UC Berkeley
UC Berkeley Previously Published Works

Title

Quantitative Stable-Isotope Probing (qSIP) with Metagenomics Links Microbial Physiology and Activity to Soil Moisture in Mediterranean-Climate Grassland Ecosystems

Permalink

Journal

ISSN

Authors

Greenlon, Alex
Sieradzki, Ella
Zablocki, Olivier
et al.

Publication Date

DOI

Quantitative Stable-Isotope Probing (qSIP) with Metagenomics Links Microbial Physiology and Activity to Soil Moisture in Mediterranean-Climate Grassland Ecosystems

Alex Greenlon,[a] Ella Sieradzki,[a] Olivier Zablocki,[b,c] Benjamin J. Koch,[d,e] Megan M. Foley,[d,e] Jeffrey A. Kimbrel,[f] Bruce A. Hungate,[d,e] Steven J. Blazewicz,[f] Erin E. Nuccio,[f] Christine L. Sun,[b,c] Aaron Chew,[a] Cynthia-Jeanette Mancilla,[a] Matthew B. Sullivan,[b,c,i] Mary Firestone,[a] Jennifer Pett-Ridge,[f,g] Jillian F. Banfield[a,h]

[a]Department of Environmental Science, Policy and Management, University California, Berkeley, Berkley, California, USA
[b]Department of Microbiology, Ohio State University, Columbus, Ohio, USA
[c]Center of Microbiome Science, Ohio State University, Columbus, Ohio, USA
[d]Center for Ecosystem Science and Society, Northern Arizona University, Flagstaff, Arizona, USA
[e]Department of Biological Sciences, Northern Arizona University, Flagstaff, Arizona, USA
[f]Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, California, USA
[g]Life & Environmental Sciences Department, University of California, Merced, Merced, California, USA
[h]Department of Earth and Planetary Science, University of California, Berkeley, Berkley, California, USA
[i]Department of Civil, Environmental and Geodetic Engineering, Ohio State University, Columbus, Ohio, USA

ABSTRACT The growth and physiology of soil microorganisms, which play vital roles in biogeochemical cycling, are shaped by both current and historical soil environmental conditions. Here, we developed and applied a genome-resolved metagenomic implementation of quantitative stable isotope probing (qSIP) with an $H_2^{18}O$ labeling experiment to identify actively growing soil microorganisms and their genomic capacities. qSIP enabled measurement of taxon-specific growth because isotopic incorporation into microbial DNA requires production of new genome copies. We studied three Mediterranean grassland soils across a rainfall gradient to evaluate the hypothesis that historic precipitation levels are an important factor controlling trait selection. We used qSIP-informed genome-resolved metagenomics to resolve the active subset of soil community members and identify their characteristic ecophysiological traits. Higher year-round precipitation levels correlated with higher activity and growth rates of flagellar motile microorganisms. In addition to heavily isotopically labeled bacteria, we identified abundant isotope-labeled phages, suggesting phage-induced cell lysis likely contributed to necromass production at all three sites. Further, there was a positive correlation between phage activity and the activity of putative phage hosts. Contrary to our expectations, the capacity to decompose the diverse complex carbohydrates common in soil organic matter or oxidize methanol and carbon monoxide were broadly distributed across active and inactive bacteria in all three soils, implying that these traits are not highly selected for by historical precipitation.

IMPORTANCE Soil moisture is a critical factor that strongly shapes the lifestyle of soil organisms by changing access to nutrients, controlling oxygen diffusion, and regulating the potential for mobility. We identified active microorganisms in three grassland soils with similar mineral contexts, yet different historic rainfall inputs, by adding water labeled with a stable isotope and tracking that isotope in DNA of growing microbes. By examining the genomes of active and inactive microorganisms, we identified functions that are enriched in growing organisms, and showed that different functions were selected for in different soils. Wetter soil had higher activity of motile organisms, but activity of pathways for degradation of soil organic carbon compounds, including simple carbon substrates, were comparable for all three soils. We identified many labeled, and thus active bacteriophages (viruses that infect bacteria), implying that the cells they

killed contributed to soil organic matter. The activity of these bacteriophages was significantly correlated with activity of their hosts.

Soils are among the most diverse microbial ecosystems, and microbial communities modulate the properties of soil that define its capacity to support terrestrial macro ecosystems, and human agro-ecosystems (1). Microbial communities contribute to global biogeochemical cycles (2, 3) as well as soil ecosystem services (e.g., moisture retention, nutrient availability, and structure) (4). Understanding microbial communities and traits along environmental gradients is foundational to predicting how soil biogeochemical processes will be altered by climate change (5). Microbial traits underlie soil organic matter biogenesis and turnover via the production and degradation of polymers (e.g., extracellular polysaccharides, chitin) and the varied processes that contribute to cell death (6). In fact, emerging paradigms of soil organic carbon (SOC) biogenesis suggest that microbial necromass constitutes much of SOC (7–9). Yet critical gaps remain in our capacity to link microbial functional capabilities to *in situ* measurements of microbial growth and mortality, and to differentiate the active versus inactive viral populations that drive microbial community dynamics and responses to environmental variation.

Soil ecology studies frequently attempt to link surveys of microbial community composition with measurements of environmental parameters. For example, amplification and sequencing of 16S rRNA gene sequences from environmental DNA was used to show that soil pH contributes strongly to microbial community structure across diverse soils (10), and that many 16S rRNA gene-based microbial phylotypes abundant in soils are common across global soil biomes and edaphic factors (11). Sequencing of phylogenetic marker genes and whole genomes of cultured soil microbial populations has also revealed the effects of latitude (12, 13) and soil parent material (14) on microbial community composition (15).

A limitation of phylogenetic marker-based studies is their inability to robustly predict microbial traits of actively growing taxa, i.e., the functional capacities of soil microbial communities that are relevant to the biogeochemical properties of their broader soil ecosystem. Further, marker genes such as 16S rRNA tend to be slow evolving; thus, any evolutionary traits they predict would likely be ancient. More fast-evolving traits are more likely to be useful in defining microbial niches (16). Genome-resolved metagenomics provides a route to access these more rapidly evolving traits and enables predictions of the sets of capacities of individual soil organisms without the requirement for cultivation (17, 18), including for organisms missed in marker-gene amplicon studies (19) and viruses (20, 21). However, functional inferences from many typical "meta-omics" studies are limited due to their lack of information about which organisms are active. Although tools such as iRep (19) can use DNA sequence coverage to provide insight into replication rates, these methods have limitations (22) and are not very effective in soil studies due to genome completeness and coverage requirements. Metaproteomics measurements can detect abundant proteins, and thus infer bacterial activity (17), but the insights are often limited by extraction bias and identification of proteins from only a small subset of the most abundant soil microbes. Soil metatranscriptomics, a more encompassing and taxonomically informative analysis, may reveal how functions are expressed in space and time, e.g., that carbohydrate decomposition is conducted by distinct guilds of taxa that operate in different soil niches (16). However, gene expression cannot be directly linked to growth rates (23). Bio-orthogonal noncanonical amino-acid tagging (BONCAT) can tag cells that are actively synthesizing proteins, sort them, and sequence marker genes. This method has been applied to soils and revealed that as many as 34% of cells in soil are translationally active at any time (24), but it cannot link directly to substrate usage and biosynthesis. In

contrast to the approaches listed above, stable isotope probing (SIP) tracks isotopically labeled substrates into microbial populations that consume them, and is method for linking resource utilization to activity as a function of environmental conditions and overall community structure.

In SIP, DNA from isotopically labeled organisms is separated by density gradient centrifugation and identified with marker gene amplification and/or metagenomics. For example, SIP experiments have been combined with metagenome sequencing to infer horizontal gene transfer events responsible for conferring isoprene degradation among novel phyllosphere taxa (25), functional diversification between cellulose and lignin degradation in forest soils (26), microbes acting in consortia for the full degradation of polycyclic aromatic hydrocarbons in contaminated soils (27), as well as in seawater (28), and linkages of uncultured microbial populations to rhizosphere carbon cycling (29, 30). The quantitative stable isotope probing (qSIP) approach additionally estimates population-specific growth and death rates by tracking compositional information for 16S rRNA genes (31, 32). Here, we applied qSIP-informed genome resolved metagenomics, following an $H_2{}^{18}O$ addition experiment, to differentiate active from inactive soil microbes, define their metabolic capacities, and evaluate the potential roles of phages in bacterial cell lysis and carbon cycling among Mediterranean-climate grassland soils that occur across a natural rainfall gradient. Our goal was to characterize the total versus growing (active) bacterial/archaeal communities (and their genomic attributes) across these sites at a time of year when water was not limiting.

RESULTS

Site characteristics. To identify genomic traits associated with active microbes under varying historical precipitation patterns, we selected sites at three geographically dispersed Mediterranean California grasslands spanning two orders of magnitude in mean annual precipitation: Sedgwick Reserve, Hopland Research and Extension Center, and Angelo Coast Range Reserve (388, 956, 2,833 mm H_2O, respectively). The soils developed on similar parent material, primarily sedimentary rock including sandstone and shale, and vary only slightly in mineralogy and texture, with the driest site, Sedgwick, containing the highest proportion of clay and the highest effective cation exchange capacity (33). Sedgwick soils also reach lower water potentials at higher moisture contents than those from Hopland or Angelo (Fig. S2). All three sites had similar nitrogen and carbon content (Table S1; Foley et al. [33]).

Density fractionation effects on genome recovery. To evaluate the effect of density fractionation and isotope labeling on metagenome assembly, we assembled and binned individual fractions, sliding windows of co-assemblies of three adjacent fractions, and co-assemblies of full density gradients. For one sample (Hopland replicate soil core 1, both ^{18}O and ^{16}O incubations), we compared genome recovery outcomes using DNA sequences from all fractions versus three adjacent fractions on the density gradient and found that the all-fraction co-assemblies yielded the largest number of high-quality genome bins (Fig. S3A and B). Where genomes were recovered by both approaches, the highest quality bins (defined by maximum completeness and minimum contamination) tended to be those from the full co-assembly and most good bins that were recovered only once came from the all-fraction co-assemblies. We clustered genome bins from each assembly at 99% sequence identity, and found that of 65 genomes representative of the clusters, 44 with the highest quality (based on dRep scores) were from the all-fraction assembly. Twenty-eight clusters contained metagenome-assembled genomes (MAGs) recovered in multiple cross-fraction co-assemblies, in 15 of which, the highest quality genome was from the all-fraction assembly (Fig. S3C and D). Among the 37 clusters that we recovered in only one co-assembly, 29 were assembled in the all-fraction co-assembly (Fig. S3A and B). Read mapping indicated that genome bins had coverage across the density gradient, consistent with the additive effects of combining density-fraction libraries contributing to improved assemblies (Fig. S4). In addition, relative coverage across the density gradient varied systematically (with low GC regions having relatively higher coverage at lower densities). Based on
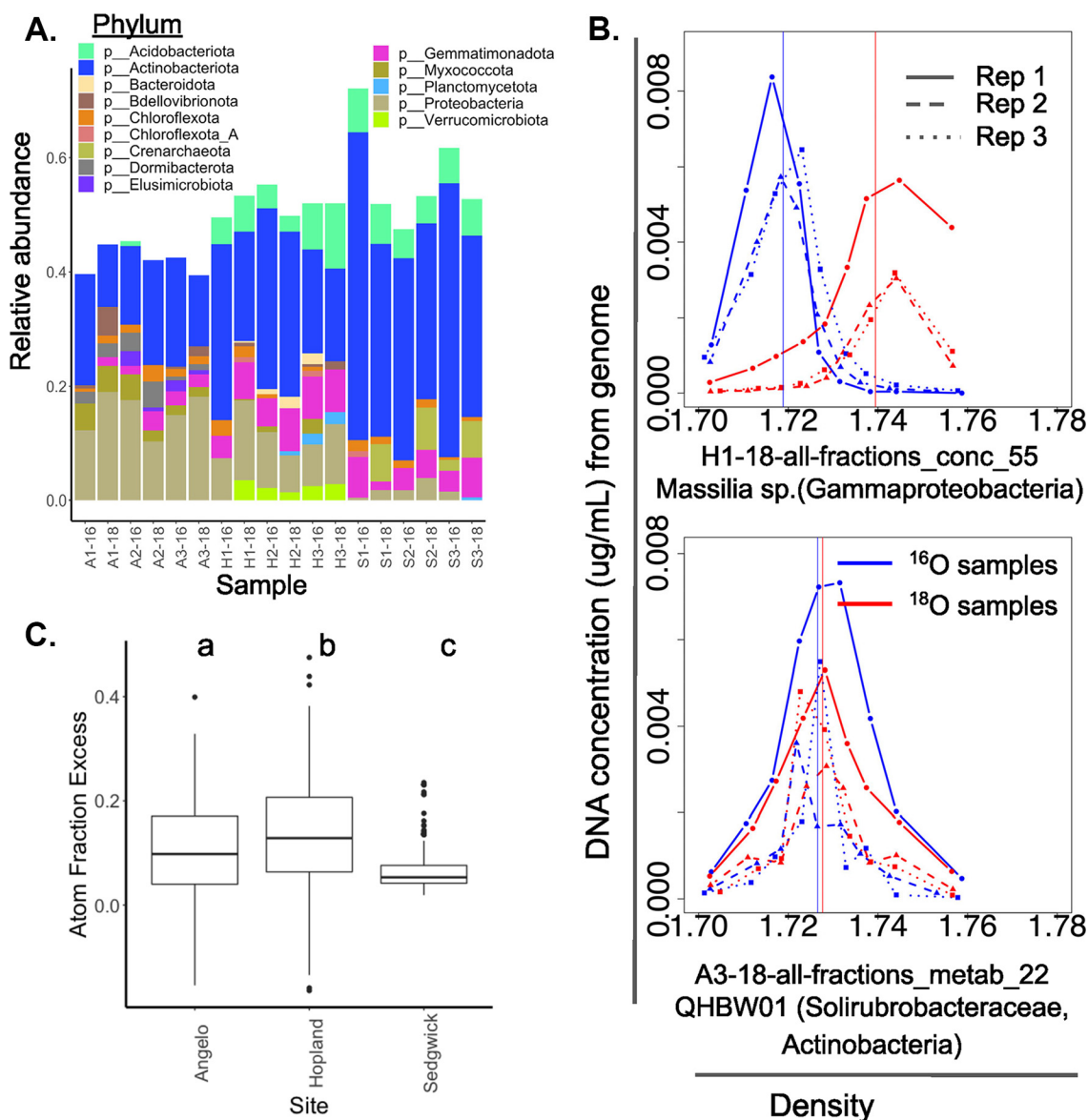
FIG 1 (A) Relative abundance (from 0 to 1, calculated as coverage in unfractionated-DNA library normalized to total sequence from that sample) and taxonomy of medium and high-quality genome bins assembled from density gradient metagenome co-assemblies following an $H_2^{18}O$ stable isotope probing incubation in three CA annual grassland soils (each bar represents one ^{16}O or ^{18}O incubation of one replicate soil core, A = Angelo, H = Hopland, S = Sedgwick, e.g., A1-16 = Angelo sample 1, ^{16}O-H2O incubation). (B) Illustrative plots representing the DNA density distribution for a single organism from one site that is either active (upper panel) or inactive (lower). Vertical lines indicated the weighted-mean density (WMD) of the pictured genome in ^{16}O or ^{18}O samples. (C) Boxplots showing the distribution of ^{18}O atom fraction excess (AFE) values for all genome bins assembled at each site. Boxes represent the range from the 25th to 75th percentile AFE for all bins; the bold horizontal line marks the median AFE value for bins from that site. Isotopic incorporation rates across genome bins at each site are significantly statistically different from each other based on Tukey's HSD test (indicated by the letters above boxes).

the superior performance of the all-fraction co-assemblies, we proceeded with these assemblies for each sample for genome binning, annotation, isotopic-label quantification, and statistical analyses.

Metagenome assembly and binning. From the all-fraction co-assemblies from all samples, we reconstructed 433 nonredundant genome bins with estimated completeness > 75% and contamination < 25%, representing a diverse array of common soil associated microbial taxa, with Actinobacteria predominating at all three sites (Fig. 1A). Diverse Proteobacteria are abundant at Angelo and Hopland reserves. Less-abundant organisms diverge between sites: we detect Gemmatimonadetes only at Angelo; Bacteroidetes and Chloroflexi predominantly

at Hopland; and Chrenarchaeota only at Sedgwick. *Bdellovibrio* was observed in Angelo and Hopland soils.

Quantifying isotopic incorporation in metagenomic sequences. To associate microbial ecophysiological traits with metabolic activity and population turnover, we calculated atom fraction excess (AFE) of ^{18}O in DNA sequences assembled from qSIP metagenomes (Fig. 1B). Isotopic enrichment represents new incorporation of oxygen into biomass, and AFE is therefore proportional to metabolic activity and population growth of the organisms from which the sequence assembled (34). Estimated AFE ranged from −0.16 (reflecting experimental error) to 0.47, with the range and distribution of AFE varying significantly by site (Fig. 1C; see additional supplemental tables). We refer to organisms with AFE lower than average for the site where it was assembled as "low activity" and those with higher than average AFE as "high activity." In the data set of ^{18}O-SIP 16S rRNA gene amplicon sequences that parallel our ^{18}O-SIP metagenomic libraries, we found a significant positive correlation between activity estimates from the two data types (Fig. 2A and B).

The index of replication (iRep) provides an orthogonal measure of *in situ* microbial growth calculated from metagenomic sequence data (19). However, only three bins exceeded the coverage filtering threshold required for iRep. Of these three bins, iRep was inversely related to AFE calculated for each (see additional supplemental tables), suggesting the difference in measuring instantaneous replication rates (as in iRep) versus gross population growth over time (qSIP).

AFE from qSIP expresses the observed shift in density as a proportion of the maximum theoretical density shift for an organism's genome. This shift is calculated by subtracting an organism's density in unlabeled samples (determined by its GC content), from its density in labeled samples. However, if the observed density of an organism's unlabeled genome matches the theoretical density calculated from the genome's GC values, we should be able to infer an organism's isotopic incorporation purely from the observed density of its isotopically labeled genome (Fig. S2B and C). With our data set, we calculated AFE for each genome bin using only read coverage data from ^{18}O-enriched libraries and found a very strong and significant positive correlation between AFE values calculated from ^{18}O and ^{16}O libraries ($P < 2.2 \times 10^{-16}$, $R^2 = 0.93$; Fig. S2C). On average, AFE values calculated solely with 18O samples are 0.08 lower than when calculated using ^{18}O samples, but the discrepancy approaches 0 for lower GC genomes. This is likely because at higher densities sampled in the density gradient, fractions representing wider density ranges were combined to yield enough DNA for sequencing, meaning the accuracy of mean-weighted density estimates decreases at higher density (i.e., for genomes with high GC and high ^{18}O incorporation) (Fig. S2D). Nevertheless, the tight correlation between AFE estimated using both ^{16}O and ^{18}O samples with AFE estimated with ^{18}O alone suggests it will be possible that unlabeled samples need not be sequenced in future SIP-metagenomics studies as long as sequencing depth enables substantial genome reconstruction.

We found ^{18}O AFE varied widely for genomes within the same phylum, both within and between sites (Fig. 2C), with community-level AFE distributions reflecting measured respiration rates between sites (Fig. S2C). Actinobacteria, the most abundant phylum observed at all three sites, had a particularly wide range in activity, including the highest and lowest AFE values at each site. Most Actinobacteria genomes had similar AFE distributions relative to the broader microbial community at each site, while members of the family Nocardioidaceae had higher activity levels than other Actinobacteria at all three sites.

Members of several less common groups of bacteria had more consistent activity levels compared to Actinobacteria. Chloroflexi and related phyla (Chloroflexota_A, also known as Rif_CHLX, and bacteria of the phylum Dormibacterota) at all three sites as well as Chrenarchaeota from Sedgwick had consistently very low activities. Members of the phylum Planctomycota had low activity levels (AFE < 0.1) at both Hopland and Sedgewick sites.
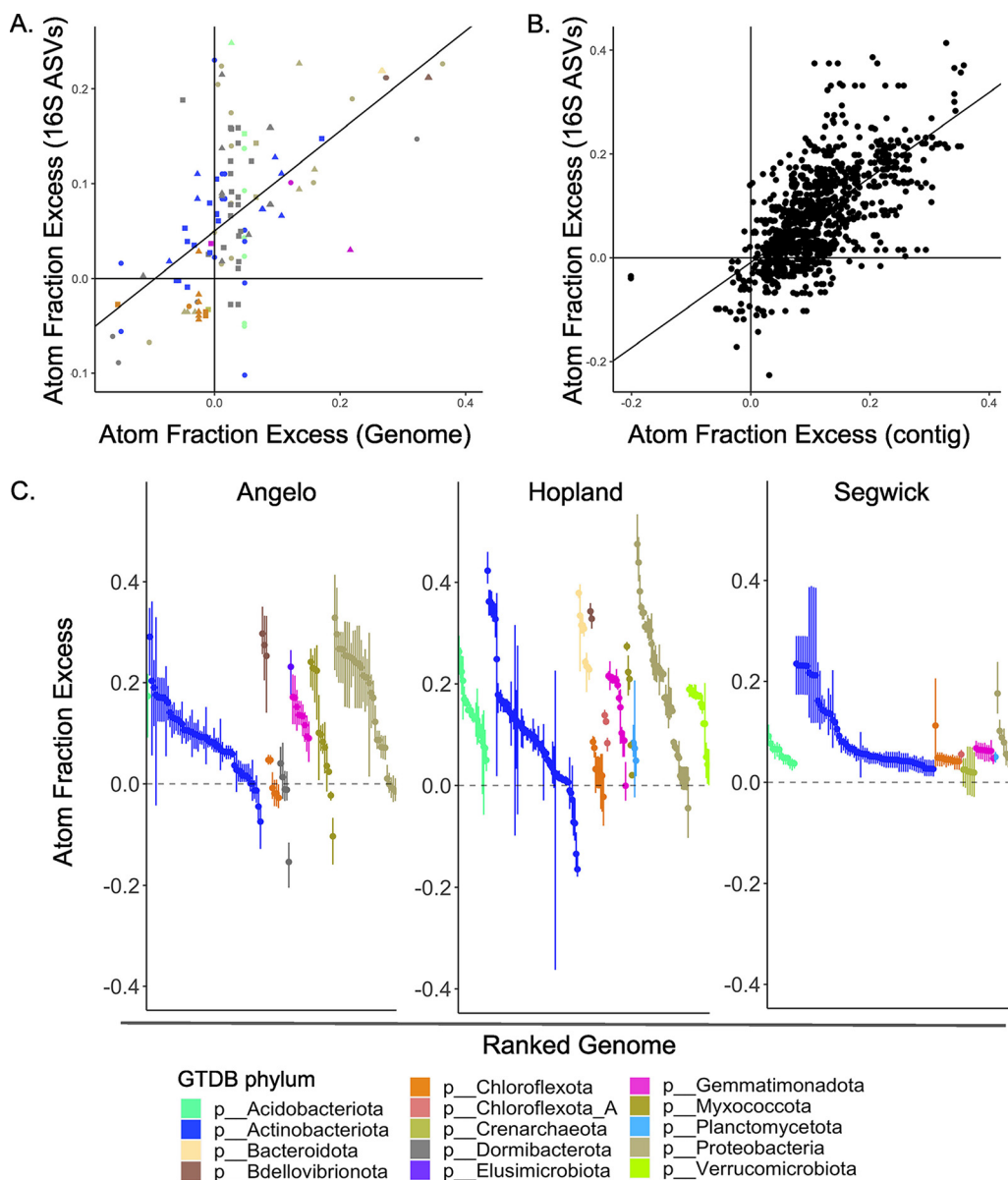
FIG 2 Microbial atom fraction excess (AFE) patterns following an H$_2$^{18}O stable isotope probing incubation in three CA annual grassland soils collected during the wet winter season. (A) Regression analysis of AFE measured in 16S rRNA ASVs (SIP-amplicon analysis from Foley et al. [33]) versus the subset of shotgun sequence metagenomic bins containing 16S sequences that match 16S-amplicon ASVs (X axis). (B) Regression of AFE estimated for 16S amplicon ASVs (Y axis) versus AFE calculated for metagenomic contigs containing 16S rRNA genes (X axis). (C) Mean and 95% confidence intervals of ^{18}O atom fraction excess (AFE) for all genome bins measured in the three CA grassland soils, colored and ranked by phylum.

Bdellovibrio (known as "predatory" bacteria for their obligate intracellular parasitism of other bacteria), represented by four genomes from Angelo and Hopland, were among the most active organisms at each site. Consistent with obligate parasitism, the genomes all contain loci for the type IV pilus known to be involved in host attachment, and exhibit many auxotrophies in amino-acid biosynthesis (35).

Statistical testing distinguishes metabolisms and ecophysiological traits by activity across sites. To assess drivers of microbial community differences between each site, we conducted constrained ordination of microbial communities from each sample in terms of beta diversity and activity (as expressed by pairwise Bray-Curtis distance between samples calculated from relative abundance of genomes as well as AFE, respectively). For both microbial diversity and activity, mean annual precipitation at each site was the only environmental variable statistically significantly ($P < 0.05$)
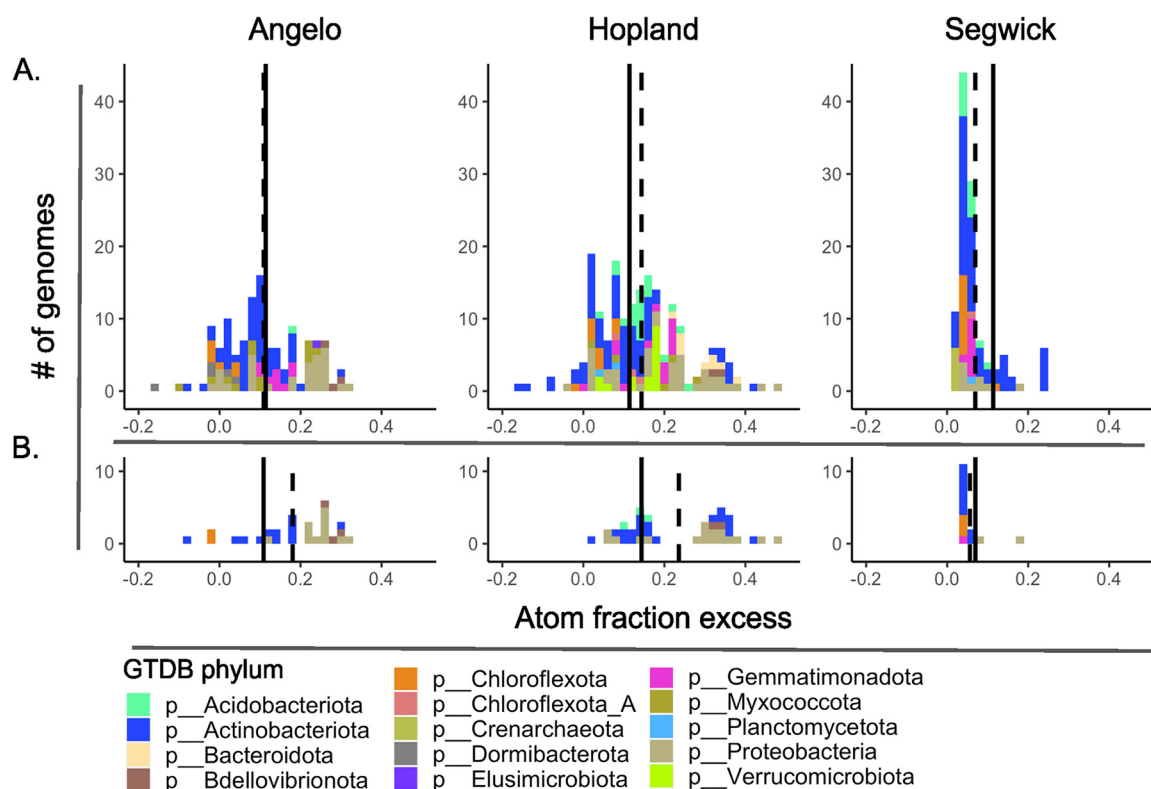
FIG 3 (A) Distributions of AFE values for all bins at each site. A solid line represents the average AFE across all bins at all sites; a dotted line represents average AFE across bins from the site shown in that figure panel. (B) Distribution of AFE values for genome bins encoding flagellin at each site. These genomes are significantly enriched at Angelo and Hopland relative to average bins at each site. A solid line represents the average AFE from all bins at that site; a dotted line is the average AFE of genomes encoding flagellin genes at that site.

explaining clustering of microbial populations (additional variables tested were soil pH, soil water potential at the permanent wilting point, soil moisture at sample collection, and soil moisture during the incubation).

To identify microbial traits associated with activity and how these patterns were framed by our sites' soil and environmental characteristics, we used a statistical test to identify traits where the AFE significantly increased at each site relative to the total microbial population (Fig. 3A; Fig. S5C). Motility emerged as the predictor of growth with the closest positive relationship to mean annual precipitation across sites. Putatively motile organisms (genome bins encoding full or mostly full suites of genes for flagellar biosynthesis) were statistically overrepresented among active organisms from the wetter Angelo and Hopland sites (organisms with flagella-encoding genes had AFE 8.1 and 9.2 higher, and 1.3% lower than all organisms assembled at Angelo, Hopland, and Segwick, respectively; Fig. 3B and 4A; Fig. S5D). This is partially due to the presence of flagellar genes in the highly active Bdellovibrionata at Angelo and Hopland. Similarly, active Nocardioideaceae (phylum: Actinobacteriota) at Angelo and Hopland have the capacity for flagellar motility whereas those from Sedgwick were not.

We hypothesized that organisms with limited mobility would tend to be more versatile in terms of the substrates that they can metabolize. We found that the nonmotile Nocardioideacae from Sedgwick encoded higher numbers of polysaccharide degradation pathways. These organisms also had higher numbers of enzymes for nitrate reduction and nitrite reduction, whereas Nocardioideacae from Angelo and Hopland only had the capacity for nitrite reduction.

Genes for carbon monoxide and methanol dehydrogenases were broadly distributed phylogenetically, geographically, and across activity levels measured by AFE (Fig. 4E). Carbon monoxide dehydrogenase (coxL) was primarily confined to Actinobacteria across
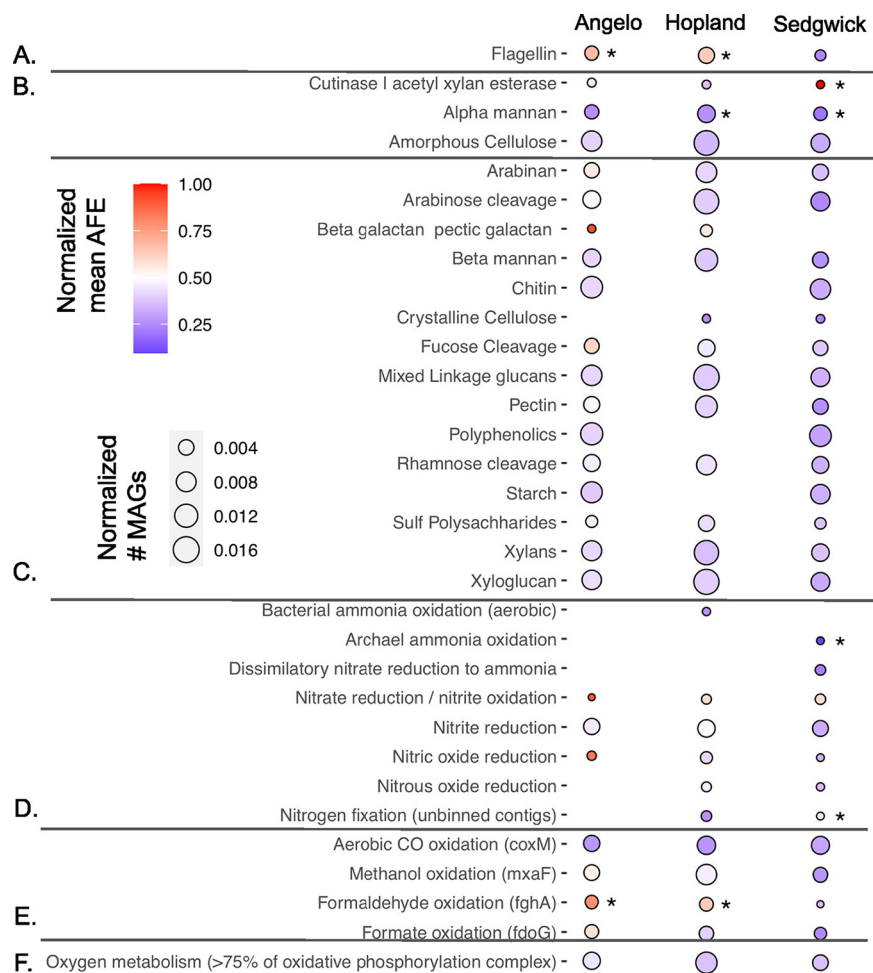
FIG 4 Heatmap of number of genomes with annotated functions and the activity (reflected by [18]O atom fraction excess-AFE) of those genomes, at three grassland sites in northern California, for genomes encoding: (A) flagellin (significantly active at high moisture sites); (B) cutinase (significantly active at low moisture site, Sedgwick); (C) polysaccharide metabolism; (D) nitrogen metabolism; (E) C1 metabolism; (F) oxygen metabolism. For all panels, the size of each point is proportional to the number of genomes with that trait. Circle color represents the average AFE of genomes with that trait, per site. An asterisk next to a point indicates significant difference from the average AFE of all bins at that site.

sites (Fig. S5A and E). Calcium-dependent methanol dehydrogenase (mxaF) was broadly distributed among Acidobacteria, Gemmatimonadetes, and Proteobacteria (Fig. S5B and F). Genomes annotated with mxaF were often highly active at Angelo (0.151 for mxaF bins, average AFE of 0.107 for all Angelo bins), and to a lesser extent at Hopland (0.174 compared with 0.135) but bacteria with mxaF from Segwick were relatively inactive (0.063 versus 0.110).

We found significant positive correlation between annotated metacyc degradation pathways and AFE-based activity levels at Angelo, but not the other soils (Fig. S6A). Across sites, we aggregated genomes based on phylum affiliation and found no significant correlation between substrate diversity and activity levels. We also examined the distribution of degradation pathways for several complex carbohydrates and found that, of genomes possessing any of 13 polysaccharide degradation pathways curated in the DRAM genome annotation pipeline (36), none differed significantly in inferred activity levels from the total community represented by the assembled metagenome at any site (Fig. 4C). We observed significant correlations between the diversity of polysaccharide degradation pathways encoded in a genome and the high activity levels for that genome in some phyla, but the phyla that displayed this pattern often varied across sites. For example, we found a significant positive correlation between the number of polysaccharide-degradation pathways

encoded in Proteobacteria genomes and organism activity (AFE) across all sites ($r = 0.47$, $P < 0.05$) whereas for Gemmatimonadetes, there was only a significant correlation between polysaccharide-degradation pathways and activity at Angelo ($r = 0.33$, $P < 0.05$) (Fig. S6B). Conversely Chrenarchaeota at Sedgwick and Acidobacteria and Bacteroidetes from Hopland exhibited higher activity correlated with a lower diversity of polysaccharide degradation pathways.

MAGs from the three sites varied in the presence of specific nitrogen cycle enzymes, in the abundance of nitrogen-related genes, and in the inferred activity levels of bacteria involved in nitrogen compound transformations (Fig. 4D). All three sites had organisms whose genomes encoded the first three steps of denitrification (reduction of nitrate to nitrite; nitrite to nitric oxide; nitric oxide to nitrous oxide). The capacity for nitrate reduction was only observed in genomes of highly active bacteria from the wettest site, Angelo, whereas genes for nitrite reduction were broadly distributed among organisms with varied activity levels at all three sites. Nitric oxide reduction capacity was also encoded in genomes of the most highly active bacteria from the Angelo site. The capacity for nitrous oxide reduction to molecular nitrogen was only observed in genome bins at Hopland and Sedgwick (organisms potentially capable of N_2O-reduction were close to average activity at both sites). The capacity for aerobic ammonia oxidation was predicted for Acidobacteria from Hopland, and Crenarchaeota from Sedgwick, but the inferred activity levels of these organisms were low. In addition, genes for nitrogenase were encoded on unbinned contigs from the Hopland and Sedwick data sets. A few nitrogenase operons were encoded on high AFE contigs from Sedgwick.

We annotated genome bins for EPS biosynthetic gene clusters for a class of polysaccharides produced through the synthase-dependent biosynthetic system, including poly-N-acetylglucosamine (PNAG), cellulose and acetylated cellulose, alginate, and Pel. PNAG was the most commonly observed polysaccharide biosynthesis cluster, assembling in a broad range of taxa with varied activity levels at all sites (Fig. S7). Notably, Actinobacteria of the family Nocardioideaceae—that were apparently highly active at each site—are not predicted to synthesize PNAG.

Quantifying activity in phage. Many studies have documented viruses in soil (e.g., 21, 37–39), but given that viruses can be deactivated in soil through multiple mechanisms (e.g., by sorption to minerals), a grand challenge is to assess what fraction of viruses is active. We identified potential phage sequences in metagenomic assemblies (total of 119,253 viral contigs across all all-fraction assemblies; clustered into 8,617 viral populations), and examined patterns of activity measured by AFE, focusing on phage contigs for which we confidently predicted hosts. As with microbial genomes, putative phage contigs demonstrated a range of AFE at each site. We find a significant positive relationship between the AFE of a host genome and its putative phage genome (P value $< 2.45e-14$, $R^2 = 0.683$; Fig. 5). For example, we identified a circular 38-kbp genome for a phage predicted to infect a *Bdellovibrio* at Hopland. Circularization indicates that the sequences were derived from phage particles, not from prophage. Both the putative *Bdellovibrio* phage and the only *Bdellovibrio* genome assembled from Hopland are predicted to have been highly active (AFE ~0.3 to 0.35 range). However, notably, for this phage-host pair and in general, hosts predicted to be highly active had lower relative abundances. High activity of hosts represents high growth rates, and we might expect that this would lead to high abundances. However, it is possible that the most active phages had infected and lysed their hosts (leading to low abundance) before samples were collected. Alternatively, the phages may simply have infected low abundance yet active hosts.

From all three sites, we reconstructed and annotated seven genomes of phages that encode alginate lyase carbohydrate-active enzymes (CAZYmes). Alginates are polysaccharides that may mask receptor sites used by phages during infections, and the phage lyase may circumvent this defense. When aligned to the refseq database, genes within these putative phage contigs that align to annotated genes from known organisms have highest homology to sequences from Pseudonocardiaceae, consistent with these phage infecting
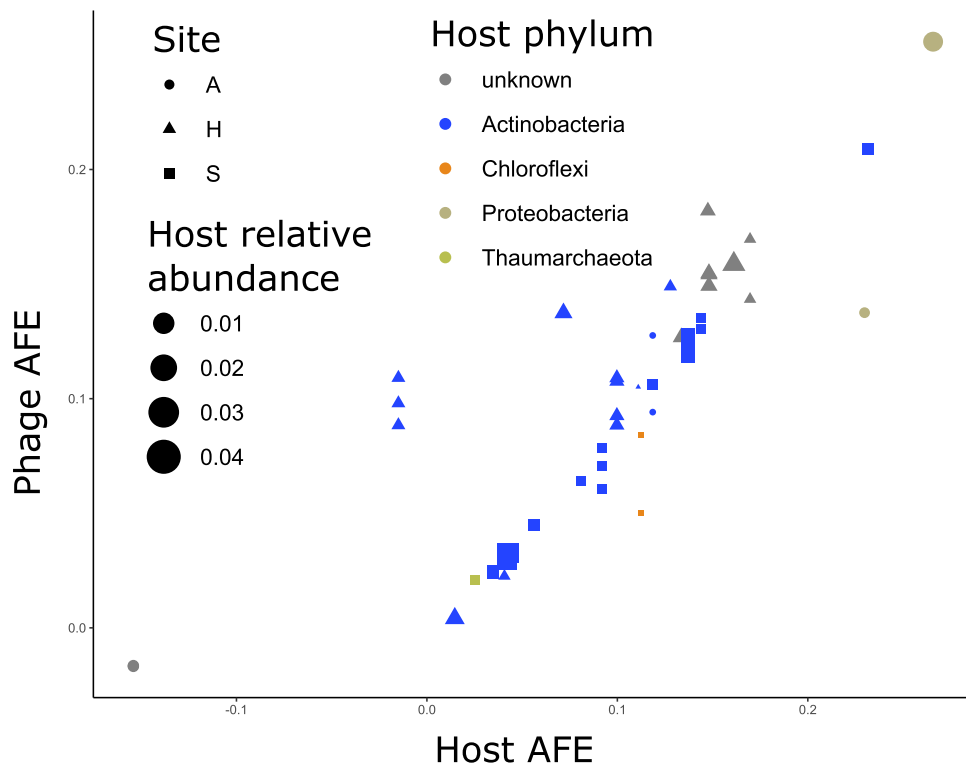
FIG 5 Taxon-specific activity (measured by ^{18}O AFE) of grassland soil bacterial genomes versus phage predicted to infect each bacterial host based on matching CRISPR spacer sequences. Sites include Angelo (A), Hopland (H), and Sedgwick (S), which exist along a rainfall gradient in northern California. Host relative abundance is indicated by the size of datapoints.

Pseudonocardiaceae cells. Interestingly, only five Pseudonocardiaceae from Sedgwick have operons implicated in alginate synthesis whereas Pseudonocardiaceae genomes from Hopland are not predicted to produce alginate (Fig. S8). Putative Pseudonocardiaceae-infecting phages from Segwick that encode alginate lyase occur on metagenomic contigs that are not circularized. These contigs have higher AFE than the predicted alginate-producing Pseudonocardiaceae host genomes assembled from Segwick, consistent with their existence as phage particles that replicated in an actively growing host (Fig. S9).

DISCUSSION

Using qSIP-informed genome-resolved metagenomics, we quantified the activities of bacteria, archaea, and phages from the three annual grasslands that exist along a strong precipitation gradient. Previous SIP metagenome studies have typically sequenced only the heavy fraction of DNA density gradients for labeled and unlabeled samples. At best, this approach can make a binary distinction between organisms that have or have not incorporated the stable isotope label, and many organisms that incorporate the isotope are missed due to lower yet significant isotope incorporation. Additionally, regions of the same genome with differences in GC content distribute across the density gradient, reducing assembly quality if only subsets of the density gradient are assembled. Finally, as the mean weighted density is determined by both GC content and isotope incorporation, many low GC genomes with high AFE may be missed if only the heavy fraction was sequenced. By sequencing across the density gradient, the approach used here, we recovered more and higher quality genomes and quantified isotopic enrichment in microbial and phage genomes. Estimates of AFE for metagenomic contigs and genomes containing 16S rRNA genes align with AFE estimates for identical 16S rRNA gene sequences calculated from the paired data set from the same samples. This approach allowed us to deduce quantitative relationships, to test for a correlation between abundance and growth, and to

use a statistical framework to associate microbial traits with activity as a function of historical water inputs.

The strongest statistical signal we saw relating a microbial trait to differences across the annual precipitation gradient represented by our sample sites was flagellar motility. Genomes predicted to belong to motile organisms were proportionally most active at the wettest site (Angelo), followed by the more moderate precipitation site (Hopland), whereas the seasonally driest site (Sedgwick) had motile bacteria with the lowest activity under our experimental conditions. Previous modeling has predicted that spatial heterogeneity of soil selects organisms capable of rapidly responding to changes in availability of a wide diversity of substrates (and against alternative nutrient acquisition strategies such as chemotactic motility) (40), whereas we found selective pressure favored increased motility in soils with a historical pattern of higher soil moisture availability.

The higher soil moisture levels at Angelo and Hopland could make motility advantageous (except for those bacteria that rely on soluble organic compounds, as diffusion rates would increase with increased soil moisture). Interestingly, the motile bacteria from family Nocardioidaceae at Angelo and Hopland have relatively few enzymes for degradation of insoluble carbohydrate compounds such as arabinan, and xylan, as well as polysaccharides containing fucose or rhamnose, compared with nonmotile bacteria from this family at Sedgwick. Thus, the Actinobacteria at higher-moisture sites appear to be relatively specialized (from the perspective of carbohydrate degradation) compared with those from Sedgwick. These findings support the premise of some ecological models (40) that suggest homogeneity associated with wetter soils should select for organisms with relatively specialized metabolisms, whereas heterogeneous environments such as drier soils should select for versatile heterotrophs. In wetter soils, microbes that are motile can relocate to sites where resources they can use are found, whereas in dry soils, motility is restricted and, therefore, it is beneficial to be a versatile heterotroph. During the soil incubations, water potential varied little between sites. The observed patterns in microbial activity therefore reflect the effects of historical differences between sites in soil moisture. Long-term climate shifts toward decreased precipitation might select for less motile organisms and their associated lifestyles.

The higher activity of motile organisms in wetter soils was driven in part by the activity of motile *Bdellovibrio* (representing 20% of flagella-encoding genomes at Angelo and 9.7% at Hopland). *Bdellovibrio* are often intracellular parasites of other bacteria. If the *Bdellovibrio* in the soils studied here are also parasites, then it may be reasonable to hypothesize that motility enabled by higher moisture levels enhanced the success of parasitic bacteria. There is increasing evidence for the predominance of microbial necromass in SOC (7); therefore, we speculate that *Bdellovibrio*-induced lysis could contribute substantially to SOC pools in wetter soils (41). The lack of difference in SOC abundances in the three soils may be due to increased access to the lysate as a C source in wetter soils. Climate-change-driven reduction in moisture levels may select against *Bdellovibrio*, lowering the contribution of bacterial predation to cell death.

Our SIP-metagenome data show that despite the precipitation gradient among our sites, metabolic capacities such as aerobic respiration, CO oxidation, and methanol oxidation were linked to genomes whose isotopic enrichment values fell close to average community enrichment values at all three sites. Thus, we conclude that these are widespread capacities in grasslands during seasonally high soil moisture.

Metagenomic qSIP enabled us to track activity of phage populations. The observation that AFE of phage genomes is closely linked to the AFE of their predicted microbial hosts might suggest that phage lysis rates are sometimes predicted by microbial growth rates (42). However, a subset of these may have been prophage and were isotopically labeled during host replication. The observed low relative abundances of host bacteria in soil may be a consequence of phage replication leading to host bacterial death, but other explanations (including bacterial and eukaryotic predation) likely also contribute. It is interesting to note that phage predation may have reduced the abundances of bacteria that were

recently highly active to undetectable levels, precluding deduction of their recent high activity via qSIP-informed genome-resolved metagenomics. The detection of phages with high ^{18}O values but no apparent host may be an indication of prior host growth and a useful signature of high rates of phage-induced mortality, as well as an indicator of top-down trophic influence in soil microbiomes (41).

The fast replication of organisms encoding enzymes involved in nitrate, nitrite, and nitric oxide reduction suggests denitrification to N_2O likely occurs in Angelo soils during the wet season. The genomic capacity of active organisms for nitrous oxide reduction, leading to full denitrification to dinitrogen gas, was more prevalent at Hopland and Sedgwick compared with Angelo. Given that all three sites have similar nitrogen compound concentrations, we suggest that activity differences might have occurred because the sites were not equally wet at the time we sampled (Table S1). Thus, long-term differences in yearly moisture availability likely had a strong impact on microbial transformations in soil nitrogen cycles. The high activity of genomes encoding denitrification steps at Angelo (historically the wettest site) is consistent with previous results showing that high soil moisture leads to increased respiration rates and increased anoxia, selecting for organisms that can rapidly use nitrate as an alternative electron acceptor (43). Furthermore, denitrifiers at Angelo and Hopland also encoded flagella, suggesting motile denitrifiers have an advantage, perhaps because oxygen is depleted more rapidly. Although the bacteria capable of reduction of nitrous oxide were not particularly active at the time of sampling, these bacteria could limit emissions of N_2O resulting from partial denitrification from the drier soils under other conditions. More importantly perhaps, the ability to reduce N_2O to N_2 is most prevalent in bacteria from both Sedgwick and Hopland that are nonmotile. Thus, future decreases in soil moisture levels could lead to decreased emissions of this greenhouse gas. Conversely, the lack of genomes encoding genes for reduction of N_2O at the Angelo site may predict a higher capacity for release of this greenhouse gas from Angelo compared with other soils, especially under wetter conditions. This is consistent with previous observations of higher N_2O fluxes from wetter soils (44).

Methodological advances and implications. Past metagenomic SIP studies have claimed that either the density fractionation or stable-isotope labeling itself improves metagenomic assembly by reducing the complexity of the microbial community sampled within each individual fraction (26, 45). In contrast, in our study we observed that co-assembling all nine sequenced fractions from the same sample almost always improved assembly quality, probably because it increased the coverage per genome. This suggests a possible SIP metagenomics experimental design wherein assemblies primarily come from unfractionated DNA sequenced to high depth where density fractions are sequenced at low coverage for identifying bins through differential abundance and for estimating isotopic incorporation.

In the future, this genome-resolved approach could be greatly expanded by using substrates labeled with ^{13}C (e.g., 46) or ^{15}N labeled (e.g., 47, 48), as demonstrated by 16S rRNA gene amplicon qSIP studies. Use of ^{13}C and ^{15}N labeled substrates could be used as in-depth analysis of soil carbon biogeochemistry, because compounds predicted from genomes to be consumed or synthesized by isotopically enriched bacteria can be quantified with spectroscopic techniques that can differentiate isotopically labeled molecules (e.g., HPLC, NMR). Furthermore, sequencing unfractionated metagenomic DNA from not one (as was done here) but many time series points would constrain the relative abundances of specific bacteria, potentially enabling direct estimates of microbial net growth (i.e., birth and death) rates (31, 32). This would complement measurements of isotope incorporation into phage populations that are predicted to infect specific microbial populations, allowing linkage of cell death to phage predation.

Genome-resolved SIP can be cost- and labor-intensive. Given that the mean-weighted densities for organisms in natural-abundance isotope-treated samples are close to those estimated purely from GC of the same sequence assembled in heavy-isotope labeled samples, it may be possible to conduct future SIP experiments with few or no natural-

abundance isotope controls. This would double, for the same cost, the number of samples that could be included in metagenomic qSIP experiments, enabling deeper replication coupled with increased statistical power (49) and/or more diverse treatments.

Conclusions. The qSIP-enabled implementation of genome-resolved metagenomics that involved sequencing multiple density fractions enabled us to probe the activities and functional potential of bacteria and phages in soil microbial communities. We differentiate phenomena associated with soils that have experienced a gradient of annual precipitation levels and identify traits that may change in prevalence as climate changes, including a subset that could impact soil greenhouse gas emissions.

MATERIALS AND METHODS

Sample collection, isotopic enrichment, and fractionation. Triplicate 0- to 10-cm soil cores were collected from northern California annual grassland sites at Sedgwick Reserve, Hopland Research and Extension Center, and Angelo Coast Range Reserve between February 2018 and March 2018 (the period when water is most available at each site annually). Soil at these sites developed on similar parent material, are overlaid by annual grasses, including *Avena* spp., and experience a rainfall gradient of 388 mm yr^{-1} to 2,833 mm yr^{-1}. Each soil core was homogenized and separated into 5-g subsamples that were dried at room temperature over 24 h to 1.5% gravimetric soil moisture, re-wetted to 25% to 30% moisture with either natural abundance ^{16}O-H_2O or 98.15 atom% ^{18}O-H_2O, and then incubated for 8 days at room temperature in the dark in 500-mL glass Mason jars. Five μg of DNA extracted from each sample was subjected to ultracentrifugation in a cesium chloride density gradient (final average density 1.730 g/mL) in 5.2-mL tubes, then separated into 36 density ($\sim$200 μL each) using a semiautomated robotic SIP protocol (50). The fractions for each sample were binned into nine groups based on density (1.6900 to 1.7099 g/mL, 1.7100 to 1.7149 g/mL, 1.7150 to 1.7199 g/mL, 1.7200 to 1.7249 g/mL, 1.7250 to 1.7299 g/mL, 1.7300 to 1.7349 g/mL, 1.7350 to 1.7399 g/mL, 1.7400 to 1.7468 g/mL, 1.7469 to 1.7720 g/mL), and fractions within a binned group were combined and sequenced. Soil water retention curves were generated for Sedgwick, Angelo, and Hopland field sites using a tensiometer (HYPROP) and dew point potentiometer (WP4C) as previously described; full details of sample collection and processing are provided in Foley et al. (33).

DNA sequencing, assembly, annotation, and binning. DNA sequencing libraries were generated using the Kapa HyperPrep kit (Roche) from each density fraction, as well as unfractionated DNA from each incubated soil sample. Paired-end, 150-bp reads were generated with two lanes of the NovaSeq platform (Illumina), to an average depth of 7 gbp per library. Illumina adapter and Phix sequences were removed with BBtools (https://jgi.doe.gov/data-and-tools/bbtools/), and low-quality sequences were trimmed or discarded with sickle (51). Quality-filtered reads were assembled with Megahit (version v1.2.9) with parameters "–k-min 21 –k-step 6 –k-max 255" (52) for each individual density-fraction library as well as unfractionated-DNA libraries; co-assemblies of all sliding windows of every three adjacent density fractions (1 + 2 + 3, 2 + 3 + 4, 3 + 4 + 5, etc.) for each incubated soil sample; and co-assemblies of all density fractions from each incubated sample (replicate samples, i.e., cores, from the same site, were assembled and binned separately). Assemblies were filtered to remove contigs shorter than 1 kb.

Contigs from each assembly were annotated using multiple sources. Open reading frames (ORFs) were predicted from assembled contigs using Prodigal v2.6.3 (53) with the parameters "-m -p meta." We used USEARCH to identify sequences homologous to predicted ORFs in the Uniprot, Uniref90, and KEGG (54) databases. We predicted 16S rRNA gene sequences using the 16SfromHMM.py script, and tRNA genes using tRNAscan-SE (55). All-fraction co-assemblies were also annotated using the METABOLIC pipeline (version 1.0) (56).

We separated metagenomic contigs greater than 2.5 kb into bins representing genomes from distinct microbial populations based on sequence signatures and differential abundance across samples. Quality-filtered reads from all libraries of each sample from the same site were mapped to each all-fraction co-assembly from that site using bbmap (https://jgi.doe.gov/data-and-tools/software-tools/bbtools/bb-tools-user-guide/bbmap-guide/) with the parameters "fast=t ambig=random minid = 0.98." The abundance of contigs across samples was calculated using the jgi_summarize_bam_contig_depths script from the Metabat2 software package (57). Contigs from each co-assembly were sorted into genome bins using Metabat2, Maxbin2 (58), and Concoct2 (59). Genome bins generated by each binning algorithm were aggregated using the Bin_refinement module of metawrap (60). Aggregated bins from each all-fraction co-assemblies were dereplicated into representative nonredundant genomes using dRep (61) using a 99% sequence identity threshold. Dereplicated high-quality bins were manually inspected for phylogenetic coherence using the ggkbase tool (http://ggkbase.berkeley.edu). The same read-mapping procedure was performed for individual-fraction assemblies, and sliding window 3-fraction co-assemblies for $H_2^{18}O$ and $H_2^{16}O$ incubations from soil core 2 from Hopland reserve. Metabat2 was used to generate genome bins from individual-fraction, 3-fraction, and all-fraction assemblies for these two samples, and these bins were dereplicated with dRep at 99% average nucleotide identity to determine which assembly strategy would yield the most high-quality genome bins.

We re-annotated the aggregated, dereplicated genome bins by predicting ORFs using prodigal with parameters "-p single." Predicted ORFs were again annotated using USEARCH (62) against Uniprot, Uniref, and KEGG, as well as METABOLIC (version 1.0) (62) and DRAM (version 1.0) (36). Dereplicated genome bins were also assessed for the capacity for extracellular polysaccharide production through the synthase-dependent pathway by first identifying all putative secondary-metabolic biosynthetic gene

clusters using the antiSMASH (v 5.0) (63) pipeline with strictness set to "loose." Putative saccharide biosynthesis clusters were further classified as synthase-dependent and by class of polysaccharide using the criteria developed by Bundalovic-Torma et al. (64). In brief, genes from biosynthetic gene clusters that antiSMASH identified as saccharide were searched against HMM profiles for gene families from known synthase-dependent polysaccharide biosynthetic gene clusters. Clusters were classified with known polysaccharide synthesis operons if the cluster had HMM hits with e-value $< 10^{-5}$ for three or more genes from that known operon, including for Polysaccharide synthase. We then clustered all predicted genes from all extracellular polysaccharide (EPS) biosynthesis pathways identified above into putative gene families using the protein clustering pipeline described in (65).

We replicated an analysis from Nunan et al. (40) correlating rRNA copy number with the number of metabolic pathways annotated "degradation" among genomes in the metacyc database. For this purpose, we manually parsed the metacyc database pathways information to identify all pathways containing the term "Degradation," analogous to Nunan et al. We identified metacyc pathways in genomes assembled from this study using MinPath (version 1.4) (66). Whereas Nunan et al. use 16S rRNA gene copy number as a proxy for maximum growth rate to compare with metabolic diversity, we used atom fraction excess (AFE) value—calculated from the observed density shift of metagenome bins—as a direct measure of population growth in our replication of Nunan et al.'s analysis (40).

Virus identification and host assignment. Phage-related contigs were identified with VirSorter (67) and deepVirFinder (68). VirSorter was run in "virome decontamination" mode, and only virus contigs identified as category 1, 2, 4, and 5 were kept. These viral contigs were compared with those identified by deepVirFinder, in which contigs were considered viral if they obtained a score $\geq$ 0.9 and a P value $\leq$ 0.05. Across all fraction assemblies, a total of 119,253 viral contigs were identified. To gain approximate "species-level" taxonomic resolution (i.e., a viral population), the contigs were dereplicated through clustering at 95% average nucleotide identity (ANI) and 80% coverage (69).

Viral populations were linked to their putative microbial hosts using a scoring approach (VirMatcher; 70) that was previously applied to study human gut viruses (71). The putative microbial hosts used to establish these linkages were retrieved from 443 high-confidence MAGs assembled in this study, and used as the host database. The different bioinformatic methods used in our scoring approach include viral matches to (i) host CRISPR-spacers, (ii) integrated prophages in MAG contigs, (iii) host tRNA sequences, and (iv) host k-mer composition (72), with the details of each method and associated scores described in Gregory et al. (71). The host assignments shown here only include the high- and intermediate-confidence predictions with a final score of $\geq$1.5.

Calculating isotopic enrichment of microbial populations. Reads from each density-fraction metagenome library were mapped to each all-fraction co-assembly for all samples from the same site. For each contig, mean coverage was multiplied by contig length and divided by the total sequencing yield of that library in order to calculate relative abundance of each assembled sequence in each density fraction (for bins, this measure of relative abundance was summed across all contigs in the bin). Relative abundances were multiplied by ng/μL of DNA recovered from ultracentrifugation for that density fraction to estimate the proportion of DNA in each density fraction belonging to each assembled sequence. These ng/μL DNA concentrations across each incubation's density gradient for a given sequence were used to calculate mean weighted density values for that sequence in each incubated sample. Isotopic enrichment for each bin was calculated following the methods described in Hungate et al. (73) with the following modifications. The observed GC content of each sequence was used to estimate the oxygen content of the DNA comprising the sequence, and from there the maximum mean weighted density of the sequence if all oxygens were substituted with ^{18}O. Additionally, we used the concentration of DNA in each recovered density fragment to normalize the abundance (measured by read mapping) of each bin or metagenomic contig (23, 49). The difference between the average mean weighted density for a given assembled sequence in all three natural-abundance ^{16}O replicates from that site was subtracted from the average mean weighted density in all ^{18}O replicates. The proportion of this observed density shift relative to the estimated maximum density shift represents the sequence's atom fraction excess ^{18}O (AFE) (Fig. 1C). In reality, no organism could reach the maximum possible density shift estimated because ^{18}O atom fraction excess was not 100% in the ^{18}O-H$_2$O incubated samples (Table S1). This can be accounted for by calculating the fraction of maximum potential enrichment as in Foley et al. (23). To calculate confidence intervals for AFE estimates for each genome bin, we implemented a bootstrapping procedure as follows: for each genome bin, weighted-mean density (WMD) shifts were calculated for each pair of ^{16}O/^{18}O incubated samples from each replicate soil core from the site where the bin was assembled; three WMD shifts were randomly subsampled with replacement from the WMD shifts calculated for each replicate soil cores. This subsampling procedure was repeated 1,000 times for each genome and average AFE was calculated for the three subsampled WMD shifts for each bootstrap; 95% confidence intervals for AFE for each genome were calculated from the distribution of AFE estimates across the 1,000 bootstraps.

In marker-gene amplicon-based qSIP, GC is estimated based on the mean weighted density of an amplicon sequence variant (ASV) in natural-abundance isotope samples (32, 73). To calculate AFE on metagenomic contigs and genome bins, we used the GC content of the assembled sequence, rather than estimating GC from the mean weighted density of the sequence in ^{16}O samples (we found that both methods yielded near equivalent values). In order to test the feasibility of SIP metagenomics experiments without natural-abundance samples, we further calculated AFE on each metagenomic bin using the genome's observed GC content and observed mean weighted density in ^{18}O samples (using GC to impute both maximum mean-weighted density and natural-abundance mean-weighted density) and compared these ^{18}O-only AFE values with those we had calculated using ^{16}O samples.

Hungate et al. (73) use the calculation $GC = \frac{1}{0.083506} \cdot (W_{Lighti} - 1.640657)$ where GC is the average GC (as a proportion of total DNA) of genome$_i$ and W$_{Lighti}$ is the mean weighted density of the 16S rRNA sequence from genome$_i$ in natural-abundance samples (73, see Equation 5). This is based on an empirically derived relationship between mean weighted density and GC for DNA from microbial cultures with genomes of known GC proportions. Based on the GC content of genome bins assembled from our samples and calculated mean weighted densities in ^{16}O-H$_2$O incubated samples (Fig. S1A), we find a relationship of $GC = \frac{1}{0.088} \cdot (W_{Lighti} - 1.6689)$. We used this formula to estimate the mean weighted density of genomes without isotopic labeling in order to calculate AFE using densities only from ^{18}O-H$_2$O incubated samples.

To evaluate the contribution of soil and environmental characteristics on microbial diversity at each site, we calculated Bray-Curtis distance between each sample using relative abundance of bins and then partial distance-based redundancy analysis on pairwise distances using stepwise model selection evaluating the contributions to microbial community structure of mean annual precipitation, soil pH, soil water potential at the permanent wilting point, soil moisture at sample collection, and soil moisture at during the incubation each as covariates. This analysis was conducted using the package "vegan" in the R computing language with the functions "vegdist," "capscale," and "ordistep" (74, 75). The analysis was repeated using AFE instead of relative abundance for each genome to calculate Bray-Curtis distance between samples.

For each annotation feature (key enzymes and pathways for core metabolism, stress, complex carbohydrate degradation, physiological traits, including motility, viral defense, and polysaccharide biosynthesis inferred from multiple annotation packages and custom pipelines), we conducted a two-way Kolmogorov-Smirnov test comparing the AFE distribution of genomes with that trait at each site versus the AFE distribution of all genomes at the same site, correcting for multiple comparisons with the Benjamani-Hochberg procedure. We hierarchically clustered annotation features by mean AFE for genomes possessing that feature at each site (see additional supplemental tables). The findings for the three distinct sites were then compared.

We also calculated the index of replication (iRep, 19) for each dereplicated metagenomic bin by mapping reads from unfractionated DNA from each of the 18 samples to genomes co-assembled from density-fraction libraries from the same sample.

Comparison to 16S-rRNA marker gene qSIP. Density-fractionated DNA was also used to generate 16S rRNA V4-5 variable region amplicons as described by Foley et al. (33). Libraries were sequenced on an Illumina MiSeq instrument at Northern Arizona University's Genetics Core Facility. Paired-end reads were filtered to remove phiX and other contaminants with bbduk v38.56 (https://jgi.doe.gov/data-and-tools/software-tools/bbtools/bb-tools-user-guide/bbduk-guide/) and Fastq files were trimmed for quality and used to generate ASVs with DADA2 v1.10 and phyloseq v1.26 (76, 77). Chimeric sequences were removed using removeBimeraDenovo from DADA2. ^{18}O AFE of bacterial and archaeal 16S rRNA gene amplicons was quantified following a modified version of the procedure (Tag-SIP) described in Hungate et al. (73). Here, average DNA concentration was used rather than 16S-rRNA copy number (23, 49) to normalize the relative abundance of taxa within each density fraction. A WMD was then calculated for each taxon based on the distribution of its DNA across the CsCl density gradient following incubation with either natural abundance or isotopically heavy water.

Representative ASV sequences from amplicon libraries were compared with 16S rRNA gene sequences identified from metagenomic contigs using blast, filtering for any hits that aligned at 99% across the full 250-bp assembled amplicons. We calculated AFE of individual 16S rRNA gene-containing contigs using the procedure described above, and compared AFE values for 16S rRNA gene sequences calculated from amplicon sequencing and metagenomic assembly using published qSIP calculations.

Data availability. Sequence data generated for this manuscript is available at NCBI under Bioproject PRJNA718849. Microbial genomes bins and annotations are available at: https://ggkbase.berkeley.edu/wsip-metawrap-drep-bins/organisms. Scripts for the computation and analysis used to generate results for this manuscript are available at https://github.com/alexgreenlon/wsip/tree/master. Additional supplemental tables are available to https://github.com/alexgreenlon/wsip/blob/master/Greenlon-et-al.2022%20additional%20supplemental%20data.xlsx.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, PDF file, 0.7 MB.
FIG S2, PDF file, 0.4 MB.
FIG S3, PDF file, 0.5 MB.
FIG S4, PDF file, 0.4 MB.
FIG S5, PDF file, 0.3 MB.
FIG S6, PDF file, 0.3 MB.
FIG S7, PDF file, 0.6 MB.
FIG S8, PDF file, 1.1 MB.
FIG S9, PDF file, 1.4 MB.
TABLE S1, XLSX file, 0.01 MB.

ACKNOWLEDGMENTS

REFERENCES

1. Fierer N. 2017. Embracing the unknown: disentangling the complexities of the soil microbiome. Nat Rev Microbiol 15:579–590. https://doi.org/10.1038/nrmicro.2017.87.

2. Li Z, Tian D, Wang B, Wang J, Wang S, Chen HYH, Xu X, Wang C, He N, Niu S. 2019. Microbes drive global soil nitrogen mineralization and availability. Glob Chang Biol 25:1078–1088. https://doi.org/10.1111/gcb.14557.

3. Xu X, Thornton PE, Post WM. 2013. A global analysis of soil microbial biomass carbon, nitrogen and phosphorus in terrestrial ecosystems. Glob Ecol Biogeogr 22:737–749. https://doi.org/10.1111/geb.12029.

4. Roberson EB, Firestone MK. 1992. Relationship between desiccation and exopolysaccharide production in a soil Pseudomonas sp. Appl Environ Microbiol 58:1284–1291. https://doi.org/10.1128/aem.58.4.1284-1291.1992.

5. Cavicchioli R, Ripple WJ, Timmis KN, Azam F, Bakken LR, Baylis M, Behrenfeld MJ, Boetius A, Boyd PW, Classen AT, Crowther TW, Danovaro R, Foreman CM, Huisman J, Hutchins DA, Jansson JK, Karl DM, Koskella B, Mark Welch DB, Martiny JBH, Moran MA, Orphan VJ, Reay DS, Remais JV, Rich VI, Singh BK, Stein LY, Stewart FJ, Sullivan MB, van Oppen MJH, Weaver SC, Webb EA, Webster NS. 2019. Scientists' warning to humanity: microorganisms and climate change. Nat Rev Microbiol 17:569–586. https://doi.org/10.1038/s41579-019-0222-5.

6. Sokol NW, Slessarev E, Marschmann GL, Nicolas A, Blazewicz SJ, Brodie EL, Firestone MK, Foley MM, Hestrin R, Hungate BA, Koch BJ, Stone BW, Sullivan MB, Zablocki O, LLNL Soil Microbiome Consortium. 2022. Life and death in the soil microbiome: how ecological processes influence biogeochemistry. Nature Rev Microbiology 20:415–430. https://doi.org/10.1038/s41579-022-00695-z.

7. Kallenbach CM, Frey SD, Grandy AS. 2016. Direct evidence for microbial-derived soil organic matter formation and its ecophysiological controls. Nat Commun 7:13630. https://doi.org/10.1038/ncomms13630.

8. Kallenbach CM, Grandy AS, Frey SD, Diefendorf AF. 2015. Microbial physiology and necromass regulate agricultural soil carbon accumulation. Soil Biol Biochem 91:279–290. https://doi.org/10.1016/j.soilbio.2015.09.005.

9. Liang C, Schimel JP, Jastrow JD. 2017. The importance of anabolism in microbial control over soil carbon storage. Nat Microbiol 2:17105. https://doi.org/10.1038/nmicrobiol.2017.105.

10. Fierer N, Jackson RB. 2006. The diversity and biogeography of soil bacterial communities. Proc Natl Acad Sci U S A 103:626–631. https://doi.org/10.1073/pnas.0507535103.

11. Delgado-baquerizo M, Oliverio AM, Brewer TE, Benavent-gonzález A, Eldridge DJ, Bardgett RD, Maestre FT, Singh BK, Fierer N. 2018. A global atlas of the dominant bacteria found in soil. Science 359:320–325. https://doi.org/10.1126/science.aap9516.

12. Choudoir MJ, Buckley DH. 2018. Phylogenetic conservatism of thermal traits explains dispersal limitation and genomic differentiation of Streptomyces sister-taxa. ISME J 12:2176–2186. https://doi.org/10.1038/s41396-018-0180-3.

13. Choudoir MJ, Doroghazi JR, Buckley DH. 2016. Latitude delineates patterns of biogeography in terrestrial Streptomyces. Environ Microbiol 18:4931–4945. https://doi.org/10.1111/1462-2920.13420.

14. Greenlon A, Chang PL, Damtew ZM, Muleta A, Carrasquilla-Garcia N, Kim D, Nguyen HP, Suryawanshi V, Krieg CP, Yadav SK, Patel JS, Mukherjee A, Udupa S, Benjelloun I, Thami-Alami I, Yasin M, Patil B, Singh S, Sarma BK, von Wettberg EJB, Kahraman A, Bukun B, Assefa F, Tesfaye K, Fikre A, Cook DR. 2019. Global-level population genomics reveals differential effects of geography and phylogeny on horizontal gene transfer in soil bacteria. Proc Natl Acad Sci 2019:201900056. https://doi.org/10.1073/pnas.1900056116.

15. Woodcroft BJ, Singleton CM, Boyd JA, Evans PN, Emerson JB, Zayed AAF, Hoelzle RD, Lamberton TO, McCalley CK, Hodgkins SB, Wilson RM, Purvine SO, Nicora CD, Li C, Frolking S, Chanton JP, Crill PM, Saleska SR, Rich VI, Tyson GW. 2018. Genome-centric view of carbon processing in thawing permafrost. Nature 560:49–54. https://doi.org/10.1038/s41586-018-0338-1.

16. Nuccio EE, Starr E, Karaoz U, Brodie EL, Zhou J, Tringe SG, Malmstrom RR, Woyke T, Banfield JF, Firestone MK, Pett-Ridge J. 2020. Niche differentiation is spatially and temporally regulated in the rhizosphere. ISME J 14:999–1014. https://doi.org/10.1038/s41396-019-0582-x.

17. Butterfield CN, Li Z, Andeer PF, Spaulding S, Thomas BC, Singh A, Hettich RL, Suttle KB, Probst AJ, Tringe SG, Northen T, Pan C, Banfield JF. 2016. Proteogenomic analyses indicate bacterial methylotrophy and archaeal heterotrophy are prevalent below the grass root zone. PeerJ 4:e2687. https://doi.org/10.7717/peerj.2687.

18. Diamond S, Andeer PF, Li Z, Crits-Christoph A, Burstein D, Anantharaman K, Lane KR, Thomas BC, Pan C, Northen TR, Banfield JF. 2019. Mediterranean grassland soil C–N compound turnover is dependent on rainfall and depth, and is mediated by genomically divergent microorganisms. Nat Microbiol 4:1356–1367. https://doi.org/10.1038/s41564-019-0449-y.

19. Brown CT, Olm MR, Thomas BC, Banfield JF. 2016. Measurement of bacterial replication rates in microbial communities. Nat Biotechnol 34:1256–1263. https://doi.org/10.1038/nbt.3704.

20. Starr EP, Nuccio EE, Pett-Ridge J, Banfield JF, Firestone MK. 2019. Metatranscriptomic reconstruction reveals RNA viruses with the potential to shape carbon cycling in soil. Proc Natl Acad Sci U S A 116:25900–25908. https://doi.org/10.1073/pnas.1908291116.

21. Emerson JB, Roux S, Brum JR, Bolduc B, Woodcroft BJ, Jang HB, Singleton CM, Solden LM, Naas AE, Boyd JA, Hodgkins SB, Wilson RM, Trubl G, Li C, Frolking S, Pope PB, Wrighton KC, Crill PM, Chanton JP, Saleska SR, Tyson GW, Rich VI, Sullivan MB. 2018. Host-linked soil viral ecology along a permafrost thaw gradient. Nat Microbiol 3:870–880. https://doi.org/10.1038/s41564-018-0190-y.

22. Long AM, Hou S, Ignacio-Espinoza JC, Fuhrman JA. 2021. Benchmarking microbial growth rate predictions from metagenomes. ISME J 15:183–195. https://doi.org/10.1038/s41396-020-00773-1.

23. Papp K, Hungate BA, Schwartz E. 2018. Microbial rRNA synthesis and growth compared through quantitative stable isotope probing with H218O. Appl Environ Microbiol 84:e02441-17.

24. Couradeau E, Sasse J, Goudeau D, Nath N, Hazen TC, Bowen BP, Chakraborty R, Malmstrom RR, Northen TR. 2019. Probing the active fraction of soil microbiomes using BONCAT-FACS. Nat Commun 10:2770. https://doi.org/10.1038/s41467-019-10542-0.

25. Crombie AT, Larke-Mejia NL, Emery H, Dawson R, Pratscher J, Murphy GP, McGenity TJ, Murrell JC. 2018. Poplar phyllosphere harbors disparate isoprene-degrading bacteria. Proc Natl Acad Sci U S A 115:13081–13086. https://doi.org/10.1073/pnas.1812668115.

26. Wilhelm RC, Singh R, Eltis LD, Mohn WW. 2019. Bacterial contributions to delignification and lignocellulose degradation in forest soils with metagenomic and quantitative stable isotope probing. ISME J 13:413–429. https://doi.org/10.1038/s41396-018-0279-6.

27. Thomas F, Corre E, Cébron A. 2019. Stable isotope probing and metagenomics highlight the effect of plants on uncultured phenanthrene-degrading bacterial consortium in polluted soil. ISME J 13:1814–1830. https://doi.org/10.1038/s41396-019-0394-z.

28. Sieradzki ET, Morando M, Fuhrman JA. 2021. Metagenomics and quantitative stable isotope probing offer insights into metabolism of polycyclic aromatic hydrocarbon degraders in chronically polluted seawater. mSystems 6. https://doi.org/10.1128/mSystems.00245-21.

29. Starr EP, Shi S, Blazewicz SJ, Probst AJ, Herman DJ, Firestone MK, Banfield JF. 2018. Stable isotope informed genome-resolved metagenomics

reveals that Saccharibacteria utilize microbially-processed plant-derived carbon. Microbiome 6. https://doi.org/10.1186/s40168-018-0499-z.

30. Starr EP, Shi S, Blazewicz SJ, Koch BJ, Probst AJ, Hungate BA, Pett-Ridge J, Firestone MK, Banfield JF. 2021. Stable-isotope-informed, genome-resolved metagenomics uncovers potential cross-kingdom interactions in rhizosphere soil. MSphere 6. https://doi.org/10.1128/mSphere.00085-21.

31. Blazewicz SJ, Hungate BA, Koch BJ, Nuccio EE, Morrissey E, Brodie EL, Schwartz E, Pett-Ridge J, Firestone MK. 2020. Taxon-specific microbial growth and mortality patterns reveal distinct temporal population responses to rewetting in a California grassland soil. ISME J 14:1520–1532. https://doi.org/10.1038/s41396-020-0617-3.

32. Koch BJ, McHugh TA, Hayer M, Schwartz E, Blazewicz SJ, Dijkstra P, Gestel N, Marks JC, Mau RL, Morrissey EM, Pett-Ridge J, Hungate BA. 2018. Estimating taxon-specific population dynamics in diverse microbial communities. Ecosphere 9. https://doi.org/10.1002/ecs2.2090.

33. Foley MM, Blazewicz SJ, McFarlane KJ, Greenlon A, Hayer M, Kimbrel JA, Koch BJ, Monsaint-Queeney V, Morrison K, Morrissey E, Hungate BA, Pett-Ridge J. 2022. Active populations and growth of soil microorganisms are framed by mean annual precipitation in three California annual grasslands. bioRxiv. https://doi.org/10.1101/2021.12.06.471491.

34. Blazewicz SJ, Schwartz E. 2011. Dynamics of 18O incorporation from H2 18O into soil microbial DNA. Microb Ecol 61:911–916. https://doi.org/10.1007/s00248-011-9826-7.

35. Rendulic S, Jagtap P, Rosinus A, Eppinger M, Baar C, Lanz C, Keller H, Lambert C, Evans KJ, Goesmann A, Meyer F, Sockett RE, Schuster SC. 2004. A predator unmasked: life cycle of Bdellovibrio bacteriovorus from a genomic perspective. Science 303:689–692. https://doi.org/10.1126/science.1093027.

36. Shaffer M, Borton MA, McGivern BB, Zayed AA, La Rosa SL, Solden LM, Liu P, Narrowe AB, Rodríguez-Ramos J, Bolduc B, Gazitúa MC, Daly RA, Smith GJ, Vik DR, Pope PB, Sullivan MB, Roux S, Wrighton KC. 2020. DRAM for distilling microbial metabolism to automate the curation of microbiome function. Nucleic Acids Res 48:8883–8900. https://doi.org/10.1093/nar/gkaa621.

37. Williamson KE, Fuhrmann JJ, Wommack KE, Radosevich M. 2017. Viruses in soil ecosystems: an unknown quantity within an unexplored territory. Annu Rev Virol 4:201–219. https://doi.org/10.1146/annurev-virology-101416-041639.

38. Trubl G, Kimbrel JA, Liquet-Gonzalez J, Nuccio EE, Weber PK, Pett-Ridge J, Jansson JK, Waldrop MP, Blazewicz SJ. 2021. Ecology of active viruses and their bacterial hosts in frozen Arctic peat soil revealed with H218O stable isotope probing metagenomics. Microbiome 9. https://doi.org/10.1186/s40168-021-01154-2.

39. ter Horst AM, Santos-Medellín C, Sorensen JW, Zinke LA, Wilson RM, Johnston ER, Trubl GG, Pett-Ridge J, Blazewicz SJ, Hanson PJ, Chanton JP, Schadt CW, Kostka JE, Emerson JB. 2021. Minnesota peat viromes reveal terrestrial and aquatic niche partitioning for local and global viral populations. Microbiome 9. https://doi.org/10.1186/s40168-021-01156-0.

40. Nunan N, Schmidt H, Raynaud X. 2020. The ecology of heterogeneity: soil bacterial communities and C dynamics. Philos Trans R Soc Lond B Biol Sci 375. https://doi.org/10.1098/rstb.2019.0249.

41. Hungate BA, Marks JC, Power ME, Schwartz E, Jan van Groenigen K, Blazewicz SJ. 2021. The functional significance of bacterial predators. mBio 12:e00466.

42. Lindell D, Jaffe JD, Coleman ML, Futschik ME, Axmann IM, Rector T, Kettler G, Sullivan MB, Steen R, Hess WR, Church GM, Chisholm SW. 2007. Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. Nature 449:83–86. https://doi.org/10.1038/nature06130.

43. Coskun ÖK, Özen V, Wankel SD, Orsi WD. 2019. Quantifying population-specific growth in benthic bacterial communities under low oxygen using H218O. ISME J 13:1546–1559. https://doi.org/10.1038/s41396-019-0373-4.

44. Chen H, Mothapo N. v, Shi W. 2015. Soil moisture and pH control relative contributions of fungi and bacteria to N2O production. Microb Ecol 69:180–191. https://doi.org/10.1007/s00248-014-0488-0.

45. Barnett SE, Buckley DH. 2020. Simulating metagenomic stable isotope probing datasets with MetaSIPSim. BMC Bioinformatics 21:37. https://doi.org/10.1186/s12859-020-3372-6.

46. Zwetsloot MJ, Ucros JM, Wickings K, Wilhelm RC, Sparks J, Buckley DH, Bauerle TL. 2020. Prevalent root-derived phenolics drive shifts in microbial community composition and prime decomposition in forest soil. Soil Biol Biochem 145:107797. https://doi.org/10.1016/j.soilbio.2020.107797.

47. Wilhelm RC, DeRito CM, Shapleigh JP, Madsen EL, Buckley DH. 2021. Phenolic acid-degrading Paraburkholderia prime decomposition in forest soil. ISME Commun 1:1–12.

48. Morrissey EM, Mau RL, Schwartz E, Koch BJ, Hayer M, Hungate BA. 2018. Taxonomic patterns in the nitrogen assimilation of soil prokaryotes. Environ Microbiol 20:1112–1119. https://doi.org/10.1111/1462-2920.14051.

49. Sieradzki ET, Koch BJ, Greenlon A, Sachdeva R, Malmstrom RR, Mau RL, Blazewicz SJ, Firestone MK, Hofmockel KS, Schwartz E, Hungate BA, Pett-Ridge J. 2020. measurement error and resolution in quantitative stable isotope probing: implications for experimental design. mSystems 5. https://doi.org/10.1128/mSystems.00151-20.

50. Nuccio EE, Blazewicz SJ, Lafler M, Campbell AN, Kakouridis A, Kimbrel JA, Wollard J, Vyshenska D, Riley R, Tomatsu A, Hestrin R, Malmstrom RR, Firestone M, Pett-Ridge J. 2022. HT-SIP: a semi-automated stable isotope probing pipeline identifies interactions in the hyphosphere of arbuscular mycorrhizal fungi. bioRxiv. https://doi.org/10.1101/2022.07.01.498377.

51. Joshi NA, Fass JN. 2011. Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33). https://github.com/najoshi/sickle.

52. Li D, Liu CM, Luo R, Sadakane K, Lam TW. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31:1674–1676. https://doi.org/10.1093/bioinformatics/btv033.

53. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119. https://doi.org/10.1186/1471-2105-11-119.

54. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2016. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res 44:D457–D462. https://doi.org/10.1093/nar/gkv1070.

55. Chan PP, Lowe TM. 2019. tRNAscan-SE: searching for tRNA genes in genomic sequences. Methods Mol Biol 1962:1–14. https://doi.org/10.1007/978-1-4939-9173-0_1.

56. Zhou Z, Tran PQ, Breister AM, Liu A, Kieft K, Cowley ES, Karaoz U, Anantharaman K. 2020. METABOLIC: high-throughput profiling of microbial genomes for functional traits, biogeochemistry, and community-scale metabolic networks. bioRxiv. https://doi.org/10.1101/761643.

57. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z. 2019. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ 7:e7359. https://doi.org/10.7717/peerj.7359.

58. Wu YW, Simmons BA, Singer SW. 2016. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics 32:605–607. https://doi.org/10.1093/bioinformatics/btv638.

59. Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, Lahti L, Loman NJ, Andersson AF, Quince C. 2014. Binning metagenomic contigs by coverage and composition. Nat Methods 11:1144–1146. https://doi.org/10.1038/nmeth.3103.

60. Uritskiy GV, DiRuggiero J, Taylor J. 2018. MetaWRAP-a flexible pipeline for genome-resolved metagenomic data analysis. Microbiome 6:158. https://doi.org/10.1186/s40168-018-0541-1.

61. Olm MR, Brown CT, Brooks B, Banfield JF. 2017. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME J 11:2864–2868. https://doi.org/10.1038/ismej.2017.126.

62. Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460–2461. https://doi.org/10.1093/bioinformatics/btq461.

63. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T. 2019. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res 47:W81–W87. https://doi.org/10.1093/nar/gkz310.

64. Bundalovic-Torma C, Whitfield GB, Marmont LS, Lynne HP, Parkinson J. 2020. A systematic pipeline for classifying bacterial operons reveals the evolutionary landscape of biofilm machineries. PLoS Comput Biol 16:e1007721. https://doi.org/10.1371/journal.pcbi.1007721.

65. Méheust R, Burstein D, Castelle CJ, Banfield JF. 2019. The distinction of CPR bacteria from other bacteria based on protein family content. Nat Commun 10:4173. https://doi.org/10.1038/s41467-019-12171-z.

66. Ye Y, Doak TG. 2009. A parsimony approach to biological pathway reconstruction/inference for genomes and metagenomes. PLoS Comput Biol 5:e1000465–8. https://doi.org/10.1371/journal.pcbi.1000465.

67. Roux S, Enault F, Hurwitz BL, Sullivan MB. 2015. VirSorter: mining viral signal from microbial genomic data. PeerJ 3:e985. https://doi.org/10.7717/peerj.985.

68. Ren J, Song K, Deng C, Ahlgren NA, Fuhrman JA, Li Y. 2018. Identifying viruses from metagenomic data by deep learning. http://arxiv.org/abs/1806.07810.

69. Gregory AC, Zayed AA, Conceição-Neto N, Temperton B, Bolduc B, Alberti A, Ardyna M, Arkhipova K, Carmichael M, Cruaud C, Dimier C, Domínguez-Huerta G, Ferland J, Kandels S, Liu Y, Marec C, Pesant S, Picheral M, Pisarev S, Poulain J, Tremblay J-É, Vik D, Babin M, Bowler C, Culley AI, de Vargas C, Dutilh BE, Iudicone D, Karp-Boss L, Roux S, Sunagawa S, Wincker P, Sullivan MB, Tara Oceans Coordinators. 2019. Marine DNA viral macro- and microdiversity from pole to pole. Cell 177:1109–1123.e14. https://doi.org/10.1016/j.cell.2019.03.040.

70. Bolduc B, Zayed A. 2020. BitBucket repository. https://bitbucket.org/MAVERICLab/virmatcher/.

71. Gregory AC, Zablocki O, Zayed AA, Howell A, Bolduc B, Sullivan MB. 2020. The gut virome database reveals age-dependent patterns of virome diversity in the human gut. Cell Host Microbe 28:724–740. https://doi.org/10.1016/j.chom.2020.08.003.

72. Galiez C, Siebert M, Enault F, Vincent J, Söding J. 2017. WIsH: who is the host? Predicting prokaryotic hosts from metagenomic phage contigs. Bioinformatics 33:3113–3114. https://doi.org/10.1093/bioinformatics/btx383.

73. Hungate BA, Mau RL, Schwartz E, Caporaso JG, Dijkstra P, van Gestel N, Koch BJ, Liu CM, McHugh TA, Marks JC, Morrissey EM, Price LB. 2015. Quantitative microbial ecology through stable isotope probing. Appl Environ Microbiol 81:7570–7581. https://doi.org/10.1128/AEM.02280-15.

74. Oksanen J, Guillaume Blanchet M, Friendly F, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Szoecs H, Wagner HH. 2020. vegan: Community Ecology Package. R package version 2.5–7. https://CRAN.R-project.org/package=vegan.

75. Legendre P, Anderson MJ. 1999. Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. Ecological Monographs 69:1–24. https://doi.org/10.1890/0012-9615(1999)069[0001:DBRATM]2.0.CO;2.

76. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: high-resolution sample inference from Illumina amplicon data. Nat Methods 13:581–583. https://doi.org/10.1038/nmeth.3869.

77. McMurdie PJ, Holmes S. 2013. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One 8:e61217. https://doi.org/10.1371/journal.pone.0061217.