Environmental proteomics reveals taxonomic and functional changes in an enriched aquatic ecosystem

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

| Citation      | Northrop, Amanda C., Rachel K. Brooks, Aaron M. Ellison, Nicholas J. Gotelli, and Bryan A. Ballif. 2017. “Environmental Proteomics Reveals Taxonomic and Functional Changes in an Enriched Aquatic Ecosystem.” Ecosphere 8 (10) (October): e01954. doi:10.1002/ecs2.1954. |
|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Published Version | 10.1002/ecs2.1954                                                                                                                                                                                                                                            |
| Citable link  | http://nrs.harvard.edu/urn-3:HUL.InstRepos:34389684                                                                                                                                                                                                          |
| Terms of Use  | This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA |
Environmental proteomics reveals taxonomic and functional changes in an enriched aquatic ecosystem

Authors: Amanda C. Northrop, Rachel Brooks, Aaron M. Ellison, Nicholas J. Gotelli, and Bryan A. Ballif

Affiliations:

1 Department of Biology, University of Vermont, Burlington, VT 05405, USA.
2 Harvard Forest, Harvard University, Petersham, MA 01366, USA.

*Corresponding authors. E-mail: bbalif@uvm.edu (B.A.B); ngotelli@uvm.edu (N.J.G.)

Conflict of Interest

The authors declare no conflict of interest.
Abstract
Aquatic ecosystem enrichment can lead to distinct and irreversible changes to undesirable states. Understanding changes in active microbial community function and composition following organic-matter loading in enriched ecosystems can help identify biomarkers of such state changes. In a field experiment, we enriched replicate aquatic ecosystems in the pitchers of the northern pitcher plant, *Sarracenia purpurea*. Shotgun metaproteomics using a custom metagenomic database identified proteins, molecular pathways, and contributing microbial taxa that differentiated control ecosystems from those that were enriched. The number of microbial taxa contributing to protein expression was comparable between treatments; however, taxonomic evenness was higher in controls. Functionally active bacterial composition differed significantly among treatments and was more divergent in control pitchers than enriched pitchers. Aerobic and facultative anaerobic bacteria contributed most to identified proteins in control and enriched ecosystems, respectively. The molecular pathways and contributing taxa in enriched pitcher ecosystems were similar to those found in larger enriched aquatic ecosystems and are consistent with microbial processes occurring at the base of detrital food webs. Detectable differences between protein profiles of enriched and control ecosystems suggest that a time series of environmental proteomics data may identify protein biomarkers of impending state changes to enriched states.

Key words: aquatic ecosystems; bacterial communities; environmental proteomics; model ecosystem; organic matter enrichment; *Sarracenia purpurea*. 
Introduction

Chronic and directional environmental drivers such as nutrient and organic matter enrichment are causing state changes in many ecosystems (Rabalais et al. 2009, Scheffer 2009). Mitigating or preventing these state changes requires predicting them with sufficient lead-time (Biggs et al. 2009). Current prediction methods rely on the statistical signature of “critical slowing down” (Scheffer et al. 2009) – an increase in the variance or temporal autocorrelation of a state variable (Dakos et al. 2015). However, such indicators usually require long time series of data with frequent sampling of an appropriate state variable (Bestelmeyer et al. 2011, Levin and Mollmann 2015). Even when such data are available, the signature of critical slowing down may not provide enough lead-time for intervention (Biggs et al. 2009, Contamin and Ellison 2009).

In aquatic systems, water quality indicators such as total suspended solids (Hargeby et al. 2007), submersed macrophyte vegetation cover (Dennison et al. 1993, Sondergaard et al. 2010), diatom composition (Pan et al. 1996), and phytoplankton biomass (Carpenter et al. 2008) often are used as state variables. However, whether top-down or bottom-up forces initiate the change, the proximate cause of eutrophication in many freshwater aquatic ecosystems is microbial processes associated with the breakdown of detritus (Chrost and Siuda 2006). A primary reason that it has been difficult to forecast shifts with sufficient lead-time may be that changes in monitored variables lag behind the microbial processes that underlie state changes. We hypothesize that biomarkers linked closely to microbial function, such as proteins, may serve as better early warning signals of impending state changes than traditional aquatic ecosystem biomarkers.
One of the challenges to studying aquatic ecosystem state changes is the lack of replicable natural ecosystems that can be ethically manipulated. Recently, we have identified the aquatic ecosystem that assembles in the cup-shaped leaves of the northern pitcher plant *Sarracenia purpurea* as a model system for identifying whole-ecosystem microbial processes associated with detrital enrichment. Each leaf functions as an independent ecosystem that can be experimentally enriched and monitored through time in the field or lab (Srivastava et al. 2004). Arthropod prey, mostly ants and flies, form the base of a “brown” food web that includes dipteran larvae, protozoa, mites, rotifers, and a diverse assemblage of bacteria that decompose and mineralize nearly all the captured prey biomass (Ellison et al. 2003, Butler et al. 2008, Koopman and Carstens 2011, Gray et al. 2012). Even in the absence of macroinvertebrates, the dominant transfer of nutrients to the plant occurs via microbial activity (Butler et al. 2008). With excess organic matter loading, microbial activity increases, pitcher fluid becomes turbid, and oxygen levels collapse to hypoxic conditions even during daytime photosynthesis (Sirota et al. 2013). Such consequences are similar to those seen in larger aquatic ecosystems that have switched from a green to a brown food web dominated by detritivores, as an initial increase in primary production leads to internal organic-matter loading and increasing biological oxygen demand as primary producers decompose (Correll 1998).

In the last decade, environmental proteomics has emerged as a powerful tool to measure microbial community function in a variety of aquatic habitats, including contaminated groundwater (Benndorf et al. 2007), coastal upwelling systems (Sowell et al. 2011), estuaries (Colatriano et al. 2015), and meromictic lakes (Lauro et al. 2011). Additionally, environmental proteomics has promise as a tool for identifying biomarkers
of changing environmental conditions, including aquatic pollution (Campos et al. 2012, Ullrich et al. 2016). Environmental proteomics looks at the complete set of proteins expressed in an ecosystem at a single time point and gives insight into the function of a community. While metatranscriptomics also serves as an important tool for understanding community function, mRNA and protein levels are generally not strongly correlated (Vogel and Marcotte 2012); this is especially true for bacteria in perturbed systems (Jayapal et al. 2008). Therefore, metaproteomics may provide a more accurate picture of bacterial community function in enriched aquatic habitats.

As a first step toward determining the utility of microbial protein biomarkers as early warning signals of state changes we conducted an environmental proteomics screen of the aquatic ecosystem in *S. purpurea* pitchers enriched with organic matter to determine whether there are detectable differences between the proteins, associated molecular pathways, and taxa contributing to expressed proteins in microbial (nonviral organisms <30 µm) communities in enriched vs. control ecosystems. We hypothesized that an environmental proteomics survey would reveal detectable differences in taxa contributing to protein expression, proteins, and functional pathways between enriched and control ecosystems. We expected to find differences between control and enriched pitchers in pathways related to respiration and decomposition, changes in the oxygen requirement of microbes contributing to expressed proteins, and shifts in the taxonomic composition of microbes contributing to protein expression. Specifically, we predicted an abundance of contributing anaerobic bacteria in enriched pitchers relative to controls. We identified and found detectable differences in taxa, proteins, and pathways common to a wide range of aquatic ecosystems. Our results suggest that environmental proteomics can
be a useful tool for detecting alternative enriched and unenriched states in aquatic ecosystems and may serve as a means to identify protein biomarkers of impending shifts between such states.

Methods

Enrichment Experiment

The field experiment was conducted in Tom Swamp, a nutrient-poor fen located at the northern end of Harvard Pond (42.51° N, −72.21° W) at Harvard Forest, Worcester County, Massachusetts. Newly opened pitchers were identified and randomly assigned to an ambient control or detritus-enriched treatment (Appendix S1). Previous work by Peterson et al. (2008) using culture-independent methods revealed that newly opened pitchers are sterile and impermeable to bacteria, so we are reasonably sure that our experimental pitchers did not harbor diverse bacterial communities prior to the start of the experiment. Detritus-enriched pitchers received 1 mg ml⁻¹ d⁻¹ of oven-dried, finely ground wasps (Dolichovespula maculata) (Appendix S1), which have elemental ratios (C:N, 5.99:1, N:P:K, 10.7:1.75:1.01) similar to those of Sarracenia’s natural ant prey (C:N, 5.9:1; N:P:K, 12.1:1.52:0.93) (Farnsworth and Ellison 2008). Proteomic analysis of the ground wasp (not reported here) failed to identify microbial proteins, so we are confident that microbial contribution to enriched pitchers from the wasps was minimal. Enrichment treatments were applied for 14 consecutive days; all pitchers were otherwise unmanipulated. Pitcher fluid was sampled on the first and last days of the experiment, filtered to remove microbes > 30 µm, pelleted, and stored at −80 °C until processed (Appendix S1).
Protein Extraction, SDS-Page, and Mass Spectrometry

Six of ten replicate microbial pellets from each treatment yielded enough protein for analysis via tandem mass spectrometry. All replicates were analyzed separately using SDS-PAGE and Coomassie staining (Fig. 1, Appendix 1: Fig. S1a, and Appendix 1: Fig. S1b). All six of the enriched pitchers and five of the six control pitchers had visible protein staining levels and were chosen for mass spectrometry. Proteins were subjected to a tryptic digest (Appendix S1) and to LC-MS/MS as previously described (Cheerathodi and Ballif 2011) using a linear ion trap mass spectrometer (Thermo Electron, Waltham, MA, USA). MS/MS spectra were matched to peptides in a custom protein database using SEQUEST software as described below.

Custom Metagenomic Databases

We generated a custom protein database from a six-frame forward and reverse translation of a metagenomic database constructed from microbial communities of three previously collected pitchers that had captured diverse amounts of prey (Appendix 1: Fig. S2). Pitchers were collected from Molly Bog, an ombrotrophic bog located in Morristown, VT (44.50° N, -72.64° W) on 18 August 2008 and transported in a cooler directly from the field to the University of Vermont. Microbial pellets were obtained immediately as described above. DNA was extracted, prepared, and sent for library construction, sequencing, and assembly to Genome Québec (Montréal, QC, Canada) with the 454 GS-FLX Titanium Sequencing System (Roche) (Appendix S1). Contigs were assembled de-novo with Roche’s Newbler assembler v2.3 (release 091027_1459) using default
parameters (minimum Read Length = 20; overlap Seed Step = 12; overlap Seed Length = 16; overlap Min Seed Count = 1; overlap Seed Hit Limit = 70; overlap Min Match Length = 40; overlap Min Match Identity = 90; overlap Match Ident Score = 2; overlap Match Diff Score = -3; overlap Match Unique Thresh = 12; map Min Contig Depth = 1; all Contig Thresh = 100), with the exception of minimum read length (20 bp) and overlap Hit Position Limit (1,000,000). The assembled contigs were imported into MG-RAST 4.0.2 to assess functional and taxonomic potential (Meyer et al. 2008). Taxonomic assignments were visualized using the Krona plugin and the following cutoffs were applied to both taxonomic and subsystem functional category assignments: minimum identity = 60%, e-value of 1 x 10^{-5} or less, and a minimum alignment length of 15 bp (Appendix 1: Fig. S3). We calculated Hurlbert’s probability of an interspecific encounter (\textit{PIE}) to estimate the evenness of bacterial classes in the metagenome (Hurlbert 1971) (Appendix S1). KEGG pathways (level 2 and level 3) were assigned to contigs using the KEGG database via MG-RAST (we report only the top 73 level 3 pathways here)(Appendix 1: Fig. S4).

A metaproteomic database was created with a six-frame forward and reverse translation of the assembled metagenome using open-source Ruby software. Sequences with greater than 100 amino acids (n=184,128) in length were retained. A decoy database was constructed by reversing the retained sequences and concatenating them to the forward database to allow for an estimation of the false discovery rate as has been described (Elias and Gygi 2007).

\textit{Protein Orthologue Identification}
Peptide and protein identifications were made via a SEQUEST search of the tandem mass spectral data against the custom pitcher-plant microbial community protein database described above (Appendix S1). The number of protein hits varied substantially among replicates, so to have enough proteins for treatment comparisons, peptides and proteins from the five control samples and six enriched samples were pooled after LC-MS/MS and the SEQUEST search into a single control and a single enriched sample dataset. The doubly- and triply-charged peptide ions were further considered and each dataset was filtered by first adjusting the cutoffs for XCorr and ΔCn until the false discovery rate was < 10%. The final filters were: Xcorr ≥ 3.0 for doubly-charged ions, Xcorr ≥ 3.3 for triply-charged ions and unique Δcorr ≥ 0.15. The resulting list of protein hits for each treatment was then ranked by unique number of peptides and the top 220 proteins from each treatment were selected so that the false discovery rate for control and enriched treatments were 6.6% and 0%, respectively. These top 220 proteins and their associated peptides are found in Data Supplement S1.

In the list of control peptides, a protein hit from the decoy database was represented by 25 total peptides; therefore, we suspected that this hit was a true positive not represented in our target database. However, a BLAST search of the full amino acid sequence did not yield an identical match, so we cannot definitively claim it is a true positive; therefore, we removed this peptide from our top 220 list of control peptides. With this peptide removed, the false discovery rate for the control treatment was 4.3%.

All peptide hits were pooled within treatments and mapped back to their source sequences in the custom protein database. Those source sequences were imported in fasta file format into blast2go v.2.8.0 (Conesa et al. 2005) for identification and annotation.
using the following configuration settings: blastp program, Blast Expect Value of 1.0E-3, 10 Blast Hits, Annotation CutOff >55, GO Weight >5.

Analysis of the Top Proteins Shared Between Treatments

A randomization test was done using R Studio (v. 0.98.1059) to test the hypothesis that there was a single common protein pool for both the control and enriched treatments and that the number of observed shared proteins between treatments reflects chance effects resulting from random draws from this single protein pool (Appendix S1). We conducted an additional simulation in R to determine the likelihood of a Type I error in our randomization test (Appendix S1).

Comparison of the Top 20 Proteins from Each Treatment

We downloaded the sequence by annotation file from the blast2go search for each treatment to get the protein names associated with each protein hit (sequence description in blast2go). Each of the top 220 identified proteins in each treatment, ordered by the number of total peptides associated with the protein hit, was matched to a protein name using R software. If multiple protein hits within a treatment matched a single protein name, the protein names were merged in silico and the total peptides representing them were summed. Protein names were ranked in order of the abundance of total peptides for each treatment.

Taxonomic Analysis
To determine the taxonomic composition of the microbes contributing to identified proteins in our treatments, we conducted a BLAST homology search of the metagenomic sequence data for protein hits. All peptides from the top 220 identified proteins in each treatment were mapped back to their contigs of origin to obtain nucleotide sequences. Because contigs were at least 500 base pairs in length, we felt confident that a BLAST search of the nucleotide sequences would yield correct taxonomic identifications at course taxonomic levels and acknowledge that ambiguity can remain in the taxonomic identification from a metacommunity at genus and species levels. The top BLAST hit was retained for each nucleotide sequence associated with an identified protein and linked to a bacterial class (Appendix S1). For each bacterial class identified, a 2×2 contingency table was created with treatments as columns and the number of peptides associated and not associated with the taxon as rows. A chi-square test was then used to determine if the abundance of the bacterial class was significantly different between treatments. All P values were adjusted using the Benjamini-Hochberg method (Benjamini and Hochberg 1995) (Table 1). Species composition was visualized using Krona (Ondov et al. 2011) (Appendix 1: Fig. S5). In addition to the BLAST homology search, we used Unipept (Mesuere 2016) to map tryptic peptides to the UniprotKB database and retrieve the least common taxonomic ancestor (= most derived shared taxonomic node) associated with each peptide for pooled replicates (Appendix 1: Fig. S6). We calculated Hurlbert’s PIE to estimate the evenness of bacterial classes contributing to expressed proteins in control and enriched pitchers (Hurlbert 1971) (Appendix S1).

Functional Analysis
Functional pathways (two levels) associated with each identified protein from each treatment were retrieved using the KEGG (Kyoto Encyclopedia of Genes and Genomes) (Kanehisa et al. 2014) mapping function of blast2go v.2.8.0. Each pathway was weighted by the total number of peptides associated with protein hits, or the number of spectral counts, mapping to that pathway (Appendix 1: Fig. S7). For each pathway identified, a 2×2 contingency table was created with treatments as columns and the number of peptides associated and not associated with the pathway as rows. A chi-square test was used to determine if each pathway was significantly over- or under-represented in enriched pitchers relative to controls. All P values were adjusted using the Benjamini-Hochberg method (Benjamini and Hochberg 1995) (Appendix 1: Table S1).

To determine whether bacteria contributing to expressed proteins in control and enriched ecosystems differed in their O₂ requirements, we mapped each bacterial species identified in our BLAST search to its O₂ requirement using data from the Integrated Microbial Genomes database (IMG) (Timinskas et al. 2014, Reddy et al. 2015) (Appendix S1, Fig. ). The IMG database contains 6 classes of O₂ requirements: aerobe, anaerobe, facultative, microaerophillic, obligate aerobe, and obligate anaerobe. The latter three categories make up less than 7% of the database. We merged any species classified as obligate aerobes or obligate anaerobes into the aerobe and anaerobe classes, respectively.

Analysis of Unpooled Data

In addition to analyzing pooled data, we used ordination and permutation analyses to determine the effect of enrichment on microbial community protein expression,
taxonomic contribution to expressed proteins at the class and family levels, and KEGG pathways. We tested the similarity within and among replicates of control and enriched microbial communities using ADONIS, a nonparametric permutation test in the ‘vegan’ package (v. 2.4.1) in R (Oksanen et al. 2016). We used a multivariate homogeneity of group dispersions test (betadisper function in the ‘vegan’ package) to determine if the composition of contributing microbial taxa was more divergent in control replicates than in enriched replicates. The permutation tests used 999 permutations and were done using total peptide counts associated with protein identifications, microbial classes, microbial families, and KEGG pathways (Table 2). To visualize the similarities among replicate ecosystems, we used the ‘vegan’ package function metaMDS to perform non-metric multidimensional scaling (NMDS) ordination using Bray-Curtis distances. Data were square-root transformed and standardized using Wisconsin double standardization. To determine which taxa contributed the most to Bray-Curtis dissimilarity of taxa contributing to protein expression between the treatments, we did a similarity percentages test using the simper function in the ‘vegan’ package.

Results

From 243 Mb of DNA sequence information, roughly 54% of 567,549 filtered reads (median read length=482 bp) were assembled into 26,713 contigs ranging from 500 to 43,200 bp (N50=1135) (Appendix 1: Fig. S2b, Appendix 1: Fig. S2c). All the contigs passed MG-RAST quality control. The metagenome was dominated by bacteria (99.11%) at the domain level. The top five bacterial classes were Betaproteobacteria (31.99%), Alphaproteobacteria (19.42%), Sphingobacteria (13.32%), Gammaproteobacteria...
Burkholderia (8.87%), Variovorax (6.50%), Pedobacter (5.24%), Mucilaginibacter (4.04%) and Lutiella (3.91%). Within the metagenome, 23% of aligned contigs were mapped to the order Burkholderiales while only 7% mapped to Neisserialies (Appendix 1: Fig. S3). Taxonomic evenness of the metagenome, calculated using Hurlbert’s PIE, was equal to 0.79.

Representation of the contigs mapping to functional pathways was dominated by amino acid metabolism (20.6%), followed by membrane transport (12.9%), carbohydrate metabolism (11.9%), translation (7.2%), and metabolism of cofactors and vitamins (6.4%). Within amino acid metabolism, pathways were represented primarily by glycine, serine, and threonine metabolism (17.1%), alanine, aspartate, and glutamate metabolism (13.8%), and valine, leucine, and isoleucine degradation (12.7%). Membrane transport was represented by ABC transporters (78.2%), bacterial secretion system (19.4%), and phosphotransferase system (PTS) (2.4%). Carbohydrate metabolism was dominated by pyruvate metabolism (13.9%), glycolysis/glucogenes (12.6%), and pentose phosphate pathway (11.6%). Overall, the top 5 level 3 KEGG categories included ABC transporters (10.1%), two-component system (4.8%), aminoacyl-tRNA biosynthesis (3.8%), glycine, serine, and threonine metabolism (3.5%), and ribosome (3.3%) (Appendix 1: Fig. S4).

We identified a total of 986 proteins in the enriched treatment and 616 proteins in the control treatment. Of the 220 most abundant protein identifications for each treatment, 65 were shared between treatments leaving 155 unique to each treatment (Fig 2a). The randomization test revealed significantly fewer protein hits shared between the treatments than expected by chance (Fig 2b). In both treatments, the top three of the 20
most abundant proteins, as measured by the total number of matched peptides (spectral counts), were the same in the control and enriched treatments. However, the relative abundances of the remaining 17 proteins in this top list differed strongly between treatments, with only seven of the 20 proteins unique to each treatment (Fig. 2c).

The majority of identified proteins were associated with bacteria. The most common microbial class contributing to identified proteins in both treatments was Betaproteobacteria, but the contribution was higher in enriched (84.4%) versus control (50.3%) treatments (Table 1, Fig. 3a, Appendix 1: Fig. S5, Appendix 1: Fig. S6). This difference was driven by a higher abundance of Alphaproteobacteria in multiple families, including Sphingobacteriaceae, Phyllobacteriaceae, Xanthomonadaceae, and Rhizobiaceae, in control ecosystems relative to the enriched ecosystems. The similarity percentages test identified Betaproteobacteria (38.8%) and Alphaproteobacteria (9.9%) as the main contributors to dissimilarity of active microbial class composition between treatments and Neisseriaceae (23.8%) and Comamonadaceae (9.7%) as the main contributors to active microbial family dissimilarity between treatments. Although both treatments yielded similar numbers of identified microbial classes (control = 12, enriched = 11), taxonomic evenness of microbial classes contributing to identified proteins was substantially higher in the controls ($\text{PIE} = 0.71$) than in the enriched pitchers ($\text{PIE} = 0.31$). Similar taxonomic profiles were obtained using Unipept’s search for the least common taxonomic ancestors of the pooled data (Appendix 1: Fig. S6). For the unpooled data, taxonomic and functional variability among treatments was greater than variability among replicate ecosystems within treatments (Fig. 3, Fig. 4). Multivariate analysis of group dispersion revealed that composition of microbes contributing to protein
expression was significantly more variable in control replicates than in enriched replicates at both the family \((P = 0.003)\) and class \((P = 0.023)\) levels.

The BLAST search yielded taxonomic assignments for 191 and 173 of the 220 sequences in enriched and control treatments, respectively, and all E-values were less than \(10^{-5}\). Of top species hits identified in the BLAST search, *Variovorax paradoxus* and *Chromobacterium violaceum* were the only two of the most six abundant “species” contributing to identified proteins common to both treatments. *Novosphingobium aromaticivorans*, *Starkeya novella*, *Sphingomonas wittichii*, and *Sphingomonas sp.* were among the six most abundant contributors in control pitchers. *Pseudogulbenkiania sp.*, *Rhodanobacter denitrificans*, *Janthinobacterium sp.*, and *Dechlorosoma suillum* were among the six most abundant contributors in enriched pitchers (Appendix 1: Table S2).

Obligate aerobic bacteria contributed the most to identified proteins in the control pitchers, while facultative anaerobic bacteria contributed the most in enriched pitchers (Fig. 5b).

Functional pathways represented by the top 220 expressed microbial proteins also differed between control and enriched pitchers. We detected significant differences in metabolic pathways, including those involved in the metabolism of amino acids, carbohydrates, lipids, secondary metabolites, cofactors & vitamins, and terpenoids & polyketides (Appendix 1: Table S1, Appendix 1: Fig. S7, Appendix 1: Fig. S8a) and, at courser pathway levels, energy metabolism, nucleotide metabolism and amino acid metabolism (Figure 5a). In the control treatment, 161 of the top 220 protein hits were not assigned to a KEGG pathway (represented by 906 total peptides). Of the 220 top protein
hits in the enriched treatment, 129 were not assigned to a pathway (represented by 2,375
total peptides).

Discussion

We hypothesized that there would be detectable differences in the taxonomic
composition of microbes contributing to expressed proteins. Indeed, we observed striking
differences between unenriched and enriched ecosystems in the taxonomic composition
of the microbes contributing to identified proteins (Fig. 3). The taxonomic composition of
bacteria contributing to protein expression in our study, and in our metagenome, is
consistent with findings of previous studies of bacterial communities in Sarracenia
species. S. purpurea pitchers contain more than 1,000 species of bacteria and a negligible
amount of archaea (Paisie et. al, 2014). One genomic study of S. alata pitcher bacterial
communities revealed an abundance of Proteobacteria (primarily Gammaproteobacteria).
Taxonomic groups within the Betaproteobacteria had relative abundances similar to our
metagenome and to control pitcher communities in our experiment, with a high
percentage of sequences derived from Burkholderiales and a lower proportion from the
Neisseriales (Koopman et al. 2010). A study of sub-habitats in S. purpurea revealed an
abundance of Betaproteobacteria (primarily Burkholderiales) on the pitcher walls and in
the sediment, co-dominance in pitcher liquid by Beta- and Alphaproteobacteria, and the
presence of Bacteroidetes and Firmicutes, though in a low proportion, in the sediment,
fluid, and pitcher walls (Krieger and Kourtev 2012). This finding is fairly consistent with
the taxonomic potential revealed by our metagenome, in which 35%, 23%, 14%, and 1%
of identified contigs were mapped to Betaproteobacteria, Alphaproteobacteria,
Bacteroidetes, and Firmicutes, respectively. Grey et al. (2012) found that *S. purpurea* pitchers were composed primarily of Proteobacteria and Bacteroidetes, with Gammaproteobacteria, Alphaproteobacteria, or Betaproteobacteria dominating within the Proteobacteria, but that taxonomic composition varied from pitcher to pitcher within and across geographic regions.

The composition of bacteria contributing to protein expression in our experiment varied between control replicates, much more so than between enriched pitcher communities. This pattern is likely the result of a combination of factors. First, pitchers contain distinct sub-habitats that vary in light availability and concentration of dissolved oxygen and organic matter and therefore provide multiple habitats for a diverse set of microbes (Krieger and Kourtev 2012). As organic matter enrichment increases biological oxygen demand, the subsequent decline in dissolved oxygen may create a more homogenous oxygen environment such that microbes sensitive to oxygen conditions can no longer compete against low-oxygen tolerant bacteria, decreasing bacterial diversity.

Low bacterial diversity in enriched pitchers echoes findings in larger enriched aquatic ecosystems. Analysis of the 16S rRNA gene product of bacterial communities in nutrient-enriched salt marsh sediments revealed that the bacterial diversity of active bacteria decreased relative to that of communities in unenriched sediments (Kearns et al. 2016). Similarly, enrichment of heterotrophic stream biofilm communities yielded lowered diversity; however, in contrast to our enriched pitcher communities, the stream biofilm communities diverged in composition (Van Horn et al. 2011).

The composition of microbes contributing to protein expression in *S. purpurea* pitchers was similar to the composition of larger freshwater aquatic ecosystems.
Betaproteobacteria dominated microbes contributing to protein expression in both enriched and control pitchers though in higher abundances in enriched pitchers relative to control pitchers. Betaproteobacteria are generally the most abundant class of bacteria in freshwater lakes (Percent et al. 2008, Newton et al. 2011) and dominate contaminated sediments (Haller et al. 2011) and organic aggregates in eutrophic lakes (Tang et al. 2009). Betaproteobacteria populations associated with the beta II clade have been shown to increase rapidly with the addition of organic carbon in humic lakes (Burkert et al. 2003, Kent et al. 2006). Furthermore, experimental dissolved organic matter additions to microcosms containing alpine lake bacteria cultures led to a near-dominance of Betaproteobacteria, suggesting that these bacteria are good competitors in enriched aquatic ecosystems (Perez and Sommaruga 2006). These results suggest that bacterial communities in *S. purpurea* pitchers are structured and behave like bacterial communities in larger lakes and ponds in response to enrichment. It is important to note that most existing literature on freshwater bacteria and *S. purpurea* bacterial communities rely primarily on genomic methods for identification and therefore are likely capturing functionally active and inactive bacteria, whereas our methods are capturing only the functionally active bacteria. As a result, we use caution when directly comparing the results of our study to those in larger aquatic ecosystems. However, the Unipept search of our identified tryptic peptides and NCBI Blast search of their contigs of origin yielded remarkably similar results (Fig 3a, Appendix 1: Fig. S6), suggesting that tryptic peptides could be used to correctly identify microbes contributing to identified proteins, though at coarser taxonomic levels than can be achieved by nucleic acid analysis.
We hypothesized that there would be detectable differences in the function of microbial communities in control and enriched pitchers. We measured function in two ways: first, we mapped identified bacterial classes associated with proteins to their oxygen requirements and second, we mapped peptides to functional KEGG pathways. Oxygen requirements differed significantly between taxa contributing to protein expression in control and enriched microbial communities. Bacteria contributing to protein expression in control pitchers were predominately aerobic whereas bacteria contributing to protein expression in enriched pitchers were primarily facultatively anaerobic. The difference in oxygen requirement of contributing bacteria between the two treatments was driven largely by two taxa: the obligate aerobe *Variovorax paradoxus* (28.4% of total peptides in the control treatment and 7.2% in the enriched treatment) and the facultative anaerobe *Chromobacterium violaceum* (53.3% of total peptides in the enriched treatment and 6.6% in the control treatment) (Appendix 1: Table S2). Peptides that mapped to *C. violaceum* in the BLAST search mapped in the Unipept search to *Aquitalea magnusonii*, a betaproteobacteria most closely related to *C. violaceum*, isolated from a humic lake in Wisconsin, USA (Lau et al. 2006). Although we did not measure dissolved oxygen during the field experiment, enriched pitchers in a subsequent experiment enriched with the same concentration of organic matter became hypoxic within 48 hours, suggesting that pitchers in the field were likely hypoxic (Sirota et al. 2013). Dissolved oxygen concentration is one of three primary drivers of bacterial community composition in eutrophic, dimictic lakes (Shade et al. 2007) and appears to also drive the composition of functionally active bacteria in enriched *S. purpurea* pitchers.
We expected to see a high proportion of obligate anaerobic bacteria in enriched pitchers. Bacteroidetes and Firmicutes, to a lesser degree, have been found to inhabit *S. purpurea* pitchers (Krieger and Kourtev 2012); however, we identified very few proteins associated with these taxa. Of the 3008 and 969 peptides associated with the top 220 proteins in enriched and control treatments, respectively, we found only 17 peptides associated with obligate anaerobes in the enriched pitchers (7 of which were associated with Firmicutes) and 13 associated with obligate anaerobes in the control pitchers (3 of which were associated with Firmicutes). Though we did find a higher number of peptides associated with Bacteroidetes (74 peptides in control pitchers and 89 in enriched pitchers), they were facultative anaerobes and not strict anaerobes. It is likely that the low numbers of identified peptides associated with these taxa in experimental pitchers are the result of a skewed protein database. Our database was built using metagenomic data from pitchers in the field, the majority of which are oxygen-rich (Adlassnig et al. 2011), and likely contained nucleotide sequences primarily from aerobic and facultative anaerobic bacteria. Additionally, pitchers are generally oxygen-rich due to photosynthetic activity of the plant and therefore primarily harbor aerobic inquilines (Adlassnig et al. 2011). Even when dissolved oxygen is low, there is a constant flux of oxygen into the pitcher fluid and so the pitchers are rarely ever truly anoxic. It is not surprising, therefore, that peptides associated with anaerobic bacteria were rare. In the absence of a fully representative database, we feel that the higher number of proteins represented by facultative bacteria in enriched pitchers relative to control pitchers is a good indicator of changing oxygen conditions. These results are consistent with the shift to a hypoxic state when *S. purpurea* is enriched with additional prey (Sirota et al. 2013).
We assigned KEGG pathways to contigs in the metagenome and to protein identifications in the metaproteomes to compare microbial community function between control and enriched pitchers, and between the metaproteomes and functional potential in the metagenome. Not surprisingly, the functional potential revealed by the metagenome differed from function revealed by the metaproteomes. Amino acid metabolism and carbohydrate metabolism were represented in the top five rank-ordered pathways in both the metaproteomes and the metagenome; however, carbohydrate metabolism was ranked first in the metaproteomes (~34-40% of total peptides) and third in the metagenome (~12% of mapped contigs). Nucleotide metabolism and energy metabolism were represented in the top five in the metaproteomes (~18% of total peptides in controls and ~34% of total peptides in enriched pitchers), but were ranked 9th (~4%) and 7th (~5%) in the metagenome, respectively. Such differences could be a result of not all nucleotide sequences being transcribed and translated to proteins, but may also be an artifact of only including 220 proteins from each treatment in the metaproteome analysis.

We hesitate to hypothesize broader relevance of our functional pathway results for two reasons. First, we are most interested in the identification of proteins that can serve as biomarkers of aquatic ecosystem state changes. Whereas we expect that functional information will be useful for determining the utility and generality of such biomarkers, it is not necessary for finding useful biomarkers. Second, it seems impossible, with our limited data, to identify a complete set of functions. With that caveat, we found that coarse KEGG pathway assignments differed between control and enriched microbial communities. Enriched pitchers contained significantly more microbial biomass, as evidenced by the size of the microbial pellets post-centrifugation.
When samples were pooled and total peptide counts were normalized, chi-square analysis revealed an enrichment of peptides associated with energy metabolism in enriched pitchers.

These results are consistent with patterns seen in larger aquatic ecosystems: mineralization of organic matter, an effect of microbial energy metabolism, has been shown to increase along trophic gradients, with bacteria contributing most to mineralization in eutrophic freshwater lakes (Simcic 2005). Not surprisingly, peptides associated with processes requiring oxygen including oxidative phosphorylation and the citric acid cycle were enriched in oxygen-rich control pitcher microbial communities. One protein associated with the citric acid cycle, isocitrate lyase, was present in the top 20 rank ordered protein identifications in the enriched treatment, but not in the control treatment. This protein, which has been found to be upregulated during periods of oxygen depletion in *M. tuberculosis* (Wayne and Lin 1982), could be a candidate biomarker for an impeding tipping point in the *S. purpurea* microecosystem. Though we did not find a significant difference in lipid metabolism pathways between control and enriched pitcher proteins, there was a trend for increased pathway representation of unsaturated fatty acid biosynthesis and fatty acid elongation in enriched pitchers. Such an increase has been found in bacteria in low-oxygen or anaerobic conditions, primarily resulting from an increase in membrane lipids (Lemmer et al. 2015). While these differences do not immediately reveal a functional explanation, it is promising that there were signatures of detectable differences in the protein profiles between treatments. Such differences imply that there are changes in the expression of the most abundant proteins in the most abundant taxa related to organic matter loading.
In larger aquatic systems, traditional water quality indicators may not provide enough lead-time to forecast a tipping point (Contamin and Ellison 2009), especially if they lag behind changes in the microbial community. We hypothesize that microbial proteins may be more sensitive and timely indicators of impending tipping points than traditional chemical markers of water quality. We argue that even though metatranscriptomic and metagenomic methods have superior throughput, metaproteomic methods can inexpensively and rapidly simultaneously characterize the function and (indirectly) composition of the active microbial community members responsible for processes related to aquatic ecosystem state changes. Our study includes a semi-quantitative small initial sampling at only a single time point and therefore does not yet enable a comprehensive enough proteomic analysis to determine the identity of biomarkers or place them in an ecological context. Future studies using more sensitive instrumentation will allow for the identification of a larger number of proteins. Time series of environmental proteomics data and quantitative analysis of changes in protein abundances prior to state changes will allow for the identification and ecological characterization of tipping point biomarkers.

Acknowledgments

This work was funded by the National Science Foundation (grant numbers 1144055 and 1144056). Proteomic analysis was funded by the Vermont Genetics Network through U.S. National Institutes of Health Grant 8P20GM103449 from the INBRE program of the NIGMS. The authors thank Hailee Tenander for assisting with preparation of samples for mass spectrometry analysis.
This Whole Genome Shotgun project has been deposited at DDB/ENA/GenBank under the accession NMRC01000000. The version described in this paper is version NMRC01000000. The protein database and all code used to analyze the data is freely available on the Harvard Forest Data Archive under ID number HF295.
References

Adlassnig, W., M. Peroutka, and T. Lendi. 2011. Traps of carnivorous pitcher plants as a habitat: composition of the fluid, biodiversity and mutualistic activities. Annals of Botany 107:181-194.

Benjamini, Y., and Y. Hochberg. 1995. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B-Methodological 57:289-300.

Benndorf, D., G. U. Balcke, H. Harms, and M. von Bergen. 2007. Functional metaproteome analysis of protein extracts from contaminated soil and groundwater. Isme Journal 1:224-234.

Bestelmeyer, B. T., A. M. Ellison, W. R. Fraser, K. B. Gorman, S. J. Holbrook, C. M. Laney, M. D. Ohman, D. P. C. Peters, F. C. Pillsbury, A. Rassweiler, R. J. Schmitt, and S. Sharma. 2011. Analysis of abrupt transitions in ecological systems. Ecosphere 2.

Biggs, R., S. R. Carpenter, and W. A. Brock. 2009. Turning back from the brink: Detecting an impending regime shift in time to avert it. Proceedings of the National Academy of Sciences of the United States of America 106:826-831.

Burkert, U., F. Warnecke, D. Babenzien, E. Zwirnmann, and J. Pernthaler. 2003. Members of a readily enriched beta-proteobacterial clade are common in surface waters of a humic lake. Applied and Environmental Microbiology 69:6550-6559.
Butler, J. L., N. J. Gotelli, and A. M. Ellison. 2008. Linking the brown and green: Nutrient transformation and fate in the Sarracenia microecosystem. Ecology 89:898-904.

Campos, A., S. Tedesco, V. Vasconcelos, and S. Cristobal. 2012. Proteomic research in bivalves Towards the identification of molecular markers of aquatic pollution. Journal of Proteomics 75:4346-4359.

Carpenter, S. R., W. A. Brock, J. J. Cole, J. F. Kitchell, and M. L. Pace. 2008. Leading indicators of trophic cascades. Ecology Letters 11:128-138.

Cheerathodi, M., and B. A. Ballif. 2011. Identification of CrkL-SH3 Binding Proteins from Embryonic Murine Brain: Implications for Reelin Signaling during Brain Development. Journal of Proteome Research 10:4453-4462.

Chrost, R. J., and W. Siuda. 2006. Microbial production, utilization, and enzymatic degradation of organic matter in the upper trophogenic layer in the pelagial zone of lakes along a eutrophication gradient. Limnology and Oceanography 51:749-762.

Colatriano, D., A. Ramachandran, E. Yergeau, R. Maranger, Y. Gélinas, and D. A. Walsh. 2015. Metaproteomics of aquatic microbial communities in a deep and stratified estuary. Proteomics 15:3566-3579.

Conesa, A., S. Gotz, J. M. Garcia-Gomez, J. Terol, M. Talon, and M. Robles. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674-3676.
Contamin, R., and A. M. Ellison. 2009. Indicators of regime shifts in ecological systems: What do we need to know and when do we need to know it? Ecological Applications 19:799-816.

Correll, D. L. 1998. The role of phosphorus in the eutrophication of receiving waters: a review. Journal of Environmental Quality 27:261-266.

Dakos, V., S. R. Carpenter, E. H. van Nes, and M. Scheffer. 2015. Resilience indicators: prospects and limitations for early warnings of regime shifts. Philosophical Transactions of the Royal Society B-Biological Sciences 370.

Dennison, W. C., R. J. Orth, K. A. Moore, J. C. Stevenson, V. Carter, S. Kollar, P. W. Bergstrom, and R. A. Batiuk. 1993. Assessing water-quality with submersed aquatic vegetation. Bioscience 43:86-94.

Elias, J. E., and S. P. Gygi. 2007. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nature Methods 4:207-214.

Ellison, A. M., N. J. Gotelli, J. S. Brewer, D. L. Cochran-Stafira, J. M. Kneitel, T. E. Miller, A. C. Worley, and R. Zamora. 2003. The evolutionary ecology of carnivorous plants. Advances in Ecological Research, Vol 33:1-74.

Farnsworth, E. J., and A. M. Ellison. 2008. Prey availability directly affects physiology, growth, nutrient allocation and scaling relationships among leaf traits in 10 carnivorous plant species. Journal of Ecology 96:213-221.

Gray, S. M., D. M. Akob, S. J. Green, and J. E. Kostka. 2012. The Bacterial Composition within the Sarracenia purpurea Model System: Local Scale Differences and the Relationship with the Other Members of the Food Web. Plos One 7.
Haller, L., M. Tonolla, J. Zopfi, R. Peduzzi, W. Wildi, and J. Pote. 2011. Composition of bacterial and archaeal communities in freshwater sediments with different contamination levels (Lake Geneva, Switzerland). Water Research 45:1213-1228.

Hargeby, A., I. Blindow, and G. Andersson. 2007. Long-term patterns of shifts between clear and turbid states in Lake Krankesjon and Lake Takern. Ecosystems 10:29-36.

Hurlbert, S. H. 1971. The nonconcept of species diversity: A critique and alternative parameters. Ecology 52:577-586.

Jayapal, K. P., R. J. Philp, Y. J. Kok, M. G. S. Yap, D. H. Sherman, T. J. Griffin, and W. S. Hu. 2008. Uncovering Genes with Divergent mRNA-Protein Dynamics in Streptomyces coelicolor. Plos One 3.

Kanehisa, M., S. Goto, Y. Sato, M. Kawashima, M. Furumichi, and M. Tanabe. 2014. Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Research 42:D199-D205.

Kearns, P. J., J. H. Angell, E. M. Howard, L. A. Deegan, R. H. R. Stanley, and J. L. Bowen. 2016. Nutrient enrichment induces dormancy and decreases diversity of active bacteria in salt marsh sediments. Nature Communications 7.

Kent, A. D., S. E. Jones, G. H. Lauster, J. M. Graham, R. J. Newton, and K. D. McMahon. 2006. Experimental manipulations of microbial food web interactions in a humic lake: shifting biological drivers of bacterial community structure. Environmental Microbiology 8:1448-1459.

Koopman, M. M., and B. C. Carstens. 2011. The microbial phyllogeography of the carnivorous plant Sarracenia alata. Microbial Ecology 61:750-758.
Koopman, M. M., D. M. Fuselier, S. Hird, and B. C. Carstens. 2010. The Carnivorous Pale Pitcher Plant Harbors Diverse, Distinct, and Time-Dependent Bacterial Communities. Applied and Environmental Microbiology 76:1851-1860.

Krieger, J. R., and P. S. Kourtev. 2012. Bacterial diversity in three distinct sub-habitats within the pitchers of the northern pitcher plant, Sarracenia purpurea. Fems Microbiology Ecology 79:555-567.

Lau, H. T., J. Faryna, and E. W. Triplett. 2006. Aquitalea magnusonii gen. nov., sp nov., a novel Gram-negative bacterium isolated from a humic lake. International Journal of Systematic and Evolutionary Microbiology 56:867-871.

Lauro, F. M., M. Z. DeMaere, S. Yau, M. V. Brown, C. Ng, D. Wilkins, M. J. Raftery, J. A. E. Gibson, C. Andrews-Pfannkoch, M. Lewis, J. M. Hoffman, T. Thomas, and R. Cavicchioli. 2011. An integrative study of a meromictic lake ecosystem in Antarctica. Isme Journal 5:879-895.

Lemmer, K. C., A. C. Dohnalkova, D. R. Noguera, and T. J. Donohue. 2015. Oxygen-Dependent Regulation of Bacterial Lipid Production. Journal of Bacteriology 197:1649-1658.

Levin, P. S., and C. Mollmann. 2015. Marine ecosystem regime shifts: challenges and opportunities for ecosystem-based management. Philosophical Transactions of the Royal Society B-Biological Sciences 370.

Mesuere, B., Williams, T., Van der Jeugt, F., Devreese, B., Vandamme, P., Dawyndt, P. 2016. Unipept web services for metaproteomics analysis. Bioinformatics.

Meyer, F. Paarmann, D., D’Souza, M., Olson, R., Glass, E.M., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilke, A., Wilkening, J., Edwards, R.A. 2008. The
Metagenomics RAST Server – a Public Resource for the Automatic Phylogenetic and Functional Analysis of Metagenomes. BMC Bioinformatics 9: 386.

Newton, R. J., S. E. Jones, A. Eiler, K. D. McMahon, and S. Bertilsson. 2011. A Guide to the Natural History of Freshwater Lake Bacteria. Microbiology and Molecular Biology Reviews 75:14-49.

Oksanen, J., F. Blanchet, R. Kindt, R. Legendre, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. Henry, H. Stevens, E. Szöcs, and H. Wagner. 2016. Vegan: community ecology package, R Package Version 2.4-1 edn. Oksanen J, Blanchet FG, Kindt R, Legendre R, Minchin PR, O'Hara RB et al. (2012). Vegan: community ecology package, R Package Version 2.1-17 edn.

Ondov, B. D., N. H. Bergman, and A. M. Phillippy. 2011. Interactive metagenomic visualization in a Web browser. Bmc Bioinformatics 12.

Pan, Y. D., R. J. Stevenson, B. H. Hill, A. T. Herlihy, and G. B. Collins. 1996. Using diatoms as indicators of ecological conditions in lotic systems: A regional assessment. Journal of the North American Benthological Society 15:481-495.

Percent, S. F., M. E. Frischer, P. A. Vescio, E. B. Duffyy, V. Milano, M. McLellan, B. M. Stevens, C. W. Boylen, and S. A. Nierzwicki-Bauer. 2008. Bacterial community structure of acid-impacted lakes: What controls diversity? Applied and Environmental Microbiology 74:1856-1868.

Perez, M. T., and R. Sommaruga. 2006. Differential effect of algal- and soil-derived dissolved organic matter on alpine lake bacterial community composition and activity. Limnology and Oceanography 51:2527-2537.
Peterson, C. N., S. Day, B. E. Wolfe, A. M. Ellison, R. Kolter, and A. Pringle. 2008. A keystone predator controls bacterial diversity in the pitcher-plant (Sarracenia purpurea) microecosystem. Environmental Microbiology 10:2257-2266.

Rabalais, N. N., R. E. Turner, R. J. Diaz, and D. Justic. 2009. Global change and eutrophication of coastal waters. Ices Journal of Marine Science 66:1528-1537.

Reddy, T. B. K., A. D. Thomas, D. Stamatis, J. Bertsch, M. Isbandi, J. Jansson, J. Mallajosyula, I. Pagani, E. A. Lobos, and N. C. Kyrpides. 2015. The Genomes OnLine Database (GOLD) v.5: a metadata management system based on a four level (meta)genome project classification. Nucleic Acids Research 43:D1099-D1106.

Scheffer, M. 2009. Critical Transitions in Nature and Society. Princeton University Press.

Scheffer, M., J. Bascompte, W. A. Brock, V. Brovkin, S. R. Carpenter, V. Dakos, H. Held, E. H. van Nes, M. Rietkerk, and G. Sugihara. 2009. Early-warning signals for critical transitions. Nature 461:53-59.

Shade, A., A. D. Kent, S. E. Jones, R. J. Newton, E. W. Triplett, and K. D. McMahon. 2007. Interannual dynamics and phenology of bacterial communities in a eutrophic lake. Limnology and Oceanography 52:487-494.

Simcic, T. 2005. The role of plankton, zoobenthos, and sediment in organic matter degradation in oligotrophic and eutrophic mountain lakes. Hydrobiologia 532:69-79.

Sirota, J., B. Baiser, N. J. Gotelli, and A. M. Ellison. 2013. Organic-matter loading determines regime shifts and alternative states in an aquatic ecosystem.
Sondergaard, M., L. S. Johansson, T. L. Lauridsen, T. B. Jorgensen, L. Liboriussen, and E. Jeppesen. 2010. Submerged macrophytes as indicators of the ecological quality of lakes. Freshwater Biology 55:893-908.

Sowell, S. M., P. E. Abraham, M. Shah, N. C. Verberkmoes, D. P. Smith, D. F. Barofsky, and S. J. Giovannoni. 2011. Environmental proteomics of microbial plankton in a highly productive coastal upwelling system. Isme Journal 5:856-865.

Srivastava, D. S., J. Kolasa, J. Bengtsson, A. Gonzalez, S. P. Lawler, T. E. Miller, P. Munguia, T. Romanuk, D. C. Schneider, and M. K. Trzcinski. 2004. Are natural microcosms useful model systems for ecology? Trends in Ecology & Evolution 19:379-384.

Tang, X. M., G. Gao, B. Q. Qin, L. P. Zhu, J. Y. Chao, J. J. Wang, and G. J. Yang. 2009. Characterization of bacterial communities associated with organic aggregates in a large, shallow, eutrophic freshwater lake (Lake Taihu, China). Microbial Ecology 58:307-322.

Timinskas, K., M. Balvociute, A. Timinskas, and C. Venclovas. 2014. Comprehensive analysis of DNA polymerase III alpha subunits and their homologs in bacterial genomes. Nucleic Acids Research 42:1393-1413.

Ullrich, N., P. Casper, A. Otto, and M. O. Gessner. 2016. Proteomic evidence of methanotrophy in methane-enriched hypolimnetic lake water. Limnology and Oceanography 61:S91-S100.
Van Horn, D. J., R. L. Sinsabaugh, C. D. Takacs-Vesbach, K. R. Mitchell, and C. N. Dahm. 2011. Response of heterotrophic stream biofilm communities to a gradient of resources. Aquatic Microbial Ecology 64:149-161.

Vogel, C., and E. M. Marcotte. 2012. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nature Reviews Genetics 13:227-232.

Wayne, L.G., and K. Lin. 1982. Glyoxylate Metabolism and Adaptation of *Mycobacterium tuberculosis* to Survival under Aerobic Conditions. Infection and Immunity 37:1042-1049.
Table 1. Results of chi-square analysis of bacterial classes in control and enriched pitchers. Bolded values represent those in which the adjusted $P$ value is <0.05.

| Class              | Control Peptides | Enriched Peptides | Adjusted chi-square |
|--------------------|------------------|-------------------|---------------------|
| Acidobacteria      | 6                | 0                 | 0.000               |
| Actinobacteria     | 32               | 3                 | 0.000               |
| Alphaproteobacteria| 276              | 196               | 0.000               |
| Bacteroidia        | 12               | 16                | 0.059               |
| Betaproteobacteria | 469              | 2448              | 0.000               |
| Chloroflexi        | 0                | 3                 | 0.816               |
| Clostridia         | 3                | 7                 | 0.959               |
| Cytophagia         | 14               | 17                | 0.021               |
| Deltaproteobacteria| 2                | 0                 | 0.132               |
| Flavobacteriia     | 0                | 3                 | 0.816               |
| Gammaproteobacteria| 50               | 146               | 0.816               |
| Gloeobacteria      | 8                | 0                 | 0.000               |
| Sphingobacteriia   | 48               | 53                | 0.000               |
| Spirochaetia       | 11               | 9                 | 0.006               |
Table 2. Effect of treatment on microbial proteins, contributing taxa (class and family), and pathways. Bolded values represent those in which the adjusted $P$ value is <0.05.

| Proteins | Taxa (Class) | Taxa (Family) | KEGG Pathways |
|----------|-------------|---------------|---------------|
| df | $F$ | $R^2$ | $P$ | df | $F$ | $R^2$ | $P$ | df | $F$ | $R^2$ | $P$ |
| Treatment | 1 | 4.217 | 0.319 | **0.004** | 1 | 3.766 | 0.295 | **0.022** | 1 | 4.218 | 0.319 | **0.003** | 1 | 4.753 | 0.373 | **0.024** |
| Residuals | 9 | 0.681 | 9 | 0.705 | 9 | 0.681 | 9 | 0.627 |
| Total | 10 | 1.000 | 10 | 1.000 | 10 | 1.000 | 10 | 1.000 |
Fig. 1. Pipeline for data collection and analysis. Proteins from the microbial communities in experimentally enriched and ambient control pitcher fluid were processed using SDS-PAGE, tryptic digest, LC-MS/MS, and a SEQUEST search of a custom metagenomic database. The composition of microbial communities was determined using a BLAST homology search of metagenomic data associated with identified proteins. Protein identity and annotation was determined via a blastp search to identify orthologs and blast2go.
Fig. 2. Protein identifications differed between control and enriched pitchers. (a) Protein hits shared between control and enriched treatments. (b) Results of a randomization test in which 220 protein hits were randomly assigned to each treatment and the number of shared protein hits was calculated. Red line indicates the actual shared number of proteins. Grey probability density function indicates the 95% confidence interval for the simulated shared protein hit values. (c) Top 20 proteins in rank order for each treatment. Proteins are ranked by the number of total peptides associated with them (in parentheses). Identical proteins in both treatments are connected by lines. Blue lines indicate proteins unique to the top 20 in control pitchers (C) and brown lines indicate proteins unique to the top 20 in enriched pitchers (E).
Fig. 3. Distinctly different microbial communities contributed to protein expression in control and enriched pitchers. The proportion of total peptides from the top 220 proteins associated with particular microbial classes (a) and families (b) in all enriched and control replicate pitchers.
Fig. 4. Microbial communities in control and enriched pitchers differ in the proteins they produce, taxa that contribute to protein expression, and function. Ordination of Bray-Curtis dissimilarities of total peptides shows clustering of pitcher microbial communities by treatment for protein hits (adonis $P=0.004$), microbial classes (adonis $P=0.022$), microbial families (adonis $P=0.003$) and KEGG pathways (adonis $P=0.003$) as a function of treatment (control or enriched)
Fig. 5. Microbial function differed between control and enriched pitchers. (a) Heat map of the proportional representation of course-level KEGG pathways between control pitchers (C) and enriched pitchers (E) and individual control (H4, H6, E3, C4) and enriched (H1, H2, H3, 2C, 5B, 5A) replicates. Significantly different pathways between pooled control and enriched samples are indicated with “*”. (b) Oxygen requirement of microbial classes contributing to protein expression as a proportion of all peptides in control (C) and enriched (E) pitchers.
Appendix S1

Detailed Methods

Field Experiment

Starting June 10, 2011, we selected newly opened, and therefore sterile (Peterson et al., 2008), S. purpurea pitchers for five days until 20 pitchers were selected. One pitcher from each group was randomly assigned to one of two treatments—ambient control and detritus-enriched. The average pitcher length, measured from the base of the pitcher along the back of the keel to the top of the hood, was 12.4±2.3 cm. Pitcher volume was not measured during the experiment. The final average volume of fluid in the pitchers was 9.6±5.8 mL.

After the first rain, initial samples of 1.5 ml of pitcher fluid were drawn from all pitchers and replaced with 1.5 ml of deionized water. In the detrital enrichment treatment, each pitcher received 1mg of detritus per day between 7:00 am and 9:00 am. Wasps were ground in a coffee grinder, dried for 48-72 hours in an oven at 70 °C, weighed, and stored in a −20 °C freezer until used.

Sampling

Initial 1.5-ml samples and the entire final contents of each pitcher were drawn independently through the frit of separate Bio-Rad (Hercules, California, USA) Poly-Prep chromatography columns to remove any organisms larger than 30 microns. For each sample, the filtrate was centrifuged in 2ml aliquots at 13,000g to concentrate the microbial assemblage and the supernatant was removed. The resulting microbial pellet was stored at −80 °C. All frozen samples were transported on dry ice from Harvard
Forest to the University of Vermont (June 29, 2011), where they were stored at −80 °C until processed.

**Metagenome Extraction and Sequencing**

We used the DNeasy Blood and Tissue kit (Qiagen) to extract DNA from the microbial pellets of three pitchers using the Purification of Total DNA from Animal Tissues Spin-Column Protocol (pages 28-30 of the handbook dated 07/2006). Samples were pre-treated with proteinase K (as described on page 45 of the booklet). For each pitcher, one pellet was also pre-treated with lysozyme during the extraction. Five percent of genomic DNA preparation was loaded on a 1% agarose gel (Appendix 1: Fig. S2a). Samples from all six preparations were pooled and the DNA was precipitated with 0.1 volume 3M sodium acetate and two volumes of absolute ethanol. The pooled samples were then centrifuged and the precipitated DNA was washed with 75% ethanol and then resuspended in water. More than 10 µg of total DNA was sent for library construction, sequencing and assembly to Genome Quebéc (Montréal, QC, Canada) using the 454 GS-FLX Titanium Sequencing System (Roche).

**Protein Extraction, SDS-Page, and Mass Spectrometry**

Microbial pellets were resuspended in 100 µl of bromophenol blue sample buffer (150mM Tris pH 6.8, 2% SDS, 5% β-mercaptoethanol, 7.8% glycerol) and boiled at 95 °C for five minutes. All samples were diluted proportional to their pellet size to obtain similar staining levels. After centrifugation, samples were loaded into separate lanes of a 10% polyacrylamide (37.5:1 acrylamide:bis-acrylamide) gel and subjected to SDS-PAGE and Coomassie staining (Fig. 1, Appendix 1: Fig. S1a, and Appendix 1: Fig. S1b).
All six of the enriched pitchers and five of the six control pitchers had visible protein staining levels and were chosen for mass spectrometry. These 11 sample lanes were each divided into five regions (Appendix 1: Fig. S1b) and each region was diced into 1 mm³ pieces. Gel cubes were rinsed with HPLC-grade water, incubated at 37 °C for 30 minutes in 1 ml of destain solution (50 mM ammonium bicarbonate, 50% acetonitrile), and dehydrated in 100% acetonitrile for 10 min in order to remove the Coomassie stain. This destain procedure was repeated a second time to ensure complete removal of the stain.

An in-gel tryptic digest was performed by submerging the dehydrated gel pieces in ice-cold sequencing-grade modified trypsin (6 ng/µl) (Promega, Fitchburg, WI, USA) for 15 minutes, adding ice-cold 50 mM ammonium bicarbonate solution, letting the gel pieces swell on ice, and then incubating the pieces overnight at 37 °C. Digests were centrifuged at 13,000 g for five minutes and the peptide-containing supernatant transferred to a .6 ml tube. Peptides were further extracted from the gel pieces by adding 100 µl of 50% acetonitrile and 2.5% formic acid, centrifuging for 15 minutes at 13,000 x g, and dehydrating in 100% acetonitrile. All extracted peptides were pooled, dried in a SpeedVac for 1 hour, and stored at -80°C.

**Custom Metagenomic and Protein Databases**

We generated a custom protein database from a six-frame forward and reverse translation of a metagenomic database constructed from microbial communities of three previously collected pitchers that had captured diverse amounts of prey (Appendix 1: Fig. S2). Pitchers were collected from Molly Bog, an ombrotrophic bog located in Morristown, VT.
on August 18, 2008 and transported in a cooler directly from the field to
the University of Vermont. Microbial pellets were obtained immediately as described
above.

We used the DNeasy Blood and Tissue kit (Qiagen) to extract DNA from the
microbial pellets of three pitchers using the Purification of Total DNA from Animal
Tissues Spin-Column Protocol (pages 28-30 of the handbook dated 07/2006). Samples
were pre-treated with proteinase K (as described on page 45 of the booklet). For each
pitcher, one pellet was also pre-treated with lysozyme during the extraction. Five percent
of genomic DNA preparation was loaded on a 1% agarose gel (Appendix 1: Fig. S2a).
Samples from all six preparations were pooled and the DNA was precipitated with 0.1
volume 3M sodium acetate and two volumes of absolute ethanol. The pooled samples
were then centrifuged and the precipitated DNA was washed with 75% ethanol and then
re-suspended in water. More than 10 µg of total DNA was sent for library construction,
sequencing and assembly to Genome Quebéc (Montréal, QC, Canada) using the 454 GS-
FLX Titanium Sequencing System (Roche). From 243 Mb of sequence information,
roughly 54% of 567,549 filtered reads (median read length = 482 bp) were assembled
into 26,713 contigs of length greater than 500 bp (Appendix 1: Fig. S2b, Appendix 1:
Fig. S2c).

A custom metaproteomic database was created from the metagenome database
using open-source Ruby programming software. Each contig was translated to an amino
acid sequence in all six reading frames. Of the resulting amino acid sequences, only
sequences with greater than 100 amino acids in length were retained. Those 184,128
sequences were written to new fasta files and retained their original description line. If
multiple amino acid sequences came from a single contig, the resulting description lines included unique letter identifiers. As such, all amino acid sequences could be mapped back to a single nucleotide sequence greater than 300 bp in length. To create a decoy database, all retained protein sequences were reversed and then concatenated to the forward database. The decoy database allowed for an estimate of the false identification rate during the database search process as has been described (Elias & Gygi, 2007).

**SEQUEST Search Parameters**

The following search parameters were used during the SEQUEST search: peptides were required to be tryptic; peptide precursor mass tolerance was set at plus or minus 2 Da; and differential oxidation of methionine (15.9949 Da) and differential acrylamidation of cysteine (71.0371 Da) were permitted.

**Randomization Test of Shared Proteins**

A pool of identified proteins was generated by combining the total protein hits from the top 220 protein hits in both treatments. We chose to analyze only the top 220 proteins in each group because the identification status of proteins that are rarer is less certain and because including many rare proteins in the test was likely to add noise caused by the sampling of rare elements. With enough noise added from rarity, there is a danger that the real signal of differences among the common proteins will be swamped by this noise. Two hundred twenty protein hits were randomly drawn and assigned to each of the two treatments, without replacement. Each protein hit in the original pool was weighted by sum of the total number of peptides associated with that protein hit in the two treatments.
For each simulation ($N=1000$), the number of shared protein hits between treatments was calculated, yielding a probability distribution of the expected number of shared protein hits. The observed shared number of protein hits was calculated by finding the intersection of the list of top 220 control protein hits and the top 220 enriched protein hits. Whether protein hits were drawn with or without replacement, the number of shared proteins was less than expected by chance supporting the alternative hypothesis that the protein pools from the two treatments are distinct from one another (Fig 2b).

We conducted an additional simulation experiment (programmed in R) to test for the possibility of a Type I error (incorrectly rejecting a true null hypothesis) in our randomization test to determine the expected number of shared proteins between enriched and control pitchers. We first simulated a single source protein pool consisting of 10,000 distinct protein types. Next, we created two sets to represent control and treatment groups. For each group, we sampled with replacement from the protein source pool until we had accumulated enough proteins so that there were exactly 200 proteins represented in each group (typically this necessitated sampling somewhere between 200 and 210 individual proteins because there were occasional duplicates observed). As you would expect, there are usually no proteins shared or only a small number between these two samples.

Next, we followed the procedure that we described in our randomization test. Namely, we reshuffled these proteins between the two groups, and calculated the number of shared proteins between them. We used 100 replicates per simulated set of proteins and repeated this procedure for 100 trials (preliminary runs showed that the results were just as precise using only 100 replicates instead of the full 1000 employed in the analysis.
of the real data). If our algorithm is behaving properly, less than 5% of such trial simulations should yield a statistically significant result. We conducted two variants of this test. In the first variant, each of the 10,000 proteins was equally abundant. In the second variant, the protein abundances followed an exponential distribution, in which there are a few relatively abundant proteins and a large number of relatively rare proteins. We simulated this distribution by drawing elements from a beta distribution with parameters shape1 = 0.5, shape2 = 1.0.

Of the 100 trials with equally abundant proteins, there was only 1 simulation in which the null hypothesis was rejected. Of the 100 trials with an exponential distribution of protein abundances, none of the trial data sets rejected the null hypothesis. We conclude from this exercise that the null model test that we used has good Type I error properties, and does not lead to spurious rejection of the null hypothesis when both treatments are sampled from a single protein pool.

**Taxonomic Analysis**

To determine the taxonomic composition of the microbes contributing to identified proteins in our treatments, we conducted a BLAST homology search of the metagenomic sequence data for protein hits. All peptides from the top 220 identified proteins in each treatment were mapped back to their contigs of origin to obtain nucleotide sequences. Each nucleotide sequence was repeated by the number of associated peptides and searched via BLAST (NCBI), allowing us to obtain a weighted hit table for each treatment. The GI number from the top blast hit was extracted from the hit table for each query sequence for each treatment. The resulting GI numbers were then searched against
the NCBI Nucleotide Database via a script that returned organism subfield values (i.e. species name), yielding a list of species names for each treatment, associated with the top blast hit.

Hurlbert’s Probability of an Interspecific Encounter ($PIE$) was calculated for each treatment using the following equation:

$$ PIE = \left( \frac{N}{N-1} \right) \left( 1 - \sum_{i=1}^{s} p_i^2 \right) $$

where $N$ is the total number of peptides identified in a treatment, $p_i$ is the proportion of peptides in a treatment represented by bacterial class $i$, and $s$ is the number of bacterial classes identified in a treatment.

$O_2$ Requirements

We mapped each bacterial species identified in our BLAST search to its $O_2$ requirement using data from the Integrated Microbial Genomes database (IMG) (Reddy et al., 2015; Timinskas et al., 2014) The IMG database contains 6 classes of $O_2$ requirements: aerobe, anaerobe, facultative, microaerophilic, obligate aerobe, and obligate anaerobe. The latter three categories make up less than 7% of the database. We merged any species classified as obligate aerobes or obligate anaerobes into the aerobe and anaerobe classes, respectively.
| Pathway                                           | Control | Enriched | Adjusted |
|--------------------------------------------------|---------|----------|----------|
| Aflatoxin biosynthesis                          | 9       | 0        | 0.000    |
| Alanine aspartate and glutamate metabolism      | 15      | 196      | 0.000    |
| alpha Linolenic acid metabolism                 | 0       | 10       | 0.294    |
| Aminoacyl tRNA biosynthesis                     | 0       | 21       | 0.069    |
| Aminobenzoate degradation                       | 1       | 23       | 0.120    |
| Arginine and proline metabolism                 | 21      | 170      | 0.000    |
| Ascorbate and aldarate metabolism               | 4       | 10       | 0.824    |
| Benzoate degradation                            | 7       | 34       | 0.697    |
| beta Alanine metabolism                         | 4       | 190      | 0.000    |
| Biosynthesis of unsaturated fatty acids          | 2       | 37       | 0.048    |
| Biotin metabolism                               | 6       | 8        | 0.172    |
| Butanoate metabolism                            | 21      | 118      | 0.120    |
| C5 Branched dibasic acid metabolism             | 0       | 20       | 0.077    |
| Caprolactam degradation                         | 7       | 15       | 0.465    |
| Carbon fixation in photosynthetic organisms      | 47      | 53       | 0.000    |
| Carbon fixation pathways in prokaryotes          | 51      | 152      | 0.354    |
| Chloroalkane and chloroalkene degradation       | 8       | 67       | 0.064    |
| Citrate cycle TCA cycle                         | 92      | 140      | 0.000    |
| Cyanoamino acid metabolism                      | 0       | 7        | 0.447    |
| Cysteine and methionine metabolism              | 68      | 39       | 0.000    |
| Drug metabolism cytochrome P450                 | 2       | 14       | 0.660    |
| Fatty acid biosynthesis                         | 4       | 6        | 0.400    |
| Fatty acid degradation                          | 22      | 51       | 0.168    |
| Fatty acid elongation                           | 2       | 56       | 0.004    |
| Geraniol degradation                            | 3       | 67       | 0.004    |
| Glutathione metabolism                          | 14      | 73       | 0.346    |
| Glycerolipid metabolism                         | 2       | 111      | 0.858    |
| Glycerophospholipid metabolism                  | 0       | 19       | 0.085    |
| Glycine serine and threonine metabolism         | 0       | 45       | 0.004    |
| Glycolysis and gluconeosis                      | 33      | 169      | 0.120    |
| Glyoxylate and dicarboxylate metabolism         | 81      | 239      | 0.186    |
| Histidine metabolism                            | 5       | 33       | 0.369    |
| Inositol phosphate metabolism                   | 0       | 70       | 0.000    |
| Limonene and pinene degradation                 | 21      | 67       | 0.781    |
| Lysine degradation                              | 21      | 53       | 0.298    |
| Metabolism of xenobiotics by cytochrome P450    | 2       | 6        | 1.000    |
| Methane metabolism                              | 28      | 76       | 0.331    |
| Naphthalene degradation                         | 0       | 4        | 0.741    |
| Nitrogen metabolism                             | 21      | 169      | 0.000    |
| Novobiocin biosynthesis                         | 39      | 0        | 0.000    |
| One carbon pool by folate                       | 0       | 16       | 0.120    |
| Oxidative phosphorylation                       | 18      | 98       | 0.194    |
| Pantothenate and CoA biosynthesis               | 14      | 19       | 0.021    |
| Pentose and glucuronate interconversions        | 3       | 39       | 0.075    |
| Pentose phosphate pathway                       | 40      | 2        | 0.000    |
| Phenylalanine tyrosine and tryptophan biosynthesis | 30  | 3        | 0.000    |
| Phenylalanine metabolism                        | 3       | 49       | 0.026    |
| Phenylpropanoid biosynthesis                    | 0       | 27       | 0.031    |
| Phosphatidylinositol signaling system           | 0       | 5        | 0.637    |
| Porphyrin and chlorophyll metabolism            | 3       | 4        | 0.027    |
| Primary bile acid biosynthesis                  | 3       | 20       | 0.554    |
| Propanoate metabolism                           | 22      | 144      | 0.027    |
| Purine metabolism                               | 91      | 346      | 0.767    |
| Metabolism                                    | Total Genes | Total Molecules | p-value |
|-----------------------------------------------|-------------|-----------------|---------|
| Pyrimidine metabolism                         | 94          | 257             | 0.048   |
| Pyruvate metabolism                           | 62          | 58              | 0.000   |
| Retinol metabolism                            | 0           | 3               | 0.858   |
| Selenocompound metabolism                     | 4           | 0               | 0.004   |
| Streptomyacin biosynthesis                    | 0           | 3               | 0.858   |
| Sulfur metabolism                             | 4           | 0               | 0.004   |
| Synthesis and degradation of ketone bodies    | 5           | 5               | 0.120   |
| Taurine and hypotaurine metabolism            | 6           | 7               | 0.120   |
| Terpenoid backbone biosynthesis               | 2           | 3               | 0.741   |
| Tetracycline biosynthesis                     | 2           | 0               | 0.120   |
| Thiamine metabolism                           | 3           | 11              | 1.000   |
| Toluene degradation                           | 11          | 17              | 0.085   |
| Tryptophan metabolism                         | 20          | 203             | 0.000   |
| Tyrosine metabolism                           | 13          | 14              | 0.007   |
| Valine leucine and isoleucine biosynthesis    | 10          | 89              | 0.021   |
| Valine leucine and isoleucine degradation     | 37          | 222             | 0.013   |
Table S2. Species, oxygen requirements, and bacterial classes identified in control and enriched pitchers in a BLAST search of nucleotide sequences associated with the top 220 proteins in each treatment, weighted by total peptides. NA values represent species that were non-bacterial.

| Species Name                  | Oxygen Requirement | Class            | Control Peptides | Enriched Peptides |
|-------------------------------|--------------------|------------------|------------------|-------------------|
| *Achromobacter xylosoxidans*  | Aerobe             | Betaproteobacteria | 9                | 0                 |
| *Acidiphilium multivorum*     | Aerobe             | Alphaproteobacteria | 0                | 7                 |
| *Acidovorax avenae*           | Aerobe             | Betaproteobacteria | 12               | 4                 |
| *Acidovorax citrulli*         | Aerobe             | Betaproteobacteria | 13               | 32                |
| *Agrobacterium radiobacter*   | Aerobe             | Alphaproteobacteria | 3                | 0                 |
| *Alicycilphilus denitrificans*| Facultative        | Betaproteobacteria | 9                | 2                 |
| *Alkalilimnicola ehrlichii*   | Anaerobe           | Gammaproteobacteria | 0                | 2                 |
| *Azoarcus sp*                 | Facultative        | Betaproteobacteria | 4                | 0                 |
| *Azorhizobium caulinodans*    | Unclassified       | Alphaproteobacteria | 3                | 3                 |
| *Azospirillum sp*              | Facultative        | Alphaproteobacteria | 0                | 3                 |
| *Bordetella pertussi*         | Aerobe             | Betaproteobacteria | 0                | 2                 |
| *Bradyrhizobium sp*           | Aerobe             | Alphaproteobacteria | 10               | 0                 |
| *Burkholderia cenocepecia*    | Facultative        | Betaproteobacteria | 0                | 6                 |
| *Burkholderia cepaci*         | Aerobe             | Betaproteobacteria | 2                | 0                 |
| *Burkholderia fungoru*        | Aerobe             | Betaproteobacteria | 14               | 0                 |
| *Burkholderia gladiol*        | Aerobe             | Betaproteobacteria | 0                | 24                |
| *Chitinophaga pinensis*       | Aerobe             | Sphingobacteria   | 17               | 14                |
| *Chromobacterium violaceum*   | Facultative        | Betaproteobacteria | 62               | 1549              |
| *Clavibacter michiganensis*   | Aerobe             | Actinobacteria    | 2                | 0                 |
| *Clostridium saccharbutylicum*| Anaerobe           | Clostridia        | 3                | 7                 |
| *Collimonas fungivorans*      | Aerobe             | Betaproteobacteria | 10               | 0                 |
| *Corynebacterium halotolerans*| Aerobe             | Actinobacteria    | 3                | 0                 |
| *Croceicoccus naphthovoran*   | Unclassified       | Alphaproteobacteria | 4                | 0                 |
| *Cupriavidus taiwanensis*     | Facultative        | Betaproteobacteria | 2                | 0                 |
| *Dechloromonas aromatica*     | Facultative        | Betaproteobacteria | 2                | 5                 |
| *Dechlorosoma suillum*        | Anaerobe           | Betaproteobacteria | 17               | 61                |
| *Delftia acidovorans*         | Aerobe             | Betaproteobacteria | 3                | 0                 |
| **Delftia sp** | Aerobe | Betaproteobacteria | 0 | 2 |
| **Desulfovibrio vulgaris** | Anaerobe | Deltaproteobacteria | 2 | 0 |
| **Draconibacterium oriental** | Facultative | Bacteroidia | 12 | 16 |
| **Dyadobacter fermentans** | Aerobe | Cytophagia | 0 | 2 |
| **Dyella jiangningensi** | Aerobe | Gammaproteobacteria | 2 | 17 |
| **Emticicia oligotrophica** | Aerobe | Cytophagia | 4 | 11 |
| **Flavobacteriaceae bacterium** | Aerobe | Flavobacteriia | 0 | 3 |
| **Gloeobacter violaceus** | Aerobe | Gloeobacteria | 8 | 0 |
| **Hymenobacter sp** | Aerobe | Cytophagia | 8 | 4 |
| **Janthinobacterium agaricidamnosum** | Unclassified | Betaproteobacteria | 0 | 4 |
| **Janthinobacterium sp** | Unclassified | Betaproteobacteria | 13 | 82 |
| **Laribacter hongkongensis** | Anaerobe | Betaproteobacteria | 0 | 8 |
| **Leifsonia xyli** | Aerobe | Actinobacteria | 14 | 0 |
| **Leptospira interrogans** | Aerobe | Spirochaetia | 11 | 9 |
| **Leptothrix cholodnii** | Aerobe | Betaproteobacteria | 2 | 0 |
| **Mesorhizobium cici** | Aerobe | Alphaproteobacteria | 11 | 0 |
| **Methylobacterium aquaticu** | Aerobe | Alphaproteobacteria | 4 | 5 |
| **Methylobacterium populi** | Aerobe | Alphaproteobacteria | 2 | 4 |
| **Methylobacterium radiotolerans** | Aerobe | Alphaproteobacteria | 17 | 0 |
| **Methylovorus sp** | Facultative | Betaproteobacteria | 4 | 0 |
| **Microbacterium testaceum** | Aerobe | Actinobacteria | 13 | 3 |
| **Niabelli soli** | Aerobe | Sphingobacteriia | 7 | 6 |
| **Novosphingobium aromaticivorans** | Aerobe | Alphaproteobacteria | 56 | 40 |
| **Oxalis latifoli** | NA | NA | 2 | 2 |
| **Pedobacter heparinus** | Aerobe | Sphingobacteriia | 16 | 27 |
| **Polaromonas naphthalenivorans** | Aerobe | Betaproteobacteria | 0 | 6 |
| **Polymorphum gilvum** | Facultative | Alphaproteobacteria | 2 | 0 |
| **Pseudogulbenkiania sp** | Facultative | Betaproteobacteria | 15 | 445 |
| **Pseudomonas denitrificans** | Aerobe | Gammaproteobacteria | 0 | 7 |
| **Pseudomonas entomophila** | Aerobe | Gammaproteobacteria | 0 | 7 |
| **Pseudomonas knackmussii** | Unclassified | Gammaproteobacteria | 0 | 10 |
| Species                        | Life Form | Class          | Gammaproteobacteria | Alphaproteobacteria | Betaproteobacteria | Acidobacteria | Chloroflexi | Cytophagia | Gammaproteobacteria |
|--------------------------------|-----------|----------------|---------------------|---------------------|--------------------|--------------|-------------|------------|---------------------|
| *Pseudomonas pseudoalcaligene* | Aerobe    | Gammaproteobacteria | 0                   | 3                   |                    |              |             |           |                     |
| *Pseudomonas putida*           | Aerobe    | Gammaproteobacteria | 0                   | 12                  |                    |              |             |           |                     |
| *Pseudomonas rhizosphaera*     | Aerobe    | Gammaproteobacteria | 21                  | 0                   |                    |              |             |           |                     |
| *Pseudopedobacter saltans*     | Aerobe    | Sphingobacteria   | 8                   | 6                   |                    |              |             |           |                     |
| *Pseudoxanthomonas spadix*     | Aerobe    | Gammaproteobacteria | 2                   | 0                   |                    |              |             |           |                     |
| *Pusillimonas sp*              | Unclassified | Betaproteobacteria | 2                   | 0                   |                    |              |             |           |                     |
| *Ramlibacter tataouinensis*    | Aerobe    | Betaproteobacteria | 8                   | 6                   |                    |              |             |           |                     |
| *Rhizobium etli*               | Aerobe    | Alphaproteobacteria | 22                  | 31                  |                    |              |             |           |                     |
| *Rhizobium sp*                 | Aerobe    | Alphaproteobacteria | 0                   | 5                   |                    |              |             |           |                     |
| *Rhizophagus intraradice*      | NA        | NA              | 2                   | 0                   |                    |              |             |           |                     |
| *Rhodanobacter denitrifican*   | Facultative | Gammaproteobacteria | 22                  | 88                  |                    |              |             |           |                     |
| *Rhodopseudomonas palustris*   | Facultative | Alphaproteobacteria | 2                   | 31                  |                    |              |             |           |                     |
| *Roseiflexus sp*               | Facultative | Chloroflexi     | 0                   | 3                   |                    |              |             |           |                     |
| *Runella slithyformus*         | Aerobe    | Cytophagia       | 2                   | 0                   |                    |              |             |           |                     |
| *Sinorhizobium fredii*         | Aerobe    | Alphaproteobacteria | 8                   | 0                   |                    |              |             |           |                     |
| *Sphingobium chlorophenolicum* | Aerobe    | Alphaproteobacteria | 0                   | 5                   |                    |              |             |           |                     |
| *Sphingobium japonicum*        | Aerobe    | Alphaproteobacteria | 2                   | 0                   |                    |              |             |           |                     |
| *Sphingobium sp*               | Unclassified | Alphaproteobacteria | 7                   | 0                   |                    |              |             |           |                     |
| *Sphingomonas sanxanigenens*   | Aerobe    | Alphaproteobacteria | 3                   | 0                   |                    |              |             |           |                     |
| *Sphingomonas sp*              | Aerobe    | Alphaproteobacteria | 38                  | 3                   |                    |              |             |           |                     |
| *Sphingomonas tax*             | Aerobe    | Alphaproteobacteria | 4                   | 0                   |                    |              |             |           |                     |
| *Sphingomonas wittichii*       | Aerobe    | Alphaproteobacteria | 30                  | 11                  |                    |              |             |           |                     |
| *Sphingopyxis alaskensis*      | Aerobe    | Alphaproteobacteria | 7                   | 7                   |                    |              |             |           |                     |
| *Starkeya novella*             | Aerobe    | Alphaproteobacteria | 41                  | 41                  |                    |              |             |           |                     |
| *Stenotrophomonas rhizophil*   | Unclassified | Gammaproteobacteria | 3                   | 0                   |                    |              |             |           |                     |
| *Terriglobus roseus*           | Aerobe    | Acidobacteria     | 4                   | 0                   |                    |              |             |           |                     |
| *Terriglobus saanensis*        | Aerobe    | Acidobacteria     | 2                   | 0                   |                    |              |             |           |                     |
| *Variovorax paradoxus*         | Aerobe    | Betaproteobacteria | 266                 | 210                 |                    |              |             |           |                     |
Fig. S1. Microbial proteins in control (C) and enriched (E) pitchers. (a) Three replicate pitchers of each treatment were initially processed in November 2012. (b) The remaining replicates were processed in May 2013. Lanes 4, 5, and 7 represent enriched pitchers. Lanes 9, 11, and 13 represent control pitchers. The replicate in lane 13 was omitted from the study due to a lack of protein. Letters a-e represent the regions that each lane was cut into for MS/MS analysis.
Fig. S2. (a) Agarose gel electrophoresis of metagenomic DNA from three pitcher plant microbial communities. One pellet from each pitcher was treated with lysozyme. All samples were pooled prior to sequencing. (b) Frequency distribution of the read lengths in the sequenced metagenomic data. The median read length was 482 bp. (c) Frequency distribution of assembled contig lengths in the metagenomic database. All contigs were 500 bp or greater in length.
**Fig. S3** Taxonomic assignments of metagenome, as visualized by Krona. The rings, from the center outward represent Kingdom (Bacteria), Phylum, Class, Order.
Fig. S4. Functional potential of the metagenome. Rank abundance of the proportion of mapped contigs assigned to a) level 2 KEGG pathways and b) level 3 KEGG pathways.
Fig. S5 Taxonomic assignments of bacterial proteins, as visualized by Krona, differed between control (a) and enriched (b) pitchers. Sunburst diagrams were constructed using nucleotide sequences from the metagenomic data associated with identified proteins in the custom protein database. Nucleotide sequences were weighted by the total number of peptides associated with each sequence. Replicates were pooled for each treatment. Figures feature only matches to bacteria. The rings, from the center outward represent Kingdom (Bacteria), Phylum, Class, Order, Family.
Fig. S6. Taxonomic assignments of bacterial proteins, as visualized by Unipept, differed between (a) control and (b) enriched pitchers. The rings, from the center outward represent Kingdom (dark blue = Bacteria), Phylum (white = Proteobacteria), Class (red = Alphaproteobacteria, light blue = Betaproteobacteria), Order (dark blue = Burkholderiales, rose = Neisseriales, light blue = Sphingomonadales).
Fig. S7. Pathway representation of the proportion of total peptides associated with KEGG pathways differed between control (blue) and enriched (brown) pitchers.
Figure S8. KEGG pathway assignments differed between control and enriched pitchers. (a) Heat map of the proportional representation of pathways between control pitchers (C) and enriched pitchers (E) and individual control (H4, H6, E3, C4) and enriched (H1, H2, H3, 2C, 5B, 5A) replicates. Significantly different pathways between pooled control and enriched samples are indicated with “**”.
Data S1. Table in .csv file form listing the proteins from the top 220 analyzed in control and enriched treatments and their associated peptides.