Disentangling the rhizosphere community through stable isotope informed genome-resolved metagenomics and assembled metatranscriptomes

By

Evan P. Starr

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Microbiology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Jillian F. Banfield, Co-Chair
Professor Mary K. Firestone, Co-Chair
Professor Britt Koskella

Summer 2019

Abstract

Disentangling the holistic rhizosphere community through stable isotope informed genome-resolved metagenomics and assembled metatranscriptomes

By

Evan P Starr

Doctor of Philosophy in Microbiology

University of California, Berkeley

Professor Jillian Banfield, Co-Chair

Professor Mary Firestone, Co-Chair

The functioning, health, and productivity of soil is intimately tied to the complex network of interactions in the rhizosphere. Because of this, the rhizosphere has been rigorously studied for over a century, but due to technical limitations many aspects of soil biology have been overlooked. In order to better understand rhizosphere functioning, my work has focused on the less explored organisms and interactions in microbial communities, this includes unculturable bacteria along with viruses and eukaryotes. Only by considering soil biology more holistically can we better understand the functioning of this enigmatic yet critical ecosystem. Knowledge about these interactions could direct how we think about plant-microbe relationships, soil carbon stabilization and the roles of understudied organisms in biogeochemical cycling.

The transformation of plant photosynthate into soil organic carbon and its recycling to CO_2 by soil microorganisms is one of the central components of the terrestrial carbon cycle. There are currently large knowledge gaps related to which soil-associated microorganisms take up plant carbon in the rhizosphere and the fate of that carbon. Additionally, understanding about obligate symbionts such as members of the Candidate Phyla Radiation (CPR) in soil is severely limited, both from the perspective of their genomic potential and their interactions with the greater soil community. We conducted an experiment in which common wild oats (*Avena fatua)* were grown in a $^{13}CO_2$ atmosphere and the rhizosphere and non-rhizosphere soil was sampled for genomic analyses. Density gradient centrifugation of DNA extracted from soil samples enabled distinction of microbes that did and did not incorporate the ^{13}C into their DNA. A 1.45-Mbp genome of a Saccharibacteria (TM7) was identified and, despite the microbial complexity of rhizosphere soil, curated to completion. The genome lacks many biosynthetic pathways, including genes required to synthesize DNA de novo. Rather, it acquires externally derived nucleic acids for DNA and RNA synthesis. Given this, we conclude that rhizosphere-associated Saccharibacteria recycle DNA from bacteria that live off plant exudates and/or phage that acquired ^{13}C because they preyed upon these bacteria and/or directly from the labeled plant DNA. Isotopic labeling

indicates that the population was replicating during the 6-week period of plant growth. Interestingly, the genome is ~ 30% larger than other complete Saccharibacteria genomes from non-soil environments, largely due to more genes for complex carbon utilization and amino acid metabolism. Given the ability to degrade cellulose, hemicellulose, pectin, starch, and 1,3-β-glucan, we predict that this Saccharibacteria generates energy by fermentation of soil necromass and plant root exudates to acetate and lactate. The genome also encodes a linear electron transport chain featuring a terminal oxidase, suggesting that this Saccharibacteria may respire aerobically. The genome encodes a hydrolase that could breakdown salicylic acid, a plant defense signaling molecule, and genes to interconvert a variety of isoprenoids, including the plant hormone zeatin. We propose that isotopically labeled CO_2 is incorporated into plant-derived carbon and then into the DNA of rhizosphere organisms capable of nucleotide synthesis, and the nucleotides are recycled into Saccharibacterial genomes.

We collected paired rhizosphere and non-rhizosphere soil at six and nine weeks of plant growth and extracted DNA that was separated by density gradient centrifugation. The separate fractions were sequenced, assembled, and binned to generate 55 unique microbial genomes that were >70% complete. Evidence for close interaction between bacteria, micro-eukaryotes and plant roots includes the ability to modulate plant signaling hormones, abundant plant pathogenicity factors and production of cyanide and insecticidal toxins. We reconstructed eukaryotic 18S rRNA sequences and identified micro-eukaryotic bacterivores and fungi in the rhizosphere soil. In addition, we reconstructed two complete genomes for phage that were among the most highly ^{13}C-enriched entities in our study. CRISPR locus targeting connected a phage to a Burkholderiales host predicted to be a plant pathogen and a possible plant growth promoting *Catenulispora* may serve as the host for another phage. Thus, ^{13}C could be tracked from the atmosphere into plant roots, soil and through the rhizosphere food web.

Viruses impact nearly all organisms on Earth, with ripples of influence in agriculture, health and biogeochemical processes. We previously investigated DNA phage, however, very little is known about RNA viruses in an environmental context, and even less is known about their diversity and ecology in the most complex microbial system, soil. Here, we assembled 48 individual metatranscriptomes from four habitats within a soil sampled over a 22-day time series: rhizosphere alone, detritosphere alone, a combination of the two, and unamended soil (four time points and three biological replicates per time point). We resolved the RNA viral community, uncovering a high diversity of viral sequences. We also investigated possible host organisms by analyzing metatranscriptome marker gene content. Based on viral phylogeny, much of the diversity was *Narnaviridae* that parasitize fungi or *Leviviridae* that infect Proteobacteria. Both host and viral communities appear to be highly dynamic, and rapidly diverged depending on experimental conditions. The viral communities were structured based on the presence of litter, while putative hosts appeared to be impacted by both the presence of litter and roots. A clear time signature from *Leviviridae* and their hosts indicated that viruses were replicating. With this time-resolved analysis, we show that RNA viruses are diverse, abundant and active in soil. Their replication causes host cell death, mobilizing carbon in a process that represents a largely overlooked component of carbon cycling in soil.

By combining state of the art techniques, stable isotope probing, genome-resolved metagenomics and assembled metatranscriptomics, we advanced knowledge about the interplay between

understudied players in the rhizosphere and provided some clues for the fate of plant derived carbon in the soil microbial ecosystem. The use of genome resolved metagenomics is the only current way to determine the lifestyle of uncultured microbes. Complete genomes are still difficult to reconstruct however, they contain extensive of information, both regarding the presence and the absence of capabilities. Stable isotope probing allowed us to follow plant fixed carbon into the microbial community and in several cases across multiple trophic levels. This study demonstrates the power of stable isotope-informed genome-resolved metagenomics to resolve aspects of the complex rhizosphere food web. The approach will find broad application for study of other soils and different ecosystems.

Table of Contents

Introduction **ii**

Acknowledgments **iv**

1 Stable isotope informed genome-resolved metagenomics reveals that Saccharibacteria utilize-microbially processed plant-derived carbon **1**
1.2 Introduction 2
1.3 Materials and methods 3
1.4 Results 4
1.5 Discussion 11
1.6 Conclusions 12
1.7 Figures 13

2 Metatranscriptomic reconstruction reveals RNA viruses with the potential to shape carbon cycling in soil **17**
2.2 Introduction 18
2.3 Results 19
2.4 Discussion 24
2.5 Conclusions 26
2.6 Methods 26
2.7 Figures 30

3 Stable isotope informed genome-resolved metagenomics uncovers trophic interactions in rhizosphere soil **34**
3.1 Abstract 34
3.2 Introduction 34
3.3 Results 36
3.4 Discussion 42
3.5 Conclusions 44
3.6 Methods 44
3.7 Figures 47
3.8 Supplemental figures 53

Conclusions **59**

References **62**

Introduction

My appreciation for the complexity of life brought me to the study of Biology, and the sheer density of organisms, connections, and relevant outputs rooted me in the study of soils. Soils are some of the most complex biological systems in the world, made up of bacteria, archaea, eukaryotes of all sizes, viruses, phage (viruses of bacteria), not to mention the physical complexity of the soil matrix itself. Our understanding of soil is vitally important for addressing climate change, agricultural practices, and ecosystem health. In my opinion one must consider all interacting players to better understand soil ecology. I am especially interested in the less well-known members of the soil community: uncultured bacteria, phage, RNA viruses and microeukaryotes. These organisms have been largely ignored due to technological short comings, however, with recent advances in sequencing technologies, we are finally able to explore the life of these mysterious beings. Soil is a complex pool of interacting organisms, and through genome analysis I hope to generate hypotheses for how these organisms could be interacting with one another.

Genome resolved metagenomics allows us to uncover not only the identity of the present organisms, but also their genetic potential. By analyzing the genomes of the sequenced bacteria, viruses and eukaryotes, we can begin to piece together how they may be interacting with one another. This includes looking for plant hormone modulating machinery, antibacterial compounds and CRISPR-Cas systems to name a few. From the genome analysis we can begin to predict whether a certain bacterium will act as a plant pathogen or a plant growth promoting bacteria or how they may defend themselves against micro-eukaryote grazing. These genomes may contain valuable biomolecules, insecticides, antibiotics, antifungals and novel enzymes. The genetic capabilities provide the context to explore inter-organismal interactions, however this does not speak to the reality of their connectedness. By following stable isotopes we can get information about the physical movement of carbon through the soil community, which bacteria consumed plant derived carbon, and which organisms may have predated or parasitized the bacteria. In Chapter One I combine these techniques to show that a bacteria with a labelled genome but no capability to synthesize nucleotides may have collected or stolen necessary biomolecules from other labelled rhizosphere organisms. In my work, I show that the less well known members of the soil community are present, active and likely influencing the ecology of the rhizosphere and soil in general.

One aspect of the soil community which is currently understudied is viruses, especially RNA based viruses. The study of viruses has been difficult because soils cannot be filtered, and many bacteria cannot be cultured. I am especially intrigued by the impacts of these small, but powerful, parasites on their hosts and the system as a whole. Chapter Three includes work on complete DNA phage genomes and indications of their possible hosts. In one case, we were able to connect the flow of carbon from the plant, into a bacterial plant pathogen, and finally into an infecting phage. Chapter Two represents the first study of diversity or ecology of eukaryotic RNA viruses or RNA phage in soil. We show that key soil habitats influence the composition of RNA viral communities and the RNA viral communities change over time, indicating an active

and virulent population. Through this work we greatly increased the known diversity of RNA viruses and their sheer abundance spoke to their possible large impacts on the soil community.

Acknowledgments

I would first like to thank my two wonderful professors, Jill Banfield and Mary Firestone. I could not have done this without your support, guidance and encouragement. I am extremely grateful for the countless hours you both have devoted to mentoring my scientific growth. The time spent wondering about obscure questions, exploring data, talking about weird and rare members of the soil community has taught me to remain curious and ask better questions. The both of you have been instrumental in my growth as a scientist and as a person, thank you. I cannot repay you the work you have put in, I can only hope to be as good a mentor as the two of you have been to me.

Megan, you are the light of my life. Without your support, encouragement and inspiration this would not have been possible. Thank you for the emotional support during the stressful times, the fun distractions when I needed them, and most of all for being my friend. I cannot express how lucky I am to have you in my life.

To my family, thank you all for the love and support. You made me who I am today and I am eternally grateful. Thank you mom, dad, Susan and Nat. Thank you all for the love and support. I am inspired by your dedication to the people and callings that you all have found. I could not ask for a better family and I cannot thank you enough.

I would like to thank all the people who have helped me through this time. Thank you to all the members of the Banfield and Firestone labs for their help and friendship. You all have made my time here so fun and supportive. Thank you for teaching me bioinformatics, soil science, and ecology from the ground up. I could not ask for two more welcoming and friendly labs who also do amazing, ground breaking work. Every talk, grant writing session and lab meeting has been a delight. Thank you for the support and help from the entire PMB department, especially my cohort and Rocio Sanchez.

Thank you to all the people who have helped get me here and who have been so supportive. Jessica Taylor, my high school biology teacher who sparked my interest in biology and helped get me my first job in a research lab. The professors at Carleton College, especially Raka Mitra, Mark McKone and Matt Rand who taught me so much, in so little time, and made my career in science a reality. Thank you Britt Koskella for the helpful comments on my dissertation. I am thankful for my amazing collaborators: everyone at the Eel River CZO, the Livermore crew (JP, Steve, and Erin), Alex Probst, Shengjing, and the Hungate lab (especially Benjamin Koch). Thank you to my friends who have supported me this entire time: Bill, Micah, Soren, Ellie, Amanda, Alex and Gail.

1 Stable isotope informed genome-resolved metagenomics reveals that Saccharibacteria utilize-microbially processed plant-derived carbon

Starr, Evan P., Shi Shengjing, Blazewicz, Steven J., Probst, Alexander J., Herman, Donald J. Firestone, Mary K. and Banfield, Jillian F.

Published in *Microbiome*, July 2018

1.1.1 Background

The transformation of plant photosynthate into soil organic carbon and its recycling to CO_2 by soil microorganisms is one of the central components of the terrestrial carbon cycle. There are currently large knowledge gaps related to which soil-associated microorganisms take up plant carbon in the rhizosphere and the fate of that carbon.

1.1.2 Results

We conducted an experiment in which common wild oats (*Avena fatua)* were grown in a $^{13}CO_2$ atmosphere and the rhizosphere and non-rhizosphere soil was sampled for genomic analyses. Density gradient centrifugation of DNA extracted from soil samples enabled distinction of microbes that did and did not incorporate the ^{13}C into their DNA. A 1.45 Mbp genome of a Saccharibacteria (TM7) was identified and, despite the microbial complexity of rhizosphere soil, curated to completion. The genome lacks many biosynthetic pathways, including genes required to synthesize DNA *de novo*. Rather, it requires externally-derived nucleotides for DNA and RNA synthesis. Given this, we conclude that rhizosphere-associated Saccharibacteria recycle DNA from bacteria that live off plant exudates and/or phage that acquired ^{13}C because they preyed upon these bacteria and/or directly from the labeled plant DNA. Isotopic labeling indicates that the population was replicating during the six-week period of plant growth. Interestingly, the genome is ~30% larger than other complete Saccharibacteria genomes from non-soil environments, largely due to more genes for complex carbon utilization and amino acid metabolism. Given the ability to degrade cellulose, hemicellulose, pectin, starch and 1,3-β-glucan, we predict that this Saccharibacteria generates energy by fermentation of soil necromass and plant root exudates to acetate and lactate. The genome also encodes a linear electron transport chain featuring a terminal oxidase, suggesting that this Saccharibacteria may respire aerobically. The genome encodes a hydrolase that could breakdown salicylic acid, a plant defense signaling molecule, and genes to interconvert a variety of isoprenoids, including the plant hormone zeatin.

1.1.3 Conclusions

Rhizosphere Saccharibacteria likely depend on other bacteria for basic cellular building blocks. We propose that isotopically labeled CO_2 is incorporated into plant-derived carbon and then into the DNA of rhizosphere organisms capable of nucleotide synthesis, and the nucleotides are recycled into Saccharibacterial genomes.

1.2 Introduction

The Candidate Phyla Radiation (CPR) comprises a large fraction of the bacterial domain (Brown et al. 2015, Hug et al. 2016). Within the CPR, Saccharibacteria, formerly TM7, is one of the most captivating phyla because of the wide diversity of habitats in which it is found, including activated sludge, human and dolphin oral cavities, seawater, aquifer sediment, soil, and cockroach guts (Hugenholtz et al. 2001, Marcy et al. 2007, Dewhirst et al. 2010, Schauer et al. 2012, Kantor et al. 2013, He et al. 2014, Kindaichi et al. 2016, Dudek et al. 2017). Much has been learned about the metabolism of Saccharibacteria and their influence on their environment, for instance the possible immunosuppressive capabilities of an episymbiont of *Actinomyces odontolyticus* in human mouths and bulking issues caused in wastewater treatment plants (Thomsen et al. 2002, He et al. 2014). There have been previously proposed subdivisions within the Saccharibacteria phylum, but currently the majority of our genomic data comes from organisms which fall within Subdivision 3 and there exists very little genomic representation of Subdivision 1 (Hugenholtz et al. 2001, Ferrari et al. 2014). Organisms in Subdivision 1 are predominantly found in environmental samples (Dinis et al. 2011). One of the reasons why there may be less genomic data available for Saccharibacteria in Subdivision 1 is because of the high microbial diversity of environmental samples, especially soil, which makes sequencing efforts more challenging. Based on 16S ribosomal RNA gene surveys we know that Saccharibacteria occur in the rhizosphere, but little is known about their metabolism and how they differ from related organisms growing in other environments (Correa-Galeote et al. 2016, Beckers et al. 2017).

There are six previously reported, closed and circularized genomes for Saccharibacteria (one from a wastewater treatment plant (Albertsen et al. 2013), human mouth (He et al. 2014), two from sediment (Kantor et al. 2013, Brown et al. 2015), and two from a thiocyanate remediation wastewater reactor (Kantor et al. 2015)) and some partial genomes. Based on genomic analyses, it was previously predicted that Saccharibacteria are anaerobic fermenters, despite the documentation of one genome encoding a single-subunit of NADH: ubiquinone dehydrogenase and a complete ubiquinol oxidase, which was explained as an oxygen scavenging mechanism (Kantor et al. 2013). It has also been documented that Saccharibacteria are able to grow in the presence of oxygen (He et al. 2014, Bor et al. 2016, Kindaichi et al. 2016). Previous soil SIP work has shown that Saccharibacteria DNA became labeled in the presence of ^{13}C cellulose and toluene (Luo et al. 2009, Wilhelm et al. 2017). Labeling experiments have indicated that $^{13}CO_2$ can be traced from plant fixation into the DNA of rhizosphere growing microbes (Haichar et al. 2008, Hernández et al. 2015). Here, we found that a Saccharibacteria population in rhizosphere soil incorporated ^{13}C into its genomic DNA, indicating that this population was active and dividing. We report the complete genome for this organism and show how analysis of its metabolism sheds light on the pathway by which the label was incorporated. We compare the genome to those of other complete Saccharibacteria to address the question of how this bacterium is adapted to live in soil and specifically in the rhizosphere.

1.3 Materials and methods

1.3.1 Labeling

Soil (0-10cm) was collected from the University of California Hopland Research and Extension Center (Hopland, CA, USA), from an area where *Avena* spp. are a common grass. Microcosms were constructed and plant growth conditions were regulated as described previously (Shi et al. 2015). For this analysis we used samples obtained from a single microcosm. A sample of T0 soil was collected after microcosm preparation and before planting. A single *A. fatua* plant was grown in a microcosm in a labeling chamber maintained at 400 uL/L CO_2, with native CO_2 replenished with 99 atom% $^{13}CO_2$. After 6 weeks a single sample of rhizosphere soil and a single sample of bulk soil was sampled at the vegetative stage; the microcosm was destructively harvested for rhizosphere soil and from bulk soil mesh bags which excluded root ingrowth. Rhizosphere soil was washed off the roots and DNA was extracted from 0.5g of soil using a phenol:chloroform extraction detailed in (Shi et al. 2015). We used bulk soil samples to control for the direct incorporation of $^{13}CO_2$ into biomass. We used two different types of samples to control for non-plant related $^{13}CO_2$ influence: 1) DNA extracted from pre-planted microcosm soil which never received $^{13}CO_2$ treatment and 2) DNA extracted from bulk soil collected at the same times as root harvesting (6 weeks) within the same microcosm (Shi et al. 2015). The pre-planted sample provides a clean natural abundance bulk soil and the T6 bulk soil sample controls for any direct microbial carbon fixation. Plant isotopic composition was determined on an Isoprime 100 isotope ratio mass spectrometer (Elementar, Langenselbold, Germany).

1.3.2 Stable isotope probing

To separate isotopically enriched DNA from unenriched DNA, each sample was separated based on density in a CsCl density gradient formed in an ultracentrifuge. Gradients were generated according to the method previously described (Blazewicz et al. 2014). Briefly 5.5 µg of DNA was added to the gradient buffer to create a solution with a density of 1.735 g/mL. Then 5.2 mL of the solution was transferred to an ultracentrifuge tube (Beckman Coulter Quick-Seal, 13 X 51 mm). Tubes were spun in an Optima L-90K ultracentrifuge (Beckman Coulter, Brea, California, USA) using a VTi65.2 rotor at 44000 rpm (176284 RCF_{avg}) at 20°C for 109 h with maximum acceleration and braking of the rotor to maintain the integrity of the density separations. Then the content of the ultracentrifuge tube was separated into ~32 fractions using a syringe pump to deliver light mineral oil at 0.25mL/min to displace the gradient solution from the pierced bottom of the tube. Each fraction was approximately 12 drops (~144µL). The density of each fraction was measured using an AR200 digital refractometer (Reichert Inc., Depew, New York, USA). The DNA for each fraction was precipitated and quantified as previously described (Blazewicz et al. 2014). Fractions were then binned based on density and by comparison between the rhizosphere samples and the associated bulk soil (light= 1.692-1.737 g/mL; middle= 1.738-1.746 g/mL; heavy= 1.747-1.765 g/mL) (Supplementary Figure S1.1 and Supplemental Table S1.1).

1.3.3 Sequencing

Each fraction, light, middle, and heavy was then sent to the UC Davis Genome Center DNA Technologies Core for sequencing. Each sample was sequenced using an Illumina HiSeq 3000

(Illumina Inc., Hayward, California, USA) with paired-end libraries prepared with the Kapa Hyper protocol and a read length of 150 bp.

1.3.4 Genome reconstruction, annotation and analysis

Reads were trimmed using Sickle (https://github.com/najoshi/sickle); BBtools (https://sourceforge.net/projects/bbmap/) was used to remove Illumina adapters and trace contaminants; finally reads were individually assembled using IDBA-UD (-step 20, -maxk 140, -mink 40) (Peng et al. 2012). A single 1.45 Mb scaffold was recovered and was able to be circularized, then scaffolding errors and completeness were assessed as described in (Brown et al. 2015) and scaffolding gaps were fixed manually by mapping reads to the scaffold using Bowtie2 on default settings (Langmead and Salzberg 2012). The scaffold was visualized in Geneious (Kearse et al. 2012). Genes were predicted using Prodigal (Hyatt et al. 2010). Predicted ORFs were functionally described using a multidatabase search pipeline. Sequence similarity searches were performed with USEARCH (Edgar 2010) against UniRef100 (Suzek et al. 2007), Uniprot (Magrane and Consortium 2011), and the KEGG database (Ogata et al. 1999). Additional gene annotations were assigned using HMMs that were constructed based on KEGG Orthologies (Ogata et al. 1999). All the proteins assigned to a KO were clustered using MCL (Van Dongen 2008) with inflation parameter (-I) of 1.1, based on global percent identity. Clusters were aligned using MAFFT v7 (Katoh and Standley 2013a), and HMMs were constructed using the HMMER suite (Finn et al. 2011). Carbohydrate-active enzymes were identified using dbCAN (Yin et al. 2012). Domain level functional annotations were done using InterProScan (Zdobnov and Apweiler 2001). tRNAs were predicted using tRNAScan-SE (Lowe and Eddy 1996) and cellular localization was predicted using PSORTb v3.0.2 with the gram-positive setting, which was demonstrated though previous imaging (Hugenholtz et al. 2001, Yu et al. 2010). Twin-arginine translocation signal peptide finding was done using TATFIND 1.4 (Dilks et al. 2003). The GC skew of the genome was calculated based on previously published tools (Brown et al. 2016). Protein modelling was done with Swiss Model and Ndh modelling incorporated the protein from Caldalkalibacillus thermarum and cytochrome bo3 ubiquinol terminal oxidase subunit I utilized a protein model from E. coli (Iwata et al. 2000, Heikal et al. 2014, Biasini et al. 2014). Representative Saccaribacteria 16S rRNA sequences were obtained from NCBI and aligned using ssu-align (Nawrocki et al. 2009) then a maximum-likelihood tree was constructed with RAxML by using the GTRCAT model with 1000 bootstraps (Supplementary Figure S1.2).

1.4 Results

After 6 weeks in labeling chambers, *Avena fatua* shoots were highly labeled (~94 atom% ^{13}C). DNA was extracted from the rhizosphere and bulk soil samples. By comparing the density separation of the rhizosphere community DNA to the bulk community DNA we were able to define un-enriched (light), partially ^{13}C-enriched (middle) and highly ^{13}C-enriched (heavy) fractions (**Figure 1.1**). Based on the cutoff values for these fractions, 32 density-separated fractions for the rhizosphere sample were then combined, generating light, middle, and heavy fractions (the bulk sample only contains light and middle fractions due to the absence of ^{13}C-enriched DNA) (**Figure 1.1**).

The light, middle, and heavy density separated fractions from the rhizosphere and bulk samples were sequenced and subjected to genome-resolved metagenomic analyses **(**Supplementary Table S1.1). From the rhizosphere middle fraction we assembled 210 Mbp of scaffolds larger than 1 Kbp. One especially large scaffold was assembled *de novo* and could be circularized. Local assembly errors were identified and corrected and three scaffolding gaps were filled by manual curation. Manual curation made use of unplaced paired reads that were mapped back to the gap boundaries to fill gaps. The complete, closed genome is 1.45 Mb in length with a GC content of 49.95%. We were able to recover a single chromosome and we detected no integrated phage or plasmids. The genome was most abundant in the rhizosphere middle fraction at 15x coverage, but was also present in the rhizosphere heavy and light fractions at ~3x normalized coverage. The genome had less than 1x coverage in the non-rhizosphere soil (Supplementary Table S1.1).

DNA buoyant density in a cesium chloride is a function of both the extent of isotopic enrichment and the GC content. Low GC has a lower buoyant density as compared to higher GC DNA (Schildkraut et al. 1962). The completed, closed genome has a lower GC content (49.95%) than the rest of the rhizosphere middle fraction assembly (average GC content of scaffolds larger than 1000 bp is 66%), which indicates ^{13}C may have been incorporated into the DNA. The rhizosphere middle fraction where the genome was mainly detected had a density of 1.737-1.747 g/ml. Given that natural abundance DNA with 49.95% GC content would have a density of ~1.71 g/ml (Creeth 2002), we estimate that the DNA from which the genome was assembled was at least 50% enriched in ^{13}C.

The genome has 1531 protein coding sequences (Supplementary Table S1.2) and a full complement of tRNAs (46 in total). The 5S rRNA, 23S rRNA and 16S rRNA genes are in a single locus that also includes Ala and Ile tRNA genes. Based on the sequence of the 16S rRNA gene, the genome was assigned to be a member of the Saccharibacteria phylum. The closest 16S rRNA gene sequences in NCBI are from the rhizosphere of *Pinus massoniana* **(Figure 1.2)** (Shi et al. 2014). The most closely related genomically described organism is Candidatus *Saccharimonas aalborgensis* from activated sludge with 84% identity across the full-length 16S rRNA gene (Albertsen et al. 2013). We propose the name *Candidatus* "Teamsevenus rhizospherense" for the organism described here, given the derivation of the genome from the rhizosphere. In accordance with the phylogenetic analysis, we renamed Subdivision 1 to *Candidatus* Soliteamseven because this genome is the first described Candidatus species of this clade. The representatives of this clade are mostly found in soil, therefore we propose the complete taxonomic descriptor: Phylum: *Candidatus* Saccharibacteria, Class: *Candidatus* Soliteamseven, Order: *Candidatus* Teamsevenales, Family: *Candidatus* Teamsevenaceae, Genus: *Candidatus* Teamsevenus, Species: *Candidatus* rhizospherense.

We calculated the GC skew and cumulative GC skew across the closed *T. rhizospherense* genome and found the symmetrical pattern typical for bacteria, with a single peak and trough indicative of the terminus and origin of replication (Supplementary Figure S1.2). This result both validates the accuracy of the circularized genome and confirms that Saccharibacteria use the typical bacterial pattern of bi-directional replication from a single origin to the terminus (as do some Peregrinibacteria, another group of CPR bacteria (Anantharaman et al. 2016)). The start of the genome was adjusted to correspond to the predicted origin, which lies between the DNA polymerase III subunit beta and the chromosomal replication initiator protein. It has a full set of

ribosomal proteins, except for L30, which is uniformly absent in CPR bacteria (Brown et al. 2015).

1.4.1 Biosynthetic pathways

The *T. rhizospherense* genome encodes a number of enzymes for the conversion of nucleotides to NMP, NDP and NTP and formation of RNA. In addition, we identified genes to phosphorylate G, C and U. However, the organism lacks the genes required to synthesize 5-phospho-alpha-D-ribose-1-diphosphate (PRPP). Further, it lacks essentially all of the steps for synthesis of nucleotide bases and the pathways that would convert PRPP to inosine monophosphate or uridine monophosphate. *T. rhizospherense* may have a novel nucleotide biosynthesis pathway, but this is unlikely as nucleotide biosynthesis remains highly conserved across domains (Peregrín-Alvarez et al. 2009) and the pathway can be recognized in some CPR bacteria (Anantharaman et al. 2016). Thus, we infer that *T. rhizospherense* did not *de novo* synthesize its nucleotides but rather acquired them from an external source. The genome encodes several nucleases, an external micrococcal nuclease, and an oligoribonuclease for the breakdown of externally-derived DNA and RNA (**Figure 1.3**). The mechanism for DNA and RNA import is unknown, as we did not identify nucleotide transporters. However, there are a number of transporters with unidentified specificity that could be involved in DNA or nucleotide uptake or the type IV pili could be responsible for this function. A large portion of the genome is dedicated to DNA and RNA repair mechanisms. There are twenty 8-oxo-dGTP diphosphatase genes that prevent the incorporation of oxidized nucleotides. These enzymes may be required given that the nucleotides may be scavenged from dead cells and could have accumulated extensive DNA damage. Access to damaged DNA may be a consequence of life in a mostly aerobic environment, a seemingly unusual condition for members of the Saccharibacteria phylum. All other complete genomes were found in mostly anaerobic environments and encode fewer genes with this function.

The *T. rhizospherense* genome does not encode the ability to synthesize any amino acids *de novo* from central metabolites. However, it encodes genes to generate amino acids from precursors (e.g., valine and leucine from 2-oxoisovalerate, isoleucine from 2-methyl-2-oxopentanoate, and histidine from L-histidinol phosphate) and to interconvert some amino acids (serine and glycine). We identified genes for proteases that could breakdown externally-derived proteins. No amino acid specific transporters were annotated, but several transporters of unknown function could import the amino acids. There is little evidence to suggest that externally derived amino acids are broken down for use in the TCA (the only TCA cycle gene identified is a fumarate reductase subunit) or other cycles.

T. rhizospherense appears unable to synthesize fatty acids, yet it encodes three copies of the 3-oxoacyl-(acyl-carrier protein) reductase, five copies of acyl-CoA thioesterase I, and two copies of SGNH hydrolase indicating the capacity for fatty acid hydrolysis and conversion. We found that the genome contains a number of genes for sequential steps in the glycerophospholipid metabolism pathway. *T. rhizospherense* may incorporate phosphatidylcholine (possibly derived from eukaryotes), 1,2 diacyl sn-glycerol-3P (from bacteria), or phosphatidylethanolamine (the main bacterial phosopholipid) and may be able to interconvert the compounds using a gene annotated as phospholipase D. We identified a putative phosphatidate cytidylyltransferase that

could add a head group to 1,2 diacyl sn-glycerol-3P forming CDP-diacylglycerol. This may be able to be converted into three products: phosphatidylglycerophosphate (via CDP-diacylglycerol-glycerol-3-phosphate 3-phosphatidyltransferase), or to cardiolipin (via cardiolipin synthase) or to phosphatidyl-1D-myo-inositol (via CDP-diacylglycerol-inositol 3-phosphatidyltransferase).

Interestingly, phosphatidyl-1D-myo-inositol is the precursor for generation of phosphatidylinositol mannosides, glycolipids that are decorated by a chain of mannose molecules and that are found in the cell walls of *Mycobacterium* (Kaur et al. 2009). There are several genes for the first step in phosphatidylinositol mannosides biosynthesis, which involves modification of phosphatidyl-1D-myo-inositol by addition of mannose. These include phosphatidylinositol alpha-mannosyltransferase (three copies) and a single copy of alpha-1,6-mannosyltransferase. Subsequently, a polyprenol-P-mannose α-1,2-mannosyltransferase (CAZy glycosyltransferase family 87) adds another mannose group. Other mannose additions may involve the three copies of dolichol-phosphate mannosyltransferase, which transfer mannose from GDP-mannose to dolichol phosphate a mannose carrier involved in glycosylation (Baulard et al. 2003). Thus, although *T. rhizospherense* appears to be unable to synthesize fatty acids, it appears to encode a number of genes that may be involved in the interconversion of membrane lipids, including phosphatidylinositol mannosides, if provided 1,2 diacyl sn-glycerol-3P.

1.4.2 Central metabolism and energy generation

Interestingly, the *T. rhizospherense* genome encodes a simple, two-subunit cellulosome that may be used to attach to and degrade plant or microbially-derived cellulose to cellobiose. Cellobiose is likely converted to D-glucose via one of 14 different glycosyl hydrolases. The genome also encodes genes for the production of D-glucose via breakdown of starch/glycogen and trehalose. The genome contains several genes for hydrolysis of 1,3-β-glucan, one of the most common fungal cell wall polysaccharides (Douglas 2001), to D-glucose. A significant portion of plant root exudation are sugars that could be fed directly into the *T. rhizospherense* metabolism (Sasse et al. 2018). D-glucose can be converted to D-glucose-6P and D-fructose-6P and fed into the glycolysis pathway. The genome lacks a key step in the glycolysis pathway the 6-phosphofructokinase gene, which takes fructose-6P to fructose 1,6-bisphosphate. However, the missing step is compensated for by the genes in the pentose phosphate pathway that convert fructose-6P to glyceraldehyde-3P which can then continue in the glycolysis pathway. This is a common workaround strategy in many members of the CPR, which frequently lack 6-phosphofructokinase (Wrighton et al. 2012). The product of the glycolysis pathway with the pentose phosphate pathway workaround is pyruvate. The pyruvate is then likely converted to D-lactate by the predicted D-lactate dehydrogenase protein, which would replenish the NAD^+ pool under fermentative conditions. Pyruvate appears not to be converted to acetyl-CoA, since the genome lacks pyruvate dehydrogenase and pyruvate ferredoxin oxidoreductase. Acetyl-CoA metabolism also appears to be lacking in the Saccharibacteria represented by the six other publicly available complete genomes.

The *T. rhizospherense* genome encodes a xylulose-5-phosphate/fructose-6-phosphate phosphoketolase that may convert D-xylose-5P derived from the pentose phosphate pathway to acetyl-P. The neighboring gene is annotated as an acetate kinase, which may convert acetyl-P to

acetate with the production of ATP. Thus, we predict growth via fermentation under anaerobic conditions. ATP also can be produced by a F-type H^+ transporting ATPase.

In glycolysis, NAD^+ is consumed to form NADH that can be regenerated via what appears to be a linear electron transport chain that includes a single subunit NADH dehydrogenase (*ndh*). We identified several predicted active site residues expected for function (Supplementary Figure 1.3A). Modeling revealed a close secondary structure match between the *T. rhizospherense* protein and the Ndh characterized from *Caldalkalibacillus thermarum* (Heikal et al. 2014). The large evolutionary distance between *T. rhizospherense* and *C. thermarum* likely accounts for differences in some active site residues.

The genome encodes a few genes in an incomplete pathway for production of quinone-based molecules, two of which are in multicopy (five copies of genes annotated as *ubiG* and two copies as *ubiE*). We suspect that quinone is scavenged from an external source. We identified genes for a cytochrome bo_3 ubiquinol terminal oxidase (*cyo*). Subunit I contains some functional residues as well as the residues that distinguish it from cytochrome-c oxidases (Supplementary Figure 1.3B), again this discrepancy in the functional residues may be due to the evolutionary distance between *E. coli* and Saccharibacteria (Iwata et al. 2000). These genes were also reported in a previous study (Kantor et al. 2013). The terminal oxidase requires heme to function, however only the final step of heme biosynthesis is predicted in the genome with a protein annotated as protoheme IX farnesyltransferase. We believe *T. rhizospherense* may scavenge heme from the environment as we propose it does for nucleotides, amino acids and quinones. Five FNR family transcriptional regulators may serve to detect O_2. If O_2 is available, it may be possible for electrons to be passed linearly from the Ndh to the cytochrome bo_3 ubiquinol terminal oxidase, which pumps four protons with the reduction of oxygen (Bekker et al. 2009).

Several predicted proteins such as a blue-copper protein and NADH-quinone oxidoreductase subunit L (*nuoL*)-related protein also may be involved in electron transfer, NuoL is known to contribute to membrane potential in Ndh systems (Mayer et al. 2015). The *nuoL* gene was found in the same region as the *ndh* and ATPase. Three cytosolic NADPH:quinone oxidoreductase genes were identified. These may reduce semiquinone (SQ) formed under conditions of high O_2 availability to prevent reaction of SQ with O_2 to form oxygen radicals (Patridge and Ferry 2006). We identified a novel protein that we predict may be involved in production of an electrochemical gradient. It contains two of the domains found in separate subunits of the Na^+-translocating NADH-quinone reductase (*na(+)-NQR*), contains an FeS cluster domain and is likely associated with the cytoplasmic membrane. Although speculative, we hypothesize that this protein is a part of the electron transport chain, converting NADH to NAD^+. One domain of the protein is similar to *na(+)-nqr* subunit F-like domain may oxidize NADH and transfer electrons to the iron-sulfur domain and the *na(+)-nqr* subunit B-like domain could form the Na^+ translocating channel (Steuber et al. 2014).

1.4.3 Other predicted capacities indicative of lifestyle

The genome lacks a CRISPR-Cas defense system, though one has been noted in the genome of a separate Saccharibacteria (Dudek et al. 2017). Additionally, we did not find any associated phage

or mobile elements, although there were a number of labeled phage contigs in the same sample which will be detailed in a separate publication.

We predict the capacity for twitching motility due to the presence of genes required for type IV pilus assembly and *pilT,* the twitching motility gene. Type IV pili may be involved in DNA uptake or attaching to other cells, root surfaces or solids. There are several genes for pseudo-pili, and an autotransporter adhesin that may also be involved in cellular attachment. Also, annotated were two CAZy carbohydrate-binding module family 44 genes for binding to cellulose and the capacity for biosynthesis of cellulose that could be used to attach to plant surfaces (Rodríguez-Navarro et al. 2007).

We identified genes encoding for laccase and pyranose 2-oxidase, which may be used for lignin breakdown or detoxification of phenolics, and genes for the detoxification of lignin byproducts, including 3-oxoadipate enol-lactonase and 4-carboxymuconolactone decarboxylase (Billings et al. 2015).

Interestingly, despite the small size of the genome and lack of many core biosynthetic pathways, we identified genes whose roles may be to modulate plant physiology, consistent with a close relationship with the plant. For example, we identified genes for the production of cis-zeatin (a plant hormone) from isoprenoid precursors (but a pathway for formation of the precursor isopentenyl-PP was not present, so the precursors are likely scavenged). The genome encodes a protein that appears to be cytokinin riboside 5'-monophosphate phosphoribohydrolase also known as "Lonely Guy," a cytokinin (a plant growth hormone) activating enzyme. We also found a gene encoding salicylate hydroxylase, which breaks down salicylic acid, a plant defense signaling molecule.

In addition to genes involved in plant interaction, we predict the capability to interact with other soil microbes. The genome contains N-acyl homoserine lactone hydrolase, a gene for quorum quenching of other soil microbes and a gene to form 3',5'-cyclic-AMP, which may be involved in intracellular signaling. We predict the presence of genes that confer resistance to bacterially produced antibiotics (beta-lactams, streptomycin, oleandomycin, methylenomycin A, vancomycin, and general macrolides) based on sequence similarity. A gene annotated as phosphatidylglycerol lysyltransferase may produce lysylphosphatidylglycerol, a membrane lipid involved in cationic antimicrobial peptide resistance (Sohlenkamp et al. 2007). We also found a gene that may confer resistance to fusaric acid, an antibiotic made by a common fungal pathogen of grass (Gagkaeva et al. 2017). This fungus was found to be growing in the rhizosphere (data not shown). Interestingly, there is a possible secreted toxin gene that encodes a 2,487 amino acid protein, it contains domains found in polymorphic toxins, rearrangement hotspot repeats, YD repeats, a PA14 domain and a galactose binding domain and the neighboring gene encodes an immunity protein (Jamet and Nassif 2015).

1.4.4 Comparative genomics

The *T. rhizospherense* genome is 28% larger than the largest reported complete Saccharibacteria genome, which is from an anaerobic bioreactor. It is 32.4% larger than the average size of all complete Saccharibacteria genomes (**Table 1.1**). By re-annotating and analyzing these

previously reported complete genomes we found that the *T. rhizospherense* genome encodes nearly 40% more unannotated genes as the other Saccharibacteria.

There are 130 *T. rhizospherense* functional annotations that were not found in any other Saccharibacteria (152 genes have these annotations, with some annotated to have the same function). A few appear to be involved in amino acid metabolism, others in transcriptional regulation, DNA repair, sugar metabolism, transport, non-homologous end-joining, three genes annotated as NADPH:quinone reductase and a dihydropteroate synthase for use in folate synthesis.

The genome appears to encode a nickel superoxide dismutase that may be used for oxidative stress response. The 49_20 scnpilot genome also encodes a gene with this function (but it is a Fe-Mn family superoxide dismutase). *T. rhizospherense* may have the ability to convert methylglyoxal to lactate, a two protein pathway (lactoylglutathione lyase and hydroxyacylglutathione hydrolase) that is absent in the other analyzed Saccharibacteria. This pathway is important in detoxification of methylglyoxal.

T. rhizospherense has a notably larger repertoire of carbohydrate active enzymes than occurs in the other Saccharibacteria, including 58 genes with 27 unique annotations. Included in the set and not found in other Saccharibacteria genomes are genes predicted to confer the ability to hydrolyze hemicellulose (AG-oligosaccharides and mannooligosaccharides), amino sugars (galactosaminide), and pectin (oligogalacturonides). The genome encodes a gene for transaldolase, a protein in the pentose phosphate pathway, which is absent in all other analyzed Saccharibacteria genomes.

T. rhizospherense is the only Saccharibacteria with a complete twin-arginine protein translocation system (*tatBC*). Four genes were predicted have a TAT motif, two genes of unknown function, a phosphatidylglycerol lysyltransferase (discussed above and not present in any other analyzed genomes), and a MFS transporter, DHA2 family, methylenomycin A resistance protein.

Of the *T. rhizospherense* genes involved in phosphatidylinositol mannoside production, phosphatidylinositol alpha-mannosyltransferase is found only in *T. rhizospherense* and 47_87 scnpilot and alpha-1,6-mannosyltransferase is found only in Candidatus *T. rhizospherense* and Candidatus *S. aalborgensis.* None of the other genomes encode the plant hormone related genes: "Lonely Guy" or salicylate hydroxylase.

The *T. rhizospherense* genome lacks several annotated genes found in most other Saccharibacteria. *T. rhizospherense* is unable to make (p)ppGpp, an alarmone that down regulates gene expression. Also lacking is dihydrolipoyl dehydrogenase, a subunit of pyruvate dehydrogenase (the function of which is unclear in these bacteria). Additionally not found are a *recJ* gene that encodes for a single stranded DNA exonuclease, a glutamine amidotransferase involved in cobyric acid synthase, the *murC* gene involved in peptidoglycan synthesis, and the *pilW* gene involved in pilus stability.

1.5 Discussion

The *T. rhizospherense* genome does not encode the capacity to generate the ribose backbone of DNA or bases and thus it must acquire nucleotides from external sources. Importantly, its DNA is labeled with ^{13}C that likely originated from $^{13}CO_2$ fixed by the plant over the six-week study period. There are several sources of labeled DNA that might have been available to *T. rhizospherense*. We suspect that it obtained DNA from bacteria that lived off plant-derived carbon. However, some of the *T. rhizospherense* DNA might have been from phage that killed bacteria that grew on the ^{13}C-labeled plant exudates. This is plausible because we identified labeled phage DNA (data not shown). Alternatively, *T. rhizospherense* may acquire DNA from fungi or the plant, but we consider this less likely given the predicted obligate association between Saccharibacteria and other bacteria.

As Saccharibacteria depend on extracellular nucleosides, they may have difficulty regulating the correct nucleoside concentrations. This is known to be a mutagenic condition and a possible reason for the large number of DNA repair mechanisms (Nordman and Wright 2008). Further, multi-copy enzymes that prevent incorporation of damaged nucleotides may reflect dependence of externally derived DNA that may have been damaged during residence in the soil environment. The lifestyle of *T. rhizospherense* is novel compared to those of other CPR in that it is predicted to have the capacity to grow aerobically. Consistent with this, it has a number of oxidative stress response genes. A close relationship with plant roots also may explain the presence of these genes, as plants release reactive oxygen species in response to pathogens and stress. It is the first member of the Saccharibacteria phylum to be described from the rhizosphere and it appears to be adapted to life there via an expanded genetic repertoire for plant hormone modulation and carbohydrate degradation relative to other Saccharibacteria.

It is possible that *T. rhizospherense* is a symbiont of soil dwelling bacteria, with a lifestyle analogous to that of the Saccharibacteria TM7x that was cultured as an epibiont of a mouth dwelling Actinobacteria (He et al. 2014, Bor et al. 2016). A similar association is indicated by the antibiotic resistance mechanisms that are encoded by *T. rhizospherense* that protect against compounds produced by Actinobacteria, methylenomycin A, oleandomycin, vancomycin and streptomycin. Additionally, the predicted ability for *T. rhizospherense* to produce phosphatidylinositol mannosides provides an intriguing connection between this organism and the clades of Actinobacteria that produce phosphatidylinositol mannosides, *Mycobacterium* being the best studied. In *Mycobacterium*, this cell wall component may protect the cell from antibiotics and act as a virulence factor during infection (Kaur et al. 2009).

We suspect that *T. rhizospherense* is a symbiont of a rhizosphere dwelling organism. However, we do not know the extent of labeling, other than DNA, of cellular components of *T. rhizospherense.* These components may have derived from plant root exudates and would therefore be labeled, but we cannot say for certain as we did not measure the isotopic composition of lipids, proteins and metabolites. The presence of machinery to degrade a variety of complex carbohydrates including cellulose and fungal cell walls suggests that this organism may not rely solely on plant exudates, but can also degrade biomass and necromass in the

rhizosphere. *T. rhizospherense* has many genes for attachment that could be used to connect to a microbial host or plant surfaces. It may directly interact with the plant, given genes for the modulation of signaling molecules zeatin, cytokinin, and salicylic acid.

1.6 Conclusions

To our knowledge, this is the first study to generate a complete, circularized genome from a soil dwelling bacteria *de novo* from any soil metagenome. It is the first stable isotope-informed genome-resolved metagenomic study in the rhizosphere and thus the first soil microbiome study to make use of stable isotope probing to track carbon from atmospheric CO_2 into the complete genome of a novel rhizosphere organism. The evidence indicates that ^{13}C was incorporated into the *T. rhizospherense* DNA through its use of externally derived nucleic acids that may have been synthesized by associated organisms that grew on plant exudates. *T. rhizospherense* is the first genomically described rhizosphere-associated member of the Saccharibacteria phylum and many of its predicted metabolic capacities distinguish this organism from related Saccharibacteria that live in anaerobic environments. The rhizosphere soil may alternate between aerobic and anaerobic conditions, and it appears that *T. rhizospherense* may have adapted to these cycles by having the capacity to perform fermentation and to grow aerobically using an unusual electron transport chain.

1.7 Figures

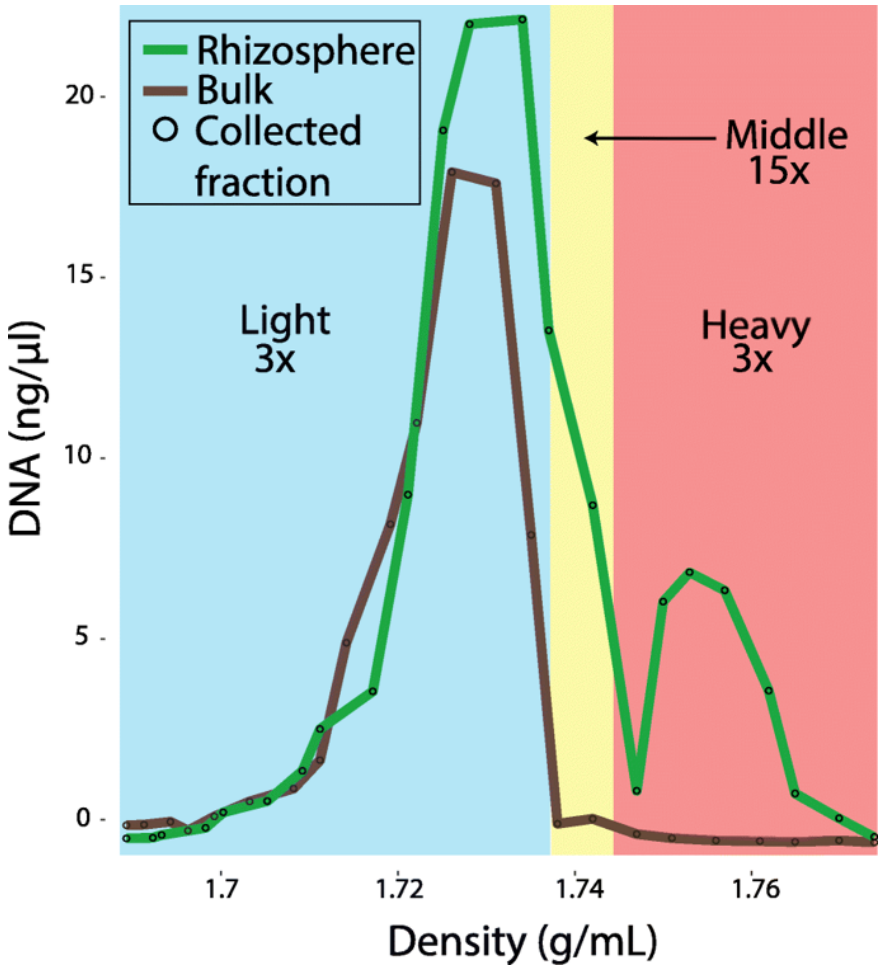


Figure 1.1 Stable isotope fraction determination. This plot shows the distribution of densities and concentrations of DNA extracted from week 6 rhizosphere and bulk soil following density centrifugation. The black circles on the curves represent individual fraction measurements. The three fractions are designated as light (blue shading), middle (yellow shading), and heavy (red shading). The top numbers indicate the normalized coverage of the *T. rhizospherense* genome in each fraction. The *T. rhizospherense* genome had < 1× coverage in each bulk soil fraction.

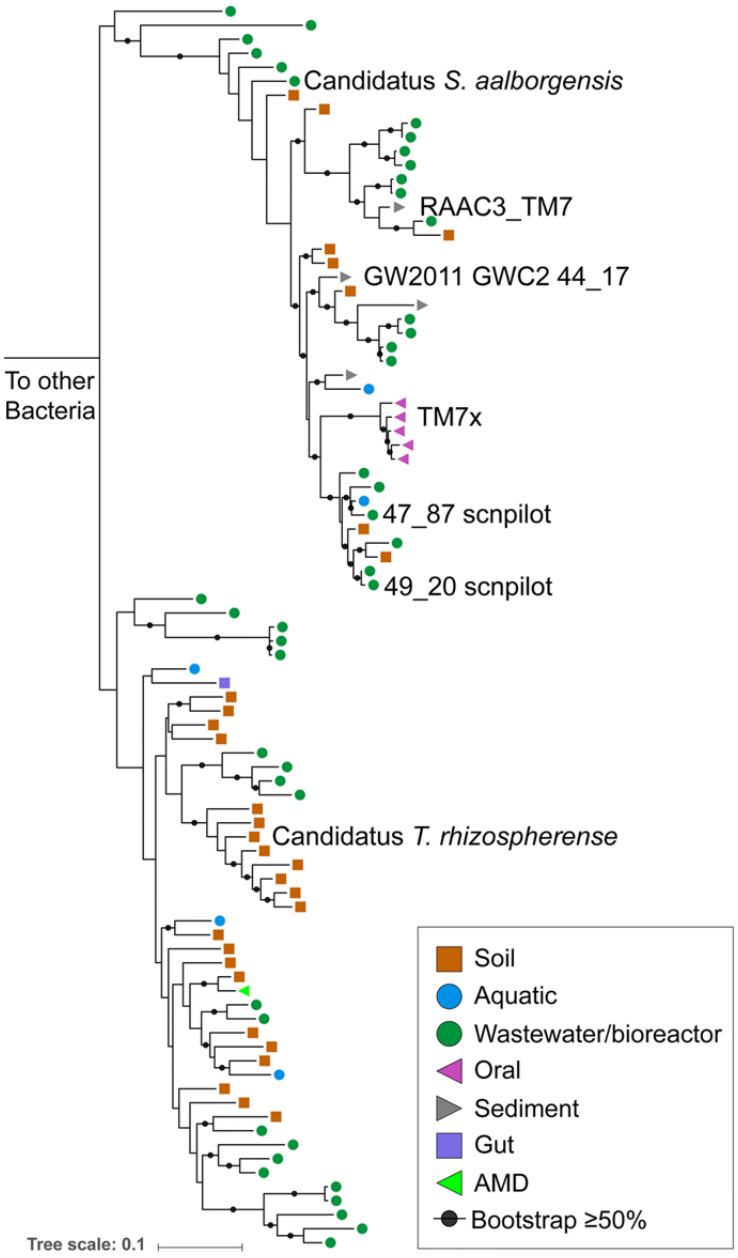


Figure 1.2 Phylogeny of Saccharibacteria based on 16S rRNA gene sequences. The maximum-likelihood tree shown was constructed from an alignment containing representative Saccharibacteria. Symbols indicate the environmental origin of the NCBI sequence. Named branches indicate the complete genomes included in this study. The tree scale bar indicates nucleotide substitutions per site. Bootstrap values ≥ 50% are indicated by black dots.

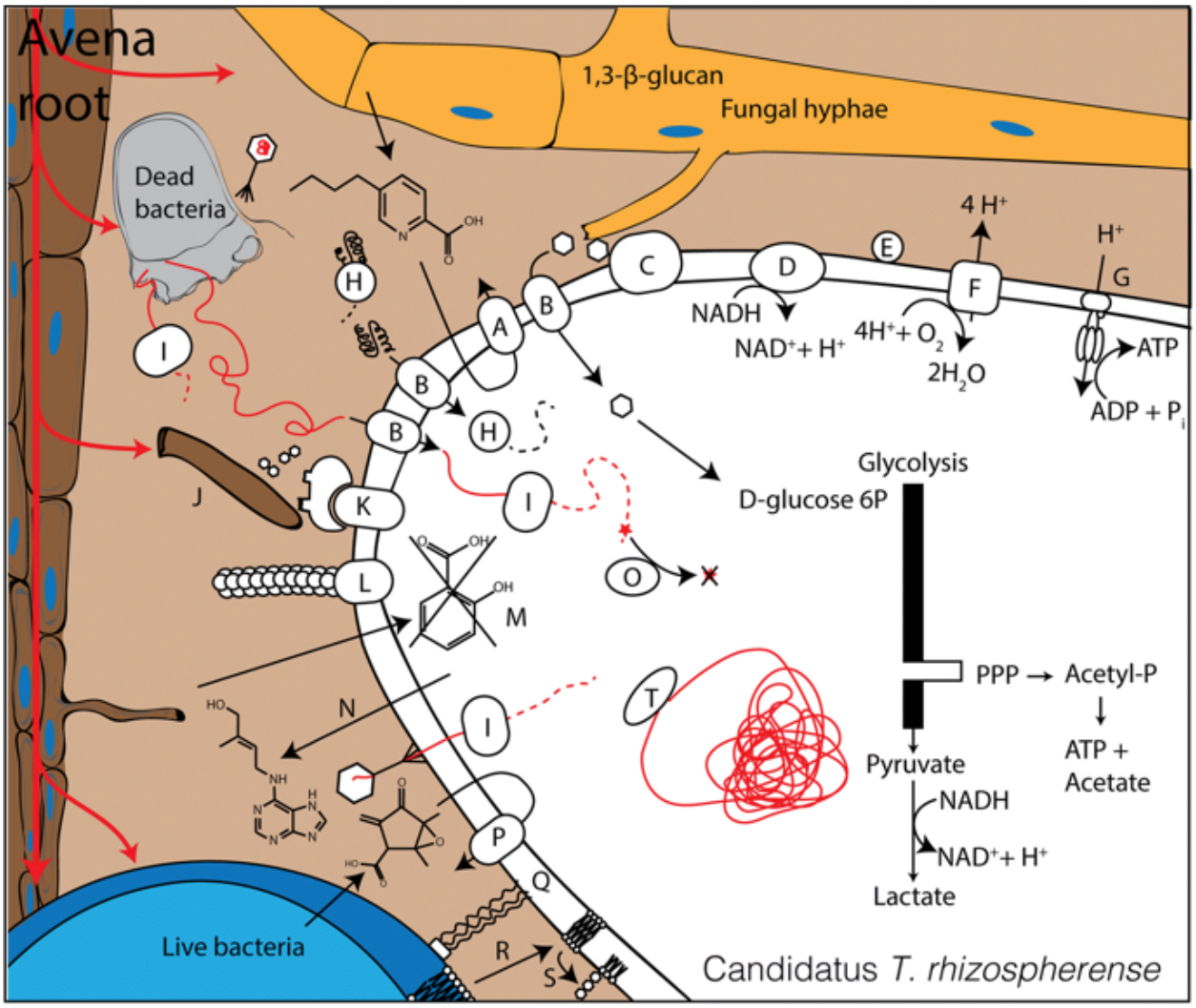


Figure 1.3. Cell diagram of T. rhizospherense. (A) fusaric acid resistance machinery, (B) unidentified importer, (C) glucan 1,3-beta-glucosidase, (D) NADH dehydrogenase II, (E) blue-copper protein, (F) cytochrome bo3 ubiquinol terminal oxidase, (G) F-type H+-transporting ATPase, (H) peptidase, (I) nuclease, (J) root hair, (K) cellulosome, (L) type IV pilus, (M) salicylate hydroxylase, (N) zeatin production, (O) removal of oxidized nucleotides, (P) various antibiotic resistance mechanisms, (Q) intercellular attachment, (R) scavenging of lipids, (S) production of phosphatidyl myo-inositol mannosides, (T) DNA repair machinery.

Table 1.1 Genome statistics for *T. rhizospherense* and other complete Saccharibacteria genomes

	Environment	Genome size (bp)	Gene number	Unannotated genes	Average gene size (bp)	Total coding sequence (bp)	Coding density (%)	CAZy genes	Unduplicated CAZy genes	Publication
Candidatus *T. rhizospherense*	Rhizosphere	1,450,269	1531	704	843.3	1,280,911	88	58	27	This study
47_87 scnpilot	Wastewater	900,471	918	376	896	822,552	91	30	17	[17]
49_20 scnpilot	Wastewater	904,897	933	362	891.5	831,756	92	28	14	[17]
GW2011 GWC2 44_17	Aquifer	1,038,683	1093	482	860.3	939,465	90	40	20	[1]
RAAC3_TM7	Aquifer	845,464	921	355	835.5	769,452	91	28	17	[8]
TM7x	Human mouth	705,138	711	241	919.2	653,574	93	30	17	[7]
Candidatus *S. aalborgensis*	Wastewater	1,013,781	1056	481	876.5	925,533	91	30	17	[16]
Difference from average		32.4%	33.2%	39.1%	− 3.7%	30.6%	− 3%	39.9%	31.7%	

For supplemental figures, tables, and information for Chapter 1, see https://doi.org/10.1186/s40168-018-0499-z

2 Metatranscriptomic reconstruction reveals RNA viruses with the potential to shape carbon cycling in soil

Evan P. Starr, Erin E. Nuccio, Jennifer Pett-Ridge,
Jillian F. Banfield and Mary K. Firestone

In review at *PNAS* July 2019

2.1.1 Abstract

Viruses impact nearly all organisms on Earth, with ripples of influence in agriculture, health and biogeochemical processes. However, very little is known about RNA viruses in an environmental context, and even less is known about their diversity and ecology in soil, one of the most complex microbial systems. Here, we assembled 48 individual metatranscriptomes from four habitats within a planted soil sampled over a 22-day time series: rhizosphere alone, detritosphere alone, rhizosphere with added litter, and unamended soil (four time points and three biological replicates). We resolved the RNA viral community, uncovering a high diversity of viral sequences. We also investigated possible host organisms by analyzing metatranscriptome marker genes. Based on viral phylogeny, much of the diversity was *Narnaviridae* that may parasitize fungi or *Leviviridae* which may infect Proteobacteria. Both host and viral communities appear to be highly dynamic, and rapidly diverged depending on experimental conditions. The viral and host communities were structured based on the presence of root litter. A clear time signature from *Leviviridae* and their hosts indicated that viruses were replicating. With this time-resolved analysis, we show that RNA viruses are diverse, abundant and active in soil. Their replication could cause host cell death, mobilizing carbon in a process that may represent an overlooked component of soil carbon cycling.

2.1.2 Statement of importance

The diversity and ecology of RNA viruses is severely understudied in complex environments. Here we investigate the diversity and community patterns of soil RNA viruses by analyzing assembled metatranscriptomes. We tracked RNA viral and host communities for 22 days in two soil environments central to carbon cycling, the rhizosphere and detritosphere. This work is an important step toward understanding the factors which drive RNA viral communities. The main hosts in our system may be Fungi and Proteobacteria, this is in contrast to the ocean where diatoms and other single cellular eukaryotes are primary hosts for RNA viruses. These results greatly expand the known diversity of viruses and contribute to understanding microbial interactions in soil with major implications for carbon cycling.

2.2 Introduction

We have much to learn about soil, the world's most diverse and enigmatic microbial habitat. With the advent of meta-omics techniques, the diversity, ecology, and impact of bacteria and archaea in soil is being rapidly revealed (Butterfield et al. 2016, Diamond et al. 2019). However, we have only begun to study viruses in soil despite community interest and their probable importance to soil functioning and carbon storage (Williamson et al. 2017). Recent work on double stranded DNA bacteriophage (phage) in soil has begun to explore their diversity and host interactions (Williamson et al. 2005, Emerson et al. 2018, Trubl et al. 2018). However, other viral members of the soil community, such as RNA-based phage and eukaryotic RNA viruses, are largely unknown. RNA viruses, those with genomes encoded on RNA, are less studied in environmental contexts because they are not captured in the more common DNA sequencing studies. In some systems, the impact of RNA viruses on community and ecosystem processes has been proposed to rival or exceed the impact of DNA viruses (Steward et al. 2013, Moniruzzaman et al. 2017). The majority of soil RNA viral work has been single-host focused, for agriculturally relevant crops (Andika et al. 2016) or crop pathogens (Zhang et al. 2014). Environmental RNA virus studies have focused on less complex or more tractable systems such as marine environments and the human and animal gut (Day et al. 2010, Culley et al. 2014, Gustavsen et al. 2014, Zeigler Allen et al. 2017). To our knowledge, no sequencing-based RNA viral community analyses have investigated soil.

Viruses are major players in biogeochemical cycling. However, much of what is known about viral impacts on elemental cycling comes from aquatic systems. Marine phages can lyse up to one-third of bacteria in ocean waters per day, releasing a huge amount of carbon (Fuhrman 1999, Suttle 2007, Lang et al. 2009). The released components include dissolved organic carbon (DOC) that is readily metabolized by heterotrophic bacteria, but largely inaccessible to eukaryotic grazers and higher trophic levels (Wilhelm and Suttle 2006). This phenomenon, termed "the viral shunt" (Wilhelm and Suttle 2006, Brum and Sullivan 2015), is thought to sustain up to 55% of heterotrophic bacterial production in marine systems (Winget et al. 2011). However, some organic particles released through viral lysis aggregate and sink to the deep ocean where they are sequestered from the atmosphere (Guidi et al. 2016). Most studies investigating viral impacts on carbon cycling have focused on DNA phages, while the extent and contribution of RNA viruses on remains unclear in most ecosystems. Several studies have indicated RNA viruses may outnumber DNA viruses in some instances, a possible hint to their influence in select ecosystems (Steward et al. 2013, Moniruzzaman et al. 2017, Stough et al. 2018). We suspect that, by analogy, lysis of organisms by RNA viruses may represent a large contribution to carbon flow in soils. The process of cell lysis and biomass carbon liberation likely stimulates heterotrophic consumption, with a substantial portion of the carbon lost to the atmosphere as CO_2. The liberated cellular debris may be protected from microbial access, through interaction with mineral surfaces and/or occlusion within soil aggregates (generation of mineral-organic associations) (Kawahigashi et al. 2006, Dondini et al. 2009). These processes could conceivably result in carbon stabilization and ultimately long-term soil carbon persistence, although in truth we have much to learn about the processes influencing soil carbon stabilization.

Understanding the diversity and ecology of soil viruses may contribute to advances in biotechnology. Viral genomes have been mined for biopesticides and self-assembling

nanomaterials (Glare et al. 2012, Wen and Steinmetz 2016). Viruses have been proposed, and used, as biocontrol agents for culling invasive organisms including fire ants, rabbits and moths (Harrison et al. 2014, Di Giallonardo and Holmes 2015, Valles et al. 2018). Viruses are also being investigated as biocontrol agents for devastating plant pathogens such as *Fusarium sp.*, *Botrytis cinereal* and *Rosellinia necatrix* (Wu et al. 2007, 2010, Zhang et al. 2014, Martínez-Álvarez et al. 2014, Wang et al. 2016, Osaki et al. 2016, Mu et al. 2018). Novel, environmentally derived viruses may be a source for new biotechnology tools and biocontrol agents.

Here, we used assembled metatranscriptomic data from a California annual grassland soil to reconstruct RNA viral genome sequences and constrain their possible hosts. We searched the assembled sequences for the RNA-dependent RNA polymerase (RdRp), a conserved gene found in all RNA viruses that lack a DNA stage (generally referred to as Baltimore type III-V RNA viruses). Although recent work tracing the deep evolutionary history of RNA viruses has merged viral families into supergroups (Koonin 1991, Shi et al. 2016a, Wolf et al. 2018), our focus was not on resolving deep phylogenetic viral branches. Rather, our objectives were to investigate fine scale diversity of RNA viruses, their putative hosts, and their possible functional importance in soil. Thus, we relied heavily on Shi et al. (2016), the most comprehensive metatranscriptomic study of RNA virus diversity so far, for taxonomic classification and to delineate viruses from previously undocumented clades.

The diversity of soil eukaryotes that may serve as hosts for RNA viruses remains largely understudied. Generally, RNA viruses are known to infect fungi, plants, animals and the many clades of single-celled eukaryotes, in addition to some Proteobacteria. Soil eukaryotic studies have relied heavily on primer based sequencing or visual classification, methods that can impart biases and miss novel organisms. Using the genomic information contained in our assembled metatranscriptomes, we identified many clades of eukaryotes without reliance on primers or microscopy. In contrast to many other viral studies which sequence a virome, the extracted viral particles, we sequenced whole samples which included viral genomes and possible host transcripts. This method allowed us to explore both extracellular viruses which would be sequenced in a standard virome and viruses within host cells, such as during the infection process or latent infections. We tracked both viral and eukaryotic communities in key soil habitats: the rhizosphere (soil influenced by the root), detritosphere (soil influenced by decaying particulate organic matter; bulk+litter), a combination of the two (rhizosphere+litter), and unamended soil (bulk) over time. Using a relatively short sampling time scale (3, 6, 12, and 22 days) allowed us to investigate viral and host community dynamics. Our research places direct constraints on the timescales for virus dynamics and provides the first genomic view of RNA viruses in soil.

2.3 Results

2.3.1 Experimental design

This analysis utilized data generated by a previous study on microbial niche differentiation in soil habitats (Nuccio et al. 2019). Briefly, wild oat (*Avena fatua*), an annual grass common in Mediterranean climates, was grown in microcosms with a sidecar, allowing us to track root age and sample directly from the rhizosphere (Jaeger et al. 1999). The experimental setup used microcosms, half of which included soil mixed with dried ground *A. fatua* root litter and the other half contained soil without litter amendment. All microcosms contained bulk soil bags,

which excluded roots. Once *A. fatua* was mature, roots were allowed to enter the sidecar and the growth of individual roots was tracked. We destructively harvested rhizosphere soil (and paired bulk soil) that had been in contact with the root for 3, 6, 12, and 22 days. In total, we sampled paired rhizosphere and bulk samples from four time points with two treatments (with and without litter), with three biological replicates, for a total of 48 samples for metatranscriptome sequencing. Using rRNA-depleted RNA, we sequenced a total of 408 Gbp (average of 8.7 Gbp per sample).

2.3.2 Eukaryotic RNA viruses

We used profile hidden Markov models (HMM) to search our assembled metatranscriptomes and found a total of 3,884 unique viral RdRp sequences (dereplicated at 99% amino acid sequence identity; AAI). This includes 1,350 predicted RNA bacteriophage (phage), viruses that infect bacteria, and 2,534 predicted viruses which infect eukaryotes.

Our eukaryotic viruses group into 15 major clades that span the majority of known viral diversity (**Figure 2.1**). Many were included into the supergroup-like clades defined by Shi et al. (2016). For the remainder, we constructed phylogenetic trees to define additional viral families.

Overall, in trees that include both existing and newly generated sequences, we noted a strong grouping of our RNA viral sequences into 'fans of diversity', much like the seed head of a dandelion. Many of these fans included a single reference sequence previously used to propose a new viral family. For example, in the Hepe-Virga, sequences grouped with the *Agaricus bisporus virus 16* (proposed family *Ambsetviridae* (Deakin et al. 2017)), in the Astro-Poty clade sequences group with *Bufivirus UC1* from wastewater(Wolf et al. 2018) and in the Partiti-Picobirna, sequences cluster with *Purpureocillium lilacinum nonsegmented virus 1* (Herrero 2016). We substantially expanded the *Barnaviridae* (Lueto-Sobemo superfamily) and *Mymonaviridae* (Tombus-noda superfamily) families (Afonso et al. 2016, Cañizares et al. 2018), which replicate in fungi, and the newly proposed *Zhaovirus* and *Qinvirus* families, which are found in invertebrates (Shi et al. 2016a).

We predict that fungi are the most common hosts for many of our newly reported RNA viruses (mycoviruses; supplemental figure 2.1). We are most confident when they fall into the *Barnaviridae*, *Megabirnaviridae, Quadriviridae,* and mitoviruses, groups currently thought to only infect fungi (Adams et al. 2017, Roossinck 2019). The most frequently encountered virus in our dataset, accounting for over 50% of the eukaryotic viral strains identified, came from the mitovirus genus in the *Narnaviridae* family. Mitoviruses are linear single stranded RNA viruses that replicate within fungal mitochondria and spread vertically through spores and horizontally through hyphal fusion (Wu et al. 2010, Silva et al. 2017). We suspect that most of the mitoviruses we detect are infecting fungi because, although some mitoviral sequences have been found integrated into the genomes of plants, they are frequently truncated and not transcribed (Silva et al. 2017, Nibert et al. 2018). Mitoviruses were also the most abundant viral clade in every sample (supplemental figure 2.2). One specific mitovirus was ~30X more abundant that the most abundant fungal taxon in the same sample, based on coverage normalized for sequencing depth and scaffold size.

Until recently, the *Narnaviridae* group, which includes mitoviruses, were thought to only encode an RdRp, but a recent discovery suggests some narnaviruses (only distantly related to mitoviruses) encode additional proteins, including capsids and helicases (Shi et al. 2016a). We identified some mitovirus genomes with possible additional genes which could play a part in infection efficiency or infection of new hosts. However, the majority of mitoviruses we identified contained only a single RdRp gene. We predicted several additional proteins on some near-complete mitoviral genomes (supplemental figure 2.3). These putative genes are small (average 79 amino acids) and novel, and functions could not be predicted for them. Sometimes the additional gene(s) are transcribed in the same direction as RdRp and in other cases they are transcribed in the opposite direction (supplemental figure 2.3).

We reconstructed many sequences for viruses that likely infect eukaryotes other than fungi. For the Picorna–Calici, hosts are likely vertebrates, insects, algae and plants, based on viral phylogenetic placements (Le Gall et al. 2008). In the Tombus-Noda tree, many of the RdRps group with umbraviruses, well recognized plant viruses. Other RdRp sequences group with sequences from complex environmental samples where the hosts are unknown.

RNA viruses from some previously defined superfamilies were conspicuously absent. We did not identify any soil RNA viruses belonging in the *Nidovirales*-like, *Reoviridae*-like or *Orthomyxoviridae*-like superfamilies. The absence of *Reoviridae* is interesting, as these RNA viruses can infect fungi (Hillman et al. 2003). We also did not confidently identify any *Ortervirales*, retroviruses that replicate with a DNA intermediate. We identified many reverse transcriptases, but none of the corresponding scaffolds encoded capsid proteins, so the sequences may be retrotransposons or fragments of retroviruses.

We classified two new clades within the Bunya-Arena viral super-family as novel, given that they exhibit sequence divergence comparable to that which separates known RNA viral families (see supplementary information). In both cases, the scaffolds only encode the polymerase with a Bunyavirus RNA dependent RNA polymerase domain. This genome structure is shared by *Hubei myriapoda virus 6* (Shi et al. 2016a) and *Ixodes scapularis associated virus 3* (Tokarz et al. 2018).

2.3.3 Eukaryotic hosts

As viruses replicate within their host, changes in virus abundance levels implicate introduction of new vectors or hosts, shifts in infected and uninfected host abundance levels, changes in infected host physiology or changes in host susceptibility to infection. To better understand the diversity and ecology of RNA viral hosts and carriers in soil, we used the mitochondrially encoded cytochrome C oxidase subunit 1 (Cox1) gene as a marker to define the present eukaryotic populations present (Hebert et al. 2003, Robba et al. 2006, Min and Hickey 2007, Damon et al. 2010, Robideau et al. 2011, Shi et al. 2017). However, mitochondrial transcription is determined by metabolic activity of a cell (Amiott and Jaehning 2006), and can be impacted by the switch to a symbiotic state (e.g. in arbuscular mycorrhizal fungi (Tamasloukht 2003)) or nutrient sensing and hormonal signals (e.g. in animals(Blomain and McMahon 2012)). Thus, the number of reads mapped to the Cox1 genes is determined by the number of organisms present and the organism's activity.

We identified 726 eukaryotic Cox1 sequences and clustered them at 98% AAI to approximate a species level view. The eukaryotes we discovered span the diversity of known soil organisms. Not surprisingly, sequences from *A. fatua* were the most abundant Cox1 transcripts in many samples. In some samples, the most abundant Cox1 were from an *Enchytraeidae* related worm or an *Amoebozoa sp.* Other samples were dominated by a mixture of fungi, Amoebozoa, Viridiplantae, and unknown eukaryotes.

Amoebozoa were the most diverse eukaryotic clade in this dataset, >25% of the identified eukaryotes, followed by species of fungi and unknown eukaryotes. The many unknown eukaryotes reflects the lack of environmental Cox1 sequences from micro- and meso-eukaryotes in public databases. For this reason, we reconstructed 18S rRNA gene sequences for classification, as they are better represented than Cox1 in public databases. However, the transcripts we analyzed had been depleted in bacterial and plant ribosomal RNA (via Ribo-Zero rRNA Removal kits) which may have influenced the composition and relative amount of other rRNA sequences. As such, the identified 18S rRNA sequences could not be used for abundance measurement or as a quantitative measure of diversity, only for identification of eukaryotes not identified using COX1. In total we identify 521 distinct species based on 18S rRNA sequences (after clustering at 98% nucleic acid identity to approximate species groups; **Table 2.1**). The results revealed the presence of Centroheliozoa, Malawimonadidae and Jakobida not identified on the Cox1 analyses and many more species of Heterolobosea, Euglenozoa and Rhizaria.

2.3.4 Eukaryotic virus and host ecology

The presence of root litter shaped the soil's eukaryotic community structure, as measured by Cox1 gene transcript abundance with *A. fatua* Cox1 sequences removed (**Figure 2.2A**; PERMANOVA R^2 0.17, P< 9.999e-05). Separation of the communities was evident by the time of the first sampling (3 days) and persisted to the end of the 22-day long experiment (supplemental figure 2.4).

We compared the abundances of eukaryotic Cox1 transcripts in rhizosphere+litter, bulk+litter, and rhizosphere samples to the bulk samples in order to identify species statistically enriched in each case (supplemental figure 2.5). The results showed that the presence of root litter and rhizosphere enriched for many Amoebozoa and fungi. However, enrichment patterns indicate that litter had a greater selective force on more individual species than the presence of a growing root.

The eukaryotic RNA viral community was influenced by some of the experimental treatments. Root litter had a significant effect on the eukaryotic RNA viral community (PERMANOVA P< 9.999e-05); the presence of growing roots had no detectable impact (PERMANOVA P< 0.07; **Figure 2.2B**). The distinction between added litter vs. no litter samples was evident within 3 days, implying that the viral and eukaryotic communities changed at similar rates. Differences persisted to the end of the 22-day long experiment. We observed the same patterns even when the *Narnaviridae* were removed from the dataset, indicating that these numerous, unusual host-bound viruses were not causing the separation of the eukaryotic viral community based on the presence of root litter (supplemental figure 2.6).

We identified a small number of eukaryotic RNA viruses enriched in certain treatments compared the number of eukaryote enrichments **(**supplemental figure 2.7**)**. This may be due to the substantial heterogeneity in viral abundance patterns, even within replicates. The RNA viruses displayed a higher level of microheterogeneity than the eukaryotes, 2% of the eukaryotic RNA viruses identified were present in only one sample. In contrast, transcripts for individual eukaryotes were more ubiquitous; only 0.1% of eukaryotes were found in only a single sample. The viral strains most frequently enriched relative to bulk soil were from the mitoviruses. This is unsurprising as the abundance of the obligately inter-mitochondrial mitoviruses is tied to the abundance and activity of the host and fungi responded strongly to the specific treatments.

To explore possible connections between the hosts and RNA viruses we created a co-occurrence network. The network was constructed from positive co-occurrences between the RNA viruses and hosts **(**supplemental figure 2.8). We noted the total number of edges, statistically significant undirected vertices between eukaryotes and RNA virus in the network, between the possible hosts and RNA viruses (**Table 2.1**), some eukaryotes and RNA viruses have multiple edges which explains the high numbers seen in some less diverse clades, such as the Nucleariidae.

RNA phage, potential hosts and ecology

Two families of RNA viruses are known to infect bacteria, the Cystoviridae which infect Pseudomonas sp. (Mäntynen et al. 2018, Elbeaino et al. 2018), and the Leviviridae which infect Gammaproteobacteria and Alphaproteobacteria (Klovins et al. 2002, Kazaks et al. 2011, Krishnamurthy et al. 2016). We used a marker gene approach to find RdRps for these RNA phage in the assembled transcripts. After dereplication, we identified 12 Cystoviridae RdRp sequences and 1,338 unique Leviviridae RdRp sequences. This is a significant increase in the known diversity of this group, as there are currently just over 200 Leviviridae RdRp sequences in public databases. Some of our sequences group with allolevivirus and levivirus, well-studied genera, but importantly, other novel Leviviridae sequences resolved new clades (**Figure 2.1**).

The specific Leviviridae taxa enriched in bulk soil are relatively phylogenetically novel (supplemental figure 2.8) and have unconventional genetic architectures. The most abundant Leviviridae sequence in 46 of 48 samples groups into a predominantly novel branch. The near-complete 4,668 bp genome of this abundant RNA phage encodes four, nonoverlapping genes—as opposed to the frequently overlapping and fewer genes found in other Leviviridae (supplemental figure 2.9 top). Like Enterobacteria phage M, it may also encode a +1 frameshift lysis protein inside its RdRp gene (Rumnieks and Tars 2012). The genome of a related Leviviridae is even more complex, with five genes and possibly two frameshift lysis proteins (supplemental figure 2.9 bottom). However, additional biochemical work will be needed to resolve the function of these predicted proteins.

We used ribosomal protein S3 (rpS3) phylogeny to identify possible hosts for Leviviridae viruses. All previously documented Leviviridae infect Proteobacteria, so we narrowed our analysis to these bacteria, which represent a possible pool of hosts for the Leviviridae, although other clades cannot be ruled out. In our samples, we identified 355 species of Proteobacteria, which differed in abundance between rhizosphere, bulk, rhizosphere+litter, and bulk+litter treatments (PERMANOVA P< 9.999e-05; supplemental figure 2.10A) after only three days of root growth (supplemental figure 2.10B**)**. The Leviviridae communities also diverged within

three days and changed over time (PERMANOVA P< 0.02; **Figure 2.4A**). The communities separated based on the presence or absence of root litter (PERMANOVA P< 9.999e-05), whereas the influence of growing roots was undetectable (PERMANOVA P< 0.09; **Figure 2.4B**).

2.4 Discussion

To our knowledge, no prior study has genomically investigated the ecology and diversity of RNA viruses in soils. Using assembled metatranscriptomes, we uncovered a vast diversity of RNA viruses (>3,000 sequences) from a single grassland soil and discovered several possible new families. Some viruses grouped phylogenetically with viral families previously proposed based on only a single member, indicating soils may hold much of the diversity of these understudied groups. The viruses we reconstructed came from nearly all known groups of RNA viruses. We identified hundreds of eukaryotes and bacteria that may represent RNA viral hosts or vectors that can pass infections among plants and other organisms. We know that the transfer of viral agents can have agriculturally relevant effects (Flores and Chapman 1968, Whitfield et al. 2015, Ruark et al. 2018) but there are likely many unexplored impacts on the soil community and carbon cycling.

Many of the RNA viruses we identified group phylogenetically with fungal viruses (mycoviruses), and fungi appear to be one of the main RNA viral hosts in this soil. The apparent dominance of mycoviruses in our dataset provides a striking contrast to aquatic systems, where *Picornavirales* dominate (Zeigler Allen et al. 2017). The most diverse mycoviruses in our samples are mitoviruses (*Narnaviridae*) which, despite their simplicity, can have significant impacts on fungal fitness and physiology (Rogers et al. 1987, Ghabrial and Suzuki 2009, Wu et al. 2010).

Previous studies of environmental RNA virus ecology have primarily addressed marine systems. The RNA viruses of oceans appear to be dominated by *Picornavirales*, thought to infect diatoms and other single-cell eukaryotes (Moniruzzaman et al. 2017, Zeigler Allen et al. 2017, Culley 2018). While our samples did contain *Picornavirales*, they did not appear to be dominant in the soil we sampled. In addition to *Picornavirales*, some aquatic systems contain plant, human and livestock RNA viruses that may have been introduced through runoff; we identified no viruses known to infect humans (Djikeng et al. 2009, Zeigler Allen et al. 2017). Some aquatic studies have noted viruses related to RNA fungal viruses, however not at the abundances or diversity we have in our soil (Moniruzzaman et al. 2017). Again, in contrast to the diversity of RNA phage that we found in soil, there is only a single report of an RNA phage from an aquatic ecosystem (Culley 2018). Compared to marine systems, soil virology, especially investigations of RNA viruses, is in its infancy.

In our soil, both eukaryotic RNA viral and eukaryotic communities were impacted by the presence of decaying root litter. By the first sampling time point, the eukaryotic and eukaryote-associated RNA viral communities were different in the presence and absence of litter. Saprophytic fungi appeared to respond favorably to the dead litter biomass. In contrast, dissolved organic compounds exuded by roots may be less accessible to non-fungal soil eukaryotes. Accordingly, roots had little impact on the non-plant eukaryotic community composition.

Addition of decaying plant biomass promoted the activity of detritivores, which likely immigrated into the litter from surrounding soil, increased transcription levels, grew, and/or germinated from spores. However, the shifts in the eukaryotic RNA viral community we observed are likely due to proliferation following increased activity of their hosts, although they may have also been transported by vectors.

Within the virus-host co-occurrence network the fungi had the greatest number and diversity of co-occurrence links with the *Narnaviridae*, which provides another piece of evidence that fungi may be major hosts for RNA viruses in this soil. Other eukaryotes had co-occurrence links with the *Narnaviridae* as well, but it is not possible to tell if these viruses are infecting or simply co-occurring with the eukaryote. The Amoebozoa had the largest number of co-occurrence linkages, including three with *Leviviridae*, which are unlikely to infect the Amoebozoa themselves but may be associated with a co-occurring bacteria. We know of only one RNA virus that infects Amoebozoa, a *Mononegavirales* (Mono-Chu) (Bird and Mc Caul 1976). These results suggest a possible link between Amoebozoa and the viral Bunya-Arena clade. The largest module (cluster of linages) contained a phylogenetically diverse array of eukaryotes and viruses including single virus types connected to multiple clades of hosts; this may represent a group of co-occurring entities which respond to similar environmental or biotic conditions (supplemental figure 2.11). Some of the individual pairs of co-occurring hosts and viruses may represent tenable connections; for instance, an Agaricales fungi co-occurred with a relative of Entoleucca bunyavirus 1, a fungal virus. Significant co-occurrence between a virus and a possible host is not evidence for infection; predator-prey interactions commonly exhibit asynchrony, patterns of co-occurrence can result from a shared niche, a eukaryote may be a consumer or a carrier of a virus, among other possibilities. Furthermore, our data may not be ideally suited for co-occurrence analysis as we are comparing RNA viral genome coverage to the expression of eukaryotic Cox1 and prokaryotic rpS3 genes. The results of this co-occurrence analysis exercise provides interesting directions for further work.

RNA phage, which replicate in bacteria, have receive relatively little study across all environments. Here, we provide the first glimpse into their ecology and diversity in soil. Given that over half of all the identified RdRp sequences in our dataset were *Leviviridae*, we encourage additional studies (in other soil types, land uses, tillage practices and cover vegetations) to assess whether soils generally host a large diversity of *Leviviridae*. As *Leviviridae* do not have a known lysogenic life stage, most of their population changes likely reflect replication in their hosts. Since Leviviridae communities and their hosts became distinctly different within the first 3 days of our experiment, we conclude that these RNA viruses infected and replicated within days. Infections over this timescale likely have dramatics effects on host communities, and thus soil ecology.

RNA viruses and RNA phage were heterogeneously distributed across samples, including replicates. In contrast, eukaryotes and bacteria appeared to have more even transcript abundances across samples. In combination, these observations suggest that virus and phage abundances are not solely determined by the presence of hosts, but also factors—such as sporadic blooms or variation in viral resistance levels in the host population —that lead to patchy viral distribution patterns in soil. These patterns may be analogous to those documented in marine virus blooms, though the exact cause of these events are still unknown (Seymour et al. 2006, Gustavsen et al.

2014). Also, viruses have different release rates and persistence times in soil—which may be impacted by a variety of factors, such as soil type, viral strain, host physiology, microsite of release, clay content and extracellular enzymes. This effect would be most pronounced for viruses with an extracellular life stage and less impactful on obligate intracellular viruses like the *Narnaviridae*.

In general, the magnitude of viral impacts on the soil carbon cycle is underexplored. Little research has been done on phage-induced bacterial lysis in soil and even less on viral-induced death of fungi and other eukaryotes, which can contain an equal or greater biomass compared to bacteria in some soils (Bar-On et al. 2018). In addition to lysis, viruses of fungi can have subtle yet profound effects on fungal biology, pathogenesis, mating success, toxin production, symbiotic relationships and other physiological effects (Zhang et al. 1998, Marquez et al. 2007, Rodríguez-Cousiño et al. 2011). These effects of viruses on fungal biology could have extensive impacts for fungal functioning and thus soil performance as a whole. In our samples, the high diversity and abundance of identified RNA viruses, combined with their dynamic population changes, indicates that there was substantial viral replication. We hypothesize that proliferation of these lytic RNA phage and RNA viruses will have substantial impacts on the form, abundance and distribution of carbon compounds in soil and on the biology and fitness of soil fungi. For example, lysis of host bacterial cells and viral-induced cell death of eukaryotes will cause release of dissolved low molecular weight carbon compounds. These will likely be quickly consumed by nearby bacteria and much of the carbon respired back to the atmosphere, mostly as CO_2. However, as happens when marine snow settles into the deep ocean, a portion of the cellular and viral debris and soluble carbon released may be stabilized in soil. For example, bacterial and eukaryotic lipids and polysaccharides could adhere to mineral surfaces or become occluded within soil aggregates.

2.5 Conclusions

By reconstructing soil metatranscriptomes drawn from multiple soil habitats and timepoints within the context of a controlled experiment we greatly increased the known diversity of RNA viruses. Phylogenetic analyses suggest that fungi are the most common hosts for RNA viruses in the studied grassland soil. While the diversity of hosts for RNA phage remain to be definitively identified, they likely also include Proteobacteria. Shifts in eukaryote, RNA phage and RNA viral abundances over a few day period reveal that entire soil communities can rapidly respond to altered resource availability. Our experiments indicate that the form of carbon inputs (root-derived low molecular weight C inputs versus macromolecular carbon compound in litter) may impact eukaryotic and bacterial abundance patterns, and that these in turn may be a major determinant of RNA viral and RNA phage dynamics.

2.6 Methods

2.6.1 Experimental design

Wild oat (*Avena fatua*) was grown in microcosms with a clear sidecar designed to allow access and visual tracking of the soil and rhizosphere (Jaeger et al. 1999, DeAngelis et al. 2009). Experimental soil was collected from Hopland Research and Extension Center (Hopland, CA.

USA). The soil is a fine-loamy, mixed, active, mesic Typic Haploxeralf, supporting dominant stands of *Avena* species (Placella et al. 2012, Whitman et al. 2018). Microcosms were filled with soil at field bulk density and seedlings were planted and grown for six weeks in the main chamber before the start of the experiment. Six days before the start of the experiment the divider separating the main chamber and the sidecar was removed and the sidecar was packed with soil. The litter-amended microcosms received 0.4g of dried *A. fatua* root litter mixed with 50g of Hopland soil. Bulk soil was contained in 1 μm mesh bags embedded into the microcosm, which allowed solutes to pass but not roots. Experimental design and soil edaphic characteristics are fully explained in a companion publication, which provides additional details regarding sample collection protocol, RNA extraction and processing, and sequencing (Nuccio et al. 2019); what follows is a brief description of the methods.

2.6.2 Sample collection

The ages of individual roots were tracked to collect rhizosphere soil which had been influenced by the root for 3, 6, 12 and 22 days. Three replicate microcosms were destructively harvested for paired rhizosphere and bulk soil. Rhizosphere soil was cut out from the rest of the soil along the edge of the root hair zone (<2 mm from the main root). Root sections and adhering soil were placed immediately in ice cold Lifeguard Soil Preservation Reagent (MoBio), soil was agitated off the root by vortexing for 2 min on medium speed and pelleted, bulk soil was treated in the same manner. Pelleted soils were frozen on dry ice and stored at -80°C.

2.6.3 RNA extraction

RNA was extracted from 0.5 g of frozen soil using a phenol-chloroform extraction protocol (Barnard et al. 2013). Then Qiagen AllPrep kits to were used to separate DNA from RNA. RNA was treated with TURBO DNase (Thermo Fisher Scientific) following the manufacturer's protocol to remove residual DNA, and was concentrated using an ethanol precipitation.

2.6.4 Sequencing

Metatranscriptome libraries were prepared and sequenced at the Joint Genome Institute. Ribosomal RNA was depleted from 1 μg of total RNA using the Ribo-Zero rRNA Removal Kit (Epicentre) for Plants and Bacteria. RNA was sequenced on the Illumina HiSeq sequencing platform utilizing a TruSeq paired-end cluster kit, v3, and Illumina's cBot instrument to generate a clustered flowcell for sequencing. Sequencing of the flowcell used a TruSeq SBS sequencing kit, v3, following a 2x150 indexed run recipe.

2.6.5 Sequence analysis

Reads were trimmed using Sickle (https://github.com/najoshi/sickle) and BBtools (https://sourceforge.net/projects/bbmap/) was used to remove Illumina adapters and trace contaminants. The reads from all 48 samples were assembled individually using IDBA-UD with default settings (Peng et al. 2012). Genes were predicted using Prodigal in the anonymous mode (Hyatt et al. 2010).

To find the host marker genes, Cox1 and rpS3, we used an HMM profile from Pfam (El-Gebali et al. 2019), Cox1 (PF00115), and HMMs for Bacteria and Archaea (https://github.com/AJProbst/rpS3_trckr) and searched using hmmsearch (-E 0.00001) from the

HMMER suite (Finn et al. 2011). The identified proteins were classified using both NCBI blast and by making trees. Once classified we predicted genes again using Prodigal in the single mode with the appropriate translation table. The Cox1 protein sequences were dereplicated and clustered at 98% AAI representing an estimated species level designation (Courty et al. 2008, Leray et al. 2013). Assembly errors were found and fixed in the scaffolds containing the Cox1 using ra2.py (https://github.com/christophertbrown/fix_assembly_errors) (Brown et al. 2015). The bacterial and archaeal rps3 genes was clustered at 99% AAI, representing species level differences (Sharon et al. 2015), using USEARCH (-cluster_fast) (Edgar 2010). The Cox1 and rpS3 trees were generated using references from NCBI, the protein sequences were then aligned using MAFFT v7.402 (Katoh and Standley 2013b) with the E-INS-i option on Cipres (Miller et al. 2010). Then alignments were trimmed for the conserved domain manually on Geneious and automatedly trimmed using trimAl (Capella-Gutiérrez et al. 2009, Kearse et al. 2012). The tree was made using RAxML (Stamatakis 2014) with the JTT protein substitution model (Jones et al. 1992). The trees were analyzed and figures were generated using iTOL (Letunic and Bork 2016). The 18s genes were found and the alignment was generated using ssu_tree.py (https://github.com/christophertbrown/bioscripts27) which searches for rRNA genes using an hmm method then the sequences were dereplicated and clustered at 98% nucleic acid identity, again representing a possible species level designation (Hadziavdic et al. 2014, Wu et al. 2015), and aligned using SSU-ALIGN (Nawrocki 2009). Assembly errors were found and fixed in the scaffolds containing the 18S using ra2.py (https://github.com/christophertbrown/fix_assembly_errors) (Brown et al. 2015). The tree was generated using the approach described above.

We identified the RNA viral scaffolds using a combined profile HMM approach, using HMMs from Pfam (El-Gebali et al. 2019) for different types of RdRps. The *Leviviridae* were identified using RNA_replicase_B (PF03431) and scaffolding errors were identified with ra2.py. There were no useable hmms for the *Cystoviridae*, so we generated our own. We aligned public RdRp sequences from the *Cystoviridae* using MAFFT v7.402 (E-INS-i) on Cipres and generated an hmm using hmmbuild from the HMMER suite with default settings (Miller et al. 2010, Finn et al. 2011, Katoh and Standley 2013b). The *Cystoviridae* hmm is publicly available on the figshare repository cooresponding to this manuscript. We used many publicly available HMMs to find the RdRp of eukaryotic RNA viruses (Mononeg_RNA_pol (PF00946), RdRP_5 (PF07925), Flavi_NS5 (PF00972), Bunya_RdRp (PF04196), Mitovir_RNA_pol (PF05919), RdRP_1 (PF00680), RdRP_2 (PF00978), RdRP_3 (PF00998), RdRP_4 (PF02123), RVT_1 (PF00078), RVT_2 (PF07727), Viral_RdRp_C (PF17501) and Birna_RdRp (PF04197)). The RdRp genes were initially classified using a blast and tree building method to determine the correct translation table to use for Prodigal in the single mode (Hyatt et al. 2010). For mitoviruses we used translation table 4, which to our knowledge is used by all mitoviruses. However, the additional mitoviral genes we predicted may have been incorrectly called if these genomes use a modified genetic code, as occurs in some fungi (Laforest et al. 1997, Forget et al. 2002). We examined gene predictions for indications of this (e.g., interruption of the RdRp gene) but as this phenomenon was not identified, no alternative codes were used for gene prediction. RdRp amino acid sequences were dereplicated and clustered at 99% amino acid identity using USEARCH (Edgar 2010). We used previously published alignments for many viral families (Shi et al. 2016a) and added our sequences and key references using the Mafft v7.402 with the --seed (Katoh and Standley 2013b) and E-INS-i options on Cipres (Miller et al. 2010). For viral clades

without published alignments or where the published alignment was inappropriate (*Fusariviridae*, *Narnaviridae*, *Leviviridae* and *Cystoviridae*) we generated our own alignments using reference sequences from NCBI and the same alignment and tree building steps as described above.

To identify lysis proteins in the *Leviviridae* genomes we used the Geneious ORF prediction to find all possible open reading frames (Kearse et al. 2012). The amino acid sequences were run through PSORTb v3.0.2 with the gram negative setting to find possible lysis protein(s) (Yu et al. 2010).

To obtain coverage values for the viruses we mapped against the entire scaffold containing the RdRp gene. For the presumed hosts we only mapped reads to the open reading frame, for the rpS3 and Cox1. Raw reads from all samples were mapped using Bowtie2 (--sensitive and --rfg 200,300 options) then reads were filtered for two mismatches using calculate_breadth.py (https://github.com/banfieldlab/mattolm-public-scripts/blob/master/calculate_breadth.py) (Langmead and Salzberg 2012). Coverage values were converted to read counts and normalized with DESeq2 (Love et al. 2014). Ordinations were generated from DESeq2 normalized count data in R. The data was ordinated using non-metric multidimensional scaling (R package: vegan) and significantly different clusters were determined using adonis (Oksanen et al. 2019). Using DESeq2 we determined the viral and eukaryotic habitat enrichments in the treatments (rhizosphere+litter, bulk+litter and rhizosphere) relative to bulk soil at each time point. We also conducted the opposite test, identifying enrichments in bulk soil by comparing normalized abundance to each treatment (rhizosphere+litter, bulk+litter and rhizosphere) at each time point. P values were corrected for multiple comparisons using a Benjamini-Hochberg correction (Benjamini and Hochberg 1995).

Network analysis was conducted using the DeSeq2 normalized coverage for the eukaryotic viruses and bacterial viruses and the eukaryotic Cox1 and bacterial rpS3. These values were filtered for coverage greater than 1,000. Spearman's rank correlation coefficient values were computed in R using the Hmisc package, the p-values were corrected for multiple testing (Benjamini and Hochberg 1995, Harrell 2019). The data was then stringently filtered to contain only organisms present in 20 or greater samples, a corrected p-value of 0.0001 or less, a correlation value of 0.7 or greater and removed all host-host correlations. This data was then fed into Cytoscape and a network was constructed using default settings (Shannon et al. 2003). Only clusters which contained at least one virus as the dependent variable and one host as the independent variable were analyzed and presented.

2.7 Figures

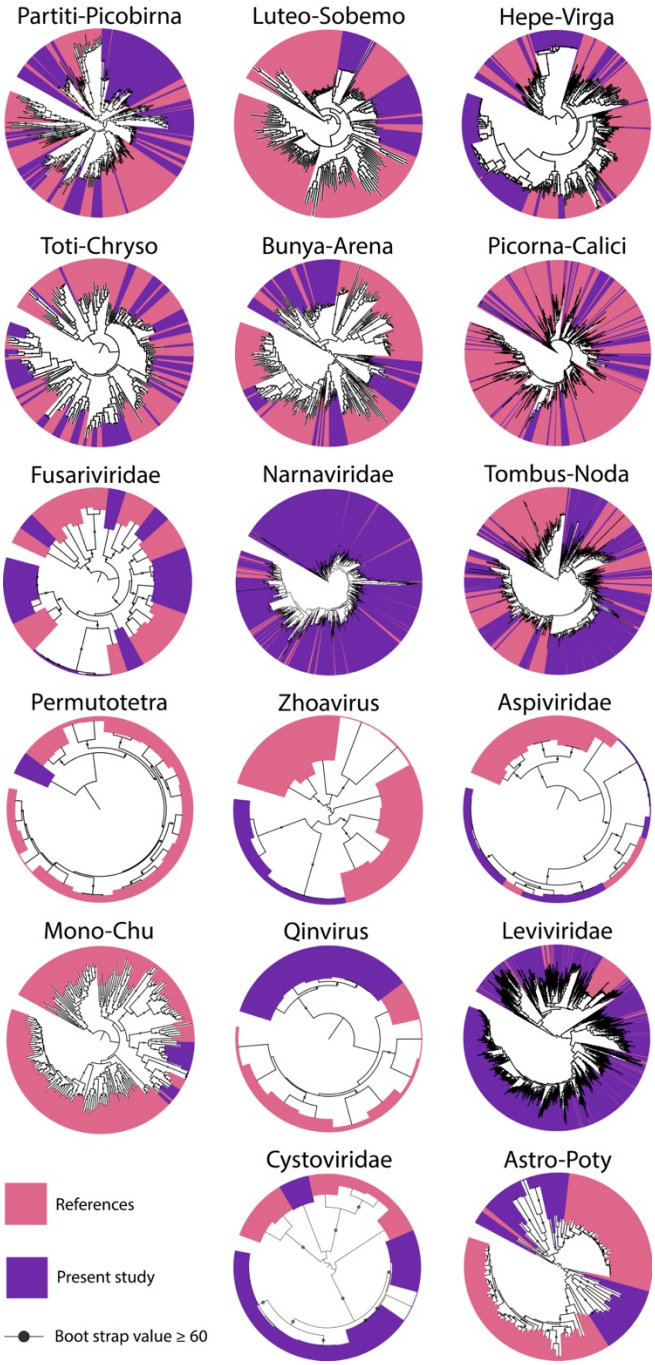


Figure 2.1. Phylogenetic trees representing clades of RNA viruses (based on RdRp) identified in our CA annual grassland experimental soil. Within each tree, the RdRp sequences we identified are colored purple and previously described sequences are in pink. Trees are all midpoint rooted. Trees with predicted fungal infecting clades are presented in Supplementary Figure 2.1.

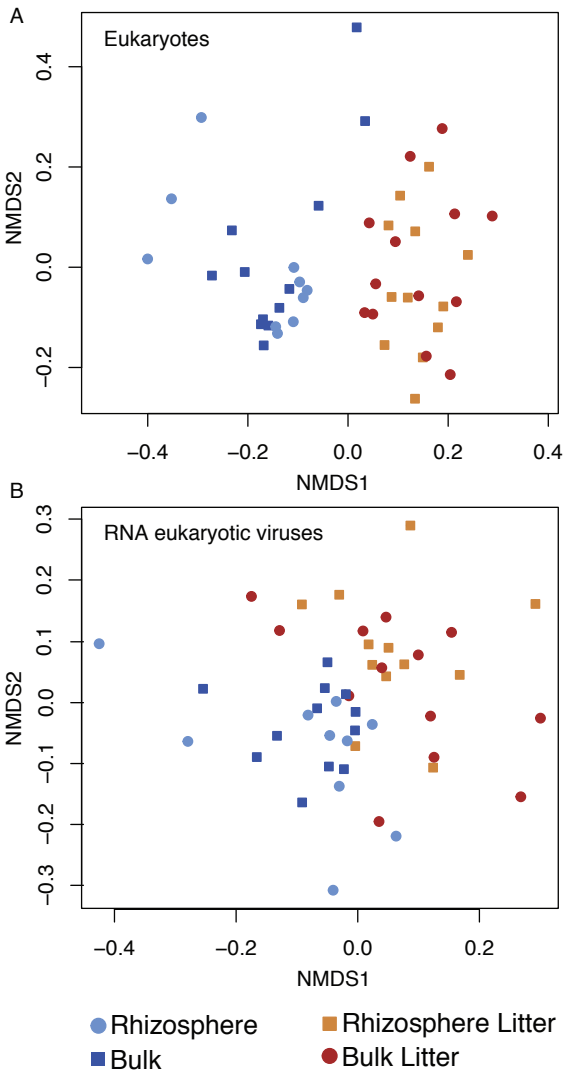


Figure 2.2. Nonmetric multidimensional scaling ordination of soil eukaryotic communities in response to four soil resource treatments, based on coverage of the Cox1 gene (*A. fatua* sequences removed) (A), and ordination of eukaryotic RNA viral communities with coverage calculated at the scaffold level (B).

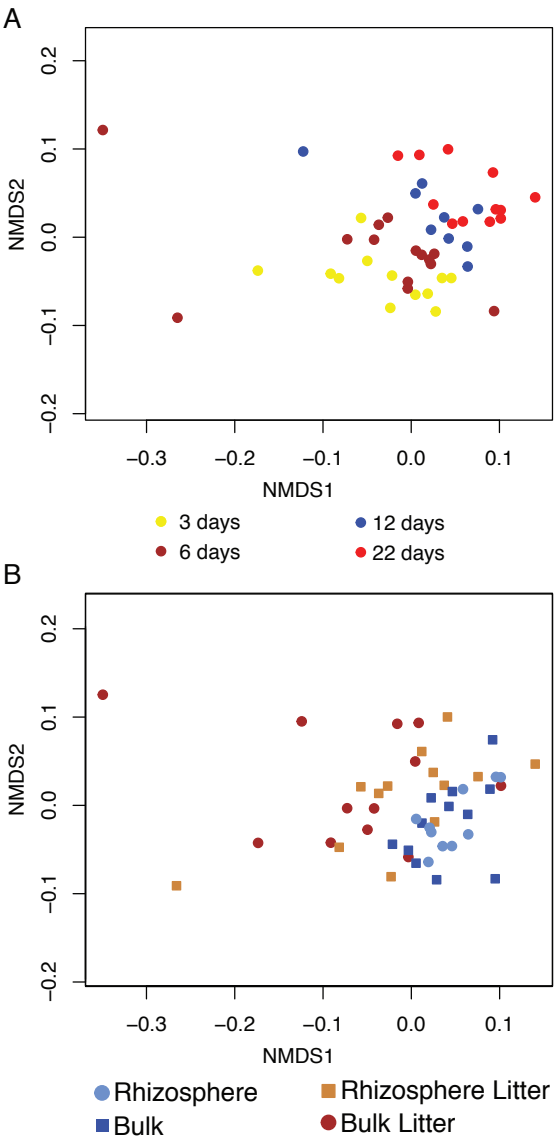


Figure 2.3. Nonmetric multidimensional scaling ordination of the *Leviviridae* community identified in an annual grassland soil exposed to four resource treatments, colored by time (A) and treatment (B).

Table 2.1. Number of eukaryotes identified based on marker genes and significant correlations between eukaryotes and RNA viruses. The number of statistically significant correlations is the number of positive co-occurrence edges between a eukaryote and virus derived from the network analysis (supplementary Figure 2.11). *Note that the absolute values may be skewed by rRNA depletion.

			Significant Co-occurrences							
Clade	Cox1	18S rRNA*	Narnaviridae	Bunya-Arena	Tombus-Noda	Luteo-Sobemo	Toti-Chryso	Picorna-Calici	Leviviridae	Parti-Picobirna
Amoebozoa	193	295	14	16	3	2			3	1
Fungi	174	33	15	11	2	1	2	2		
Unknown eukaryotes	108	14	1					1		
Metazoa	85	28								
Stramenopiles	55	17	11	6	3		2	1		
Alveolata	54	14								
Heterolobosea	11	41								
Nucleariidae	11	0	11	2		5				1
Cryptophyta	9	1								
Euglenozoa	7	28								
Viridiplantae	6	2								
Choanoflagellida	5	8	1	3	1					
Rhizaria	3	21								
Apusozoa	2	1								
Alveidia	2	0								
Ichthyosporea	1	0								
Centroheliozoa	0	11								
Malawimonadidae	0	6								
Jakobida	0	1								

For supplemental figures, tables, and information for Chapter 2, see https://www.biorxiv.org/content/10.1101/597468v1

3 Stable isotope informed genome-resolved metagenomics uncovers trophic interactions in rhizosphere soil

Evan P. Starr, Shengjing Shi, Steven J. Blazewicz, Benjamin J. Koch, Alexander J. Probst, Bruce A. Hungate, Mary K. Firestone* and Jillian F. Banfield*

3.1 Abstract

The functioning, health, and productivity of soil is intimately tied to the complex network of interactions in zone surrounding plant roots (the rhizosphere). We conducted a stable isotope-informed, genome-resolved metagenomic study to trace carbon from a common annual grass, *Avena fatua*, grown in a $^{13}CO_2$ atmosphere into the soil ecosystem. We collected paired rhizosphere and non-rhizosphere soil at six and nine weeks of plant growth and extracted DNA from soil that was then separated by density gradient centrifugation. The separate fractions of different densities were sequenced, assembled, and binned to generate 55 unique microbial genomes that were >70% complete. Evidence for close interactions between bacteria, micro-eukaryotes and plant roots includes the ability to modulate plant signaling hormones, abundant plant pathogenicity factors and production of cyanide and insecticidal toxins. We reconstructed eukaryotic 18S rRNA sequences and identified micro-eukaryotic bacterivores and fungi in the rhizosphere soil. In addition, we reconstructed two complete genomes for phage that were among the most highly ^{13}C-enriched entities in our study. CRISPR locus targeting connected a phage to a Burkholderiales host predicted to be a plant pathogen and a possible plant growth promoting *Catenulispora* may serve as the host for another phage. Thus, ^{13}C could be tracked from the atmosphere into plant roots, soil and through the rhizosphere food web.

3.2 Introduction

Plant derived carbon is a central hub that supports an intricate web of life in soil, among the world's most complex microbial ecosystems. Soil organisms are sustained primarily by the

carbon that is fixed by plants, released into the rhizosphere and eventually flows into the surrounding soil, feeding other organisms or becoming soil organic matter.

The soil ecosystem is characterized by interactions between organisms across trophic levels, which may direct the fate of plant derived carbon in soil. Soil ecology has been difficult to investigate because of the high physical and chemical heterogeneity and resulting vast scale of biological diversity. Much of the recent work on soil biology has been sequencing-based and focused on generating inventories of bacteria and archaea using 16S rRNA gene fragments (Delgado-Baquerizo et al. 2018, Bahram et al. 2018). Because all DNA is extracted together, it is possible to comprehensively profile soil communities, potentially with genomic resolution. This is important because genomes provide not only phylogenetic information but also a cache of functional predictions. However, relatively few studies have achieved genomic resolution of soil, largely due to enormous strain complexity and relatively even abundance levels (White et al. 2016, Butterfield et al. 2016, Kroeger et al. 2018, Diamond et al. 2019, Xue et al. 2019). Bacterial genomes encode a huge variety of previously unknown genes, including those that comprise biosynthetic gene clusters (BGC). Such biosynthetic pathways hold technological relevance but are also of fundamental interest from the perspective of soil ecology, as their products could mediate inter-organismal interactions, including offensive interactions via antibiotics, inter-organism and mineral-interactions via siderophores, and signaling compounds (Tyc et al. 2017, Crits-Christoph et al. 2018). We must note that the presence of annotated genes only indicates a possible capacity, not the true activity of the enzyme. Other genomically little-studied soil community members include phage and viruses. In fact, there has been a surge of interest in viral diversity and ecology in soil (Emerson et al. 2018, Trubl et al. 2018, Starr et al. 2019). Also present in soil are fungi, protists and larger organisms that have been documented via 18S rRNA gene sequencing and classical methods. Less understood are networks of interactions among these community members, although some studies have documented bipartite interactions, such as fungi-bacteria (Warmink and van Elsas 2009, Deveau et al. 2016), bacteria-phage (Ashelford et al. 2003, Emerson et al. 2018), plant-fungi (Bonfante and Genre 2010, Yang et al. 2015). Understanding of complex interactions across trophic levels remains a key challenge.

Lacking from almost all studies of soil to date is information about how nutrients flow among soil community members. Stable isotope probing (SIP) has been used for decades to identify interactions and trace the flow of elements through communities. SIP studies have been done in a variety of ecosystems from hot springs to the human gut and used many isotopes and labelled substrates (Egert et al. 2007, Godwin et al. 2014, Schubotz et al. 2015, Herrmann et al. 2017). Single substrate SIP studies can focus on the turnover and interactions around very specific compounds and have proven especially useful for investigating bioremediation (Uhlik et al. 2013). Alternatively, SIP studies can utilize carbon fixing organisms to generate biomass and complex compounds to investigate more general processes such as decomposition of litter or life in the rhizosphere (Murase et al. 2012, Haichar et al. 2012, Wilhelm et al. 2019, Hünninghaus et al. 2019). Stable isotopes can also be traced between trophic levels allowing the study of microbial predation and phage activity (Lee et al. 2012, Haig et al. 2015). Recent developments generate hypothetical quantitative isotopic enrichment for each taxon, termed qSIP. This is done by comparing the taxon-specific density for each sequenced entity in the labelled and unlabeled fractions (Hungate et al. 2015).

To better understand inter-organismal interactions in soil, initiated by the input of plant photosynthate we combined stable isotope probing and genome resolved metagenomics. We grew a plant, *Avena fatua,* for six and nine weeks in a $^{13}CO_2$ enriched atmosphere. The fixed plant carbon was presumably released to the surrounding soil community by exudation, decay of root biomass and pathogen attack. Over the course of the experiment the carbon is expected to move through trophic levels, including primary consumers, predators, parasites, saprotrophs and their phage and viruses. By separating the DNA based on density we could infer which organisms consumed the isotopically heavy plant-derived carbon and incorporated it into their genomes during replication. The assembly and binning of genomes enables analysis of the encoded proteins and better understanding of how organisms might interact with one another. Signatures of inter-organismal interaction include genes for the production of plant hormones and modulation of hormone concentrations, secretion systems, secondary metabolites and carbohydrate active enzymes. These predictions can be made without cultivation of individual organisms, which is important given that cultivation of the vast majority of soil organisms has not been achieved. The genomes also enable prediction of whether an organism is a plant growth promoting bacterium (PGPB) or a pathogen. Our approach also allowed us to reconstruct complete phage genomes and genetic markers for soil eukaryotes. We identified some community members linked via the flow of isotopically labeled carbon, demonstrating the potential for the combination of stable isotope probing and genome resolved metagenomics to unravel food webs in soil.

3.3 Results

3.3.1 Labelling

Two *A. fatua* plants were grown in a continuously regenerated $^{13}CO_2$ environment. The first plant was harvested after six weeks and the second after 9 weeks. We also collected a sample of soil prior to the start of the plant growth from a separately prepared microcosm and samples from soil regions from which roots were excluded (see Methods); these are referred to as “bulk” samples. After six weeks of growth, the plant shoots were highly labelled (~ 94 atom% ^{13}C) (Starr et al. 2018). DNA was extracted from the rhizosphere and bulk soil samples. We compared the density separation of the rhizosphere DNA to the bulk soil to DNA to determine un-enriched (light), partially ^{13}C-enriched (middle), and highly ^{13}C-enriched (heavy; Supplemental figure 3.1 and Supplemental table 3.1).

3.3.2 Soil community

Given that it is impossible to bin genomes for all organisms, yet relatively extensive reconstruction of genome fragments is possible, we used an assembled marker gene approach to approximate microbial community composition. For this analysis we used ribosomal protein S3 (rpS3), which is in single copy on bacterial and archaeal genomes and has been used to profile microbial communities for phylogeny and abundance (Sharon et al. 2015, Probst et al. 2018). From each sample we identified the rpS3 and dereplicated the sequences to the species level (99% nucleic acid identity) (Sharon et al. 2013). The resulting 314 distinct rpS3 sequences represent bacteria from both well characterized and understudied clades (**Figure 3.1**).

To study community dynamics across samples and fractions we mapped the reads from each sample to each scaffold containing the rpS3 gene. This analysis allowed us to determine if the microbes that took up the labelled carbon were a separate community, a subsection of the bulk organisms, or if the carbon labelled all organisms living in the rhizosphere. We use the coverage of the scaffolds as a proxy for the organism's relative abundance. The fractions and samples show a clear community separation (**Figure 3.2**). The bulk samples separate into a light fraction (tan) cluster and a heavier fraction (brown) cluster. This is due to DNA from the lower GC organisms occurring in the light fraction and the higher GC DNA making up the heavier fraction. The bulk light fraction and rhizosphere light samples group together, indicating a similar microbial composition. The rhizosphere middle fraction separates from the light fractions (both rhizosphere light and bulk light fractions) and the bulk soil heavier fraction as it contains DNA from the high GC organisms that did not incorporate the label and the lower GC organisms that were labelled sufficiently to increase the density of their DNA. Finally, the rhizosphere heavy fractions separate from all other fractions. The bacteria in this sample are almost entirely high GC organisms that incorporated ^{13}C into their DNA, thus are distinct from bulk samples comprised of DNA from unlabeled bacteria (Supplemental figure 1). There is no difference between the time points. Subtle compositional shifts may have occurred, but the resolution of the method is not sufficient to detect them. Taken together it appears that the carbon coming from the *Avena* root was taken up by a subset of the bacterial community that is distinct from the background bulk soil organisms.

We reconstructed and binned 55 partial bacterial genomes that were > 70% complete with <10% contamination, as measured by the inventory of 51 single copy genes (**Figure 3.3**; Supplemental figure 3.2). We recovered genomes from most of the phyla with rpS3 genes, but no Bacteroidetes, Firmicutes or Chloroflexi were represented by genomes (**Figure 3.1**). One Saccharibacteria genome was curated to completion, one of the first complete genomes from a soil metagenome (Starr et al. 2018).

In addition to bacteria we detected a number of eukaryotes. We identified eukaryotes from assembled 18S rRNA, in total we reconstructed 27 complete 18S rRNA gene sequences from soil eukaryotes (**Figure 3.3**; Supplemental figure 3.3). While this number does not scratch the surface for soil eukaryotic diversity, the assembled metagenomes provide complete 18S rRNA sequences without the primer bias inherent with tag based methods. Apart from the *A. fatua* and a single moss, the remaining 25 micro-eukaryotes fall into a variety of soil clades including Amoebozoa, Fungi, Metazoa (nematodes and rotifers), Rhizaria, and Alveolata.

We identified phage derived DNA in our samples. In total we reconstructed 10 complete, circularized phage genomes (**Figure 3.3)**. We infer that complete phage genomes derive from phage particles or phage in the process of replication, rather than integrated (pro)phage. Five of the phage may be lysogenic based on the identification of integrase genes. The remaining phage may be lytic based on the absence of annotated integrase genes or plasmid partitioning genes frequently found in non-integrating prophages (Canchaya et al. 2003, Stella et al. 2013). In addition to phage, we reconstructed 476 kbp of a fragmented Pandoraviridae eukaryotic virus genome.

We utilized the quantitative stable isotope probing (qSIP) method to determine a hypothetical atom percent excess (APE) of ^{13}C in each of the organisms. qSIP relies on tracking the shift in density, calculated by coverage in the different fractions, of a genome between the unlabeled (bulk) and labelled (rhizosphere) samples. We used the coverage of the scaffolds containing the 18S rRNA as a proxy for eukaryotic genome coverage. Because of the lack of replicates and the low number of fractions we chose a conservative cutoff for determining if a genome is labelled **(Figure 3.4)**. For each sequence we report the rank in ascending order of APE in **Figure 3.3**. Many of the phage we identified appear to be highly labelled **(Figure 3.4)**.

We found possible evidence for inter-organismal interactions encoded within the nearly all identified bacterial genomes. The possible machinery involved in interactions provide the foundation for a conceptual diagram of the carbon flow and multitrophic interactions in these soil samples (**Figure 3.5**).

3.3.3 Plant-soil community

Although the plant is likely the main source of labelled carbon for the soil community, other organisms may be able to fix CO_2. However, we found no evidence for carbon fixation pathways in the bacterial genomes and the lack of density shift in the bulk samples indicates that this effect was undetectable (Supplemental figure 3.1). Many of the bacteria encoding plant interaction mediating genes and pathways were the most labelled of the organisms, indicating an intimate relationship between the plant roots and bacteria.

Bacteria that closely associate with plants degrade plant hormones to avoid detection, which triggers a pathogen defense cascade. Many of the genomes encoded the ability to hydrolyze salicylic acid (**Figure 3.6**), a common phenolic plant hormone used in pathogen defense signaling (Wang et al. 2007). However, this protein can also be used to degrade other phenolic compounds (Balashova et al. 2001). Thus, we examined the nearby genomic regions for clues about the function of the gene. In one instance , the salicylate hydroxylase gene in a *Microbacterium* genome is surrounded by a variety of glycosyl hydrolases and esterases that act on plant cell wall polymers. This region of the genome may be devoted to digesting living plant material and preventing a defense cascade by hydrolyzing the hormone. Additionally, many of the rhizosphere dwelling organisms encode the ability to degrade another pathogen defense hormone, nitric oxide gas (Mur et al. 2006).

Bacteria, especially PGPB, promote plant growth through the production of hormones and other compounds. Two of the bacterial genomes we reconstructed encoded the pathway for producing indole-3-acetic acid, which increases plant growth and induces a variety of other physiological changes (**Figure 3.6**). Eight genomes encode 1-aminocyclopropane-1-carboxylate (ACC) deaminase (**Figure 3.6**), which prevents ACC from being converted to ethylene in the plant. This could be important because ethylene decreases growth by activating the plant stress response. However, ACC deaminase is also involved in the generation of propionate. In a *Streptomyces sp.* genome the ACC deaminase gene is surrounded by pectin lyases, pectinesterases, and the biosynthetic pathway for producing ectoine, a compatible solute common in PGPB (Czech et al. 2018). Bacterial produced volatile organic compounds (VOC) can act as growth promoting factors. The VOCs diffuse through soil and may promote plant growth and stimulate the plant

systemic defense, increasing their resistance to plant pathogens, although this pathway can also be involved in the anaerobic fermentation of glucose (Ryu et al. 2003, Farag et al. 2006). We identified 18 genomes with the pathways for the production of acetoin and/or 2,3-butanediol from pyruvate (**Figure 3.6**). Both the *Streptomyces sp.* and a *Catenulispora sp.* appear to be able to produce indole-3-acetic acid, ACC deaminase and 2,3-butanediol, meaning these bacteria could act as PGPB.

Microbes also promote plant growth through nutrient generation or mobilization. We did not identify any N_2 fixing pathways in the genomes or on the unbinned scaffolds. Some PGPB are able to mobilize phosphorus for use by the plant, but the genes cannot be distinguished from central metabolism and common enzymes. However, microbially produced phytases release phosphorus from phytate, a phosphorus storage compound common in soil but inaccessible to mature plants (**Figure 6**) (Alori et al. 2017). Several of the genomes encoded biosynthetic pathways to produce siderophores (Supplemental figure 4), which can be important for plant growth via provision of iron under iron limiting conditions. However siderophores are also important for plant pathogen virulence (Sharma et al. 2003, Pandey et al. 2017).

Some of the genomes appeared to encode secretion systems associated with bacteria-eukaryotic interactions. Six genomes encode many of the genes of type III secretion systems, which are known to be important in symbiotic colonization and infection of plants, fungi, single-celled eukaryotes, and animals (Nazir et al. 2017). We do not know the intended targets of the type III secretion systems because of the diversity of possible hosts. However, the type III secretion system encoded on SM_S39_Burkholderiales_62_29's genome is likely intended for injection into plant cells because the genome encodes over 50 type III effector proteins with sequence homology to known plant pathogen effector proteins (mostly from *Ralstonia, Pseudomonas* and *Xanthomonas*; **Figure 3.6**). We also identified 13 genomes with probable type IV secretion systems (**Figure 3.6**), which are used in conjugation and injection of proteins or DNA/protein complexes into eukaryotic cells (Green and Mecsas 2016).

In addition to bacteria-plant interactions we noted some possible plant-eukaryote interactions. We identified a *Fusarium sp*. in the rhizosphere samples that may have infected the *A*. *fatua* and obtained carbon directly from the plant, although the genome is not detectibly labelled. Either the fungus did not incorporate plant DNA or the label was diluted in the large cell volume and genome. We are unable to determine if this hypothetical relationship was beneficial or detrimental for the plant, as some *Fusarium* are now recognized as favorable endophytes (Karpouzas et al. 2011). The remaining fungi are from phylogenetically and ecologically diverse clades, and may act as saprobes or directly as pathogens.

3.3.4 Bacteria-Microeukaryotes

Several of the bacterial genomes encode systems for that may be intended for interacting with soil fungi. Ten of the partially complete genomes encoded fusaric acid resistance proteins, which protect from the antibiotic produced by *Fusarium*, a fungus identified in the rhizosphere (**Figure 3.6**). The *Variovorax sp.* genome appeared to encode three fusaric acid resistance modules, one was near an esterase and phospholipase C which hydrolyzes phosphatidylcholine, an important fungal phospholipid (Chen et al. 2018). In one *Streptomyces sp.* genome the fusaric acid

resistance module is near two glycoside hydrolases which may act on fungal cell walls. Included in this region is a ceramidase, which hydrolyzes glucosylceramides, a necessary metabolite for *Fusarium* pathogenesis and morphology (Rittenour et al. 2011). Another fungal antibiotic that these bacteria may encounter is penicillin, possibly produced by the *Eurotiomycetes* detected in the soil. Many bacterial genomes encoded penicillin resistance genes.

One possible indication of bacterial-fungal relationships is bacterial genes for decomposition of fungal compounds. A *Streptomyces* genome encodes 39 enzymes that target fungal biomass, 30 of which had an identifiable secretion signal indicating they are exported extracellularly (**Figure 3.6**). Most of the genomes encode some fungal biomass decomposing carbohydrate-active enzymes (CAZy) precited to act on chitin and other fungal polymers. Thus, these compounds may be a commonly utilized as carbon source for bacteria in soil.

We also detected sequences from larger eukaryotes, including animals like nematodes and rotifers and single celled eukaryotes such as rhizaria, alveolata and amoebozoa. Many of the micro-eukaryotes have bacterivorous lifestyles, and likely feed off bacteria living in the rhizosphere and surrounding soil. Nematodes and other micro-eukaryotes have genomes much larger than bacteria (the *Caenorhabditis elegans* genome is over 100 Mbp). Thus, in order for their genomes to become detectably labelled they would have to have consumed a large number of highly labelled bacteria. In addition, they would have to replicate their genomes to incorporate the label into their DNA. We found evidence for one possibly labelled micro-eukaryote. A Bdelloidea rotifer was the 32nd most labelled entity, however the coverage of the 18S rRNA containing scaffold is low which means stochasticity has a greater impact on the label quantification (**Figure 3.3**). If the genome is detectably labelled then the rotifer must have consumed numerous rhizosphere bacteria, since the genome could be as large as 244 Mbp (Flot et al. 2013).

There is some evidence for bacterially produced defenses against grazing. The majority of work on bacterial grazing defense has focused on nematodes but we assume similar strategies could be effective on other micro-eukaryotes. Bacteria often use biofilm production, specific secondary metabolites, and active infection to deter grazing (Matz and Kjelleberg n.d.). Biofilm production is common in soil bacteria but difficult information to infer from genomes. One indication of biofilm formation is the production of proteinaceous adhesins (Berne et al. 2015). Many of the genomes we investigated encoded adhesins, with two *Verrucomicrobia* and a *Variovorax* genome encoding the most (**Figure 3.6**). Biofilm production likely has many benefits apart from predation defense, including desiccation protection, phage defense, plant symbiosis, antibiotic protection, nutrient trapping and other purposes. Several genomes encode the pathway to produce hydrogen cyanide (HCN), a trait found in PGPB. HCN acts as a nematocidal agent and has deleterious effects on fungi (**Figure 3.6**) (Nandi et al. 2017, Kang et al. 2018). Some insecticidal toxins act on nematodes but have not been tested on other soil predators (Wei et al. 2003). Several of the genomes encoded the insecticidal toxin subunit TcC which is lethal to certain insects on its own (**Figure 3.6**) (Chen et al. 2014). However, one Acidobacterial genome encoded two complete Tc insecticidal toxin modules. In addition, in the same genome we identified an insecticidal crystal protein related to the bt toxin from *Bacillus thuringiensis* (**Figure 3.6**).

3.3.5 Bacteria-Bacteria

In addition to bacterial interaction, cooperation, and competition with eukaryotes there may be indications of inter-bacterial connections in soil. Previously, we showed that a rhizosphere Saccaribacteria accessed nutrients and nucleotides from other rhizosphere bacteria that took up plant-derived organic carbon (Starr et al. 2018). In other bacterial genomes we identified genes and systems that may mediate interspecies, inter-strain, or intercellular interactions. Some of the best characterized mediators of inter-organismal interactions are signaling molecules including acyl-homoserine lactones, autoinducing peptides, indoles, gamma-butyrolactones, and a variety of other compounds (Supplemental figure 3.4). We identified many of these in the genomes, although their roles in signaling cannot be verified. In addition, many genomes encode one or more quorum quenching genes, which may act either act to degrade self-produced quorum molecules or as a means to disrupt other bacterial species communications.

The high number of biosynthetic gene clusters identified in rhizosphere dwelling bacteria, especially polyketide and nonribosomal peptide biosynthetic gene clusters, may indicate production of antibiotics for competition for resources or space (Supplemental figure 4). Several *Streptomyces sp.* encoded many BGCs and several Burkholderiales and Acidobacteria genomes also encoded a high number of secondary metabolite clusters, though this is less surprising in light of recent findings in other soils (Esmaeel et al. 2016, Crits-Christoph et al. 2018). Some of the BGCs were related to known clusters, such as those that produce the broad spectrum antibiotic bicornutin from *Xenorhabdus budapestensis* and the sesquiterpene antibiotic albaflavenone, but many were novel. Some of the biosynthetic gene clusters may be involved in completion for plant colonization. For instance, one *Streptomyces sp.* encoded a three BGC near ten plant cell wall hydrolysis proteins, meaning these secondary metabolites may be coregulated with plant cell wall depolymerization. The compounds could aid defeating plant defenses or perhaps more likely protect the now soluble pool of sugars or open plant wound niche from other bacteria. Evidence for competition between closely related strains comes in the form of bacteriocins, antibiotics that act on closely related bacteria. These were found in nearly 40% of the genomes. Ten genomes encoded type VI secretions systems, which inject effectors into neighboring bacteria (**Figure 3.6**). Many genomes had multiple VgrG proteins, the tip of the needle and effector transporter, near large proteins with Rhs repeat domains which may function as bacterial toxins (**Figure 3.6**) (Ho et al. 2014). The Rhs proteins were not limited to the bacteria with type VI secretion systems and may be secreted in other ways.

3.3.6 Bacteria-Phage

Many of the phage we identified appeared to be highly labelled **(Figure 3.4)**. In the week 6 sample the two most labelled entities were phage **(Figure 3.3)**. The extensive labeling of these phage indicates that their genomes were replicated during the study period in bacteria living off of plant derived carbon. We focused on circularized phage genomes (not integrated into the host genome where they could become ^{13}C labelled through host growth) as these are most likely the product of active infection during our experiment.

We identified a Burholderiales as the possible host for one of the most highly labelled phage based on the match between a Burholderiales CRISPR-Cas spacer and the complete phage genome. The spacer hit the large terminase subunit with two mismatches (33 bp total). It appears

the phage may be capable of a lysogenic lifecycle because of the presence of a serine recombinase and a possible induction region consisting of a histone-like protein H-NS and an AT rich region flanking a lambda repressor-like gene.

The other phage-host connection we identified was based on a possible recent lateral gene transfer event. Again, the phage was highly labelled, being the 1st and 9th most labelled entity in week 6 and 9 respectively **(Figure 3.3)**. The probable host, a *Catenulispora sp*., encodes a glycoside hydrolase (GH25) which is very similar to one found in a complete phage genome, possibly indicating the phage may have horizontally acquired the gene. The two proteins share 77% amino acid identity across the protein and a phylogenetic tree of the proteins indicates that the phage and bacterial proteins are more related to one that to other publicly available bacterial sequences (Supplemental figure 5). Homologs exist in many *Catenulispora* genomes but not in closely related phage genomes, consistent with recent acquisition of the gene by lateral transfer. The acquired gene may be a lysis factor, as GH25 breaks down peptidoglycan (the possible host genome encodes nearly the full peptidoglycan biosynthesis pathway). The possible host does not encode any CRISPR-Cas or restriction modification systems. The phage may have a lysogenic lifecycle based on the presence of a gene annotated as a tyrosine recombinase and induction regulation genes, a lambda repressor-like gene and a gene containing a GntR family transcriptional regulator domain and an UTRA domain which regulates the activity of transcription factors in response to small molecules (Aravind and Anantharaman 2003, Gordon et al. 2008).

The remaining eight complete phage genomes cannot be linked to a specific host bin. Many encode DNA methylation genes that may protect the phage DNA from detection or destruction by host antiviral systems. Some phage have additional genes encoding complete restriction modification systems. Several phage genomes encode the ability to manipulate the composition of the bacterial nucleotide pool, including one phage that may synthesize 5-hydroxymethyldeoxyuridylate, a special nucleotide which replaces thymidine in the phage genome to evade antiviral defense systems. One phage may encode auxiliary genes, proteins related to those involved in glycosylation and biosynthesis of phosphatidylinositol mannosides. Of the 55 partial bacterial genomes, only two contained identifiable CRISPR-Cas systems, the Burholderiales described above and one Acidobacteria, which encoded two Cas type III-B systems and a type I-C system.

3.4 Discussion

By combining stable isotope probing with genome resolved metagenomics have begun to investigate possible interactions among soil dwelling organisms and their phage/viruses. We identified DNA from bacteria, phage, eukaryotic viruses, fungi, single celled eukaryotes, nematodes, and plants. Only through the generation of genome bins were we able to provide clues regarding the ecological role of these organisms and to develop hypotheses about the ecology of the system. Stable isotope probing may generate higher quality metagenomic assemblies and more complete genome bins. The labelled fractions may have contained less sequence diversity, since they were made up of most labelled microbes, this improved the

assemblies which in turn allows for better binning. Context provides valuable information when analyzing genome bins, as gene neighborhoods can constrain gene function. For instance, an ACC deaminase gene surrounded by cellulases should be interpreted differently than an ACC deaminase gene near propanoate metabolism genes, as the gene can serve as a connection between methionine and propanoate metabolism pathways.

A better understanding of holistic rhizosphere ecology may help us develop better agricultural techniques and biotechnological products. Some of the bacteria in our system may be acting as PGPB, through their ability to release nutrients, help in water stressed conditions or by repelling plant pathogens. By understanding the different trophic networks we may begin to develop multi-domain plant growth promoting consortia that could include phage of plant pathogens, fungi and microeukaryotes. From only 55 partial genomes we detected 400 BGC, many of these compounds likely play important roles is soil microbial ecosystems as antibiotics, antifungals, insecticides, herbicides, and may also have value as pharmaceuticals.

Based on our analysis it appears that many organisms we identified living in the rhizosphere, especially those that took up root derived carbon, encode the ability to interact with the plant itself and may span the spectrum from symbiont to pathogen. SM_S39_Burkholderiales_62_29 may be a plant pathogen, which may have led the assimilation of large amount plant derived carbon making this genome the 17th most labelled entity in week 6. The analysis of uncultured rhizosphere organisms may uncover previously unknown and unaddressed pathogens. We hypothesize that SH_S37_Streptomyces_71_13 may increase the plant growth by producing IAA, 2,3-butanediol and siderophores and may help the plan survive stressful conditions, especially drought, through the breakdown of ACC and biosynthesis of ectoine. *Variovorax sp* appears resistant to fusaric acid and may repel fungi and nematodes with the production of HCN, decrease plant stress response by regulating nitric oxide concentrations and improve the soil water holding capacity though the production of biofilm. SH_S37_Stenotrophomonas_67_7 encodes no clear plant interacting machinery, meaning either we have not identified the pertinent systems or perhaps it does not interact directly with the plant to acquire large amounts of plant derived carbon.

One of the most intriguing genomes we reconstructed is for an acidobacterium. This genome encodes many defensive capabilities, three CRISPR Cas systems, two complete Tc insecticidal toxin protein complexes, a crystal protein and 40 BGC. The insecticidal proteins may indicate either an active role in insect or nematode infection or a grazing defense strategy that may benefit one of the more abundant organisms in the bulk soil.

In both instances where we identified a putative phage-host connection, the phage is more highly labelled than the host, possibly because during infection new nucleotides are shunted to the phage genome. However, the mismatches between the bacterial CRISPR spacer and the phage genome may indicate that this bacterium is infected by a different but related strain of this phage. Alternatively, the phage genome may have acquired escape mutations since the spacer was acquired.

By tracing the movement of the carbon from the plant into the soil community we can begin to understand rhizosphere ecology, which in turn informs us about the carbon cycle in soil. The

possible interactions in rhizosphere have the ability to shape the flow and stabilization of carbon in soil. Lysis of bacteria by phage or interbacterial killing systems may release easily metabolized compounds that could be respired and returned to the atmosphere. Bacteria may contribute to soil carbon stabilization through the production of biofilm which may increase soil aggregation and PGPB could enable the plant to fix more CO_2, ultimately increasing the amount of carbon shunted into the soil. Only though a better understanding of the interdomain interactions present in soil can we begin to see its functioning as a whole.

3.5 Conclusions

Genome resolved metagenomics applied in the context of a SIP study generated insights into the possible interactive capabilities encoded on the genomes of bacteria, including the subset that could be linked to uptake of plant-derived carbon compounds. We provided an overview of the soil community, including eukaryotes and phage, and present inferences regarding how plant derived carbon moves through the soil system.

3.6 Methods

3.6.1 Labelling

Growing and labelling procedures have been described previously (Shi et al. 2015, 2016b, Starr et al. 2018). Briefly, soil was collected from the University of California Hopland Research and Extension Center (Hopland, CA, USA), from a field dominated by *Avena spp.* The soil is a fine-loamy, mixed, active, mesic Typic Haploxeralf (Placella et al. 2012, Whitman et al. 2018). Microcosms and plant growth conditions have been documented prior (Shi et al. 2015). A soil sample was collected after microcosm preparation abut before planting to represent our time 0 sample. A single *A. fatua* plant was grown in a labeling chamber maintained at 400 μL/L CO_2, with native CO_2 replenished with 99 atom% $^{13}CO_2$. After six and nine weeks microcosms were destructively harvested. The rhizosphere sample comprised of soil attached to the root after shaking and the bulk soil was collected from a root exclusion mesh bag (1 μm). Rhizosphere soil was washed off the root and DNA was extracted from 0.5g of soil phenol:chloroform extraction (Shi et al. 2015).

3.6.2 Density separation

We used a CsCl density gradient centrifugation to separate the DNA based on density. Methods have been described previously (Blazewicz et al. 2014). Briefly, 5.5 μg of DNA was incorporated into a gradient buffer with a density of 1.735 g/mL. The solution was spun in ultracentrifuge tubes (Beckman Coulter Quick-Seal, 13 × 51 mm) in an Optima L-90K ultracentrifuge (Beckman Coulter, Brea, California, USA) using a VTi65.2 rotor at 44,000 rpm (176,284 RCFavg) at 20 °C for 109 h with maximum acceleration and braking of the rotor to maintain the integrity of the density separations. The gradient was then gently separated in to ~32 fractions using a syringe pump delivering light mineral oil. Each fraction (~144 μL) was measured for density using an AR200 digital refractometer (Reichert Inc., Depew, New York,

USA). Then DNA was precipitated and quantified as previously described (Blazewicz et al. 2014). Fractions were then binned based on density and by comparison between the rhizosphere samples and the associated bulk soil (light = 1.692–1.737 g/mL; middle = 1.738–1.746 g/mL; heavy = 1.747–1.765 g/mL; Supplemental figure 3.1 and Supplemental table 3.1).

3.6.3 Sequencing

Each fraction was sequenced at the UC Davis Genome Center on an Illumina HiSeq 3000 (Illumina Inc., Hayward, California, USA) with paired-end libraries prepared with the Kapa Hyper protocol and a read length of 150 bp.

3.6.4 Sequence preparation and analysis

Reads were trimmed using Sickle (https://github.com/najoshi/sickle) and BBtools (https://sourceforge.net/projects/bbmap/) was used to remove Illumina adapters and trace contaminants. Each sample was assembled individually using IDBA-UD (-step 20, -maxk 140, -mink 40) (Peng et al. 2012). Only scaffolds larger than 1000 bp were included in future analyses. Genes were predicted using Prodigal (Hyatt et al. 2010). The ORFs were the annotated using a combined approach. Sequence similarity searches were preformed using USEARCH (Edgar 2010)again UniRef100 (Suzek et al. 2007), Uniprot (Magrane and Consortium 2011) and the KEGG (Ogata et al. 1999) databases. Additional gene annotations were done using HMMs that were constructed based on KEGG Orthologies. All proteins assigned to a KO were clustered using MCL (Van Dongen 2008) with inflation parameter (-I) of 1.1, based on global percent identity. Clusters were aligned using MAFFT v7 (Katoh and Standley 2013b), and HMMs were constructed using the HMMER suite (Finn et al. 2011). Protein domain level analysis was conducted using InterProScan (Zdobnov and Apweiler 2001). Carbohydrate active enzymes were identified using dbCAN (Yin et al. 2012). Secondary metabolite clusters were found using antiSMASH (Medema et al. 2011). tRNAs were predicted using tRNAScan-SE (Lowe and Eddy 1996). The 16S and 18S rRNA sequences were found and aligned using ssu_tree.py (https://github.com/christophertbrown/bioscripts27). Eukaryotic 18S rRNA genes were the genes were dereplicated and clustered at 98% nucleic acid identity representing a possible species level designation (Hadziavdic et al. 2014, Wu et al. 2015), aligned using SSU-ALIGN (Nawrocki 2009) and trees were generated using RAxML on CIPRES (Miller et al. 2010, Stamatakis 2014). Genomes were binned using a combined approach. We used abawaca (https://github.com/CK7/abawaca), MaxBin2 (Wu et al. 2016), and MetaBAT (Kang et al. 2015), the best bins were chosen using DAS Tool (Sieber et al. 2018). Further genome curation and binning was done manually on ggKbase (http://www.ggkbase.berkeley.edu). Bins were dereplicated to species level based on rpS3 and 16S rRNA sequence phylogeny.

The rpS3 genes were identified, dereplicated to species level, and the longest scaffold was chosen using rpS3_trckr (https://github.com/AJProbst/rpS3_trckr). Each sample was mapped to each scaffold using Bowtie2 (--sensitive and --rfg 200,300) and the reads were filtered for two mismatches using and the coverage was calculated using calculate_breadth.py (https://github.com/banfieldlab/mattolm-public-scripts/blob/master/calculate_breadth.py) (Langmead and Salzberg 2012). The coverage values for the rpS3 scaffolds were normalized for total read depth from the corresponding sample. The principal coordinate analysis was conducted in the R programming environment with the vegan package (R Core Team 2014, Oksanen et al.

2019). The R script is publicly available (Weinmaier et al. 2015). RpS3 sequences were aligned using MAFFT v7.402 (Katoh and Standley 2013b) with the E-INS-i options on Cipres (Miller et al. 2010). Trees were visualized and figures generated using iTOL (Letunic and Bork 2016).

Phage genomes were found using VirSorter and manually on ggKbase (Roux et al. 2015). Phage genome completeness was checked by mapping reads and visualizing on Geneious (Kearse et al. 2012). Circularized phage genomes will have reads paired across the entire span of the scaffold without repetitive elements on the end of the scaffold which could cause long paired reads instead of a circular sequence. CRISPR spacers were found using CRISPRDetect (Biswas et al. 2016).

3.6.5 qSIP

We estimated the atom percent excess (APE) ^{13}C enrichment for each taxon following the procedures detailed in Hungate et al. (2015), with the following adjustments for metagenome-assembled genomes instead of 16S rRNA genes. Previous qSIP experiments have relied on amplicon sequencing of the 16SrRNA gene to estimate the relative abundances of bacterial and archaeal taxa. Those relative abundances were then converted to estimates of absolute abundance by multiplying by the total number of 16S rRNA gene copies using the universal 16S rRNA primer for qPCR for each density fraction in each replicate gradient. Here, we used the relative coverage (i.e., coverage normalized for read depth) of metagenome-assembled genomes as a proxy for relative abundance and total DNA concentration in place of total 16S copies in order to calculate a metric of abundance (y) for each taxon (i) in each density fraction (k), of each replicate (j) as:

$$y_{ijk} = p_{ijk} \cdot f_{jk}$$

where p is relative coverage and f is total DNA concentration. This metric of taxon-specific abundance (y) is imperfect because total DNA concentration in the rhizosphere presumably represents many more taxa beyond the metagenome-assembled genomes and eukaryotic scaffolds. In addition, the input sequences varied widely in size, and therefore contributed unequally to bulk DNA concentration. To account for this, we also calculated a 'scaled' metric of abundance (y') by dividing by the genome size (B, with units of base pairs) of each taxon (i):

$$y'_{ijk} = \frac{\left(p_{ijk} \cdot f_{jk}\right)}{B_i}$$

Analyses were conducted separately for the week 6 and week 9 data. For each time point, we used the bulk soil data from that week as the unlabeled treatment. Because the *Avena* ^{13}C labeling experiment did not contain true replicates, we could not generate formal estimates of uncertainty in ^{13}C enrichment or abundance for the metagenome-assembled genomes. Therefore, to distinguish enrichment values from zero, we identified the lowest negative enrichment value (-13.9% APE ^{13}C) and assumed that to be our detection limit. All taxa with enrichment values < +13.9% were assumed to be no different from zero.

3.7 Figures

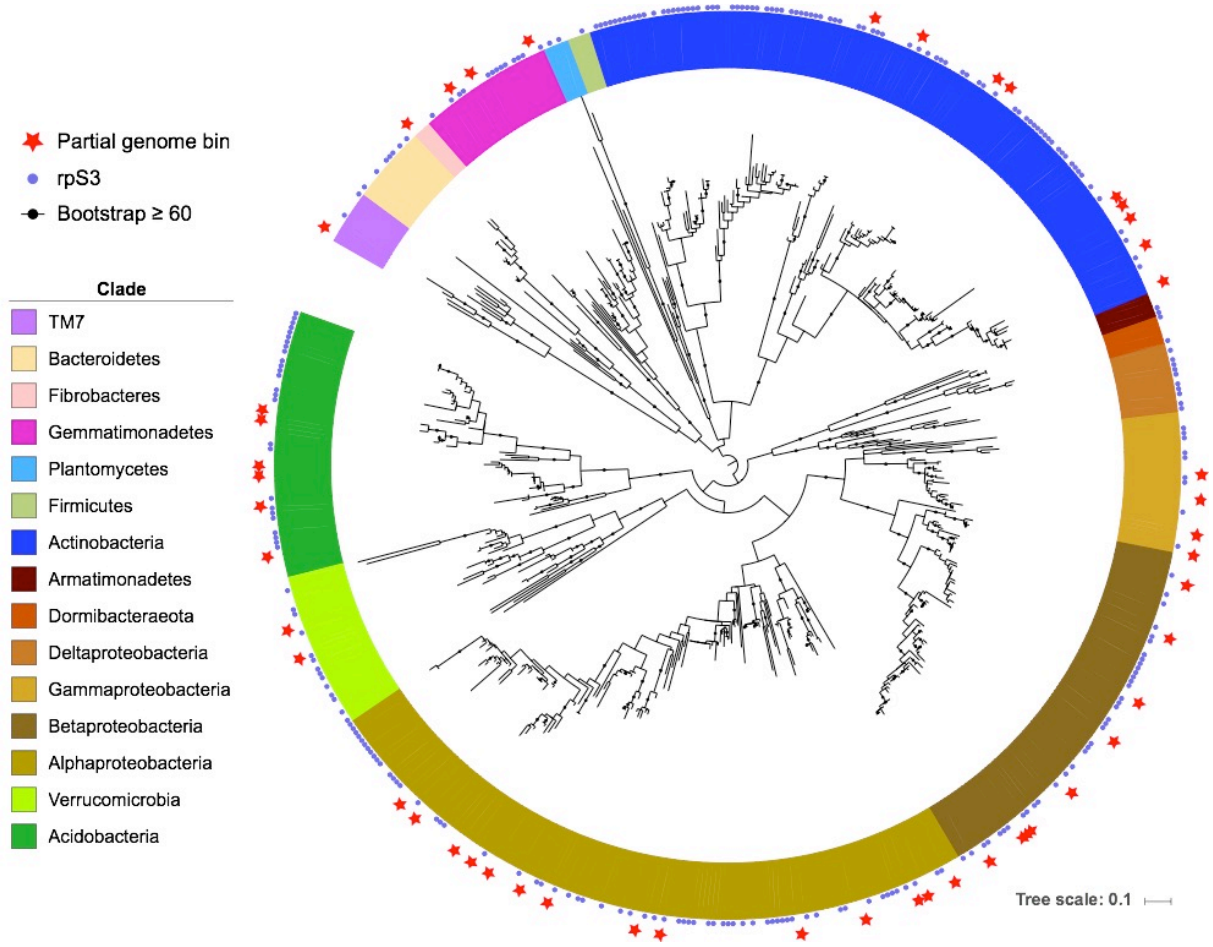


Figure 3.1. Phylogenetic tree of bacterial rpS3. Soil and rhizosphere derived metagenomic bacterial bins with an rpS3 gene and unbinned scaffolds are highlighted. Bacterial clades are highlighted in different colors.

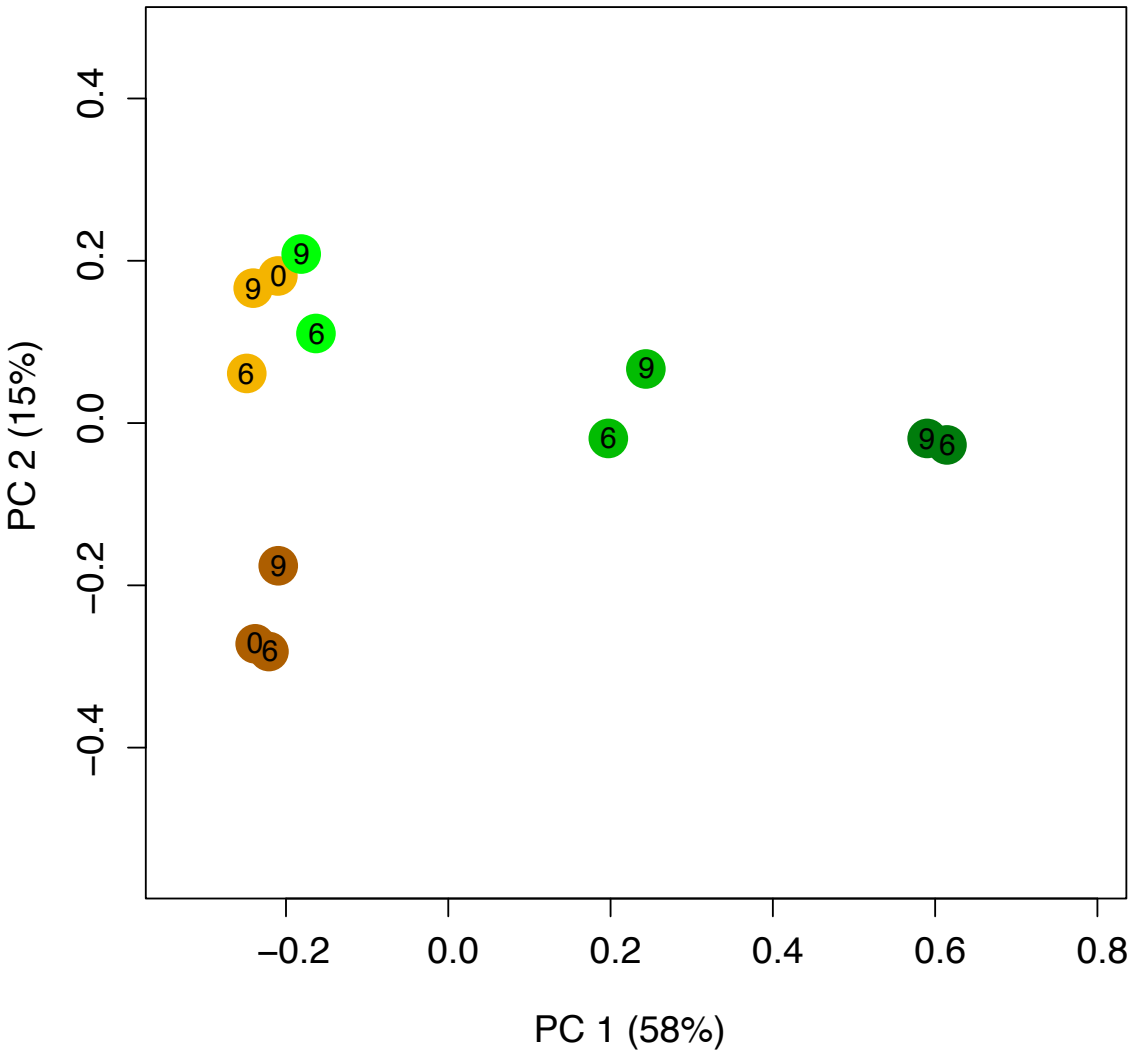
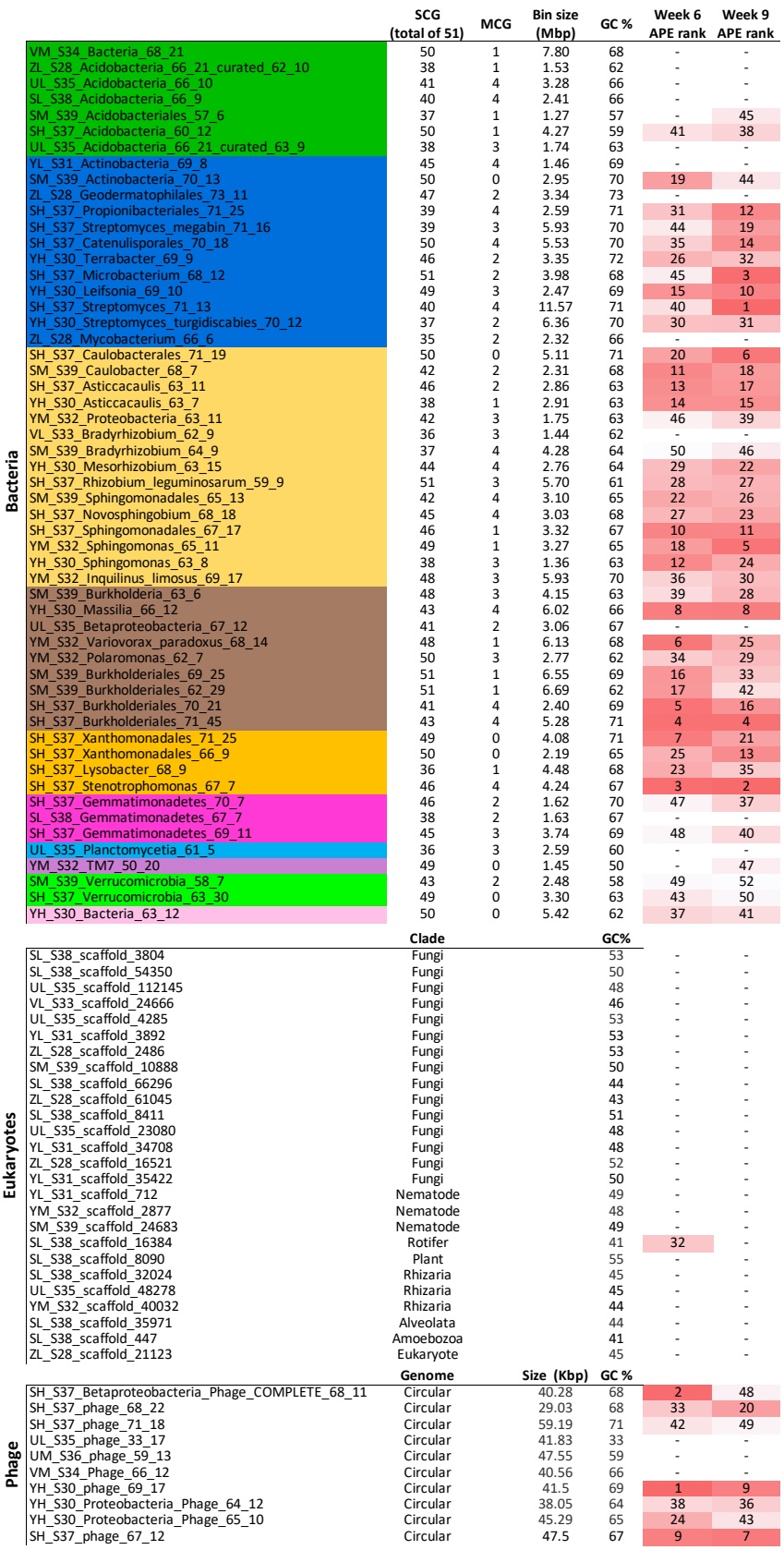


Figure 3.2. PCoA of bacterial rpS3 from different soil samples and fractions. The brown dots represent bulk samples and the greed dots represent rhizosphere samples. The darker the color the more dense the sample. Numbers inside the dots correspond to the week of sampling.



Bacteria

	SCG (total of 51)	MCG	Bin size (Mbp)	GC %	Week 6 APE rank	Week 9 APE rank
VM_S34_Bacteria_68_21	50	1	7.80	68	-	-
ZL_S28_Acidobacteria_66_21_curated_62_10	38	1	1.53	62	-	-
UL_S35_Acidobacteria_66_10	41	4	3.28	66	-	-
SL_S38_Acidobacteria_66_9	40	4	2.41	66	-	-
SM_S39_Acidobacteriales_57_6	37	1	1.27	57	-	45
SH_S37_Acidobacteria_60_12	50	1	4.27	59	41	38
UL_S35_Acidobacteria_66_21_curated_63_9	38	3	1.74	63	-	-
YL_S31_Actinobacteria_69_8	45	4	1.46	69	-	-
SM_S39_Actinobacteria_70_13	50	0	2.95	70	19	44
ZL_S28_Geodermatophilales_73_11	47	2	3.34	73	-	-
SH_S37_Propionibacteriales_71_25	39	4	2.59	71	31	12
SH_S37_Streptomyces_megabin_71_16	39	3	5.93	70	44	19
SH_S37_Catenulisporales_70_18	50	4	5.53	70	35	14
YH_S30_Terrabacter_69_9	46	2	3.35	72	26	32
SH_S37_Microbacterium_68_12	51	2	3.98	68	45	3
YH_S30_Leifsonia_69_10	49	3	2.47	69	15	10
SH_S37_Streptomyces_71_13	40	4	11.57	71	40	1
YH_S30_Streptomyces_turgidiscabies_70_12	37	2	6.36	70	30	31
ZL_S28_Mycobacterium_66_6	35	2	2.32	66	-	-
SH_S37_Caulobacterales_71_19	50	0	5.11	71	20	6
SM_S39_Caulobacter_68_7	42	2	2.31	68	11	18
SH_S37_Asticcacaulis_63_11	46	2	2.86	63	13	17
YH_S30_Asticcacaulis_63_7	38	1	2.91	63	14	15
YM_S32_Proteobacteria_63_11	42	3	1.75	63	46	39
VL_S33_Bradyrhizobium_62_9	36	3	1.44	62	-	-
SM_S39_Bradyrhizobium_64_9	37	4	4.28	64	50	46
YH_S30_Mesorhizobium_63_15	44	4	2.76	64	29	22
SH_S37_Rhizobium_leguminosarum_59_9	51	3	5.70	61	28	27
SM_S39_Sphingomonadales_65_13	42	4	3.10	65	22	26
SH_S37_Novosphingobium_68_18	45	4	3.03	68	27	23
SH_S37_Sphingomonadales_67_17	46	1	3.32	67	10	11
YM_S32_Sphingomonas_65_11	49	1	3.27	65	18	5
YH_S30_Sphingomonas_63_8	38	3	1.36	63	12	24
YM_S32_Inquilinus_limosus_69_17	48	3	5.93	70	36	30
SM_S39_Burkholderia_63_6	48	3	4.15	63	39	28
YH_S30_Massilia_66_12	43	4	6.02	66	8	8
UL_S35_Betaproteobacteria_67_12	41	2	3.06	67	-	-
YM_S32_Variovorax_paradoxus_68_14	48	1	6.13	68	6	25
YM_S32_Polaromonas_62_7	50	3	2.77	62	34	29
SM_S39_Burkholderiales_69_25	51	1	6.55	69	16	33
SM_S39_Burkholderiales_62_29	51	1	6.69	62	17	42
SH_S37_Burkholderiales_70_21	41	4	2.40	69	5	16
SH_S37_Burkholderiales_71_45	43	4	5.28	71	4	4
SH_S37_Xanthomonadales_71_25	49	0	4.08	71	7	21
SH_S37_Xanthomonadales_66_9	50	0	2.19	65	25	13
SH_S37_Lysobacter_68_9	36	1	4.48	68	23	35
SH_S37_Stenotrophomonas_67_7	46	4	4.24	67	3	2
SH_S37_Gemmatimonadetes_70_7	46	2	1.62	70	47	37
SL_S38_Gemmatimonadetes_67_7	38	2	1.63	67	-	-
SH_S37_Gemmatimonadetes_69_11	45	3	3.74	69	48	40
UL_S35_Planctomycetia_61_5	36	3	2.59	60	-	-
YM_S32_TM7_50_20	49	0	1.45	50	-	47
SM_S39_Verrucomicrobia_58_7	43	2	2.48	58	49	52
SH_S37_Verrucomicrobia_63_30	49	0	3.30	63	43	50
YH_S30_Bacteria_63_12	50	0	5.42	62	37	41

Eukaryotes

	Clade	GC%		
SL_S38_scaffold_3804	Fungi	53	-	-
SL_S38_scaffold_54350	Fungi	50	-	-
UL_S35_scaffold_112145	Fungi	48	-	-
VL_S33_scaffold_24666	Fungi	46	-	-
UL_S35_scaffold_4285	Fungi	53	-	-
YL_S31_scaffold_3892	Fungi	53	-	-
ZL_S28_scaffold_2486	Fungi	53	-	-
SM_S39_scaffold_10888	Fungi	50	-	-
SL_S38_scaffold_66296	Fungi	44	-	-
ZL_S28_scaffold_61045	Fungi	43	-	-
SL_S38_scaffold_8411	Fungi	51	-	-
UL_S35_scaffold_23080	Fungi	48	-	-
YL_S31_scaffold_34708	Fungi	48	-	-
ZL_S28_scaffold_16521	Fungi	52	-	-
YL_S31_scaffold_35422	Fungi	50	-	-
YL_S31_scaffold_712	Nematode	49	-	-
YM_S32_scaffold_2877	Nematode	48	-	-
SM_S39_scaffold_24683	Nematode	49	-	-
SL_S38_scaffold_16384	Rotifer	41	32	-
SL_S38_scaffold_8090	Plant	55	-	-
SL_S38_scaffold_32024	Rhizaria	45	-	-
UL_S35_scaffold_48278	Rhizaria	45	-	-
YM_S32_scaffold_40032	Rhizaria	44	-	-
SL_S38_scaffold_35971	Alveolata	44	-	-
SL_S38_scaffold_447	Amoebozoa	41	-	-
ZL_S28_scaffold_21123	Eukaryote	45	-	-

Phage

	Genome	Size (Kbp)	GC %		
SH_S37_Betaproteobacteria_Phage_COMPLETE_68_11	Circular	40.28	68	2	48
SH_S37_phage_68_22	Circular	29.03	68	33	20
SH_S37_phage_71_18	Circular	59.19	71	42	49
UL_S35_phage_33_17	Circular	41.83	33	-	-
UM_S36_phage_59_13	Circular	47.55	59	-	-
VM_S34_Phage_66_12	Circular	40.56	66	-	-
YH_S30_phage_69_17	Circular	41.5	69	1	9
YH_S30_Proteobacteria_Phage_64_12	Circular	38.05	64	38	36
YH_S30_Proteobacteria_Phage_65_10	Circular	45.29	65	24	43
SH_S37_phage_67_12	Circular	47.5	67	9	7

Figure 3.3. Soil derived bacterial bin, eukaryotic scaffold and phage genome statistics. Genome statistics for the bacteria (colored by clade with color scheme from Figure 3.1), eukaryote scaffolds containing 18S rRNA genes and complete, circularized phage genomes. Also noted is the rank APE from the qSIP calculation in ascending order for week 6 and 9.

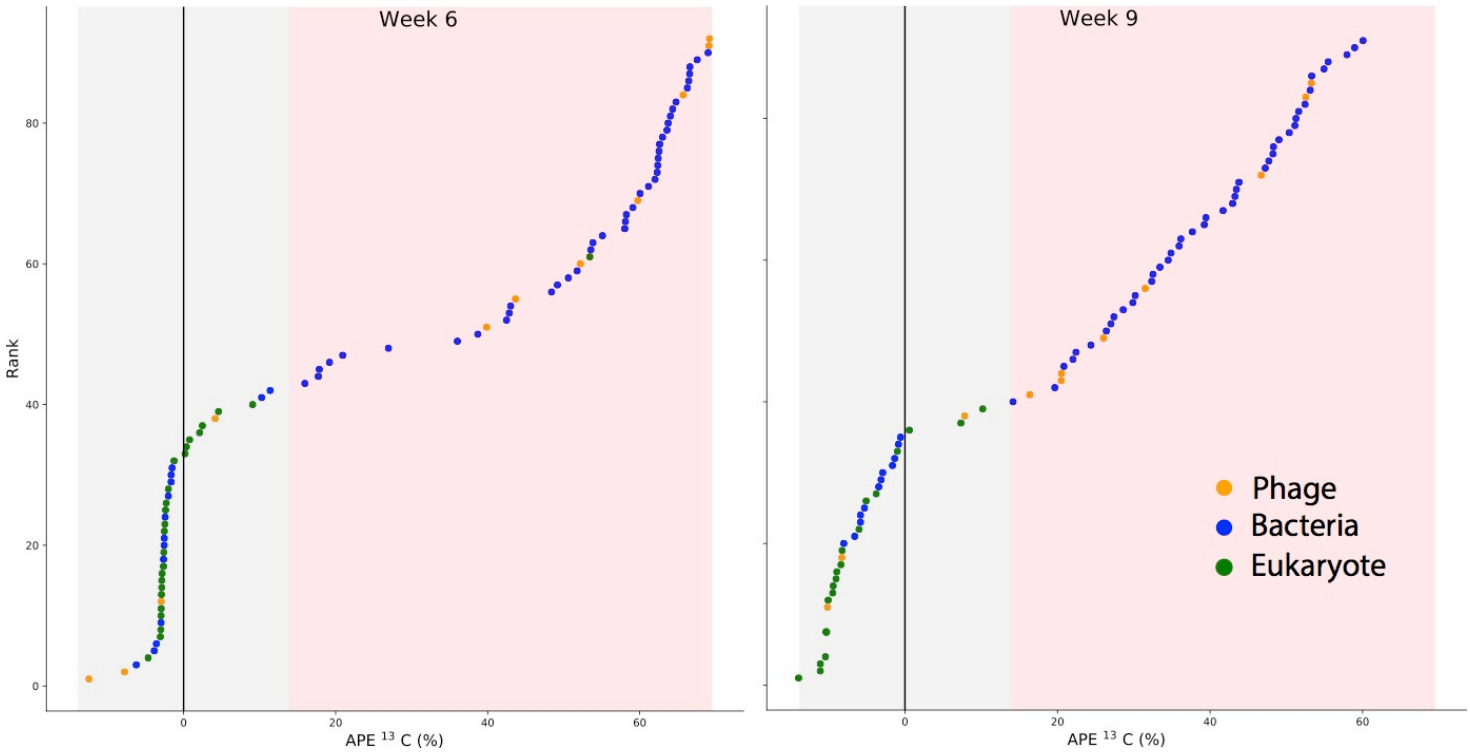


Figure 3.4. Rank of entities in order of their atom percent excess. The rank of soil derived phage genomes, bacterial genome bins and scaffolds encoding eukaryotic 18S rRNA genes in week 6 and 9 in order of their APE from the qSIP calculation. The grey highlight indicates unlabeled entities and the red indicated possibly labelled DNA, the labelling cutoff is explained in the methods section.

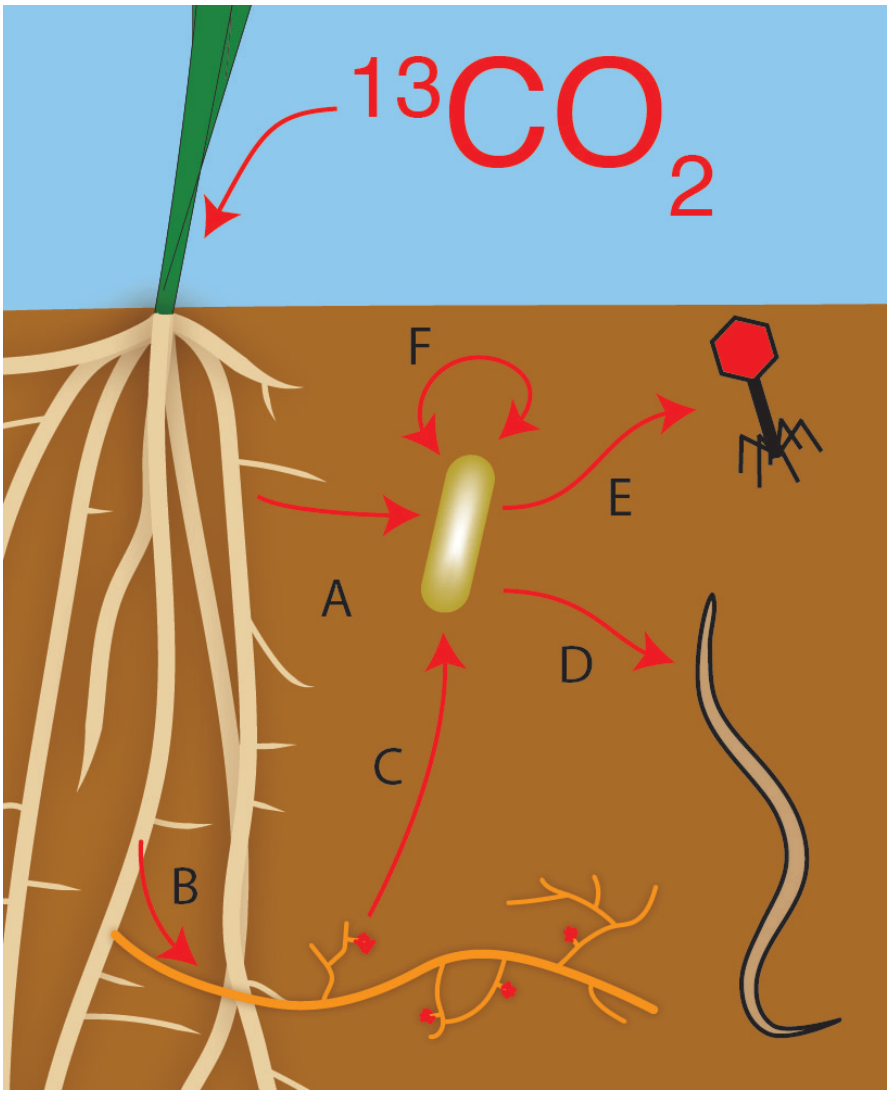


Figure 3.5. Conceptual diagram of carbon flow and community interactions in the rhizosphere. Red lines indicate possible flow of ^{13}C label into and through DNA. Bacteria consume plant exudates or actively infect the plant to receive carbon (A). Fungi may also be infecting and/or decomposing dead plant material (B). Rhizosphere dwelling bacteria may be interacting with live fungi or breaking down dead fungal biomass (C). Bacterivorous micro-eukaryotes may have consumed rhizosphere bacteria. (D). Bacteriophage may have infected rhizosphere dwelling bacteria (E). Bacterial genomes were also rife systems made for communication and competition between bacteria (F).

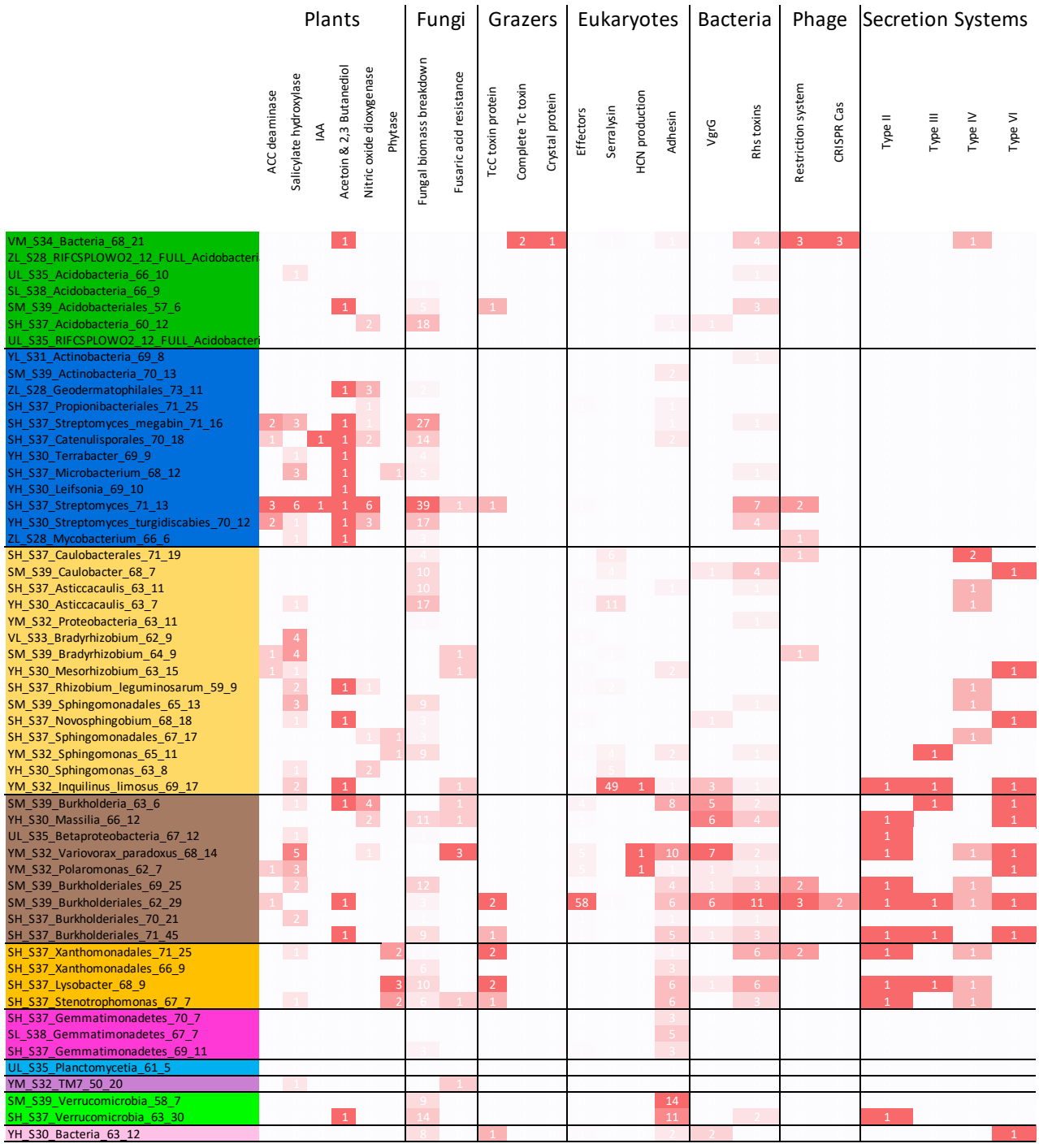
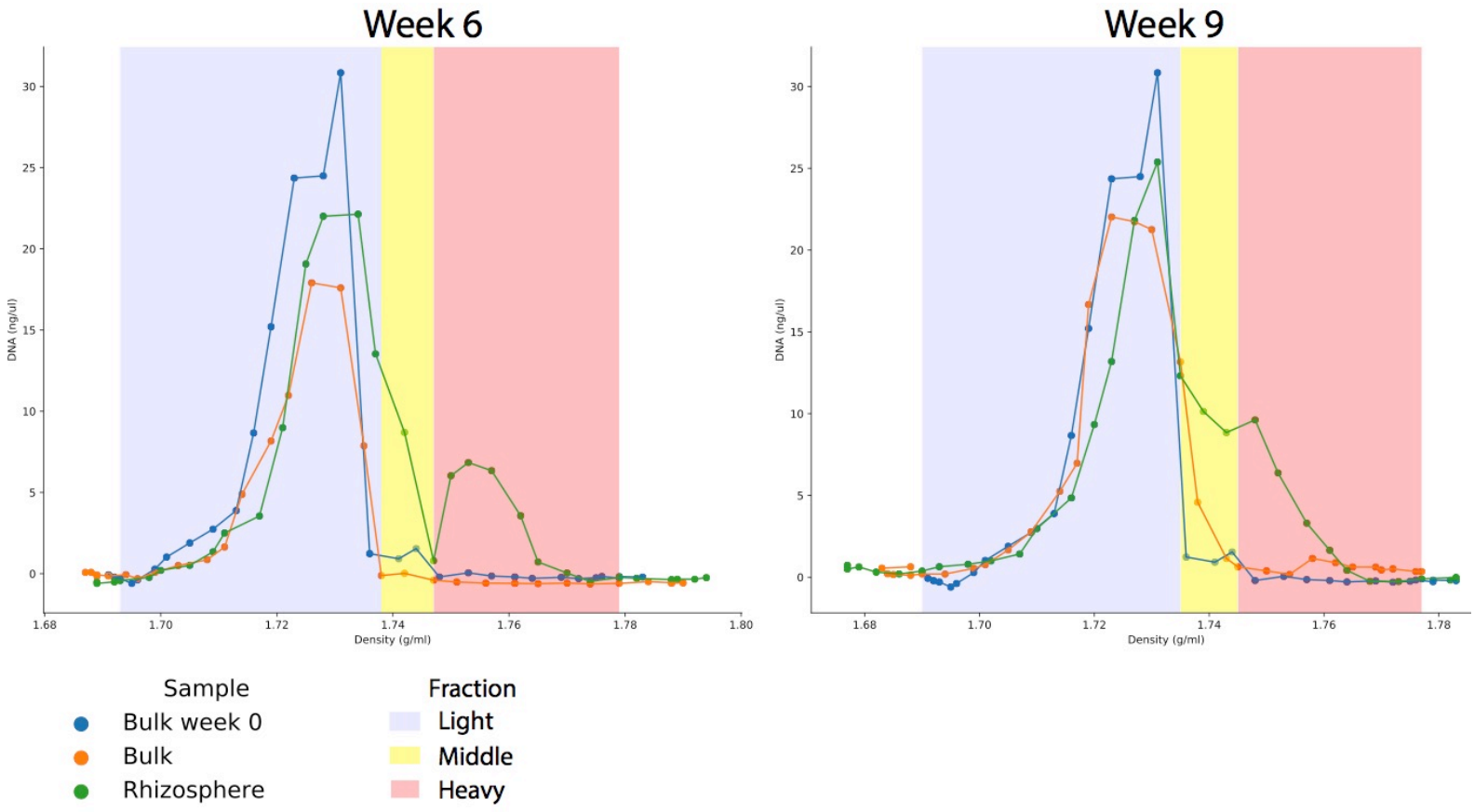
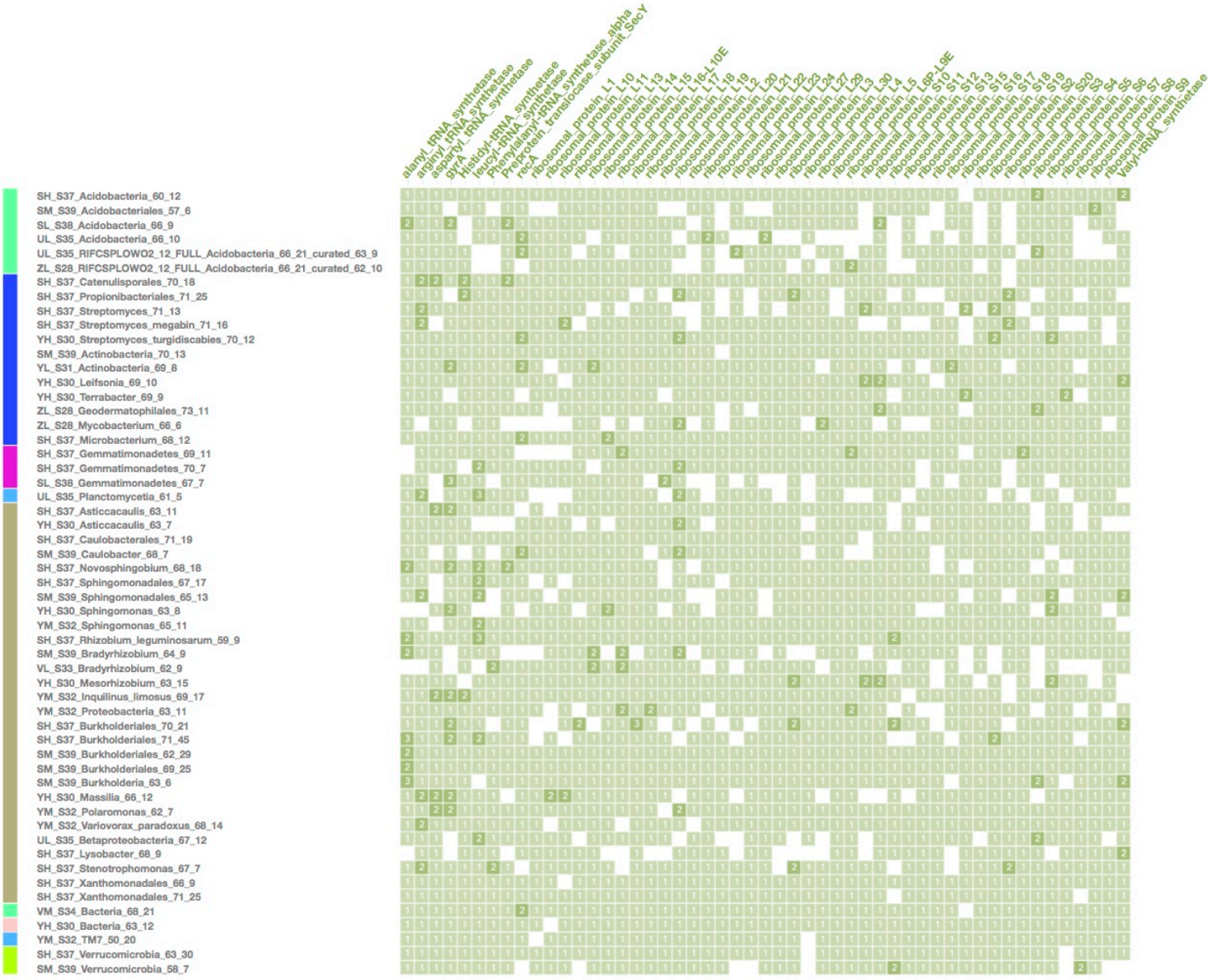


Figure 3.6. Possible inter-organismal interactions encoded on the soil derived bacterial genome bins. Numbers indicate individual genes or nearly complete pathways predicted to be used in inter-organismal interaction.

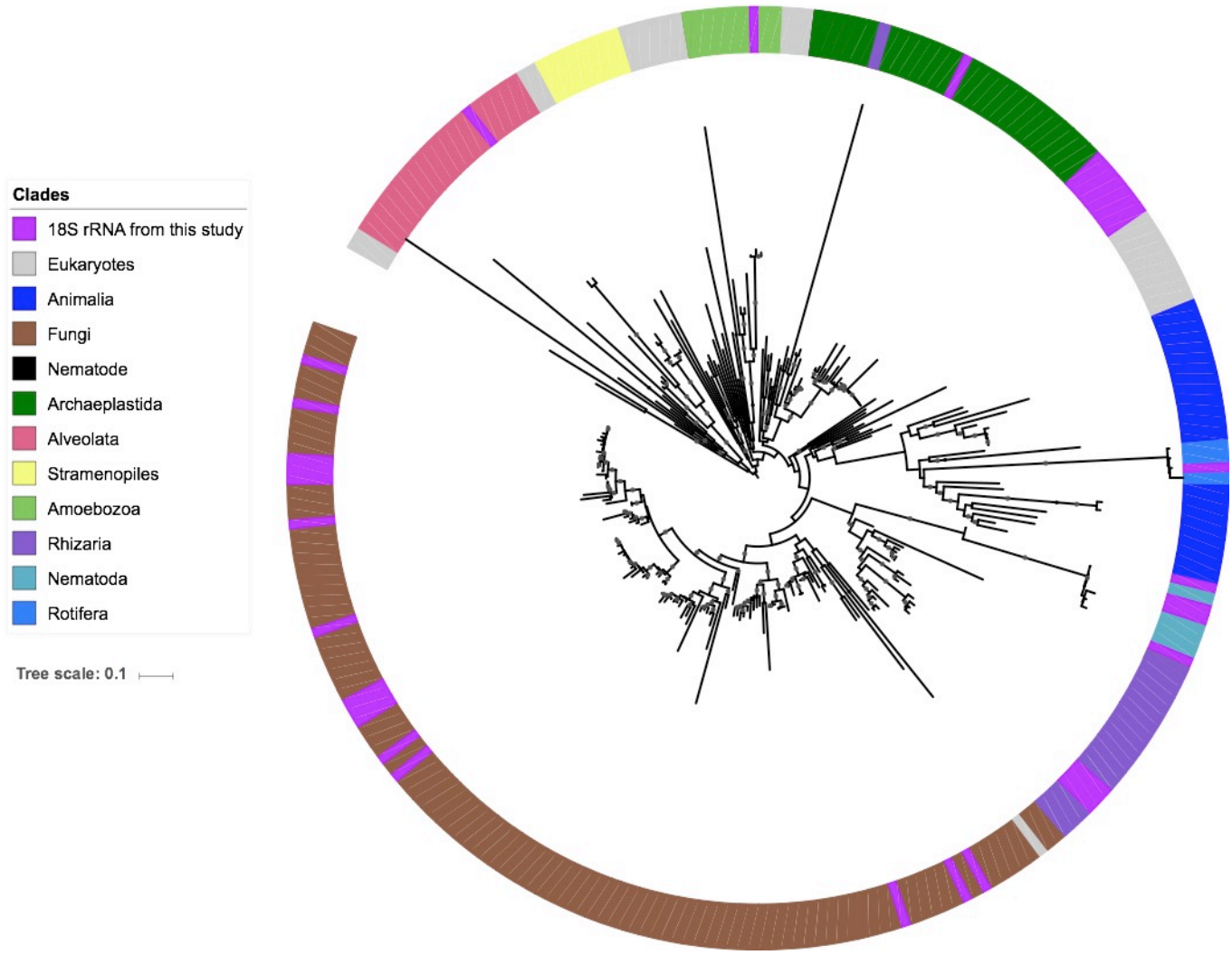
3.8 Supplemental figures



Supplemental figure 3.1. Stable isotope separation in bulk and rhizosphere soil. The density and concentration of DNA for each collection is marked to generate the SIP separation graph.



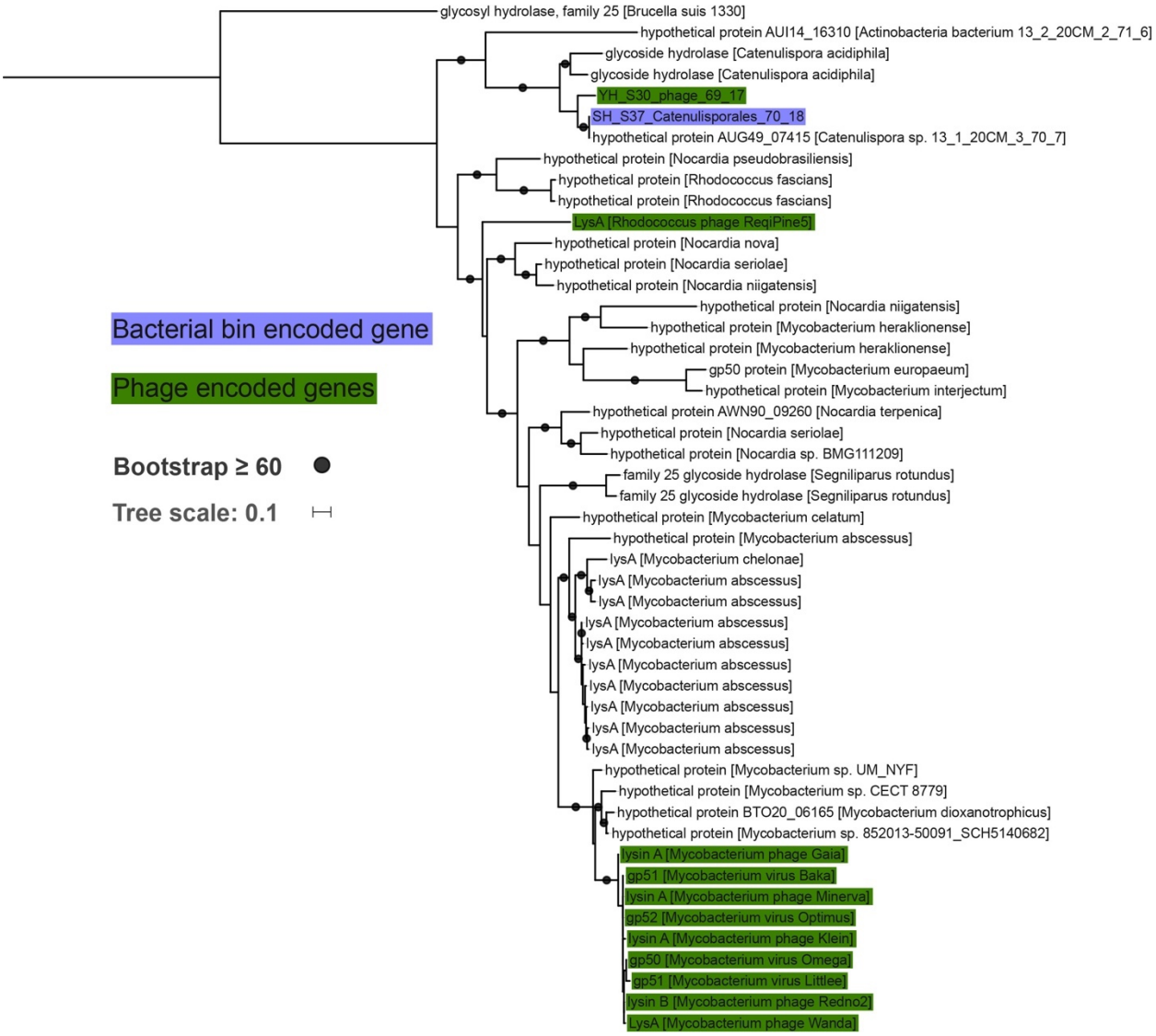
Supplemental figure 3.2. Count of bacterial single copy genes in the partial bins. Color bar represents clade of bin, colored according to Figure 3.1.



Supplemental figure 3.3. Soil metagenome derived 18S rRNA tree colored by clade. Black dots represent bootstrap values ≥ 60. Long leaves were removed for ease of visualization.

	NRPS & NRPS-like	PKS	Bacteriocin	Lanthipeptide	Phosphonate	Indole	Signalling	Terpene	Siderophore	Ectoine	Pigments	Other
VM_S34_Bacteria_68_21	20	3	3	8				1				5
ZL_S28_RIFCSPLOWO2_12_FULL_Acidobacteria_66_21_curated_62_10	1											
UL_S35_Acidobacteria_66_10	1											
SL_S38_Acidobacteria_66_9	1				1							
SM_S39_Acidobacteriales_57_6												1
SH_S37_Acidobacteria_60_12	1	1	2					3				1
UL_S35_RIFCSPLOWO2_12_FULL_Acidobacteria_66_21_curated_63_9								2				
YL_S31_Actinobacteria_69_8												1
SM_S39_Actinobacteria_70_13								1				1
ZL_S28_Geodermatophilales_73_11	5	1						2			1	
SH_S37_Propionibacteriales_71_25	1											
SH_S37_Streptomyces_megabin_71_16	2	9	2					7	2	1		
SH_S37_Catenulisporales_70_18	2	3		2					1			2
YH_S30_Terrabacter_69_9	1	1										
SH_S37_Microbacterium_68_12	1								1			2
YH_S30_Leifsonia_69_10	1	1						1				
SH_S37_Streptomyces_71_13	11	4	2	2			3	6	4	1	1	9
YH_S30_Streptomyces_turgidiscabies_70_12	7	3	2	1				3	2	1	1	2
ZL_S28_Mycobacterium_66_6	3	3	1									
SH_S37_Caulobacterales_71_19		1	1									
SM_S39_Caulobacter_68_7												
SH_S37_Asticcacaulis_63_11			2			1						
YH_S30_Asticcacaulis_63_7			1			1						2
YM_S32_Proteobacteria_63_11	1							1				
VL_S33_Bradyrhizobium_62_9								1				
SM_S39_Bradyrhizobium_64_9	5		1					3				
YH_S30_Mesorhizobium_63_15		1					1					
SH_S37_Rhizobium_leguminosarum_59_9		1	2				1	1		1		
SM_S39_Sphingomonadales_65_13	1							2				2
SH_S37_Novosphingobium_68_18	1	1						1				1
SH_S37_Sphingomonadales_67_17	2						2	1				
YM_S32_Sphingomonas_65_11	1	1					1	1				
YH_S30_Sphingomonas_63_8							1					
YM_S32_Inquilinus_limosus_69_17	3	2			1			3				
SM_S39_Burkholderia_63_6		1	1		1			3				
YH_S30_Massilia_66_12			2					1	1			
UL_S35_Betaproteobacteria_67_12	1							1				1
YM_S32_Variovorax_paradoxus_68_14	9	1	1				1	1			2	2
YM_S32_Polaromonas_62_7			1					1				
SM_S39_Burkholderiales_69_25	30		2			1	2	1			1	3
SM_S39_Burkholderiales_62_29	15	1	2				1	1				2
SH_S37_Burkholderiales_70_21	1						1				1	
SH_S37_Burkholderiales_71_45								1			2	1
SH_S37_Xanthomonadales_71_25	6		2				1				1	4
SH_S37_Xanthomonadales_66_9	1										1	1
SH_S37_Lysobacter_68_9	9		1	3								
SH_S37_Stenotrophomonas_67_7	2		2								1	
SH_S37_Gemmatimonadetes_70_7								1				
SL_S38_Gemmatimonadetes_67_7												
SH_S37_Gemmatimonadetes_69_11	2							1				
UL_S35_Planctomycetia_61_5		1	1					3				
YM_S32_TM7_50_20												
SM_S39_Verrucomicrobia_58_7	1							3				
SH_S37_Verrucomicrobia_63_30		1										
YH_S30_Bacteria_63_12		1	2					1				

Supplemental figure 3.4. Biosynthetic gene clusters encoded in soil derived bacterial bins. Signaling compounds includes homoserine lactone clusters, N-acyl amino acid clusters and butyrolactone cluster.



Supplemental figure 3.5. Phylogenetic tree of cazy GH25 gene. The cazy GH25 genes from phages are highlighted in green and blue highlights the bacterial genome bin from this study, the remaining cazy GH25 genes come from publicly available genomes.

Supplemental table 3.1. Stable isotope probing and sequencing statistics.

Sample	Time point (weeks)	Fraction	Total sequenced (Gbp)	Density (g/ml)	DNA concentration (ng/ul)	Total DNA (ng)
T0 bulk soil	0	Light	18.4	1.695-1.731	271	2519
T0 bulk soil	0	Middle	16.8	1.732-1.744	22	176
Bulk soil	6	Light	17.3	1.694-1.735	291	2039
Bulk soil	6	Middle	17.4	1.736-1.742	11	109
Rhizosphere soil	6	Light	16.8	1.692-1.737	267	2421
Rhizosphere soil	6	Middle	15.2	1.738-1.746	17	159
Rhizosphere soil	6	Heavy	17.3	1.747-1.765	137	1200
Bulk soil	9	Light	18.3	1.694-1.735	113	2260
Bulk soil	9	Middle	16.4	1.736-1.745	6	127
Rhizosphere soil	9	Light	18.9	1.69-1.731	82	1635
Rhizosphere soil	9	Middle	16.6	1.732-1.743	31	626
Rhizosphere soil	9	Heavy	19.5	1.744-1.768	21	427

Conclusions

The study of soils is fundamentally important to our understanding of global biogeochemical cycles, terrestrial life and food production. Plant roots represent one of the main sources for carbon coming into soil. The carbon fixed by plants and exuded into the rhizosphere supports a complex web of life and interactions. These interactions have direct consequences for plant health, soil fertility, soil carbon stabilization and ecosystem functioning. Through a better understanding of these processes we increase our predictive capabilities and may be able influence the fate of these ecosystem outputs. In my work I have focused on understanding the less well studied organisms in soil and hope to show that they are critical members of the soil community and could have drastic effects on soil functioning.

This work represents small, but real, steps forward in our understanding of soil microbiology. In Chapter One I present one of the first circularized, complete bacterial genome derived from a metagenome. This is a considerable step forward in our metagenomic capabilities in one of the world's most complex microbial habitats. In metagenomics we should strive for completing genomes. Complete genomes are the only way to responsibly say that an organism lacks a gene or pathway (if the gene or pathway is identifiable). In our case we could say with certainty that the Saccharibacteria genome that we reconstructed did not encode canonical nucleotide biosynthesis pathways and must have incorporated the labelled nucleotides from elsewhere, likely from one or more labelled rhizosphere bacteria. Circularized genomes should become the gold standard for metagenomics, however further development in sequencing technology and assembly algorithms is needed.

I believe our ability to complete a genome bin from soil was aided by the SIP method. By stable isotope probing the rhizosphere community we drastically improved the assemblies from the middle and heavy fractions. The SIP process reduced the organismal and thus sequence diversity and this may have generated much higher quality assemblies and better bins. Most of the high quality genome bins analyzed in Chapter Three came from the rhizosphere middle and heavy fractions despite three times more sequencing depth in the other samples. This simplification of the rhizosphere community was likely only possible because of the long labelling process. In the rhizosphere heavy samples there was a clear new peak of labelled DNA. With a shorter or less efficient labeling process the shift would be less dramatic and simplification of the community would be less apparent. It appears that by combining SIP with metagenomics we can improve the quality of the recoverable genome bins. In addition to being higher quality, the SIP genome bins may also represent the more relevant rhizosphere organisms. There has been much talk about DNA in soil coming from dead organisms, called relic DNA. In Chapter Three we show that only a fraction of the microbial DNA is labelled even after living near a root which is exuding labelled carbon for nine weeks. The unlabeled DNA could represent those organisms that grow too slowly to be detected using SIP, organisms which did not consume plant derived carbon, whether due to metabolic constraints or microhabitats isolated from the root's influence, or relic DNA. The genomes we reconstructed in Chapter Three were also unexpectedly more relevant in a tandem metatranscriptome study. It appeared that transcripts from the same soil but from a different experiment mapped best to the bins from Chapter Three when compared to isolates,

single cell genomes, and non-SIP metagenome bins all from the same study soil. This may reflect the higher quality bins obtained from the SIP process, but it could also mean that there was a higher likelihood that these were the metabolically active microbes in the soil.

The SIP process supported our hypothesis that phage are actively infecting bacteria in the rhizosphere. We traced carbon fixed by the plant into a possible plant pathogen and then into an infecting phage. It is a true accomplishment that we were able to identify the movement of carbon through two trophic levels in a complex community. This result was aided by our analysis of only complete phage genomes. Much like bacterial genomes, when uncomplete genomes are analyzed we do not know what genes are missing. But unlike bacteria the phage can be integrated into the bacterial genome. This means that in our qSIP analysis we could mistake a prophage integrated into a labelled bacterial genome for a phage that lysed its host and became labelled. In addition to this danger, many of the genomes we analyzed contained prophage-like regions which could represent defunct or domesticated prophage that are unlikely to kill their host in the future. One downside to focusing on circularized phage genomes is the lack of complete linear phage genomes. In working on Chapter Three I encountered several phage contigs which displayed an unusual phenomenon. The reads did not have the usual logarithmic curve of increasing coverage from the end of the contig toward the center, but instead all the reads stacked at the ends of the contig. This phenomenon could represent a way to identify complete linear phage genomes.

There are some key differences in the viral information obtained from Chapter Two and Chapter Three. The most striking finding is the high diversity and abundance of eukaryotic and bacterial RNA viruses. We identified many more eukaryotic RNA viruses (3,884 from 48 samples) compared to eukaryotic DNA viruses (1 from five biological samples) in the soil. This is inherently biased since there are many more types of eukaryotic RNA viruses, however it is an important finding regarding eukaryotic virology in soil. The more surprising result was the large number of RNA phage (1,350 contigs from 48 samples) compared to DNA phage (10 complete genomes and hundreds of predicted contigs from five biological samples). This means we should devote more time and energy to studying RNA viruses and phage. The DNA phage have much larger genomes and can have diverse metabolic and ecological influences on their hosts; however the sheer number of RNA phage may speak to their importance in our soil and possibly other soils. The amount of identifiable eukaryotes and RNA viruses in each sample speaks to their importance in the ecology of the rhizosphere. Eukaryotic RNA viruses may play a bigger role in soil communities than previously thought both in terms of ecology, but also carbon cycling through the lysis of eukaryotic cells and fitness drain caused by chronic infections. This work represents an important first step: the first analysis of RNA viruses from an assembled soil metatranscriptome and I feel the field of RNA virology in complex systems will only grow.

The difference in information obtained from the RNA assemblies (Chapter Two) compared to the DNA assemblies (Chapter Three) was not limited to viruses. The amount of eukaryotes we identified using the two techniques differed greatly as well. We identified 27 eukaryotic species based on assembled 18S rRNA in the DNA samples and we reconstructed 521 18S rRNA species level sequences from the assembled metatranscriptome, even after bacterial and plant rRNA depletion. This means that we may be able to uncover a greater level of eukaryotic diversity using assembled metatranscriptomes compared to metagenomes. From the metagenome we were

unable to recover many eukaryotic scaffolds or any genome bins, while we were able to identify many eukaryotic scaffolds from the assembled metatranscriptome. This may speak to the inherent difference in bacterial and eukaryotic DNA to RNA ratios within a cell. In addition, we noted some differences in the bacteria and archaea detected in the metatranscriptome compared to the metagenome. There was high transcriptional activity of Chlamydiae and Thaumarchaeota which, when combined, had the highest level of rpS3 transcription in 11 of 48 samples. However, these organisms were undetectable in the metagenome samples. This could partially be explained by the difference in the study, but some organisms may have a low cellular abundance (thus low DNA abundance) yet maintain a high transcriptional activity. It appears that assembled metatranscriptomes could provide an avenue for studying organisms with a low DNA:RNA ratios in complex communities.

By considering nearly all the members of the rhizosphere community, I hope that we can better understand the ecology of soil and the importance of previously overlooked members and functions in soil. The soil community is made up of bacteria, archaea, eukaryotes, viruses and phages all interacting with one another, and by studying all members of the soil community, we may come to know our soils more deeply. There is much to be learned from studying how organisms interact with one another, not only for the advancement of ecology but also for the improvement of agricultural and biotechnological practices. By incorporating our knowledge about interacting members of the rhizosphere community we may be able to develop better and more sustainable biocontrol or plant growth promoting consortium made up of bacteria, eukaryotes, phage and RNA viruses. The analysis of inter-microbial interaction may provide clues to the function of biosynthetic gene clusters or toxin modules. Yet to be cultured entities may provide new biotechnological tools including novel insecticidal proteins and phage derived simple antibacterials.

The combination of genome resolved metagenomics, assembled metatranscriptomes and stable isotope probing provides powerful tools to investigate the holistic identity and functioning of organisms in complex communities. By following stable isotopes through the community we can begin to understand the connectedness and flow of nutrients in soil, and genome analysis provides the context and clues to the ways in which the organisms were connected.

References

Adams, M. J., E. J. Lefkowitz, A. M. Q. King, B. Harrach, R. L. Harrison, N. J. Knowles, A. M. Kropinski, M. Krupovic, J. H. Kuhn, A. R. Mushegian, M. Nibert, S. Sabanadzovic, H. Sanfaçon, S. G. Siddell, P. Simmonds, A. Varsani, F. M. Zerbini, A. E. Gorbalenya, and A. J. Davison. 2017. Changes to taxonomy and the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses (2017). Archives of Virology 162:2505–2538.

Afonso, C. L., G. K. Amarasinghe, K. Bányai, Y. Bào, C. F. Basler, S. Bavari, N. Bejerman, K. R. Blasdell, F. X. Briand, T. Briese, A. Bukreyev, C. H. Calisher, K. Chandran, J. Chéng, A. N. Clawson, P. L. Collins, R. G. Dietzgen, O. Dolnik, L. L. Domier, R. Dürrwald, J. M. Dye, A. J. Easton, H. Ebihara, S. L. Farkas, J. Freitas-Astúa, P. Formenty, R. A. M. Fouchier, Y. Fù, E. Ghedin, M. M. Goodin, R. Hewson, M. Horie, T. H. Hyndman, D. Jiāng, E. W. Kitajima, G. P. Kobinger, H. Kondo, G. Kurath, R. A. Lamb, S. Lenardon, E. M. Leroy, C. X. Li, X. D. Lin, L. Liú, B. Longdon, S. Marton, A. Maisner, E. Mühlberger, S. V. Netesov, N. Nowotny, J. L. Patterson, S. L. Payne, J. T. Paweska, R. E. Randall, B. K. Rima, P. Rota, D. Rubbenstroth, M. Schwemmle, M. Shi, S. J. Smither, M. D. Stenglein, D. M. Stone, A. Takada, C. Terregino, R. B. Tesh, J. H. Tian, K. Tomonaga, N. Tordo, J. S. Towner, N. Vasilakis, M. Verbeek, V. E. Volchkov, V. Wahl-Jensen, J. A. Walsh, P. J. Walker, D. Wang, L. F. Wang, T. Wetzel, A. E. Whitfield, J. Xiè, K. Y. Yuen, Y. Z. Zhang, and J. H. Kuhn. 2016. Taxonomy of the order Mononegavirales: update 2016. Archives of Virology 161:2351–2360.

Albertsen, M., P. Hugenholtz, A. Skarshewski, K. L. Nielsen, G. W. Tyson, and P. H. Nielsen. 2013. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. Nature Biotechnology 31:533–538.

Alori, E. T., B. R. Glick, and O. O. Babalola. 2017. Microbial Phosphorus Solubilization and Its Potential for Use in Sustainable Agriculture. Frontiers in Microbiology 8:971.

Amiott, E. A., and J. A. Jaehning. 2006. Mitochondrial Transcription Is Regulated via an ATP "Sensing" Mechanism that Couples RNA Abundance to Respiration. Molecular Cell 22:329–338.

Anantharaman, K., C. T. Brown, D. Burstein, C. J. Castelle, A. J. Probst, B. C. Thomas, K. H. Williams, and J. F. Banfield. 2016. Analysis of five complete genome sequences for members of the class Peribacteria in the recently recognized Peregrinibacteria bacterial phylum. PeerJ 4:e1607.

Andika, I. B., H. Kondo, and L. Sun. 2016. Interplays between soil-borne plant viruses and RNA silencing-mediated antiviral defense in roots. Frontiers in Microbiology 7:1458.

Aravind, L., and V. Anantharaman. 2003. HutC/FarR-like bacterial transcription factors of the GntR family contain a small molecule-binding domain of the chorismate lyase fold. FEMS Microbiology Letters 222:17–23.

Ashelford, K. E., M. J. Day, and J. C. Fry. 2003. Elevated abundance of bacteriophage infecting

bacteria in soil. Applied and Environmental Microbiology 69:285–289.
Bahram, M., F. Hildebrand, S. K. Forslund, J. L. Anderson, N. A. Soudzilovskaia, P. M. Bodegom, J. Bengtsson-Palme, S. Anslan, L. P. Coelho, H. Harend, J. Huerta-Cepas, M. H. Medema, M. R. Maltz, S. Mundra, P. A. Olsson, M. Pent, S. Põlme, S. Sunagawa, M. Ryberg, L. Tedersoo, and P. Bork. 2018. Structure and function of the global topsoil microbiome. Nature 560:233–237.
Balashova, N. V, A. Stolz, H.-J. Knackmuss, I. A. Kosheleva, A. V Naumov, and & A. M. Boronin. 2001. Purification and characterization of a salicylate hydroxylase involved in 1-hydroxy-2-naphthoic acid hydroxylation from the naphthalene and phenanthrene-degrading bacterial strain Pseudomonas putida BS202-P1. Page Biodegradation.
Bar-On, Y. M., R. Phillips, and R. Milo. 2018. The biomass distribution on Earth. Proceedings of the National Academy of Sciences 115:6506–6511.
Barnard, R. L., C. A. Osborne, and M. K. Firestone. 2013. Responses of soil bacterial and fungal communities to extreme desiccation and rewetting. ISME Journal 7:2229–2241.
Baulard, A. R., S. S. Gurcha, J. Engohang-Ndong, K. Gouffi, C. Locht, and G. S. Besra. 2003. In vivo interaction between the Polyprenol phosphate mannose synthase Ppm1 and the integral membrane protein Ppm2 from Mycobacterium smegmatis revealed by a bacterial two-hybrid system. Journal of Biological Chemistry 278:2242–2248.
Beckers, B., M. O. De Beeck, N. Weyens, W. Boerjan, and J. Vangronsveld. 2017. Structural variability and niche differentiation in the rhizosphere and endosphere bacterial microbiome of field-grown poplar trees. Microbiome 5:25.
Bekker, M., S. De Vries, A. Ter Beek, K. J. Hellingwerf, and M. J. Teixeira De Mattos. 2009. Respiration of Escherichia coli can be fully uncoupled via the nonelectrogenic terminal cytochrome bd-II oxidase. Journal of Bacteriology 191:5510–5517.
Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: A Practical and powerful approach to multiple testing. J.Roy.Statist.Soc. 57:289–300.
Berne, C., A. Ducret, G. G. Hardy, and Y. V Brun. 2015. Adhesins Involved in Attachment to Abiotic Surfaces by Gram-Negative Bacteria. Microbiology spectrum 3.
Biasini, M., S. Bienert, A. Waterhouse, K. Arnold, G. Studer, T. Schmidt, F. Kiefer, T. G. Cassarino, M. Bertoni, L. Bordoli, and T. Schwede. 2014. SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Research 42:W252–W258.
Billings, A. F., J. L. Fortney, T. C. Hazen, B. Simmons, K. W. Davenport, L. Goodwin, N. Ivanova, N. C. Kyrpides, K. Mavromatis, T. Woyke, and K. M. DeAngelis. 2015. Genome sequence and description of the anaerobic lignin-degrading bacterium Tolumonas lignolytica sp. nov. Standards in Genomic Sciences 10:106.
Bird, R. G., and T. F. Mc Caul. 1976. The rhabdoviruses of Entamoeba histolytica and Entamoeba invadens. Annals of Tropical Medicine and Parasitology 70:81–93.
Biswas, A., R. H. J. Staals, S. E. Morales, P. C. Fineran, and C. M. Brown. 2016. CRISPRDetect: A flexible algorithm to define CRISPR arrays. BMC Genomics 17:356.
Blazewicz, S. J., E. Schwartz, and M. K. Firestone. 2014. Growth and death of bacteria and fungi underlie rainfall-induced carbon dioxide pulses from seasonally dried soil. Ecology 95:1162–1172.
Blomain, E. S., and S. B. McMahon. 2012. Dynamic regulation of mitochondrial transcription as a mechanism of cellular adaptation. Biochimica et Biophysica Acta - Gene Regulatory Mechanisms 1819:1075–1079.

Bonfante, P., and A. Genre. 2010. Mechanisms underlying beneficial plant–fungus interactions in mycorrhizal symbiosis. Nature Communications 1:48.
Bor, B., N. Poweleit, J. S. Bois, L. Cen, J. K. Bedree, Z. H. Zhou, R. P. Gunsalus, R. Lux, J. S. McLean, X. He, and W. Shi. 2016. Phenotypic and Physiological Characterization of the Epibiotic Interaction Between TM7x and Its Basibiont Actinomyces. Microbial Ecology 71:243–255.
Brown, C. T., L. A. Hug, B. C. Thomas, I. Sharon, C. J. Castelle, A. Singh, M. J. Wilkins, K. C. Wrighton, K. H. Williams, and J. F. Banfield. 2015. Unusual biology across a group comprising more than 15% of domain Bacteria. Nature 523:208–211.
Brown, C. T., M. R. Olm, B. C. Thomas, and J. F. Banfield. 2016. In situ replication rates for uncultivated bacteria in microbial communities. Nat Biotechnol 34:1256–1263.
Brum, J. R., and M. B. Sullivan. 2015. Rising to the challenge: Accelerated pace of discovery transforms marine virology. Nature Reviews Microbiology 13:147–159.
Butterfield, C. N., Z. Li, P. F. Andeer, S. Spaulding, B. C. Thomas, A. Singh, R. L. Hettich, K. B. Suttle, A. J. Probst, S. G. Tringe, T. Northen, C. Pan, and J. F. Banfield. 2016. Proteogenomic analyses indicate bacterial methylotrophy and archaeal heterotrophy are prevalent below the grass root zone. PeerJ 4:e2687.
Canchaya, C., C. Proux, G. Fournous, A. Bruttin, and H. Brüssow. 2003. Prophage genomics. Microbiology and molecular biology reviews : MMBR 67:238–76, table of contents.
Cañizares, M. C., F. J. López-Escudero, E. Pérez-Artés, and M. D. García-Pedrajas. 2018. Characterization of a novel single-stranded RNA mycovirus related to invertebrate viruses from the plant pathogen Verticillium dahliae. Archives of Virology 163:771–776.
Capella-Gutiérrez, S., J. M. Silla-Martínez, and T. Gabaldón. 2009. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972–1973.
Chen, W.-J., F.-C. Hsieh, F.-C. Hsu, Y.-F. Tasy, J.-R. Liu, and M.-C. Shih. 2014. Characterization of an insecticidal toxin and pathogenicity of Pseudomonas taiwanensis against insects. PLoS pathogens 10:e1004288.
Chen, Y., B. Li, K. Cen, Y. Lu, S. Zhang, and C. Wang. 2018. Diverse effect of phosphatidylcholine biosynthetic genes on phospholipid homeostasis, cell autophagy and fungal developments in *Metarhizium robertsii*. Environmental Microbiology 20:293–304.
Correa-Galeote, D., E. J. Bedmar, A. J. Fernández-González, M. Fernández-López, and G. J. Arone. 2016. Bacterial Communities in the Rhizosphere of Amilaceous Maize (Zea mays L.) as Assessed by Pyrosequencing. Frontiers in Plant Science 7:1016.
Courty, P. E., A. Franc, J. C. Pierrat, and J. Garbaye. 2008. Temporal changes in the ectomycorrhizal community in two soil horizons of a temperate oak forest. Applied and Environmental Microbiology 74:5792–5801.
Creeth, J. M. 2002. Centrifugal Separations in Molecular and Cell Biology. Page FEBS Letters. Butterworths.
Crits-Christoph, A., S. Diamond, C. N. Butterfield, B. C. Thomas, and J. F. Banfield. 2018. Novel soil bacteria possess diverse genes for secondary metabolite biosynthesis. Nature 558:440–444.
Culley, A. 2018. New insight into the RNA aquatic virosphere via viromics. Virus Research 244:84–89.
Culley, A. I., J. A. Mueller, M. Belcaid, E. M. Wood-Charlson, G. Poisson, and G. F. Steward. 2014. The Characterization of RNA Viruses in Tropical Seawater Using Targeted PCR and Metagenomics. mBio 5.

Czech, L., L. Hermann, N. Stöveken, A. Richter, A. Höppner, S. Smits, J. Heider, and E. Bremer. 2018. Role of the Extremolytes Ectoine and Hydroxyectoine as Stress Protectants and Nutrients: Genetics, Phylogenomics, Biochemistry, and Structural Analysis. Genes 9:177.

Damon, C., G. Barroso, C. Férandon, J. Ranger, L. Fraissinet-Tachet, and R. Marmeisse. 2010. Performance of the COX1 gene as a marker for the study of metabolically active Pezizomycotina and Agaricomycetes fungal communities from the analysis of soil RNA. FEMS Microbiology Ecology 74:693–705.

Day, J. M., L. L. Ballard, M. V. Duke, B. E. Scheffler, and L. Zsak. 2010. Metagenomic analysis of the turkey gut RNA virus community. Virology Journal 7:313.

Deakin, G., E. Dobbs, J. M. Bennett, I. M. Jones, H. M. Grogan, and K. S. Burton. 2017. Multiple viral infections in Agaricus bisporus - Characterisation of 18 unique RNA viruses and 8 ORFans identified by deep sequencing. Scientific Reports 7:2469.

DeAngelis, K. M., E. L. Brodie, T. Z. DeSantis, G. L. Andersen, S. E. Lindow, and M. K. Firestone. 2009. Selective progressive response of soil microbial community to wild oat roots. ISME Journal 3:168–178.

Delgado-Baquerizo, M., A. M. Oliverio, T. E. Brewer, A. Benavent-González, D. J. Eldridge, R. D. Bardgett, F. T. Maestre, B. K. Singh, and N. Fierer. 2018. A global atlas of the dominant bacteria found in soil. Science (New York, N.Y.) 359:320–325.

Deveau, A., S. Antony-Babu, F. Le Tacon, C. Robin, P. Frey-Klett, and S. Uroz. 2016. Temporal changes of bacterial communities in the Tuber melanosporum ectomycorrhizosphere during ascocarp development. Mycorrhiza 26:389–399.

Dewhirst, F. E., T. Chen, J. Izard, B. J. Paster, A. C. R. Tanner, W. H. Yu, A. Lakshmanan, and W. G. Wade. 2010. The human oral microbiome. Journal of Bacteriology 192:5002–5017.

Diamond, S., P. F. Andeer, Z. Li, A. Crits-Christoph, D. Burstein, K. Anantharaman, K. R. Lane, B. C. Thomas, C. Pan, T. R. Northen, and J. F. Banfield. 2019. Mediterranean grassland soil C–N compound turnover is dependent on rainfall and depth, and is mediated by genomically divergent microorganisms. Nature Microbiology:1.

Dilks, K., R. W. Rose, E. Hartmann, and M. Pohlschröder. 2003. Prokaryotic utilization of the twin-arginine translocation pathway: A genomic survey. Journal of Bacteriology 185:1478–1483.

Dinis, J. M., D. E. Barton, J. Ghadiri, D. Surendar, K. Reddy, F. Velasquez, C. L. Chaffee, M. C. W. Lee, H. Gavrilova, H. Ozuna, S. A. Smits, and C. C. Ouverney. 2011. In search of an uncultured human-associated TM7 bacterium in the environment. PLoS ONE 6.

Djikeng, A., R. Kuzmickas, N. G. Anderson, and D. J. Spiro. 2009. Metagenomic Analysis of RNA Viruses in a Fresh Water Lake. PLoS ONE 4:e7264.

Dondini, M., K.-J. Van Groenigen, I. Del Galdo, and M. B. Jones. 2009. Carbon sequestration under Miscanthus: a study of 13C distribution in soil aggregates. GCB Bioenergy 1:321–330.

Van Dongen, S. 2008. Graph Clustering Via a Discrete Uncoupling Process. SIAM Journal on Matrix Analysis and Applications 30:121–141.

Douglas, C. M. 2001. Fungal B (1,3)-D-glucan synthesis. Medical Mycology 39:55–66.

Dudek, N. K., C. L. Sun, D. Burstein, R. S. Kantor, D. S. Aliaga Goltsman, E. M. Bik, B. C. Thomas, J. F. Banfield, and D. A. Relman. 2017. Novel Microbial Diversity and Functional Potential in the Marine Mammal Oral Microbiome. Current Biology 27:3752–3762.e6.

Edgar, R. C. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics

26:2460–2461.
Egert, M., A. A. De Graaf, A. Maathuis, P. De Waard, C. M. Plugge, H. Smidt, N. E. P. Deutz, C. Dijkema, W. M. De Vos, and K. Venema. 2007. Identification of glucose-fermenting bacteria present in an in vitro model of the human intestine by RNA-stable isotope probing. FEMS Microbiology Ecology 60:126–135.
El-Gebali, S., J. Mistry, A. Bateman, S. R. Eddy, A. Luciani, S. C. Potter, M. Qureshi, L. J. Richardson, G. A. Salazar, A. Smart, E. L. L. Sonnhammer, L. Hirsh, L. Paladin, D. Piovesan, S. C. E. Tosatto, and R. D. Finn. 2019. The Pfam protein families database in 2019. Nucleic Acids Research 47:D427–D432.
Elbeaino, T., M. Digiaro, N. Mielke-Ehret, H. P. Muehlbach, and G. P. Martelli. 2018. ICTV virus taxonomy profile: Fimoviridae. Journal of General Virology 99:1478–1479.
Emerson, J. B., S. Roux, J. R. Brum, B. Bolduc, B. J. Woodcroft, H. Bin Jang, C. M. Singleton, L. M. Solden, A. E. Naas, J. A. Boyd, S. B. Hodgkins, R. M. Wilson, G. Trubl, C. Li, S. Frolking, P. B. Pope, K. C. Wrighton, P. M. Crill, J. P. Chanton, S. R. Saleska, G. W. Tyson, V. I. Rich, and M. B. Sullivan. 2018. Host-linked soil viral ecology along a permafrost thaw gradient. Nature Microbiology 3:870–880.
Esmaeel, Q., M. Pupin, N. P. Kieu, G. Chataigné, M. Béchet, J. Deravel, F. Krier, M. Höfte, P. Jacques, and V. Leclère. 2016. Burkholderia genome mining for nonribosomal peptide synthetases reveals a great potential for novel siderophores and lipopeptides synthesis. MicrobiologyOpen 5:512–26.
Farag, M. A., C.-M. Ryu, L. W. Sumner, and P. W. Paré. 2006. GC–MS SPME profiling of rhizobacterial volatiles reveals prospective inducers of growth promotion and induced systemic resistance in plants. Phytochemistry 67:2262–2268.
Ferrari, B., T. Winsley, M. Ji, and B. Neilan. 2014. Insights into the distribution and abundance of the ubiquitous candidatus Saccharibacteria phylum following tag pyrosequencing. Scientific Reports 4:3957.
Finn, R. D., J. Clements, and S. R. Eddy. 2011. HMMER web server: Interactive sequence similarity searching. Nucleic Acids Research 39:W29-37.
Flores, H., and R. A. Chapman. 1968. Population development of Xiphinema americanum in relation to its role as a vector of tobacco ringspot virus. Phytopathology 58:814–817.
Flot, J.-F., B. Hespeels, X. Li, B. Noel, I. Arkhipova, E. G. J. Danchin, A. Hejnol, B. Henrissat, R. Koszul, J.-M. Aury, V. Barbe, R.-M. Barthélémy, J. Bast, G. A. Bazykin, O. Chabrol, A. Couloux, M. Da Rocha, C. Da Silva, E. Gladyshev, P. Gouret, O. Hallatschek, B. Hecox-Lea, K. Labadie, B. Lejeune, O. Piskurek, J. Poulain, F. Rodriguez, J. F. Ryan, O. A. Vakhrusheva, E. Wajnberg, B. Wirth, I. Yushenova, M. Kellis, A. S. Kondrashov, D. B. Mark Welch, P. Pontarotti, J. Weissenbach, P. Wincker, O. Jaillon, and K. Van Doninck. 2013. Genomic evidence for ameiotic evolution in the bdelloid rotifer Adineta vaga. Nature 500:453–457.
Forget, L., J. Ustinova, Z. Wang, V. A. R. Huss, and B. F. Lang. 2002. Hyaloraphidium curvatum: A linear mitochondrial genome, tRNA editing, and an evolutionary link to lower fungi. Molecular Biology and Evolution 19:310–319.
Fuhrman, J. A. 1999. Marine viruses and their biogeochemical and ecological effects. Nature 399:541–548.
Gagkaeva, T. Y., O. P. Gavrilova, A. S. Orina, E. V. Blinova, and I. G. Loskutov. 2017. Response of wild Avena species to fungal infection of grain. Crop Journal 5:499–508.
Le Gall, O., P. Christian, C. M. Fauquet, A. M. Q. King, N. J. Knowles, N. Nakashima, G.

Stanway, and A. E. Gorbalenya. 2008. Picornavirales, a proposed order of positive-sense single-stranded RNA viruses with a pseudo-T = 3 virion architecture. Archives of Virology 153:715–727.
Ghabrial, S. A., and N. Suzuki. 2009. Viruses of Plant Pathogenic Fungi. Annual Review of Phytopathology 47:353–384.
Di Giallonardo, F., and E. C. Holmes. 2015. Viral biocontrol: Grand experiments in disease emergence and evolution. Trends in Microbiology 23:83–90.
Glare, T., J. Caradus, W. Gelernter, T. Jackson, N. Keyhani, J. Köhl, P. Marrone, L. Morin, and A. Stewart. 2012. Have biopesticides come of age? Trends in Biotechnology 30:250–258.
Godwin, S., A. Kang, L.-M. Gulino, M. Manefield, M.-L. Gutierrez-Zamora, M. Kienzle, D. Ouwerkerk, K. Dawson, and A. V Klieve. 2014. Investigation of the microbial metabolism of carbon dioxide and hydrogen in the kangaroo foregut by stable isotope probing. The ISME Journal 8:1855–1865.
Gordon, B. R. G., R. Imperial, L. Wang, W. W. Navarre, and J. Liu. 2008. Lsr2 of Mycobacterium represents a novel class of H-NS-like proteins. Journal of Bacteriology 190:7052–7059.
Green, E. R., and J. Mecsas. 2016. Bacterial Secretion Systems: An Overview. Microbiology spectrum 4.
Guidi, L., S. Chaffron, L. Bittner, D. Eveillard, A. Larhlimi, S. Roux, Y. Darzi, S. Audic, L. Berline, J. R. Brum, L. P. Coelho, J. C. I. Espinoza, S. Malviya, S. Sunagawa, C. Dimier, S. Kandels-Lewis, M. Picheral, J. Poulain, S. Searson, L. Stemmann, F. Not, P. Hingamp, S. Speich, M. Follows, L. Karp-Boss, E. Boss, H. Ogata, S. Pesant, J. Weissenbach, P. Wincker, S. G. Acinas, P. Bork, C. De Vargas, D. Iudicone, M. B. Sullivan, J. Raes, E. Karsenti, C. Bowler, and G. Gorsky. 2016. Plankton networks driving carbon export in the oligotrophic ocean. Nature 532:465–470.
Gustavsen, J. A., D. M. Winget, X. Tian, and C. A. Suttle. 2014. High temporal and spatial diversity in marine RNA viruses implies that they have an important role in mortality and structuring plankton communities. Frontiers in Microbiology 5:703.
Hadziavdic, K., K. Lekang, A. Lanzen, I. Jonassen, E. M. Thompson, and C. Troedsson. 2014. Characterization of the 18s rRNA gene for designing universal eukaryote specific primers. PLoS ONE 9:e87624.
Haichar, F. E. Z., C. Marol, O. Berge, J. I. Rangel-Castro, J. I. Prosser, J. Balesdent, T. Heulin, and W. Achouak. 2008. Plant host habitat and root exudates shape soil bacterial community structure. ISME Journal 2:1221–1230.
Haichar, F. el Z., M.-A. Roncato, and W. Achouak. 2012. Stable isotope probing of bacterial community structure and gene expression in the rhizosphere of Arabidopsis thaliana. FEMS Microbiology Ecology 81:291–302.
Haig, S.-J., M. Schirmer, R. D'Amore, J. Gibbs, R. L. Davies, G. Collins, and C. Quince. 2015. Stable-isotope probing and metagenomics reveal predation by protozoa drives E. coli removal in slow sand filters. The ISME journal 9:797–808.
Harrell, F. E. 2019. Package "Hmisc."
Harrison, R. L., M. A. Keena, and D. L. Rowley. 2014. Classification, genetic variation and pathogenicity of Lymantria dispar nucleopolyhedrovirus isolates from Asia, Europe, and North America. Journal of Invertebrate Pathology 116:27–35.
He, X., J. S. McLean, A. Edlund, S. Yooseph, A. P. Hall, S.-Y. Liu, P. C. Dorrestein, E. Esquenazi, R. C. Hunter, G. Cheng, K. E. Nelson, R. Lux, and W. Shi. 2014. Cultivation of

a human-associated TM7 phylotype reveals a reduced genome and epibiotic parasitic lifestyle. Proceedings of the National Academy of Sciences 112:244–249.
Hebert, P. D. N., A. Cywinska, S. L. Ball, and J. R. DeWaard. 2003. Biological identifications through DNA barcodes. Proceedings of the Royal Society B: Biological Sciences 270:313–321.
Heikal, A., Y. Nakatani, E. Dunn, M. R. Weimar, C. L. Day, E. N. Baker, J. S. Lott, L. A. Sazanov, and G. M. Cook. 2014. Structure of the bacterial type II NADH dehydrogenase: A monotopic membrane protein with an essential role in energy generation. Molecular Microbiology 91:950–964.
Hernández, M., M. G. Dumont, Q. Yuan, and R. Conrad. 2015. Different bacterial populations associated with the roots and rhizosphere of rice incorporate plant-derived carbon. Applied and Environmental Microbiology 81:2244–2253.
Herrero, N. 2016. A novel monopartite dsRNA virus isolated from the entomopathogenic and nematophagous fungus Purpureocillium lilacinum. Archives of Virology 161:3375–3384.
Herrmann, E., W. Young, D. Rosendale, V. Reichert-Grimm, C. U. Riedel, R. Conrad, and M. Egert. 2017. RNA-Based Stable Isotope Probing Suggests *Allobaculum* spp. as Particularly Active Glucose Assimilators in a Complex Murine Microbiota Cultured In Vitro. BioMed Research International 2017:1–13.
Hillman, B. I., S. Supyani, H. Kondo, and N. Suzuki. 2003. A Reovirus of the Fungus Cryphonectria parasitica That Is Infectious as Particles and Related to the Coltivirus Genus of Animal Pathogens. Journal of Virology 78:892–898.
Ho, B. T., T. G. Dong, and J. J. Mekalanos. 2014. A view to a kill: The bacterial type VI secretion system. Cell Host and Microbe 15:9–21.
Hug, L. A., B. J. Baker, K. Anantharaman, C. T. Brown, A. J. Probst, C. J. Castelle, C. N. Butterfield, A. W. Hernsdorf, Y. Amano, K. Ise, Y. Suzuki, N. Dudek, D. A. Relman, K. M. Finstad, R. Amundson, B. C. Thomas, and J. F. Banfield. 2016. A new view of the tree of life. Nature Microbiology 1:16048.
Hugenholtz, P., G. W. Tyson, R. I. Webb, A. M. Wagner, and L. L. Blackall. 2001. Investigation of candidate division TM7, a recently recognized major lineage of the domain Bacteria, with no known pure-culture representatives. Applied and Environmental Microbiology 67:411–419.
Hungate, B. A., R. L. Mau, E. Schwartz, J. G. Caporaso, P. Dijkstra, N. van Gestel, B. J. Koch, C. M. Liu, T. A. McHugh, J. C. Marks, E. M. Morrissey, and L. B. Price. 2015. Quantitative microbial ecology through stable isotope probing. Applied and environmental microbiology 81:7570–81.
Hünninghaus, M., D. Dibbern, S. Kramer, R. Koller, J. Pausch, B. Schloter-Hai, T. Urich, E. Kandeler, M. Bonkowski, and T. Lueders. 2019. Disentangling carbon flow across microbial kingdoms in the rhizosphere of maize. Soil Biology and Biochemistry 134:122–130.
Hyatt, D., G. L. Chen, P. F. LoCascio, M. L. Land, F. W. Larimer, and L. J. Hauser. 2010. Prodigal: Prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119.
Iwata, S., M. Wikström, J. Abramson, S. Riistama, G. Larsson, A. Jasaitis, M. Svensson-Ek, L. Laakkonen, and A. Puustinen. 2000. The structure of the ubiquinol oxidase from *Escherichia coli* and its ubiquinone binding site. Nature Structural Biology 7:910–917.
Jaeger, C. H., S. E. Lindow, W. Miller, E. Clark, and M. K. Firestone. 1999. Mapping of sugar

and amino acid availability in soil around roots with bacterial sensors of sucrose and tryptophan. Applied and Environmental Microbiology 65:2685–2690.
Jamet, A., and X. Nassif. 2015. New Players in the Toxin Field: Polymorphic Toxin Systems in Bacteria. mBio 6:1–8.
Jones, D. T., W. R. Taylor, and J. M. Thornton. 1992. The rapid generation of mutation data matrices. Bioinformatics 8:275–282.
Kang, B. R., A. J. Anderson, and Y. C. Kim. 2018. Hydrogen Cyanide Produced by Pseudomonas chlororaphis O6 Exhibits Nematicidal Activity against Meloidogyne hapla. The plant pathology journal 34:35–43.
Kang, D. D., J. Froula, R. Egan, and Z. Wang. 2015. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ 3:e1165.
Kantor, R. S., K. C. Wrighton, K. M. Handley, I. Sharon, L. A. Hug, C. J. Castelle, B. C. Thomas, and J. F. Banfield. 2013. Small Genomes and Sparse Metabolisms of Sediment-Associated Bacteria from Four Candidate Phyla. mBio 4.
Kantor, R. S., A. W. van Zyl, R. P. van Hille, B. C. Thomas, S. T. L. Harrison, and J. F. Banfield. 2015. Bioreactor microbial ecosystems for thiocyanate and cyanide degradation unravelled with genome-resolved metagenomics. Environmental Microbiology 17:4929–4941.
Karpouzas, D. G., A. Karatasas, E. Spiridaki, C. Rousidou, F. Bekris, M. Omirou, C. Ehaliotis, and K. K. Papadopoulou. 2011. Impact of a beneficial and of a pathogenic Fusarium strain on the fingerprinting-based structure of microbial communities in tomato (Lycopersicon esculentum Milll.) rhizosphere. European Journal of Soil Biology 47:400–408.
Katoh, K., and D. M. Standley. 2013a. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Molecular Biology and Evolution 30:772–780.
Katoh, K., and D. M. Standley. 2013b. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Molecular Biology and Evolution 30:772–780.
Kaur, D., M. E. Guerin, H. Škovierová, P. J. Brennan, and M. Jackson. 2009. Chapter 2 Biogenesis of the Cell Wall and Other Glycoconjugates of Mycobacterium tuberculosis. Advances in Applied Microbiology 69:23–78.
Kawahigashi, M., K. Kaiser, A. Rodionov, and G. Guggenberger. 2006. Sorption of dissolved organic matter by mineral soils of the Siberian forest tundra. Global Change Biology 12:1868–1877.
Kazaks, A., T. Voronkova, J. Rumnieks, A. Dishlers, and K. Tars. 2011. Genome Structure of Caulobacter Phage phiCb5. Journal of Virology 85:4628–4631.
Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Meintjes, and A. Drummond. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649.
Kindaichi, T., S. Yamaoka, R. Uehara, N. Ozaki, A. Ohashi, M. Albertsen, P. H. Nielsen, and J. L. Nielsen. 2016. Phylogenetic diversity and ecophysiology of Candidate phylum Saccharibacteria in activated sludge. FEMS Microbiology Ecology 92:1–11.
Klovins, J., G. P. Overbeek, S. H. E. van den Worm, H. W. Ackermann, and J. van Duin. 2002. Nucleotide sequence of a ssRNA phage from Acinetobacter: Kinship to coliphages. Journal of General Virology 83:1523–1533.
Koonin, E. V. 1991. The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses. Journal of General Virology 72:2197–2206.

Krishnamurthy, S. R., A. B. Janowski, G. Zhao, D. Barouch, and D. Wang. 2016. Hyperexpansion of RNA Bacteriophage Diversity. PLoS Biology 14:e1002409.
Kroeger, M. E., T. O. Delmont, A. M. Eren, K. M. Meyer, J. Guo, K. Khan, J. L. M. Rodrigues, B. J. M. Bohannan, S. G. Tringe, C. D. Borges, J. M. Tiedje, S. M. Tsai, and K. Nüsslein. 2018. New Biological Insights Into How Deforestation in Amazonia Affects Soil Microbial Communities Using Metagenomics and Metagenome-Assembled Genomes. Frontiers in Microbiology 9:1635.
Laforest, M. J., I. Roewer, and B. Franz Lang. 1997. Mitochondrial tRNAs in the lower fungus Spizellomyces punctatus: TRNA editing and UAG "stop" codons recognized as leucine. Nucleic Acids Research 25:626–632.
Lang, A. S., M. L. Rise, A. I. Culley, and G. F. Steward. 2009. RNA viruses in the sea. FEMS Microbiology Reviews 33:295–323.
Langmead, B., and S. L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. Nature methods 9:357–9.
Lee, C. G., T. Watanabe, Y. Fujita, S. Asakawa, and M. Kimura. 2012. Heterotrophic growth of cyanobacteria and phage-mediated microbial loop in soil: Examination by stable isotope probing (SIP) method. Soil Science and Plant Nutrition 58:161–168.
Leray, M., J. Y. Yang, C. P. Meyer, S. C. Mills, N. Agudelo, V. Ranwez, J. T. Boehm, and R. J. Machida. 2013. A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: Application for characterizing coral reef fish gut contents. Frontiers in Zoology 10:34.
Letunic, I., and P. Bork. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic acids research 44:W242–W245.
Love, M. I., W. Huber, and S. Anders. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology 15:550.
Lowe, T. M., and S. R. Eddy. 1996. TRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Research 25:955–964.
Luo, C., S. Xie, W. Sun, X. Li, and A. M. Cupples. 2009. Identification of a novel toluene-degrading bacterium from the candidate phylum TM7, as determined by DNA stable isotope probing. Applied and Environmental Microbiology 75:4644–4647.
Magrane, M., and U. P. Consortium. 2011. UniProt Knowledgebase: A hub of integrated protein data. Database 2011:bar009-bar009.
Mäntynen, S., L. R. Sundberg, and M. M. Poranen. 2018. Recognition of six additional cystoviruses: Pseudomonas virus phi6 is no longer the sole species of the family Cystoviridae. Archives of Virology 163:1117–1124.
Marcy, Y., C. Ouverney, E. M. Bik, T. Losekann, N. Ivanova, H. G. Martin, E. Szeto, D. Platt, P. Hugenholtz, D. A. Relman, and S. R. Quake. 2007. Dissecting biological "dark matter" with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth. Proceedings of the National Academy of Sciences 104:11889–11894.
Marquez, L. M., R. S. Redman, R. J. Rodriguez, and M. J. Roossinck. 2007. A Virus in a Fungus in a Plant: Three-Way Symbiosis Required for Thermal Tolerance. Science 315:513–515.
Martínez-Álvarez, P., E. J. Vainio, L. Botella, J. Hantula, and J. J. Diez. 2014. Three mitovirus strains infecting a single isolate of Fusarium circinatum are the first putative members of the family Narnaviridae detected in a fungus of the genus Fusarium. Archives of Virology 159:2153–2155.
Matz, C., and S. Kjelleberg. (n.d.). Off the hook-how bacteria survive protozoan grazing.

Mayer, S., W. Steffen, J. Steuber, and F. Götz. 2015. The staphylococcus aureus nuoL-like protein MpsA contributes to the generation of membrane potential. Journal of Bacteriology 197:794–806.
Medema, M. H., K. Blin, P. Cimermancic, V. de Jager, P. Zakrzewski, M. A. Fischbach, T. Weber, E. Takano, and R. Breitling. 2011. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Research 39:W339–W346.
Miller, M. A., W. Pfeiffer, and T. Schwartz. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Page 2010 Gateway Computing Environments Workshop, GCE 2010.
Min, X. J., and D. A. Hickey. 2007. Assessing the effect of varying sequence length on DNA barcoding of fungi: Barcoding. Molecular Ecology Notes 7:365–373.
Moniruzzaman, M., L. L. Wurch, H. Alexander, S. T. Dyhrman, C. J. Gobler, and S. W. Wilhelm. 2017. Virus-host relationships of marine single-celled eukaryotes resolved from metatranscriptomics. Nature Communications 8.
Mu, F., J. Xie, S. Cheng, M. P. You, M. J. Barbetti, J. Jia, Q. Wang, J. Cheng, Y. Fu, T. Chen, and D. Jiang. 2018. Virome characterization of a collection of Sclerotinia sclerotiorum from Australia. Frontiers in Microbiology 8:2540.
Mur, L. A. J., T. L. W. Carver, and E. Prats. 2006. NO way to live; the various roles of nitric oxide in plant–pathogen interactions. Journal of Experimental Botany 57:489–505.
Murase, J., M. Shibata, C. G. Lee, T. Watanabe, S. Asakawa, and M. Kimura. 2012. Incorporation of plant residue-derived carbon into the microeukaryotic community in a rice field soil revealed by DNA stable-isotope probing. FEMS Microbiology Ecology 79:371–379.
Nandi, M., C. Selin, G. Brawerman, W. G. D. Fernando, and T. de Kievit. 2017. Hydrogen cyanide, which contributes to Pseudomonas chlororaphis strain PA23 biocontrol, is upregulated in the presence of glycine. Biological Control 108:47–54.
Nawrocki, E. P. 2009. Structural RNA Homology Search and Alignment using Covariance Models. Ph.D. thesis:282.
Nawrocki, E. P., D. L. Kolbe, and S. R. Eddy. 2009. Infernal 1.0: Inference of RNA alignments. Bioinformatics 25:1335–1337.
Nazir, R., S. Mazurier, P. Yang, P. Lemanceau, and J. D. van Elsas. 2017. The ecological role of type three secretion systems in the interaction of bacteria with fungi in soil and related habitats is diverse and context-dependent. Frontiers in Microbiology 8:38.
Nibert, M. L., M. Vong, K. K. Fugate, and H. J. Debat. 2018. Evidence for contemporary plant mitoviruses. Virology 518:14–24.
Nordman, J., and A. Wright. 2008. The relationship between dNTP pool levels and mutagenesis in an Escherichia coli NDP kinase mutant. Proceedings of the National Academy of Sciences 105:10197–10202.
Nuccio, E. E., E. Starr, U. Karaoz, E. L. Brodie, J. Zhou, S. Tringe, R. R. Malmstrom, T. Woyke, J. F. Banfield, M. K. Firestone, and J. Pett-Ridge. 2019. Niche differentiation is spatially and temporally regulated in the rhizosphere. bioRxiv:611863.
Ogata, H., S. Goto, K. Sato, W. Fujibuchi, H. Bono, and M. Kanehisa. 1999. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Research 27:29–34.
Oksanen, J., F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. Mcglinn, P. R. Minchin, R. B. O'hara, G. L. Simpson, P. Solymos, M. Henry, H. Stevens, E. Szoecs, and H. W.

Maintainer. 2019. Package “vegan” Title Community Ecology Package.
Osaki, H., A. Sasaki, K. Nomiyama, and K. Tomioka. 2016. Multiple virus infection in a single strain of Fusarium poae shown by deep sequencing. Virus Genes 52:835–847.
Pandey, S. S., P. K. Patnana, R. Rai, and S. Chatterjee. 2017. Xanthoferrin, the α-hydroxycarboxylate-type siderophore of Xanthomonas campestris pv. campestris, is required for optimum virulence and growth inside cabbage. Molecular Plant Pathology 18:949–962.
Patridge, E. V, and J. G. Ferry. 2006. Is an NAD (P) H : Quinone Oxidoreductase. Society 188:3498–3506.
Peng, Y., H. C. M. Leung, S. M. Yiu, and F. Y. L. Chin. 2012. IDBA-UD: A de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics 28:1420–1428.
Peregrín-Alvarez, J. M., C. Sanford, and J. Parkinson. 2009. The conservation and evolutionary modularity of metabolism. Genome Biology 10:R63.
Placella, S. A., E. L. Brodie, and M. K. Firestone. 2012. Rainfall-induced carbon dioxide pulses result from sequential resuscitation of phylogenetically clustered microbial groups. Proceedings of the National Academy of Sciences 109:10931–10936.
Probst, A. J., B. Ladd, J. K. Jarett, D. E. Geller-Mcgrath, C. M. K. Sieber, J. B. Emerson, K. Anantharaman, B. C. Thomas, R. R. Malmstrom, M. Stieglmeier, A. Klingl, T. Woyke, M. C. Ryan, and J. F. Banfield. 2018. Differential depth distribution of microbial function and putative symbionts through sediment-hosted aquifers in the deep terrestrial subsurface. Nature Microbiology 3:328–336.
R Core Team. 2014. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
Rittenour, W. R., M. Chen, E. B. Cahoon, and S. D. Harris. 2011. Control of Glucosylceramide Production and Morphogenesis by the Bar1 Ceramide Synthase in Fusarium graminearum. PLoS ONE 6:e19385.
Robba, L., S. J. Russell, G. L. Barker, and J. Brodie. 2006. Assessing the use of the mitochondrial cox1 marker for use in DNA barcoding of red algae (Rhodophyta). American Journal of Botany 93:1101–1108.
Robideau, G. P., A. W. A. M. De Cock, M. D. Coffey, H. Voglmayr, H. Brouwer, K. Bala, D. W. Chitty, N. Désaulniers, Q. A. Eggertson, C. M. M. Gachon, C. H. Hu, F. C. Küpper, T. L. Rintoul, E. Sarhan, E. C. P. Verstappen, Y. Zhang, P. J. M. Bonants, J. B. Ristaino, and C. André Lévesque. 2011. DNA barcoding of oomycetes with cytochrome c oxidase subunit I and internal transcribed spacer. Molecular Ecology Resources 11:1002–1011.
Rodríguez-Cousiño, N., M. Maqueda, J. Ambrona, E. Zamora, R. Esteban, and M. Ramírez. 2011. A new wine Saccharomyces cerevisiae killer toxin (Klus), encoded by a double-stranded rna virus, with broad antifungal activity is evolutionarily related to a chromosomal host gene. Applied and environmental microbiology 77:1822–32.
Rodríguez-Navarro, D. N., M. S. Dardanelli, and J. E. Ruíz-Saínz. 2007. Attachment of bacteria to the roots of higher plants. FEMS Microbiology Letters 272:127–136.
Rogers, H. J., K. W. Buck, and C. M. Brasier. 1987. A mitochondrial target for double-stranded RNA in diseased isolates of the fungus that causes Dutch elm disease. Nature 329:558–560.
Roossinck, M. J. 2019. Evolutionary and ecological links between plant and fungal viruses. New Phytologist 221:86–92.
Roux, S., F. Enault, B. L. Hurwitz, and M. B. Sullivan. 2015. VirSorter: mining viral signal from

microbial genomic data. PeerJ 3:e985.
Ruark, C. L., M. Gardner, M. G. Mitchum, E. L. Davis, and T. L. Sit. 2018. Novel RNA viruses within plant parasitic cyst nematodes. PLoS ONE 13:e0193881.
Rumnieks, J., and K. Tars. 2012. Diversity of pili-specific bacteriophages: genome sequence of IncM plasmid-dependent RNA phage M. BMC microbiology 12:1.
Ryu, C.-M., M. A. Farag, C.-H. Hu, M. S. Reddy, H.-X. Wei, P. W. Pare, and J. W. Kloepper. 2003. Bacterial volatiles promote growth in Arabidopsis. Proceedings of the National Academy of Sciences 100:4927–4932.
Sasse, J., E. Martinoia, and T. Northen. 2018. Feed Your Friends: Do Plant Exudates Shape the Root Microbiome? Trends in Plant Science 23:25–41.
Schauer, C., C. L. Thompson, and A. Brune. 2012. The Bacterial Community in the Gut of the Cockroach Shelfordella lateralis Reflects the Close Evolutionary Relatedness of Cockroaches and Termites. Applied and Environmental Microbiology 78:2758–2767.
Schildkraut, C. L., J. Marmur, and P. Doty. 1962. Determination of the base composition of deoxyribonucleic acid from its buoyant density in CsCl. Journal of Molecular Biology 4:430–443.
Schubotz, F., L. E. Hays, D. R. Meyer-Dombard, A. Gillespie, E. L. Shock, and R. E. Summons. 2015. Stable isotope labeling confirms mixotrophic nature of streamer biofilm communities at alkaline hot springs. Frontiers in Microbiology 6:42.
Seymour, J. R., L. Seuront, M. Doubell, R. L. Waters, and J. G. Mitchell. 2006. Microscale patchiness of virioplankton. Journal of the Marine Biological Association of the United Kingdom 86:551–561.
Shannon, P., A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker. 2003. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Research 13:2498–2504.
Sharma, A., B. . Johri, A. . Sharma, and B. . Glick. 2003. Plant growth-promoting bacterium Pseudomonas sp. strain GRP3 influences iron acquisition in mung bean (Vigna radiata L. Wilzeck). Soil Biology and Biochemistry 35:887–894.
Sharon, I., M. Kertesz, L. A. Hug, D. Pushkarev, T. A. Blauwkamp, C. J. Castelle, M. Amirebrahimi, B. C. Thomas, D. Burstein, S. G. Tringe, K. H. Williams, and J. F. Banfield. 2015. Accurate, multi-kb reads resolve complex populations and detect rare microorganisms. Genome Research 25:534–543.
Sharon, I., M. J. Morowitz, B. C. Thomas, E. K. Costello, D. A. Relman, and J. F. Banfield. 2013. Time series community genomics analysis reveals rapid shifts in bacterial species, strains, and phage during infant gut colonization. Genome Research 23:111–120.
Shi, C., C. Wang, X. Xu, B. Huang, L. Wu, and D. Yang. 2014. Comparison of bacterial communities in soil between nematode-infected and nematode-uninfected Pinus massoniana pinewood forest. Applied Soil Ecology 85:11–20.
Shi, M., X. D. Lin, J. H. Tian, L. J. Chen, X. Chen, C. X. Li, X. C. Qin, J. Li, J. P. Cao, J. S. Eden, J. Buchmann, W. Wang, J. Xu, E. C. Holmes, and Y. Z. Zhang. 2016a. Redefining the invertebrate RNA virosphere. Nature 540:539–543.
Shi, M., P. Neville, J. Nicholson, J.-S. Eden, A. Imrie, and E. C. Holmes. 2017. High-Resolution Metatranscriptomics Reveals the Ecological Dynamics of Mosquito-Associated RNA Viruses in Western Australia. Journal of Virology 91.
Shi, S., E. E. Nuccio, Z. J. Shi, Z. He, J. Zhou, and M. K. Firestone. 2016b. The interconnected rhizosphere: High network complexity dominates rhizosphere assemblages. Ecology Letters

19:926–936.
Shi, S., E. Nuccio, D. J. Herman, R. Rijkers, K. Estera, J. Li, U. N. da Rocha, Z. He, J. Pett-Ridge, E. L. Brodie, J. Zhou, and M. Firestone. 2015. Successional Trajectories of Rhizosphere Bacterial Communities over Consecutive Seasons. mBio 6.
Sieber, C. M. K., A. J. Probst, A. Sharrar, B. C. Thomas, M. Hess, S. G. Tringe, and J. F. Banfield. 2018. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. Nature Microbiology 3:836–843.
Silva, S. R., D. O. Alvarenga, Y. Aranguren, H. A. Penha, C. C. Fernandes, D. G. Pinheiro, M. T. Oliveira, T. P. Michael, V. F. O. Miranda, and A. M. Varani. 2017. The mitochondrial genome of the terrestrial carnivorous plant Utricularia reniformis (Lentibulariaceae): Structure, comparative analysis and evolutionary landmarks. PloS one 12:e0180484.
Sohlenkamp, C., K. A. Galindo-Lagunas, Z. Guan, P. Vinuesa, S. Robinson, J. Thomas-Oates, C. R. H. Raetz, and O. Geiger. 2007. The Lipid Lysyl-Phosphatidylglycerol Is Present in Membranes of Rhizobium tropici CIAT899 and Confers Increased Resistance to Polymyxin B Under Acidic Growth Conditions . Molecular Plant-Microbe Interactions 20:1421–1430.
Stamatakis, A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313.
Starr, E. P., E. E. Nuccio, J. Pett-Ridge, J. F. Banfield, and M. K. Firestone. 2019. Metatranscriptomic reconstruction reveals RNA viruses with the potential to shape carbon cycling in soil. bioRxiv:597468.
Starr, E. P., S. Shi, S. J. Blazewicz, A. J. Probst, D. J. Herman, M. K. Firestone, and J. F. Banfield. 2018. Stable isotope informed genome-resolved metagenomics reveals that Saccharibacteria utilize microbially-processed plant-derived carbon. Microbiome 6:122.
Stella, E. J., J. J. Franceschelli, S. E. Tasselli, and H. R. Morbidoni. 2013. Analysis of Novel Mycobacteriophages Indicates the Existence of Different Strategies for Phage Inheritance in Mycobacteria. PLoS ONE 8:e56384.
Steuber, J., G. Vohl, M. S. Casutt, T. Vorburger, K. Diederichs, and G. Fritz. 2014. Structure of the V. cholerae Na+-pumping NADH:quinone oxidoreductase. Nature 516:62–67.
Steward, G. F., A. I. Culley, J. A. Mueller, E. M. Wood-Charlson, M. Belcaid, and G. Poisson. 2013. Are we missing half of the viruses in the ocean? ISME Journal 7:672–679.
Stough, J. M. A., M. Kolton, J. E. Kostka, D. J. Weston, D. A. Pelletier, and S. W. Wilhelm. 2018. Diversity of Active Viral Infections within the Sphagnum Microbiome. Applied and Environmental Microbiology 84:1124–1142.
Suttle, C. A. 2007. Marine viruses--major players in the global ecosystem. Nature reviews. Microbiology 5:801–12.
Suzek, B. E., H. Huang, P. McGarvey, R. Mazumder, and C. H. Wu. 2007. UniRef: Comprehensive and non-redundant UniProt reference clusters. Bioinformatics 23:1282–1288.
Tamasloukht, M. 2003. Root Factors Induce Mitochondrial-Related Gene Expression and Fungal Respiration during the Developmental Switch from Asymbiosis to Presymbiosis in the Arbuscular Mycorrhizal Fungus Gigaspora rosea. Plant Physiology 131:1468–1478.
Thomsen, T. R., B. V. Kjellerup, J. L. Nielsen, P. Hugenholtz, and P. H. Nielsen. 2002. In situ studies of the phylogeny and physiology of filamentous bacteria with attached growth. Environmental Microbiology 4:383–391.
Tokarz, R., S. Sameroff, T. Tagliafierro, K. Jain, S. H. Williams, D. M. Cucura, I. Rochlin, J. Monzon, G. Carpi, D. Tufts, M. Diuk-Wasser, J. Brinkerhoff, and W. I. Lipkin. 2018.

Identification of Novel Viruses in Amblyomma americanum, Dermacentor variabili, and Ixodes scapularis Ticks. mSphere 3.
Trubl, G., H. Bin Jang, S. Roux, J. B. Emerson, N. Solonenko, D. R. Vik, L. Solden, J. Ellenbogen, A. T. Runyon, B. Bolduc, B. J. Woodcroft, S. R. Saleska, G. W. Tyson, K. C. Wrighton, M. B. Sullivan, and V. I. Rich. 2018. Soil Viruses Are Underexplored Players in Ecosystem Carbon Processing. mSystems 3:e00076-18.
Tyc, O., C. Song, J. S. Dickschat, M. Vos, and P. Garbeva. 2017. The Ecological Role of Volatile and Soluble Secondary Metabolites Produced by Soil Bacteria. Trends in microbiology 25:280–292.
Uhlik, O., M.-C. Leewis, M. Strejcek, L. Musilova, M. Mackova, M. B. Leigh, and T. Macek. 2013. Stable isotope probing in the metagenomics era: a bridge towards improved bioremediation. Biotechnology advances 31:154–65.
Valles, S. M., S. D. Porter, and L. A. Calcaterra. 2018. Prospecting for viral natural enemies of the fire ant Solenopsis invicta in Argentina. PLoS ONE 13:e0192377.
Wang, D., K. Pajerowska-Mukhtar, A. H. Culler, and X. Dong. 2007. Salicylic acid inhibits pathogen growth in plants through repression of the auxin signaling pathway. Current biology : CB 17:1784–90.
Wang, L., J. Zhang, H. Zhang, D. Qiu, and L. Guo. 2016. Two novel relative double-stranded RNA mycoviruses infecting Fusarium poae strain SX63. International Journal of Molecular Sciences 17.
Warmink, J. A., and J. D. van Elsas. 2009. Migratory Response of Soil Bacteria to Lyophyllum sp. Strain Karsten in Soil Microcosms. Applied and Environmental Microbiology 75:2820–2830.
Wei, J.-Z., K. Hale, L. Carta, E. Platzer, C. Wong, S.-C. Fang, and R. V Aroian. 2003. Bacillus thuringiensis crystal proteins that target nematodes. Proceedings of the National Academy of Sciences of the United States of America 100:2760–5.
Weinmaier, T., A. J. Probst, M. T. La Duc, D. Ciobanu, J.-F. Cheng, N. Ivanova, T. Rattei, and P. Vaishampayan. 2015. A viability-linked metagenomic analysis of cleanroom environments: eukarya, prokaryotes, and viruses. Microbiome 3:62.
Wen, A. M., and N. F. Steinmetz. 2016. Design of virus-based nanomaterials for medicine, biotechnology, and energy. Chemical Society Reviews 45:4074–4126.
White, R. A., E. M. Bottos, T. Roy Chowdhury, J. D. Zucker, C. J. Brislawn, C. D. Nicora, S. J. Fansler, K. R. Glaesemann, K. Glass, and J. K. Jansson. 2016. Moleculo Long-Read Sequencing Facilitates Assembly and Genomic Binning from Complex Soil Metagenomes. mSystems 1:e00045-16.
Whitfield, A. E., B. W. Falk, and D. Rotenberg. 2015. Insect vector-mediated transmission of plant viruses. Virology 479–480:278–289.
Whitman, T., R. Neurath, A. Perera, I. Chu-Jacoby, D. Ning, J. Zhou, P. Nico, J. Pett-Ridge, and M. Firestone. 2018. Microbial community assembly differs across minerals in a rhizosphere microcosm. Environmental Microbiology 20:4444–4460.
Wilhelm, R. C., E. Cardenas, H. Leung, A. Szeitz, L. D. Jensen, and W. W. Mohn. 2017. Corrigendum: Long-term enrichment of stress-tolerant cellulolytic soil populations following timber harvesting evidenced by multi-Omic stable isotope probing [(2017). Front. Microbiol. 8:537.] doi: 10.3389/fmicb.2017.00537. Frontiers in Microbiology 8:537.
Wilhelm, R. C., R. Singh, L. D. Eltis, and W. W. Mohn. 2019. Bacterial contributions to delignification and lignocellulose degradation in forest soils with metagenomic and

quantitative stable isotope probing. The ISME Journal 13:413–429.
Wilhelm, S. W., and C. A. Suttle. 2006. Viruses and Nutrient Cycles in the Sea. BioScience 49:781–788.
Williamson, K. E., J. J. Fuhrmann, K. E. Wommack, and M. Radosevich. 2017. Viruses in Soil Ecosystems: An Unknown Quantity Within an Unexplored Territory. Annual Review of Virology 4:201–219.
Williamson, K. E., M. Radosevich, and K. E. Wommack. 2005. Abundance and diversity of viruses in six Delaware soils. Applied and environmental microbiology 71:3119–25.
Winget, D. M., R. R. Helton, K. E. Williamson, S. R. Bench, S. J. Williamson, and K. E. Wommack. 2011. Repeating patterns of virioplankton production within an estuarine ecosystem. http://www.pnas.org/cgi/doi/10.1073/pnas.1101907108.
Wolf, Y. I., D. Kazlauskas, J. Iranzo, A. Lucía-Sanz, J. H. Kuhn, M. Krupovic, V. V. Dolja, and E. V. Koonin. 2018. Origins and Evolution of the Global RNA Virome. mBio 9.
Wrighton, K. C., B. C. Thomas, I. Sharon, C. S. Miller, C. J. Castelle, N. C. VerBerkmoes, M. J. Wilkins, R. L. Hettich, M. S. Lipton, K. H. Williams, P. E. Long, and J. F. Banfield. 2012. Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. Science 337:1661–1665.
Wu, M. D., L. Zhang, G. Q. Li, D. H. Jiang, M. S. Hou, and H.-C. Huang. 2007. Hypovirulence and Double-Stranded RNA in Botrytis cinerea . Phytopathology 97:1590–1599.
Wu, M., L. Zhang, G. Li, D. Jiang, and S. A. Ghabrial. 2010. Genome characterization of a debilitation-associated mitovirus infecting the phytopathogenic fungus Botrytis cinerea. Virology 406:117–126.
Wu, S., J. Xiong, and Y. Y. Yu. 2015. Taxonomic resolutions based on 18S rRNA Genes: A case study of subclass Copepoda. PLoS ONE 10:e0131498.
Wu, Y. W., B. A. Simmons, and S. W. Singer. 2016. MaxBin 2.0: An automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics 32:605–607.
Xue, Y., I. Jonassen, L. Øvreås, and N. Taş. 2019. Bacterial and Archaeal Metagenome-Assembled Genome Sequences from Svalbard Permafrost. Microbiology resource announcements 8:e00516-19.
Yang, A., N. Liu, Q. Tian, W. Bai, M. Williams, Q. Wang, L. Li, and W. H. Zhang. 2015. Rhizosphere bacterial communities of dominant steppe plants shift in response to a gradient of simulated nitrogen deposition. Frontiers in Microbiology 6:789.
Yin, Y., X. Mao, J. Yang, X. Chen, F. Mao, and Y. Xu. 2012. DbCAN: A web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Research 40:W445–W451.
Yu, N. Y., J. R. Wagner, M. R. Laird, G. Melli, S. Rey, R. Lo, P. Dao, S. Cenk Sahinalp, M. Ester, L. J. Foster, and F. S. L. Brinkman. 2010. PSORTb 3.0: Improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics 26:1608–1615.
Zdobnov, E. M., and R. Apweiler. 2001. InterProScan - An integration platform for the signature-recognition methods in InterPro. Bioinformatics 17:847–848.
Zeigler Allen, L., J. P. McCrow, K. Ininbergs, C. L. Dupont, J. H. Badger, J. M. Hoffman, M. Ekman, A. E. Allen, B. Bergman, and J. C. Venter. 2017. The Baltic Sea Virome: Diversity and Transcriptional Activity of DNA and RNA Viruses. mSystems 2.
Zhang, L., R. A. Baasiri, and N. K. Van Alfen. 1998. Viral repression of fungal pheromone

precursor gene expression. Molecular and cellular biology 18:953–9.
Zhang, R., S. Liu, S. Chiba, H. Kondo, S. Kanematsu, and N. Suzuki. 2014. A novel single-stranded RNA virus isolated from a phytopathogenic filamentous fungus, Rosellinia necatrix, with similarity to hypo-like viruses. Frontiers in Microbiology 5.