water w/ 5 mM ammonium acetate and 5 µM methylene- di-phosphonic acid) for 10 min, then down to 70% B over 4.75 min, then down to 20% B over 0.5 min, and then isocratically held at 20% B for 2.25 min (Table S4).

Metabolite Atlas software analyzed the metabolomics data, providing extracted ion chromatograms and peak heights for each metabolite (11). Authentic chemical standards were used to confirm metabolite identification. Included in Table S4 are annotations in cases where there are multiple unresolvable peaks with the same m/z, retention time, and MS/MS. A total of 97 level 1 unique metabolites and two metabolites with putative identifications were identified. 27 metabolites were identified in both positive and negative modes and to avoid redundancy only identification in single mode was used for future analysis based on the peak height and quality. All identified metabolites were detected in at least four out of six replicates from at least one treatment (Table S4).

Network Analysis of Rhizosphere ASV-Metabolite Associations

To determine covariance between metabolites and microbial ASVs, we constructed a correlation network based on the relative abundances of metabolites and relative abundances of ASVs across all treatments. To prevent false positives, we only included ASVs with non-zero abundances in at least 15 of the 25 samples. Spearman correlations were calculated for each of the metabolite-metabolite, ASV-ASV and metabolite-ASV pairs. A Random Matrix Theory (RMT)-based approach determined 0.710 as the correlation coefficient cutoff that controls false discovery in our network by separating noise vs. non-random correlations (12). This RMT-based approach has been previously used to construct correlation-based networks of complex microbial systems (13, 14), and is available through the Molecular Ecological Network Analysis (MENA) pipeline (http://ieg4.rccc.ou.edu/MENA/). To construct the networks, we included only metabolite-ASV pairs with an abundance correlation coefficient above the threshold of 0.710, and discarded links within metabolite species or within 16S ASVs to construct the network. Discarding non-metabolite-ASV links partly alleviates the potential bias caused by compositional data, as the abundances of ASVs and metabolites were independently derived; also, compositional data bias should be a minor problem in high-diversity communities (15). Positive and negative correlations correspond to positive and negative links, respectively.

Based on previously developed criteria (14, 16, 17), we defined putative 'keystone metabolites' or 'keystone ASVs' from an individual node's role in network topology as follows: the network was separated into modules using a fast greedy algorithm, and the within-module connectivity (zi) and among-module connectivity (pi) were calculated for each node (16). Nodes with zi >2.5 are designated as module hubs, while nodes with pi >0.62 were designated as connectors among different modules. Nodes with both zi >2.5 and pi > 0.62 stretch among the whole network were designated as network hubs (17). These module hubs and network hubs were defined as keystone metabolites or keystone ASVs based on the node's identity. Correlation calculations, network construction, and network topology analysis were conducted with the igraph package (18). The network was visualized using Cytoscape (19). In addition, to identify associations between metabolites, microbial communities, and treatments, we performed hierarchical clustering analysis using the vegan package (10). For the analysis we selected differentially abundant ASVs determined with the DESeq2, that had more than three significant positive or negative Spearman's rank correlations with metabolites ($r > 0.7$, $P < 0.05$) and metabolites with more than one significant positive or negative correlation with ASVs ($r > 0.7$, $P < 0.05$).

Plant phenotype response to serotonin

To test the effect of serotonin on plant growth, surface-sterilized Alamo switchgrass seeds were sown on ⅕ Murashige and Skoog (MS) basal salt mixture M524 (Phyto Technology Laboratories) (0.87 g/L MS salts, pH 7.3, and 8 g/L agar). Nine biological replicates (n = 9) of seven-day-old seedlings were transferred to ⅕ MS agar plates supplemented with 0.1 mM serotonin (Sigma-Aldrich) or with purified H_2O. Plates with seedlings were incubated at 24 °C on a 16 h/8 h day/night cycle, with humidity maintained at 70% and irradiance at 250 μE m^{-2} s^{-1}. After twenty-five days, switchgrass plants were harvested, and root and shoot biomass were measured. Root length and root number were quantified using the SmartRoot plugin (version 4.21) in ImageJ (version 2.0.0) (20). Significant differences between treatments were determined using an ANOVA test ($P < 0.05$).

Microbial growth curves with serotonin

To test the effect of serotonin on microbial growth responses, we selected six bacterial isolates from common rhizosphere genera: *Reyranella*, *Mucilaginibacter*, *Methylobacterium*, *Paenarthrobacter*, *Burkholderia*, and *Mesorhizobium*. Growth curves were performed in 1/10 R2A

medium for 130 hours with either 0, 0.1, or 0.5 mM serotonin. Four replicates of each isolate were inoculated in a 96-well plate and grown at 30 °C, shaking once per hour at 200 rpm before optical density measurement at 600 nm (OD_{600}). After 130 hours of isolate growth with 0, 0.1 or 0.5 mM serotonin, the culture OD_{600} was compared to that of a control treatment without serotonin (0 mM). Optical density responses were analyzed using a Kruskal-Wallis test after the OD_{600} of uninoculated blanks was subtracted from the inoculated samples. The growth curve experiments have been repeated twice.

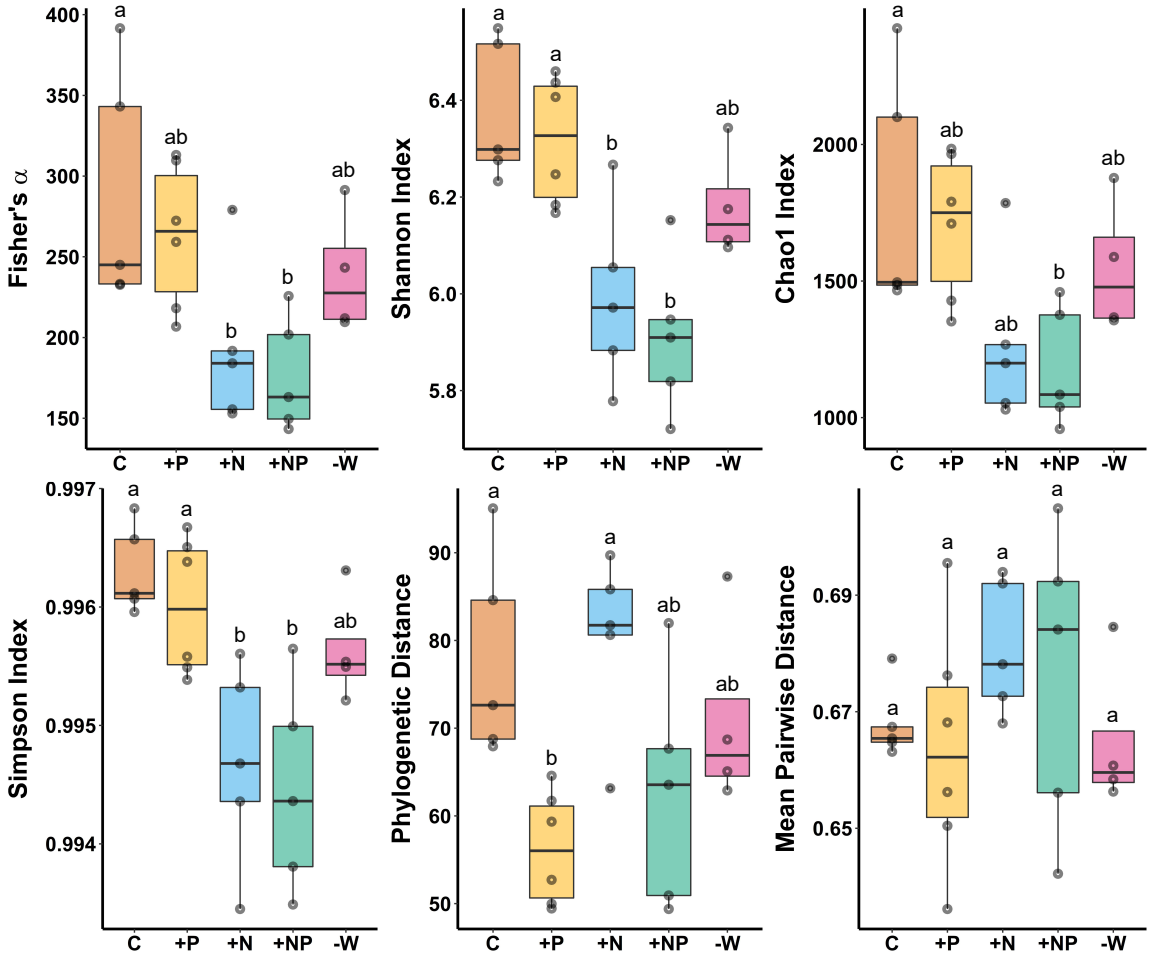

Fig. S1. Box-whisker plots (median and 25-75% quartiles) of 16S amplicon sequence variant α-diversity metrics for switchgrass rhizosphere microbial communities grown under five treatments: controls ('C') with nutrient-poor marginal soil, '+P', '+N', and '+NP' mesocosms with phosphorus and/or nitrogen amendments in the top soil horizon, and '-W' mesocosms which received 50% less water relative to the other treatments. Letters represent significantly different post hoc pairwise comparisons via Tukey's test ($P < 0.05$, n=6).

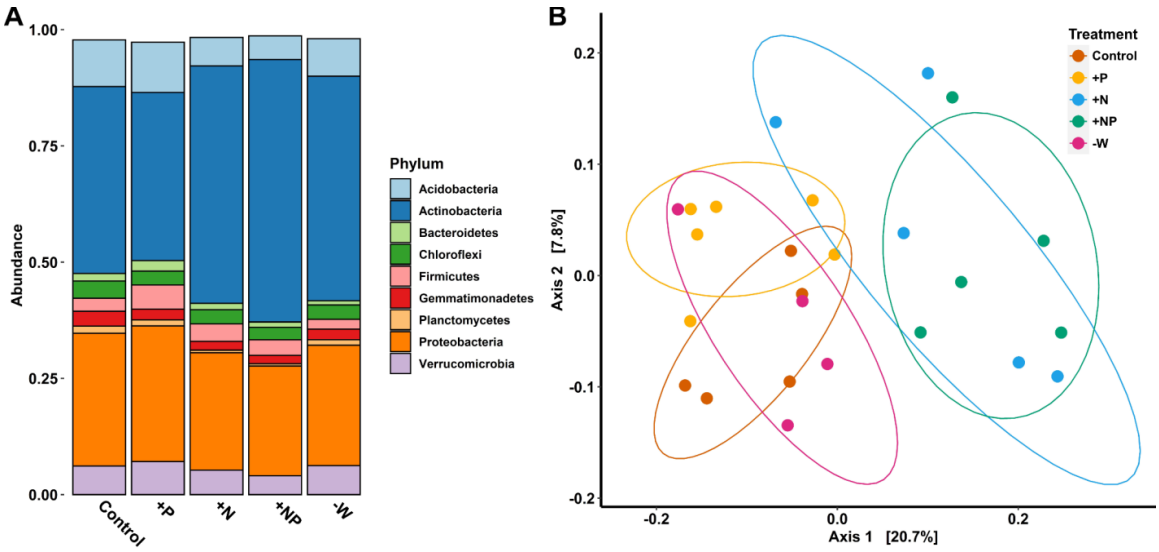

Fig. S2. Differences in microbial community composition by treatment, illustrated by (*A*) a barplot of relative abundances of all bacterial phyla that made up more than 1% of the total rhizosphere community and (*B*) principal components analysis of Unifrac distance matrices drawn from the distribution of ASVs in each sample (ellipses indicate 75% confidence intervals).

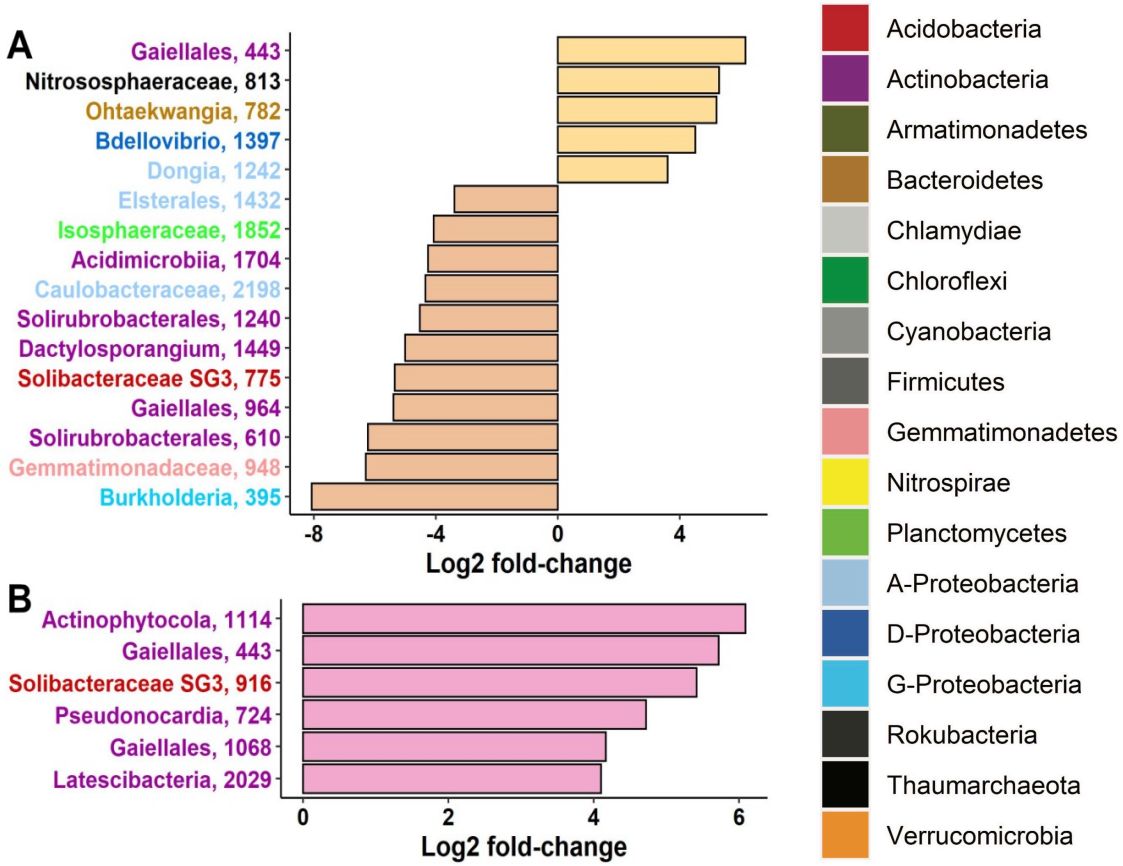

Fig. S3. Influence of (*A*) phosphorus amendment and (*B*) reduced watering on switchgrass rhizosphere microbial community structure assessed by DESeq2 analysis (adjusted $P < 0.01$). ASVs that increased (+ Log2 fold-change) versus decreased in prevalence (- Log2 fold-change) in response to +P or -W treatment are shown. ASVs are presented at the highest available taxonomic resolution, and are colored by class for *Proteobacteria* and by phylum for all other phyla.

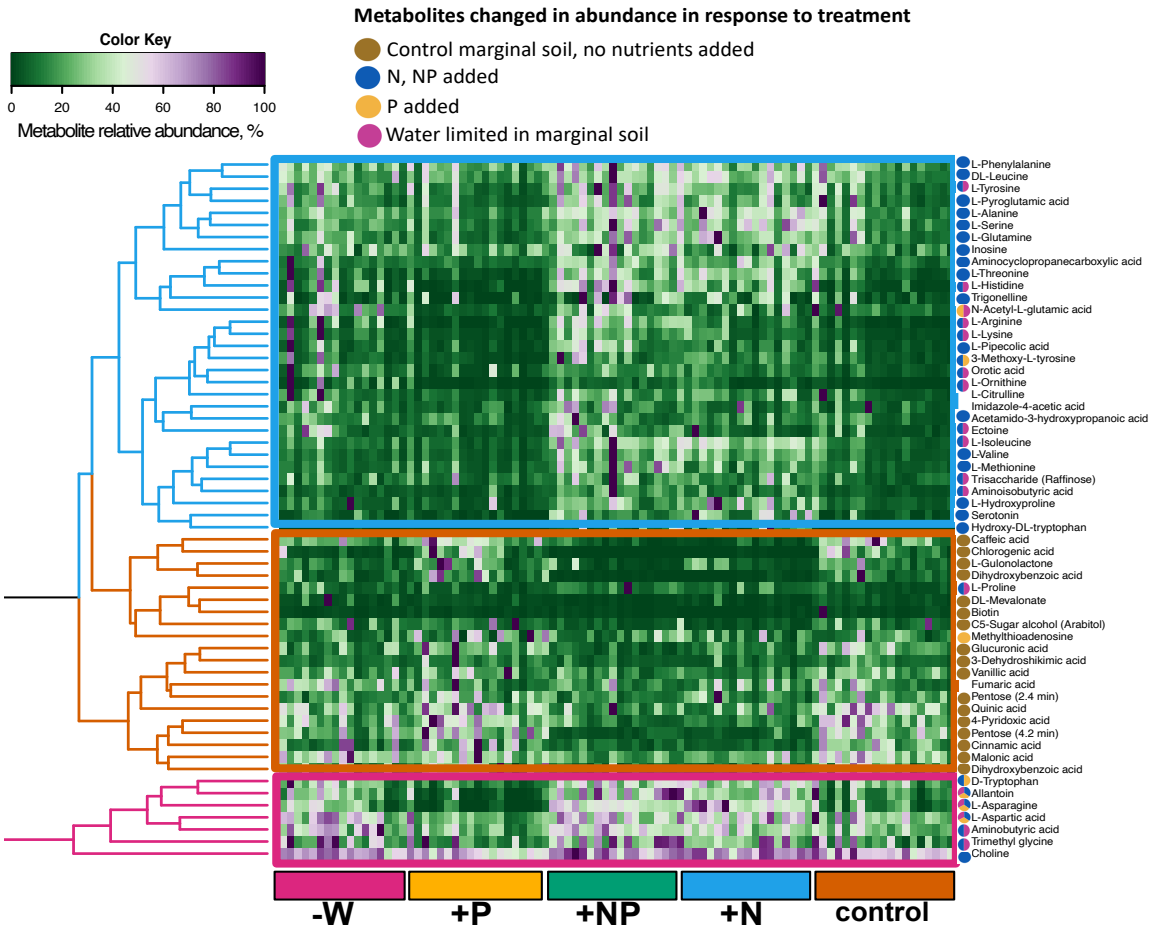

Fig. S4. Heatmap of metabolite relative abundances in the rhizosphere of switchgrass grown with five soil nutrient and water treatments. Five treatments include (i) 'control' with nutrient-poor marginal soil, (ii) '+P', (iii) '+N', and (iv) '+NP', and (v) '-W' treatment, which received 50% reduced water relative to the other treatments. Hierarchical clustering shows three main clusters representing (i) metabolites that were more abundant in +N, +NP treatments (blue lines); (ii) metabolites that were more abundant in no N added treatments (brown lines); (iii) metabolites that were more abundant in +N, +NP, and reduced water treatments (pink lines).

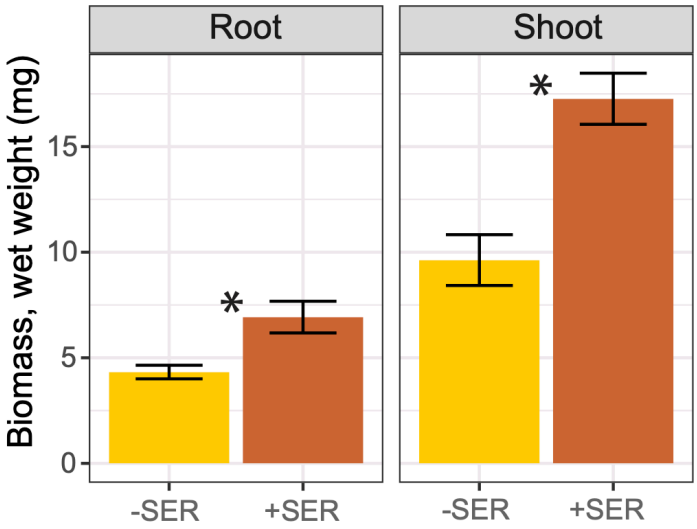

Fig. S5. Serotonin effects on switchgrass root and shoot biomass of 25 day-old switchgrass seedlings (n=9) grown with exogenous application of 0.1 mM of serotonin or controls. Significant differences between added-serotonin (+SER) and controls (-SER) were assessed by ANOVA; asterisks reflect $P < 0.05$.

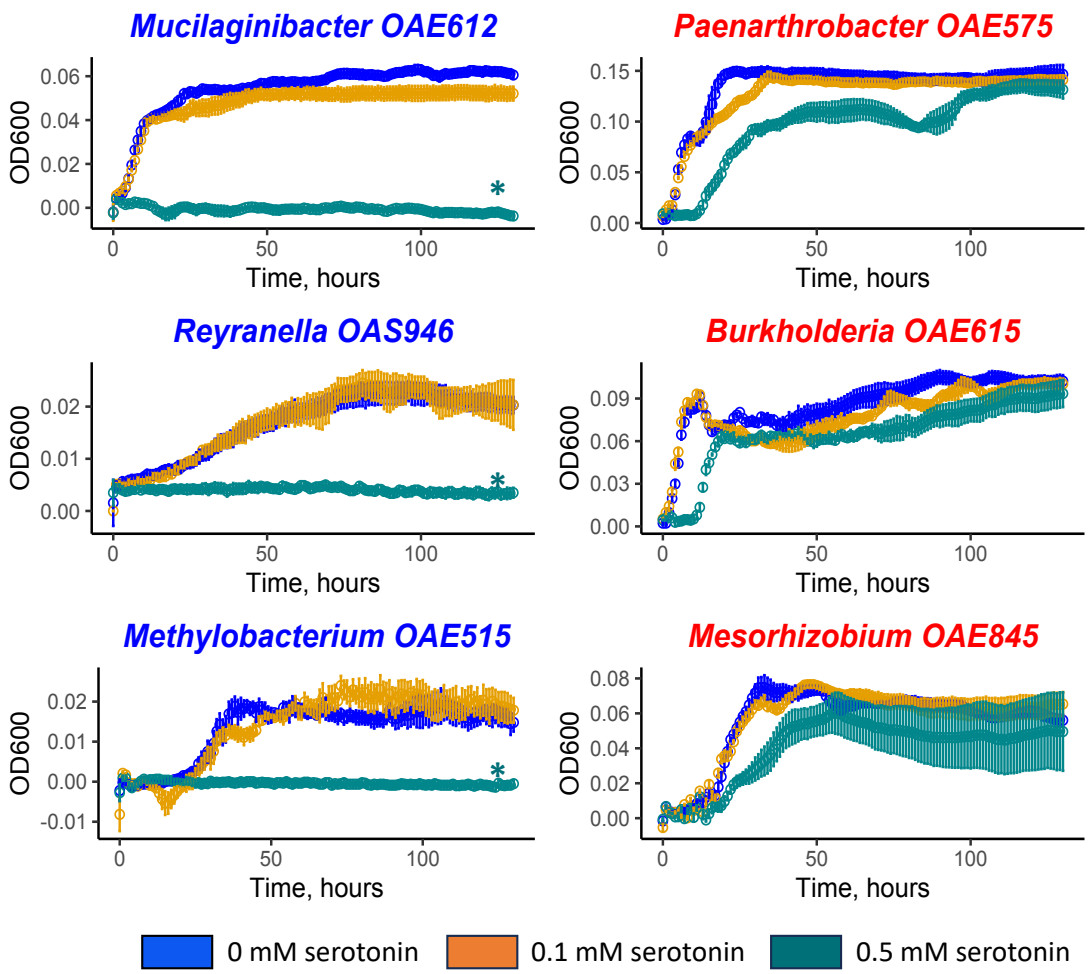

Fig. S6. Growth curves of isolates grown in 1/10 R2A medium with 0 (blue), 0.1 (orange), or 0.5 mM (green) serotonin added. 16S rRNA gene sequences of heterotrophic bacteria isolated from marginal fields cultivated with switchgrass, have been compared to ASVs that showed significant Spearman correlations (both positive and negative) with serotonin relative abundances in this experiment. Isolates with a match (E_{values} <1 × 10^{-10} and ≥97% of gene sequence homology) to the ASVs with a significant negative Spearman correlation with serotonin are highlighted in blue color and isolates matched to the ASVs with a positive correlation represented in red color. The growth curve experiments have been repeated twice to ensure the reproducibility of the effect of serotonin. Asterisk indicates significantly different optical density (OD_{600}) at 130 hours of isolate growth in serotonin treatments and a control treatment without serotonin (0 mM) at $P < 0.05$ by means of Kruskal-Wallis test. Error bars show the standard error of the mean.

Dataset S1 (separate file). Mean relative abundance of rhizosphere ASVs by phylum across all samples.

Dataset S2 (separate file). Pair-wise comparisons for treatment differences in β-diversity for switchgrass rhizosphere 16S ASV communities exposed to N, P, and moisture stress in the greenhouse.

Dataset S3 (separate file). DESeq2 responsive ASVs, their taxonomy, and the treatments they responded to.

Dataset S4 (separate file). Metabolite identifications and response to treatments.

Dataset S5 (separate file). List of network connectors, module hubs, network hub and their topological features; correlation strengths between network nodes.

SI References

1. K. M. DeAngelis, *et al.*, Selective progressive response of soil microbial community to wild oat roots. *ISME J.* **3**, 168–178 (2009).
2. R. I. Griffiths, A. S. Whiteley, A. G. O'Donnell, M. J. Bailey, Rapid method for coextraction of DNA and RNA from natural environments for analysis of ribosomal DNA- and rRNA-based microbial community composition. *Appl. Environ. Microbiol.* **66**, 5488–5491 (2000).
3. A. Untergasser, DNA Miniprep using CTAB. *Untergasser's Lab*. (2008). http://www.untergasser.de/lab/protocols/miniprep_dna_ctab_v1_0.htm
4. V. L. Singer, L. J. Jones, S. T. Yue, R. P. Haugland, Characterization of PicoGreen reagent and development of a fluorescence-based solution assay for double-stranded DNA quantitation. *Anal. Biochem.* **249**, 228–238 (1997).
5. S. J. Blazewicz, *et al.*, Taxon-specific microbial growth and mortality patterns reveal distinct temporal population responses to rewetting in a California grassland soil. *ISME J.* **14**, 1520–1532 (2020).
6. A. E. Parada, D. M. Needham, J. A. Fuhrman, Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ. Microbiol.* **18**, 1403–1414 (2016).
7. J. G. Caporaso, *et al.*, Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad. Sci.* **108**, 4516–4522 (2011).
8. B. J. Callahan, *et al.*, DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).
9. P. J. McMurdie, S. Holmes, phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* **8**, e61217 (2013).
10. J. Oksanen, *et al.*, vegan: Community Ecology Package. R package version 2.5-7. (2020) https://CRAN.R-project.org/package=vegan
11. Y. Yao, *et al.*, Analysis of Metabolomics Datasets with High-Performance Computing and Metabolite Atlases. *Metabolites* **5**, 431–442 (2015).
12. Y. Deng, *et al.*, Molecular ecological network analyses. *BMC Bioinformatics* **13**, 113–20 (2012).
13. S. Shi, *et al.*, The interconnected rhizosphere: High network complexity dominates rhizosphere assemblages. *Ecol. Lett.* **19**, 926–936 (2016).
14. J. Zhou, Y. Deng, F. Luo, Z. He, Y. Yang, Phylogenetic molecular ecological network of soil microbial communities in response to elevated CO2. *mBio* **2**, e00122-11 (2011).
15. S. Weiss, *et al.*, Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. *ISME J.* **10**, 1669–1681 (2016).
16. R. Guimerà, L. A. Nunes Amaral, Functional cartography of complex metabolic networks. *Nature* **433**, 895–900 (2005).

17. J. M. Olesen, J. Bascompte, Y. L. Dupont, P. Jordano, The modularity of pollination networks. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 19891–19896 (2007).
18. G. Csardi, T. Nepusz, The Igraph Software Package for Complex Network Research. *InterJournal* **Complex Systems**, 1695 (2005).
19. P. Shannon, *et al.*, Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
20. G. Lobet, L. Pagès, X. Draye, A novel image-analysis toolbox enabling quantitative analysis of root system architecture. *PLANT Physiol.* **157**, 29–39 (2011).