

Omnitrophota encompasses diverse and hyperactive nanobacteria: Potential metabolisms and host-dependent lifestyles

Cale Seymour

University of Nevada Las Vegas

Marike Palmer

University of Nevada Las Vegas

Eric Becraft

University of North Alabama

Ramunas Stepanauskas

Bigelow Laboratory for Ocean Sciences <https://orcid.org/0000-0003-4458-3108>

Ariel Friel

University of Nevada, Las Vegas

Frederik Schulz

DOE Joint Genome Institute <https://orcid.org/0000-0002-4932-4677>

Tanja Woyke

U.S. Department of Energy Joint Genome Institute <https://orcid.org/0000-0002-9485-5637>

Emiley Eloë-Fadrosch

DOE Joint Genome Institute <https://orcid.org/0000-0002-8162-1276>

Dengxun Lai

University of Nevada, Las Vegas

Jian-Yu Jiao

Sun Yat-sen University

Zheng-Shuang Hua

University of Science and Technology of China

Lan Liu

Sun Yat-sen University <https://orcid.org/0000-0001-8139-7246>

Wen-Jun Li

Sun Yat-Sen University <https://orcid.org/0000-0002-1233-736X>

Maria Chuvochina

University of Queensland

Brianna Finley

Northern Arizona University

Benjamin Koch

Northern Arizona University <https://orcid.org/0000-0002-1914-7681>

Egbert Schwartz

Northern Arizona University

Paul Dijkstra

Northern Arizona University

Duane Moser

Desert Research Institute

Bruce Hungate

Northern Arizona University <https://orcid.org/0000-0002-7337-1887>

Brian Hedlund (✉ brian.hedlund@unlv.edu)

University of Nevada Las Vegas <https://orcid.org/0000-0001-8530-0448>

Article

Keywords:

Posted Date: March 2nd, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1352086/v1>

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 Nature Microbiology on March 16th, 2023. See the published version at <https://doi.org/10.1038/s41564-022-01319-1>.

Abstract

Candidate bacterial phylum *Omnitrophota* has never been grown in axenic culture and is poorly understood. Here, we combined analysis of 421 *Omnitrophota* genomes representing six classes and 276 species and show that they are prevalent in water, sediments, and soils globally. Fluorescence-activated cell sorting and differential size filtration showed ultra-small ($\sim 0.2 \mu\text{m}$) cells to be common across the phylum. Reduced genomes in all six classes maintained major biosynthetic and energy conservation pathways, particularly the acetogenic Wood-Ljungdahl pathway or diverse aerobic and anaerobic respirations. However, most genomes also encoded multiple systems typical of bacterial predators and intracellular parasites, suggesting possible predatory or parasitic lifestyles. In support of this, quantitative stable-isotope probing revealed three families with high isotope uptake rates comparable to obligate bacterial predators in diverse soils. Based on their ubiquity, small cell size, high metabolic activity, and genomic repertoire, many *Omnitrophota* are likely to be ecologically important in a wide range of ecosystems, possibly as predators or parasites.

Main Text

The bacterial candidate phylum *Omnitrophota* (synonyms: OP3, *Omnitrophica*, *Omnitrophicaeota*) has been observed in 16S rRNA gene surveys globally^{1,2}, particularly in water and sediment environments. Both 16S rRNA gene^{2,3} and genome-based^{2,4-24} studies indicate they belong to the *Planctomycetes-Verrucomicrobia-Chlamydiae* (PVC) superphylum. However, none have been isolated and only two have been microscopically observed: (i) *Candidatus Omnitrophus magneticus* SKK-01^{25,26}, a large magnetic bacterium containing sulfur inclusions; and (ii) “*Candidatus Vampirococcus archaeovorax*” LiM²⁷, a very small ($\sim 0.2 \mu\text{m}$) parasitic coccus. The latter attaches to other prokaryotes^{27,28} and depends on a host cell for optimal metabolism and reproduction, suggesting LiM may have a predatory lifestyle similar to known *Bdellovibrio* and like organisms (BALOs)²⁷. Interpretations of metabolic capabilities of these and other *Omnitrophota* single amplified genomes (SAGs) and metagenome-assembled genomes (MAGs), totaling seven, hinted at heterotrophic aerobic respiration or acetogenesis^{4,6,26,27}. Recent studies also reported interpretations of 14 *Omnitrophota* MAGs from an Antarctic lake and a single MAG from Black Sea sediments and suggested all were obligate fermenters^{5,29}. Although these observations provide valuable insights into individual species, the nature of the *Omnitrophota* as a whole has been unclear, as there has been no systematic attempt to extrapolate these data to the rest of the phylum.

Here, we expanded genomic representation of the phylum to 75 SAGs and 346 MAGs and combined genomic information with cell size and *in situ* activity data to work toward a comprehensive view of the biology of *Omnitrophota*. Our results show small cell size ($\sim 0.2 \mu\text{m}$) to be common across the phylum. Genomes were reduced yet encoded relatively complete biosynthetic and energy conservation pathways. However, most genomes also contained genes associated with parasitism or predation, and diverse *Omnitrophota* were highly active in quantitative stable-isotope probing (qSIP) experiments in diverse soils, similar to obligate bacterial predators in the same datasets. Together, these data suggest potentially

widespread predation and parasitism. This work sets the foundation for a better understanding of this ubiquitous and diverse lineage.

Results And Discussion

***Omnitrophota* genomes from a broad range of biomes comprise six classes and 276 species**

421 genomes classified as *Omnitrophota* were collected from various sources (**Figure 1; Supplementary Text; Supplementary Table 1**). The 75 SAGs and 346 MAGs originated exclusively from environmental biomes, with a plurality from lake or river water (111) or groundwater (97), followed by geothermal sediments (64), bulk soils (59), wastewater (37), and marine or otherwise saline sediments (30).

Systematic organization of the genomes revealed high diversity. Using a 95% average nucleotide identity (ANI) threshold³⁰, the genomes represent 276 putative species clusters, with 204 including a high- or medium-quality assembly according to Genomic Standards Consortium criteria³¹ (**Figure 1b**). The highest quality member of each species cluster was subjected to phylogenetic analysis using four marker sets: the “bacteria-specific” marker sets BAC120³², UBCG³³, BCG110³⁴, and the “universal” marker set UNI56³⁵ (**Supplementary Text; Supplementary Fig. 1-11**). Concordance among the bacteria-specific marker trees was used to revise the GTDB taxonomy^{32,36} to eliminate poly- and paraphyletic taxa across all three conserved marker gene trees. Finally, new ranks between genus and class were proposed by evaluating relative evolutionary distance, implemented in GTDB-Tk, and average amino acid identity (AAI)³⁷ (**Supplementary Fig. 12**). This refined taxonomic classification resulted in a total of six classes, 25 orders, 52 families, 146 genera, and 204 species represented by a medium- or high-quality genome, which is significantly expanded over the previously published 12 Candidatus genera and species in the *Omnitrophota* (**Supplementary Text; Supplementary Fig. 1-11**).

To improve communication, a system of nomenclature was developed based only on high-quality genomic assemblies, with a goal of naming taxa under the nascent SeqCode³⁸ (**Supplementary Text; Supplementary Tables 1-3**). In some cases, special considerations were made to resolve synonymy and retain historical names. A total of 27 species were named using this approach, providing a structure to name four classes: *Vampiricoccoidia*, *Omnitrophia*, *Makaraimicrobia*, and *Aquiviventia* (**Figure 1**). For the two other classes, no full-length 16S rRNA gene could be matched to an otherwise high-quality genome. Thus, these classes retain the alphanumeric designations 2-02-FULL-51-18 and 4484-213. Nine species clusters were phylogenetically unstable and were assigned *incertae sedis* status at the class level.

***Omnitrophota* are prevalent in soils and aquatic environments at low abundance**

16S rRNA gene sequences recovered from *Omnitrophota* genomes were mated with a 16S rRNA gene phylogeny from the SILVA database¹ version 138 (**Supplementary Fig. 13**) to develop a QIIME 2³⁹-compatible Naïve-Bayesian classifier for *Omnitrophota*. Intra- and inter-taxon 16S rRNA gene distance summaries are available in **Supplementary Fig. 14**. Application of the classifier to 25,744 samples from the Earth Microbiome Project (EMP)⁴⁰ release 1 data showed *Omnitrophota* to be common, with 65% of environmental samples containing *Omnitrophota* sequence variants (SV) (**Figure 2**). *Omnitrophota* were especially prevalent in non-saline environments, including waters (70% of EMP samples), sediments (94% of EMP samples), and soils (73% of EMP samples), with enrichment in rhizosphere soils (96% of EMP samples), although typically at low relative abundance (<0.1%). *Omnitrophota* were nearly absent from host-associated samples (**Figure 2**); the rare cases where they were relatively abundant in animal-associated samples were from the alimentary tract of sediment-consuming benthic fishes in the genus *Fundulus* (**Figure 2a**).

The wide distribution of nearly every class of *Omnitrophota* among EMP biomes agrees with the provenance of the genomic assemblies; where *Omnitrophota* are found, multiple classes tend to coexist. In EMP samples, *Vampiricoccoidia* and *Omnitrophia* are the most widely distributed classes, followed by *Makaraimicrobia* and class 2-02-FULL-51-18, with *Aquiviventia* being the least common. Non-saline sediments displayed the lowest taxonomic specificity, with each of the four common classes occurring in >70% of EMP samples. Soils displayed the highest taxonomic specificity; *Omnitrophia*, *Vampiricoccoidia*, and class 2-02-FULL-51-18 occurred in soils at much higher frequencies than *Makaraimicrobia* or *Aquiviventia*. *Omnitrophia*, *Vampiricoccoidia*, and *Makaraimicrobia* occurred in higher relative abundance in anoxic aquatic environments relative to oxic waters. In all, environmental data associated with EMP data and genomic assemblies (**Supplementary Table 1**) suggest a wide physicochemical niche for *Omnitrophota* as ubiquitous members of the rare biosphere.

Most known *Omnitrophota* are nanobacteria

Previous amplicon and metagenomic analyses suggest some *Omnitrophota* cells are small^{41,42}. Here, we report several lines of evidence that small cell size is typical of *Omnitrophota*, regardless of class: (i) 36 SAGs with size measurements taken during FACS⁴³; (ii) two microscopically observed species²⁵⁻²⁸; (iii) 112 MAGs derived from serial-filtered groundwater^{24,44,45} (**Figure 3a**); and (iv) 16S rRNA gene Illumina tag surveys following serial filtration of source water from three springs in the Spring Mountains, Nevada (**Figure 3b-c; Supplementary Table 4**). Cell diameters estimated during FACS for 36 SAGs representing five classes indicated most cells are ~0.2 µm (**Figure 3a**), similar to well-known nano-sized *Patescibacteria* and DPANN archaea⁴³. *Vampiricoccoidia* cells (12/12) and class 2-02-FULL-51-18 cells (4/4) were universally <0.3 µm, as were most FACS-sorted cells belonging to classes *Aquiviventia* (9/10), *Omnitrophia* (3/4), and *Makaraimicrobia* (1/2). Only a few cells of *Aquiviventia* (1/10), *Omnitrophia* (1/4), and *Makaraimicrobia* (1/2) were >0.3 µm. The corresponding SAGs were not contaminated, based on

several marker gene sets (**Supplementary Fig. 6-9**), suggesting these were single, free-living cells $>0.3 \mu\text{m}$, although it is possible they were dividing cells or small aggregates. Similarly, MAGs from serially filtered cells revealed a few species small enough to pass through a $0.2 \mu\text{m}$ filter in the *Vampiricoccoidia* (3), *Makaraimicrobia* (3), and *Omnitrophia* (2) (**Figure 3a**). Like the SAGs, a few MAGs were recovered from larger size fractions ($>0.65 \mu\text{m}$) from across the phylum, including *Makaraimicrobia* (8 MAGs), *Vampiricoccoidia* (5 MAGs), *Omnitrophia* (3 MAGs), *Aquiviventia* (1 MAG), and class 2-02-FULL-51-18 (1 MAG); however, whether these represent single cells $>0.65 \mu\text{m}$ or cells attached to other cells or particulate matter is uncertain.

16S rRNA gene amplicon sequencing of abundant *Omnitrophota* populations from serial-filtered groundwater from Cave Spring, Kiup Spring, and Grapevine Spring, all in the Spring Mountains of Nevada, indicated similar results: all five classes and 13/14 families were more abundant on $0.2 \mu\text{m}$ filters than $0.45 \mu\text{m}$ filters in all springs (**Figure 3b-c**). Together, these results show that cells of all major lineages of *Omnitrophota* are frequently among the smallest known cells.

Genomes are reduced in size, but biosynthetic pathways are conserved and complete

Prokaryotes with small cell size typically have genomes that are reduced in both size and function. The predicted genome sizes of *Omnitrophota* range between 1 and 3.5 Mb (**Figure 1a**, **Figure 4** and **Supplementary Fig. 15**). Optimal growth temperatures (OGT) of free-living bacteria are negatively correlated with genome size⁴⁶ and positively correlated with IVYWREL content of proteins they encode⁴⁷. When provided with OGT predicted from the IVYWREL content of *Omnitrophota* genomes, a linear model trained on OGT from 829 free-living bacterial species⁴⁸ predicted significantly larger genome sizes ($p < 2.2 \cdot 10^{-16}$, paired t-test) than observed (**Supplementary Fig. 15**).

Despite overall genome size reductions, no single Clusters of Orthologous Groups (COG) category was reduced in richness or percentage in *Omnitrophota* genomes compared to PVC genomes (**Supplementary Fig. 16**), and biosynthetic pathways were generally complete (**Supplementary Fig. 17, 18**); instead, genes not mapping to COGs were significantly reduced in both richness and percentage ($p < 0.05$, anova and post-hoc Tukey's HSD) (**Supplementary Fig. 16**). This pattern results in streamlined genomes that retain most or all genes essential for a free-living lifestyle, including energy conservation.

Classes of *Omnitrophota* encode either acetogenic or respiratory potential

Genetic potential for energy conservation was ubiquitous in *Omnitrophota* genomes, and each class tended to follow either a respiratory or acetogenic scheme (**Figure 4, 5; Supplementary Fig. 18, 19**). Genes suggestive of acetogenesis dominate in *Makaraimicrobia*, *Vampiricoccoidia*, and class 4484-213. These genomes encode acetate kinase (ack) and phosphotransacetylase (pta) (**Figure 4, 5a, and Supplementary Fig. 20**), the combination of which serves to catalyze the ATP-yielding hydrolysis of acetyl-CoA to acetate⁴⁹. Most ack- and pta-encoding genomes also encode components of the Wood-Ljungdahl pathway; a formate-tetrahydrofolate ligase (K01938) was common, but never a formate dehydrogenase (K05299, K15022) (**Supplementary Table 5**), suggesting that many *Omnitrophota* lack the capability for autotrophic fixation of CO₂ and instead may incorporate formate⁵⁰ via the methyl branch of the Wood-Ljungdahl pathway (**Figure 4, 5a**). These enzymes can also catalyze the reverse reaction, acetate to acetyl-CoA, and can serve to ligate coenzyme A to propionate or other short-chain fatty acids (SFCA)⁵¹. Direct utilization of propionate via acetate kinase (**Figure 5a**) and phosphate transacetylase is therefore feasible in some *Omnitrophota*. This interpretation is consistent with the enrichment of *Omnitrophota* in an anaerobic reactor community fed propionate at a high dilution rate⁵². All *Omnitrophota* 16S rRNA phylotypes in the reactor community mapped to the order *Profunditerraquicolales* within the *Vampiricoccoidia* (Genomes 318-418, **Figure 1, Supplementary Table 1**), which encodes this pathway (**Figure 4**). The genomic potential to fix carbon monoxide into acetyl-CoA⁵³ (K00194, K00197, K00198, K00297, K01938, K14138, K15023) (**Figure 4, 5a; Supplementary Table 5**) is common among the *Vampiricoccoidia*, however genes encoding carbon monoxide dehydrogenase (K00192, K00195, K00196, K00198) (**Figure 4, 5a; Supplementary Table 5**) are rare. These results suggest that species within the *Makaraimicrobia* and *Vampiricoccoidia* may also be capable of directly fixing carbon monoxide via the carbonyl branch of the Wood-Ljungdahl pathway.

These genomes also encode a highly conserved Rnf complex. When used by acetogenic bacteria, Rnf complexes serve to restore NAD⁺ and oxidized ferredoxin pools⁵⁴, or in reverse, to generate an electrochemical gradient capable of powering an ATPase⁵⁵. These same genomes also frequently encode genes containing 4g NiFe membrane-bound hydrogenase domains (**Figure 4, 5a, Supplementary Fig. 21**). The exact role of group-4g hydrogenases remains unexplored, but hydrogenases from this group have been proposed to couple the oxidation of H₂ or methylated organic compounds to ferredoxin reduction for CO₂ fixation⁵⁶. Other group-4 hydrogenases reverse this process, however, and couple the production of H₂ to the oxidation of carbon monoxide or formate to CO₂⁵⁷. Which, if any of these roles are filled by these genes is currently unclear. Nevertheless, the frequent presence of a putative hydrogenase, acetate kinase, and phosphotransacetylase, and near-complete Wood-Ljungdahl and gluconeogenesis pathways (**Supplementary Table 5**), together suggest that most *Makaraimicrobia*, *Vampiricoccoidia*, and class 4484-213 conserve energy via acetogenesis. Exceptions to this pattern were the *Vampiricoccoidia* order *Fredricksoniimonadales* (Genomes 238-314, e.g., *Fredricksoniimonas* spp. **Figures 1, 4**) and the *Makaraimicrobia* order *Pluralincolimonadales* (e.g., *Pluralincolimonas frigidipaludosa*; Genomes 109-114; **Figures 1, 4**), which instead encode components suggestive of diverse anaerobic respirations or, in the case of some *Fredricksoniimonadales* species (Genomes 238-314), a facultatively anaerobic lifestyle.

Genomes in the classes *Omnitrophia*, *Aquiviventia*, and class 2-02-FULL-51-18 suggest diverse respiratory capabilities. These lineages lack the Rnf complex and Wood-Ljungdahl genes possessed by their counterparts in the other lineages, instead encoding an array of electron-transport chain components. Genes encoding the 49 kDa subunit of respiratory complex I, NADH dehydrogenase⁵⁸ (PF00346) (**Supplementary Table 5**), are present in these three lineages plus some genomes in the orders *Fredricksoniimonadales* (Genomes 238-314) and *Pluralincolimonadales* (Genomes 109-114). However, genes encoding the rest of the NADH dehydrogenase complex (M00144) (**Figure 4**) and succinate dehydrogenase (M00149) (**Figure 4**) are only present in the *Omnitrophia*, class 2-02-FULL-51-18, *Aquiviventia*, and *Fredricksoniimonadales* (Genomes 238-314). This complex is also rarely complete; genes encoding the membrane anchor component of complex II (K00242, K18859, or K18860) are missing from all *Omnitrophota* except members of *Aquiviventia* (**Supplementary Table 5**). Genes encoding the succinate dehydrogenase complex catalytic subunits are nevertheless present in many *Omnitrophia* and class 2-02-FULL-51-18, suggesting this pathway may be functional. Still, respiratory complexes II and III are frequently incomplete, which suggests either functional reduction or a propensity toward a simplified respiratory chain like those observed in some oligotrophic organisms⁵⁹. Central carbon metabolism, including largely complete glycolytic pathways and partial tri-carboxylic acid pathways are discussed in **Supplementary Text**.

While some genes encoding respiratory complexes and central carbon metabolism tend to be conserved in the respiratory *Omnitrophota*, potential terminal electron acceptors vary. Members of *Aquiviventia*, class 2-02-FULL-51-18, and some *Omnitrophia* possess genes encoding the cytochrome c oxidase complex (M00155) (**Figure 4, 5**), the low-affinity, oxygen-reducing terminal oxidase, suggesting these lineages may be aerophilic. One species of *Omnitrophia* and some members of the order *Fredricksoniimonadales* (Genomes 238-314) possess a near-complete cytochrome bd ubiquinol complex (M00153) (**Figure 4, Supplementary Table 5**), the high-affinity complex favored under low oxygen concentrations⁶⁰. The presence of cytochrome bd ubiquinol complexes in *Fredricksoniimonadales* (Genomes 238-314) suggests they are microaerophiles or facultative anaerobes.

Alternative terminal oxidases in the respiratory lineages suggest a variety of options for terminal electron acceptors. A few *Omnitrophia* genomes encode a nickel-iron group-4f hydrogenase. While the precise function of group-4f hydrogenases is unclear, they are predicted to couple formate oxidation to hydrogen production⁵⁷. Denitrification genes are also encoded by some *Omnitrophia*, *Aquiviventia*, and class 2-02-FULL-51-18, although none encode a complete denitrification pathway (**Figure 4, Supplementary Table 5**), suggesting syntrophic denitrification with other organisms. Genomes of some *Omnitrophia* encode homologs of periplasmic cytochromes thought to be used by *Desulfovibrio ferrophilus* for dissimilatory metal reduction^{61,62}. Conductive pilins have been observed to facilitate direct electron transfer between syntrophic partners or to mineral surfaces⁶³. Putatively conductive pilins⁶⁴ are predicted sparsely across *Omnitrophia*, *Makaraimicrobia*, and *Vampirococcoidia*. However, known metal reduction pathways that would imply the use of these pilins for metal respiration are much less conserved, especially in the *Vampirococcoidia*.

Genes encoding components of the F-type ATPase complex (KEGG:M00157) are common (**Figure 4, 5b**). One set of these genes is mostly vertically transferred (**Figure 4, 5b, Supplementary Fig. 22, 23**). Aside from its predicted role in respiration, many homoacetogens make use of this ATPase; concurrent with its function to restore oxidized ferredoxin pools, the Rnf complex could generate a Na⁺ gradient⁵⁵, which is then harnessed by the F₀F₁ ATPase to generate ATP⁶⁵. The presence of F₀F₁ ATPase genes and other ETC complexes encoded by the *Omnitrophia*, *Aquiviventia*, and class 2-02-FULL-51-18 might represent conserved potential for oxidative phosphorylation. Further information on the predicted carbon metabolism of *Omnitrophota* are available in **Supplementary Text**.

Genomic data suggest predatory or parasitic functions

Despite near-complete biosynthetic and energy conservation capacity, genes that suggest roles in predatory or parasitic interactions are common in *Omnitrophota* genomes. However, these genes differ among classes in predicted function and origin, suggesting different forms of predation or parasitism. This seems particularly likely in the *Vampirococcoidia* given the known predatory lifestyle of “*Candidatus Vampirococcus archaeovor*” LiM, herein renamed *Vampirococcoides archaeovor* LiM²⁷ to avoid synonymy⁶⁶, and the general conservation of genes associated with that lifestyle in *Vampirococcoidia*. In all BALOs, the tight-adherence (Tad) complex is used to attach to and/or enter host bacterial cells⁶⁷, and a similar function was proposed for the Tad complex of *Vampirococcoides archaeovor* LiM⁷ (Genome 417, GCA_004102945.1, CP019384.1 locus 12382-29832). However, Tad complexes are found not only in the *Vampirococcoidia*, but also in some genomes in the *Makaraimicrobia*, *Aquiviventia*, *Omnitrophia*, and the single medium-quality genome of class 4484-213 (**Figure 4, 6, and Supplementary Fig. 20**). Phylogenetically, TadB (**Supplementary Fig. 24**) and TadC (**Figure 6, Supplementary Fig. 25**) from *Vampirococcoidia*, *Aquiviventia*, and class 4484-213 form a monophyletic cluster with homologs from members of the *Bdellovibrionales*, specifically *Halobacteriovorax marinus*, *Bacteriovorax stolpii*, and an unclassified species of *Bdellovibrio*. RcpA proteins from these three classes, encoding the largest component of the dedicated multimeric outer membrane secretion channel, are not directly adjacent to those from *Bdellovibrionales*, but nonetheless form a supported group with a cluster of homologs from the predatory genera *Stigmatella*, *Vulgatibacter*, and *Lysobacter* (**Supplementary Fig. 26**). These genes, and TadZ (**Supplementary Fig. 27**), were also found in *Omnitrophia* and class 2-02-FULL-51-18 but they were phylogenetically distant, suggesting two separate origins for the Tad complex in *Omnitrophota*. Although the frequency of Tad complexes in *Omnitrophota* as a whole is not higher than in the related phylum *Planctomycetes* (**Supplementary Table 6**), the near-universal presence of Tad complexes in *Vampirococcoidia* and their phylogenetic affiliation with those from known bacterial predators suggests a common role in attachment, and possibly entry into, other cells.

Vampirococcoides archaeovor LiM also expresses type-4a pili²⁷. Type-4a pili can function in cell-cell attachment and are considered necessary for epibiontic predatory lifestyles⁶⁸, or alternatively for

twitching or gliding motility. Regardless of taxonomic affiliation, all *Omnitrophota* genomes encode multiple copies of a type-4a pilus complex (**Figure 4, 6; Supplementary Table 6, 7**), similar to known BALOs and at much higher frequencies than non-predatory bacteria (**Supplementary Table 6**). In many cases, including *Vampirococcoides archaeovor* LiM, a second, non-respiratory F-type ATPase with a distinct evolutionary history (**Figure 6c, Supplementary Fig. 22, 23**) is found adjacent to one of the type-4a pilus gene clusters (**Figure 4, 6b**). The α -subunit of this second ATPase is phylogenetically separate from prototypical ATP-fixing genes, instead neighboring the type-2 and type-3 F_0F_1 homologs used by several pathogenic species of *Mycoplasma* to power gliding motility on the surfaces of eukaryotic cells^{69,70}. The *Omnitrophota* β -subunit forms a supported cluster with two proteins from *Candidatus Saccharimonadia* (TM7)^{71,72} and one from “bacterium AB1_lowgc”⁷³, a relative of *Candidatus Dependientiae* (TM6)⁷⁴ with an extremely small genome⁷³. These relationships suggest *Omnitrophota* and other nanobacteria may use a “parasitic” F-type ATPase complex to power type-4a pilus-mediated gliding motility in the context of predation or parasitism.

Omnitrophia and *Aquiviventia* genomes also often encode an ATP/ADP translocase (K03301) (**Figure 4, 6; Supplementary Fig. 28**), as does one species of *Makaraimicrobia*, *P. frigidipaludosa* (Genome 114). Translocases of this type are used by intracellular parasites in the *Rickettsiae* and *Chlamydiae* to import cytoplasmic ATP while invading eukaryotic cells⁷⁵; however, ATP/ADP translocases are also common in the genomes of BALOs and may serve a similar function (**Supplementary Table 6**). Within the *Omnitrophota*, this translocase from *Omnitrophia* and *Aquiviventia* forms a well-supported, monophyletic cluster, with the single sequence recovered from *P. frigidipaludosa* grouping separate (**Supplementary Fig. 28**). The *Omnitrophia* and *Aquiviventia* ATP/ADP translocases group with those from *Flavobacteriaceae*, a lineage that includes intracellular parasites of *Acanthamoebae*⁷⁶, albeit without strong support. The sequence from *P. frigidipaludosa* forms a supported group with a putative ATP/ADP translocase from *Candidatus Babelia massiliensis*⁷⁷, an obligate intracellular parasite of *Acanthamoebae* belonging to “*Candidatus Dependientiae*”⁷⁴, and both are basal to known *Chlamydiae* ATP translocases. The neighboring taxa of both lineages of *Omnitrophota* ATP/ADP translocases contain facultative intracellular parasites of amoebae, suggesting that some members of *Omnitrophia*, *Aquiviventia*, and *P. frigidipaludosa* may be intracellular parasites of micro-eukaryotes.

Giant open-reading frames (ORFs) are also present in many *Omnitrophota* genomes, including *Vampirococcoides archaeovor* LiM, and may further facilitate cell adhesion (**Figure 4, 6**). Giant ORFs are uncommon but widely distributed in bacteria and archaea. A search of two different complete-genome databases of isolated bacteria and archaea^{48,78}, together representing 1,500 genomes, revealed fewer than 14% with an ORF longer than 20 kb (6,666 aa) (**Supplementary Table 7**), whereas at least 45% of *Omnitrophota* genomes contain such genes (**Figure 4, Supplementary Fig. 29, 30**). More striking, 137 (~36.1% of 379 total) of all *Omnitrophota* assemblies encode a complete ORF longer than 45 kb (15,000 aa) or a partial ORF longer than 22.5 kb (7,500 aa). These ORFs are distributed across *Omnitrophota*; 71 (~37.2% of 191 total) of *Omnitrophota* species groups included at least one assembly with a giant ORF, with a further 26 species lacking a complete giant ORF but possessing a partial ORF homologous to a

large ORF from another species (psiblast^{79,80} e-value < 1e-30 and pident > 50% on a truncated ORF longer than 1,500 aa over an alignment region of at least 250 aa). Most species possess only one giant ORF, but some genomes encode multiple. The species with the highest total number of giant ORFs is a member of *Omnitrophia*, *Danuiimicrobium aquiferis*, for which five separate genomes (Genomes 181-185) are available with up to seven distinct giant ORFs in a single genome. Depending on the assembly, these giant ORFs accounted for between 220 and 357 kilobases, amounting to roughly 10 percent of the ~3 Mb genomes of this species. Overall, 212 giant ORFs (>20 kbp) and 312 partial ORFs homologous to giant ORFs were observed across the *Omnitrophota*. This is necessarily an underestimation of the true number and size of giant ORFs across *Omnitrophota* due to the fragmented and incomplete nature of the SAGs and MAGs used as well as the strict criteria for inclusion. No catalytic RNA domains were observed within the giant ORFs, suggesting they may be transcribed as a single mRNA. Despite these ambiguities, some ORFs exceed 100 kb (33,333 aa), the longest being a 140,576 nucleotide (46,858 aa) partial ORF encoded by a *Vampiricoccoidia* species, UBA10174 sp003528115 (Genome 270, GCA_001804395.1 locus MHFX01000013.1 3-140576). This is the largest prokaryotic ORF reported to date. If fully transcribed and translated, the product of this ORF would dwarf the largest documented product of the human titin gene (~36,000 aa).

Other prokaryotic genomes encode giant ORFs for secreted hemolysins, adhesins, or non-ribosomal polypeptide synthesis proteins⁷⁸. Many *Planctomycetes* genomes also encode giant ORFs that are also speculated to mediate cell adhesion. The utility of large ORFs to *Omnitrophota* is unclear because they encode few annotated protein domains. Most are predicted to code for transmembrane helices with large predicted extracellular domains, suggesting many encode surface-presented proteins. However, annotated catalytic and structural domains are extremely variable, even between members of the same genus. Fast evolution of surface-presented proteins is well-known in pathogens and other bacteria to evade host detection systems and viral attachment sites. Giant ORFs from *Omnitrophia* frequently encode domains used by eukaryotes for cell adhesion^{81,82}. For example, ORFs from *Multiplicimicrobium inquinatum* (Genome 171) and *Omnitrophus fodinae* (Genome 190) encode discoidin (PF00754) domains and ORFs from *M. inquinatum* and SAG-3300015153 (Genome 196) encode laminin_G_3 (PF13385) domains (**Supplementary Fig. 29**). Similarly, an ORF from *Aquinicolibacterium lacustre* (Genome 157) encodes a giant ORF with nine noncellulosomal dockerin (PF00404) domains that may serve as adhesins⁸³ (**Figure 6**). ORFs of similar character, albeit smaller size, were recently found in the genomes of parasitic/symbiotic members of the *Patescibacteria*⁶⁶ and *Nanohaloarchaeota*⁸⁴ and suggested to serve as adhesins to attach to host cells and form pores in the S-layer or membrane to gain access to the cytoplasm. Large ORFs with a putative role in adhesion are also present in the genome of *Chlorobium chlorochromatii*⁸⁵, a symbiont and epibiont of *Candidatus "Symbiobacter mobilis"*⁸⁶. The implication that *Omnitrophota* may use giant proteins for adhesion to and or form pores in other cells suggests a broader context for giant ORFs in the otherwise reduced genomes of bacterial symbionts and parasites.

Taken together, the completeness of major biosynthetic and energy conservation pathways, combined with several systems suggesting parasitism or predation, suggest complex lifestyles, possibly involving shifts between mutualism, predation, and parasitism might be common in *Omnitrophota*, as broadly suggested for *Patescibacteria* and DPANN archaea⁸⁷. Given the conservation of genes related to predation in most *Vampirococcoidia*, we suggest most *Vampirococcoidia* may be epibiotic predators similar to known BALOs, as exemplified by *Vampirococcoides archaeovorius*^{27,28}. *Vampirococcoides archaeovorius* LiM was observed attaching to and predated on a variety of cells in a methanogenic limonene-degrading culture, similar to BALOs^{27,28}. However, the LiM cells were also frequently found as free unattached cocci but with low metabolic activity and no apparent cell replication. This observation, and the biosynthetic capacity and syntrophic energy conservation pathways encoded by *Vampirococcoidia* genomes, including *Vampirococcoides archaeovorius* LiM, suggest these organisms might persist in nature as individual cells, especially given the low maintenance energy associated with small cell size. This interpretation is consistent with our recovery of 31 SAGs from groundwater and anoxic lakes via FACS exclusively as small cells, with no co-sorts of *Vampirococcoidia* or any other *Omnitrophota* with other cells. Which, if any, *Omnitrophota* have specific interactions with host species or genera is unknown; however, *Aquiviventia* genomes shared a disproportionate fraction of genes of actinobacterial origin, especially *Streptomycetaceae*, suggesting a possible partnership (**Supplementary Fig. 32**).

Despite the abundance of genes suggesting predation or parasitism within *Omnitrophota*, small groups of genomes across the *Omnitrophota* seem to lack such genes (**Figure 4**), suggesting these may be free-living organisms. With few exceptions, the *Makaraimicrobia* lack Tad or ADP/ATP translocase genes (**Figure 4**), suggesting this class may be a primarily free-living exception to the predatory or parasitic model we present here for *Omnitrophota*. In support of this “*Candidatus* *Omnitrophus magneticus* SKK-01”^{25,26}, herein renamed *Thioglobulicalix magnus* to avoid synonymy, is the only confirmed large, free-living species in the phylum *Omnitrophota*, and many *Makaraimicrobia* were retained on large size filters (>0.65 µm; **Figure 3**), suggesting large cell size might be common in *Makaraimicrobia*, although serial filtration and FACS suggested both large and small cell sizes in the class.

High stable isotope incorporation extended to *Omnitrophia* and class 2-02-FULL-51-18

Bacterial predators and parasites have been shown to incorporate labels in stable isotope probing experiments faster than those with other feeding behaviors⁸⁸. This trend is magnified among obligate predators. Our qSIP experiments in three geologically distinct aerobic soils revealed high intrinsic ¹⁸O-H₂O incorporation rates by the families *Aquicolibacteriaceae* and *Taenariivivaceae* (class *Omnitrophia*) and the family 2-02-FULL-51-18 (class 2-02-FULL-51-18) (**Figure 7**). Since the ¹⁸O from water exchanges with ¹⁶O atoms in free nucleotide and nucleoside pools, but not DNA, these high ¹⁸O rates are consistent with high rates of DNA synthesis and/or high rates of consumption of biomass from labeled cells, either

through predation, parasitism, or necromass consumption. Relative ^{18}O incorporation rates from these *Omnitrophota* were not significantly different from BALOs in the *Bdellovibrionales* and *Vampirovibrionales* in the same dataset ($p > 0.05$, anova with post-hoc Tukey's HSD test) but were significantly higher than the facultative predators *Lysobacter*, *Myxococcales*, and *Streptomycetaceae*, and free-living bacteria *en masse* ($p < 0.05$, anova with post-hoc Tukey's HSD test, **Supplementary Fig. 31**). Family 2-02-FULL-51-18 also assimilated ^{13}C -labeled glucose and oxalate at high rates, although long incubation times and high soil community complexity complicate interpretation of carbon source utilization.

These high isotope incorporation rates, along with those previously observed in Black Sea sediments²⁹, document high metabolic activity and/or consumption of labeled cell mass in four of the six classes of *Omnitrophota* in both anoxic sediments (Black Sea sediments²⁹ hosting highly labeled *Vampiricoccoidia* and *Makaraimicrobia*) and oxic soils (diverse soils hosting highly labeled *Omnitrophia* and class 2-02-FULL-51-18). We note that *Vampiricoccoidia* and *Makaraimicrobia* were present at very low abundance in the soils studied here and they did not incorporate the stable isotopes, as would be expected for these mostly anaerobic lineages. Similarly, *Aquiviventia* were not present in the soils, as they are rare and typically inhabit aquatic systems.

Summary

The *Omnitrophota* are a diverse and ubiquitous bacterial phylum that shares genomic capacity for a free-living chemoheterotrophic lifestyle with genomic markers often observed in parasites and predatory bacteria, including: (i) tight-adherence (Tad) complexes used for attachment and penetration by BALOs; (ii) type-2 and type-3 F_0F_1 -type ATPase homologs and co-located type-4a pili used for surface motility by *Mycoplasma*; (iii) ATP/ADP translocases used to deplete infected cells of intracellular ATP stores by *Chlamydia*; and (iv) giant ORFs $\geq 45,000$ amino acids, as has been observed in some parasitic *Patescibacteria*⁶⁶ and *Nanohaloarchaeota*⁸⁴. The environmental distribution of *Omnitrophota* precludes them from being pathogens of macroscopic hosts, suggesting other microorganisms as likely hosts.

If many *Omnitrophota* are predatory or parasitic, at least during some stages of a complex lifestyle, then potential hosts and mechanisms of parasitism may be lineage-specific. Given the known BALO lifestyle of *Vampiricoccoides archaeovorius* LiM and highly conserved nature of predation-related genes in *Vampiricoccoidia*, we suggest that a BALO lifestyle may be common in that class. Conversely, the variable presence of predation-related genes and the specific presence of ATP/ADP translocases suggest many *Omnitrophia* and *Aquiviventia* may fulfill a role as intracellular parasites. The high isotope incorporation rates in *Vampiricoccoidia*, *Omnitrophia*, and class 2-02-FULL-51-18 are consistent with these interpretations. Direct experimental evidence will be needed to test this intriguing hypothesis in future studies.

It is important to note that another group of nanobacteria that is widely resistant to laboratory cultivation – the *Patescibacteria* – has been interpreted as predominantly obligate parasites⁸⁹ or as free-living cells

that lack respiration⁴³. Key to the latter interpretation was the lack of genes encoding electron transport systems and the lack of specific cell-cell associations detected by FACS during integrated FACS and single-cell genomics. Integrated FACS and single-cell genomics pipelines have been used successfully to detect specific associations between *Nanoarchaeota* and archaeal hosts^{90,91} in geothermal environments and between Candidatus *Saccharibacteria* and actinobacterial hosts^{92,93}. The absence of evidence for cell-cell associations between *Omnitrophota* and any other species may either suggest that *Omnitrophota* cell-cell interactions are weak, they mostly free-living, or that *Omnitrophota* persist in the environment as free-living nanobacteria that are either facultatively predatory/parasitic or obligately predatory/parasitic but with a free-living phase, which has been suggested for *Vampirococcoides archaeovorans* LiM^{27,28} and broadly for *Patescibacteria* and DPANN archaea^{66,84}. Either possibility is consistent with the largely complete biosynthetic and energy conservation potential of *Omnitrophota*.

Knowledge of potential energy conservation pathways and possible lifestyles conserved within the *Omnitrophota* may contribute to their future cultivation and study, whether or not they eventually yield axenic or defined co-cultures. We believe the framework provided by this study promotes more incisive studies of the biology of these organisms and the roles they play in aquatic and terrestrial biomes globally.

Methods

MAG and SAG sources.

MAGs were obtained from metagenomes from a variety of geothermal springs in Tengchong, China or from data repositories. SAGs were obtained from a variety of locations as described previously⁴³. The generation, identification, sequencing and de novo assembly of SAGs was performed at the Bigelow Laboratory for Ocean Sciences Single Cell Genomics Center-SCGC (scgc.bigelow.org), with cell sizes estimated using calibrated index FACS⁹⁴. Genomes used in the study, their provenance, and associated methods are summarized in **Supplementary Table 1** and **Supplementary text**.

MAG and SAG QC, species delineation, and genome-based phylogeny.

The quality of each genome was assessed using CheckM⁹⁵. Genomes greater than 50% complete and less than 10% contaminated were considered medium quality of those greater than 90% with less than 5% estimated contamination, and presence of the 23S, 16S, and 5S rRNA genes and at least 18 tRNAs as high quality according to Bowers et al³¹. Those 10-50% complete and less than 5% contaminated were considered low quality but were retained to be phylogenetically placed. Those of less than 10% completeness or greater than 10% contamination were omitted. This yielded a total of 73 SAGs and 343 MAGs. Genomes were grouped into species based on their membership within single-linkage clusters

based on a threshold of 95% average nucleotide identity (ANI) according to FastANI³⁰. Where a high or medium quality representative was available, the most complete genome in each cluster was used for phylogenetic analysis. Clusters were classified to the genus level if their members included a genome present in the Genome Taxonomy Database (GTDB)³². Genomes lacking a representative in GTDB were classified using Genome Taxonomy Database Toolkit (GTDB-Tk)³⁶ with the r202 reference. The classification step of GTDB-Tk was modified so that placement was constrained to the subtree one node above the MRCA of *Omnitrophota* (p__Omnitrophota) and *Ratteibacteria* (r202: p__Ratteibacteria). Pplacer⁹⁶ was then used to place genomes on the subtree using the same reference package model parameters as the full tree but using the reduced alignment. The subtree was then trimmed by one node and then grafted to replace the MRCA of the *Omnitrophota* and *Ratteibacteria* on the reference tree. These steps were taken to reduce the memory requirement for phylogenetic placement onto the GTDB-Tk reference tree and to exclude genomes outside of these two lineages. The classification step then proceeded as normal using the subtree-grafted reference tree. Conserved marker gene alignments were used to construct the phylogeny of *Omnitrophota*. The full-length bac120 alignment³² was obtained from GTDB-Tk. The bcg110 marker set was generated using bcgTree 1.1.0³⁴. An additional alignment of 56 universally conserved COGs³⁵ were identified using hmmer 3.3.1⁹⁷. Up-to-date bacterial core genome (UBCG) marker genes were identified and aligned using the UBCG software³³ version 3. UBCG and Bac120 marker alignments were reduced using the gappyout function of trimal⁹⁸ v1.4.rev22. Phylogenetic trees were constructed from reduced alignments using IQ-Tree⁹⁹ version. Individual gene trees were aligned using mafft¹⁰⁰ 7.453, then reduced using gappyout function of trimal⁹⁸. Substitution model testing was performed using IQ-Tree's ModelFinder¹⁰¹, restricted to WAG, LG, JTT, JTTDCMUT, and PMB. Node support was based on 1000 SH-like aLRT (alrt) test and 1000 ultrafast bootstrap¹⁰² rounds. Bac120 alignment sequences with greater than 1000 residues in the alignment corresponding to retained low-completeness assemblies were masked using the same filter produced by gappyout. These sequences were then placed onto the Bac120 marker set tree using epa-ng¹⁰³ v0.3.8 using a custom script to generate a reference package from IQ-tree outputs. The GTDB taxonomy was refined based on the Bac120, UBCG, and BCG marker set trees to find concordance between topologies. AAI was calculated between species using the procedure from "aai.rb" of the enveomics¹⁰⁴ script collection, reimplemented in the R programming language. Genus-level assignments were modified or added where no taxonomy existed to produce consistent intra- and inter-genus AAI values in each family. All trees were rendered in R using the ape¹⁰⁵, phytools¹⁰⁶, and ggtree¹⁰⁷ software packages.

Genome annotation.

Gene functional annotation was performed using kofamscan (DB version dated 2020-02-02)¹⁰⁸, txsscan^{109,110} METABOLIC¹¹¹, FeGenie⁶¹, TMHMM¹¹², and hmmer⁹⁷. The hmmsearch function of hmmer 3.3.1 was used to annotate a protein sequence file created from the concatenated sequences of all *Omnitrophota* genomes, using the noise cutoff (NC) as the threshold for a positive result. Profiles used

for annotation were from the PFAM database version 32^{113,114}, TIGRFam¹¹⁵ version 15, METABOLIC¹¹¹, and FeGenie⁶¹. 16S and 23S rRNA gene sequences were identified and extracted using Metaxa2¹¹⁶ version 2.2.1. COGs were annotated using rpsblast version 2.9.0^{79,80} against COG position-specific scoring matrices provided by the NCBI CDD database with an evaluate threshold of 1e-2. tRNAs were annotated using tRNAscan-SE¹¹⁷ version 2.0.7. Other RNA motifs were annotated using the Rfam database¹¹⁸ version 14.5 with Infernal¹¹⁹ version 1.1.4.

Inference of genome size and reduction.

Overall genome reduction was inferred from the discrepancy between genome size estimates based on genome quality and those modeled from optimal growth temperature (OGT). In free-living bacteria, OGT is negatively correlated with genome size⁴⁶. A regression model was constructed using the OGT of bacteria in the Tempura database⁴⁸. Optimal growth temperature was estimated from the IVYWREL content of *Omnitrophota* genomes using the regression model calculated by Zeldovich et al.⁴⁷. Predicted OGT of *Omnitrophota* and provided to the model to generate a prediction of genome size based on OGT. Completed genome size was estimated for each *Omnitrophota* genome using the formula: $Size_{est} = (Size_{obs} - (Size_{obs} * contamination)) / (completeness)$, where $Size_{obs}$ is the total size of the assembly and contamination and completeness are genome quality estimates calculated from single-copy marker genes in a general bacterial marker set excluding markers systematically absent from the phylum. Genome sizes predicted according to genome quality and genome sizes modeled via OGT were compared using a paired t-test using an alternative hypothesis that the OGT-modeled genome size is greater than the quality-estimated genome size. The same procedure, including the prediction of OGT using IVYWREL-content, was repeated for the reference genomes upon which the OGT model was based.

16S rRNA gene phylogeny and Qiime2 classifier construction.

Omnitrophota 16S rRNA genes were aligned against the SILVA 99-percent identity nonredundant 16S rRNA gene (SILVA 99nr) database¹ version 138 database using the arb-silva.de web-interface ACT (Alignment-Classification-Tree) tool. Any residues outside of the aligned region were removed. Aligned sequences were combined with an additional 657 aligned 16S rRNA gene sequences from *Omnitrophota* and neighboring phyla exported from the same database. Aligned sequences with fewer than 1000 unambiguous residues in the alignment or those with a sequence, alignment, or quality lower than 75 were omitted. For phyla other than *Omnitrophota*, one sequence belonging to each genus was retained, while sequences belonging to *Omnitrophota* were clustered at a 99% threshold using vsearch¹²⁰ v2.18.0. Sequences were then filtered according to the Lane mask¹²¹ computed from the version provided by the authors of Mothur¹²². Duplicated sequences were omitted at this point. A phylogenetic tree was constructed from the masked sequences using IQ-Tree^{99,102} version 1.6.8. Sequences within 99% clusters were considered to be from the same species-level group. Taxonomy was assigned to nodes on the

phylogenetic tree based on the consensus of each node's children. This taxonomy along with the unmasked versions of sequences represented on the phylogenetic tree (**Supplementary Table 8**) were then imported into Qiime2³⁹ version 2020.8. The sequences were clipped to the V4-V5 region using the feature-classifier¹²³ plugin from Qiime2 by using the EMP 515F/806R¹²⁴ primers as parameters. A naïve-Bayesian sequence classifier was generated from the clipped sequences and corresponding taxonomy using the feature-classifier¹²³ plugin from Qiime2. 16S rRNA gene distances were calculated using Mothur¹²² 1.44.3.

Earth Microbiome Project meta-analysis.

Earth Microbiome Project⁴⁰ biom files were downloaded from the EMP ftp server. Sequence variants were trimmed to 90nt and dereplicated to 3,664,846, then classified according to the SILVA 99nr database¹ version 138 using the classifier obtained from the developers of Qiime2¹²⁵. Sequences classified as *Omnitrophota* (in SILVA: *Verrucomicrobia;Omnitrophia*) were classified once more using the *Omnitrophota*-specific classifier created here, yielding 29,249 *Omnitrophota* sequence variants. SVs unclassified at the domain level were removed. Sequence variant (SV) tables corresponding to samples from release 1 of the EMP dataset, excluding negative controls or blanks were merged into a single table. This table was used to calculate the environmental and geographic distribution of *Omnitrophota*.

Tandem filtration and 16S rRNA gene amplicon analysis.

Spring water was filtered using serialized 1 µm, 0.45 µm, and 0.2 µm polyethersulfone membrane Sterivex-GP pressure filters (Millipore Sigma, Burlington, MA) through Masterflex LS-24 platinum-cured silicone tubing (Cole-Parmer, Vernon Hills, IL) with a Geopump peristaltic pump (Geotech, Denver, CO). The inlet tube was placed as close to the spring source as possible. Filters were purged of water, frozen immediately afterward, and kept on dry ice. The filters were transferred to a -80 °C freezer until DNA was extracted. Membranes from the 0.45 and 0.2 µm filters were pulverized manually. DNA was extracted from the membrane pulp using a FastDNATM SPIN Kit for Soil (MP Biomedicals, Santa Ana, CA) according to the manufacturer's instructions. DNA extracts were amplified using the primer sets OP3-LiM (F: GGGTGTAAGGGCAGGTA; R: TACGACTTAGCGCCAGTC)^{27,28} and OP3-V4 (F: CGGATTNAYTGGGYGTAAAG; R: CYCTTACACCGRGAATTCC) (this study). PCR reaction mixtures contained 1 µL DNA extract, 15 µL of molecular grade nuclease-free water, 5 µL Taq reaction buffer (1X), 1 µL of each primer (0.4 µM), 2 µL of dNTPs (0.2 mM each), and 0.65 U of GoTaq DNA polymerase (Promega). For the amplification of DNA using the OP3-LiM primer set, cycling conditions were as follows: denaturation at 94 °C for 4 minutes, then 30 cycles of denaturation (1 minute at 94 °C), primer annealing (1 minute at 50 °C), and elongation (3 minutes at 72 °C), with a final elongation step (10 minutes at 72 °C). For the amplification of DNA using the OP3-V4 primer set, cycling conditions were as follows: denaturation at 95 °C for 5 minutes, then 30 cycles of denaturation (1 minute at 95 °C), primer annealing (1 minute at 48 °C), and elongation (45 seconds at 72 °C), with a final elongation step (5

minutes at 72 °C). Amplification products were visualized using 1% agarose gel electrophoresis and ethidium bromide staining. DNA extracts were submitted to MrDNA (www.mrdnalab.com, MR DNA, Shallowater, TX) for sequencing on the Illumina MiSeq platform. The updated bacterial- and archaeal-specific 515F/806R primer set was used to amplify the V4 region of the 16S rRNA gene^{126,127}. 16S rRNA gene amplicon reads from serial filtration experiments were processed using Qiime2¹²³ version 2020.8. Paired-end reads were quality filtered using the quality-filter plugin of Qiime2 at the default settings. Quality-filtered reads were then joined, trimmed to 150 nt, and SVs were identified using deblur within Qiime2. SVs were then classified using both classifiers as previously described. Raw reads for tandem filtration experiments were submitted to the Sequence Read Archive under accession number SRAXXXXXX.

Quantitative stable isotope probing and 16S rRNA amplicon analysis.

qSIP was performed across two separate experiments: (i) utilization of ¹⁸O H₂O to assess intrinsic growth and response to carbon addition in the form of root exudates or leaf litter, as described previously⁸⁸, and (ii) utilization of ¹⁸O H₂O or ¹³C with carbon amendments of glucose and oxalic acid to soil⁸⁸. Briefly, 1.0 g dry weight soil was added to 15 mL Falcon tubes and irrigated to 60% field capacity soil moisture for five days, and addition of ¹⁸O H₂O and carbon were done as described⁸⁸. Tubes were incubated for one week, followed by storage at -80°C. DNA extracts were then fractionated, and DNA concentration was quantified using the procedure described by Hungate et al., (2015)¹²⁸ and Purcell et al., (2018)¹²⁹. For each fraction, 16S rRNA gene copy number was assessed via qPCR and amplicons were sequenced with the Illumina MiSeq platform. 16S rRNA gene amplicon reads from both qSIP experiments were processed using Qiime version 2020.8. Paired-end reads were quality filtered, joined, trimmed to 150 nt, and SVs were identified using Dada2 within Qiime2. SVs were then classified using both classifiers as previously described. AFE values were calculated using the R package qsip version 0.1.0. Utilization of labeled substrate (¹⁸O or ¹³C) was considered for taxa where both tails of the 95 % confidence interval of AFE values were greater than zero. Additional methods are provided in **Supplementary Text**. Raw reads for unpublished qSIP experiments were submitted to the Sequence Read Archive under accession number SRAXXXXXX.

Declarations

Funding.

This work was funded by US National Science Foundation grants 1826734 (BH, DM, RS, CS), DEB-1441717 (RS and DM), and 1928924 (BH, MP), and NASA Exobiology grant 80NSSC17K0548 (BH, CS, MP). TW, FS and EF were funded by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility supported under Contract No. DE-AC02-05CH11231. MC was supported by an Australian Research Council Laureate Fellowship (FL150100038).

Acknowledgements.

We are grateful to the staff from the Yunnan Tengchong Volcano and Spa Tourist Attraction Development Corporation for their assistance and the US Bureau of Land Management Las Vegas Office and the US Department of Agriculture Forest Service (permit SMA0556) for permission to sample springs in the Spring Mountains. Ash Meadows National Wildlife Refuge was sampled under US National Park Service permit DEVA-2009-SCI-005 and US Fish and Wildlife Service (USFWS) permit 84550-15-03.

References

1. Pruesse, E. *et al.* SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* **35**, 7188–7196 (2007).
2. Glöckner, J. *et al.* Phylogenetic diversity and metagenomics of candidate division OP3. *Environ Microbiol* **12**, 1218–1229 (2010).
3. Hugenholtz, P., Pitulle, C., Hershberger, K. L. & Pace, N. R. Novel Division Level Bacterial Diversity in a Yellowstone Hot Spring. *Journal of Bacteriology* **180**, 366–376 (1998).
4. Rinke, C. *et al.* Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**, 431–437 (2013).
5. Williams, T. J., Allen, M. A., Berengut, J. F. & Cavicchioli, R. Shedding Light on Microbial “Dark Matter”: Insights Into Novel Cloacimonadota and Omnitrophota From an Antarctic Lake. *Frontiers in Microbiology* **12**, 2947 (2021).
6. Momper, L., Jungbluth, S. P., Lee, M. D. & Amend, J. P. Energy and carbon metabolisms in a deep terrestrial subsurface fluid microbial community. *ISME J* **11**, 2319–2333 (2017).
7. Parks, D. H. *et al.* Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat Microbiol* **2**, 1533–1542 (2017).
8. Vavourakis, C. D. *et al.* A metagenomics roadmap to the uncultured genome diversity in hypersaline soda lake sediments. *Microbiome* **6**, 1–18 (2018).
9. Vavourakis, C. D. *et al.* Metagenomes and metatranscriptomes shed new light on the microbial-mediated sulfur cycle in a Siberian soda lake. *BMC Biol* **17**, 1–20 (2019).
10. Seitz, K. W., Lazar, C. S., Hinrichs, K.-U., Teske, A. P. & Baker, B. J. Genomic reconstruction of a novel, deeply branched sediment archaeal phylum with pathways for acetogenesis and sulfur reduction. *ISME J* **10**, 1696–1705 (2016).

11. Woodcroft, B. J. *et al.* Genome-centric view of carbon processing in thawing permafrost. *Nature* **560**, 49–54 (2018).
12. Seitz, K. W. *et al.* Asgard archaea capable of anaerobic hydrocarbon cycling. *Nat Commun* **10**, 1822 (2019).
13. Zhou, Z. *et al.* Genome- and Community-Level Interaction Insights into Carbon Utilization and Element Cycling Functions of Hydrothermarchaeota in Hydrothermal Sediment. *mSystems* **5**, (2020).
14. Vigneron, A. *et al.* Genomic evidence for sulfur intermediates as new biogeochemical hubs in a model aquatic microbial ecosystem. *Microbiome* **9**, 1–14 (2021).
15. Nobu, M. K. *et al.* Catabolism and interactions of uncultured organisms shaped by eco-thermodynamics in methanogenic bioprocesses. *Microbiome* **8**, 111 (2020).
16. Tian, R. *et al.* Small and mighty: adaptation of superphylum Patescibacteria to groundwater environment drives their genome simplicity. *Microbiome* **8**, 51 (2020).
17. Pedron, R. *et al.* Genomic and metagenomic insights into the microbial community of a thermal spring. *Microbiome* **7**, 8 (2019).
18. Coutinho, F. H. *et al.* New viral biogeochemical roles revealed through metagenomic analysis of Lake Baikal. *Microbiome* **8**, 163 (2020).
19. Buck, M. *et al.* Comprehensive dataset of shotgun metagenomes from oxygen stratified freshwater lakes and ponds. *Sci Data* **8**, 131 (2021).
20. Alteio, L. V. *et al.* Complementary Metagenomic Approaches Improve Reconstruction of Microbial Diversity in a Forest Soil. *mSystems* **5**, e00768-19 (2020).
21. Suzuki, S., Nealson, K. H. & Ishii, S. Genomic and in-situ Transcriptomic Characterization of the Candidate Phylum NPL-UPL2 From Highly Alkaline Highly Reducing Serpentinized Groundwater. *Front Microbiol* **9**, 3141 (2018).
22. Breister, A. M. *et al.* Soil microbiomes mediate degradation of vinyl ester-based polymer composites. *Commun Mater* **1**, 1–15 (2020).
23. Campanaro, S. *et al.* New insights from the biogas microbiome by comprehensive genome-resolved metagenomics of nearly 1600 species originating from multiple anaerobic digesters. *Biotechnol Biofuels* **13**, 25 (2020).
24. Probst, A. J. *et al.* Differential depth distribution of microbial function and putative symbionts through sediment-hosted aquifers in the deep terrestrial subsurface. *Nature Microbiology* **3**, 328–336 (2018).

25. Kolinko, S. *et al.* Single-cell analysis reveals a novel uncultivated magnetotactic bacterium within the candidate division OP3. *Environmental Microbiology* **14**, 1709–1721 (2012).
26. Kolinko, S., Richter, M., Glöckner, F.-O., Brachmann, A. & Schüler, D. Single-cell genomics of uncultivated deep-branching magnetotactic bacteria reveals a conserved set of magnetosome genes. *Environmental Microbiology* **18**, 21–37 (2016).
27. Kizina, J. Insights into the biology of Candidate Division OP3 LiM populations. (2017).
28. Rotaru, A.-E., Schauer, R., Probian, C., Mussmann, M. & Harder, J. Visualization of candidate division OP3 cocci in limonene-degrading methanogenic cultures. *J. Microbiol. Biotechnol.* **22**, 457–461 (2012).
29. Suominen, S., Dombrowski, N., Sinninghe Damsté, J. S. & Villanueva, L. A diverse uncultivated microbial community is responsible for organic matter degradation in the Black Sea sulphidic zone. *Environ Microbiol* **23**, 2709–2728 (2021).
30. Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* **9**, 1–8 (2018).
31. Bowers, R. M. *et al.* Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol* **35**, 725–731 (2017).
32. Parks, D. H. *et al.* A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* **36**, 996–1004 (2018).
33. Na, S.-I. *et al.* UBCG: Up-to-date bacterial core gene set and pipeline for phylogenomic tree reconstruction. *J Microbiol.* **56**, 280–285 (2018).
34. Ankenbrand, M. J. & Keller, A. bcgTree: automatized phylogenetic tree building from bacterial core genomes. *Genome* **59**, 783–791 (2016).
35. Eloe-Fadrosh, E. A. *et al.* Global metagenomic survey reveals a new bacterial candidate phylum in geothermal springs. *Nature Communications* **7**, 10476 (2016).
36. Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **36**, 1925–1927 (2019).
37. Rodriguez-R, L. M. & Konstantinidis, K. T. Bypassing Cultivation To Identify Bacterial Species: Culture-independent genomic approaches identify credibly distinct clusters, avoid cultivation bias, and provide true insights into microbial species. *Microbe Magazine* **9**, 111–118 (2014).
38. Murray, A. E. *et al.* Roadmap for naming uncultivated Archaea and Bacteria. *Nature Microbiology* **5**, 987–994 (2020).

39. Bolyen, E. *et al.* Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* **37**, 852–857 (2019).
40. Thompson, L. R. *et al.* A communal catalogue reveals Earth’s multiscale microbial diversity. *Nature* **551**, 457–463 (2017).
41. Proctor, C. R. *et al.* Phylogenetic clustering of small low nucleic acid-content bacteria across diverse freshwater ecosystems. *ISME J* **12**, 1344–1359 (2018).
42. Vigneron, A. *et al.* Ultra-small and abundant: Candidate phyla radiation bacteria are potential catalysts of carbon transformation in a thermokarst lake ecosystem. *Limnology and Oceanography Letters* **5**, 212–220 (2020).
43. Beam, J. P. *et al.* Ancestral Absence of Electron Transport Chains in Patescibacteria and DPANN. *Front. Microbiol.* **11**, (2020).
44. Anantharaman, K. *et al.* Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nature Communications* **7**, 13219 (2016).
45. Hug, L. A. *et al.* Aquifer environment selects for microbial species cohorts in sediment and groundwater. *ISME J* **9**, 1846–1856 (2015).
46. Sabath, N., Ferrada, E., Barve, A. & Wagner, A. Growth Temperature and Genome Size in Bacteria Are Negatively Correlated, Suggesting Genomic Streamlining During Thermal Adaptation. *Genome Biology and Evolution* **5**, 966–977 (2013).
47. Zeldovich, K. B., Berezovsky, I. N. & Shakhnovich, E. I. Protein and DNA Sequence Determinants of Thermophilic Adaptation. *PLOS Computational Biology* **3**, e5 (2007).
48. Sato, Y., Okano, K., Kimura, H. & Honda, K. TEMPURA: Database of Growth TEMPeratures of Usual and RAre Prokaryotes. *Microbes Environ* **35**, ME20074 (2020).
49. Dittrich, C. R., Bennett, G. N. & San, K.-Y. Characterization of the Acetate-Producing Pathways in *Escherichia coli*. *Biotechnology Progress* **21**, 1062–1067 (2005).
50. Crowley, P. J., Gutierrez, J. A., Hillman, J. D. & Bleiweis, A. S. Genetic and physiologic analysis of a formyl-tetrahydrofolate synthetase mutant of *Streptococcus mutans*. *Journal of Bacteriology* **179**, 1563–1572 (1997).
51. Hesslinger, C., Fairhurst, S. A. & Sawers, G. Novel keto acid formate-lyase and propionate kinase enzymes are components of an anaerobic pathway in *Escherichia coli* that degrades L-threonine to propionate. *Mol Microbiol* **27**, 477–492 (1998).

52. Tang, Y.-Q., Shigematsu, T., Morimura, S. & Kida, K. Effect of dilution rate on the microbial structure of a mesophilic butyrate-degrading methanogenic community during continuous cultivation. *Appl Microbiol Biotechnol* **75**, 451–465 (2007).
53. Ljungdahl, L., Irion, E. & Wood, H. G. Total Synthesis of Acetate from CO₂. I. Co-Methylcobyrlic Acid and Co-(Methyl)-5-methoxybenzimidazolylcobamide as Intermediates with *Clostridium thermoaceticum**. *Biochemistry* **4**, 2771–2780 (1965).
54. Westphal, L., Wiechmann, A., Baker, J., Minton, N. P. & Müller, V. The Rnf Complex Is an Energy-Coupled Transhydrogenase Essential To Reversibly Link Cellular NADH and Ferredoxin Pools in the Acetogen *Acetobacterium woodii*. *Journal of Bacteriology* **200**, (2018).
55. Kuhns, M., Trifunović, D., Huber, H. & Müller, V. The Rnf complex is a Na⁺ coupled respiratory enzyme in a fermenting bacterium, *Thermotoga maritima*. *Communications Biology* **3**, 1–10 (2020).
56. Loh, H. Q., Hervé, V. & Brune, A. Metabolic Potential for Reductive Acetogenesis and a Novel Energy-Converting [NiFe] Hydrogenase in Bathyarchaeia From Termite Guts – A Genome-Centric Analysis. *Front. Microbiol.* **11**, (2021).
57. Greening, C. *et al.* Genomic and metagenomic surveys of hydrogenase distribution indicate H₂ is a widely utilised energy source for microbial growth and survival. *ISME J* **10**, 761–777 (2016).
58. Weidner, U. *et al.* The Gene Locus of the Proton-translocating NADH: Ubiquinone Oxidoreductase in *Escherichia coli*: Organization of the 14 Genes and Relationship Between the Derived Proteins and Subunits of Mitochondrial Complex I. *Journal of Molecular Biology* **233**, 109–122 (1993).
59. Kim, Y. J. *et al.* Formate-driven growth coupled with H₂ production. *Nature* **467**, 352–355 (2010).
60. Baughn, A. D. & Malamy, M. H. The strict anaerobe *Bacteroides fragilis* grows in and benefits from nanomolar concentrations of oxygen. *Nature* **427**, 441–444 (2004).
61. Garber, A. I. *et al.* FeGenie: A Comprehensive Tool for the Identification of Iron Genes and Iron Gene Neighborhoods in Genome and Metagenome Assemblies. *Front. Microbiol.* **11**, (2020).
62. Deng, X., Dohmae, N., Nealson, K. H., Hashimoto, K. & Okamoto, A. Multi-heme cytochromes provide a pathway for survival in energy-limited environments. *Science Advances* **4**, eaao5682 (2018).
63. McGlynn, S. E., Chadwick, G. L., Kempes, C. P. & Orphan, V. J. Single cell activity reveals direct electron transfer in methanotrophic consortia. *Nature* **526**, 531–535 (2015).
64. Walker, D. J. *et al.* Electrically conductive pili from pilin genes of phylogenetically diverse microorganisms. *The ISME Journal* **12**, 48–58 (2018).

65. Poehlein, A. *et al.* An Ancient Pathway Combining Carbon Dioxide Fixation with the Generation and Utilization of a Sodium Ion Gradient for ATP Synthesis. *PLoS One* **7**, (2012).
66. Moreira, D., Zivanovic, Y., López-Archilla, A. I., Iniesto, M. & López-García, P. Reductive evolution and unique predatory mode in the CPR bacterium *Vampirococcus lugosii*. *Nat Commun* **12**, 2454 (2021).
67. Evans, K. J., Lambert, C. & Sockett, R. E. Predation by *Bdellovibrio bacteriovorus* HD100 Requires Type IV Pili. *J Bacteriol* **189**, 4850–4859 (2007).
68. Pasternak, Z. *et al.* In and out: an analysis of epibiotic vs periplasmic bacterial predators. *ISME J* **8**, 625–635 (2014).
69. Tulum, I., Kimura, K. & Miyata, M. Identification and sequence analyses of the gliding machinery proteins from *Mycoplasma mobile*. *Scientific Reports* **10**, 3792 (2020).
70. Tulum, I., Yabe, M., Uenoyama, A. & Miyata, M. Localization of P42 and F1-ATPase α -Subunit Homolog of the Gliding Machinery in *Mycoplasma mobile* Revealed by Newly Developed Gene Manipulation and Fluorescent Protein Tagging. *Journal of Bacteriology* **196**, 1815–1824 (2014).
71. Brown, C. T. *et al.* Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* **523**, 208–211 (2015).
72. Kindaichi, T. *et al.* Phylogenetic diversity and ecophysiology of Candidate phylum Saccharibacteria in activated sludge. *FEMS Microbiology Ecology* **92**, (2016).
73. Miller, I. J., Weyna, T. R., Fong, S. S., Lim-Fong, G. E. & Kwan, J. C. Single sample resolution of rare microbial dark matter in a marine invertebrate metagenome. *Scientific Reports* **6**, 34362 (2016).
74. Yeoh, Y. K., Sekiguchi, Y., Parks, D. H. & Hugenholtz, P. Comparative Genomics of Candidate Phylum TM6 Suggests That Parasitism Is Widespread and Ancestral in This Lineage. *Mol Biol Evol* **33**, 915–927 (2016).
75. Schmitz-Esser, S. *et al.* ATP/ADP Translocases: a Common Feature of Obligate Intracellular Amoebal Symbionts Related to Chlamydiae and Rickettsiae. *Journal of Bacteriology* **186**, 683–691 (2004).
76. Horn, M. *et al.* Members of the Cytophaga-Flavobacterium-Bacteroides phylum as intracellular bacteria of acanthamoebae: proposal of 'Candidatus Amoebophilus asiaticus'. *Environ Microbiol* **3**, 440–449 (2001).
77. Pagnier, I. *et al.* *Babela massiliensis*, a representative of a widespread bacterial phylum with unusual adaptations to parasitism in amoebae. *Biol Direct* **10**, (2015).

78. Reva, O. & Tümmler, B. Think big – giant genes in bacteria. *Environmental Microbiology* **10**, 768–777 (2008).
79. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
80. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
81. Sakka, M., Kunitake, E., Kimura, T. & Sakka, K. Function of a laminin_G_3 module as a carbohydrate-binding module in an arabinofuranosidase from *Ruminiclostridium josui*. *FEBS Lett* **593**, 42–51 (2019).
82. Itoh, Y. Discoidin domain receptors: Microenvironment sensors that promote cellular migration and invasion. *Cell Adhesion & Migration* **12**, 378–385 (2018).
83. Peer, A., Smith, S. P., Bayer, E. A., Lamed, R. & Borovok, I. Noncellulosomal cohesin- and dockerin-like modules in the three domains of life. *FEMS Microbiology Letters* **291**, 1–16 (2009).
84. Hamm, J. N. *et al.* Unexpected host dependency of Antarctic Nanoarchaeota. *PNAS* **116**, 14661–14670 (2019).
85. Vogl, K. *et al.* Identification and analysis of four candidate symbiosis genes from ‘*Chlorochromatium aggregatum*’, a highly developed bacterial symbiosis. *Environmental Microbiology* **10**, 2842–2856 (2008).
86. Liu, Z. *et al.* Genomic analysis reveals key aspects of prokaryotic symbiosis in the phototrophic consortium “*Chlorochromatium aggregatum*”. *Genome Biol* **14**, 1–17 (2013).
87. Paul, B. G. *et al.* Retroelement-guided protein diversification abounds in vast lineages of Bacteria and Archaea. *Nat Microbiol* **2**, 17045 (2017).
88. Hungate, B. A. *et al.* The Functional Significance of Bacterial Predators. *mBio* **12**, (2021).
89. Castelle, C. J. *et al.* Biosynthetic capacity, metabolic variety and unusual biology in the CPR and DPANN radiations. *Nature Reviews Microbiology* **16**, 629–645 (2018).
90. Jarett, J. K. *et al.* Single-cell genomics of co-sorted Nanoarchaeota suggests novel putative host associations and diversification of proteins involved in symbiosis. *Microbiome* **6**, 1–14 (2018).
91. Munson-McGee, J. H. *et al.* Nanoarchaeota, Their Sulfolobales Host, and Nanoarchaeota Virus Distribution across Yellowstone National Park Hot Springs. *Applied and Environmental Microbiology* **81**, 7860–7868 (2015).
92. Cross, K. L. *et al.* Targeted isolation and cultivation of uncultivated bacteria by reverse genomics. *Nat Biotechnol* **37**, 1314–1321 (2019).

93. He, X. *et al.* Cultivation of a human-associated TM7 phylotype reveals a reduced genome and epibiotic parasitic lifestyle. *Proc Natl Acad Sci U S A* **112**, 244–249 (2015).
94. Stepanauskas, R. *et al.* Improved genome recovery and integrated cell-size analyses of individual uncultured microbial cells and viral particles. *Nat Commun* **8**, 84 (2017).
95. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25**, 1043–1055 (2015).
96. Matsen, F. A., Kodner, R. B. & Armbrust, E. V. pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics* **11**, 538 (2010).
97. Eddy, S. R. Accelerated Profile HMM Searches. *PLOS Computational Biology* **7**, e1002195 (2011).
98. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
99. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol Biol Evol* **32**, 268–274 (2015).
100. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* **30**, 772–780 (2013).
101. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Haeseler, A. von & Jermini, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* **14**, 587–589 (2017).
102. Minh, B. Q., Nguyen, M. A. T. & von Haeseler, A. Ultrafast Approximation for Phylogenetic Bootstrap. *Mol Biol Evol* **30**, 1188–1195 (2013).
103. Barbera, P. *et al.* EPA-ng: Massively Parallel Evolutionary Placement of Genetic Sequences. *Syst Biol* **68**, 365–369 (2019).
104. Rodriguez-R, L. M. & Konstantinidis, K. T. *The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes*. <https://peerj.com/preprints/1900> (2016) doi:10.7287/peerj.preprints.1900v1.
105. Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* **20**, 289–290 (2004).
106. Revell, L. J. phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* **3**, 217–223 (2012).

107. Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T.-Y. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution* **8**, 28–36 (2017).
108. Aramaki, T. *et al.* KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* **btz859**, (2019).
109. Abby, S. S. & Rocha, E. P. C. Identification of Protein Secretion Systems in Bacterial Genomes Using MacSyFinder. *Methods Mol. Biol.* **1615**, 1–21 (2017).
110. Abby, S. S. *et al.* Identification of protein secretion systems in bacterial genomes. *Sci Rep* **6**, 1–14 (2016).
111. Zhou, Z., Tran, P., Liu, Y., Kieft, K. & Anantharaman, K. METABOLIC: A scalable high-throughput metabolic and biogeochemical functional trait profiler based on microbial genomes. *bioRxiv* 761643 (2019) doi:10.1101/761643.
112. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305**, 567–580 (2001).
113. Bateman, A. *et al.* The Pfam protein families database. *Nucleic Acids Research* **32**, D138–D141 (2004).
114. Sonnhammer, E. L. L., Eddy, S. R. & Durbin, R. Pfam: A comprehensive database of protein domain families based on seed alignments. *Proteins: Structure, Function, and Bioinformatics* **28**, 405–420 (1997).
115. Haft, D. H., Selengut, J. D. & White, O. The TIGRFAMs database of protein families. *Nucleic Acids Res* **31**, 371–373 (2003).
116. Bengtsson-Palme, J. *et al.* METAXA2: improved identification and taxonomic classification of small and large subunit rRNA in metagenomic data. *Mol Ecol Resour* **15**, 1403–1414 (2015).
117. Chan, P. P. & Lowe, T. M. tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences. *Methods Mol Biol* **1962**, 1–14 (2019).
118. Kalvari, I. *et al.* Rfam 14: expanded coverage of metagenomic, viral and microRNA families. *Nucleic Acids Research* **49**, D192–D200 (2021).
119. Nawrocki, E. P. & Eddy, S. R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933–2935 (2013).

120. Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**, e2584 (2016).
121. Lane, D. 16S/23S rRNA Sequencing. in *E. Stackebrandt and M. Goodfellow (ed.), Nucleic acid techniques in bacterial systematics* 115–175 (John Wiley and Sons, 1991).
122. Schloss, P. D. *et al.* Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Appl. Environ. Microbiol.* **75**, 7537–7541 (2009).
123. Bokulich, N. A. *et al.* Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome* **6**, 1–17 (2018).
124. Caporaso, J. G. *et al.* Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *PNAS* **108**, 4516–4522 (2011).
125. Robeson, M. S. *et al.* RESCRIPT: Reproducible sequence taxonomy reference database management for the masses. *bioRxiv* 2020.10.05.326504 (2020) doi:10.1101/2020.10.05.326504.
126. Apprill, A., McNally, S., Parsons, R. & Weber, L. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquatic Microbial Ecology* **75**, 129–137 (2015).
127. Parada, A. E., Needham, D. M. & Fuhrman, J. A. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environmental Microbiology* **18**, 1403–1414 (2016).
128. Hungate, B. A. *et al.* Quantitative Microbial Ecology through Stable Isotope Probing. *Appl. Environ. Microbiol.* **81**, 7570–7581 (2015).
129. Purcell, A. M. *et al.* Quantitative stable isotope probing with H₂18O to measure taxon-specific microbial growth. *Soil Science Society of America Journal* **84**, 1503–1518 (2020).

Figures

representatives. The number within parentheses at the end of each tip corresponds to the genome ID within **Table S1**. Dotted nodes indicate SH-aLRT support $\geq 80\%$ and UFboot support of $\geq 95\%$.

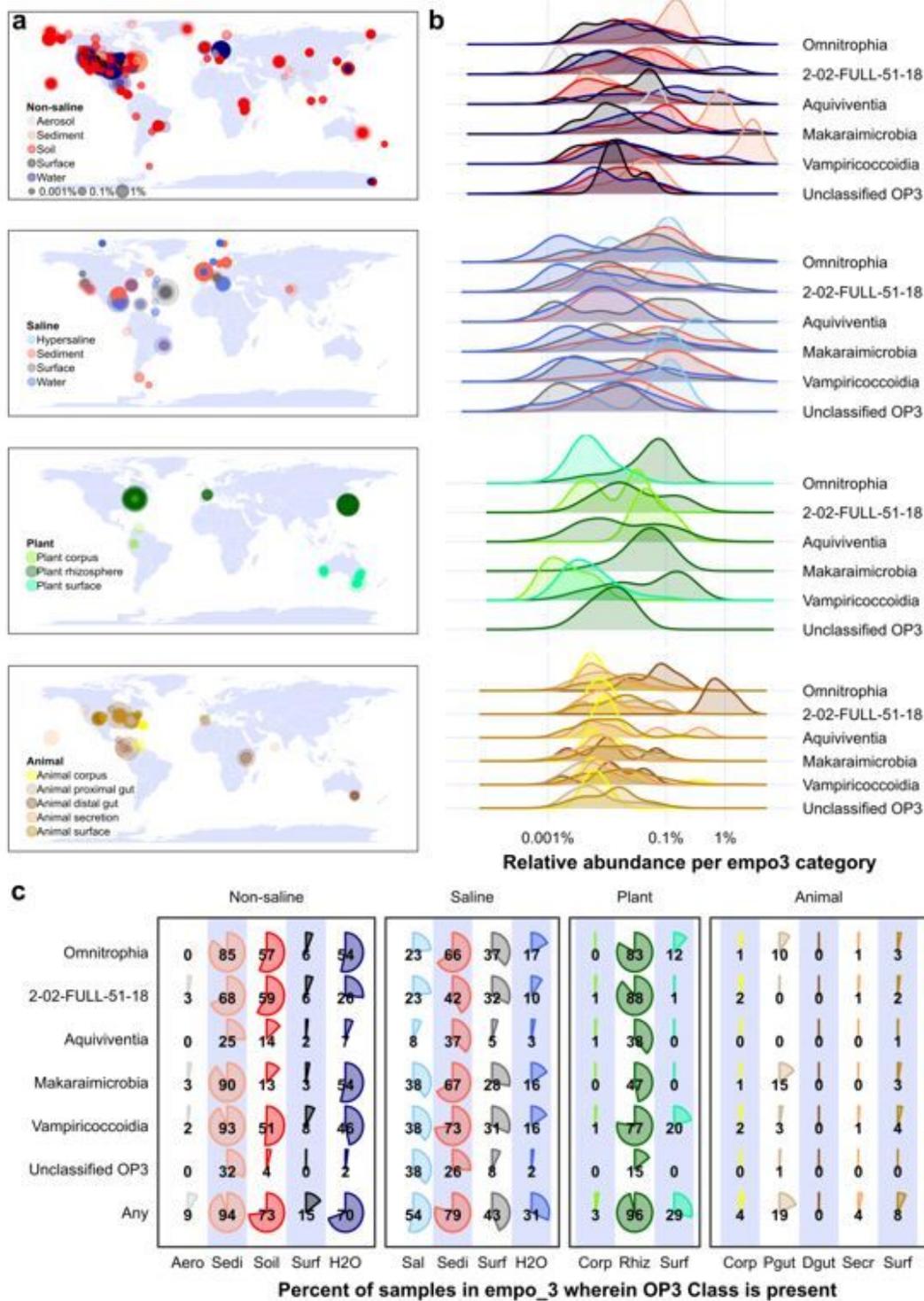


Figure 2

Global distribution of *Omnitrophota*. (a) Maps showing the coordinates of Earth Microbiome Project (EMP) samples in which *Omnitrophota* sequence variants were observed in (from top to bottom) EMP

environmental ontology (EMPO) level-2 categories: Non-saline, Saline, Plant-associated, and Animal-associated. Bubbles depict relative abundance of the phylum and are colored to indicate EMPO level 3. (b) Log₁₀-scale distribution of percent relative abundances of each class of *Omnitrophota* within each EMPO-3 category. (c) Multi-pie chart displaying the percent of EMP samples containing *Omnitrophota* sequences in each EMPO-3 category (x-axis). “Aero” refers to aerosol samples, “Sedi” sediment, “Surf” surface, “H₂O” water, “Sal” hypersaline, “Corp” corpus, “Rhiz” rhizosphere, “Pgut” proximal gut, “Dgut” distal gut, “Secr” secretion.

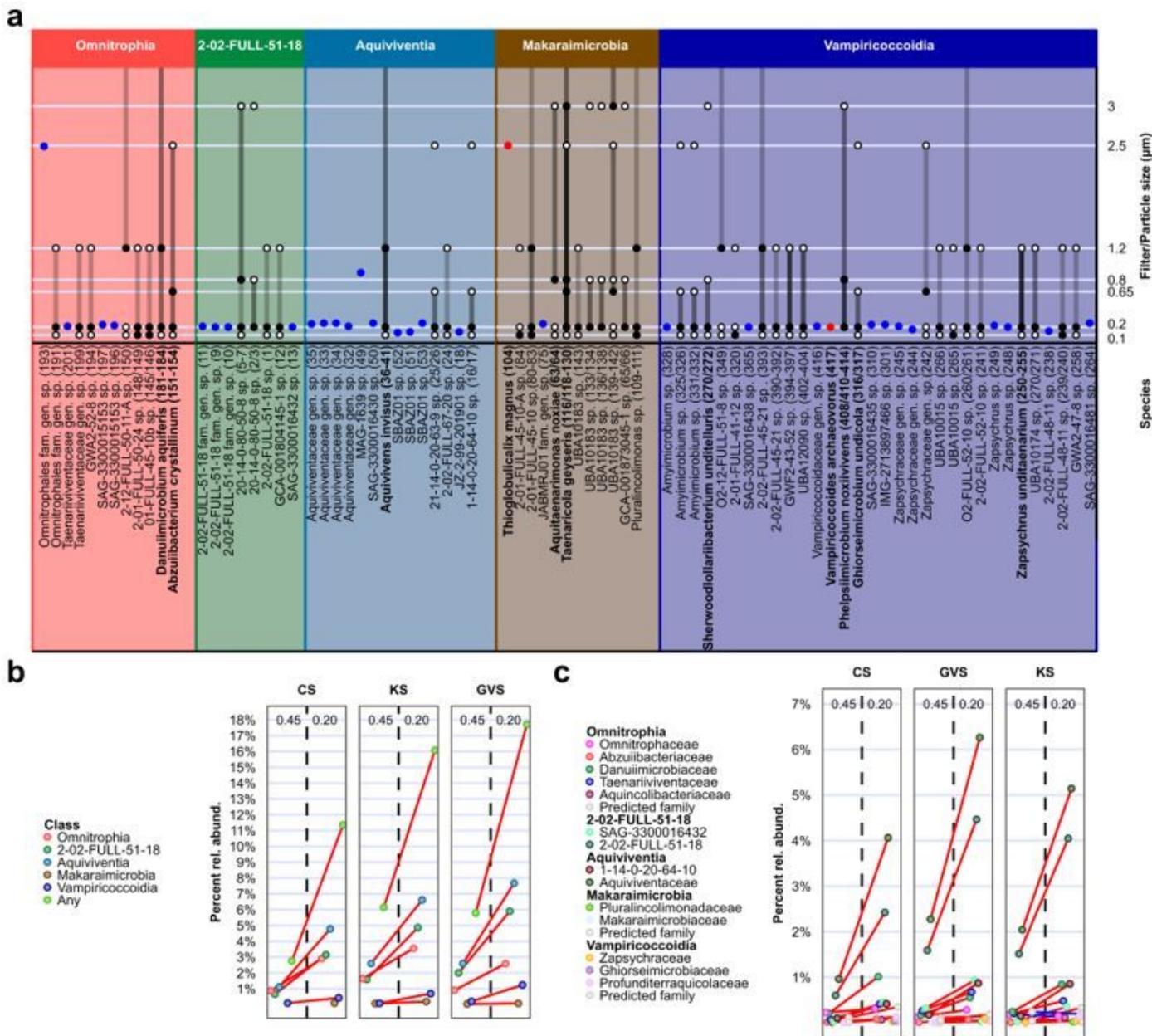


Figure 3

Omnitrophota cell size. (a) Cell size inferred from *Omnitrophota* MAGs from Rifle, CO^{44,45} and Crystal Geyser, UT²⁴ from serial filtered samples. Genomes associated with microscopically observed organisms

are indicated with red circles. SAGs with associated particle sizes are indicated as blue dots. Filters are shown as black or white circles: filters shown with filled circles retained the organism; unfilled circles did not. Lines connecting dots represent the possible range of each cell size given the observed data. (b) Relative abundance of 16S rRNA genes in filtrates from Cave Spring (CS), Kiup Spring (KS), and Grapevine Spring (GVS). Red lines indicate an increase in relative abundance from the 0.45 μm filter to the 0.2 μm filter. (c) The same calculation performed at the family level.

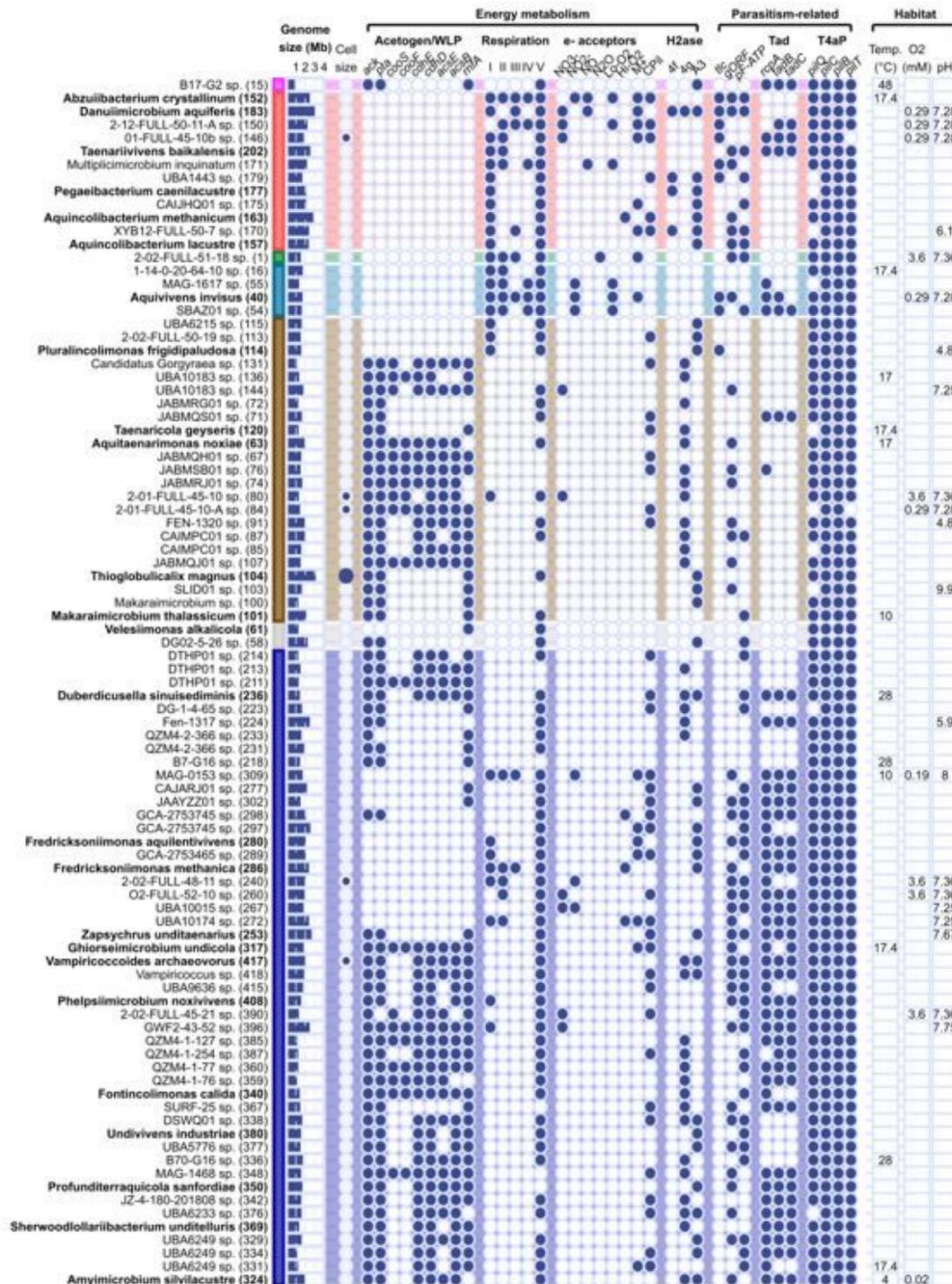


Figure 4

Summary of genomic, predicted physiology, and environmental data for high-completeness ($\geq 90\%$) genomes. Bars to the right of taxon names and in background reflect classes (**Figure 1**). “Genome size” indicates the observed size of each genome. “Cell size” indicates evidence that the genome was sequenced from small cells (<0.5 microns, filled small circle) or large cells (>0.5 microns, filled large circle). “Acetogen/WLP” (Wood-Ljungdahl pathway), “Respiration”, “e- acceptors”, and “H2ase” (hydrogenase) indicate genes predicted to encode proteins involved in energy metabolism. “Lo-O2” (cytochrome c oxidase complex); “Hi-O2” (cytochrome bd ubiquinol); “M+” (metal-reducing cytochromes); “CPil” (conductive pili). Parasitism-related genes include “tlc” (ADP/ATP translocase), “gORF” (giant ORF, indicating the presence of a large ORF), “Tad” (tight adherence pilus), and “T4aP” (type-4a pilus) indicates genes predicted to encode proteins involved in symbiosis/parasitism. “Temp”, “O2”, and “pH” indicate the observed temperature, oxygen concentration (mM), and pH of the sample from which each genome was sequenced. Numbers in parenthesis are unique genome identifiers as discussed in text. Data for additional *Omnitrophota* genomes are summarized in **Supplementary Fig. 20**.

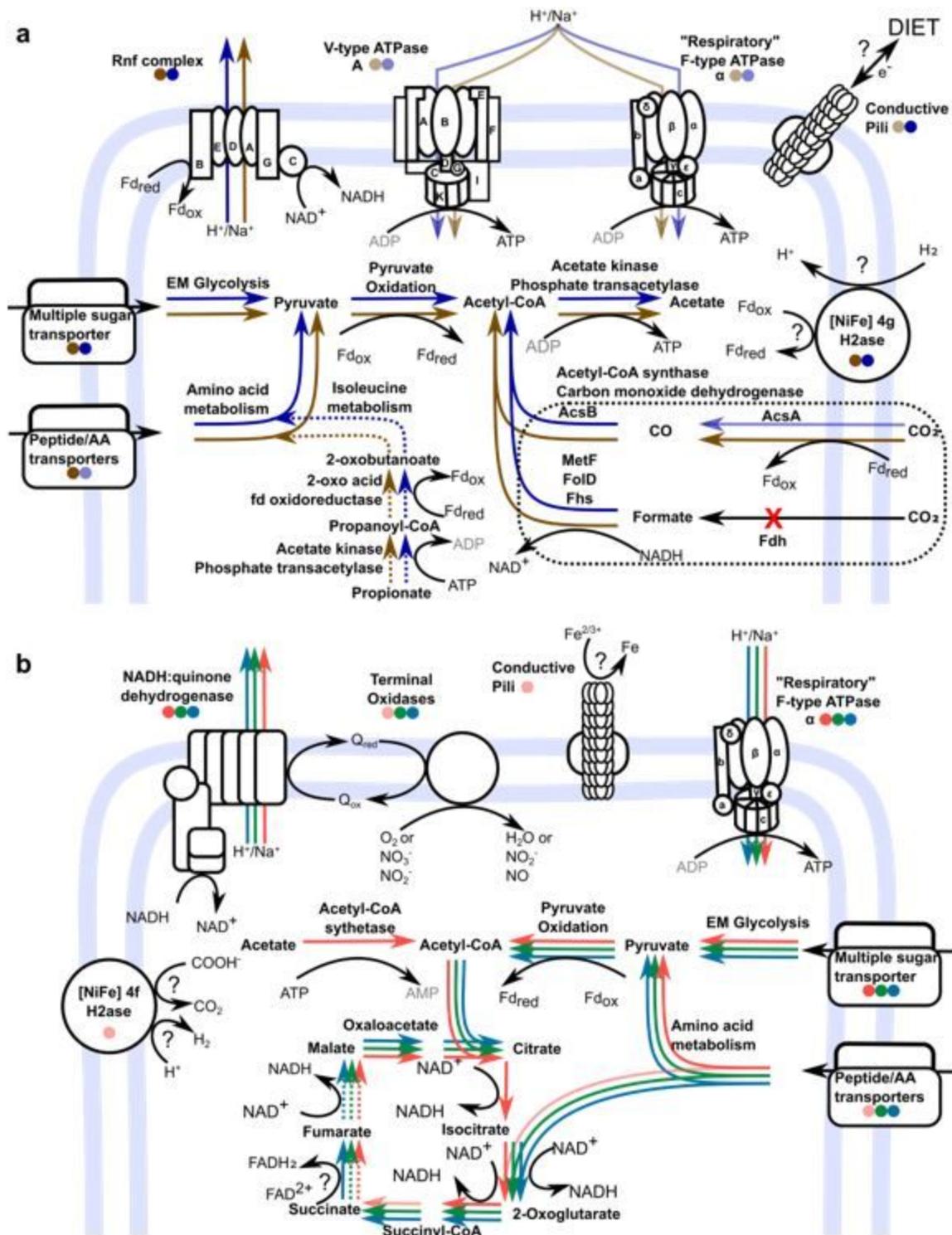


Figure 5

Conserved energy metabolism in major lineages of *Omnitrophota*. (a) Predicted metabolism of putative acetogens or syntrophs in *Vampiricoccodia* and *Makaraimicrobia*. (b) Predicted metabolism of putatively respiratory lineages *Omnitrophia*, *Aquiviventia*, and 2-02-FULL-51-18. Lines represent genes or modules as appropriate. Reactions are represented by arrows. Components of a complex are represented by colored circles. Colors correspond to each class. Shapes are opaque if gene or gene set catalyzing a

given reaction is present in the representative genomes of $\geq 50\%$ of species, transparent if > 1 and $< 50\%$ of species, or deleted if one or no species. Dotted lines represent pathways imputed from domains outside of KEGG ontology. See **Supplementary Fig. 18-21** for details of these features for ANI cluster representatives.

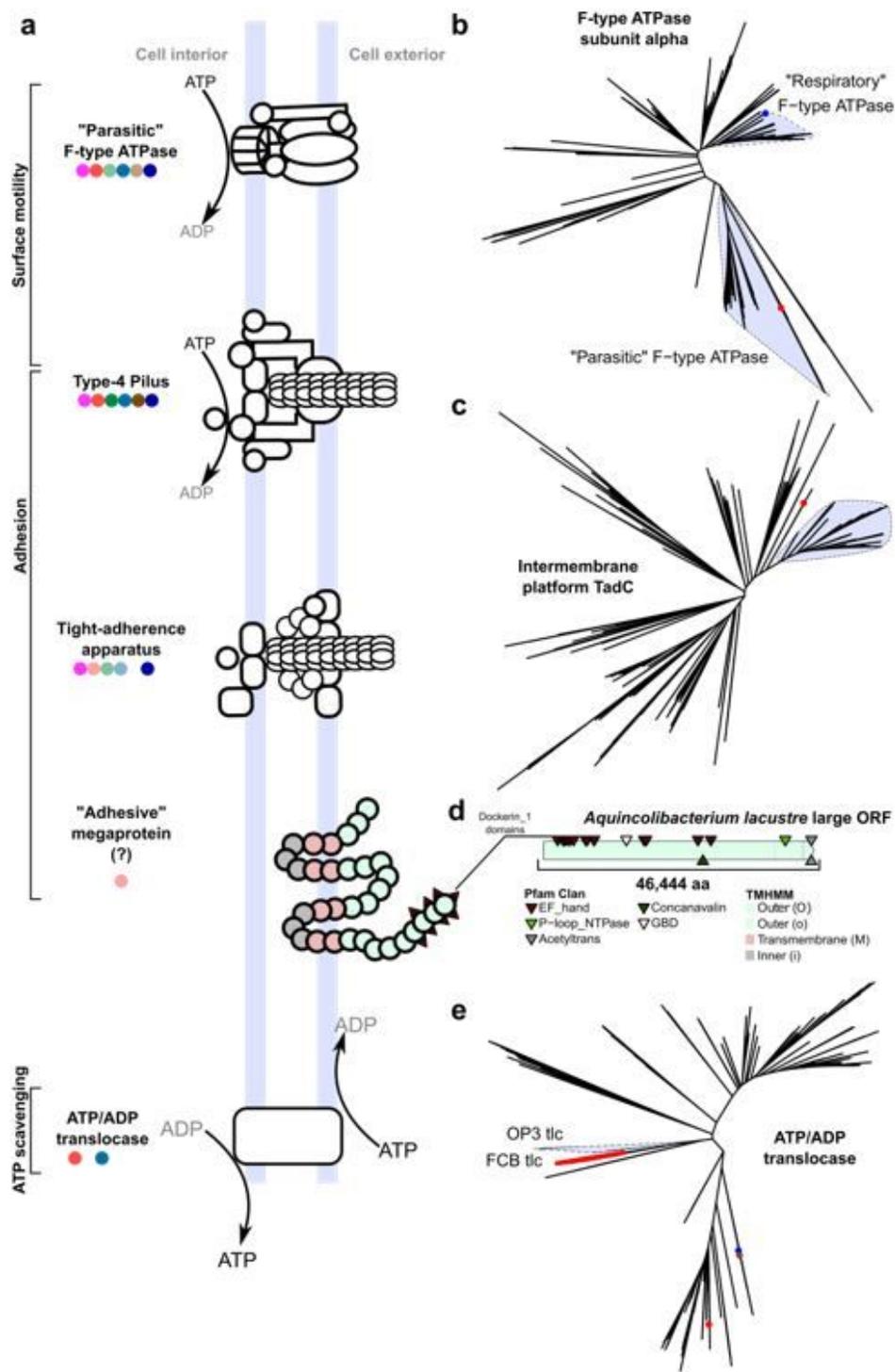


Figure 6

Summary of genomic evidence for parasitism and predation. (a) Systems related to parasitism and predation in *Omnitrophota* genomes. Circles correspond to occurrence in each class, with a lighter color indicating <50% of species in the class encoding the system. (b) Phylogeny of F-type ATP synthase subunit alpha. Highlighted clades indicate clusters of *Omnitrophota* genes neighboring putatively respiratory or parasitism-related genes. Tips corresponding to the biochemically characterized respiratory homolog from *Waddlia chondrophila* and the pathogenesis-related homolog from *Mycoplasma mobile* are represented with blue and red points respectively. (c) Phylogeny of homologs of “Tight-adherence” apparatus intermembrane platform protein TadC. The highlighted clade indicates a cluster of homologs from *Omnitrophota* genomes. The characterized TadC from *Halobacteriovorax marinus* is indicated with a red point. (d) Illustration of the largest ORF from the class *Omnitrophia*, which encodes domains possibly involved in adhesion. See **Supplementary Fig. 29** for additional views of large ORFs from *Omnitrophota* genomes. (e) Phylogeny of ATP/ADP translocase homologs. The light-blue highlighted clade represents homologs *Omnitrophota*. The red clade represents homologs from unclassified *Flavobacteriaceae*. The red point represents a homolog from *W. chondrophila*, blue from *Candidatus Babela massiliensis*, and brown from the *Omnitrophota* species *P. frigidipaludosa*. Expanded views for the phylogeny of each homolog are available in **Supplementary Fig. S22-S28**.

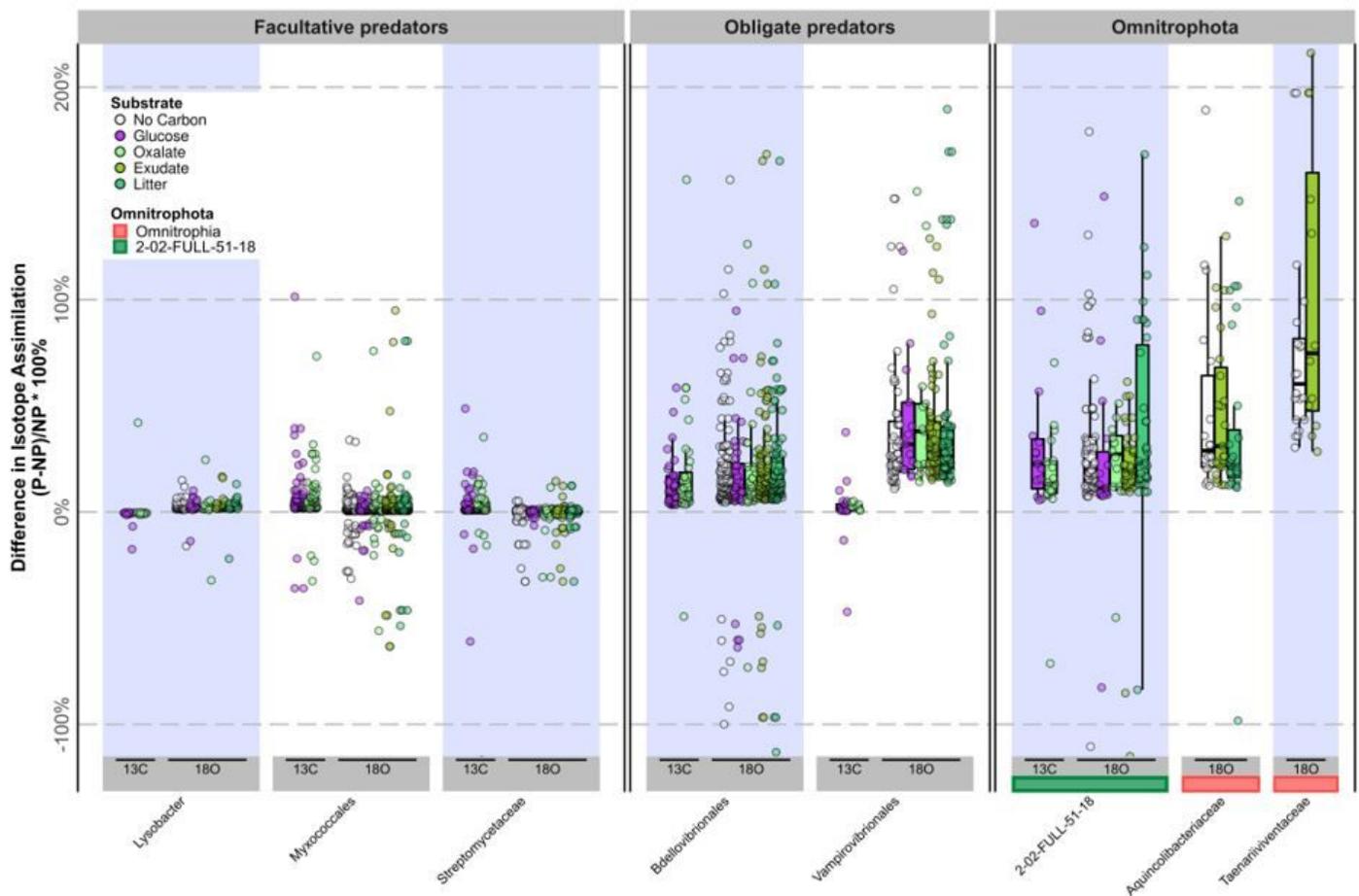


Figure 7

Family-level quantitative stable isotope probing in diverse soils. Y-axis shows the percent difference between atom fraction excess (AFE) for a given taxon (P), compared to all non-predatory (NP) taxa from the same sample. Boxes display the median and inner quartiles while whiskers extend to the 95 percent confidence interval of the distribution of AFE ratios for a given taxon within each experimental group.

Supplementary Files

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

- [SupplementNatMicro2022.pdf](#)
- [SupplementarytablesNatMicro2022.xlsx](#)