Taxonomically and metabolically distinct microbial communities with depth and across a hillslope to riparian zone transect

Adi Lavy, Paula B. Matheus Carnevali, Ray Keren, Markus Bill, Jiamin Wan,

Tetsu K. Tokunaga, Kenneth H. Williams, Susan S. Hubbard, Jillian F. Banfield

Abstract

Watersheds are important for supplying fresh water, the quality of which depends on complex interplay involving physical, chemical and biological processes. As water percolates through the soil and underlying weathering rock en route to the river corridor, microorganisms mediate key geochemical transformations, yet the distribution and functional capacities of subsurface microbial communities remain little understood. Here, we used genome resolved metagenomics to study microbial communities during late summer along a 230 m hillslope meadow to floodplain transect within the mountainous East-River watershed, Colorado. We found very limited strain/species overlap at different depths below the ground surface and at different distances along the hillslope, possibly due to restricted hydraulic connectivity after early stages of snowmelt, and show that communities are largely distinct in their metabolic capacities. Both proximity to the river and to the underlying Mancos shale apparently control species distribution and metabolic potential. Functions such as carbon fixation and selenate reduction were prevalent at multiple sites, although the lineages of organisms responsible tend to be location-specific. For example, selenate reduction within the hillslope is linked to Candidate phyla *Rokubacteria* and *Methylomirabillis*, but only *Deltaproteobacteria* are predicted to have

this capacity at the floodplain. Based on the abundance of genomically-encoded functions that utilize it as an electron donor or acceptor, sulfur is significantly more important for microbial metabolism at the floodplain compared to on the hillslope. Nitrification and methylamine oxidation are likely only occurring within the floodplain, with nitrification capacity in shallow soil and methylamine oxidation in deeper unsaturated sediment. The capacity for nitrogen fixation was mostly associated with *Geobacteraceae* and *Myxococcales* (phylum *Deltaproteobacteria*) at the floodplain. Thus, we conclude that metabolic potential and microbial community composition vary as a function of depth and distance along a mountainous watershed hillslope to floodplain transect, and that the microbiome of the floodplain compartment plays a significant role in nitrogen fixation and sulfur cycling at this site.

Keywords: metagenome, metabolism, soil, watershed, riparian zone, shale

Introduction

Watersheds, and specifically those of mountains, are important providers of freshwater for areas downstream [1]. These environments are composed of a complex array of ecological life zones and geomorphic units, such as forested alpine and montane meadows and floodplains. These are present at a multitude of geomorphic units, such as mountain tops, hillslopes, and the riparian corridor, and each unit, in turn, has specific subsurface compartments of varying saturation, including s soil, weathering zone, and bedrock. Within each subsystem, a multitude of biotic and abiotic interactions occur. Water runoff and infiltration transform and transfer solutes and particles from headwaters to streams and lakes, and from the surface to the subsurface. Currently, there is relatively little known about how hydrobiogeochemical functions are partitioned across ecosystem compartments and the extent to which environmental factors select for distinct microbial communities that mediate biogeochemical transformations and impact water quality and other watershed outputs [2]. Ecosystem gradients are also poorly understood from this perspective, for example how functions change along elevation gradients, with depth below the surface and soil properties. Some prior work examined microbial communities in soils on mountain slopes focusing on the composition of these communities across different climate zones along a transect hundreds of meters long [3–7]. While these studies are important for our understanding of climate effects on microbial ecology, they do not address differences at the spatial resolution of meters and centimeters.

Most studies often focus on shallow soil, neglecting deeper soils and weathering rock in mountain ecosystems. The gradients affecting microbial communities within these deeper zones are influenced by above-ground factors such as precipitation, but are also determined by the geochemistry of the underlying bedrock and groundwater [8, 9]. Here, we investigated a well

studied headwaters mountainous catchment in the East River of Colorado, where intersecting research activities are focused on developing a predictive understanding of geological, hydrobiogeochemical and ecological process interactions [2]. We analyzed genome resolved metagenomes to evaluate the connections between microbial composition, metabolic capacities and the spatial organization of these attributes along a hillslope to riparian zone transect and as a function of depth below ground surface.

Given the vast size of watersheds, one way to address the complexity of ecosystem processes is to use a scale adaptive approach in which repetitive elements, or systems, are studied separately [10]. The scale at which each study is conducted is determined by the hypothesis of the study as well as the methods available for answering it. Following this approach, a “representative” element of hillslope to riparian zone transect was chosen within the Upper East River watershed. From the perspective of vegetation, slope, underlying bedrock and weather conditions, the transect is typical of a substantial fraction of the watershed. The East River watershed in the Upper Colorado River Basin serves as a testing ground for the Department of Energy Watershed Function Project (watershed.lbl.gov) and its associated research community [2]. The watershed is approximately 300 km^2 with an average elevation of 3266 m. It has an average annual temperature ~0°C, with average minimum and maximum temperatures of −9.2°C and 9.8°C, respectively. The bulk of the 600 mm yr^{-1} of average precipitation is in the form of snow, which covers the surface between September and May [2, 11]. In early spring, groundwater reaches the surface, but the lowest level of the water table depends on the location along the hillslope. Before the snow melts, the water table is 3.5 m deep at the top of the hillslope whereas at the floodplain it is only ~0.75 m below the surface (Tokunaga et al., under review). The underlying rock at the study site is Cretaceous Mancos Shale, with carbonate and pyrite contents of roughly 20% and 1%, respectively [12]. Selenium,

which is present in Mancos Shale soils [13], is found at high concentration of 8 ppb in pore waters of unsaturated sediment at the riparian zone whereas sulfate is enriched within groundwater (800 mg/L) at the same location [14].

Microbial communities in soil and sediment play a key role in shaping their surroundings, from rock weathering to nitrogen and carbon fixation [15, 16]. Prior to the current study, we surveyed microbial metabolic genes from metagenomes that were analyzed without assignment to specific genomes to test the possibility that metabolism of microbial communities varied along the hillslope to floodplain transect. The analysis uncovered indications of trends related to carbon and nitrogen fixation and selenium metabolism [14]. These findings motivated the current study, in which the genomes from sample series were comprehensively resolved so that organisms could be confidently linked to functions and the presence of multiple genes required for some functionalities could be verified. Our results reveal that patterns of organism distribution and metabolic potential are shaped by both soil depth and position relative to the riparian zone, and that certain functions are only carried out by distinct organisms at specific sites. We leverage comparative genomic analyses to evaluate overlap of strains across sites, addressing the extent to which landscape position contributes to overall watershed biodiversity and constrains inter-site organism dispersal.

Results:

Overall, 6.5 million scaffolds longer than 1 Kbp were assembled from 41 samples collected from between 5 cm and 200 cm depth and distances of 50 - 280 m from the river (Table S1). Given that microbial communities in soils are diverse and therefore difficult to resolve genomically, it is unsurprising that, on average, only 27.8% (±11) of the reads could be mapped back to these scaffolds. The large fraction of unmapped reads reflects the huge

microbial diversity in soil and the prevalence of organisms at similar very low abundance. Scaffolds were binned into draft genomes, 1448 of which were assessed to be ≥60% complete based on the presence of 43 bacterial single copy genes or 38 archaeal single copy genes. The bins contain 4.05 Gbp in 625,348 scaffolds. Dereplication at 98% average nucleotide identity (ANI) and screening for genomes that are at least 80% complete resulted in 484 near-complete genomes (Table S2). These genomes were classified as high-quality draft (111 genomes) or medium-quality draft (350 genomes) according to Anvio's completeness and redundancy scores and the genome reporting standard [17]. Twenty-three draft genomes had redundancy score higher than 10 and therefore did not match any of the categories in the reporting standard. It should be noted that A'nvio is generating a redundancy value which does not completely correlate with the definition of 'contamination' as used in the genome reporting standard.

Genomes for organisms from 28 phyla and 5 proteobacterial classes were assembled from the hillslope to riparian zone transect. Among the genomes were those of organisms typically detected in soils in 16S rRNA gene surveys [18]. These include *Acidobacteria*, *Actinobacteria*, *Bacteroidetes*, *Planctomycetes*, *Proteobacteria*, and *Verrucomicrobia* (Figure 1). Genomes representing *Woesebacteria*, *Yanofskyabcteria*, and *Wolfebacteria* from the Candidate Phyla Radiation group were reconstructed. These were previously detected only at the floodplain by their rpS3 gene, and most often in samples from 10 cm above the water table [14]. Having genomes of taxa which were previously only detected by a marker gene presents an opportunity to explore their potential metabolism and environmental impacts.

Two clades of *Deltaproteobacteria* displayed a differential preference for either the floodplain or the hillslope. The first clade, present at the floodplain, consists of *Desulfuromonadales* (*Geobacteraceae*), Syntrophobacterales (*Syntrophaceae* and *Syntrophobacteraceae*), *Thermodesulfobacteriales* (*Thermodesulfobacteriaceae*), and unknown

Deltaproteobacteria mostly similar to those genomically described from Rifle, CO. These species were found to be significantly correlated with samples taken from water-saturated sediment and from 10 cm above the water table. The second clade, containing species of Myxococcales (*Cystobacterineae*, *Nannocystineae*, and *Soranglineae*), was present at all parts of the transect other than water-saturated sediment. Actinobacteria, which is routinely found in soils and represented here by 70 genomes, were significantly correlated only with hillslope sites while some genomes were detected at the riparian zone.

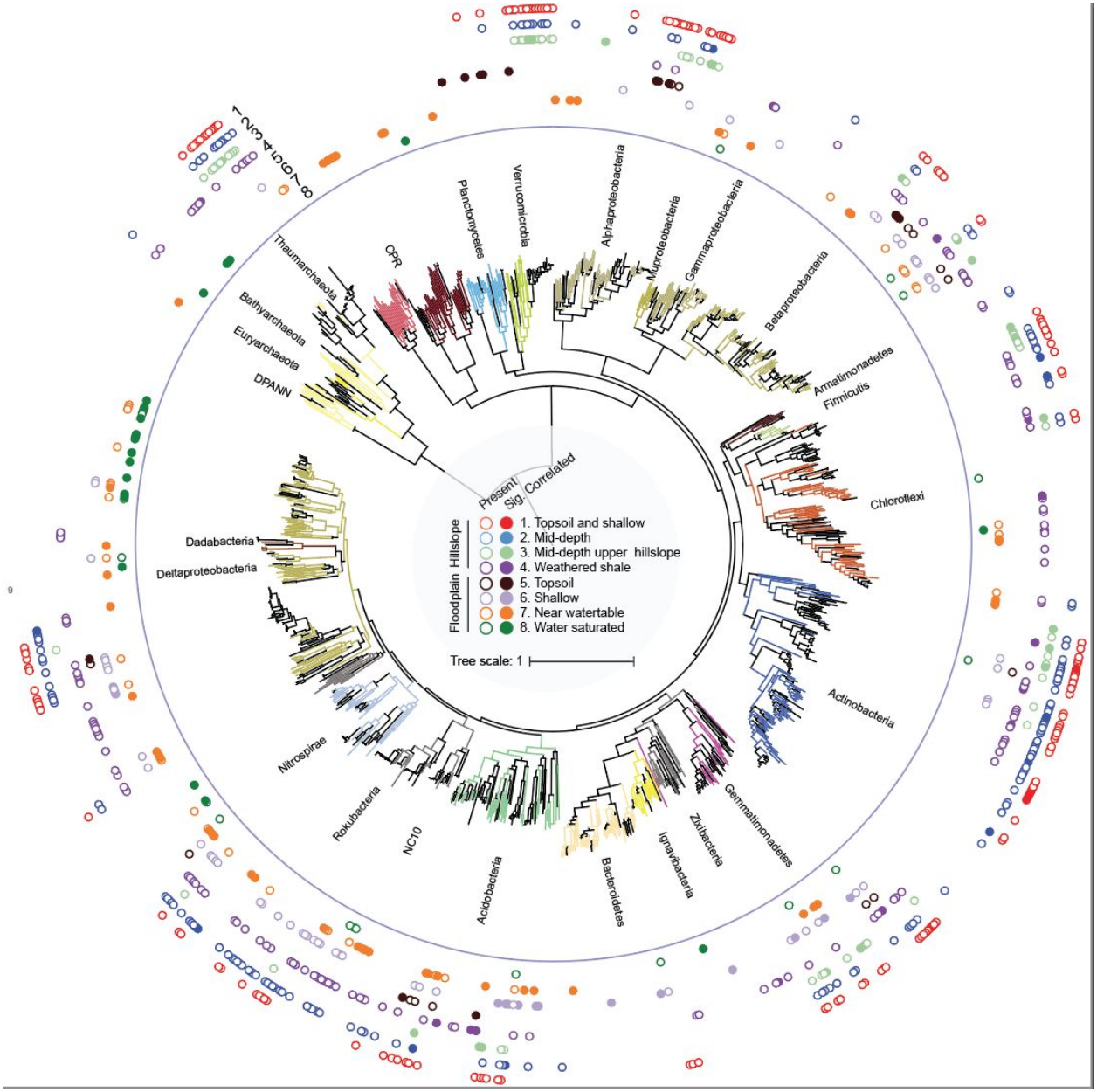

Figure 1 - Correlation of microbial species with ecosystem compartment along hillslope to Riparian-zone transect - A maximum-likelihood tree of 16 Ribosomal Proteins of 484 representative genomes (black lines) and reference genomes (colored lines). The presence of a genome at each location is denoted by an empty circle. A species found to be correlated with a specific sample is marked with a filled circle (IndVal analysis, p-value < 0.05, 9999 permutations, Benjamini-Hochberg corrected).

The presence of strain level, nearly identical genomes across the transect requires relatively recent inter-site transport, probably via downslope subsurface water flow. Conversely, lack of strain sharing could be explained either by limited inter-site dispersal or selection against genotypes adapted to conditions elsewhere on the hillslope. Only seven bacterial and two archaeal strains occurred at more than one sampling location (Figure 2). One strain of *Actinobacteria* (Ac2), was found at two sites that are not adjacent (PLM1 and PLM6), but at a similar depth, 60 and 50 cm, respectively.

Drilling and installation of sampling equipment for another experiment at each of the PLM locations revealed compositionally distinct shale layers with a 20° dip to the southwest, paralleling the prominent lithological feature expressed at the surface at PLM0. Thus, lithological variation in the Mancos shale should be considered when evaluating the effect of depth below the surface on community composition along the hillslope transect.

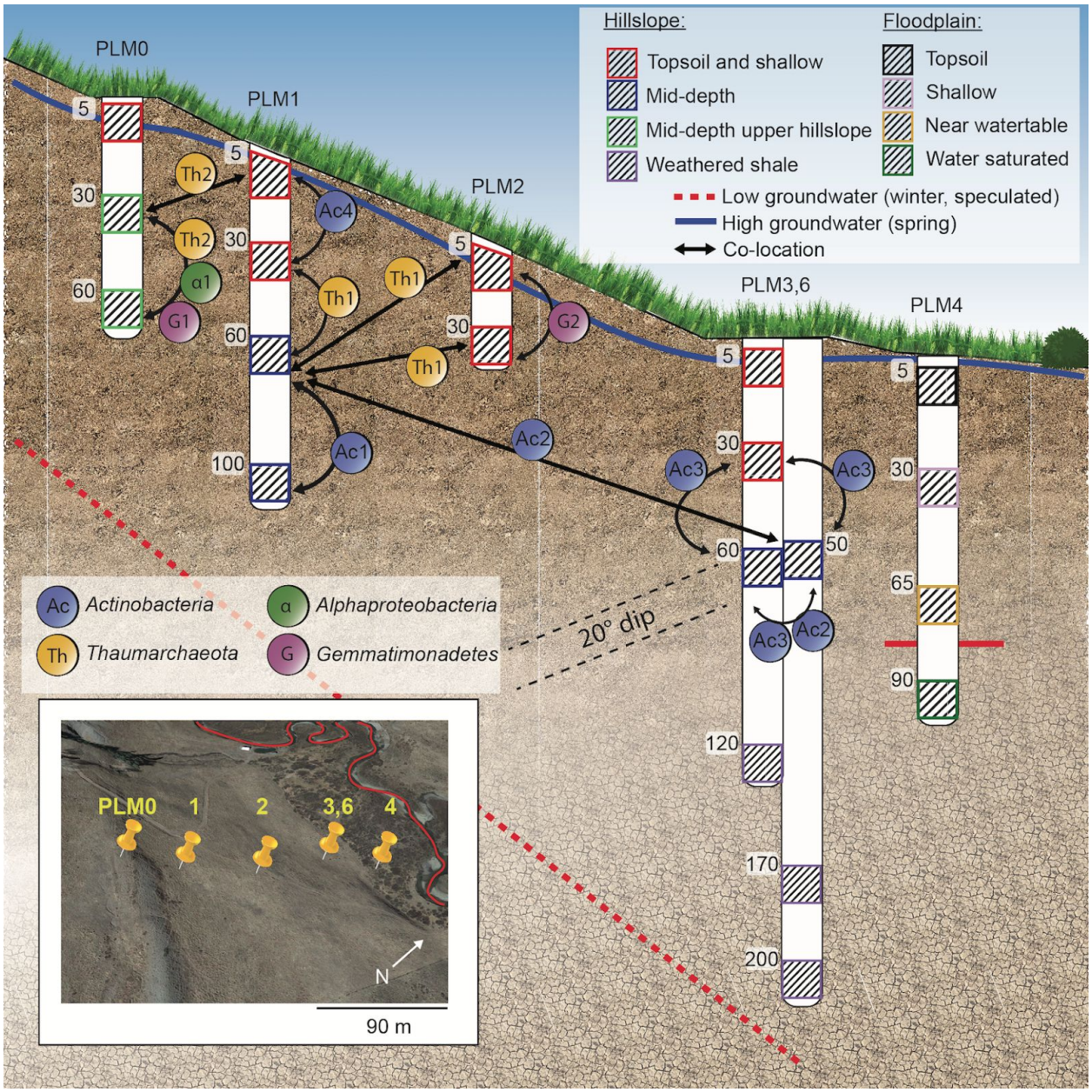

Figure 2 - Sampling scheme, connectivity between sites and samples grouping. Numbers along wells indicate depth below ground surface of sample (in cm). Specific depths that were sampled within each PLM site along the transect are marked with diagonal lines pattern. Nine strains were found to occur at more than one sample. These are from the class *Alphaproteobacteria*, and *Thaumarchaeota*, *Actinobacteria*, and *Gemmatimonadetes* phyla. Identical strains were detected within adjacent depths of a site, or between two adjacent sites. Only Ac2 co-occurred in two sites that are not adjacent (PLM1 and

PLM6) but at similar depths, 60 and 50 cm, respectively. Samples similar in community composition, as was described by Lavy et al. [14], were grouped to form a single zone which is denoted by a colored box. A 20° dip towards the South-East is depicted by dashed lines. An aerial photo of the transect is shown within the inset.

We searched the 484 representative genomes for 23 metabolic functions associated with five types of reactions, namely C1 metabolism, oxidation processes (electron donors), reduction processes (electron acceptors), carbon and nitrogen fixation, and resistance to toxic elements. The genomes were found to harbor 20 out of the 23 metabolic functions. Denitrification, methane oxidation, and respiratory arsenite oxidation were absent from the draft genomes as they lack the required genes that encode for enzymes involved in these pathways. Among all electron acceptors, organisms potentially capable of using oxygen for respiration were the most abundant. However, genomes of organisms that could potentially respire selenate and sulfite had high relative abundance within water-saturated sediment at the floodplain. Organisms with the potential for sulfur respiration were not detected at any of the hillslope sections (Figure 3).

By using genomes rather than individual genes without genomic context, we were able to show that the relative abundance of organisms capable of using sulfate and sulfite as terminal electron acceptors increases with depth at the floodplain, whereas the relative abundance of those capable of oxidizing thiosulfate, sulfur and sulfide diminish (Figure 3). This follows the general trend which shows that the relative abundance of organisms with reductive functions increases with depth while that of oxidative functions decreases. Non-respiratory arsenic resistance, either via redox or methylation or arsenic species is the third most abundant function at 10 cm above groundwater. However, its relative abundance is much lower just 20 cm below the water table, and within weathering shale at the hillslope. Interestingly, we detected

organisms capable of dissimilatory arsenate reduction using the *arr* operon, but could not detect the full sets of genes which make the *arx* operon, encoding for anoxygenic arsenic oxidation. The highest abundance of genes encoding for arsenite methylation detoxification mechanism which may follow arsenate reduction were detected within unsaturated sediments at the floodplain (Figure S1). These genes were also found in every compartment other than shallow soil of the floodplain.

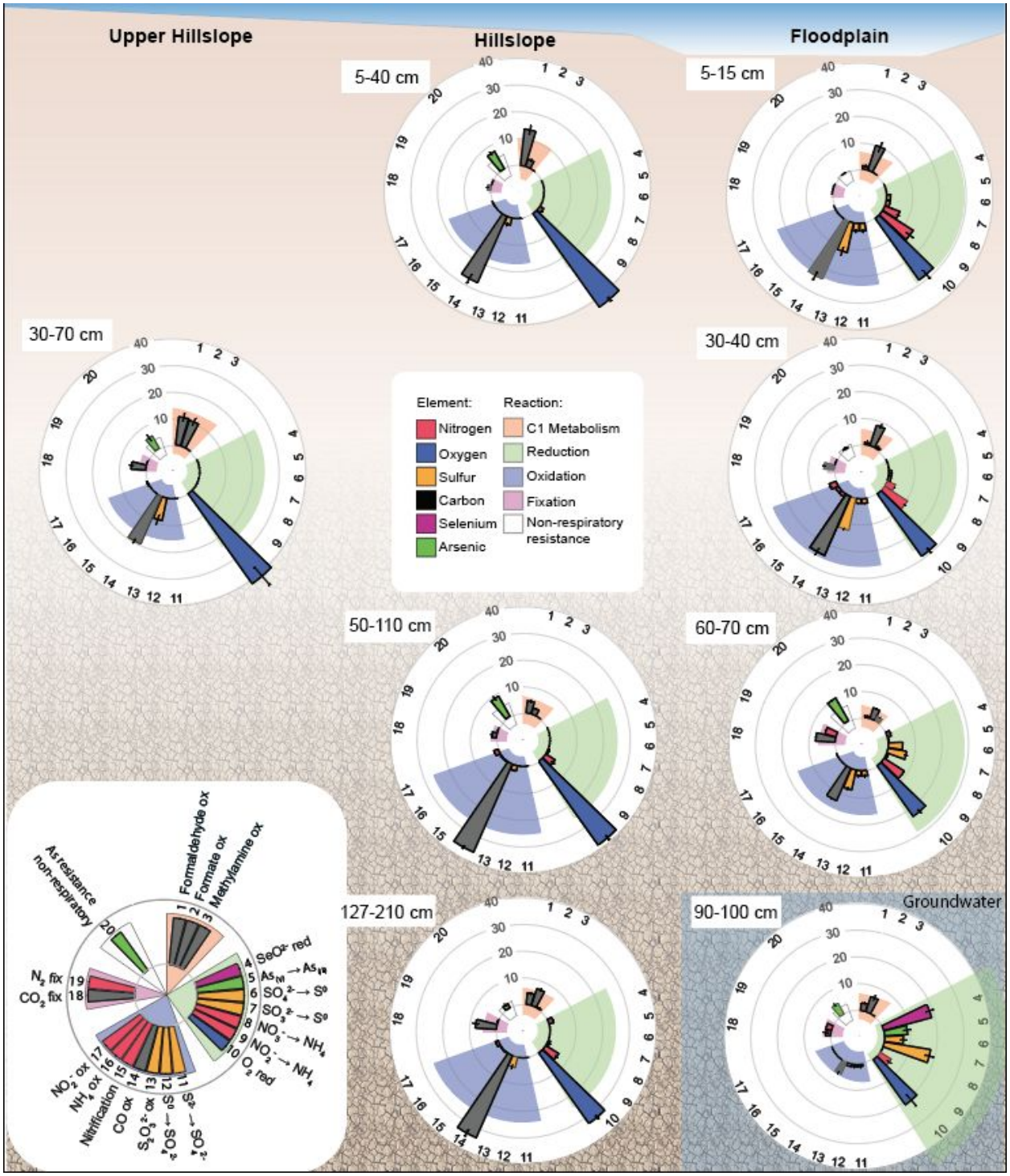

Figure 3 - Relative abundance of metabolic functions at each zone. The metabolic potential of each genome was evaluated using Hidden Markov Model for key enzymes of metabolic function. The length of

the reaction type bar (e.g. C1 metabolism) is correlated with the overall relative abundance of functions within the category and is scaled to a maximum of 60. Bars represent SE.

Applying the indicator species analysis to metabolic functions allows identifying functions that uniquely describe one or multiple sites. Organisms with nitrogen fixation genes were most abundant in samples taken from near groundwater, above and below the water table, but not in other samples. Therefore, nitrogen fixation was found to be significantly representative of these zones (Indval = 0.983, p-value = 0.002) (Figure 4). Organisms capable of oxidizing ammonia to nitrite were abundant at several zones but the process was identified as an indicator process only for floodplain associated shallow samples (Indval = 0.764, p-value = 0.033). Sulfite and sulfur reduction as well as sulfur and sulfide oxidation were indicators of any sample at the floodplain as these functions were absent from the hillslope. While organisms with the potential to fix CO_2 were detected in every sample other than the topsoil of the floodplain, these were most abundant in weathered shale and in the unsaturated zone near the water table.

Notably, organisms with the potential ability to oxidize formaldehyde were more abundant at the hillslope compared to the floodplain. While genes involved in Thiol independent, Tetrahydromethanopterin (H4MPT) and Glutathione and Bacillithiol/Mycothiol dependent formaldehyde oxidation were abundant at the hillslope sites, these were scarce at the floodplain (Figure S2). The fused gene encoding for a bi-funcitonal enzyme with formaldehyde activating enzyme (fae) and hexulose-6-phosphate synthase (hps) capacity was only found within the genomes of *Bathyarchaeota*, detected in groundwater saturated sediment at the floodplain.

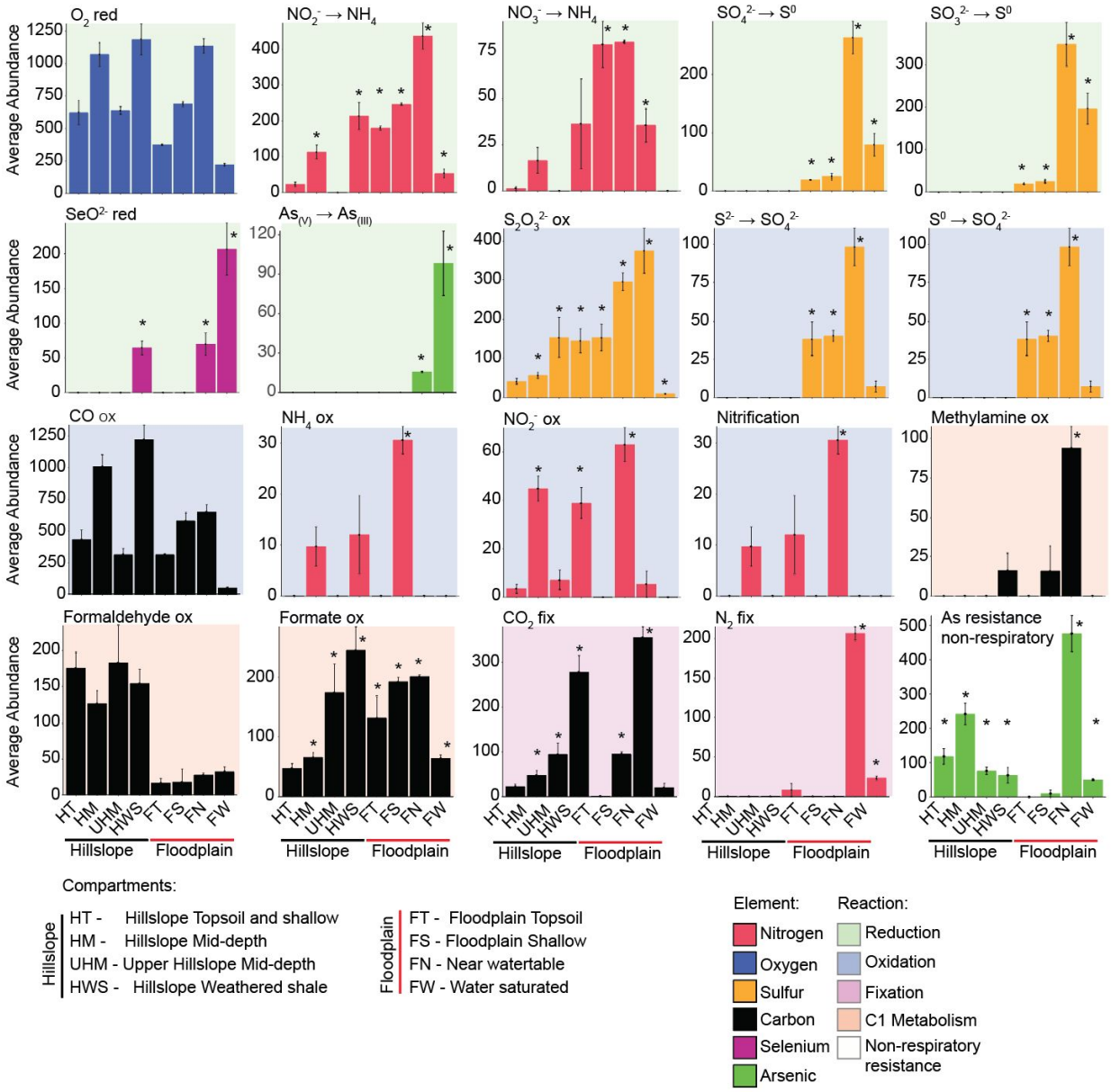

Figure 4 - The abundance of functions and indicator functions. Bars marked with stars are indicator functions for the samples. For example, nitrification is an indicator function at the near water table zone, whereas carbon fixation is an indicator function for at multiple zones (hillslope mid-depth, upper hillslope mid-depth, weathered shale, floodplain shallow samples, and floodplain near water table). Bars are SE. Representation of each sample is given in Figure S3.

While no full methanogenic pathway was observed in the microbial communities of the hillslope, evidence for the presence of methanogenic bacteria was found at the hillslope. Concentrations higher than ~3200 mg/L_{H2O} of methane were measured in the water phase at the PLM6 well from January to July 2017. Concentrations higher than ~ 1100 ppm_v were observed in the gas phase in an adjacent well (PLM3) in May and July 2017 (Table S3). Carbon and hydrogen stable isotopes of methane show mixing between acetoclastic and hydrogenotrophic methane, except for samples taken from 3.7 meters at PLM3 where methane has isotopic ratios typical of hydrogenotrophic methane (Figure 5).

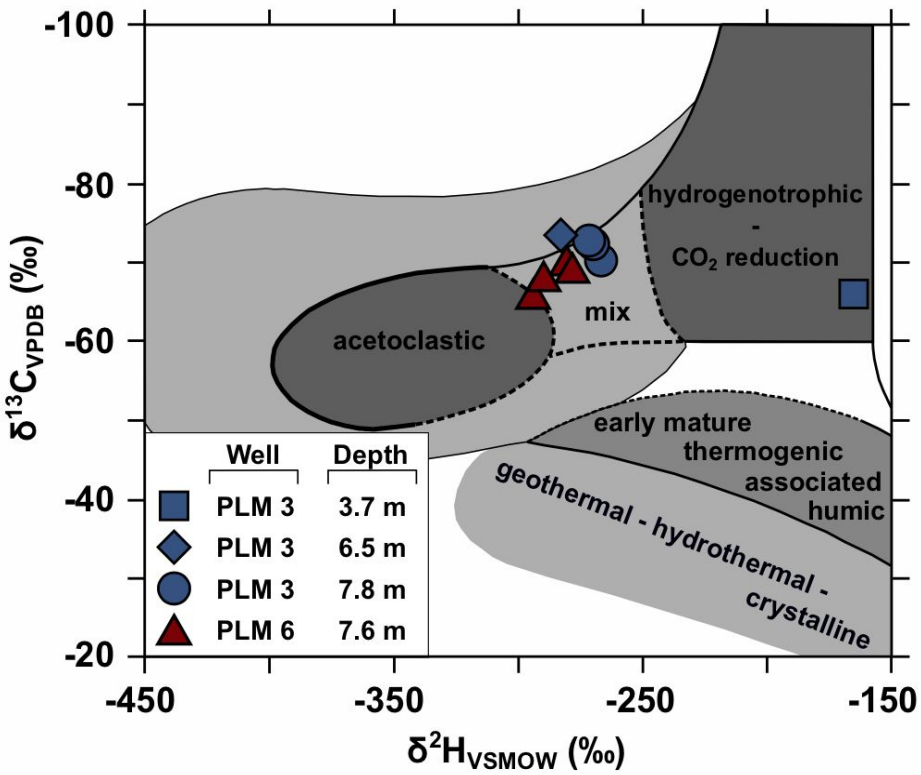

Figure 5 – **Carbon and hydrogen isotopic compositions of methane samples.** $\delta^{13}C$ and $\delta^{2}H$ values indicate mixing between acetoclastic and hydrogenotrophic methane for depths from 6.5 m to 7.8 m. Well PLM3 at 3.7 m shows hydrogenotrophic methane. This figure was adapted from Whiticar [19].

Some microbially mediated biogeochemical reactions potentially occur at multiple locations at the study site. However, these processes are apparently mediated by organisms from different taxonomic lineages. For example, *Deltaproteobacteria* (*Desulfobacca acetoxidans*, *Syntrophorhabdus aromaticivorans*, *Geobacteraceae*, and species closest to *Deltaproteobacteria* RBG genomes) are potential reducers of selenate at the hillslope, whereas Candidatus *Methylomirabilis oxyfera*, and *Rokubacteria* conduct the same role at the floodplain (Figure 6). The selenate reductase gene *srdA* in *Rokubacteria* was found on a 4.05 Kbp scaffold also carrying genes encoding a putative polysulphide reductase *NrfD* family protein, and 4Fe-4S ferredoxin-type, iron-sulfur binding domain, both of which are likely components of the microbial electron transport chain. The last coding region has no predicted function.

A similar pattern of partitioning is apparent for nitrogen fixation. All genes required for nitrogen fixation are present in the genomes of *Myxococcales* (class *Deltaproteobacteria*) in the weathered shale zone from the hillslope, whereas at the floodplain this role is taken by *Geobacteraceae* from the same class and *Nitrospirae* closely related to species from Rifle, CO [14]. The abundance of organisms that may fix CO_2 increases with depth at the hillslope as well as in the floodplain, except for within groundwater saturated samples. As in the case of nitrogen fixation and selenate reduction, multiple taxonomically distinct organisms that are potentially able to fix CO_2 at the hillslope and the floodplain. The most abundant autotrophs occur exclusively at the hillslope and are *Alphaproteobacteria, Gemmatimonadetes*, *Deltaproteobacteria, and Thaumarchaeota*. While *Thaumarchaeota* species seem to fix CO_2 via

the 3-hydroxypropionate /4-hydroxybutyrate (3HP-4HB) pathway, species of the other phyla are using the non-photosynthetic Calvin cycle for this purpose. At the floodplain, *Betaproteobacteria*, *Chloroflexi*, and Candidate phyla *Muproteobacteria* may conduct carbon fixation utilizing the non-photosynthetic Calvin cycle.

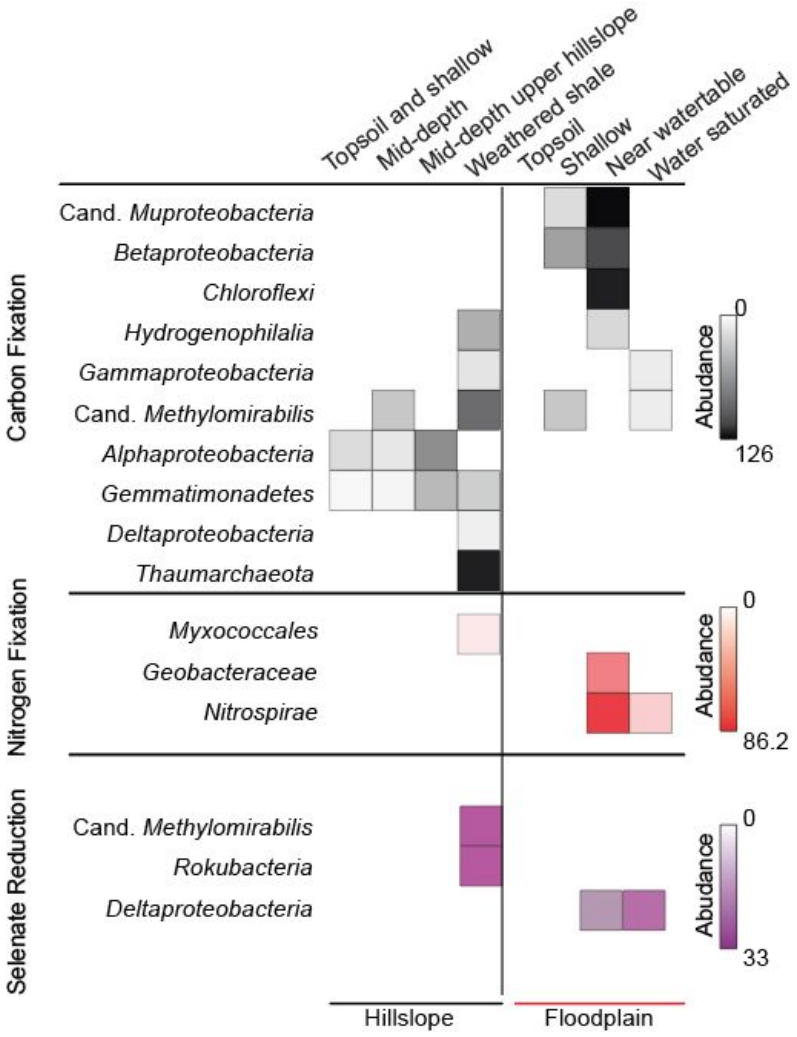

Figure 6 - Metabolic functions conducted by different taxa at the hillslope compared to the floodplain. CO_2 fixation, N_2 fixation, and selenate reduction all occur both at the hillslope and floodplain

zones. However, these functions are carried out by organisms from different taxa at each location. Heatmap shows the average abundance of organisms at each zone. Average abundance and SE are given in Figures S4-S6.

Discussion

The hillslope to riparian zone transect was selected to study the distribution of microbial metabolic capacities with depth below the surface and distance from the river. The sampling sites along the transect follow a hydraulic flow line that connects sites from a higher elevation with their downstream counterparts (Tokunaga et al., under review). The sites along the transect are occasionally connected by infiltration and lateral subsurface flow, and rarely by overland runoff; such connectivity is expected mostly in the weeks following snowmelt. During this late summer investigation, when vertical and lateral waterflow does not occur, we found that the microbial communities are largely distinct from one another. Additionally, while soil and saprolite at this site are derived from Mancos Shale, differences in the physical and chemical properties of layers in parent rock could impose different selective pressures on microbes. This effect could contribute to taxonomic variability at similar depths but at different sites. The finding of many genomes from the same species (but not strains, see Figure 1) in multiple sites suggests that similar, but not identical, environmental conditions select for similar, but distinct strains. However, 8 of 484 genomes of essentially identical strains occur at two depths or at more than one site, a result that probably requires dispersal through the soil and saprolite profile. The finding of three essentially genotypically identical microorganisms (one *Actinobacteria*, and two *Thaumarchaeota*) in sites downslope from each other probably indicates dispersal between sites across the hillslope. If transfer occurs, it is probably during restricted periods, as

downslope groundwater movement through the soil occurs at only ~10 to 20 m per month and for only a few weeks per year (Tokunaga et al., under review).

In a previous study conducted at the same site, and in which rpS3 protein sequences were used, organisms from 35 phyla were detected [14]. We reconstructed genomes for 25 of these. The phylum-level lineages for which we did not recover genomes were *Amesbacteria*, *Katanobacteria*, *Latesbacteria*, *Modulibacteria*, *Moranbacteria*, *Nealsonbacteria*, *Parcubacteria*, and *Yanofskybacteria*, all from the CPR clade. Genomes of organisms from these phyla were not recovered due to their low relative abundance, less than 0.04% based on the abundance of their rpS3 genes in the samples [14]. Many prior studies have predicted limited metabolic capacity of CPR organisms [20, 21], so their absence from the metabolic analyses would not substantially alter the results. Additionally, we did not recover genomes of *Spirochaetes* which were detected in 8 out of 41 samples and had a maximum relative abundance of ~1% based on the rpS3 gene [14]. For classes of *Proteobacteria*, 26 Alphaproteobacteria, 27 Betaproteobacteria, 5 Gammaproteobacteria, 72 Deltaproteobacteria and 2 Candidate phylum *Muproteobacteria* genomes were reconstructed. Candidate phylum *Lambdaproteobacteria* was detected based on the presence of its rpS3 gene, but a genome was not recovered. The detection of organisms for which genomes were not recovered is unsurprising, due to the limitations of resolving genomes from complex communities such as found in soils.

The strength of the genome-resolved metagenomic approach lies in the possibility to compare organisms at the very high-resolution level of their genomes and infer metabolic insights. Moreover, metabolic potential as detected from metagenomic data and which is not equatable with function in-situ measurements, provides valuable information on the functionality available at each site while minimizing temporal fluctuations.

Microbes and hydrology were suggested to jointly control the fate of carbon and nitrogen in the watershed [22]. Organisms capable of CO_2 fixation were detected throughout the hillslope, at highest abundance within weathered shale (127-200 cm depth) and linked to *Thaumarchaeota*, and at every depth other than the shallow zone at the floodplain, indicating potentially important microbial contributions to soil and sediment organic carbon compound inventories. While organic carbon from decomposition of organic material is readily available for microorganisms in shallow soil, CO_2 fixation could sustain heterotrophs in the subsurface. Autotrophically derived organic carbon may turn over rapidly, given that concentrations measured at these locations were low, 3.26 and 9.6 mg/L within the weathered shale and near groundwater, respectively [14]. Members of *Betaproteobacteria* and *Chloroflexi* were previously described to be autotrophs [23, 24], but carbon fixation by Candidatus *Muproteobacteria* is suggested here for the first time. Environmental conditions probably select for different bacteria/archaea with these roles under different conditions.

The presence of methylamine oxidizing organisms is significantly correlated with unsaturated sediment from above the water table at the floodplain, but this capacity was also detected in shallow soil at the floodplain, but at lower abundance. Methylamine, which is released from biodegradation of proteins and nitrogen-containing osmolytes [24], for example, those released from microbial cells due to repeated drying and rewetting [26], can be used as sole carbon and energy source for methylotrophs [27], as well as a nitrogen source [28]. The origin of methylamine in the deeper zones of the floodplain is unclear, but it could be sourced from the surface or formed by oxidation of compatible solutes and other compounds from microbial community members [29].

Genomes carrying all the required genes for methanogenesis were not detected at the study site. However, there is evidence that methanogenic microbes produce methane in

hillslope groundwater at depths greater than those that were sampled for metagenomic analysis. Carbon and hydrogen isotope ratios measured in methane from PLM3 and PLM6 suggest that the source of methane at these sites is biogenic. Interestingly, methane from PLM3 at 3.7 m depth was found to have a hydrogenotrophic origin while all other samples were from mixed biogenic origins. Hydrogenotrophic methanogenesis is the most widespread pathway among methanogenic metabolic pathways [30]. Organisms employing this pathway use H_2, or sometimes formate as an electron donor [30, 31]. They reduce CO_2 to CH_4 by utilizing the reductive acetyl-CoA or Wood–Ljungdahl pathway [32]. In order to conserve energy, hydrogenotrophs may couple the Wood–Ljungdahl pathway to methanogenesis. While genes responsible for formate oxidation were detected in some genomes, these organisms did not have the required genes for a complete methanogenesis. Therefore it is suggested that the methanogenic organisms are either found at low abundance at the depths that were sampled or that they are exclusive to deeper depths, potentially confided to the groundwater habitat in which methane was detected. Similarly, the capacity for methane oxidation could not be established. However, 24 genomes encode *amoABC/pmmoABC* genes that due to structural similarities between the ammonia oxidizing and the particulate methane oxidizing variants, cannot be assigned a specific function.

Some organisms have genes required to oxidize formaldehyde and formate, which may be formed by methane oxidation or by fermentation and breakdown of plant and animal products [33]. Formaldehyde oxidation could be a means of detoxification [30], but could also be used for energy production and as a carbon source via the ribulose monophosphate (RuMP) pathway [30]. The presence of a fused gene encoding for a bifunctional enzyme with formaldehyde activating enzyme (fae) and hexulose-6-phosphate synthase (hps) exclusively in

organisms from groundwater saturated sediment may indicate that formaldehyde is an important carbon source within this ecological compartment.

Nitrogen fixation often takes place in estuarine, coastal environments, and river banks sediment [23]. Specifically, at river banks, nitrogen fixation could account for a significant portion of nitrogen transported to groundwater. For example, 9.3% of the total terrigenous inorganic nitrogen transported into the Yangtze estuarine and coastal environment is attributed to biological (as opposed to industrial) nitrogen fixation [24]. At the East River study site, nitrogen fixation at the floodplain, particularly close to groundwater, could contribute to the flux of nitrogen into the river, affecting downstream water quality. The identity of nitrogen fixers is differentiated spatially, probably due to selection for other traits. Inorganic oxidized nitrogen from shallow soil layers could potentially be reduced by microorganisms in deeper subsurface, and most often at the floodplain. The reducing environment probably explains why nitrate-reducing, and to a larger extent nitrite reducing organisms were more abundant at the floodplain compared to the hillslope. Some degree of stratification involving nitrogen compound reduction capacities in shallower regions and sulfate reduction in deeper regions of the profile that spans aerobic, transitional and groundwater saturated regions of riparian zone extends similar findings related to stratification in marine or terrestrial sediments [34, 35].

A previous study at the East River watershed suggested that the loss of SO_4 across a low-gradient, meandering section of the watershed could indicate microbial sulfate reduction in floodplain sediments [36]. Furthermore, this process was suggested to modify the hillslope inputs of sulfur species to the riparian zone. The abundance of organisms in the water-saturated zone, capable of reducing sulfate and sulfite to elemental sulfur, supports this suggestion.

Arsenic is present in the Mancos shale, likely as an impurity in sulfide and selenide minerals, and is released to solution as the trivalent form, arsenite, or the pentavalent form,

arsenate, following oxidative dissolution during shale weathering. Microorganisms can alter arsenic speciation and transform arsenic compounds between biotic (i.e., methyl-arsenic and arsenosugars) and abiotic forms [37]. Arsenate may be reduced to the more toxic form, arsenite [35], with or without methylation [36, 37] prior to export from cells during detoxification. Two genomes of Deltaproteobacteria that are relatively abundant in the riparian zone sediments encode an arsenate reductase *arr* operon (a third partial *arr* operon, including the key *arrA* subunits and *arrS* was also detected). Thus, these bacteria may decrease arsenate and increase arsenite concentrations in groundwater and the river. However, genes involved in arsenite methylation which were detected at high abundance in unsaturated sediment near the water table suggest that arsenic might be methylated before being expelled into the environment. Microbially-mediated reactions in deeper riparian zone sediment may be important in determining arsenic speciation in groundwater and subsequently in the river.

Selenium (Se), an essential element for several enzymatic reactions, becomes toxic when present at a high concentration in its most oxidized form, selenate. Se occurs in Mancos shale in the form of insoluble metal selenides [13] that can be oxidized, forming mobile selenite and selenate [38, 39]. Here, we attribute reduction of selenate released by weathering of the Mancos shale to *Deltaproteobacteria* within the floodplain and to *Rokubacteria* and Candidatus *Methylomirabilis* at hillslope sites. *Geobacter* species (phylum *Deltaproteobacteria*) are known to reduce selenite [40] and Candidatus *Methylomirabilis* may be capable of coupling anaerobic methane oxidation to selenate reduction [41]. However, this is the first evidence suggesting that *Rokubacteria* is capable of selenate reduction. The activities of these three groups of bacteria have the potential to limit the export of toxic, mobile selenate from the watershed.

Riparian ecosystems provide key services to society [42]. Some studies suggest that they are particularly sensitive to climate change impacts [43–45] whereas other studies predict

that riparian zones will be relatively resilient, as they evolved under conditions of high environmental variability [46]. Within montainous watersheds such as studied here, rising temperatures and atmospheric CO_2 concentration, droughts and early snowmelt all have the potential to alter the riparian ecosystem [2]. These changes may alter water table elevations, and consequently depth of sub-surface anoxia, affecting vegetation [47] as well as microbial community composition and functioning [48]. Lowering of the water table could accelerate the release rate of solutes, such as nitrate, sulfate, iron, and selenate from soils, causing a deterioration of water quality both locally and downstream [49]. The chemical forms of these elements, as well as the reaction rates, are, in part, determined by transformations mediated by subsurface microbial organisms. The distribution of functions different organisms (and in different metabolic contexts) could provide functional flexibility as conditions change over the course of the year and support system resilience to longer-term environmental changes [50]. Our findings of functional partitioning between the hillslope and the riparian zone and further partitioning within compartments of these units constrains understanding of the total biodiversity within the watershed system.

Conclusions

Most elemental cycles near the Earth's surface are microbially mediated, but we know very little about how these biogeochemical processes are distributed across organisms and environmental compartments, especially at the watershed scale. Here, we show that such questions can be addressed across large transects, from high on hillslopes down to the river corridor, via genomic analysis of microbial communities. A similar methodology could be used across scalable transects in other environments. We find heterogeneity in the distribution of metabolic potential, the organisms involved, and presumably in the biogeochemical cycling that

occurs at each site. Thus, we conclude that different ecosystem compartments such as the riparian zone and weathered bedrock need to be identified and studied individually. However, sites must also be studied in combination to begin to predict how nutrients and contaminants released from underlying rock are modulated by spatially variable biological processes to dictate exports from watersheds.

Methods:

Sampling - The Pumphouse Lower Montane (PLM) intensive study site is located on the north-east facing slope of the East River valley near Crested Butte, Colorado, USA (38°55'12.56"N, 106°56'55.39"W). Soil and sediment samples were collected as described in Lavy et al. [14]. Briefly, samples were taken during four days in June 2016 from five sites along a hillslope transect and one site at a floodplain using manual augur and sterile plastic liners. These sites ordered from highest to lowest elevation are PLM0, PLM1, PLM2, PLM3, and PLM4. All samples other than those taken at 90 cm depths at PLM4 were from above the water table. An additional site at the floodplain was sampled using an air-cooled split-spoon Odex drill during the same period of time. The installation of sampling equipment for another experiment at the same locations revealed compositionally distinct shale layers over the hillslope with a 20° uplift to the northeast. The first five centimeters of topsoil and plant material were cleared from the surface of each sampling site before samples were taken. In total, 20 samples were collected as follows: PLM0 - 5, 30, 60 cm; PLM1 – 5, 30, 60, 100 cm; PLM2 – 5, 30 cm; PLM3 – 5, 30, 60, 127 cm; PLM6 – 50, 170, 200 cm; PLM4 – 5, 32, 65, 90 cm. Each sample was manually homogenized in a sterile Whirl-Pak bag and aliquots of 5 g cleared from rocks and roots were placed on ice for DNA extraction, as well as in 10 ml of LifeGuard Soil Preservation Solution (Qiagen, Netherlands) for RNA and DNA co-extraction.

DNA extraction and sequencing - The extraction process is described in Lavy et al. [14]. Briefly, DNA was extracted from 10 g of soil or sediment with DNeasy PowerMax Soil Kit (Qiagen, Netherlands) in two batches of 5 g each which were combined during the cleaning step. DNA was also co-extracted with RNA from 5 g of soil using RNeasy PowerSoil Total RNA Kit (Qiagen, Netherlands) and Phenol:Chloroform:Isoamyl Alcohol 25:24:1 saturated with 10 mM Tris (final pH 8.0) and 1 mM EDTA. RNeasy PowerSoil DNA Elution Kit (Qiagen, Netherlands) was used to collect DNA which was further cleaned using DNeasy PowerClean Pro Clean Up Kit (Qiagen, Netherlands). Overall, two DNA samples were produced from each sampling, one from DNA extraction and the second from the DNA that was co-extracted along with RNA. A third DNA sample was extracted from the 90 cm deep PLM4 sample, thus a total of 41 DNA samples were used for further analysis.

Metagenomic libraries were prepared at the Joint Genome Institute (JGI) after validating concentrations and DNA integrity using Qubit (Thermo Fisher Scientific) and gel electrophoresis, respectively. NEB's Ultra DNA Library Prep kit (New England Biolabs, MA) was used for library preparation for Illumina with AmpureXP bead selection aimed for an average fragment length of 500 base-pair (bp) following the manufacturer's protocol. The libraries were sequenced at JGI in an Illumina Hiseq 2500, generating 150 bp paired-end, sequences (Table S1)

Sequences analysis and genome binning- Raw reads were trimmed based on quality scores with Sickle [51] and assembled with IDBA-UD v1.1.1 [52] using kmer size range of 40-140. Only assembled scaffolds longer than 1Kbp were retained for downstream analysis. Open reading frames were identified with Prodigal v2.6.3 [53] and using the metagenomic

setting. Predicted protein sequences were annotated using usearch [54] against KEGG, UniRef100 and UniProt databases.

Genomes were binned manually using the ggKbase platform [55] as well as with the automated binners CONCOCT v0.4 [56], Maxbin2 [57], Abawaca1, Abawaca2 [20], and MetaBAT v0.32.4 [58]. All predicted bins were used as input for DAStool v1.1.0 [59] which selected for the best representative genome. Representative genomes were dereplicated with dRep v2.2.2 [60] at 98% average nucleotide identity (ANI), which selected for the best species level representatives. Each of the resulting genomes was further curated in ggKbase, and genomes considered as near-complete, having ≥ 80% of 43 bacterial single copy genes or 38 archaeal single copy gene sets [61] were selected for further curation. These genes were imported into Anvi'o v5.4 [62] for further manual inspection and refinement using differential coverage, kmer frequency, and GC content.

Genome abundance across samples - To determine the presence and abundance of genomes across samples, reads from each sample were mapped against genomes with bowtie2 [63]. The average coverage and breadth of coverage of each genome in each sample was then calculated (Olm et al., 2017). Each genome is considered to be present in at least one sample (at a minimum, the sample from which it was originally binned) but could be falsely identified in other samples due to a low breadth cutoff (i.e., false positive). Therefore, we implemented a breadth cutoff of 0.828 based on the lowest breadth cutoff that retains all genomes from the samples they were originally binned from (Figure S7). The coverage of the genomes was adjusted according to the number of basepairs that were sequenced from each sample to compensate for differences in sequencing depth.

Strains detection - Strains of the same species were identified as either genome that shares more than 99% ANI according to their dRep cluster placement [60], or by having an average breadth > 0.99 within more than one sample.

Partitioning into zones - Samples were grouped based on the results of a previous study [14] in which community composition was defined by ribosomal protein S3 (*rpS3*), and in which microbial communities at the hillslope and floodplain compartments were found to be distinct. Furthermore, a depth gradient was observed to affect community composition and potential metabolism. Therefore eight zones were defined for the purpose of the current study. At the hillslope compartment, samples were grouped into either Top and shallow soil (n=14), mid-depth zone (n=8), mid-depth at an upper hillslope location (n=4), or weathered shale (n=6) (Figure 1). At the riparian zone compartment, samples were grouped into either topsoil (n=2), shallow soil (n=2), unsaturated sediment at 10 cm above the water table (n=2), or groundwater saturated (n=3).

Metabolism prediction - Genes encoding for key enzymes in 23 metabolic functions were identified with 204 KOfam [64] and custom Hidden Markov Models [61]. Cutoffs for all custom models and for 10% of all KOfam models were tested and adjusted by searching for best matches of HMM hits against NCBI's nr database using BLASTP. The search results and protein domains were compared with a reference protein sequence from Uniprot, and a stringent cutoff was selected to prevent false positives by setting it to be above the first false positive in the search results (Table S4). In case a genome was found to have >= 0.75 of all genes required to complete a function, and the genome was representing a dRep cluster, then

the missing genes were sought for within the other genomes of the cluster using the same HMM model and thresholds. A metabolic function was considered present only if all required genes were present within a single genome (Table S4). The presence of *dsrD* gene was required in order to consider an organism, which has dsrAB genes, as a sulfide oxidizer. As some genomes lack *dsrD* due to genome incompleteness, concatenated *dsrAB* genes were aligned with a database of oxidizing and reducing *dsrAB* genes from Muller et al. [65]. The genes were aligned with MAFFT [66] with default parameters, and a tree was inferred with Fasttree [67]. Concatenated *dsrAB* genes that were located within the reductive clade were considered as potentially reductive *dsrAB*. The abundance of each function was determined by summing the average coverage of reads mapped to each organism that has all required genes to fulfill the function.

Due to structural similarities, the HMM which was designed to pick identify methane dehydrogenase (*mdh*) could also detect other alcohol dehydrogenases. The identity of the amino acid sequence was verified by placing the sequences along with a reference set into a phylogenetic tree as described by Diamond et. al. [68] and using the same reference set.

Hidden Markov Models for arsenic genes were custom made. The seed sequences for the HMMs were taken from Zhu et. al., [69] and references within. Seed sequences were added by searching for amino acid sequences of proteins with similar function and structure by searching NCBI's nr database with BLASTp to increase the taxonomic diversity. For each gene, reference sequences from the same (or related) protein families were used to verify clustering of all seed sequences. These were taken from the NCBI Conserved Domain database. Once seed sequences were selected, HMMs were built with HMMER [70]. Score cutoffs were selected as reported above. To verify the results, sequences from the current study were aligned with the

seed sequences and reference sequences using MAFFT [66] and a FastTree was constructed to assess clustering.

Taxonomy - The taxonomy of each genome was determined by comparing the concatenated sequences of 15 ribosomal proteins to a reference dataset [71].
The sequences of rpL2, rpL3, rpL4, rpL5 rpL6, rpL14, rpL15, rpL18, rpL22, rpS4, rpS8, rpS17 and rpS19 were identified with FetchMG v1.1. which is available as a standalone part of MOCAT [72]. Additionally, rpS10 and rpL24 were searched for with TIGR01049 and TIGR01079 models for bacterial, or TIGR01046 and TIGR01080 models for archaeal bins, respectively [73]. The search was conducted with hmmsearch v3.1b2 from HMMER suite using NC cutoff implement within the model files [70]. The genes were concatenated and aligned to a consensus sequence generated from a 3078 sequences reference set and placed on a reference tree using a Neighbour-Joining (NJ) algorithm implemented in pplacer v1.1.alpha [74]. Next, the 10 closest taxa to each query genome were identified and the genes of their 15 ribosomal proteins were extracted from the reference set. Each gene of the new, leaner reference set was aligned with the query gene using MAFFT [66] and a Maximum-Likelihood (ML) tree was constructed on the CIPRES Science Gateway v3.3 [75] with RAxML [76] using the LG substitution model and bootstrapping, allowing the software to halt bootstrapping once it reached a consensus. The taxonomy of each query genome was inferred based on the topology of the tree and its closest reference genome.

CH_4 concentrations were quantified on a GC-2014 Shimadzu gas chromatograph (Shimadzu Corporation, Kyoto, Japan). Using a gas-tight syringe, 4.5 mL of sample aliquot was injected into a 1 mL stainless steel loop mounted on a 10-port valve (Valco). CH_4 was separated

on a packed HayeSep-D packed column (4 m x 1/8 inch). CH_4 was quantified with a flame ionization detector (FID). For water samples, after headspace concentration measurements, methane dissolved in the water concentrations were calculated using Henry's Law constant. Analytical precision based on repeated standard analyses was ~ 3% of the reported concentrations.

Carbon and hydrogen isotope ratios of CH_4 were measured separately using a gas chromatograph and an isotope ratio mass spectrometer interfaced with a pyrolysis (GC-P-IRMS) and a combustion reactor (GC-C-IRMS, Thermo Fisher Scientific, Bremen, Germany). CH_4 was separated chromatographically on an HP-molesieve fused silica capillary column (30 m x 0.320 mm). For hydrogen isotopes, after GC separation, CH_4 was pyrolyzed in a capillary carbon-coated ceramic tube at 1450 °C and the hydrogen isotope ratios were measured in the IRMS. For carbon isotopes, after GC separation, the CH_4 was combusted to CO_2 at 1030 °C in a capillary ceramic tube loaded with Ni, Cu, and Pt wires and the carbon isotope ratio was acquired in the IRMS. Results are reported in the standard δ notation as the per mil deviation (‰) relative to Vienna Pee Dee Belemnite (VPDB) for carbon and relative to Vienna Standard Mean Ocean Water (VSMOW) for hydrogen. Repeated injections of CH_4 yield values of -39.46 ± 0.33‰ (1σ; n=11) and -167.01 ± 3.10‰ (1σ; n=14) for δ^2H.

Statistical analysis - All analysis was done R v3.4.3 [77] and Rstudio v1.1.423 [78]. Correlation between genomic bins and sampling zones was done with “multipatt” command and r.g. function in the Indicspecies package v1.7.6 [79]. Indicator function was tested with the IndVal.g function. In both cases, 9999 permutations were conducted and the p-value was

corrected for multiple testing with Benjamini-Hochberg procedure. Maps were retrieved from Google maps database using Google Earth v7.3.2.

Declarations

Ethics approval and consent to participate

Not required for this study

Consent for publication

Not required for this study

Availability of data and material

- Raw reads are available through the NCBI Short Reads Archive. Accession number for each sample is provided in Table s1.
- Sequences of the 484 Metagenomic Assembled Genomes are available at: https://ggkbase.berkeley.edu/PLM2016-NC-80p/organisms
- HMMs used in the current study (ref to the updated list, and list of ko and thresholds for KOfam) are available at https://figshare.com/projects/Supplementary_Data_for_Taxonomically_and_metabolically_distinct_microbial_communities_with_depth_and_across_a_hillslope_to_riparian_zone_transect/66182
- Phylogenetic tree of 15RP https://figshare.com/projects/Supplementary_Data_for_Taxonomically_and_metabolically_distinct_microbial_communities_with_depth_and_across_a_hillslope_to_riparian_zone_transect/66182

- Presence/absence of genes in each genome and the rules by which presence/absence of fucuntions were determined.https://figshare.com/projects/Supplementary_Data_for_Taxonomically_and_metabolically_distinct_microbial_communities_with_depth_and_across_a_hillslope_to_riparian_zone_transect/66182

Competing interests

None

Funding

The work described in the manuscript was supported as part of the Watershed Function Scientific Focus Area funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-AC02-05CH11231.

Authors' contributions

A.L designed research, performed research, analyzed data, and wrote the paper. P.B.M.C conducted fieldwork and assisted in constructing the metabolic pathway analysis. R.K created and tested Hidden Markov Models for arsenic metabolism related genes. M.B. contributed to methane data acquisition and in writing the paper. J.W assisted in designing and conducting fieldwork. T.K.T assisted in designing and

conducting fieldwork. K.H.W took part in the research design and assisted in fieldwork. S.S.H took part in the research design. J.F.B supervised the study and mentored the first author.

References

1. Viviroli D, Dürr HH, Messerli B, Meybeck M, Weingartner R. Mountains of the world, water towers for humanity: Typology, mapping, and global significance. Water Resour Res. 2007;43. doi:10.1029/2006WR005653.
2. Hubbard SS, Williams KH, Agarwal D, Banfield JF, Beller H, Bouskill N, et al. The East River, CO Watershed: A Mountainous Community Testbed for Improving Predictive Understanding of Multi-Scale Hydrological-Biogeochemical Dynamics. Vadose Zone J. 2018.
3. Bardelli T, Gómez-Brandón M, Ascher-Jenull J, Fornasier F, Arfaioli P, Francioli D, et al. Effects of slope exposure on soil physico-chemical and microbiological properties along an altitudinal climosequence in the Italian Alps. Sci Total Environ. 2017;575:1041–55.
4. Djukic I, Zehetner F, Mentler A, Gerzabek MH. Microbial community composition and activity in different Alpine vegetation zones. Soil Biol Biochem. 2010;42:155–61.
5. Klimek B, Niklińska M, Jaźwa M, Tarasek A, Tekielak I, Musielok Ł. Covariation of soil bacteria functional diversity and vegetation diversity along an altitudinal climatic gradient in the Western Carpathians. Pedobiologia. 2015;58:105–12.
6. Xu M, Li X, Cai X, Gai J, Li X, Christie P, et al. Soil microbial community structure and activity along a montane elevational gradient on the Tibetan Plateau. Eur J Soil Biol. 2014;64:6–14.
7. Zhang B, Liang C, He H, Zhang X. Variations in Soil Microbial Communities and Residues Along an Altitude Gradient on the Northern Slope of Changbai Mountain, China. PLOS ONE. 2013;8:e66184.
8. Tytgat B, Verleyen E, Sweetlove M, D'hondt S, Clercx P, Van Ranst E, et al. Bacterial community composition in relation to bedrock type and macrobiota in soils from the Sør Rondane Mountains, East Antarctica. FEMS Microbiol Ecol. 2016;92. doi:10.1093/femsec/fiw126.
9. Rempfert KR, Miller HM, Bompard N, Nothaft D, Matter JM, Kelemen P, et al. Geological and Geochemical Controls on Subsurface Microbial Life in the Samail Ophiolite, Oman. Front Microbiol. 2017;8. doi:10.3389/fmicb.2017.00056.
10. Levin SA. The Problem of Pattern and Scale in Ecology: The Robert H. MacArthur Award Lecture. Ecology. 1992;73:1943–67.
11. Pribulick CE, Foster LM, Bearup LA, Navarre-Sitchler AK, Williams KH, Carroll RWH, et al. Contrasting the hydrologic response due to land cover and climate change in a mountain headwaters system. Ecohydrology. 2016;9:1431–8.
12. Morrison SJ, Goodknight CS, Tigar AD, Bush RP, Gil A. Naturally Occurring Contamination in the Mancos Shale. Environ Sci Technol. 2012;46:1379–87.
13. Elrashidi MA. Selenium Distribution for Soils Derived from Mancos Shale in Gunnison and Uncompahgre River Basins, West-Central Colorado. Commun Soil Sci Plant Anal. 2018;49:1083–91.
14. Microbial communities across a hillslope-riparian transect shaped by proximity to the stream, groundwater table, and weathered bedrock - Lavy - - Ecology and Evolution - Wiley Online Library. https://onlinelibrary.wiley.com/doi/10.1002/ece3.5254. Accessed 14 Jun 2019.
15. Banfield JF, Nealson KH. Chapter 12: Biogeochemical Weathering of Silicate Minerals. In: Geomicrobiology: Interactions between Microbes and Minerals. Walter de Gruyter GmbH & Co KG; 2018. p. 391–428.
16. Viles HA. Microbial geomorphology: A neglected link between life and landscape. Geomorphology. 2012;157–158:6–16.
17. Bowers RM, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, Schulz F, et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled

genome (MIMAG) of bacteria and archaea. Nat Biotechnol. 2017;35:725–31.
18. Janssen PH. Identifying the Dominant Soil Bacterial Taxa in Libraries of 16S rRNA and 16S rRNA Genes. Appl Environ Microbiol. 2006;72:1719–28.
19. Whiticar MJ. Carbon and hydrogen isotope systematics of bacterial formation and oxidation of methane. Chem Geol. 1999;161:291–314.
20. Brown CT, Hug LA, Thomas BC, Sharon I, Castelle CJ, Singh A, et al. Unusual biology across a group comprising more than 15% of domain Bacteria. Nature. 2015;523:208–11.
21. Castelle CJ, Brown CT, Anantharaman K, Probst AJ, Huang RH, Banfield JF. Biosynthetic capacity, metabolic variety and unusual biology in the CPR and DPANN radiations. Nat Rev Microbiol. 2018;16:629.
22. Newcomer ME, Hubbard SS, Fleckenstein JH, Maier U, Schmidt C, Thullner M, et al. Influence of Hydrological Perturbations and Riverbed Sediment Characteristics on Hyporheic Zone Respiration of CO2 and N2. J Geophys Res Biogeosciences. 2018;123:902–22.
23. Howarth RW, Marino R, Lane J, Cole JJ. Nitrogen fixation in freshwater, estuarine, and marine ecosystems. 1. Rates and importance. Limnol Oceanogr. 1988;33:669–687.
24. Hou L, Wang R, Yin G, Liu M, Zheng Y. Nitrogen Fixation in the Intertidal Sediments of the Yangtze Estuary: Occurrence and Environmental Implications. J Geophys Res Biogeosciences. 2018;123:936–44.
25. Barrett EL, Kwan HS. Bacterial Reduction of Trimethylamine Oxide. Annu Rev Microbiol. 1985;39:131–49.
26. Halverson LJ, Jones TM, Firestone MK. Release of Intracellular Solutes by Four Soil Bacteria Exposed to Dilution Stress. Soil Sci Soc Am J. 2000;64:1630–7.
27. Anthony C. The biochemistry of methylotrophs. Academic press London; 1982.
28. Taubert M, Grob C, Howat AM, Burns OJ, Pratscher J, Jehmlich N, et al. Methylamine as a nitrogen source for microorganisms from a coastal marine environment. Environ Microbiol. 2017;19:2246–57.
29. Daly RA, Borton MA, Wilkins MJ, Hoyt DW, Kountz DJ, Wolfe RA, et al. Microbial metabolisms in a 2.5-km-deep ecosystem created by hydraulic fracturing in shales. Nat Microbiol. 2016;1:16146.
30. Thauer RK, Kaster A-K, Seedorf H, Buckel W, Hedderich R. Methanogenic archaea: ecologically relevant differences in energy conservation. Nat Rev Microbiol. 2008;6:579.
31. Liu Y, Whitman WB. Metabolic, phylogenetic, and ecological diversity of the methanogenic archaea. Ann N Y Acad Sci. 2008;1125:171–189.
32. Berghuis BA, Yu FB, Schulz F, Blainey PC, Woyke T, Quake SR. Hydrogenotrophic methanogenesis in archaeal phylum Verstraetearchaeota reveals the shared ancestry of all methanogens. Proc Natl Acad Sci. 2019;116:5037–44.
33. White D. C1 Metabolism. In: The physiology and biochemistry of prokaryotes. 3rd ed. New York: Oxford University Press; 2007. p. 359–82.
34. Wright JJ, Konwar KM, Hallam SJ. Microbial ecology of expanding oxygen minimum zones. Nat Rev Microbiol. 2012;10:381–94.
35. Smith RL, Harris, Jr. SH. Determining the Terminal Electron-Accepting Reaction in the Saturated Subsurface. In: Hurst CJ, Lipson DA, Garland JL, Stetzenbach LD, Mills AL, Crawford RL, editors. Manual of Environmental Microbiology, Third Edition. American Society of Microbiology; 2007. p. 860–71. doi:10.1128/9781555815882.ch69.
36. Carroll RWH, Bearup LA, Brown W, Dong W, Bill M, Willlams KH. Factors controlling seasonal groundwater and solute flux from snow-dominated basins. Hydrol Process. 2018;32:2187–202.
37. Oremland RS, Stolz JF. Arsenic, microbes and contaminated aquifers. Trends Microbiol.

2005;13:45–9.
38. Presser TS. Geologic origin and pathways of selenium from the California Coast Ranges to the West-Central San Joaquin Valley. In: Frankenberger WT, Benson S, editors. Selenium in the Environment. 1st edition. New York, NY: Marcel Dekker Inc.; 1994. p. 456.
39. Gebreeyessus GD, Zewge F. A review on environmental selenium issues. SN Appl Sci. 2018;1:55.
40. Pearce CI, Pattrick RAD, Law N, Charnock JM, Coker VS, Fellowes JW, et al. Investigating different mechanisms for biogenic selenite transformations: Geobacter sulfurreducens, Shewanella oneidensis and Veillonella atypica. Environ Technol. 2009;30:1313–26.
41. Luo J-H, Chen H, Hu S, Cai C, Yuan Z, Guo J. Microbial Selenate Reduction Driven by a Denitrifying Anaerobic Methane Oxidation Biofilm. Environ Sci Technol. 2018;52:4006–12.
42. González E, Felipe-Lucia MR, Bourgeois B, Boz B, Nilsson C, Palmer G, et al. Integrative conservation of riparian zones. Biol Conserv. 2017;211:20–9.
43. Tockner K, Stanford JA. Riverine flood plains: present state and future trends. Environ Conserv. 2002;29:308–30.
44. Perry LG, Andersen DC, Reynolds LV, Nelson SM, Shafroth PB. Vulnerability of riparian ecosystems to elevated CO2 and climate change in arid and semiarid western North America. Glob Change Biol. 2012;18:821–42.
45. Capon SJ, Chambers LE, Mac Nally R, Naiman RJ, Davies P, Marshall N, et al. Riparian Ecosystems in the 21st Century: Hotspots for Climate Change Adaptation? Ecosystems. 2013;16:359–81.
46. Catford JA, Naiman RJ, Chambers LE, Roberts J, Douglas M, Davies P. Predicting Novel Riparian Ecosystems in a Changing Climate. Ecosystems. 2013;16:382–400.
47. Vartapetian BB, Jackson MB. Plant Adaptations to Anaerobic Stress. Ann Bot. 1997;79 suppl_1:3–20.
48. Rühle FA, von Netzer F, Lueders T, Stumpp C. Response of Transport Parameters and Sediment Microbiota to Water Table Fluctuations in Laboratory Columns. Vadose Zone J. 2015;14. doi:10.2136/vzj2014.09.0116.
49. Freeman C, Lock MA, Reynolds B. Climatic change and the release of immobilized nutrients from Welsh riparian wetland soils. Ecol Eng. 1993;2:367–73.
50. Allison SD, Martiny JBH. Resistance, resilience, and redundancy in microbial communities. Proc Natl Acad Sci. 2008;105 Supplement 1:11512–9.
51. Joshi NA, Fass JN. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]. Available Httpsgithubcomnajoshisickle. 2011.
52. Peng Y, Leung HC, Yiu SM, Chin FY. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics. 2012;28:1420–8.
53. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010;11:119.
54. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 2010;26:2460–1.
55. Jillian F Banfield. Development of a Knowledgebase to Integrate, Analyze, Distribute, and Visualize Microbial Community Systems Biology Data. US Department of Energy; 2015.
56. Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, et al. Binning metagenomic contigs by coverage and composition. Nat Methods. 2014;11:1144–6.
57. Wu Y-W, Simmons BA, Singer SW. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics. 2016;32:605–7.
58. Kang DD, Froula J, Egan R, Wang Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ. 2015;3:e1165.

59. Sieber CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, et al. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. Nat Microbiol. 2018;3:836.
60. Olm MR, Brown CT, Brooks B, Banfield JF. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME J. 2017;11:2864–8.
61. Anantharaman K, Brown CT, Hug LA, Sharon I, Castelle CJ, Probst AJ, et al. Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. Nat Commun. 2016;7:13219.
62. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, et al. Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ. 2015;3:e1319.
63. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9.
64. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, et al. KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. bioRxiv. 2019;:602110.
65. Muller AL, Kjeldsen KU, Rattei T, Pester M, Loy A. Phylogenetic and environmental diversity of DsrAB-type dissimilatory (bi)sulfite reductases. ISME J. 2015;9:1152–65.
66. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol. 2013;30:772–80.
67. Price MN, Dehal PS, Arkin AP. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLOS ONE. 2010;5:e9490.
68. Diamond S, Andeer PF, Li Z, Crits-Christoph A, Burstein D, Anantharaman K, et al. Mediterranean grassland soil C–N compound turnover is dependent on rainfall and depth, and is mediated by genomically divergent microorganisms. Nat Microbiol. 2019;4:1356–67.
69. Zhu Y-G, Xue X-M, Kappler A, Rosen BP, Meharg AA. Linking Genes to Microbial Biogeochemical Cycling: Lessons from Arsenic. Environ Sci Technol. 2017;51:7326–39.
70. Finn RD, Clements J, Arndt W, Miller BL, Wheeler TJ, Schreiber F, et al. HMMER web server: 2015 update. Nucleic Acids Res. 2015;43:W30–8.
71. Hug LA, Castelle CJ, Wrighton KC, Thomas BC, Sharon I, Frischkorn KR, et al. Community genomic analyses constrain the distribution of metabolic traits across the Chloroflexi phylum and indicate roles in sediment carbon cycling. Microbiome. 2013;1:22.
72. Kultima JR, Sunagawa S, Li J, Chen W, Chen H, Mende DR, et al. MOCAT: A Metagenomics Assembly and Gene Prediction Toolkit. PLOS ONE. 2012;7:e47656.
73. Haft DH, Loftus BJ, Richardson DL, Yang F, Eisen JA, Paulsen IT, et al. TIGRFAMs: a protein family resource for the functional identification of proteins. Nucleic Acids Res. 2001;29:41–3.
74. Matsen FA, Kodner RB, Armbrust EV. pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. BMC Bioinformatics. 2010;11:538.
75. Miller MA, Pfeiffer W, Schwartz T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: 2010 Gateway Computing Environments Workshop (GCE). 2010. p. 1–8.
76. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30:1312–3.
77. R Development Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2012. http://www.R-project.org.
78. Rstudio Team. RStudio: Integrated Development for R. Boston, MA: RStudio Inc.; 2015.

http://www.rstudio.com/.
79. Caceres MD, Legendre P. Associations between species and groups of sites: indices and statistical inference. 2009. http://sites.google.com/site/miqueldecaceres/.

Supplement material

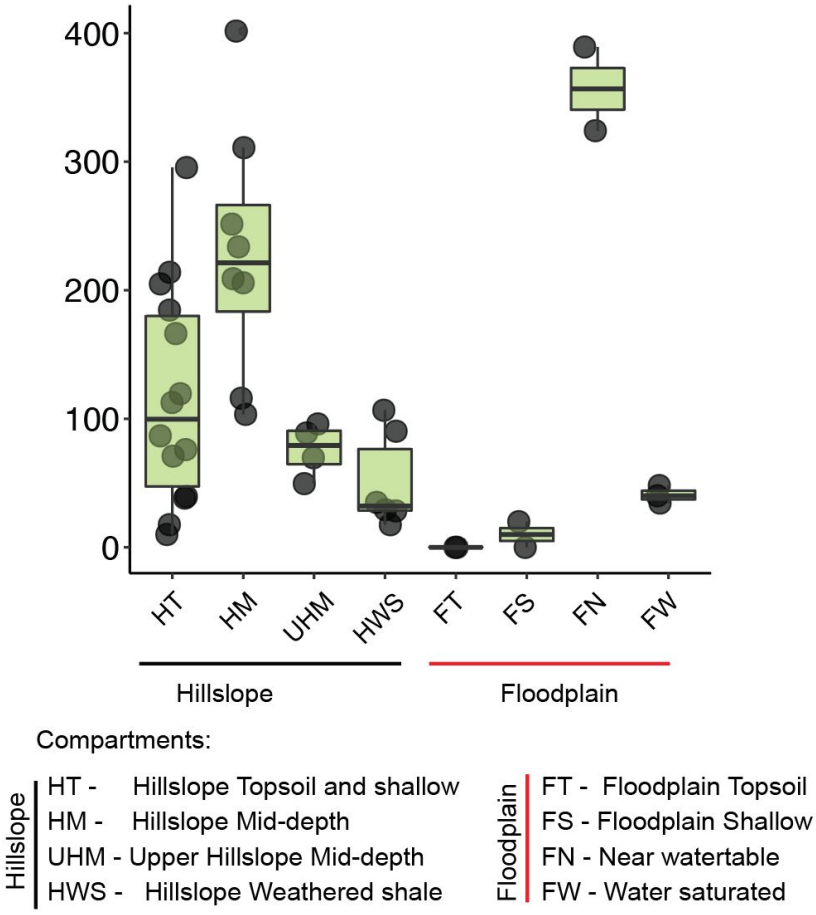

Figure S1 - Abundance of organisms with arsenic methylation related genes. Each dot represents one sample. Samples are clustered spatially by compartment

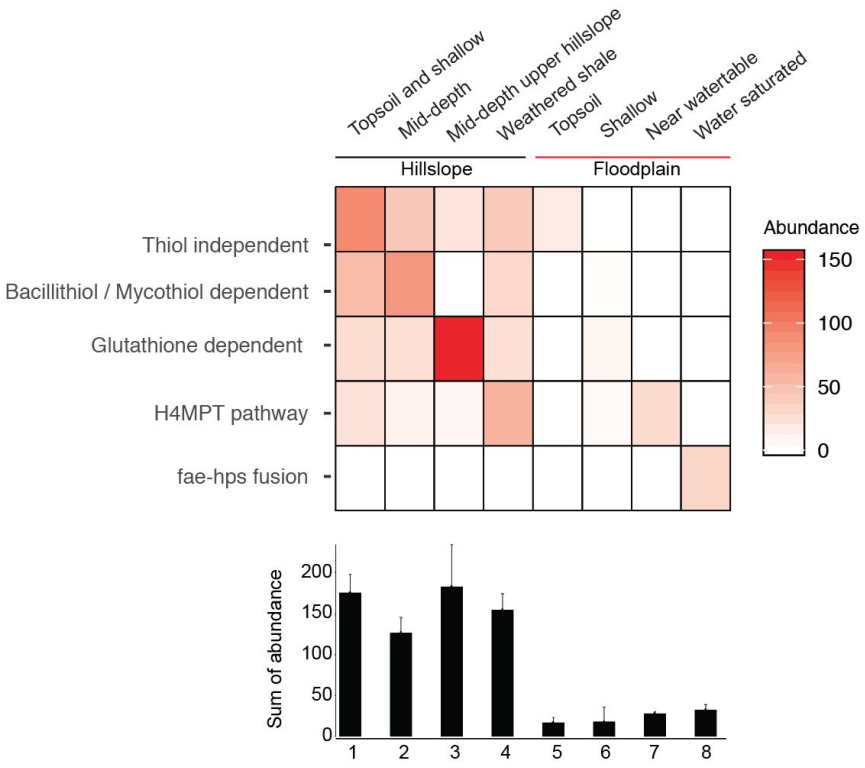

Figure S2: Abundance of organisms with the different Formaldehyde oxidation pathways detected across the hillslope to riparian zone transect.

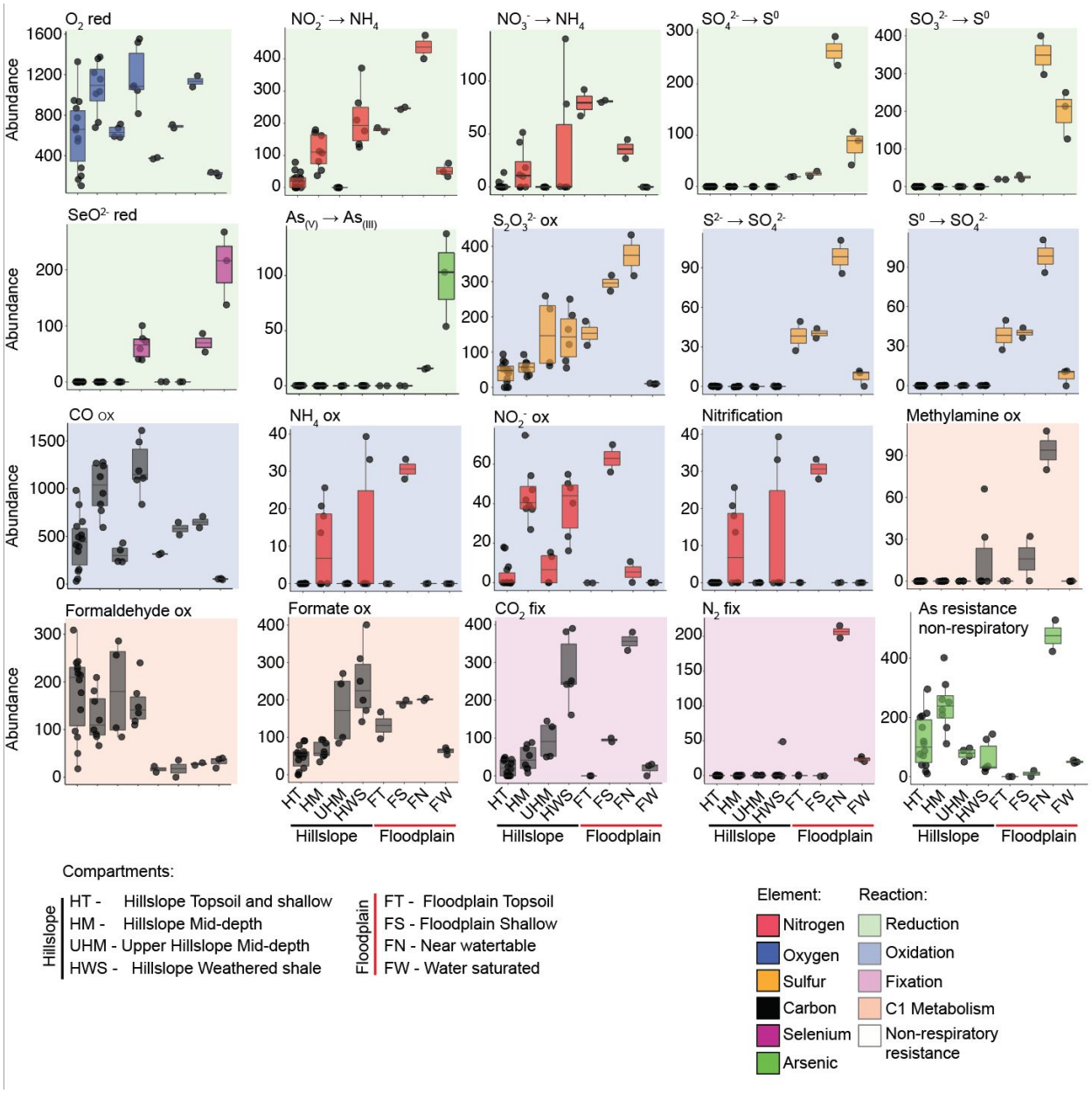

Figure S3: The abundance of metabolic functions. Middle bar of each box plot represents the mean abundance of organisms at this location.

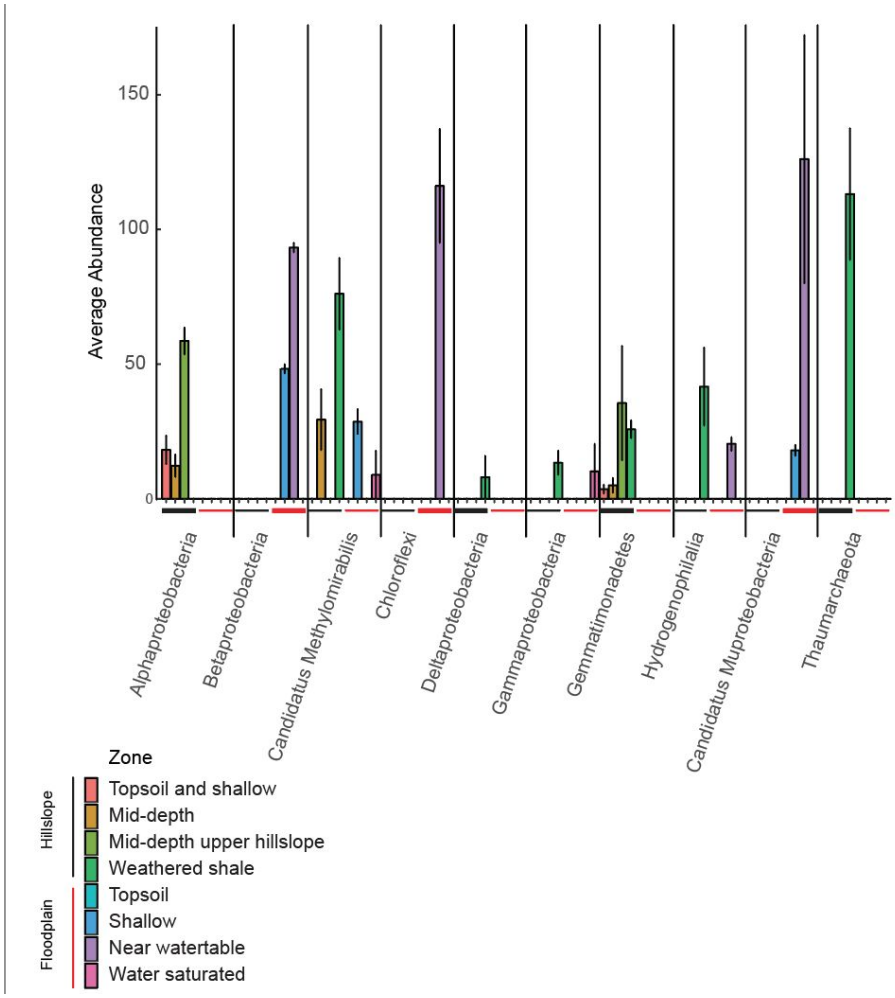

Figure S4: Average abundance of classes and phyla with potential capability for carbon fixation. Bars are SE.

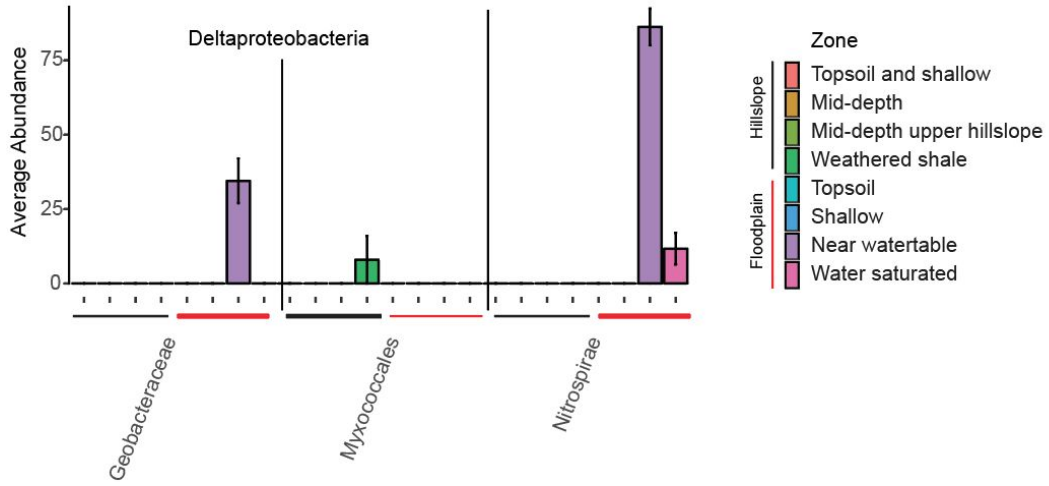

Figure S5: Average abundance of classes and phyla with potential capability for nitrogen fixation. Bars are SE.

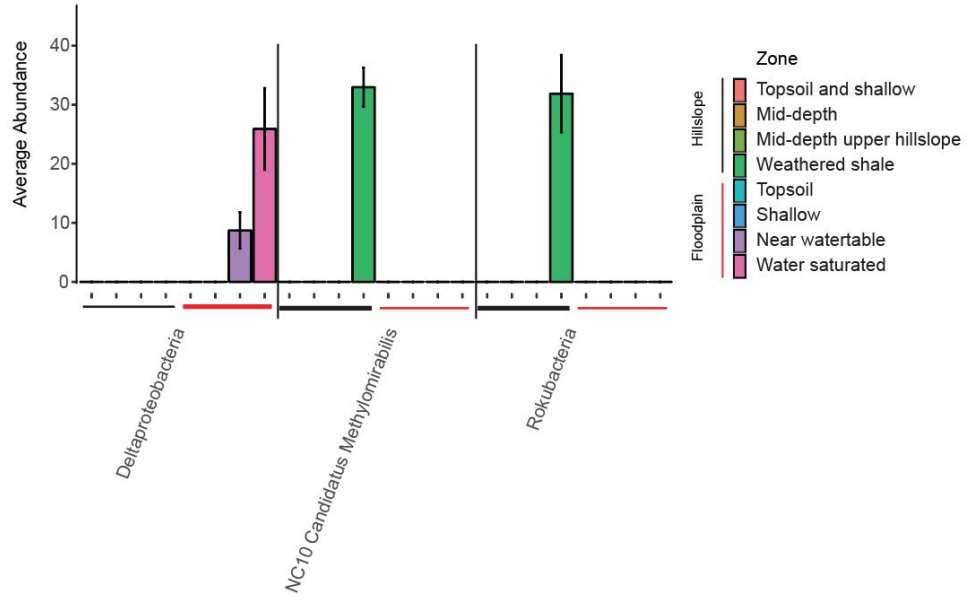

Figure S6: Average abundance of classes and phyla with potential capability for selenate reduction. Bars are SE.

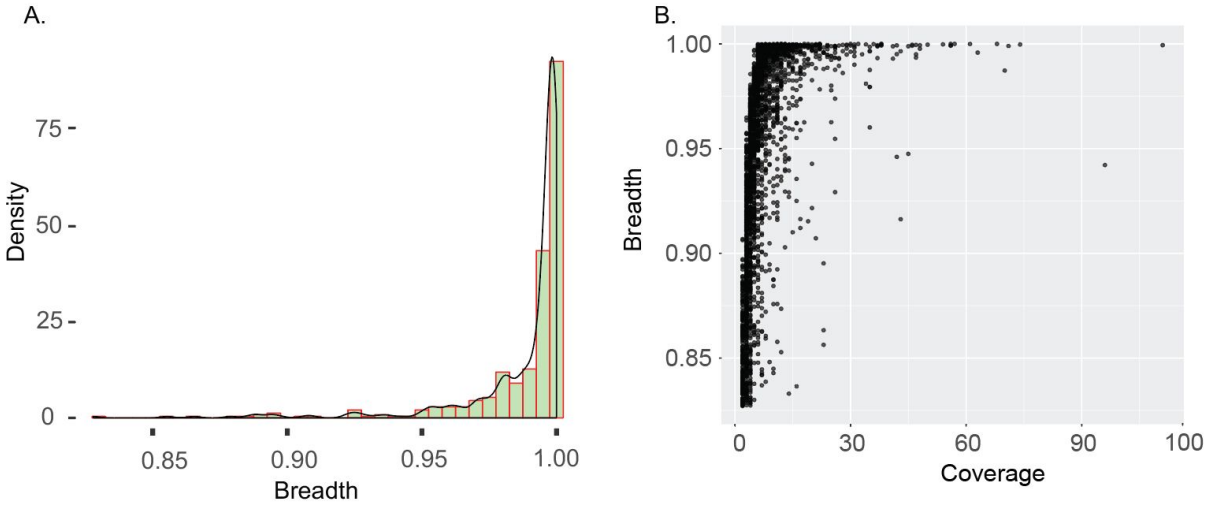

Figure S7: Minimal breadth required in order to consider a genome present in a sample. A. The breadth of genomes as calculated by mapping reads of the sample from which the genome was binned. An 82.8% breadth was found to be the minimum required to obtain a genome from its original sample. B. The breadth and coverage of all genomes passing the 82.8% breadth cutoff are shown. Each dot represents one genome.