1 Title: Microbial community function increases host plant leaf growth in a pitcher plant experimental system

2

3 Authors: Jessica R. Bernardin[1], Erica B. Young[2], Sarah M. Gray[3], Leonora S. Bittleston[1]

4

5 Addresses: [1]Department of Biological Sciences, Boise State University, Boise, Idaho, USA, [2]Department of Biological

6 Sciences and School of Freshwater Sciences, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin, US,

7 [3]University of Fribourg, Department of Biology-Ecology and Evolution, Fribourg, Switzerland

8

9 Corresponding author contact information: jessicabernardin@boisestate.edu

Abstract

Across diverse ecosystems, bacteria and their host organisms engage in complex relationships having negative, neutral, or positive effects. However, the specific effects of leaf-associated bacterial community functions on plant growth are poorly understood. This study investigated mechanistic relationships between bacterial community function and host plant growth in a carnivorous plant that relies on microbes to break down insect prey via extracellular hydrolytic enzymes. Sterile, freshly opened leaves (pitchers) of purple pitcher plants (*Sarracenia purpurea*) were inoculated with three functionally distinct bacterial communities to test the effects of bacterial function on plant growth and leaf nutrient content. Bacterial community composition and function were measured using physiological assays, metagenomics, and metatranscriptomics. Distinct bacterial functions affected plant traits, with a bacterial community enriched in decomposition and secondary metabolite production traits leading to leaves with almost double the biomass of control pitchers. Functional differences in bacterial communities, measured by physiological assays, were supported by metatranscriptomic analysis; for example, the bacterial community with the highest chitinase activity had greater relative abundance of transcripts associated with chitinase enzymes. The bacterial community associated with larger pitchers also had increased differential expression of transcripts linked to microbially-produced plant hormones. The direct relationship between bacterial community function and plant growth observed, indicates specific mechanisms for host-associated bacterial functions to support plant health and growth.

Keywords

microbial community function; metatranscriptomics; microbial; phyllosphere; microbe-plant interactions; host health; metagenomics; pitcher plant; *Sarracenia purpurea*

Introduction

Microbiomes have strong effects on the health and fitness of their hosts (1,2). Host plant growth and physiology are affected by the presence of leaf-associated (phyllosphere) microbes, with microbial taxa acting as mutualists, commensals, and antagonistic pathogens (3–6). Phyllosphere microbes live in complex microecosystems (7,8) performing functions that drive plant nutrient acquisition, phytohormone production, and stress tolerance (3). For

example, the addition of synthetic phyllosphere communities has been shown to increase fruit production in greenhouse-grown tomatoes (6) and increase leaf sugar content and metabolic diversity in field-grown *Agave tequilana* (9). Plants reap many benefits by hosting microbial communities (3,9). Nutrient cycling interactions between microbes and plants can occur directly, e.g., via bacterial nitrogen fixation, and indirectly, e.g., via increased bacterial growth that influences host plant metabolism (3,10). The majority of bacteria that engage in plant-microbe interactions also have pathways to produce phytohormones, like auxins or cytokinins, which support plant cell metabolism and growth (3,11–15). Plant growth-promoting bacteria can also mediate host stress responses, including increasing drought tolerance, thermotolerance, and UV protection (3,16,17). Despite these demonstrated benefits to host plants, we still lack research directly connecting bacterial community function with host physiological functions and benefits. We define bacterial community function as the collective activities and metabolic processes of all the individual bacterial species within a community. This study addresses this gap by examining both bacterial and plant physiology, using metabarcoding, metagenomics, and metatranscriptomics of bacterial communities and measuring plant responses over time.

The carnivorous pitcher plant, *Sarracenia purpurea*, uses water-filled, cup-shaped leaves to trap and digest insects to supplement nutrient acquisition that it is unable to obtain from low nutrient soils (18,19). These leaves or 'pitchers' host an aquatic micro-ecosystem composed of microbes, fungi, protists and invertebrates (8,20,21). The microbes, in particular, are thought to provide critical ecosystem functions to the host plants by mediating insect prey digestion through the production of extracellular hydrolytic enzymes. It is likely that the bacterial production of protease and chitinase enzymes that digest protein and chitin-rich insect prey are especially important for this carnivorous plant species (22–24). *Sarracenia purpurea* pitcher plants do not produce chitinase enzymes (22), thus they may rely heavily on the bacterial extracellular enzymes to catalyze decomposition of insect prey and support the mineralization of carbon, nitrogen, and other nutrients for the host plant and the whole micro-ecosystem. This is an ideal system to examine bacterial functions in controlled experiments, serving as an analog for other systems like the gut microbiome, where experimentation is more challenging. Although previous research has hypothesized that pitcher plant leaf-associated bacterial communities benefit the host plant (18,25,26), no studies have yet demonstrated or measured a direct benefit.

To connect the effects of bacterial function to plant growth, we designed a manipulative experiment using pitcher plant leaves and bacterial communities with different physiological functional profiles, measured as different levels of activity across critical functions. These community physiological functions, including hydrolytic enzyme activity, carbon substrate use, and respiration rates, were characterized prior to inoculation, and the effects of these functionally distinct bacterial communities on host plants was examined over eight weeks. We use the term high-functioning community to highlight communities that are highly efficient at critical functions in our system, such as high rates of chitinase activity. We asked four specific questions: (1) Do bacterial communities with different functional profiles have differential effects on plant traits, such as leaf growth and nutrients? (2) Are there specific functions, e.g., hydrolytic enzyme activity or carbon substrate use, that drive these effects on host plants? (3) Are there taxonomic characteristics of these bacterial communities that can explain the observed functional differences? (4) Can bacterial functions directly measured *in planta* be linked with differential gene expression (metatranscriptomics) to better understand the mechanisms of bacterial impacts on their host? For question 4, we hypothesized that communities with higher physiological functional measurements would show higher transcription of genes associated with degradative functions, allowing bacterial functions measured *in planta* to be linked with differential gene transcription and mechanisms underlying bacterial-host interactions.

Methods (See supplement for additional methods)

Experimental design

Prior to experiment initiation, mature greenhouse-grown *S. purpurea* plants were surface-sterilized, repotted in twice-autoclaved soil, and transferred to plant growth chambers (Hoffman Mfg. Inc. Corvallis, OR USA, SG2-22). The three distinct bacterial communities (CommA, CommB, and CommC) used for inoculation were previously collected from wild *S. purpurea* pitcher plants (27), reconstituted from cryogenic culture, and stabilized in culture (acidified cricket media, ACM) prior to the experiment. There were five treatment groups: CommA, CommB, CommC, ACM only (control), and sterile deionized (DI) water only (control) (Fig. 1). Ten individual pitcher plants were randomly assigned to each bacterial treatment group and six for each control group, for a total of 42 individual plants. On each plant, one new, unopened pitcher was manually opened using sterile gloved hands and

carefully inoculated with a bacterial community culture in ACM (experimental groups), or sterile ACM media or sterile water (control groups, no bacteria). The single pitcher on each plant that received one of the five treatments was the target pitcher (pitcher 1). All other pitchers on each plant received sterile water for the duration of the experiment. For some plant trait measurements, data was also collected on the second newest pitcher (pitcher 2). Pitcher volume varied with the size of the new pitcher, but bacterial culture concentrations were maintained across all experimental group pitchers (i.e., each pitcher got a 1:10 ratio of bacterial culture to new media at day 0 of experiment regardless of pitcher volume). Samples were collected weekly: starting one day after inoculation (day 1), 4 mL was collected (full details in Supplement) from each pitcher using sterile technique and aliquoted for amplicon sequencing, metagenomic/metatranscriptomic analysis, bacterial physiological assays, and culture preservation. All preserved samples were stored at -80°C until nucleic acid extraction.

Bacterial Physiological Assays

Chitinase and protease activities were quantified weekly on a sample from each treatment pitcher using fluorometric assays in black 96-well microplates (24,28). Fluorescence emission was measured every five minutes for one hour using a microplate reader (Biotek Synergy Mx Multi-Mode Microplate Reader). Of three types of chitinases found in the bacterial communities of *S. purpurea* pitchers, chitobiosidase activity was selected for quantification because preliminary data showed higher activity potential than β-N-acetylglucosaminidase and triacetyl chitotriosidase. Cumulative enzyme activity potential was calculated by summing the weekly enzyme activity for each pitcher over eight weeks. The “functional fingerprint” for each bacterial community was measured by examining the community’s capacity to use 31 carbon substrates. These measurements were conducted on the first and last day of the experiment (day 1 and 55) using Biolog EcoPlates that were read with a microplate reader after 72 hours of incubation at room temperature (29). See the supplemental methods for other bacterial functional measurements, including bacterial growth rates and respiration.

Plant functional traits

Pitcher morphological traits, including length, width, and aperture (30) were measured weekly for eight weeks. For each plant, the number of mature leaves was recorded at the beginning and end of the experiment. After eight

weeks, wet and dry biomass were measured for pitchers 1 and 2 from each plant. Plant physiological capacity was measured as photosynthetic CO_2 assimilation rate on day 55 using a LICOR-640 and using chlorophyll *a* fluorescence. Nutrient acquisition was assessed in plant tissue after drying at 65°C for 72 hours, stored dry before homogenization for CN elemental analysis (Thermo-Electron Flash EA1112). See supplementary methods for details on plant measurements.

16S rRNA sequencing and analysis

To characterize bacterial community composition, DNA was extracted for all samples (day 8, 15, 22, 29, 36, 43, 50, 55, n=310) using the DNAdvance Genomic DNA Isolation Kit (Beckman Coulter A48705). DNA was quantified using AccuClear Ultra HS dsDNA kit (Biotium #31028). 16S rRNA gene amplification and library preparation and sequencing was conducted by the Environmental Sample Preparation and Sequencing Facility at Argonne National Laboratory on a 151 bp x 12 bp x 151 bp Illumina MiSeq run targeting the V4 region of 16S rRNA using the 515F and 806R primers (31).

We used QIIME2 (version 2022.8) on the Boise State University computing cluster (Borah) to demultiplex amplicon sequences using no golay error correction, and the DADA2 plugin to denoise sequences and generate amplicon sequence variants (ASVs) of 150 bp length (32,33). Taxonomy was assigned using the classify-sklearn method which is a Native Bayes classifier, and a pre-trained classifier made with Silva v. 138 database containing 99% amplicon sequence variants (ASVs) from 515F/806R region (34). The phylogenetic tree was built using multiple-sequence alignment via MAFFT and phylogenetic reconstruction via FastTree, both via a QIIME2 plugin (35,36). Taxonomy and ASV tables were migrated to R for further analysis (see supplement for full methods) and visualization using qiime2R, picante, and phyloseq (37–39).

Metagenomic and metatranscriptomic sequencing and analysis

The pitcher fluid from three of the 10 pitchers in each treatment group were selected from three time points (3 plants x 3 treatments x 3 timepoints, n=27) for simultaneous DNA/RNA extractions for metagenomic and metatranscriptomic sequencing using the Qiagen AllPrep Bacteria DNA/RNA/Protein Kit (Cat 47054), according to

manufacturer directions. Nucleic acids were quantified using a fluorometer (Qubit, Invitrogen). Metagenomic libraries were prepared for a subset of these DNA samples (3 plants x 3 treatments x 2 timepoints, n=18) using NEBNext Ultra II FS DNA library prep kit for Illumina (NEB#78055). Libraries were prepared for the full 27 RNA samples using NEBNext rRNA Depletion Kit (NEB#E7850) and NEBNext Ultra II Directional RNA library prep kit for Illumina (NEB#E7760) using manufacturer protocols. Sequencing was done on a NovaSeq partial lane to a sequencing depth of 55 Gb per library at Novogene Co. (Sacramento, CA USA).

Shotgun metagenomic and metatranscriptomic sequences were processed on BSU's Borah cluster. Adapters and low quality sequences were trimmed using Trimmomatic (0.39) and viewed with MultiQC (v1.6) quality reports (40,41). Metatransciptomic sequences were further processed to remove ribosomal RNA using the default settings in Ribodetector (42). Two co-assemblies were produced (metagenomic and metatransciptomic) using MEGAHIT (v1.2.9) (43,44). For the metagenomic analysis, the "sketch", "gather", and "tax annotate" functions within sourmash (version 4.8.2) (45) using kmer=31 were used to assign and quantify taxonomy using the full GTDB database (R08-RS214 403k) (46). Contigs in the coassembly and paired-end reads were binned using MAXBIN2 (2.2.7), CONCOCT (1.1.0), and Metabat2 (2.12.1), consensus bins were produced using DASTool (1.1.6) (47–50). Completeness and contamination of each metagenome assembled genome (MAG) were calculated using the presence of single-copy marker genes with CHECKM2 v.1.0.11 (51). Comparisons and dereplication of MAGs employed dREP v.2.2.1 (52). Taxonomy was assigned to the quality-controlled MAGs using the full GTDB database within sourmash. Bins were quality ranked according to their completeness and contamination (53,54). For metatranscriptomic analyses, read mapping was conducted using Bowtie2 (2.5.1) (55). Samtools (1.6) (56) was used to calculate mapping statistics and further process the contig files for downstream analysis. Open reading frames (ORFs) were predicted using Prodigal (v2.6.3) (57). Transcript abundance quantification was done in Kallisto (v1.9.0) and resulted in transcripts per million (TPM) (58). KOFAMSCAN (1.3.0) was used to bin ORFs into KOs using the full KEGG database (59).

Statistical analyses

All statistical analyses were conducted in R version 4.2.2 (60). Alpha diversity metrics (Hill Number 1) were calculated using the *vegan* package (61) and additional processing used *tidyverse* and *janitor* (62–64). For ASV-level differential abundance explained by the treatment groups, we did an analysis of compositions of microbiomes with bias correction (*ANCOMBC*, (65)). Significantly differentially abundant ASVs were visualized with boxplots and a heatmap (66) by transforming the normalized reads: (log2(read count+0.5)).

Community similarity in ASVs and EcoPlate carbon substrate use among samples in each treatment group was determined using the vegan package in (61). Compositional differences (beta diversity) and changes across the three bacterial community treatments from day 8 to 55 were assessed using phylogenetic methods (unweighted UniFrac, 67)). Mantel tests (61) assessed correlations between bacterial composition and hydrolytic enzyme activity or pitcher biomass. To visualize the similarity in EcoPlate carbon substrate use among samples and treatments, a *metaMDS* analysis with Bray-Curtis distance at two dimensions (k=2) was utilized. In both cases, non-metric multidimensional scaling (NMDS, 61)) visualized the dissimilarities calculated between each pitcher fluid sample. A non-parametric multivariate analysis of variance test (*PERMANOVA*, *adonis2* function, by="margin") was used to assess any differences between treatments, and *pairwise adonis* was used to assess the effects of treatment and time, while accounting for repeated measures for each pitcher (61). Dispersion, or the homogeneity in dispersion between treatments, was calculated using the *betadisper* function (permutations compared using a post-hoc Tukey test, 61).

Metatranscriptomic statistical analyses began with a matrix of filtered genes (length>500bp), samples were normalized using the *calcNormFactors* function in the *variancePartition* package (1.28.9). Differential transcript abundance was quantified using *Dream* within the *variancePartition* package (68). In order to link specific taxa to differentially expressed functions, significantly differentially abundant contigs were mapped to the MAGs using a custom database built from the MAG contigs and *BLAST* (2.14.0) (69).

To isolate the impact of functionally distinct bacterial communities on plant traits (after 8 weeks *in planta*, Questions 1 and 2), we developed Bayesian generalized linear mixed models (GLMMs) using variables based on the

causal assumptions of directed acyclic graphs (DAGs). Directed acyclic graphs help identify which observed and unobserved variables to condition upon to remove non-causal associations between variables and isolate mechanisms of interest, under the assumption that the depicted pathways between variables in a research design are accurate . Therefore, GLMMs contained a variable set designed to isolate the effects of treatment and enzymatic activity from the other unspecified and unrecorded impacts of our bacterial treatments (Fig. S1) (70,71). Pitcher biomass was selected to represent the morphometrics and growth, along with leaf nitrogen content representing leaf nutrient acquisition. To consider non-independence of repeated measures of pitchers over time, models included a random intercept for pitcher identity.

Bayesian hierarchical models using a gamma (log-link) GLM were fit using the *brms* package (72) and visualized using *bayesplot*, *ggeffects*, and *tidybayes* (73–75). The continuous predictor variables were standardized and mean centered using the *standardize* package (76,77). Default uninformative priors were used, convergence and mixing of chains and unimodality in posterior predictions were visually assessed, and all R-hat values were close to 1.0 (78). The model fit was evaluated using the posterior predictive check function in *brms* (72).

Results

Starting communities were functionally distinct

Initial communities were functionally different in their chitinase activity, protease activity, bacterial respiration, and EcoPlate substrate use profiles (Fig. 1, Fig. S2). CommB had the highest chitinase activity and lowest protease activity prior to inoculation (Fig. 1A, B, statistical differences shown in Fig. S2). Chitinase and protease activity increased for all treatments except the water control throughout the experiment (Fig. S4). Bacterial respiration was initially lower in CommB, but there were no respiration rate differences among the communities by day 55 (Fig. S2). There were strong differences in substrate use profiles among three bacterial community groups, sampled 24 hours after pitcher inoculation (Table S1, Fig. 1C), which remained distinct and increased in dispersion within over time (Fig. S3, Table S1). Bacterial community composition was taxonomically different on day 8 (Table S2; Fig. 1D). There were no measured differences in bacterial community growth rates between our treatments at any time point.

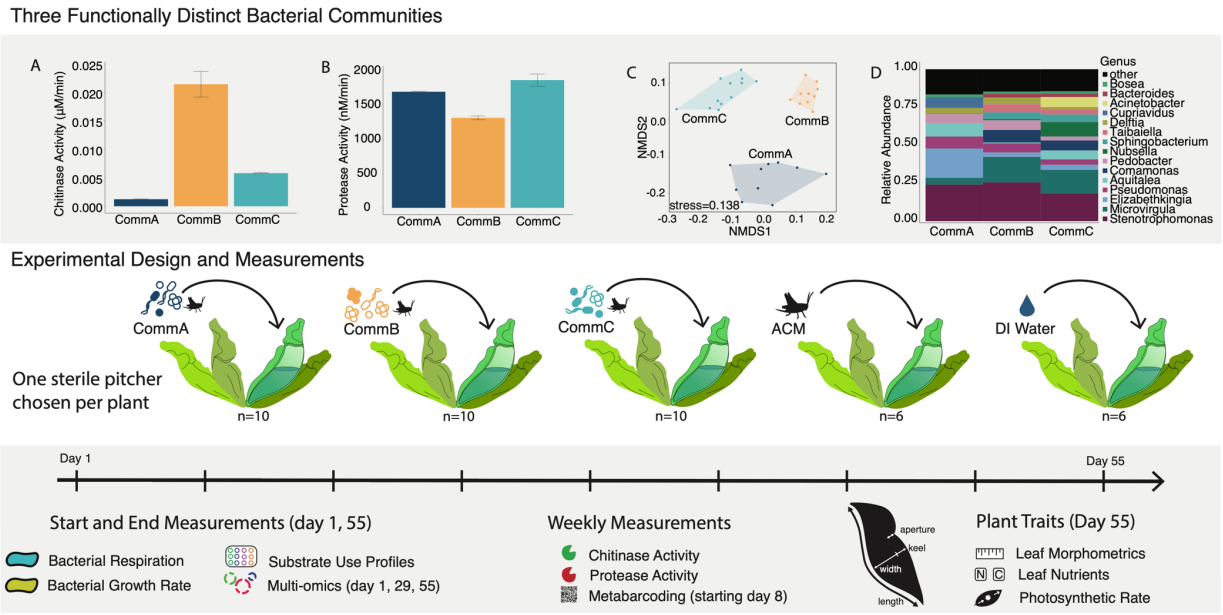

Figure 1. **Experimental design and initial functions of starting bacterial communities.** Three functionally distinct bacterial communities growing in acidified cricket media (ACM) were inoculated into sterile pitchers and incubated in growth chambers for eight weeks. ACM or sterile deionized (DI) water were added to additional pitchers as experimental controls. Bacterial communities had different physiological functions measured before inoculation: **A** CommB had higher chitinase activity and **B** lower protease activity than CommA and CommC. **C** As visualized with the NMDS (k=2, stress=0.138) of EcoPlate functional fingerprints using Bray-Curtis dissimilarity, the three starting communities had different substrate use profiles (n=31, PERMANOVA p=0.001) (statistical differences shown in Fig. S2). **D** These three communities also had distinct community compositions. Bacterial community function was characterized prior to inoculation (**A,B**), then weekly after inoculation, starting on day 1 *in planta* and continuing until day 55. Bacterial data collected included: chitinase and protease activity, community respiration, growth rate, substrate use profiles, and samples for metagenomics and metatranscriptomics. Plant functional traits assessed include morphometric measurements of pitchers, as well as end of experiment measures of leaf biomass, leaf nutrient content, and photosynthetic rate.

Bacterial function affects plant traits

To address Question 1, we assessed the effect of community function on the response of specific plant traits (pitcher biomass and nitrogen content) after hosting a bacterial community for 8 weeks. Pitchers hosting bacterial communities with different functional profiles differed in their final biomass. Specifically, hosting a bacterial community with higher chitinase activity at the start of the experiment resulted in leaves with higher biomass (Fig. 2A). Pitchers hosting the high chitinase CommB, had 1.7 times the biomass (mean biomass = 0.468 g) of pitchers with ACM only (mean biomass = 0.275 g) and double the biomass of water-only control plants (mean biomass = 0.235 g). The estimated probability of a positive effect on pitcher biomass for each group compared to the cricket media (ACM) control was 81.04% for CommA, 99.73% for CommB, and 84.32% for CommC (Fig. 2B). We saw

similar effects of treatment on all leaf morphological measures (which were correlated, Fig. S5); for example, pitchers that hosted the high chitinase bacterial community also had longer leaves and higher pitcher carbon content than other treatments (Fig. S7). Additionally, there was a positive effect of the CommC treatment on total plant growth (measured as the net production of leaves effect; 96.55% greater than 0, although the 95% CIs cross zero), plants in the CommC group had an estimated 1.4 greater net pitcher mass production than the ACM control over eight weeks (Sig. S7).

The estimated effect of the bacterial function on the total pitcher nitrogen content (Fig. 2C and 2D) showed the probability of a pitcher having increased nitrogen content was greatest when hosting a bacterial community compared to the water-only treatment; our 95% CIs cross zero (ACM baseline) but we did see a weak positive directional effect and estimated a 74% probability that bacterial treatments had a positive effect on pitcher nitrogen. There was no difference between the bacterial communities in terms of pitcher nitrogen content (mg/pitcher, estimated N content for the three communities; CommA=5.04 mg, CommB=4.9 mg, CommC=4.7 mg). There was a strong negative effect of N content in pitchers containing water only (estimated N content = 2.11 mg, 95% CIs= -1.057 -0.433, Fig. 2C and 2D). No other differences between the treatments were observed, including no differences in photosynthetic rates or photosynthetic quantum yield (F_v/F_m) between pitchers in the five treatments (Fig. S7).

To investigate specific functions driving differences in plant traits (Question 2), there were no independent effects of chitinase (95% CIs=-0.212, 0.074) or protease activity (95% CIs=-0.288, 0.146) on pitcher nitrogen content (Fig. S6C).

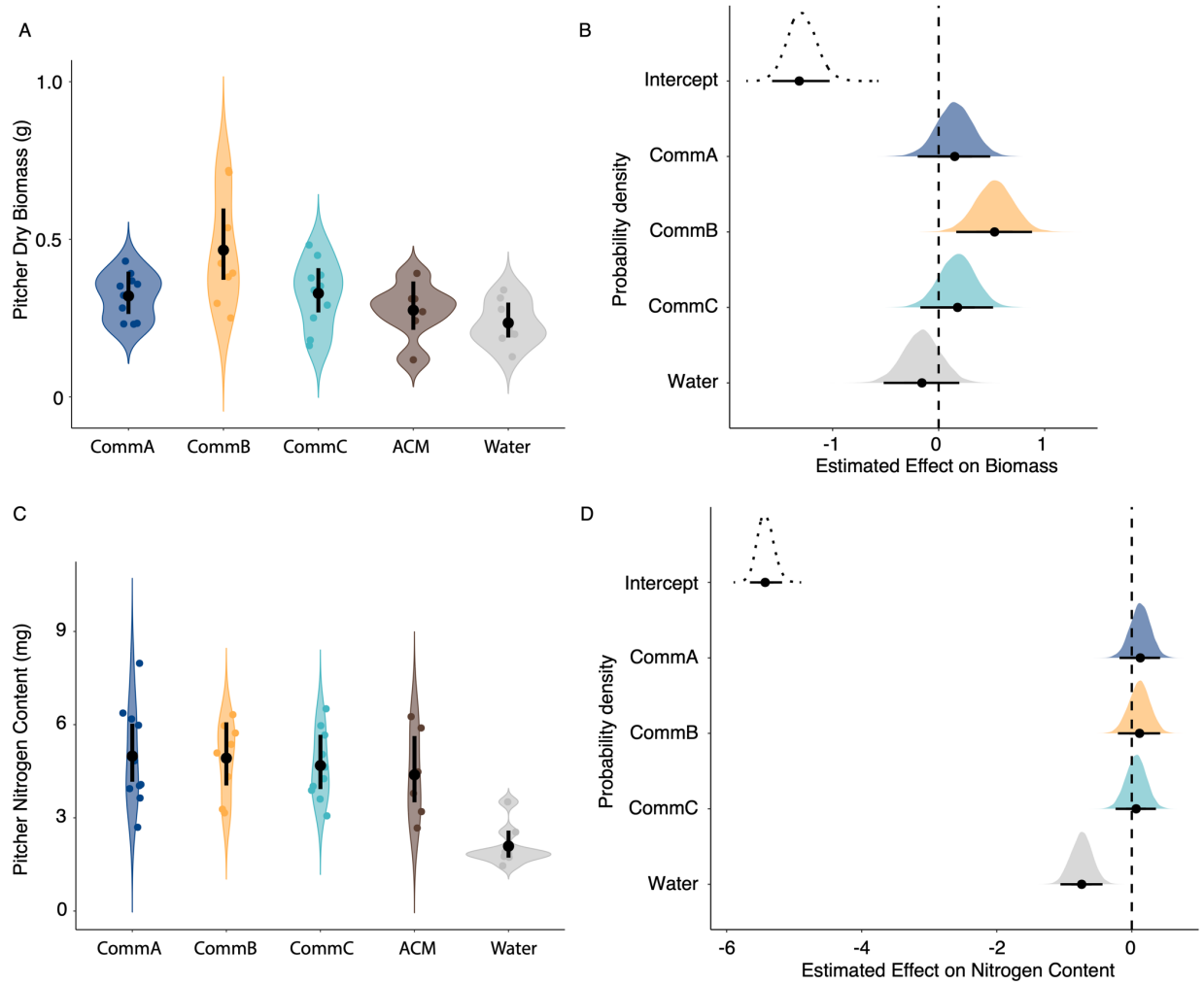

Figure 2. **Bacterial community treatment affects plant traits. A** Marginal effect of treatment on target pitcher biomass vary in relation to community function. The density distributions of pitcher biomass data used to calculate estimated effects are represented by the violin plot with colored points behind the marginal effect estimates. The predicted effects of treatment (while holding all other covariates at their means) are represented by the black line range. This point is the median and the line represents the 95% credible intervals (CIs). **B** Predicted effects of functionally distinct community treatments on pitcher biomass. Black points represent median estimates, lines represent the 95% CIs, and color-coded (by treatment) density plots indicate the full posterior. The intercept (dashed posterior) is set as the ACM baseline. **C** The density distributions of pitcher nitrogen content are shown in violin plots and points. The marginal effect estimates (while holding all other covariates at their means) are represented by the black line range, the black point is the median and the line represents the 95% credible intervals (CIs). **D** Predicted effects of functionally distinct community treatments on pitcher nitrogen (N). As with (B), black points represent median estimates, lines represent the 95% CIs, and color-coded (by treatment) density plots indicate the full posterior probability densities, and the Intercept (dashed posterior) is set as the ACM baseline.

Compositional differences between the three communities

16S Metabarcoding results

After quality control and rarefaction, 917 ASV were recovered and in a community matrix across 310 samples (6-10 pitchers x 5 treatment groups x 8 weekly timepoints) these represented 26 bacterial phyla, 192 families, and 286 genera. Of the top 20 genera averaged across pitchers within a treatment at a single time point, three were highly abundant, comprising almost 25% of total reads: *Microvirgula* (14.2%), *Stenotrophomonas* (12.4%), and *Pedobacter* (11.3%) (Fig. 3A).

Community Diversity and Differential Abundance

Addressing Question 3, we observed differences in alpha diversity between three treatments (CommA, CommB, and CommC) after one week *in planta* (Fig. 3B, Fig. S9). Between day 8 and 55, the effective number of ASVs increased (mean gain of 7.5 ASVs per sample) but community treatment differences in ASV numbers disappeared by day 55 (Fig. 3B, Fig. S9). Some bacterial ASVs were present in control treatments, with a higher alpha diversity in the ACM treatment than water controls, but all control treatment samples had lower alpha diversity than those with inoculated communities (ENS_{ACMwk8}=17.2, $ENS_{WATERwk8}$=12.2).

We found a significant difference in beta diversity (the turnover in community composition) between the three community treatments (PERMANOVA; R^2=0.262, $F_{2, 214}$=37.92, p=0.001) using unweighted UniFrac distances, shown as clear clustering of treatments in NMDS visualization (Fig. 3C) and significant pairwise differences over time (Table S2). Mantel tests (Spearman's rank correlation rho) showed significant correlations between the unweighted UniFrac distances and chitinase activity (r=0.227, p=0.001) and final pitcher length (r=0.117, p=0.001), but no significant relationship with protease activity (r=0.032, p=2.17). Using ANCOM-BC, 19 differentially abundant ASVs were identified across CommA, CommB, and CommC at day 8 and 37 ASVs at day 55. Among these ASVs, there were stronger patterns in presence and abundance between the three community treatment groups than between day 8 and day 55 (Fig. 3D).

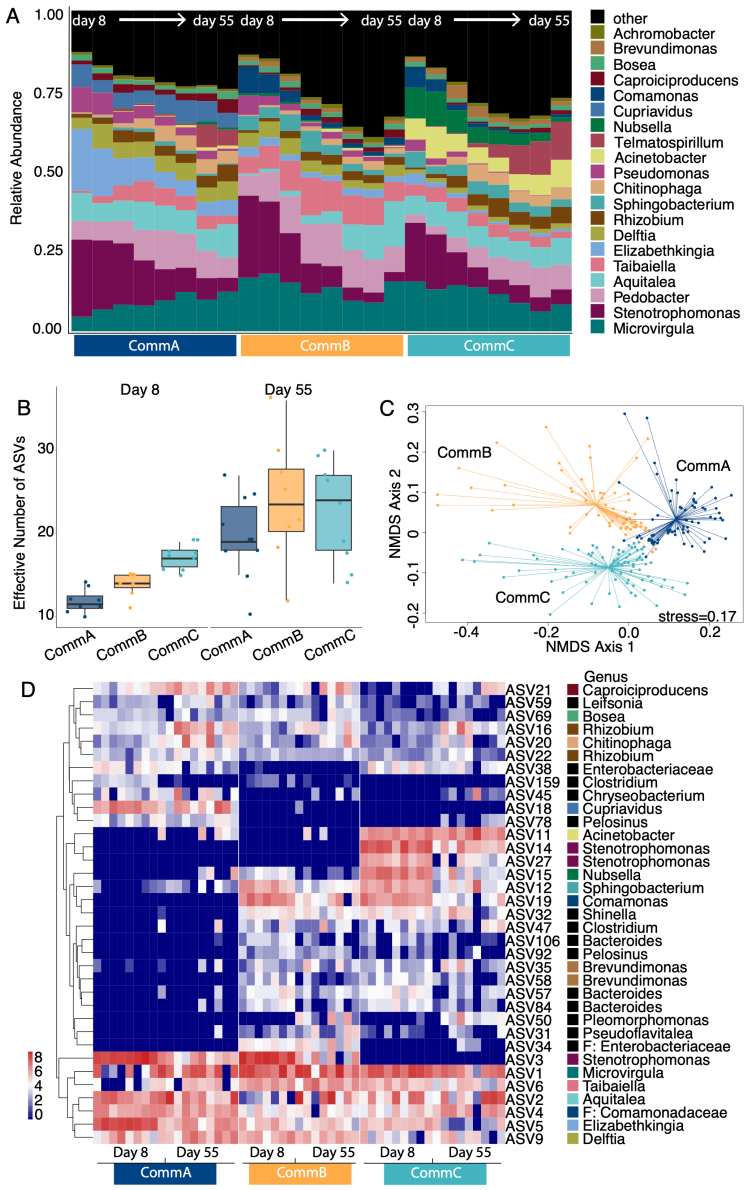

Figure 3. **Metabarcoding (16S) compositional differences between the three bacterial community treatments. A** Relative abundance of the top 20 most abundant genera from day 8 to 55, with biological replicates averaged for each timepoint. **B** Effective number of ASVs (Hill number 1) within each treatment at day 8 and day 55; each point represents a pitcher sample. **C** Composition-based NMDS calculated using unweighted Unifrac for all samples within each of the three treatments from day 8 to 55, stress=0.17, k=2. **D** Heatmap showing the abundance (log2(reads+0.5)) of significantly differentially abundant ASVs calculated using ANCOMBC at day 8 and 55, ASVs color coded according to genus in **A**.

Identifying bacterial physiological functions using metatranscriptomics

After processing, there were 83,977 transcripts from 27 metatranscriptomic samples (3 pitchers for each of 3 bacterial treatments and 3 time points). Of these, 5,732 transcripts were significantly differentially abundant across three community treatments, and 2,394 of these were associated with the KEGG “L1” level of bacterial

metabolism. We used this data to address question 4, can bacterial functions measured *in planta* be linked with differential gene expression in bacterial communities. Metatranscriptomic analysis revealed CommB had significant differential expression of transcripts (including increased abundance, log2(TPM+0.5)) mapping to chitinase enzyme production and transport (endochitinase, chitinase, endo-chitodextinase, and chitobiose transport systems, Fig. 3C). Because the metatranscriptomic analysis only involved a subset of the samples, in order to have a direct comparison we filtered the dataset containing the cumulative chitinase and protease results to only those plants that also had paired metatranscriptomic data (n=9, three plants in each of the three bacterial treatments). Based on our underpowered GLM, there was a weak difference in cumulative chitinase rate measured *in planta* (95% CIs cross zero), with CommB having the highest mean rate (95% CIs=-0.051, 0.130, Fig. 4A). Additionally, for the same subset of samples, *in planta* cumulative protease activity was different between the treatments (Fig. 4B), with the CommA samples showing higher cumulative protease activity than CommB and CommC (Fig. 4B). However, transcript abundances, which were mapped to eight different aminopeptidase KOs, were higher in CommB, than the other two treatments (Fig. 4C). Thus, there was a mismatch between protease activity measured *in planta* and expression profiles of aminopeptidase genes, and instead, the metatranscriptomic data supported that CommB was the high functioning community with greater relative abundance of transcripts coding for both types of hydrolytic enzymes.

When metatranscriptomics was related to carbon substrate-use profiles (Biolog EcoPlates) using pairwise comparisons between CommB and either CommC or CommA (Fig. 4D, E) , CommB showed higher metabolism of the substrates N-acetyl-D-glucosamine (GlcNAc), Beta Methyl D glucoside, and D-Cellobiose. Consistent with this, differentially abundant transcripts for KOs associated with the metabolism of these substrates were higher in CommB on day one for beta glucosidase (which degrades cellobiose) and for five of the six enzymes associated with metabolism of GlcNAc (a chitin monomer, Fig. 4F). No differentially abundant KOs were identified for Beta-methyl D-glucoside.

Metatranscriptomic results related to bacterial production of growth-associated plant hormones, also showed CommB had the highest abundance of transcripts related to indolepyruvate decarboxylase (IPDC), tryptophan

synthase beta chain (TrpB), tryptophanase, and tryptophanyl tRNA synthetase (TrpRS) production (Fig. S8), all of which are related to IAA (indole-3-acetic acid) biosynthesis. CommC had a higher abundance of transcripts related to 4-hydroxy-phenylpyruvate dioxygenase (HDDP) (Fig. S8) which is linked to plastoquinones in photosynthetic electron transport (79). Both CommB and CommC had higher abundances of transcript for ACC-deaminase (Fig. S8), involved in ethylene regulation, important for plant growth, stress tolerance and nutrient uptake (80).

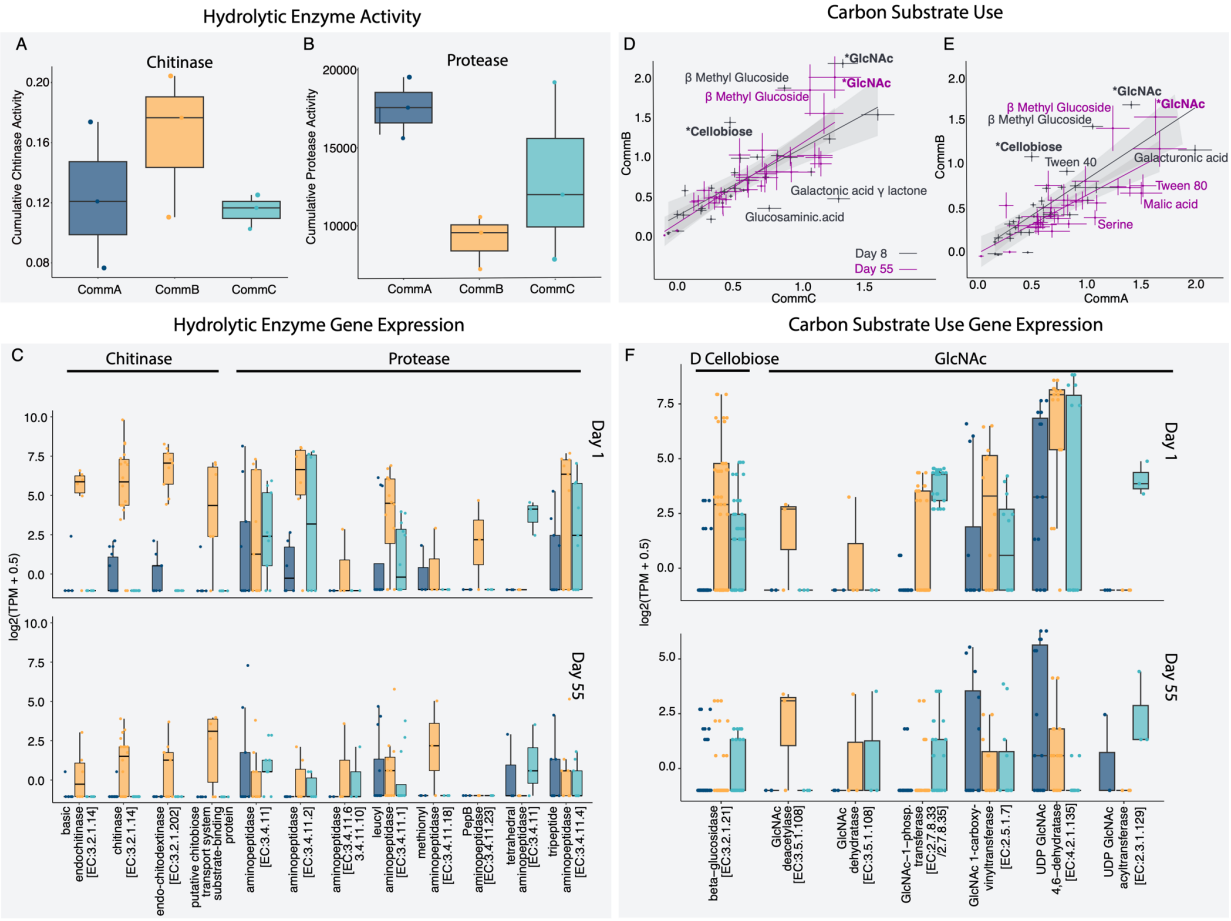

Figure 4. **Functional assays generally matched well with metatranscriptomic results. A** Cumulative chitinase activity (summed over the 8 weekly measurements) showing data from the nine pitcher fluid samples that underwent downstream metatranscriptomic analysis. **B** Cumulative protease rates for those same nine pitcher fluid samples. **C** Transcript abundance per million (TPM) filtered to those transcripts that were differentially abundant between our three bacterial community treatments (colored according to Fig 3A x-axis), then filtered and clustered based on functions related to chitinase and aminopeptidase activity, with day one on top and day 55 below. **D** Differential substrate use (EcoPlate) physiological profiles between CommB and CommC, color coded by day (gray= day 8, pink= day 55). Each point is the mean absorbance for that substrate within samples and bars indicate standard error, gray shaded areas are 95% confidence intervals, substrates falling outside those CI are significantly different between communities. **E** Differential substrate (EcoPlate) profiles between CommB and CommA, color coded by day (gray= day 8, pink= day 55). **F** Transcript abundance filtered as described above and clustered based on enzymes associated with the substrates cellobiose and GlcNAc, with day 1 on top and day 55 below.

Linking targeted and agnostic functional differences

For the 83,977 predicted transcripts, the model identified the differentially expressed transcripts for each treatment compared to the mean expression of the other two treatments combined. Compared to CommA and CommC, CommB had higher expression in pantetheine-phosphate adenylyltransferase (PanC), large subunit ribosomal proteins, and OmpA-OmpF porins from gram negative membranes (Fig. 5A). PanC is a bacterial enzyme involved in biosynthesis of coenzyme A, an essential cofactor for lipid synthesis/oxidation and amino acid metabolism, so it may contribute to the high functioning of CommB. Transcript abundances were mapped to 359 functional categories (KEGG L4) identified in each sample, and our three bacterial treatments differed in overall function based on the significantly differentially expressed transcripts identified by our model (Fig. 5B). Of the transcripts differentially expressed in CommB, a number relate to enzymes involved in chitin and GlcNAc degradation (Fig. 5B inset), corroborating our targeted analysis (Fig. 3F).

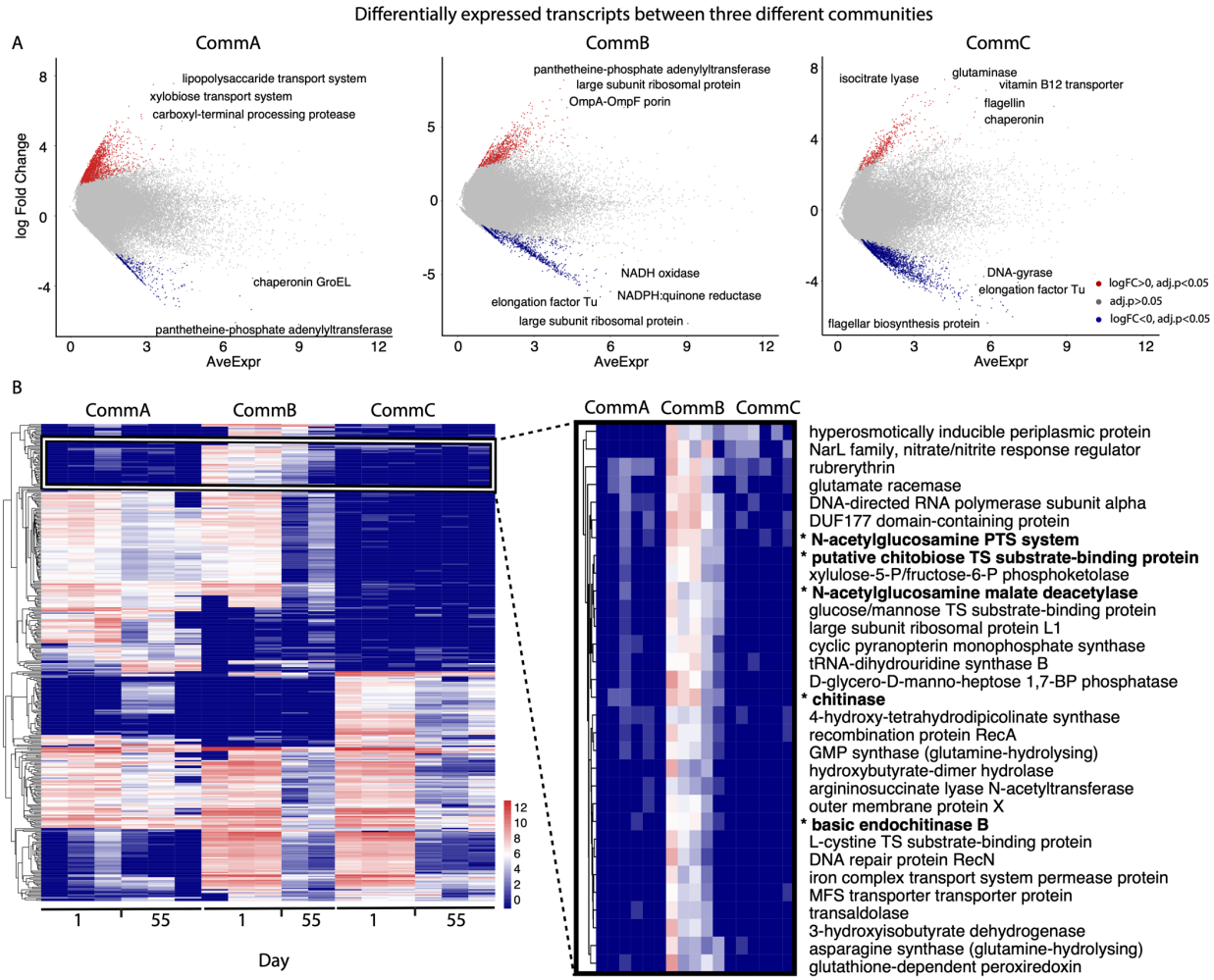

Figure 5. **Treatment communities had distinct patterns of differential gene expression. A, B** Total transcript abundance (points represent expressed genes with at least one read count) color coded by significant differential transcript abundance. **A** Red points show more highly expressed transcripts (positive log2 fold change) and blue points (negative log2 fold change) are less abundant, with p adj. < 0.05, gray are p adj > 0.05. From left to right: CommA vs. the mean of CommB and CommC, CommB vs. the mean of CommA and CommC, and CommC vs. the mean of CommA and CommB. All transcripts with |log2 fold| change greater than 6 and/or a mean of normalized counts greater than 6 were labeled when KO functional annotation was available. **B** Heatmap showing KO L4 functional pathway annotation for the significantly differentially abundant transcripts between CommA, CommB, CommC (|logFC|>4) from metatranscriptomic samples day 1 and day 55. The inset shows the subset of functions with increased transcript abundance in CommB at greater resolution (* indicates functions related to the breakdown of chitin and GlcNAC).

Connecting relevant functions to metagenome-assembled genomes

We aimed to investigate taxa with high hydrolytic enzyme production capabilities, because they might be important for pitcher plant growth. Metagenomic samples collected on day one and day 55 *in planta* (from 3 representative plants per community treatment, n=18) were taxonomically similar to 16S amplicon results, with

some differences likely due to different databases used for each dataset (Fig. S10). After quality filtering and dereplication of the initial 48 bins,36 metagenome assembled genomes (MAGs) were recovered. Of these, 26 MAGs were high quality, eight were medium quality, and two MAGs were discarded because they had a redundancy >10% (Table S3). To investigate specific functions, we mapped the mRNA functional transcripts to the MAGs and then filtered the KOs to those associated with chitinase activity or aminopeptidase (proteinase) activity (Fig. 6). MAG_09 (*Caproiciproducens* sp.) had the highest number of functional KOs associated with chitinase activity, followed by MAG_41 (*Aquitalea* sp.) and MAG_10 (*Pelosinus* sp.). Protease activity was more well-represented across the MAGs, with the highest abundance of aminopeptidase-related KOs associated with MAG_03 (*Microbacterium algeriense*), followed by MAG_38 (*Xylophilus* sp.), MAG_40 (*Cupriavidus* sp.), and MAG_33 (*Comamonas acidovorans*).

Figure 6. **MAGs enriched in functions related to chitin and protein metabolism**. The size of the bubble represents the number of unique MAG contigs mapped to functional transcripts grouped by KO (x-axis) and filtered to include only those KOs associated with targeted hydrolytic enzyme functions (chitin metabolizing KOs and aminopeptidase metabolizing KOs).

Discussion

Bacterial community functioning influences plant traits

This study clearly demonstrates a positive impact of bacterial community function on host pitcher growth. The presence of a high-functioning bacterial community, characterized by robust chitinase activity, was associated with increased pitcher growth and plant nutrient uptake. The second highest functioning community in terms of hydrolytic enzyme activity and plant hormone production was associated with a greater number of leaves. Conversely, the absence of leaf-associated bacterial inoculum was associated with lower hydrolytic enzyme activity and carbon content and resulted in reduced host plant growth. The impact of phyllosphere microbial communities have been demonstrated in other systems (e.g., 6, 90), but there has been limited research on the mechanistic effects of microbial function on host plants, partly due to challenges of sterile plant cultivation and culturing complex bacterial communities. This research utilized the unique *S. purpurea* system to compare effects of functionally distinct bacterial communities on initially sterile pitchers. These communities caused differential effects on leaf biomass and leaf number, and pitcher carbon content. These effects were measured after inoculation *in planta* over just 8 weeks, which is a rapid response for these slow-growing, long-lived plants.

Targeted bacterial functions influence plant traits

Bacterial communities can enhance nutrient cycling via secreted enzymes to decompose complex substrates and increase the availability of nutrients for host plants. The hydrolytic enzyme activity in pitcher fluid increases prey decomposition efficiency, transforming complex organic nutrient sources like protein and chitin into simpler, plant-available forms (24). Strong differences in chitinase and protease activity between bacterial communities measured *in planta* did not explain the measurable benefit to plant biomass and pitcher nutrients alone, but they likely contributed to the whole-community treatment effect.

We identified increased capacity to metabolize complex carbohydrates present in insect prey and other detritus through the screening of carbon substrate use. Two examples include GlcNAc and cellobiose. GlcNAc is a monomer of chitin, chitobiose, or chitotriose and is used for many critical metabolic functions within bacterial cells (82–84), while cellobiose is a metabolic precursor to glucose and interacts with β-glucosidase to yield two glucose molecules used in glycolysis (85). The high chitinase activity of CommB, coupled with high functional capacity for GlcNAC and cellobiose metabolism, suggests a clear role for CommB in the nutrient cycling of pitcher ecosystems. This study establishes a direct link between bacterial metabolic activity and host productivity, highlighting the role of bacterial communities in shaping nutrient dynamics within microbe-host partnerships.

Relationship between bacterial composition and function

The three bacterial communities had clear functional differences despite many shared ASVs. However, some compositional differences remained. For example, *Comamonas*, which had many aminopeptidase genes (Fig. 6C), dominated CommB, while *Elizabethkingia* was more abundant in CommA (Fig. 4A). In lake microbial communities, functional differences were only weakly related to taxonomic composition (86). In wild pitcher plants, profound functional differences were associated with modest differences in bacterial composition following manipulation of mosquito larvae abundance (Arellano, Young, Coon in review). Additionally, recent modelling work highlights specific conditions that lead to negative biodiversity ecosystem functioning relationships, where higher functioning can occur at lower species richness (87). Our results suggest that functions are not solely related to composition in these bacterial communities, but rather combinations of taxa and differential regulation of gene expression led to communities with different functional profiles.

Plants host high bacterial biodiversity, providing complex habitats and resources that create bacterial niches (88). The relatively stochastic bacterial colonization of pitcher plants results influenced by season, dispersal agents, host characteristics, and nutrient inputs (26,89,90). This study demonstrates that relatively small differences in bacterial composition within pitchers can have functional consequences for nutrient cycling from captured prey and nutrient supply to host plants. Similar patterns are emerging from human microbiome research, where surgery or drug dosage can alter gut bacterial composition, causing cascading shifts in bacterial functions which influence

host health (91). These findings underscore the relationship between bacterial community composition and function, and that functional differences are not solely dictated by composition.

Bacterial gene expression and plant growth

Using bacterial metatranscriptomics, we targeted specific genes related to known functions of interest, but also employed an agnostic approach to identify important bacterial functions that we may not have identified *a priori*. Higher expression of transcripts related to targeted chitinase, and protease enzymes were identified in CommB compared to CommA and CommC, supporting the functional measurements for chitinase related metabolism *in planta*. Protease activity was not so clearly related to transcripts of related functions. However, diverse functions of proteases unrelated to prey breakdown could complicate the relationship for proteases, along with factors including protein stability or post-transcriptional regulation (92).

Beyond nutrient transformation, bacterial communities also transcribed genes related to hormones that affect plant growth; elevated expression of genes and pathways for indole acetic acid (IAA), ACC-deaminase, cytokinins, and pectinesterase were observed in CommB. Bacterial synthesis of hormones has been shown to increase biomass (93), root growth (94), and fruit production in host plants (95). ACC-deaminase is involved in transporting and lowering ethylene levels (important for bog plants) and increasing plant growth (80) and was identified with higher expression in CommC. Bacterial production of plant growth promoting compounds may support more growth and nutrients in those pitchers.

Agnostic transcript abundance analysis allows for a more comprehensive view of the functional potential, which may be especially valuable in complex communities where diverse bacterial species are interacting. Our agnostic analysis supported functional distinctions between the three communities, and reinforced the higher expression in CommB of certain genes that might support the observed increase in pitcher growth: PanC has a crucial role in lipid and amino acid metabolism (96) so could promote N acquisition. Furthermore, higher expression of enzymes involved in the degradation and transport of chitin, chitobiose, and GlcNAc were also confirmed in CommB, complementing the targeted approach and the enzyme activity measurements.

494

495 **Conclusions**

496 In this controlled study, plant growth was mediated by bacterial community function. The most growth was
497 observed in pitchers hosting bacterial communities with high hydrolytic enzyme activity and elevated plant
498 hormone gene expression. Bacterial community functions were characterized using complementary approaches
499 and identified direct mechanistic relationships between bacterial functions, expression of genes associated with
500 these functions, and the support of plant leaf growth and nutrient acquisition. Plant-microbe interactions are
501 complex, and benefits to host productivity are likely not driven by a single functional parameter, or even a single
502 taxon expressing a function of interest. This bacterial community analysis sheds light on how measurable
503 functional traits can identify relevant bacterial effects in phyllosphere communities. Identification of these specific
504 mechanistic relationships contribute increased knowledge of the direct effects of bacterial communities on plant
505 traits and can inform other systems with applications in agriculture, conservation, and restoration.

506

507 **Acknowledgements**

508 We would like to acknowledge Stacey Pedraza for help with processing samples, the late Bill Freutel and other
509 members of the Earth, Wind, Fire Lab for use of lab space, Dr. Linda Reynard for FLASH assistance, and Franco Nero
510 and Jacob Heil for help with experiments. We offer our gratitude to Dr. Allison Simler-Williamson and the
511 Ecological Statistics Community at BSU for feedback, resources, and instruction. We would like to thank Dr. Ben
512 Baiser and Dr. Zac Freedman helpful suggestions and feedback on the final manuscript.

513

514 **Data Availability Statement**

515 Scripts and data associated with this manuscript are available at: https://github.com/jessibernardin/microbial-
516 function-plant_trait (https://doi.org/10.5281/zenodo.10055295). The raw reads for 16S rRNA amplicon
517 sequencing, raw reads for whole genome shotgun sequencing, whole genome metatranscriptomic sequencing, and
518 nucleotide sequences of MAGs have been deposited in the National Center for Biotechnology Information (NCBI)
519 under the project number PRJNA1028624.

520

Competing Interests

Authors have no competing interests to declare.

Funding Statement

JRB, LSB and EBY were supported with funding from the National Science Foundation through an Understanding the Rules of Life-Microbiome Theory and Mechanisms 2 collaborative grant (award #2025250 to LSB and award # 2025337 to EBY). LSB was also funded by an award from the Simons Foundation, LS-ECIAMEE-00001638.

References

1. Vandenkoornhuyse P, Quaiser A, Duhamel M, Le Van A, Dufresne A. The importance of the microbiome of the plant holobiont. New Phytologist. 2015;206(4):1196–206.

2. Simon JC, Marchesi JR, Mougel C, Selosse MA. Host-microbiota interactions: from holobiont theory to analysis. Microbiome. 2019;7(1):5.

3. Stone B, Weingarten E, Jackson C. The role of the phyllosphere microbiome in plant health and function. 2018 Aug 14;1:1–24.

4. Liu H, Brettell LE, Singh B. Linking the Phyllosphere Microbiome to Plant Health. Trends in Plant Science. 2020;25(9):841–4.

5. Trivedi P, Leach JE, Tringe SG, Sa T, Singh BK. Plant-microbiome interactions: from community assembly to plant health. Nat Rev Microbiol. 2020 Nov;18(11):607–21.

6. Mehlferber EC, McCue KF, Debray R, Kaulbach G, Ferrel JE, Khanna R, et al. Early phyllosphere microbial associations impact plant reproductive success. bioRxiv; 2022.

7. Delmotte N, Knief C, Chaffron S, Innerebner G, Roschitzki B, Schlapbach R, et al. Community proteogenomics reveals insights into the physiology of phyllosphere bacteria. Proceedings of the National Academy of Sciences. 2009;106(38):16428–33.

8. Vorholt JA. Microbial life in the phyllosphere. Nat Rev Microbiol. 2012;10(12):828–40.

9. Flores-Núñez V, Camarena-Pozos D, Chávez-González D, Alcalde-Vázquez R, Vázquez-Sánchez N, Hernández-Melgar A, et al. Microbiome-inspired phyllospheric synthetic communities increase microbial diversity and productivity of field-grown plants of Agave tequilana. In Review; 2022 Jul.

10. Holland MA. Nitrogen: Give and Take from Phylloplane Microbes. In: Ecological Aspects of Nitrogen Metabolism in Plants. John Wiley & Sons, Ltd; 2011. p. 215–30.

11. Kende H, Zeevaart J. The Five "Classical" Plant Hormones. Plant Cell. 1997 Jul;9(7):1197–210.

12. MacMillan J. Occurrence of Gibberellins in Vascular Plants, Fungi, and Bacteria. J Plant Growth Regul. 2001 Dec;20(4):387–442.

13. Spaepen S, Vanderleyden J, Remans R. Indole-3-acetic acid in microbial and microorganism-plant signaling. FEMS Microbiol Rev. 2007 Jul;31(4):425–48.

14. Spaepen S. Plant Hormones Produced by Microbes. In: Lugtenberg B, editor. Principles of Plant-Microbe Interactions: Microbes for Sustainable Agriculture. Cham: Springer International Publishing; 2015. p. 247–56.

15. Lievens L, Pollier J, Goossens A, Beyaert R, Staal J. Abscisic Acid as Pathogen Effector and Immune Regulator. Front Plant Sci. 2017 Apr 19;8:587.

16. Porter SS, Bantay R, Friel CA, Garoutte A, Gdanetz K, Ibarreta K, et al. Beneficial microbes ameliorate abiotic and biotic sources of stress on plants. Bennett A, editor. Funct Ecol. 2020 Oct;34(10):2075–86.

17. Aziz U, Rehmani MS, Wang L, Luo X, Xian B, Wei S, et al. Toward a Molecular Understanding of Rhizosphere, Phyllosphere, and Spermosphere Interactions in Plant Growth and Stress Response. Critical Reviews in Plant Sciences. 2021 Nov 2;40(6):479–500.

18. Bradshaw WE, Creelman RA. Mutualism between the Carnivorous Purple Pitcher Plant and its Inhabitants. The American Midland Naturalist. 1984;112(2):294–304.

19. Givnish TJ, Burkhardt EL, Happel RE, Weintraub JD. Carnivory in the Bromeliad Brocchinia reducta, with a Cost/Benefit Model for the General Restriction of Carnivorous Plants to Sunny, Moist, Nutrient-Poor Habitats. The American Naturalist. 1984 Oct;124(4):479–97.

20. Heard SB. Capture Rates of Invertebrate Prey by the Pitcher Plant, Sarracenia purpurea L. amid. 1998 Jan;139(1):79–89.

21. Miller TE, Horth L, Reeves RH. Trophic interactions in the phytotelmata communities of the Pitcher Plant,Sarracenia purpurea. COMMUNITY ECOLOGY. 2002;3(1):109–16.

22. Gallie DR, Chang SC. Signal Transduction in the Carnivorous Plant Sarracenia purpurea (Regulation of Secretory Hydrolase Expression during Development and in Response to Resources). Plant Physiology. 1997 Dec 1;115(4):1461–71.

23. Butler JL, Gotelli NJ, Ellison AM. Linking the brown and green: nutrient transformation and fate in the *Sarracenia* microecosystem. Ecology. 2008;89(4):898–904.

24. Young EB, Sielicki J, Grothjan JJ. Regulation of hydrolytic enzyme activity in aquatic microbial communities hosted by carnivorous pitcher plants. Microb Ecol. 2018 Nov;76(4):885–98.

25. Butler JL, Ellison AM. Nitrogen cycling dynamics in the carnivorous northern pitcher plant, Sarracenia purpurea. Functional Ecology. 2007;21(5):835–43.

26. Bittleston LS, Wolock CJ, Yahya BE, Chan XY, Chan KG, Pierce NE, et al. Convergence between the microcosms of Southeast Asian and North American pitcher plants. Newman DK, Weigel D, editors. eLife. 2018;7.

27. Bittleston LS, Gralka M, Leventhal GE, Mizrahi I, Cordero OX. Context-dependent dynamics lead to the assembly of functionally distinct microbial communities. Nat Commun. 2020 Dec;11(1):1440.

28. Ueno H, Miyashita K. Inductive production of chitinolytic enzymes in soil microcosms using chitin, other carbon-sources, and chitinase-producing *Streptomyces*. Soil Science and Plant Nutrition. 2000 Dec;46(4):863–71.

29. Choi KH, Dobbs FC. Comparison of two kinds of Biolog microplates (GN and ECO) in their ability to distinguish among aquatic microbial communities. Journal of Microbiological Methods. 1999 Jun 1;36(3):203–13.

30. Ellison AM, Gotelli NJ. Nitrogen availability alters the expression of carnivory in the northern pitcher plant, Sarracenia purpurea. PNAS. 2002 Apr 2;99(7):4409–12.

31. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc Natl Acad Sci USA. 2011;108:4516–22.

32. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High resolution sample inference from Illumina amplicon data. Nat Methods. 2016 Jul;13(7):581–3.

33. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol. 2019 Aug;37(8):852–7.

34. McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, et al. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. ISME J. 2012 Mar;6(3):610–8.

35. Katoh K. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research. 2002 Jul 15;30(14):3059–66.

36. Price MN, Dehal PS, Arkin AP. FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. Molecular Biology and Evolution. 2009 Jul 1;26(7):1641–50.

37. McMurdie PJ, Holmes S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLOS ONE. 2013 Apr 22;8(4):e61217.

38. Bisanz, Jordan E. qiime2R: Importing QIIME2 artifacts and associated data into R sessions. 2018;(v0.99). Available from: https://github.com/jbisanz/qiime2R

39. Kembel SW, Ackerly DD, Blomberg SP, Cornwell WK, Cowan PD, Helmus MR, et al. picante: Integrating Phylogenies and Ecology [Internet]. 2020 [cited 2023 Sep 26]. Available from: https://cran.r-project.org/web/packages/picante/index.html

40. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014 Aug 1;30(15):2114–20.

41. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016 Oct 1;32(19):3047–8.

42. Deng ZL, Münch PC, Mreches R, McHardy AC. Rapid and accurate identification of ribosomal RNA sequences via deep learning. Nucleic Acids Res. 2022 Feb 21;50(10):e60.

43. Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 2015 May 15;31(10):1674–6.

44. Li D, Luo R, Liu CM, Leung CM, Ting HF, Sadakane K, et al. MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. Methods. 2016 Jun 1;102:3–11.

45. Brown CT, Irber L. sourmash: a library for MinHash sketching of DNA. Journal of Open Source Software. 2016 Sep 14;1(5):27.

46. Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil PA, Hugenholtz P. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. Nucleic Acids Research. 2022 Jan 7;50(D1):D785–94.

47. Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, et al. Binning metagenomic contigs by coverage and composition. Nat Methods. 2014 Nov;11(11):1144–6.

48. Wu YW, Simmons BA, Singer SW. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics. 2016 Feb 15;32(4):605–7.

650 49. Sieber CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, et al. Recovery of
651 genomes from metagenomes via a dereplication, aggregation and scoring strategy. Nat
652 Microbiol. 2018 Jul;3(7):836–43.

653 50. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, et al. MetaBAT 2: an adaptive binning
654 algorithm for robust and efficient genome reconstruction from metagenome assemblies.
655 PeerJ. 2019 Jul 26;7:7359.

656 51. Chklovski A, Parks DH, Woodcroft BJ, Tyson GW. CheckM2: a rapid, scalable and accurate
657 tool for assessing microbial genome quality using machine learning. bioRxiv; 2022.

658 52. Olm MR, Brown CT, Brooks B, Banfield JF. dRep: a tool for fast and accurate genomic
659 comparisons that enables improved genome recovery from metagenomes through de-
660 replication. ISME J. 2017 Dec;11(12):2864–8.

661 53. Campbell JH, O'Donoghue P, Campbell AG, Schwientek P, Sczyrba A, Woyke T, et al. UGA is
662 an additional glycine codon in uncultured SR1 bacteria from the human microbiota.
663 Proceedings of the National Academy of Sciences. 2013 Apr 2;110(14):5540–5.

664 54. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, et al.
665 Minimum information about a single amplified genome (MISAG) and a metagenome-
666 assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol. 2017
667 Aug;35(8):725–31.

668 55. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012
669 Mar 4;9(4):357–9.

670 56. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of
671 SAMtools and BCFtools. GigaScience. 2021;10(2):8.

672 57. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene
673 recognition and translation initiation site identification. BMC Bioinformatics. 2010 Mar
674 8;11(1):119.

675 58. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq
676 quantification. Nat Biotechnol. 2016 May;34(5):525–7.

677 59. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, et al. KofamKOALA:
678 KEGG Ortholog assignment based on profile HMM and adaptive score threshold. Valencia
679 A, editor. Bioinformatics. 2020 Apr 1;36(7):2251–2.

680 60. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna,
681 Austria: R Foundation for Statistical Computing; 2019. Available from: https://www.R-
682 project.org/

61. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. vegan: Community Ecology Package [Internet]. 2019. Available from: https://CRAN.R-project.org/package=vegan

62. Wickham H. reshape2: Flexibly Reshape Data: A Reboot of the Reshape Package [Internet]. 2020 [cited 2023 Sep 26]. Available from: https://cran.r-project.org/web/packages/reshape2/index.html

63. Firke S, Denney B, Haid C, Knight R, Grosser M, Zadra J. janitor: Simple Tools for Examining and Cleaning Dirty Data [Internet]. 2023 [cited 2023 Sep 26]. Available from: https://cran.r-project.org/web/packages/janitor/index.html

64. Wickham H, RStudio. tidyverse: Easily Install and Load the "Tidyverse" [Internet]. 2023 [cited 2023 Sep 26]. Available from: https://cloud.r-project.org/web/packages/tidyverse/index.html

65. Lin H, Peddada SD. Analysis of compositions of microbiomes with bias correction. Nat Commun. 2020 Jul 14;11(1):3514.

66. Kolde R. pheatmap: Pretty Heatmaps [Internet]. 2019 [cited 2023 Sep 25]. Available from: https://cran.r-project.org/web/packages/pheatmap/index.html

67. Lozupone C, Lladser ME, Knights D, Stombaugh J, Knight R. UniFrac: an effective distance metric for microbial community comparison. ISME J. 2011 Feb;5(2):169–72.

68. Law CW, Chen Y, Shi W, Smyth GK. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biology. 2014 Feb 3;15(2):R29.

69. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990 Oct 5;215(3):403–10.

70. Cinner JE, Huchery C, MacNeil MA, Graham NAJ, McClanahan TR, Maina J, et al. Bright spots among the world's coral reefs. Nature. 2016 Jul 1;535(7612):416–9.

71. Kimmel K, Dee LE, Avolio ML, Ferraro PJ. Causal assumptions and causal inference in ecological experiments. Trends in Ecology & Evolution. 2021 Dec 1;36(12):1141–52.

72. Bürkner PC. Advanced Bayesian Multilevel Modeling with the R Package brms. The R Journal. 2018;10(1):395.

73. Lüdecke D. ggeffects: Tidy Data Frames of Marginal Effects from Regression Models. Journal of Open Source Software. 2018 Jun 29;3(26):772.

74. Gabry J, Mahr T, Bürkner PC, Modrák M, Barrett M, Weber F, et al. bayesplot: Plotting for Bayesian Models [Internet]. 2022 [cited 2023 Sep 26]. Available from: https://cran.r-project.org/web/packages/bayesplot/index.html

75. Kay M, Mastny T. tidybayes: Tidy Data and “Geoms” for Bayesian Models [Internet]. 2023 [cited 2023 Sep 26]. Available from: https://cloud.r-project.org/web/packages/tidybayes/index.html

76. Eager CD. standardize: Tools for Standardizing Variables for Regression in R [Internet]. 2021 [cited 2023 Sep 26]. Available from: https://cran.r-project.org/web/packages/standardize/

77. Gelman A, Hill J. Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press; 2006. 651 p.

78. Gelman A, Rubin DB. Inference from Iterative Simulation Using Multiple Sequences. Statistical Science. 1992;7(4):457–72.

79. Moran GR. 4-Hydroxyphenylpyruvate dioxygenase. Arch Biochem Biophys. 2005 Jan 1;433(1):117–28.

80. Herpell JB, Alickovic A, Diallo B, Schindler F, Weckwerth W. Phyllosphere symbiont promotes plant growth through ACC deaminase production. ISME J. 2023;

81. De la Vega-Camarillo E, Sotelo-Aguilar J, Rios-Galicia B, Mercado-Flores Y, Arteaga-Garibay R, Villa-Tanaca L, et al. Promotion of the growth and yield of Zea mays by synthetic microbial communities from Jala maize. Frontiers in Microbiology. 2023;14.

82. Naseem S, Parrino SM, Buenten DM, Konopka JB. Novel roles for GlcNAc in cell signaling. Commun Integr Biol. 2012 Mar 1;5(2):156–9.

83. Beier S, Bertilsson S. Bacterial chitin degradation—mechanisms and ecophysiological strategies. Frontiers in Microbiology. 2013;4.

84. Jiang WX, Li PY, Chen XL, Zhang YS, Wang JP, Wang YJ, et al. A pathway for chitin oxidation in marine bacteria. Nat Commun. 2022 Oct 6;13(1):5899.

85. Li J, Gu S, Zhao Z, Chen B, Liu Q, Sun T, et al. Dissecting cellobiose metabolic pathway and its application in biorefinery through consolidated bioprocessing in Myceliophthora thermophila. Fungal Biology and Biotechnology. 2019 Nov 13;6(1):21.

86. Langenheder S, Lindström ES, Tranvik LJ. Weak coupling between community composition and functioning of aquatic bacteria. Limnology and Oceanography. 2005;50(3):957–67.

87. D’Andrea R, Khattar G, Koffel T, Frans V, Bittleston L, Cuellar-Gempeler C. Reciprocal inhibition and competitive hierarchy cause negative biodiversity-ecosystem function relationships. Preprints; 2023.

88. Morris C, Kinkel L, Lindow S, Hecht-Poinar E, Elliott VJ. Fifty years of phyllosphere microbiology: significant contributions to research in related fields. In 2002.

749 89. Grothjan JJ, Young EB. Diverse microbial communities hosted by the model carnivorous
750 pitcher plant *Sarracenia purpurea* : analysis of both bacterial and eukaryotic composition
751 across distinct host plant populations. PeerJ. 2019 Feb 18;7:e6392.

752 90. Armitage DW. Linking the development and functioning of a carnivorous pitcher plant's
753 microbial digestive community. The ISME Journal. 2017 Nov;11(11):2439–51.

754 91. Fernando DG, Saravia FL, Atkinson SN, Barron M, Kirby JR, Kindel TL. A single, peri-
755 operative antibiotic can persistently alter the post-operative gut microbiome after Roux-
756 en-Y gastric bypass. Surg Endosc. 2023;37(2):1476–86.

757 92. López-Otín C, Bond JS. Proteases: Multifunctional Enzymes in Life and Disease. Journal of
758 Biological Chemistry. 2008;283(45):30433–7.

759 93. Patten CL, Glick BR. Bacterial biosynthesis of indole-3-acetic acid. Can J Microbiol.
760 1996;42(3):207–20.

761 94. Park S, Kim AL, Hong YK, Shin JH, Joo SH. A highly efficient auxin-producing bacterial strain
762 and its effect on plant growth. J Genet Eng Biotechnol. 2021;19:179.

763 95. Egamberdieva D, Wirth SJ, Alqarawi AA, Abd_Allah EF, Hashem A. Phytohormones and
764 Beneficial Microbes: Essential Components for Plants to Balance Stress and Fitness.
765 Frontiers in Microbiology. 2017;8.

766 96. Begley TP, Kinsland C, Strauss E. The biosynthesis of coenzyme a in bacteria. In: Vitamins &
767 Hormones. Academic Press; 2001. p. 157–71. (Cofactor Biosynthesis; vol. 61).

768