Research Square

Preprints are preliminary reports that have not undergone peer review. They should not be considered conclusive, used to inform clinical practice, or referenced by the media as validated information.

Competitive exclusion and metabolic dependency among microorganisms structure the cellulose economy of agricultural soil

Roland C Wilhelm (rcw239@cornell.edu)
https://orcid.org/0000-0003-1170-1753

Charles Pepe-Ranney
Cornell University

Pamela Weisenhorn
Argonne National Laboratory

Mary Lipton
Pacific Northwest National Laboratory

Daniel H. Buckley
Cornell University

Research

Keywords: decomposition, metagenomics, stable isotope probing, metaproteomics, metabolic dependency, competitive exclusion, surface ecology, soil carbon cycling.

Posted Date: March 17th, 2020

DOI: https://doi.org/10.21203/rs.2.23522/v2

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Version of Record: A version of this preprint was published at mBio on February 23rd, 2021. See the published version at https://doi.org/10.1128/mBio.03099-20.

Abstract

Background Microorganisms that degrade cellulose utilize extracellular processes that yield free intermediates which promote interactions with non-cellulolytic organisms. We hypothesized that these interactions determine the ecological and physiological traits governing the fate of cellulosic carbon (C) in soil. We employed metagenomic-SIP and metaproteomics to characterize the attributes of cellulolytic and non-cellulolytic microbes accessing 13C from cellulose. We hypothesized that cellulolytic taxa would exhibit competitive traits to limit access, while non-cellulolytic taxa would display metabolic dependency, such as signatures of adaptive gene loss. We tested this hypothesis by evaluating genomic traits indicative of competitive exclusion or metabolic dependency, such as antibiotic production, growth rate, surface attachment, biomass degrading potential and auxotrophy. Results The most 13C-enriched taxa were cellulolytic Cellvibrio (Gammaproteobacteria) and Chaetomium (Ascomycota), which exhibited a strategy of self-sufficiency (prototrophy), rapid growth, and competitive exclusion via antibiotic production. These ruderal taxa were common indicators of soil disturbance in agroecosystems, such as tillage and fertilization. Auxotrophy was more prevalent in cellulolytic Actinobacteria than in cellulolytic Proteobacteria, demonstrating differences in dependency among cellulose degraders. Non-cellulolytic taxa that accessed 13C from cellulose (Planctomycetales, Verrucomicrobia and Vampirovibrionales) were highly dependent, as indicated by patterns of auxotrophy and 13C-labeling (i.e. partial labelling or labeling at later-stages). Major 13C-labeled cellulolytic microbes (e.g. Sorangium, Actinomycetales, Rhizobiales and Caulobacteraceae) possessed adaptations for surface colonization (e.g. gliding motility, hyphae, attachment structures) signifying the importance of surface ecology in decomposition. Conclusions Our results demonstrate that access to cellulose was accompanied by ecological trade-offs characterized by differing degrees of metabolic dependency and competitive exclusion. These trade-offs influence microbial growth dynamics on particulate organic carbon and reveal that the fate of carbon is governed by a complex economy within the microbial community. We propose three ecological groups for microbes participating in this economy: (i) independent primary degraders, (ii) integrated primary degraders and (iii) mutualists, opportunists and parasites.

Background

The major structural component of plant biomass, cellulose, is degraded in soil by a diverse and interacting community of microorganisms [1]. Since cellulose is insoluble and highly crystalline, it cannot be transported across cell membranes. Hence, microorganisms rely on extracellular reactions to digest cellulose fibers into oligodextrins, cellobiose, and glucose for transport into the cell. Due to the structural complexity of lignocellulose, cellulose degradation is facilitated by synergistic interactions between diverse enzyme systems that differ in specific activity [2,3]. Various physiological traits also influence the colonization and deconstruction of cellulose fibers, such as hyphal growth by members of fungi and *Actinobacteria* [4,5], gliding motility in *Bacteroidetes* [6], and the formation of cellulosomes by many anaerobes [7,8]. These circumstances suggest that cellulose degradation in soil is predisposed to metabolic and ecological interactions that govern access to the soluble products of cellulose

degradation. However, most approaches to studying cellulose decomposition have overlooked the ecological interactions occurring within microbial consortia, since most cellulolytic microorganisms have been characterized in isolation or in co-culture [9–14]. Phylogenetic and functional gene-based metagenomics coupled with DNA stable isotope probing (SIP) now provides the capability to study the ecophysiological traits of the diverse microorganisms participating in the cellulose economy as it occurs in soil.

The extracellular nature of cellulose degradation creates conditions where the fitness of individuals is contingent on both competition and facilitation. Competition for cellulose and its degradation products impose fitness costs on cellulolytic organisms, promoting antagonistic interactions [15–17]. However, facilitation by commensal and mutualistic partners enhance degradation rates to the benefit of cellulolytic organisms [18]. Many cellulolytic microbes have close relatives lacking in endoglucanases, suggesting adaptive benefits from the gain or loss of these genes [19]. The beneficiaries of community metabolism should be expected to shed energetically costly traits, resulting in adaptive gene loss and evolution of metabolic dependency [20]. For example, non-cellulolytic bacteria can complement the metabolic functions of cellulolytic bacteria *in vitro*, through catabolic reactions [21,22], vitamin biosynthesis [23], amino acid biosynthesis [24], or biosynthesis of other essential metabolites [25]. Such metabolic dependency can occur through specific, tightly coupled interactions, such as those of syntrophic partners, but it can also be loosely coupled and non-specific, such as when cells obtain metabolites or other nutrients from extracellular pools replenished by mortality of community members [26,27]. Hence, we expect that the anabolic and catabolic byproducts of cellulolytic microbes influence overall community structure and function and thereby govern the fate of cellulosic carbon in soil.

Shotgun metagenomics and the recovery of metagenome-assembled genomes (MAGs) provides a cultivation-independent means of studying the phylogenetic and functional characteristics of microbial communities. This approach has been used to identify ecophysiological traits [28,29] and study metabolic dependency in environmental populations [30]. With DNA-SIP, one can identify MAGs from organisms that assimilate ^{13}C, either directly or indirectly, from ^{13}C-labeled substrates by separating and sequencing the ^{13}C-enriched DNA (Figure S1). Metagenomic-SIP proved effective in resolving traits of cellulolytic and lignolytic populations in forest soil, in part, by improving MAG recovery [14]. DNA-SIP can be used to estimate the degree of ^{13}C-labeling of individual MAGs by measuring the change in buoyant density across the CsCl gradient [31,32]. This approach offers the capacity to distinguish between highly ^{13}C-enriched DNA corresponding to taxa with primary access to cellulose-C, less ^{13}C-enriched DNA corresponding to microbes with peripheral access to cellulose-C, and unenriched DNA derived from the broader soil community. The genomic information from these groups can then be used to study differences in ecological traits, such as metabolic dependency or antibiotic production, encoded by members of the cellulose economy.

We performed gradient-resolved metagenomic-SIP on ^{13}C-labelled DNA from an agricultural soil following a 30-day incubation with ^{13}C-labeled cellulose. This experiment was designed to delineate membership in

a cellulose degrading soil consortium. We hypothesized that ^{13}C-enriched, cellulolytic taxa would display diverse ecological strategies. We expected early cellulose colonists to depend less on the products of community metabolism than later colonists, which will be evident in varying degrees of auxotrophy. We expected that cellulolytic microbes would be enriched in secondary metabolite gene clusters (SMs), such as those that synthesize antibiotics, to control access to community resources. We further hypothesized that ^{13}C-enriched, non-cellulolytic microbes would exhibit signatures of metabolic dependency based on the degree of auxotrophy and/or capacity for degrading microbial necromass (*i.e.* numbers of genes encoding nucleases, peptidases, chitinases, and other hydrolytic enzymes). Additionally, we identified traits related to surface colonization that we anticipated to be important features of community interactions on insoluble cellulosic fibres. We validated our results with reference genomes and performed metaproteomics to confirm gene expression by target groups.

Results

Overview of Metagenome Assembly and Designation of ^{13}C-Enrichment

Shotgun metagenomic sequencing of ^{13}C-labeled DNA recovered a total of 1.1 billion reads after quality filtering, estimated by nonpareil [74] to cover 80% of genomic diversity in the DNA pool. Metagenome assembly produced a total of 356,131 contigs greater than 2.5 Kb, amounting to a total length of 1.8 Gb (~ 230 genomes of 8 Mb). The degree of ^{13}C-enrichment was estimated for each contig by comparing the CsCl gradient profile to simulated natural abundance profiles (Figure S2). More than half of all contigs were designated either strongly or weakly ^{13}C-enriched (total length = 921 Mb), while the remainder were from genomes of abundant soil taxa that lacked evidence of ^{13}C-labelling (766 Mb). Contigs clustered into coherent sets according to pentanucleotide frequency which grouped by patterns of ^{13}C-labelling and taxonomy (Figure 1ab), and not average G+C content (Figure 1c). The Random Forest model used to predict enrichment status had an overall accuracy of 89.1% with high sensitivity and specificity for both strongly enriched (98.3% and 86.4%, respectively) and unenriched contigs (100% and 93%) (details in Supplementary methods; Table S4).

Phylobin Characteristics

A total of 47, 2 and 46 phylobins greater than 1 Mb were produced from strongly, weakly and unenriched contig sets, respectively (Table S5). Of the 95 total phylobins, 38 were deemed high quality (>75% completeness) and were divided into four categories based on enrichment status and cellulolytic potential (inferred from the presence of endoglucanases): strongly ^{13}C-enriched and cellulolytic (n_{strong} = 12), strongly ^{13}C-enriched and non-cellulolytic (n_{strong} = 8), weakly ^{13}C-enriched and non-cellulolytic (n_{weak} = 2), and unenriched (n = 16). Each phylobin represented a genomic-ecological unit, rather than an individual genome, encompassing genomes from, at most, four to seven genera for enriched and unenriched phylobins, respectively, based on the diversity of assembled full-length 16S rRNA genes (Figure 2). Both phylobins and representative reference genomes from ^{13}C-enriched cellulolytic taxa were

larger (μ_{PhyBin} = 24.7 Mb and μ_{Rep} = 7.1 Mb) than those from ^{13}C-enriched non-cellulolytic taxa (μ_{PhyBin} = 12.5 Mb and μ_{Rep} = 5.2 Mb; Wilcoxon test, $p \leq 0.05$) and unenriched taxa (μ_{PhyBin} = 11.4 Mb and μ_{Rep} = 5.0 Mb; $p < 0.01$). Analysis of single-copy genes and single nucleotide polymorphisms per single-copy gene indicated that the large size of phylobins resulted from natural pangenomic diversity (intra-species) and inclusion of genome fragments from closely related taxa (inter-species diversity; details in Supplementary Methods).

The Structure of a Cellulose Economy

Taxa designated as strongly ^{13}C-enriched and cellulolytic (*i.e.* encoding endoglucanases) represented the greatest proportion of unassembled SSU gene fragments in metagenomes. The most abundant were classified to well-known genera of cellulolytic soil organisms, including *Cellvibrio*, *Herpetosiphon (Chloroflexi)*, and members of the fungal order *Sordariales* (predominantly *Chaetomium*), as well as lesser-known cellulolytic genera, such as *Devosia* and *Sphingomonas* (Figure 2). Peptides from these five taxa were also abundant within the total metaproteome (n_{total} = 90,557 peptides; 33,765 unique proteins) occupying in the following percentages of total peptides: *Rhizobiales* (6.9%), *Cellvibrionales* (2.6%), *Sphingomonadales* (2.6%), *Herpetosiphon* (1.0%), and *Sordariales* (0.5%). Endoglucanases were detected in contigs classified to 22 of the 30 genera designated as strongly ^{13}C-enriched, consistent with their presence in reference genomes (Figure 2, Table S6). The 22 genera designated strongly ^{13}C-enriched and cellulolytic comprised 29% of the total SSU rRNA gene fragments. In contrast, the seven genera designated as strongly ^{13}C-enriched and non-cellulolytic (see Figure 2) comprised 2.3% of recovered SSU rRNA reads. These putative non-cellulolytic taxa included only one reference genome encoding an endoglucanase.

A diverse set of endoglucanases were recovered from phylobins, revealing a snapshot of the functional diversity of cellulolytic populations. A total of 426 unique endoglucanase homologs were identified (at a > 80% identity threshold) belonging to 39 different CAZy families/sub-families. Eighty-two of these endoglucanases were present within gene clusters that contained a carbohydrate-binding module. A total of 54 peptides in the metaproteome matched endoglucanases; the most abundant was a GH9 from *Cellvibrio* (Figure 3). The second and third most abundant endoglucanases in the metaproteome were encoded by fungi (GH131 and GH7). Overall, most endoglucanases in the metaproteome were encoded by fungi (57%), which was disproportionate to the total relative abundance of fungal peptides (1.2%) in the whole metaproteome.

Evidence for Metabolic Dependency and Competitive Exclusion

To evaluate potential interactions among ^{13}C-labeled taxa, we assessed the degree of auxotrophy (as an indicator of metabolic dependency) and presence of SM-encoding genes (antibiotic-based competition) in phylobins and their representative genomes. No phylobin or genome was fully prototrophic or auxotrophic for all biosynthetic pathways evaluated (n = 32), with the average phylobin being auxotrophic for 6 amino acids, 1 cofactor and 2 vitamins and the average representative genome

auxotrophic for 5 amino acids, 1 cofactor and 1 vitamin. The extent of auxotrophy did not differ significantly between ^{13}C-enriched cellulolytic and ^{13}C-enriched non-cellulolytic phylobins or representative genomes (Figure 4) but did vary among dominant taxa within each group.

The most prototrophic representative genomes were *Cellvibrio* (31/32 pathways detected; genome size = 5.2 Mb; designated: ^{13}C-enriched, cellulolytic), *Devosia* (31/32; 4.2 Mb; ^{13}C-enr. cellulolytic) and *Leptothrix* (31/32; 4.9 Mb; unenr. non-cellulolytic) (see Table S7). The most auxotrophic representative genomes were *Planctomyces* (13/32; 3.2 Mb; ^{13}C-enr. non-cellulolytic), *Nannocystis* (16/32; 12.1 Mb; ^{13}C-enr. non-cellulolytic) and *Vampirovibrio* (16/32; 3.0 Mb; ^{13}C-enr. non-cellulolytic). These trends were consistent in phylobins, where *Cellvibrionales* (30/32; ranked 1st in terms of biosynthetic capacity among the 38 phylobins examined) and *Rhizobiales* (28/32; ranked 3rd) were among the most prototrophic, while *Planctomycetales* (24/32; 20th), *Vampirovibrionales* (15/32; 34th), *Chthoniobacterales* (10/32; 37th; ^{13}C-enr. non-cellulolytic) and *Chloroflexales* (9/32; 38th, ^{13}C-enr. cellulolytic) were among the most auxotrophic. Overall, representative genomes from the phylum *Actinobacteria* were significantly more auxotrophic than *Proteobacteria* (μ_{actino} = 24.2 vs. μ_{proteo} = 25.3; Wilcoxon test, p = 0.04) driven by largely by *Alphaproteobacteria* (μ_{alpha} = 26.2, p = 0.003; Figure S3). This trend was consistent, but not significant, in phylobins (μ_{actino} = 23.8 versus μ_{alpha} = 26.0). Representative genomes for *Actinobacteria* and *Alphaproteobacteria* did not significantly differ in size (μ_{actino} = 5.8 Mb vs. μ_{alpha} = 5.3 Mb; Wilcoxon test, p = 0.46) or completeness (μ = 99.6% in both; p = 0.71).

The number of SM genes encoded in ^{13}C-enriched cellulolytic phylobins (μ_{rep} = 14.0; μ_{PhyBin} = 40.0) was significantly higher than in ^{13}C-enriched non-cellulolytic (μ_{rep} = 5.8; μ_{PhyBin} = 12.4; Wilcoxon test, $p < 0.01$) or unenriched phylobins (μ_{rep} = 4.8; μ_{PhyBin} = 10.0). The trend remained after normalizing to total phylobin or reference genome size: 2.2 read counts per million (rcpm) versus 1.9 rcpm and 1.3 rcpm, and 1.6 rcpm versus 1.0 rcpm and 1.0 rcpm, respectively. The genomes encoding the greatest number of SMs were *Sporocytophaga*, several *Actinobacteria* (*Streptomyces*, *Lentzea*, *Dactylosporangium* and *Kitasatospora*) and *Cellvibrio* (Table S7). Genes encoding type 1 polyketide synthases were consistently more abundant in ^{13}C-enriched cellulolytic taxa than the other two groups (Figure S4). Non-ribosomal peptide synthetases and bacteriocins were more frequently encoded in both ^{13}C-enriched groups, but only peptides matching cellulolytic taxa were present in the metaproteome (Figure S4). The metaproteome was dominated by terpene synthases from *Actinobacteria*, bacteriocins from *Cellvibrio* and non-ribosomal peptide synthetases from *Sordariales*. In contrast to trends in auxotrophy, representative genomes of ^{13}C-enriched cellulolytic *Actinobacteria* encoded significantly higher numbers of SMs (μ = 20.0; n=6; p=0.03) than the cellulolytic *Alphaproteobacteria* (μ=5.3; n=6).

The orders *Planctomycetales* and *Sphingomonadales* were represented by independent phylobins that were weakly ^{13}C-enriched, alongside those that were strongly ^{13}C-enriched and unenriched (Figure 2). Only the strongly ^{13}C-enriched *Sphingomonadales* phylobin encoded endoglucanases and also more SMs (predominantly bacteriocins) than the weakly ^{13}C-enriched and unenriched phylobins (2.1 rcpm, 0.4 rcpm

and 1.3 rcpm, respectively). Both ^{13}C-enriched *Sphingomonadales* phylobins shared the same pattern of auxotrophy (Figure S5). No *Planctomycetales* phylobin encoded endoglucanase, nor a substantial number of SMs.

Comparison of Cellulolytic and Hydrolytic Potential

The functional gene content of representative genomes explained substantial variation in enrichment status (Figure 5). The trend was driven primarily by the relative abundance of glycosyl hydrolases (GH), which were 1.5- to 3-fold higher (after normalization for genome size) in ^{13}C-enriched cellulolytic phylobins and corresponding reference genomes, respectively. This trend was evident in all gene families associated with lignocellulose degradation (GH, CBMs, AA and PL), which collectively explained 63% of variation in community functional composition along NMDS1 (Figure 5a; Table S8b). The genomes also separated along the secondary axis (NMDS2) defined by genome size, and peptidase and motility gene content, which explained 16.3, 16%, and 19% of variation, respectively (Table S8b). Degree of auxotrophy did not correlate with variation in functional gene content in either representative genomes or phylobins (Table S8a). In addition, the relative abundance of biomass-degrading enzymes (*e.g.* peptidases and nucleases) did not differ with respect to degree of ^{13}C enrichment or cellulolytic capacity, either in phylobins or representative genomes. In contrast, broad differences in functional gene categories were observed between *Actinobacteria* and *Proteobacteria* (Figure 5a).

Temporal Dynamics in Cellulose Economy

Early and late-stage colonizers of cellulose were identified according to genome-based predictions of growth rate (Table S7). Taxa designated as ^{13}C-enriched and cellulolytic were predicted to have faster generation times based on phylobins (3.0 hr) and representative genomes (3.1 hr) compared to ^{13}C-enriched non-cellulolytic taxa (5.7 hr and 5.6 hr, respectively; Figure 6a), though these differences were not significant (Kruskal-Wallis, p_{phybin} = 0.33 and p_{rep} = 0.52). However, genomes from ^{13}C-enriched non-cellulolytic taxa exhibited a bimodal distribution (Figure 6a) and the set of genomes with slower generation times (generation time > 5 hr) were significantly more auxotrophic (μ_{slow} = 15.8/32 prototrophies) than taxa with faster predicted generation times (< 3 hr; μ_{fast} = 23.8/32; Wilcoxon test, p = 0.05). The same trend was not apparent in phylobins, though predictions were unobtainable for two of the most auxotrophic bins (*Vampirovibrionales* and *Chthoniobacterales*).

The genome-based characterizations of early and late-stage colonizers were consistent with temporal patterns of taxa observed in time-course amplicon sequencing data. The highly prototrophic taxa *Cellvibrio* and *Devosia*, increased in relative abundance earliest, peaking in ^{13}C-enrichment at days 7 to 14 and declining by day 30 (Figure 6b; Figure S6a). *Chaetomium* were also early colonizers, showing ^{13}C-enrichment by day 7 (Figure S7). In contrast, the relative abundance of *Actinobacteria* was less dynamic, and these organisms tended to become labeled on, or after, day 14. Taxa predicted to be slowest growing, and identified as ^{13}C-enriched, non-cellulolytic, and highly auxotrophic (*Planctomyces*, *Sphingomonas*

and members of *Verrucomicrobia*), began to increase in relative abundance only after day 14 and were maximally ^{13}C-enriched on day 30 (Figure 6c; Figure S6c).

Surface Adhesion and Surface Motility

Phylobins from cellulolytic taxa were more likely to encode the capacity for surface adhesion and/or surface motility than other groups, including twitch motility, pili systems and fimbriae (Figure S8a). Surface attachment proteins were abundant in reference genomes from ^{13}C-enriched taxa (both cellulolytic and non-cellulolytic), and in phylobins classified as *Rhizobiales* and *Caulobacterales* (Figure S8b). Adhesion proteins used in gliding motility (*aglZ* and *sprB*) were present in reference genomes of cellulolytic taxa but absent from phylobins.

Discussion

We delineated members of a cellulolytic soil community using metagenomic-SIP to evaluate the ecological traits of microorganisms participating in the cellulose economy. Taxa identified as ^{13}C-enriched and cellulolytic had larger genomes and a greater number of genes encoding carbohydrate-active enzymes, secondary metabolites, surface motility or surface attachment, and tended to have faster generation times, when compared to ^{13}C-enriched non-cellulolytic and unenriched taxa. This evidence supports our hypothesis that the fate of cellulose carbon is mediated by ecological interdependencies among cellulolytic and non-cellulolytic taxa. Furthermore, ^{13}C-enriched cellulolytic taxa encoded diverse endoglucanases, representing 39 different subfamilies, but no single taxon encoded more than a third of these enzymes, suggesting the potential for synergistic decomposition. Auxotrophy was common among both ^{13}C-labeled cellulolytic and non-cellulolytic taxa, indicating that most taxa acquire essential metabolites from other community members. The average phylobin was auxotrophic for 9 of 32 pathways evaluated, though the highest levels of auxotrophy occurred among non-cellulolytic ^{13}C-labeled taxa.

The two most abundant cellulolytic taxa in the consortium, *Cellvibrio* and *Chaetomium*, were fast-growing and self-sufficient (i.e. prototrophic), both qualities of ruderal organisms. *Cellvibrio* dominated access to cellulosic C in two other SIP studies of agricultural soils [75] and, in one of the studies, were specific to agricultural soil [12]. The most abundant endoglucanase in our metaproteome, a GH9 from *Cellvibrio*, predominated in worm castings from agricultural soil (ACY24809) [76]. Both *Cellvibrio* and *Chaetomium* are commonly more abundant in tilled versus untilled fields (Figure S9) [77–80] and, the latter in disturbed forest soils [81]. The occurrence of *Chaetomium* in agroecosystems may be linked to nitrogen fertilization, given their enrichment in N-fertilized fields and wetlands [82–84]. The predominance of these ruderal cellulolytic taxa is indicative of the frequent soil disturbances in agroecosystems. Thus, it remains to be seen whether the cellulose economy of infrequently disturbed soils exhibits differing trends in the competitive exclusion or metabolic dependency reported here.

The ecological classes within the cellulose economy

Our results demonstrated that access to cellulosic carbon is mediated by trade-offs related to the capacity to produce carbohydrate-active enzymes, biosynthetic capacity, growth rate, and adaptation to colonize surfaces. Taken together our results suggest that, at least, three broad ecological groups of microorganism access ^{13}C from cellulose: (*i*) fast growing, biosynthetically competent cellulolytic taxa (*e.g. Cellvibrio* and *Devosia*), (*ii*) slow growing, metabolically-dependent (more auxotrophic), cellulolytic taxa (*e.g. Actinobacteria*), and (*iii*) slow growing, metabolically-dependent (highly auxotrophic), non-cellulolytic taxa (*e.g. Planctomycetales*, *Verrucomicrobia* and *Vampirovibrionales*). Certainly, a wide range of adaptive traits will affect access to cellulose carbon during decomposition, but these three categories provide a framework we can use to dissect community interactions that affect carbon cycling.

Independent primary degraders

Bacteria in the first category are first to colonize cellulosic materials based on their cellulolytic competency, self-sufficiency and rapid growth. On average, the phylobins and representative genomes of ^{13}C-enriched cellulolytic taxa were more prototrophic and had lower minimum generation times than their ^{13}C-enriched non-cellulolytic counterparts, though these results were statistically insignificant due to phylogenetic and ecological diversity within groups. *Cellvibrio* and *Devosia* were among the most enriched taxon in the ^{13}C-DNA pool (1st and 4th, respectively) and were the two most prototrophic of any genome or phylobin examined. *Cellvibrio* and *Devosia* populations peaked earlier than any other ^{13}C-enriched taxa and were in decline as dependent taxa increased in relative abundance. The yeast *Chaetomium* exhibited similar trends of early ^{13}C-enrichment, occupying upwards of 20% of the ^{13}C-DNA pool by day 7 in a sibling study at the same field site [34]. *Chaetomium* are also prototrophic, being capable of growth on cellulose in minimal media without the addition of amino acids or cofactors [85], though our methods (designed for prokaryotes) failed to accurately annotate eukaryotic genomes. The rapid growth and self-sufficiency of *Cellvibrio* and *Chaetomium* were coupled with a strategy of competitive exclusion via the production of antibiotics such as bacteriocin, likely a cellvibriocin [86], and fungicides [87–89]. We expect the competitive nature of these early colonizers and their metabolic by-products to influence the ability of non-cellulolytic taxa to access cellulosic C.

Integrated primary degraders

Bacteria in the second category, primarily *Actinobacteria* but also *Herpetosiphon* (*Chloroflexi*), were cellulolytic but exhibited higher levels of auxotrophy and SM production than early colonists. Populations of *Actinobacteria* lagged in comparison to *Cellvibrio*, with the first signs of ^{13}C-labelling appearing at day 14, and populations did not increase consistently over time. These trends suggest a greater integration with other population that exert top down (mortality driven) or bottom up (nutrient limitation as a result of competition for nutrients) control. *Actinobacteria* encoded and produced the greatest number of SMs and SM peptides, including an abundance of terpenoids which can function in interspecific signaling in soil, potentially facilitating mutualistic interactions [90]. The potential benefit of metabolic dependency for cellulolytic *Actinobacteria* was apparent in their consistent auxotrophy for four of the costliest non-aromatic amino acids to synthesize, namely: isoleucine (ranked 1st), leucine (2nd), methionine (3rd) and

lysine (4th) [91,92]. We hypothesize, based on their cellulolytic capacity; SM production, and high degree of auxotrophy, that the fitness of integrated primary degraders depends on community interactions.

Mutualists, opportunists and parasites

The third ecological group we observed, the 'MOP,' were metabolically dependent, late-stage colonizers of cellulose, characterized by the inability to degrade cellulose and high levels of auxotrophy. The MOP were comprised of *Planctomycetales*, *Vampirovibrionales, and Verrucomicrobia* (*Luteolibacter*, Candidatus *Xiphinematobacter* and *01D2Z36*), which reached maximal relative abundance after *Cellvibrio, Devosia* and *Chaetomium,* and remained abundant even after their decline. This pattern suggests dependence on products of community metabolism either through co-metabolism, the consumption of metabolic by-products or the consumption of macromolecules released during the turnover of microbial biomass. Indeed, these taxa all have traits that indicate lifestyles characterized by dependency on other microorganisms.

Planctomyces are commonly found to colonize the surfaces of marine algae, and to metabolize forms of algal polysaccharides, but not cellulose [93–95]. They purportedly assimilate oligosaccharides into their cells, indicating the ability to scavenge higher molecular weight degradation by-products [13,96–98]. The capacity of *Planctomyces* to attach to surfaces with holdfast, and their distinct tolerance to a range of antibiotics, would advantage an opportunistic lifestyle, particularly amongst antibiotic-producing primary degraders [95,99,100]. Cultured representatives for two other highly auxotrophic ^{13}C-enriched non-cellulolytic phylobins are obligate symbionts, namely *Vampirovibrio* and Candidatus *Xiphinematobacter*. The former are algal parasites that encode a range of GHs [101] but lack endoglucanases, and the latter are endobionts of nematodes, and are commonly observed in forest litter, cellulose-degrading consortia or in associated with *Basidiomycota* [102–105].

One set of phylobins provided evidence for what could be considered opportunistic 'cheating' [20]. Phylobins from *Sphingomonadales* differed in terms of weak and strong ^{13}C-enrichment yet shared the same pattern of auxotrophy. The strongly enriched phylobins encoded several endoglucanases and bacteriocins, while the equally sized weakly-enriched phylobins lacked these capabilities. These data suggest that the strongly labeled cellulolytic strain is degrading ^{13}C-cellulose extracellularly and the weakly ^{13}C-enriched strain can access degradation products as well as other sources of unlabeled carbon present in soil. The capacity of *Sphingomonas* species to degrade cellulose through the activity of extracellular enzymes is known [106,107].

The role of surface ecology in decomposition

Several major populations of microbes that accessed ^{13}C from cellulose were capable of surface-adherence and/or surface-motility. Genes encoding surface attachment were present in phylobins, or have been previously reported, in *Rhizobiaceae* (*Ensifer/ Sinorhizobium*, *Rhizobium* and *Agrobacterium*), *Hyphomicrobiaceae* (*Devosia*), *Sphingomonadaceae* (*Sphingomonas*) and *Caulobacteraceae*

(*Asticcacaulis, Brevundimonas* and *Caulobacter*), as well as in *Pseudoxanthomonas* and *Planctomycetaceae* (*Planctomyces* and *Rhodopirellula*) [108–110]. Each of these genera, except for those in *Planctomyceteceae*, are represented by isolates capable of degrading cellulose [111–118]. For these organisms, attachment would provide preferential access to the by-products of cellulose degradation. This phenomenon is exemplified by the abundance of sugar transporters located on the stalk used by *Caulobacter* to adhere to surfaces [119,120]. Attachment may also facilitate cooperation to crowd out competitors from accessing resources, as observed in the social behavior of *Caulobacter* during xylan degradation (D'Souza *et al*., *bioRxiv pre-print available soon*) or in the coordination of extracellular degradative processes by surface-gliding bacteria *Herpetosiphon* and *Sorangium* [121,122]. Social interactions and cell aggregation density were critical determinants of the rate and efficiency of decomposition of particulate carbon [18]. The dynamics of surface attachment have ramifications for ecology and evolution as well as biogeochemical cycling, which have yet been studied outside of the rumen [123,124].

Diversity at the sub-genus level in the cellulose economy

Shotgun metagenomics provided a comprehensive view of the cellulolytic consortium but was ineffective at resolving the genomes of closely related species. Phylobins were comprised of large pangenomes which limited our ability to test for adaptive gene loss among closely related species, known to be important in the evolution of metabolic dependencies [20,125]. The recovery of large single-genus phylobins for *Myxococcales* (*Sorangium*), *Cellvibrionales* (*Cellvibrio*), *Planctomycetales* (*Planctomyces*) and *Micrococcales* (*Microbacterium*), provided evidence of sizeable pangenomic genetic diversity which could reflect niche partitioning among close relatives. However, the degree of ^{13}C-enrichment within these single-genus phylobins did not differ, except for *Planctomycetales* and *Sphingomonadales* (*i.e.* 'weak' versus 'strong' phylobins). We conclude that few differences in the capacity to access cellulosic carbon had occurred among closely related populations.

Conclusions

We used metagenomic-SIP and metaproteomics to evaluate the traits of microorganisms accessing C from cellulose in an agricultural soil, which we could group into three major classes. These classes included self-sufficient cellulolytic bacteria and fungi (*e.g. Cellvibrio* and *Chaetomium*) that sought to restrict access via competitive exclusion, and other more integrated cellulolytic bacteria (*e.g. Actinobacteria* and *Herpetosiphon*) whose fitness depended on the metabolic byproducts of the community. A third class of non-cellulolytic taxa that accessed cellulosic C (*e.g. Planctomycetes*, *Vampirovibrio, Verrucomicrobia*) were characterized by dependency on community resources as well as mutualistic, opportunistic and parasitic (MOP) interactions, which have yet been fully described due to challenges in cultivability [101]. Our framework facilitates bottom-up measurement of the quantity and quality of each classes' contributions to carbon cycling. For example, the activity of independent degraders likely follows a more idealized pattern of growth and decomposition and be simpler to model. In contrast, the effects of interdependent degraders or MOP will require targeted experiments and more

specialized modeling. A better understanding of the relative effect of each class, and conditions where their contributions are greatest, is now possible through the targeted study of the taxa we have identified. Our findings emphasize that physiological traits and ecological interactions with non-cellulolytic taxa affect the degradation and fate of cellulosic carbon in soil and highlight the range of evolutionary adaptations that constitute the cellulose economy.

Methods

Sample Description and Recovery of ^{13}C-enriched DNA

DNA-SIP was performed using an agricultural soil incubated with ^{13}C-labelled cellulose for 30 days [11]. In brief, microcosms were prepared with soil from a tilled agricultural field under organic management in Penn Yan, New York, as previously described [33]. Samples were sieved (2 mm) and homogenized and pre-incubated for two weeks prior to initiation of the experiment. After soil respiration normalized, an amendment designed to approximate the composition of plant biomass was added. By weight, the mixture was comprised of 38% ^{13}C-labeled bacterial cellulose (99 atom % ^{13}C), 23% lignin alkali, 20% xylose, 3% arabinose, 1% galactose, 1% glucose, 0.5% mannose, 10.6% amino acids, and 2.9% Murashige Skoog basal salt mixture [11]. The amendment was added to soil at 2.9 mg C g^{-1} soil dry weight. After incubation, extracted DNA was subjected to CsCl density gradient centrifugation and fractionated into thirty-five 100-uL aliquots. Shotgun metagenomes were prepared from eight gradient fractions, starting at a buoyant density (BD) of 1.749 g ml^{-1} (F_6) and continuing to a BD of 1.717 g ml^{-1} (F_{13}). A schematic overview of the methods used in this study is presented in Figure S1. The 16S rRNA gene and ITS1 region amplicon data from this DNA-SIP experiment [11] and a sibling study [34] are available at the NCBI under BioProjects PRJNA317227 and PRJNA589050, respectively.

DNA and Peptide Sequencing

Shotgun metagenomes were generated by multiplexing DNA from each gradient fraction using the Nextera XT library preparation kit, then sequenced using three lanes of Illumina HiSeq 2500 (150-bp, paired-end). A subsequent round of sequencing was performed on each gradient fraction using a single lane of MiSeq (250-bp, paired-end) using a library prepared with the Illumina Nextera XT DNA Library Prep Kit (Product number: FC-131-1024, Illumina). The raw sequencing data is archived in the European Nucleotide Archive (BioProject: PRJEB23737). A full description of protein extraction, purification, digestion, mass spectroscopy, and peptide annotation are available in the Supplementary Methods. In brief, protein was extracted from soil with the NoviPure Soil Protein Kit (QIAGEN), initially separated and massed using a Waters nano-Acquity M-Class dual pumping UPLC system (Milford, MA) and a Q-Exactive HF mass spectrometer (Thermo Scientific, San Jose, CA). Twenty-four fractions were subsequently submitted for LC-MS/MS analysis using an LTQ Orbitrap Velos mass spectrometer (ThermoFisher, Waltham MA). Peptides were identified from LC-MS/MS data using predicted protein sequences from the metagenome and filtered with a false discovery rate cut-off of 1%.

Assembly and Classification of SSU RNA genes

Partial 16S and 18S rRNA gene fragments were identified in unassembled reads to estimate relative abundances. Fragments were identified using infernal [35] (v. 1.1.2) and assigned taxonomy using the mothur implementation of the RDP Classifier [36,37] with the Silva database (silva.nr_v128) as reference [38]. Full-length 16S or 18S rRNA genes were assembled using MATAM [39] also using the Silva database. We manually recovered a full-length 16S rRNA gene for *Vampirovibrio*, which were prevalent in SSU fragments, but not assembled by MATAM (details in Supplementary Methods).

Shotgun Metagenome Assembly

Metagenomes for each gradient fraction were composited and assembled using an iterative process to maximize assembly quality (Figure S1; details in Supplementary Methods). In brief, an initial *de novo* assembly was performed using megahit (v1.1.1-2-g02102e1) [40]. Contigs shorter than 2,500 bp were discarded (~7% of total). Contigs were then classified by the Lowest Common Ancestor (LCA) algorithm implemented by MEGAN (v. 6) [41] based on DIAMOND BLASTX searches [42] against the NCBI 'nr' database (downloaded Feb. 3rd, 2017). To improve assembly, two additional assemblies were performed on read sets with reduced sequence diversity. This reduction was achieved by segregating unassembled, quality-processed reads by mapping to (i) the LCA taxonomy of the initial assembly, at rank Order, and ii) to publicly available genomes represented in the full-length 16S rRNA gene library (Table S1). All assemblies were then merged using MeGAMerge [43] with the latest version of MUMMer [44] (v.4beta) designed for large datasets. Merging improved assembly statistics as determined by QUAST [45], increasing N_{50} from 4,407 to 5,419 (Table S2).

Designating ^{13}C-enrichment of Contigs with Gradient-resolved SIP

The relative abundance of every contig across the density gradient (a 'gradient profile') was determined by calculating average read depth using 'jgi summarize bam contig depths' from MetaBAT [46] (v. 2.12.1). The gradient profile of each contig was also simulated with natural abundance of ^{13}C (~1.1 atom % C) to control for variation in GC content using methods outlined in [32,47]. A Random Forest regression model was used to assign a categorical degree of ^{13}C-enrichment for each contig, namely 'strongly' and 'weakly' enriched, 'unenriched' and 'bimodal' (*i.e.* local maxima in both heavy and light portions of the gradient), and 'undetermined' (examples in Figure S2). The following features were used to build the model: the number of local maxima and minima (and the fraction in which they occurred) and the average read depth in each fraction for observed and simulated gradient profiles. Data from 600 manually curated contigs were used to train the model which was implemented in the R package 'caret' [48]. Model validation was performed on 20% of the training set (see R code in the Supplementary Data).

Genome Binning

Common tools for reconstructing MAGs, based on kmer frequency and covariance (in our case across the CsCl gradient), were prone to cross-contamination (see Supplementary Methods). In addition, MAGs

constructed using standard practices failed to recover genomes from taxa known to be abundant in the metagenome and ^{13}C-labeled, including *Chaetomium*, *Vampirovibrio* and members of *Verrucomicrobia* and *Chloroflexales*. Given these limitations, we opted to define a genomic unit based on ^{13}C-enrichment and LCA classification of contigs, which we term a 'phylobin.' Phylobins consisted of contig sets divided by ^{13}C-enrichment status (i.e. 'strong', 'weak' and 'unenriched') and by the taxonomic rank at the level of Order (e.g. 'strongly enriched *Cellvibrionales*'). We justify this approach accordingly: (i) DNA-SIP selectively enriched for a relatively narrow subset of taxa within a given Order; (ii) phylogenetically related organisms with similar enrichment status are likely to share similar genomic and ecological traits. There is no universally appropriate taxonomic rank or phylogenic depth for grouping organisms, since functional traits are conserved at various phylogenetic depths [49]. We chose the rank of Order as cutoff because LCA often fails to accurately classify to the species level taxa that are poorly represented in the NCBI 'nr' database. Hence, aggregating at the rank of Order decreases the risk of losing genomic information. Prior research has shown that aggregating microbiome data by taxonomic Order produced the greatest discriminating power of relevant soil microbial processes [50]. The loss of resolution of individual genomes was compensated for by performing all analyses in parallel on reference genomes chosen based on the similarity of full-length SSU rRNA genes recovered in our study ($\mu_{similarity}$ = 98%, n = 89) or, in some cases, by the only available representative genome for that genus or clade (n = 38; Table S1).

Functional Gene Annotation

Functional genes were annotated using curated databases relating to genes for motility, adhesion, secondary metabolite biosynthetic gene clusters (SMs), and catabolic enzymes for biomass and cellulose. SMs were annotated using the defaults settings of AntiSMASH [51] (v. 4.1.0). Genes involved in cellulolytic activity, namely glycosyl hydrolases (GH), endoglucanases (specific GH families), carbohydrate-binding modules (CBM), polysaccharide lyases (PL), and auxiliary activity enzymes (AA), were annotated using DIAMOND BLASTX searches against a local version of the CAZy database [52] (downloaded Dec. 20th, 2017). A complete list of GH families deemed to be endoglucanases can be found in Supplementary Methods. Chitinases were represented by CAZy families GH18 and GH19. Genes encoding nucleases, adhesion (curli and holdfast proteins) and motility were annotated using DIAMOND BLASTX searches against a local version of the NCBI COG database [53] (downloaded May 1st, 2018), and, in the case of motility, mapped to KEGG biosynthetic pathways for synthesizing complete motility apparatus (Table S3). Genes encoding peptidases were annotated using DIAMOND BLASTX searches against a local version of the MEROPS database [54] (downloaded July 1st, 2018). The capacity for gliding motility was assessed using canonical genes from three model organisms: the focal adhesion protein in *Myxococcus xanthus* [55] (AglZ); the SprB and RemA adhesins in *Flavobacterium johnsonia* [56,57] and Gli349 and Gli521 in *Mycoplasma mobile* [58,59]. Additional adhesion gene families were annotated using compilations of well-characterized proteins, including unipolar polysaccharide synthesis proteins (*upp*) [60] and tight adherence proteins (*tad*) [61]. All annotations were based on a sequence identity cutoff of ≥ 60% across 90% of the full-length gene.

Auxotrophies were determined for each representative genome and phylobin based on ‘genome-enabled metabolic models’ (GEM) in KBase [62] according to [63]. Briefly, flux balance analysis was performed on GEMs under two growth conditions: on a rich media containing all potential biomass precursors and on a minimal media containing only C and essential nutrients. The number of critical enzyme-catalyzed reactions were calculated for each GEM according to the following criteria: (i) the reaction was not involved in central C metabolism, (ii) was essential and carried flux only under minimal (i.e. not under rich) media conditions, and (iii) whose flux was coupled to the production of an essential compound. A genome was considered auxotrophic for a compound if the number of its critical reactions for its biosynthesis was below a compound-specific threshold or if the number of gap-filled critical reactions exceeded a compound-specific threshold. Thresholds were set based on auxotrophy profiles from a dozen well-characterized bacteria in the *Bacteroidetes*, *Firmicutes*, *Alphaproteobacteria*, and *Gammaproteobacteria*.

Statistical Analyses

Statistics were performed in R v. 3.4.2 [64] with the following packages: reshape2, ggplot2, plyr [65–67], Hmisc [68] and phyloseq [69]. Non-parametric multidimensional scaling (NMDS) was performed using ‘metaMDS’ from the R package ‘vegan’ [70]. The relative amount of variation in the primary and secondary NMDS axes explained by functional traits was calculated using the R package ‘relaimpo’ [71]. Pairwise multiple comparisons based on the Kruskal-Wallis test (‘kruskalmc’) were performed using the R package ‘pgirmess’ [72]. Minimum generation times were predicted for all phylobins and representative genomes using growthpred [73] (v. 1.07) based on codon-usage bias using ribosomal genes (identified by COG ID) as the set of highly expressed genes. All analyses can be reproduced using R scripts and data available in the Supplementary Data package.

List Of Abbreviations

AA: auxiliary activity enzyme

C: carbon

CAZy: carbohydrate active enzyme

CBM: carbohydrate-binding module

G+C: guanine and cytosine

GH: glycosyl hydrolase

MAG: metagenome-assembled genome

MOP: mutualist, opportunist or parasite. List of abbreviations

NMDS: non-metric multidimensional scale

PL: polysaccharide lyase

PhyBin: phylobin

rcpm: read count per million

Rep: representative genome

SIP: stable isotope probing

SM: secondary metabolite

SSU: small subunit of the ribosome

Declarations

Author contributions

RW performed all data analysis, research and writing. CP performed metagenomic sequencing, commented on the manuscript and provided the basic code for producing Fig. 1. PW performed predictions of prototrophy/auxotrophy. ML performed metaproteomics. DB guided the formulation of research questions and research efforts and made significant writing contributions.

Availability of data and materials

All analyses can be reproduced using R scripts and data available in the Supplementary Data package. The following data is publicly available:

Supplementary Data	Open Science Framework	doi: 10.17605/OSF.IO/TB3N4
Shotgun metagenomes	European Nucleotide Archive	PRJEB23737
16S rRNA amplicon data	European Nucleotide Archive	PRJNA317227
ITS amplicon data	European Nucleotide Archive	PRJNA589050

Funding

This work was supported by the U.S. Department of Energy, Office of Biological & Environmental Research Genomic Science Program under award numbers DE-SC0016364 and DE-SC0004486.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

Not applicable

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

References

1. Wohl DL, Abora S, Gladstone JR. Ecosystem Function in a Closed and Constant Environment. Ecology. 2004;85:1534–40.
2. Van Dyk JS, Pletschke BI. A review of lignocellulose bioconversion using enzymatic hydrolysis and synergistic cooperation between enzymes-Factors affecting enzymes, conversion and synergy. Biotechnol Adv [Internet]. Elsevier Inc.; 2012;30:1458–80. Available from: http://dx.doi.org/10.1016/j.biotechadv.2012.03.002
3. Shoseyov O, Shani Z, Levy I. Carbohydrate Binding Modules: Biochemical Properties and Novel Applications. Microbiol Mol Biol Rev. 2006;70:283–95.
4. Lynd LR, Weimer PJ, Zyl WH Van, Pretorius IS. Microbial Cellulose Utilization: Fundamentals and Biotechnology. Microbiol Mol Biol Rev. 2002;66:506–77.
5. Schwarze FWMR. Wood decay under the microscope. Fungal Biol Rev. 2007;21:133–70.
6. Berg B, van Hofsten B, Pettersson G. Electronmicroscopic Observations on the Degradation of Cellulose Fibres by *Cellvibrio fulvus* and *Sporocytophaga myxococcoides*. J Appl Bacteriol. 1972;35:215–9.
7. Bayer E a, Kenig R, Lamed R. Adherence of *Clostridium thermocellum* to cellulose. J Bacteriol. 1983;156:818–27.
8. Doi RH, Kosugi A. Cellulosomes: Plant-cell-wall-degrading enzyme complexes. Nat Rev Microbiol. 2004;2:541–51.
9. Eichorst SA, Kuske CR. Identification of cellulose-responsive bacterial and fungal communities in geographically and edaphically different soils by using stable isotope probing. Appl Environ Microbiol. 2012;78:2316–27.
10. Schellenberger S, Kolb S, Drake HL. Metabolic responses of novel cellulolytic and saccharolytic agricultural soil Bacteria to oxygen. Environ Microbiol. 2010;12:845–61.
11. Pepe-Ranney C, Campbell AN, Koechli CN, Berthrong S, Buckley DH. Unearthing the ecology of soil microorganisms using a high resolution DNA-SIP approach to explore cellulose and xylose metabolism in soil. Front Microbiol. 2016;7:1–17.
12. Y. Verastegui, J. Cheng KE. Multisubstrate Isotope Labeling and Metagenomic Analysis of Active Soil Bacterial Communities. MBio. 2014;5:1–12.

13. Wang X, Sharp CE, Jones GM, Grasby SE, Brady AL, Dunfield PF. Stable-isotope probing identifies uncultured *Planctomycetes* as primary degraders of a complex heteropolysaccharide in soil. Appl Environ Microbiol. 2015;81:4607–15.
14. Wilhelm RC, Singh R, Eltis LD, Mohn WW. Bacterial contributions to delignification and lignocellulose degradation in forest soils with metagenomic and quantitative stable isotope probing. ISME J. 2019;13:413–429.
15. Romaní AM, Fischer H, Mille-Lindblom C, Tranvik LJ. Interactions of bacteria and fungi on decomposing litter: Differential extracellular enzyme activities. Ecology. 2006;87:2559–69.
16. Folman LB, Klein Gunnewiek PJA, Boddy L, De Boer W. Impact of white-rot fungi on numbers and community composition of bacteria colonizing beech wood from forest soil. FEMS Microbiol Ecol. 2008;63:181–91.
17. Book AJ, Lewin GR, Mcdonald BR, Takasuka TE, Fox G, Currie CR. Evolution of High Cellulolytic Activity in Symbiotic *Streptomyces* through Selection of Expanded Gene Content and Coordinated Gene Expression. 2016;1–21.
18. Ebrahimi A, Schwartzman J, Cordero OX. Cooperation and spatial self-organization determine rate and efficiency of particulate organic matter degradation in marine bacteria. Proc Nat Acad Sci U S A. 2019;116:23309–16.
19. Berlemont R, Martiny AC. Phylogenetic distribution of potential cellulases in bacteria. Appl Environ Microbiol. 2013;79:1545–54.
20. Morris JJ, Lenski RE, Zinser ER. The Black Queen Hypothesis : Evolution of Dependencies through Adaptative Gene Loss. MBio. 2012;3:1–7.
21. Ng TK, Zeikus JG. Ethanol Production by Thermophilic Bacteria : Fermentation of Cellulosic Substrates by Cocultures of *Clostridium thermocellum* and *Clostridium thermohydrosulfuricum*. 1981;41:1337–43.
22. Jiao Y, Navid A, Stewart BJ, Mckinlay JB, Thelen MP, Pett-ridge J. Syntrophic metabolism of a co-culture containing *Clostridium cellulolyticum* and *Rhodopseudomonas palustris* for hydrogen production. Int J Hydrogen Energy [Internet]. Elsevier Ltd; 2012;37:11719–26. Available from: http://dx.doi.org/10.1016/j.ijhydene.2012.05.100
23. Cavedon K, Canale-Parola E. Physiological interactions between a mesophilic cellulolytic *Clostridium* and a non-cellulolytic bacterium. FEMS Microbiol Lett. 1992;86:237–45.
24. Mori Y. Characterization of symbiotic coculture of *Clostridium thermohydrosulfuricum* YM3 and *Clostridium thermocellum* YM4. Appl Environ Microbiol. 1990;56:37–42.
25. Kato S, Haruta S, Cui ZJ, Ishii M, Igarashi Y. Effective cellulose degradation by a mixed-culture system composed of a cellulolytic *Clostridium* and aerobic non-cellulolytic bacteria. 2018;51:133–42.
26. D'Souza G, Shitut S, Preussger D, Yousif G, Waschina S, Kost C. Ecology and evolution of metabolic cross-feeding interactions in bacteria. Nat Prod Rep. Royal Society of Chemistry; 2018;35:455–88.

27. Enke TN, Datta MS, Schwartzman J, Cermak N, Schmitz D, Barrere J, et al. Modular Assembly of Polysaccharide-Degrading Marine Microbial Communities. Curr Biol. 2019;29:1528-1535.e6.
28. He S, Stevens SLR, Chan LK, Bertilsson S, Glavina del Rio T, Tringe SG, et al. Ecophysiology of Freshwater Verrucomicrobia Inferred from Metagenome-Assembled Genomes. mSphere. 2017;2:e00277-17.
29. Lee STM, Kahn SA, Delmont TO, Shaiber A, Esen özcan C, Hubert NA, et al. Tracking microbial colonization in fecal microbiota transplantation experiments via genome-resolved metagenomics. Microbiome. Microbiome; 2017;5:1–10.
30. Garcia SL, Buck M, McMahon KD, Grossart HP, Eiler A, Warnecke F. Auxotrophy and intrapopulation complementary in the interactome of a cultivated freshwater model community. Mol Ecol. 2015;24:4449–59.
31. Youngblut ND, Barnett SE, Buckley DH. HTSSIP : An R package for analysis of high throughput sequencing data from nucleic acid stable isotope probing (SIP) experiments. PLoS One. 2018;1–8.
32. Barnett SE, Buckley DH. Simulating metagenomic stable isotope probing datasets with MetaSIPSim. bioRxiv. 2019;5–10.
33. Berthrong ST, Buckley DH, Drinkwater LE. Agricultural Management and Labile Carbon Additions Affect Soil Microbial Community Structure and Interact with Carbon and Nitrogen Cycling. 2013;158–70.
34. Koechli C, Campbell AN, Pepe-ranney C, Buckley DH. Assessing fungal contributions to cellulose degradation in soil by using high- throughput stable isotope probing. Soil Biol Biochem [Internet]. Elsevier; 2019;130:150–8. Available from: https://doi.org/10.1016/j.soilbio.2018.12.013
35. Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics. 2013;29:2933–5.
36. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol. 2007;73:5261–7.
37. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol. 2009;75:7537–41.
38. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, et al. SILVA: A comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res. 2007;35:7188–96.
39. Pericard P, Dufresne Y, Couderc L, Blanquart S, Touzet H. MATAM: Reconstruction of phylogenetic marker genes from short sequencing reads in metagenomes. Bioinformatics. 2018;34:585–91.
40. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 2015;1–2.
41. Huson DH, Beier S, Flade I, Górska A, El-Hadidi M, Mitra S, et al. MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. PLoS Comput Biol. 2016;12:1–12.

42. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2014;12:59–60.
43. Scholz M, Lo CC, Chain PSG. Improved assemblies using a source-agnostic pipeline for metagenomic assembly by merging (MeGAMerge) of contigs. Sci Rep. 2014;4:1–9.
44. Kurtz S, Phillippy A, Delcher A, Smoot M, Shumway M, Antonescu C, et al. Versatile and open software for comparing large genomes. Genome Biol. 2004;5:R12.
45. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: Quality assessment tool for genome assemblies. Bioinformatics. 2013;29:1072–5.
46. Kang DD, Froula J, Egan R, Wang Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ [Internet]. 2015;3:e1165. Available from: https://peerj.com/articles/1165
47. Youngblut ND, Buckley DH. Intra-genomic variation in G + C content and its implications for DNA stable isotope probing. Environ Microbiol Rep. 2014;6:767–75.
48. Kuhn M. The caret package. R-package. 2011;
49. Martiny AC, Treseder K, Pusch G. Phylogenetic conservatism of functional traits in microorganisms. ISME J [Internet]. Nature Publishing Group; 2013;7:830–8. Available from: http://dx.doi.org/10.1038/ismej.2012.160
50. Chang H, Haudenshield JS, Bowen CR, Allen R, Iii W, Parnell JJ. Metagenome-Wide Association Study and Machine Learning Prediction of Bulk Soil Microbiome and Crop Productivity. 2017;8:1–11.
51. Blin K, Medema MH, Kottmann R, Lee SY, Weber T. The antiSMASH database, a comprehensive database of microbial secondary metabolite biosynthetic gene clusters. Nucleic Acids Res. 2017;45:D555–9.
52. Cantarel BI, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics. Nucleic Acids Res. 2009;37:233–8.
53. Tatusov RL. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res [Internet]. 2000;28:33–6. Available from: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/28.1.33
54. Rawlings ND, Barrett AJ, Bateman A. MEROPS : the database of proteolytic enzymes , their substrates and inhibitors. 2012;40:343–50.
55. Islam ST, Mignot T. The mysterious nature of bacterial surface (gliding) motility: A focal adhesion-based mechanism in *Myxococcus xanthus*. Semin Cell Dev Biol. Elsevier Ltd; 2015;46:143–54.
56. Shrivastava A, Rhodes RG, Pochiraju S, Nakane D, McBride MJ. *Flavobacterium johnsoniae* RemA is a mobile cell surface lectin involved in gliding. J Bacteriol. 2012;194:3678–88.
57. Nakane D, Sato K, Wada H, McBride MJ, Nakayama K. Helical flow of surface protein required for bacterial gliding motility. Proc Natl Acad Sci [Internet]. 2013;110:11145–50. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.1219753110

58. Uenoyama A, Kusumoto A, Miyata M. Identification of A 349-Kilodalton Protein (Gli349) Responsible for Cytadherence and Glass Binding during Gliding of *Mycoplasma mobile*. J Bacteriol. 2004;186:1537–45.
59. Seto S, Uenoyama A, Miyata M. Identification of a 521-kilodalton protein (Gli521) involved in force generation or force transmission for *Mycoplasma mobile* gliding. J Bacteriol. 2005;187:3502–10.
60. Fritts RK, Lasarre B, Stoner AM, Posto AL, McKinlay JB. A *Rhizobiales*-Specific Unipolar Polysaccharide Adhesin Contributes to *Rhodopseudomonas palustris* Biofilm Formation across Diverse Photoheterotrophic Conditions. Appl Environ Microbiol. 2017;83:1–14.
61. Kachlany SC, Planet PJ, Desalle R, Fine DH, Figurski DH, Planet PJ. *Actinobacillus actinomycetemcomitans*: from plaque to plague to pond scum. 2001;9:429–37.
62. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, et al. KBase : The United States Department of Energy Systems Biology Knowledgebase. Nat Biotechnol. 2018;36.
63. Henry CS, Dejongh M, Best AA, Frybarger PM, Linsay B, Stevens RL. High-throughput generation, optimization and analysis of genome-scale metabolic models. Nat Biotechnol [Internet]. Nature Publishing Group; 2010;28:969–74. Available from: http://dx.doi.org/10.1038/nbt.1672
64. Team. RC. R: a language and environment for statistical computing. R Found Stat Comput. 2017;
65. Wickham H. Elegant Graphics for Data Analysis. Media [Internet]. 2009;35:211. Available from: http://had.co.nz/ggplot2/book
66. Wickham H. Reshaping data with the reshape package. J Stat Soft. 2007;21:1–20.
67. Wickham H. The split-apply-combine strategy for data analysis. J Stat Soft. 2009;40:1–29.
68. FE H, C. D. Hmisc: Harrell miscellaneous. R Packag. 2015;
69. McMurdie PJ, Holmes S. Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLoS One. 2013;8.
70. Oksanen J, Blanchet F, Kindt R, Legendre P, Minchin PR, O'Hara RB, et al. Vegan: community ecology package. R Packag. 2015;
71. U. G. Relative importance for linear regression in R: the package relaimpo. J Stat Softw. 2006;17:1–27.
72. Giraudoux P. pgirmess: Spatial Analysis and Data Mining for Field Ecologists [Internet]. CRAN; 2018. Available from: https://cran.r-project.org/package=pgirmess
73. Vieira-Silva S, Rocha EPC. The systemic imprint of growth and its uses in ecological (meta)genomics. PLoS Genet. 2010;6.
74. Rodriguez-R LM, Konstantinidis KT. Nonpareil: A redundancy-based approach to assess the level of coverage in metagenomic datasets. Bioinformatics. 2014;30:629–35.
75. Kramer S, Dibbern D, Moll J, Huenninghaus M, Koller R, Krueger D, et al. Resource partitioning between bacteria, fungi, and protists in the detritusphere of an agricultural soil. Front Microbiol. 2016;7:1–12.

76. Beloqui A, Nechitaylo TY, Lo N, Polaina J, Strittmatter AW, Reva O, et al. Diversity of Glycosyl Hydrolases from Cellulose-Depleting Communities Enriched from Casts of Two Earthworm Species †. 2010;76:5934–46.
77. Koechli CN. Land Management Affects Microbial Community Composition and Function in Carbon Cycling. Cornell University; 2016.
78. Degrune F, Theodorakopoulos N, Colinet G, Hiel MP, Bodson B, Taminiau B, et al. Temporal dynamics of soil microbial communities below the seedbed under two contrasting tillage regimes. Front Microbiol. 2017;8.
79. Sharma-Poudyal D, Schlatter D, Yin C, Hulbert S, Paulitz T. Long-term no-till: A major driver of fungal communities in dryland wheat cropping systems. PLoS One. 2017;12:1–17.
80. Orgiazzi A, Lumini E, Nilsson RH, Girlanda M, Vizzini A, Bonfante P, et al. Unravelling Soil Fungal Communities from Different Mediterranean Land-Use Backgrounds. 2012;7:1–9.
81. Wilhelm RC, Cardenas E, Leung HTC, Szeitz A, Jensen LD, Mohn WW. Long-Term Enrichment of Stress-Tolerant Cellulolytic Soil Populations following Timber Harvesting Evidenced by Multi-Omic Stable Isotope Probing. Front Microbiol. 2017;8:537.
82. Banerjee S, Kirkby CA, Schmutter D, Bissett A, Kirkegaard JA, Richardson AE. Network analysis reveals functional redundancy and keystone taxa amongst bacterial and fungal communities during organic matter decomposition in an arable soil. Soil Biol Biochem [Internet]. Elsevier Ltd; 2016;97:188–98. Available from: http://dx.doi.org/10.1016/j.soilbio.2016.03.017
83. Tian H, Wang H, Hui X, Wang Z, Drijber RA, Liu J. Changes in soil microbial communities after 10 years of winter wheat cultivation versus fallow in an organic-poor soil in the Loess Plateau of China. PLoS One. 2017;12:1–15.
84. Kearns PJ, Bulseco-McKim AN, Hoyt H, Angell JH, Bowen JL. Nutrient Enrichment Alters Salt Marsh Fungal Communities and Promotes Putative Fungal Denitrifiers. Microb Ecol. Microbial Ecology; 2018;1–12.
85. Vasil'chenko LG, Khromonygina V V., Karapetyan KN, Vasilenko O V., Rabinovich ML. Cellobiose dehydrogenase formation by filamentous fungus *Chaetomium* sp. INBI 2-26(-). J Biotechnol. 2005;119:44–59.
86. Halliwell G, Sweet C. Bacteriocinogenic Activity in the Genus *Cellvibrio*. J Gen Microbiol. 1973;77:363–9.
87. Tathan S, Sibounnavong P, Sibounnavong P, Soytong K, To-anun C. Biological metabolites from *Chaetomium* spp to inhibit *Drechslera oryzae* causing leaf spot of rice. 2012;8:1691–701.
88. Hung PM, Wattanachai P, Kasem S, Poeaim S. Efficacy of *Chaetomium* species as biological control agents against *Phytophthora nicotianae* root rot in citrus. Mycobiology. 2015;43:288–96.
89. Liu L, Zhang J, Cai Z, Huang X, Zhao J. Characterizing the key agents in a disease-suppressed soil managed by reductive soil disinfestation. Appl Environ Microbiol. 2019;1–15.
90. Schulz-Bohm K, Martín-Sánchez L, Garbeva P. Microbial volatiles: Small molecules with an important role in intra- and inter-kingdom interactions. Front Microbiol. 2017;8:1–10.

91. Wagner A. Energy Constraints on the Evolution of Gene Expression. Mol Biol Evol. 2005;22:1365–74.
92. Barton MD, Delneri D, Oliver SG, Rattray M, Bergman CM. Evolutionary Systems Biology of Amino Acid Biosynthetic Cost in Yeast. 2010;5.
93. Jeske O, Jogler M, Petersen J, Sikorski J, Jogler C. From genome mining to phenotypic microarrays: *Planctomycetes* as source for novel bioactive molecules. Antonie van Leeuwenhoek, Int J Gen Mol Microbiol. 2013;104:551–67.
94. Lage OM, Bondoso J. *Planctomycetes* and macroalgae, a striking association. Front Microbiol. 2014;5:1–9.
95. Ward NL, Staley JT, Schmidt JM. *Planctomyces*. Bergey's Man Syst Archaea Bact. 2015. p. 1–23.
96. Wiegand S, Jogler M, Jogler C. On the maverick *Planctomycetes*. FEMS Microbiol Rev. 2018;42:739–60.
97. Erbilgin O, Mcdonald KL, Kerfeld A. Characterization of a Planctomycetal Organelle : a Novel Bacterial Microcompartment for the Aerobic Degradation of Plant Saccharides. 2014;80:2193–205.
98. Reintjes G, Fuchs CAB. Selfish, sharing and scavenging bacteria in the Atlantic Ocean : a biogeographical study of bacterial substrate utilisation. ISME J [Internet]. Springer US; 2019;1119–32. Available from: http://dx.doi.org/10.1038/s41396-018-0326-3
99. Cayrou C, Raoult D, Drancourt M. Broad-spectrum antibiotic resistance of *Planctomycetes* organisms determined by Etest. J Antimicrob Chemother. 2010;65:2119–22.
100. Schlesner H. The Development of Media Suitable for the Microorganisms Morphologically Resembling *Planctomyces* spp., *Pirellula* spp., and other *Planctomycetales* from Various Aquatic Habitats Using Dilute Media. Syst Appl Microbiol [Internet]. Gustav Fischer Verlag, Stuttgart · Jena · New York; 1994;17:135–45. Available from: http://dx.doi.org/10.1016/S0723-2020(11)80042-1
101. Soo RM, Woodcroft BJ, Parks DH, Tyson GW, Hugenholtz P. Back from the dead; the curious tale of the predatory cyanobacterium *Vampirovibrio chlorellavorus*. PeerJ [Internet]. 2015;3:e968. Available from: https://peerj.com/articles/968
102. Gui H, Purahong W, Hyde KD, Xu J, Mortimer PE. The arbuscular mycorrhizal fungus *Funneliformis mosseae* alters bacterial communities in subtropical forest soils during litter decomposition. Front Microbiol. 2017;8:1–11.
103. Wei Y, Zhou H, Zhang J, Zhang L, Geng A, Liu F, et al. Insight into dominant cellulolytic bacteria from two biogas digesters and their glycoside hydrolase genes. PLoS One. 2015;10:1–19.
104. Vandekerckhove TTM, Willems A, Gillis M, Coomans A. Occurrence of novel verrucomicrobial species, endosymbiotic and associated with parthenogenesis in *Xiphinema americanum*-group species (*Nematoda*, *Longidoridae*). Int J Syst Evol Microbiol. 2000;50:2197–205.
105. Warmink JA, Nazir R, Elsas JD Van. Universal and species-specific bacterial ' fungiphiles ' in the mycospheres of different basidiomycetous fungi. 2009;11:300–12.
106. Li D, Feng L, Liu K. Optimization of cold-active CMCase production by psychrotrophic *Sphingomonas* sp . FLX-7 from the cold region of China. Cellulose. Springer Netherlands; 2016;23:1335–47.

107. Zhu L, Si M, Li C, Xin K, Chen C, Shi X, et al. *Sphingomonas gei* sp. nov., isolated from roots of *Geum aleppicum*. Int J Syst Evol Microbiol. 2015;65:1160–6.

108. Rosenberg E. The Prokaryotes: *Alphaproteobacteria* and *Betaproteobacteria*. 4th ed. Delong EF, Lory S, Stackegrandt E, Thompson F, editors. Springer Berlin Heidelberg; 2014.

109. Chen H, Jogler M, Rohde M, Klenk HP, Busse HJ, Tindall BJ, et al. Reclassification and emended description of *Caulobacter leidyi* as *Sphingomonas leidyi* comb. nov., and emendation of the genus *Sphingomonas*. Int J Syst Evol Microbiol. 2012;62:2835–43.

110. Kumar M, Revathi K, Khanna S. Biodegradation of cellulosic and lignocellulosic waste by *Pseudoxanthomonas* sp R-28. Carbohydr Polym. Elsevier Ltd.; 2015;134:761–6.

111. Michaud P, Belaich A, Courtois B, Courtois J. Cloning, sequencing and overexpression of a *Sinorhizobium meliloti* M5N1CS carboxymethyl-cellulase gene. Appl Microbiol Biotechnol. 2002;58:767–71.

112. Yoo S, Weon H, Kim B, Hong S, Kwon S, Cho Y, et al. *Devosia soli* sp. nov ., isolated from greenhouse soil in Korea. 2016;2689–92.

113. Kim S, Gong G, Park TH, Um Y. *Asticcacaulis solisilvae* sp. nov., isolated from forest soil. Int J Syst Evol Microbiol. 2013;63:3829–34.

114. Song N, Cai HY, Yan ZS, Jiang HL. Cellulose degradation by one mesophilic strain *Caulobacter* sp. FMC1 under both aerobic and anaerobic conditions. Bioresour Technol [Internet]. 2013;131:281–7. Available from: http://dx.doi.org/10.1016/j.biortech.2013.01.003

115. Robledo M, Rivera L, Jiménez-zurdo JI, Rivas R, Dazzo F, Velázquez E, et al. Role of *Rhizobium* endoglucanase CelC2 in cellulose biosynthesis and biofilm formation on plant roots and abiotic surfaces. 2012;1–12.

116. Matthysse AG, Thomas DL, White AR. Mechanism of cellulose synthesis in *Agrobacterium tumefaciens*. J Bacteriol. 1995;177:1076–81.

117. Hu X, Yu J, Wang C, Chen H. Cellulolytic bacteria associated with the gut of *Dendroctonus armandi* larvae (*Coleoptera*: *Curculionidae*: *Scolytinae*). Forests. 2014;5:455–65.

118. Talia P, Sede SM, Campos E, Rorig M, Principi D, Tosto D, et al. Biodiversity characterization of cellulolytic bacteria present on native Chaco soil by comparison of ribosomal RNA genes. Res Microbiol [Internet]. Elsevier Masson SAS; 2011;163:221–32. Available from: http://dx.doi.org/10.1016/j.resmic.2011.12.001

119. Ireland MME, Karty JA, Quardokus EM, Reilly JP, Brun Y V. Proteomic analysis of the *Caulobacter crescentus* stalk indicates competence for nutrient uptake. Mol Microbiol [Internet]. 2002;45:1029–41. Available from: http://doi.wiley.com/10.1046/j.1365-2958.2002.03071.x

120. Blanvillain S, Meyer D, Boulanger A, Lautier M, Guynet C, Denancé N, et al. Plant carbohydrate scavenging through TonB-dependent receptors: A feature shared by phytopathogenic and aquatic bacteria. PLoS One. 2007;2.

121. Lee N, Reichenbach H. The Genus Herpetosiphon. 3rd ed. Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E, editors. The Prokaryotes. Springer Berlin Heidelberg; 2006.

122. Schneiker S, Perlova O, Kaiser O, Gerth K, Alici A, Altmeyer MO, et al. Complete genome sequence of the myxobacterium *Sorangium cellulosum*. Nat Biotechnol. 2007;25:1281–9.
123. Shi Y, Weimer PJ. Competition for cellobiose among three predominant ruminal cellulolytic bacteria under substrate-excess and substrate-limited conditions. Appl Environ Microbiol. 1997;63:743–8.
124. Mosoni P, Fonty G, Gouet P. Competition between ruminal cellulolytic bacteria for adhesion to cellulose. Curr Microbiol. 1997;35:44–7.
125. Cordero OX, Polz MF. Explaining microbial genomic diversity in light of evolutionary ecology. Nat Rev Microbiol [Internet]. Nature Publishing Group; 2014;12:263–73. Available from: http://dx.doi.org/10.1038/nrmicro3218

Figures

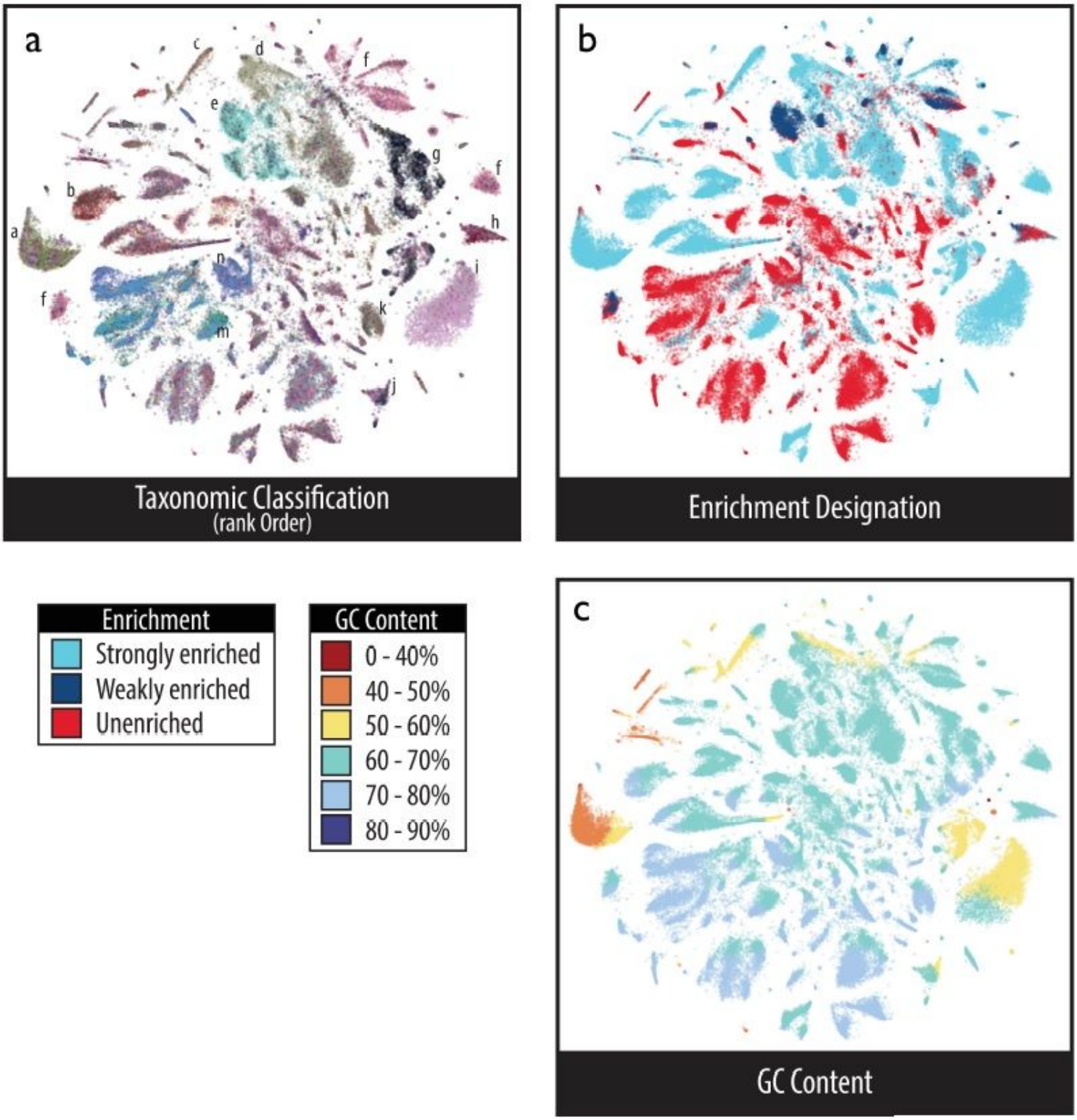

Figure 1

Contigs from the metagenome assembly cluster with respect to (a) LCA taxonomic classification, (b) 13C-enrichment designation, and (c) genome GC content. Contigs (> 3.5 Kbp) were clustered by pentanucleotide frequency using the t-SNE multi-dimension reduction algorithm. Taxonomic groups discussed in the text were labeled as follows: (a) Cellvibrionaceae, (b) Caulobacteraceae, (c) Oxalobacteraceae, (d) Rhizobiaceae, (e) Sphingomonadaceae, (f) Planctomycetaceae, (g)

Xanthomonadaceae, (h) Nannocystaceae, (i) Chaetomiaceae, (j) Verrucomicrobiaceae, (k) Rhodobacteraceae, (m) Pseudonocardiaceae/Streptomycetaceae and (n) Microbacteraceae. An interactive .html version in the Supplementary Data package allows for detailed exploration of taxonomic annotations for all contigs.

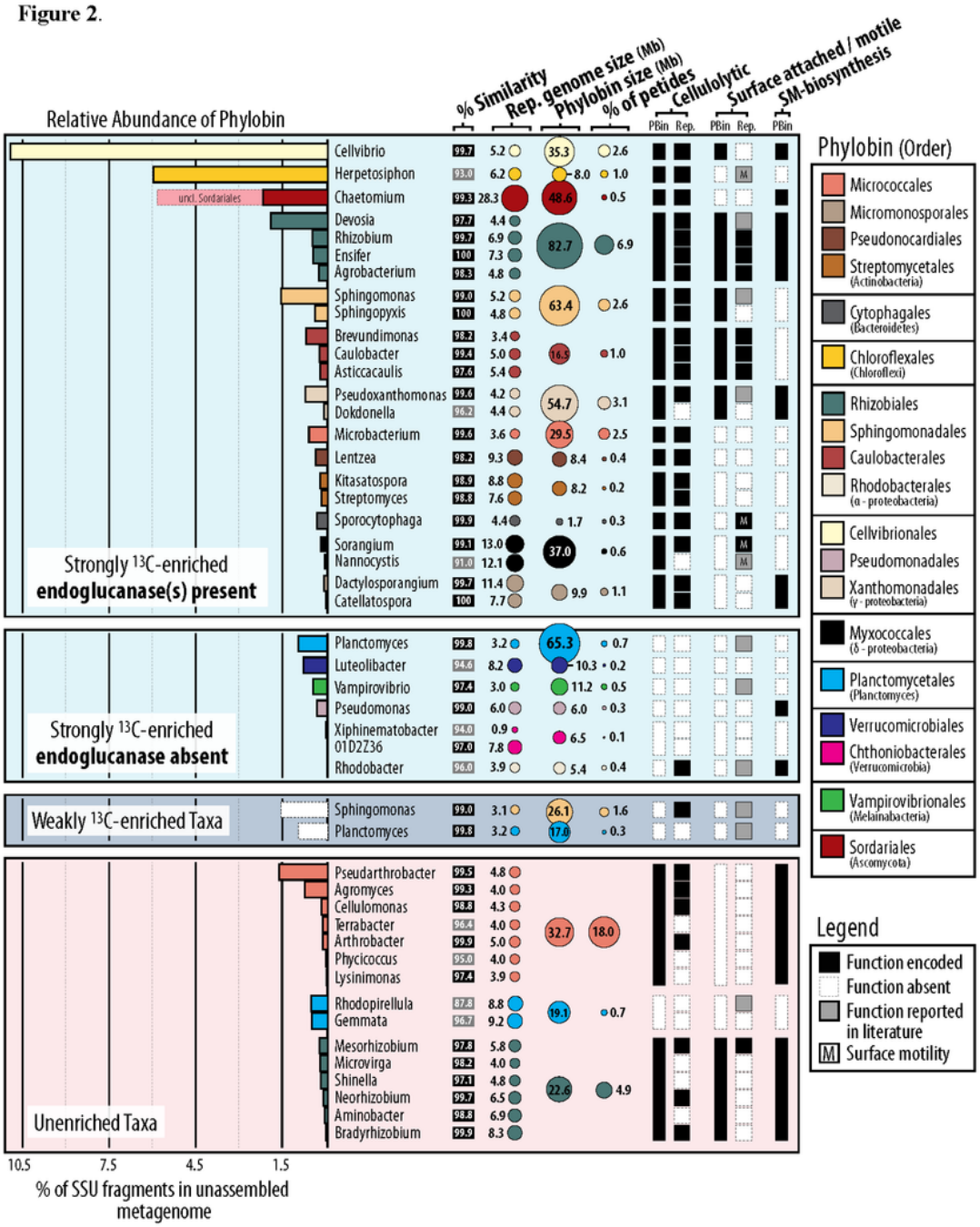

Figure 2

Members of the cellulose-degrading consortium were defined by their taxonomy and functional capabilities encoded in metagenome-assembled phylobins ('PBin') and representative genomes ('Rep'). Phylobins were categorized by their 13C-enrichment and cellulolytic capacity and ranked along the y-axis by the relative abundance SSU rRNA gene fragments recovered (indicated by barplots). Representative genomes were identified according similarity of full-length 16S rRNA gene (column 1) and were grouped with their respective phylobin. Representative genomes with less than 97% similarity to a phylobin 16S rRNA gene were shaded in grey. Several phylobins were comprised of genomes from multiple genera and the size of each (in megabases) and the percentage of peptides assigned to each phylobin are provided. The remaining columns show the presence/absence of genes for endoglucanases or those involved in surface attachment, surface motility (M), and secondary metabolite (SM) production. Boxes were shaded grey if a member of that genus reportedly possesses the ability for attachment or surface motility. Secondary metabolite production was designated if peptides corresponding to antimicrobial gene clusters were detected in the metaproteome. Only the most abundant 'unenriched' phylobins are shown (full overview in Figure S7).

Figure 3.

Order	Classification	GH families	# of Peptides
Sordariales	o_Sordariales	5(5), 5(7), **5(9)**, 5(12), **5(24)**, **5(50)**, 6, 7, 12, 16, 51, 74, **131**	**8 x 131** 3 x 5(5) 7 x 7 2 x 5(7) 6 x 6 2 x 74
Bacteroidetes	p_Bacteroidetes	5(5), 5(7)	1 x 5(5) 1 x 5(7)
Caulobacterales	g_*Asticcacaulis* g_*Caulobacter*	5(4), 5(48) 51 51	
Cellvibrionales	g_*Cellvibrio*	5(4), 5(7), **5(26)**, **5(47)**, 6, 9, 16, 26 51, 74	10 x 9 1 x 26 **2 x 5(47)**
Herpetosiphonales	g_*Herpetosiphon*	5(2), 5(5)	
Cytophagales	f_Cytophagaceae	5(13), 8, 9, 16, 26	2 x 8
Micrococcales	g_*Microbacterium* o_Micrococcales	**5(22)**, 16, 51 5(4), 5(5)	1 x 16
Micromonosporales	g_*Actinoplanes* g_*Micromonospora*	51 5(5)	
Myxococcales	o_Myxococcales	8, 16, 45	1 x 45
Pseudonocardiales	g_*Amycolatopsis*	**5(2)**, 6, 9, 12	
Rhizobiales	g_*Agrobacterium* g_*Rhizobium* g_*Sinorhizobium* o_Rhizobiales	8, 16, 26, 51 5(5), 5(7), 8, 16, 26, 51 4(48), 8, 16, 51 5(13), 51	
Sphingomonadales	g_*Novosphingobium* g_*Sphingomonas* g_*Sphingopyxis*	5(13) 9, 16, 51 51, 74	
Streptomycetales	g_*Streptomyces*	5(4), 5(18), **5(43)**, 5(51), 9 ,51	
Uncl. Actinobacteria	p_Actinobacteria	5(4), 5(5), 5(18), 5(51), 6, 9, 12, **44**, 51	**1 x 44** 1 x 5(4)

Fungi ■ Bacteria □

Figure 3

Diverse glycosyl hydrolase (GH) genes were identified in cellulolytic phylobins, many of which matched peptides detected in the metaproteome (the number and type of each peptide is indicated). Taxonomic classifications and GH family are provided for the endoglucanase gene fragments found within each phylobin (lowest LCA classified to the order (o_), family (f_), or genus (g_) level). Taxon specific

endoglucanase families are indicated in bold font. Five peptides matched to endoglucanase genes not belonging to any phylobins.

Figure 4.

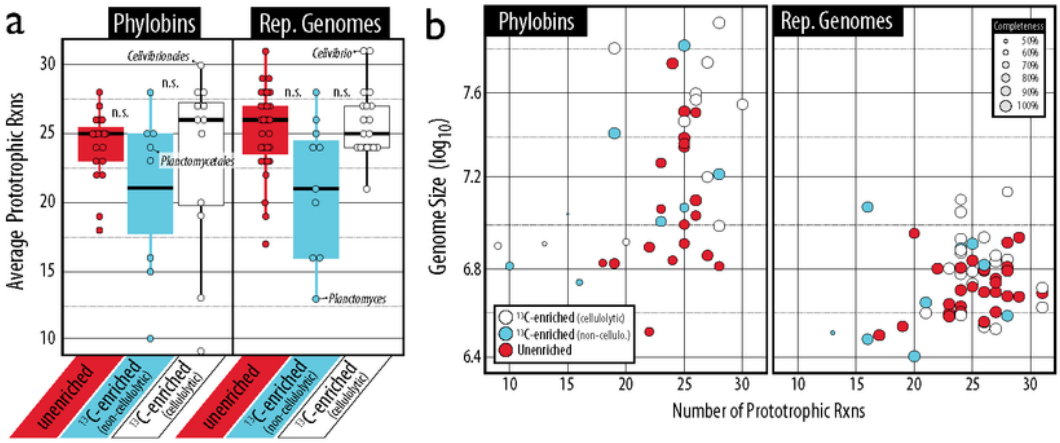

Figure 4

In (a), a comparison of the number of complete biosynthetic pathways (prototrophy) revealed that 13C-enriched non-cellulolytic (blue) phylobins and representative genomes were slightly less prototrophic than those which were either 13C-enriched cellulolytic (white) or unenriched (red). On average, these

differences were not significant (Kruskal-Wallis; phylobins, p = 0.6; Representative genomes, p = 0.2), though major populations within each group (Planctomyces versus Cellvibrio) exhibited consistent differences in accordance with hypotheses. In (b), prototrophy was significantly correlated with genomes size for phylobins (r=0.39; p = 0.01), but not representative genomes (r = 0.14; p = 0.28). A ranking of prototrophy in all phylobins and representative genomes is available in Table S7.

Figure 5.

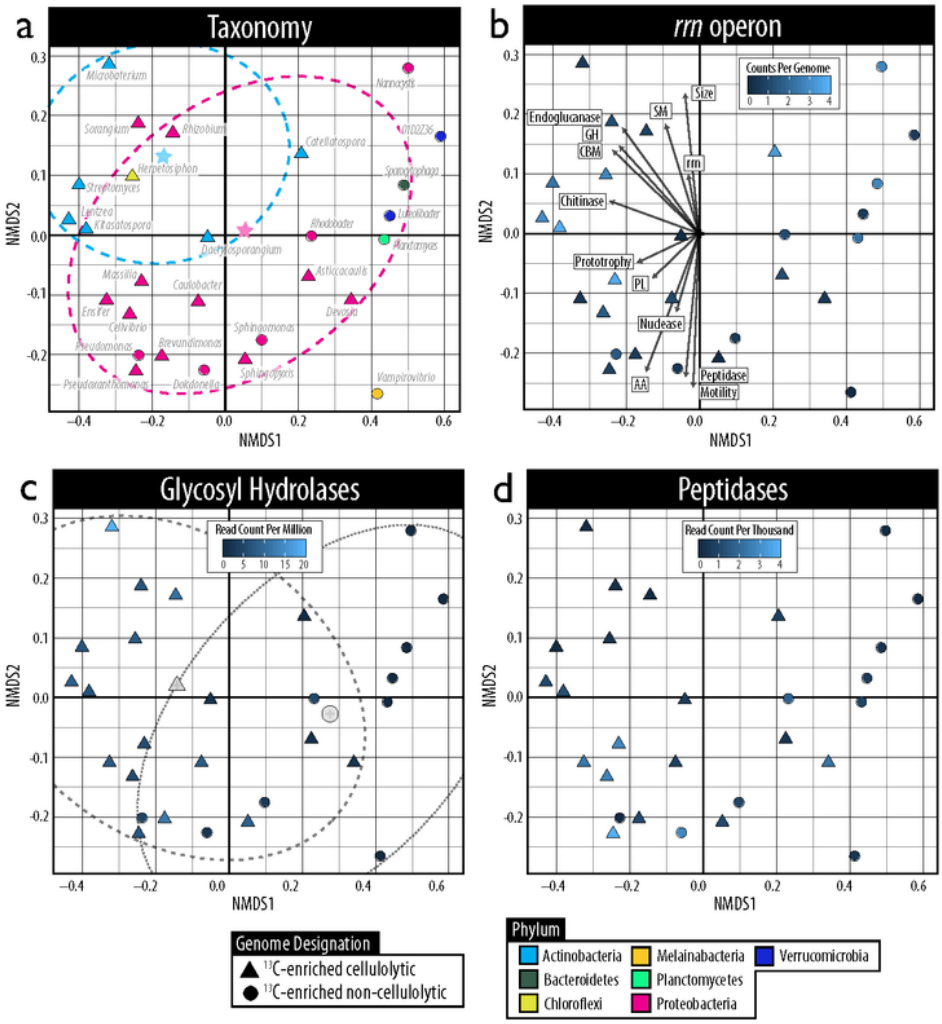

Figure 5

The functional gene content of representative genomes was compared by NMDS using the Bray-Curtis dissimilarity of gene abundances normalized to genome size. Most variation among representative genomes was attributable to carbohydrate active enzymes content (63% of NMDS 1; Table S8). Four panels showing the same ordination were colored according to (a) the taxonomic classification at the phylum level; (b) rrn operon copy number; (c) abundance of glycosyl hydrolases, and (d) peptidase abundance. Genomes formed clusters according to cellulolytic potential (ANOSIM $R = 0.498$, $p < 0.001$) with the centroid of each group displayed in (c). Genomes loosely clustered by phylogenetic differences between Actinobacteria and Proteobacteria, though lacking statistical support (ANOSIM $R = 0.1$, $p = 0.2$) with the centroid (star symbol) for each shown in (a). In (b), functional gene data were fitted to the ordination with arrow length proportional to the correlation between each variable and ordination axes. The following were abbreviated: “GH” (glycosyl hydrolase), “SM” (secondary metabolite gene cluster), “CBM” (carbohydrate-binding module), “rrn” (ribosomal operon), “PL” (polysaccharide lyase), “AA” (auxiliary activity) and “Size” (genome size).

Figure 6.

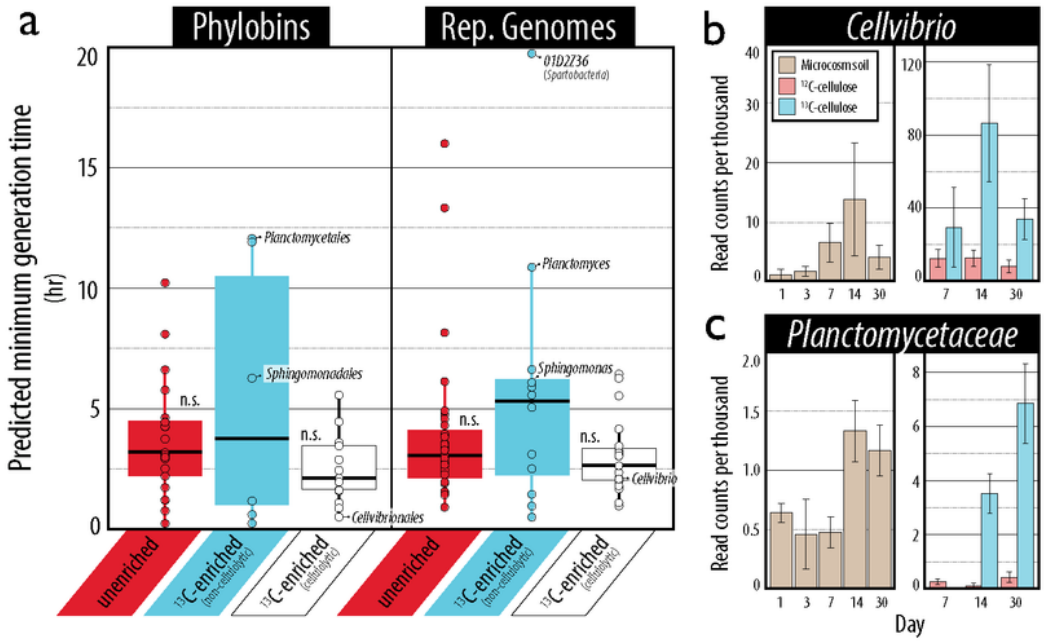

Figure 6

Genome-based predicted generation times of taxa (a) corresponded with temporal patterns in relative abundance in soil microcosms amended with cellulose (bc), where 13C-enriched cellulolytic phylobins and representative genomes tended to have faster growth rates than non-cellulolytic phylobins and their representatives. In (a), the differences in genome-based predictions of generation time were, on average, not significant (Kruskal-Wallis, pphylobin = 0.33 and prep = 0.52). However, the most abundant 13C-

enriched cellulolytic and non-cellulolytic taxa, Cellvibrio and Planctomycetaceae, respectively, exhibited characteristics of faster (b) and slower (c) growth consistent with expectations of the degree of their metabolic dependency. Several other major taxa, including Devosia, Sphingomonas and members of the Verrucomicrobia exhibited similar trends (Figure S6). In (b) and (c), each panel is divided into two datasets: one corresponding to the relative abundance of bacterial populations in whole DNA extract from soil amended with cellulose and the other corresponding to the relative abundance in 13C-enriched DNA pools from soils amended with 12C natural abundance or 13C-labeled cellulose. Error bars correspond to standard error. A complete ranking of predicted generation times for phylobins and representative genomes is available in Table S7.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Wilhelmetal.SupplementaryMethods.docx
- Wilhelmetal.SupplementaryTables.xlsx
- Wilhelmetal.SupplementaryFigures.pdf