



# Genomic Analysis of the Yet-Uncultured Binatota Reveals Broad Methylo-trophic, Alkane-Degradation, and Pigment Production Capacities

Chelsea L. Murphy,<sup>a</sup> Andriy Sheremet,<sup>b</sup> Peter F. Dunfield,<sup>b</sup>  John R. Spear,<sup>c</sup> Ramunas Stepanauskas,<sup>d</sup>  Tanja Woyke,<sup>e</sup> Mostafa S. Elshahed,<sup>a</sup>  Noha H. Youssef<sup>a</sup>

<sup>a</sup>Department of Microbiology and Molecular Genetics, Oklahoma State University, Stillwater, Oklahoma, USA

<sup>b</sup>Department of Biological Sciences, University of Calgary, Calgary, Alberta, Canada

<sup>c</sup>Civil and Environmental Engineering, Colorado School of Mines, Golden, Colorado, USA

<sup>d</sup>Bigelow Laboratory for Ocean Sciences, East Boothbay, Maine, USA

<sup>e</sup>Department of Energy Joint Genome Institute, Berkeley, California, USA

**ABSTRACT** The recent leveraging of genome-resolved metagenomics has generated an enormous number of genomes from novel uncultured microbial lineages yet left many clades undescribed. Here, we present a global analysis of genomes belonging to Binatota (UBP10), a globally distributed, yet-uncharacterized bacterial phylum. All orders in Binatota encoded the capacity for aerobic methylo-trophy using methanol, methylamine, sulfome-thanes, and chloromethanes as the substrates. Methylo-trophy in Binatota was character-ized by order-specific substrate degradation preferences, as well as extensive metabolic versatility, i.e., the utilization of diverse sets of genes, pathways, and combinations to achieve a specific metabolic goal. The genomes also encoded multiple alkane hydroxy-lases and monooxygenases, potentially enabling growth on a wide range of alkanes and fatty acids. Pigmentation is inferred from a complete pathway for carotenoids (lycopene,  $\beta$ - and  $\gamma$ -carotenes, xanthins, chlorobactenes, and spheroidenes) production. Further, the majority of genes involved in bacteriochlorophyll *a*, *c*, and *d* biosynthesis were identified, although absence of key genes and failure to identify a photosynthetic reaction center preclude proposing phototrophic capacities. Analysis of 16S rRNA databases showed the preferences of Binatota to terrestrial and freshwater ecosystems, hydrocarbon-rich habitats, and sponges, supporting their potential role in mitigating methanol and methane emis-sions, breakdown of alkanes, and their association with sponges. Our results expand the lists of methylo-trophic, aerobic alkane-degrading, and pigment-producing lineages. We also highlight the consistent encountering of incomplete biosynthetic pathways in micro-bial genomes, a phenomenon necessitating careful assessment when assigning putative functions based on a set-threshold of pathway completion.

**IMPORTANCE** A wide range of microbial lineages remain uncultured, yet little is known regarding their metabolic capacities, physiological preferences, and ecological roles in various ecosystems. We conducted a thorough comparative genomic analysis of 108 genomes belonging to the Binatota (UBP10), a globally distributed, yet-uncharacterized bacterial phylum. We present evidence that members of the order Binatota specialize in methylo-trophy and identify an extensive repertoire of genes and pathways medi-ating the oxidation of multiple one-carbon ( $C_1$ ) compounds in Binatota genomes. The occurrence of multiple alkane hydroxylases and monooxygenases in these genomes was also identified, potentially enabling growth on a wide range of alkanes and fatty acids. Pigmentation is inferred from a complete pathway for carotenoids production. We also report on the presence of incomplete chlorophyll biosynthetic pathways in all genomes and propose several evolutionary-grounded scenarios that could explain such a pattern. Assessment of the ecological distribution patterns of the Binatota indicates

**Citation** Murphy CL, Sheremet A, Dunfield PF, Spear JR, Stepanauskas R, Woyke T, Elshahed MS, Youssef NH. 2021. Genomic analysis of the yet-uncultured Binatota reveals broad methylo-trophic, alkane-degradation, and pigment production capacities. *mBio* 12: e00985-21. <https://doi.org/10.1128/mBio.00985-21>.

**Editor** Caroline S. Harwood, University of Washington

**Copyright** © 2021 Murphy et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Noha H. Youssef, [Noha@okstate.edu](mailto:Noha@okstate.edu).

**Received** 5 April 2021

**Accepted** 7 April 2021

**Published** 18 May 2021

preference of its members to terrestrial and freshwater ecosystems characterized by high methane and methanol emissions, as well as multiple hydrocarbon-rich habitats and marine sponges.

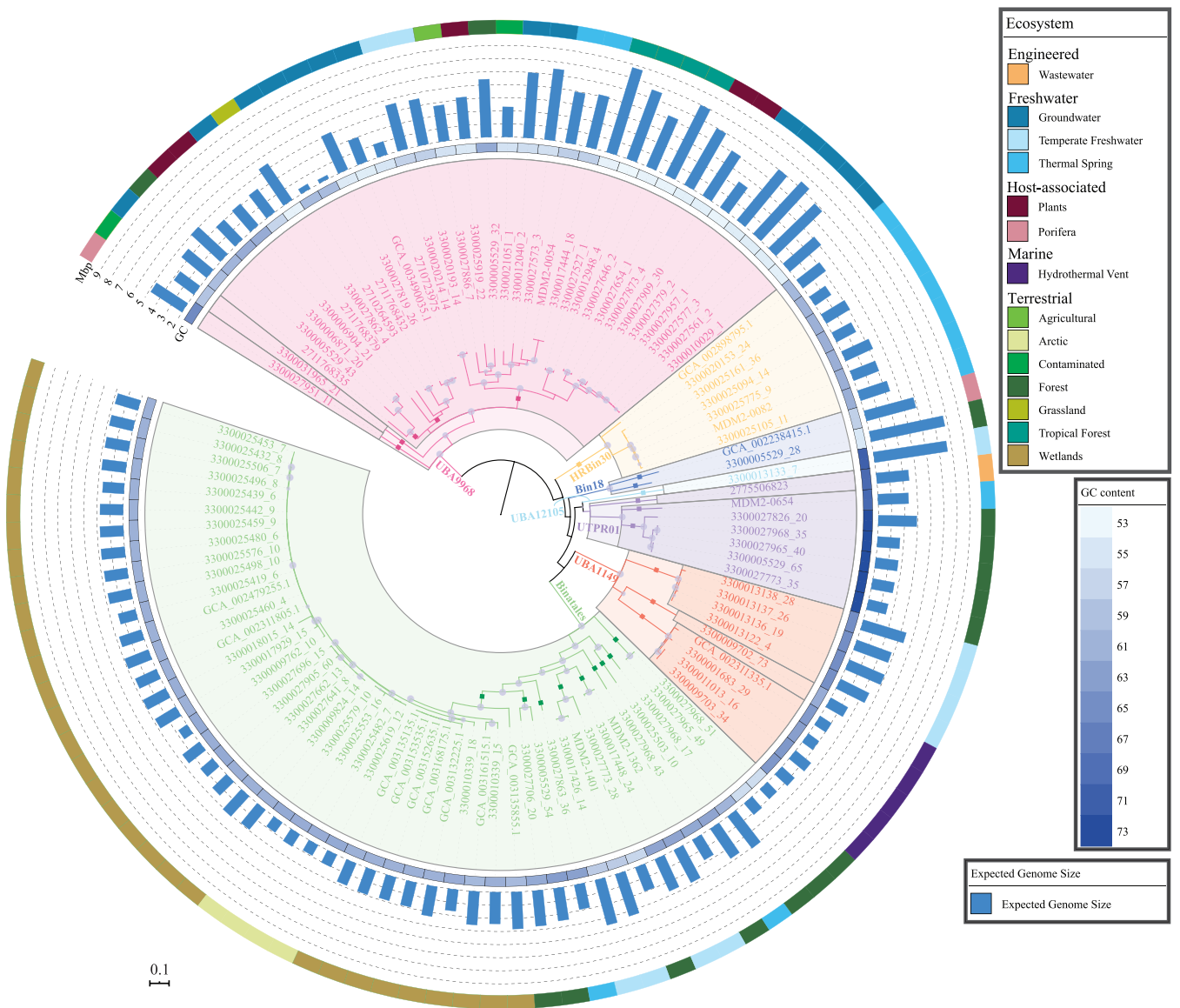
**KEYWORDS** Binatota, comparative genomics, environmental genomics, metagenomics, phylogenomics

Approaches that directly recover genomes from environmental samples, e.g., single-cell genomics and genome-resolved metagenomics, and hence bypass the hurdle of cultivation have come of age in the last decade. The resulting availability of environmentally sourced genomes, obtained as SAGs (single amplified genomes) or MAGs (metagenome-assembled genomes), is having a lasting impact on the field of microbial ecology. Distinct strategies are employed for the analysis of the deluge of obtained genomes. Site- or habitat-specific studies focus on spatiotemporal sampling of a single site or habitat of interest. Function-based studies focus on genomes from single or multiple habitats to identify and characterize organisms involved in a specific process, e.g., cellulose degradation (1) or sulfate reduction (2). Phylogeny-oriented (phylo-centric) studies, on the other hand, focus on characterizing genomes belonging to a specific lineage of interest. The aim of such studies is to delineate pan, core, and dispensable gene repertoires for a target lineage, document the lineage's defining metabolic capabilities (3, 4), understand the lineage's putative roles in various habitats (5, 6), and elucidate genomic basis underpinning the observed niche specializing patterns (7). The scope of phylo-centric studies could range from the analysis of a single genome from a single ecosystem (8) to global sampling and *in silico* analysis efforts (9, 10). The feasibility and value of phylo-centric strategies have recently been enhanced by the development of a genome-based (phylogenomic) taxonomic outline based on extractable data from MAGs and SAGs providing a solid framework for knowledge building and data communication (11), as well as recent efforts for massive, high-throughput binning of genomes from global collections of publicly available metagenomes in GenBank nr and Integrated Microbial Genomes & Microbiomes (IMG/M) database (12, 13). As such, these studies provide immensely useful information on potential metabolic capabilities and physiological preferences of yet-uncultured taxa. However, such *in silico* predictions require confirmation through enrichment, isolation, complementary cloning, and expression studies of the gene of interest or other functional genomics approaches to ascertain their phenotypic relevance.

Candidate phylum UBP10 has originally been described as one of the novel lineages recovered from a massive binning effort that reconstructed thousands of genomes from publicly available metagenomic data sets (12). UBP10 has subsequently been named candidate phylum Binatota (henceforth Binatota) in an effort to promote nomenclature for uncultured lineages based on attributes identified in MAGs and SAGs (14). The recent generation of 52,515 distinct MAGs binned from over 10,000 metagenomes (13) has greatly increased the number of available Binatota genomes. Here, we utilize a phylo-centric approach and present a comparative analysis of the putative metabolic and biosynthetic capacities and putative ecological roles of members of the candidate phylum Binatota, as based on sequence data from 108 MAGs. Our study documents aerobic methylotrophy, aerobic alkane degradation, and carotenoid pigmentation as defining traits in the Binatota. We also highlight the presence of incomplete chlorophyll biosynthetic pathways in all genomes and propose several evolutionary-grounded scenarios that could explain such a pattern.

## RESULTS

**Genomes analyzed in this study.** A total of 108 Binatota MAGs with >70% completion and <10% contamination were used for this study, which included 86 medium-quality (>50% completion, <10% contamination) and 22 high-quality (>90% completion, <5% contamination) genomes, as defined by MIMAG standards (15). Binatota genomes clustered into seven orders designated Bin18 ( $n=2$ ), Binatales ( $n=48$ ), HRBin30 ( $n=7$ ), UBA1149 ( $n=9$ ),



**FIG 1** Phylogenomic relationship between analyzed Binatota genomes. The maximum-likelihood tree was constructed in RAxML from a concatenated alignment of 120 single-copy marker genes. The tree was rooted using *Deferisoma camini* (GCA\_000526155.1) as the outgroup (not shown). Orders are shown as colored wedges: UBA9968, pink; HRBin30, tan; Bin18, blue; UBA12105, cyan; UTPRO1, purple; UBA1149, orange; and Binatales, green. Within each order, families are delineated by gray borders and genera are shown as colored squares on the branches. Bootstrap values are shown as purple bubbles for nodes with  $\geq 70\%$  support. The tracks around the tree represent (innermost-outermost) G+C content (with a heatmap that ranges from 53% [lightest] to 73% [darkest]), expected genome size (bar chart), and classification of the ecosystem from which the genome originated. All genomes analyzed in this study were  $>70\%$  complete and  $<10\%$  contaminated. Completion/contamination percentages and individual genomes assembly size are shown in Tables S2 and S3, respectively.

UBA9968 ( $n = 34$ ), UBA12105 ( $n = 1$ ), and UTPRO1 ( $n = 7$ ), encompassing 12 families and 24 genera (Fig. 1; Table S1). 16S rRNA gene sequences extracted from orders Bin18 and UBA9968 genomes were classified in SILVA (release 138) (16) as members of class bacteriap25 in the phylum *Myxococcota*, order Binatales, and order HRBin30 as uncultured phylum *RCP2-54* and orders UBA1149 and UTPRO1 as uncultured *Desulfobacterota* classes (Table S1). RDP II-classification (July 2017 release, accessed July 2020) classified all Binatota sequences as unclassified *Deltaproteobacteria* (Table S1).

**Methylotrophy in the Binatota: methanol.** With the exception of HRBin30, all orders encoded at least one type of methanol dehydrogenase (Figure 2a). Three distinct types of methanol dehydrogenases were identified (Figure 2a and b). (i) The NAD (P)-binding MDO/MNO-type methanol dehydrogenase (*mno*), typically associated with Gram-positive methylotrophic bacteria (*Actinobacteria* and *Bacillus methanolicus*) (17),





orders Bin18 and Binatales, together with the accessory XoxG c-type cytochrome and XoxJ periplasmic-binding proteins (Figure 2a). All later genomes also encoded PQQ biosynthesis. Surprisingly, none of the genomes encoded the MxaF1-type (MDH1) methanol dehydrogenase, typically encountered in model methylotrophs (19).

**Methylotrophy in the Binatota: methylamine.** All Binatota orders except UBA9968 encoded methylamine degradation capacity. The direct periplasmic route (methylamine dehydrogenase; *mau*) was more common, with *mauA* and *mauB* enzyme subunits encoded in Binatales, HRBin30, UBA1149, UBA12105, and UTPRO1 (Figure 2a; Extended Data set 1). Amicyanin (encoded by *mauC*) is the most probable electron acceptor for methylamine dehydrogenase (19) (Figure 2a). On the other hand, one Bin18 genome and two Binatales genomes (that also encode the *mau* cluster) carried the full complement of genes for methylamine oxidation via the indirect glutamate pathway (Figure 2a; Extended Data set 1).

**Methylotrophy in the Binatota: methylated sulfur compounds.** Binatota genomes encoded several enzymes involved in the degradation of dimethyl sulfone, methane sulfonic acid (MSA), and dimethyl sulfide (DMS). Nine genomes (2 Bin18 and 7 Binatales) encoded dimethyl sulfone monooxygenase (*sfnG*) involved in the degradation of dimethyl sulfone to MSA with the concomitant release of formaldehyde. Three of these nine genomes also encoded alkane sulfonic acid monooxygenase (*ssuD*), which will further degrade the MSA to formaldehyde and sulfite. Degradation of DMS via DMS monooxygenase (*dmoA*) to formaldehyde and sulfide was encountered in 13 genomes (2 Bin18, 9 Binatales, and 2 UBA9968). Further, one Binatales genome encoded the *dso* system (enzyme class [EC]: 1.14.13.245) for DMS oxidation to dimethyl sulfone, which could be further degraded to MSA as explained above (Figure 2a; Extended Data set 1).

**Methylotrophy in the Binatota: dihalogenated methane.** One Bin18 genome encoded the specific dehalogenase/glutathione S-transferase (*dcmA*) capable of converting dichloromethane to formaldehyde.

**Methylotrophy in the Binatota: methane.** Genes encoding copper membrane monooxygenases (CuMMOs), a family of enzymes that includes particulate methane monooxygenase (pMMO), were identified in orders Bin18 (2/2 genomes) and Binatales (9/48 genomes) (Figure 2a; Extended Data set 1), while genes encoding soluble methane monooxygenase (sMMO) were not found. A single copy of the three genes encoding all CuMMO subunits (A, B, and C) was encountered in 9 of the 11 genomes, while two copies were identified in two genomes. CuMMO subunit-encoding genes (A, B, and C) occurred as a contiguous unit in all genomes, with a CAB (5 genomes) and/or CAxB or CAxB (8 genomes, where x is a hypothetical protein) organization, similar to the pMMO operon structure in methanotrophic *Proteobacteria*, *Verrucomicrobia*, and “*Candidatus* Methylomirabilis” (NC10) (20–23) (Figure 2c). In addition, 5 of the above-mentioned 11 genomes also encoded a *pmoD* subunit, recently suggested to be involved in facilitating the enzyme complex assembly and/or in electron transfer to the enzyme’s active site (24, 25). Phylogenetic analysis of Binatota *pmoA* sequences revealed their affiliation with two distinct clades: the yet-uncultured cluster 2 TUSC (tropical upland soil cluster) methanotrophs (26) (2 Binatales genomes) and a clade encompassing *bmoA* sequences (putative butane monooxygenase gene A) from *Actinobacteria* (*Nocardioides* sp. strain CF8, *Mycolicibacterium*, and *Rhodococcus*) and SAR324 (“*Candidatus* Lambdaproteobacteria”) (27, 28) (Figure 2d). Members harboring these specific lineages have previously been identified in a wide range of environments, including soil (26). Previous studies have linked cluster 2 TUSC CuMMO-harboring organisms to methane oxidation based on selective enrichment on methane in microcosms derived from Lake Washington sediments (29). Binatota genomes encoding TUSC-affiliated CuMMO also harbored genes for downstream methanol and formaldehyde oxidation as well as formaldehyde assimilation (see below), providing further evidence for their putative involvement in methane oxidation. On the other hand, studies on *Nocardioides* sp. strain CF8 demonstrated its capacity to oxidize short-chain (C<sub>2</sub> to C<sub>4</sub>) hydrocarbons, but not methane, via its CuMMO, and its genome lacked

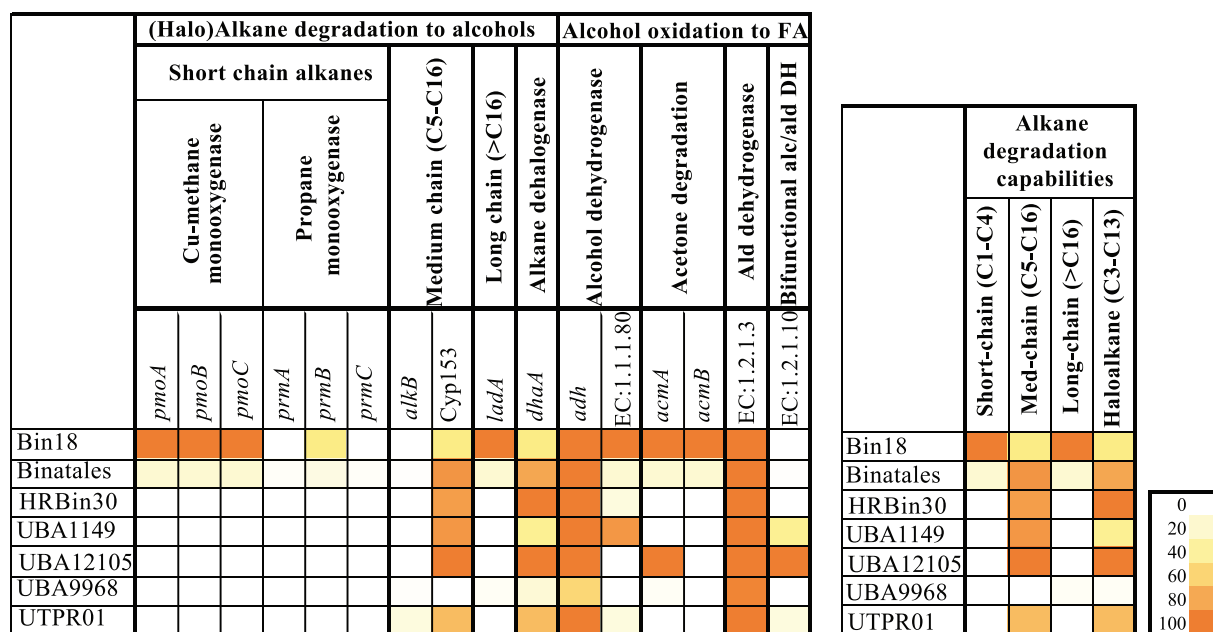
methanol dehydrogenase homologues (30). Such data favor a putative short-chain hydrocarbon degradation function for organisms encoding this type of CuMMO, although we note that five out of the nine Binatota genomes carrying SAR324/*Actinobacteria*-affiliated *pmoA* sequences also encoded at least one methanol dehydrogenase homologue. Modeling CuMMO subunits from both TUSC-type and *Actinobacteria*/SAR324-type Binatota genomes using *Methylococcus capsulatus* (Bath) 3D model (Protein Data Bank ID: 3RGB) revealed a heterotrimeric structure ( $\alpha_3\beta_3\gamma_3$ ) with the 7, 2, and 5 alpha helices of the PmoA, PmoB, and PmoC subunits, respectively, as well as the beta sheets characteristic of PmoA and PmoB subunits (Fig. S1). Recently, the location of the active site at the amino terminus of the PmoB subunit has been suggested (31). There has been recent debate as to the exact nuclearity of the Cu cofactor at the active site (31–33). Regardless of the nuclearity of the copper metal center, conserved histidine residues His<sup>33</sup>, His<sup>137</sup>, and His<sup>139</sup> (numbering following the *Methylococcus capsulatus* strain Bath PmoB subunit, Protein DataBank ID: 3RGB), thought to coordinate the Cu cofactor, were identified in all TUSC-affiliated and SAR324/*Actinobacteria*-affiliated Binatota CuMMO sequences (Fig. S1). Modeling PmoB subunits from both TUSC-type and *Actinobacteria*/SAR324-type Binatota genomes using *Methylococcus capsulatus* (Bath) PmoB subunit (Protein Data Bank [PDB] ID: 3RGB) predicted the binding pockets for Cu in Binatota sequences (Figure 2e).

As previously noted (19), methylotrophy requires the possession of three metabolic modules: C<sub>1</sub> oxidation to formaldehyde, formaldehyde oxidation to CO<sub>2</sub>, and formaldehyde assimilation. Formaldehyde generated by C<sub>1</sub> substrates oxidation is subsequently oxidized to formate and eventually CO<sub>2</sub>. Multiple pathways for formaldehyde oxidation to formate were identified in all Binatota orders (Text S1; Fig. S2). In addition, the majority of Binatota genomes encoded a formate dehydrogenase for formate oxidation to CO<sub>2</sub> (Text S1; Fig. S2). Finally, for assimilating formaldehyde into biomass, genes encoding all enzymes of the serine cycle, as well as genes encoding different routes of glyoxylate regeneration, were identified in all genomes (Text S1; Fig. S2).

**Alkane degradation in the Binatota.** In addition to methylotrophy and methanotrophy, Binatota genomes exhibited extensive short-, medium-, and long-chain alkanes degradation capabilities. In addition to the putative capacity of *Actinobacteria*/SAR324-affiliated CuMMO to oxidize C<sub>1</sub> to C<sub>5</sub> alkanes and C<sub>1</sub> to C<sub>4</sub> alkenes as described above, some Binatota genomes encoded propane-2-monoxygenase (*prmABC*), an enzyme mediating propane hydroxylation in the 2-position yielding isopropanol. Several genomes also encoded medium-chain-specific alkane hydroxylases, e.g., homologues of the nonheme iron *alkB* (34) and Cyp153-class alkane hydroxylases (35). The genomes also encoded multiple long-chain specific alkane monooxygenases, e.g., *ladA* homologues (enzyme class [EC]: 1.14.14.28) (36) (Fig. 3, Extended Data set 1). Finally, Binatota genomes encoded the capacity to metabolize medium-chain haloalkane substrates. All genomes encoded *dhaA* (haloalkane dehalogenases [EC: 3.8.1.5]) known to have a broad substrate specificity for medium-chain-length (C<sub>3</sub> to C<sub>10</sub>) mono- and dihaloalkanes, resulting in the production of their corresponding primary alcohol and haloalcohols, respectively (37) (Fig. 3, Extended Data set 1).

Alcohol and aldehyde dehydrogenases sequentially oxidize the resulting alcohols to their corresponding fatty acids or fatty acyl-CoA. Binatota genomes encode a plethora of alcohol and aldehyde dehydrogenases mediating such processes (Text S1; Fig. S3). As well, a complete fatty acid degradation machinery that enables all orders of the Binatota to degrade short-, medium-, and long-chain fatty acids to acetyl-CoA and propionyl-CoA was identified (Text S1; Fig. S3).

**Predicted electron transport chain.** All Binatota genomes encode an aerobic respiratory chain comprising complexes I and II, alternate complex III (ACIII, encoded by *actABCDEFG*), and complex IV, as well as an F-type H<sup>+</sup>-translocating ATP synthase (Text S1; Fig. 4). Binatota genomes also encode respiratory O<sub>2</sub>-tolerant H<sub>2</sub>-uptake [NiFe] hydrogenases, belonging to groups 1c (6 sequences), 1f (22 sequences), 1i (1 sequence), and 1h (4 sequences) (Fig. S4). Simultaneous oxidation of hydrogen (via type I respiratory O<sub>2</sub>-tolerant hydrogenases) and methane (via pMMO) has been shown to occur in methanotrophic *Verrucomicrobia* to

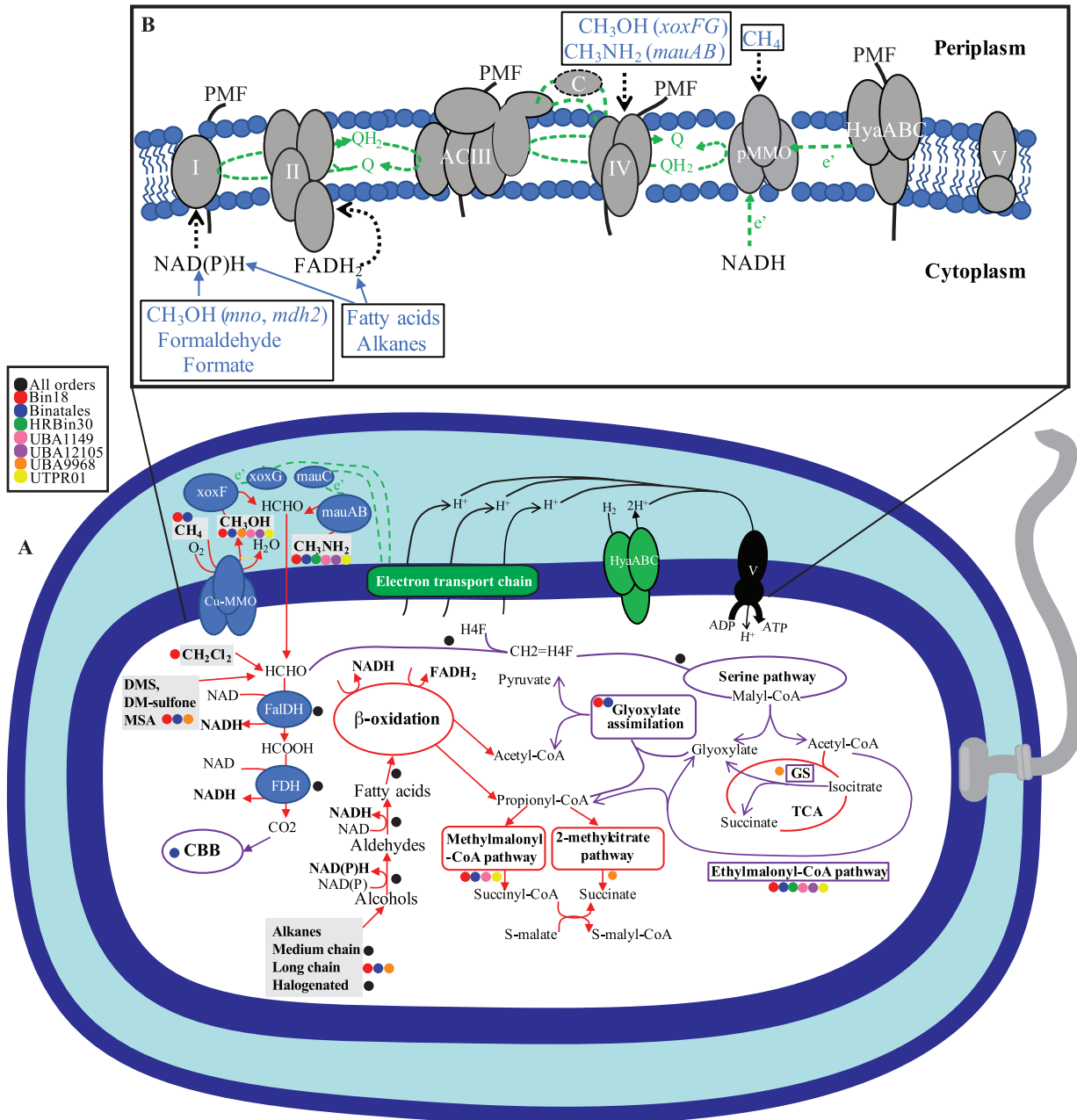


**FIG 3** Heatmap of the distribution of (halo)alkane degradation to alcohol in Binatota genomes. The heatmap colors (as explained in the key) correspond to the percentage of genomes in each order carrying a homologue of the gene in the column header. The per-order predicted alkane-degradation capacity is shown to the right as a heatmap with the colors corresponding to the percentage of genomes in each order where the full degradation pathway was detected for the substrate in the column header. These include CuMMO and/or *prmABC* for short-chain alkanes, *alkB* or Cyp153 for medium-chain alkanes, *ladA* for long-chain alkanes, and *dhaA* for haloalkanes. Ald, aldehyde.

maximize proton-motive force generation and subsequent ATP production (38). As well, some of the reduced quinones generated through H<sub>2</sub> oxidation are thought to provide reducing power for catalysis by pMMO (38) (Fig. 4). Details on the distribution of electron transport chain (ETC) components across Binatota orders are shown in Fig. S4, and the proposed electron flow under different growth conditions is presented in Text S1.

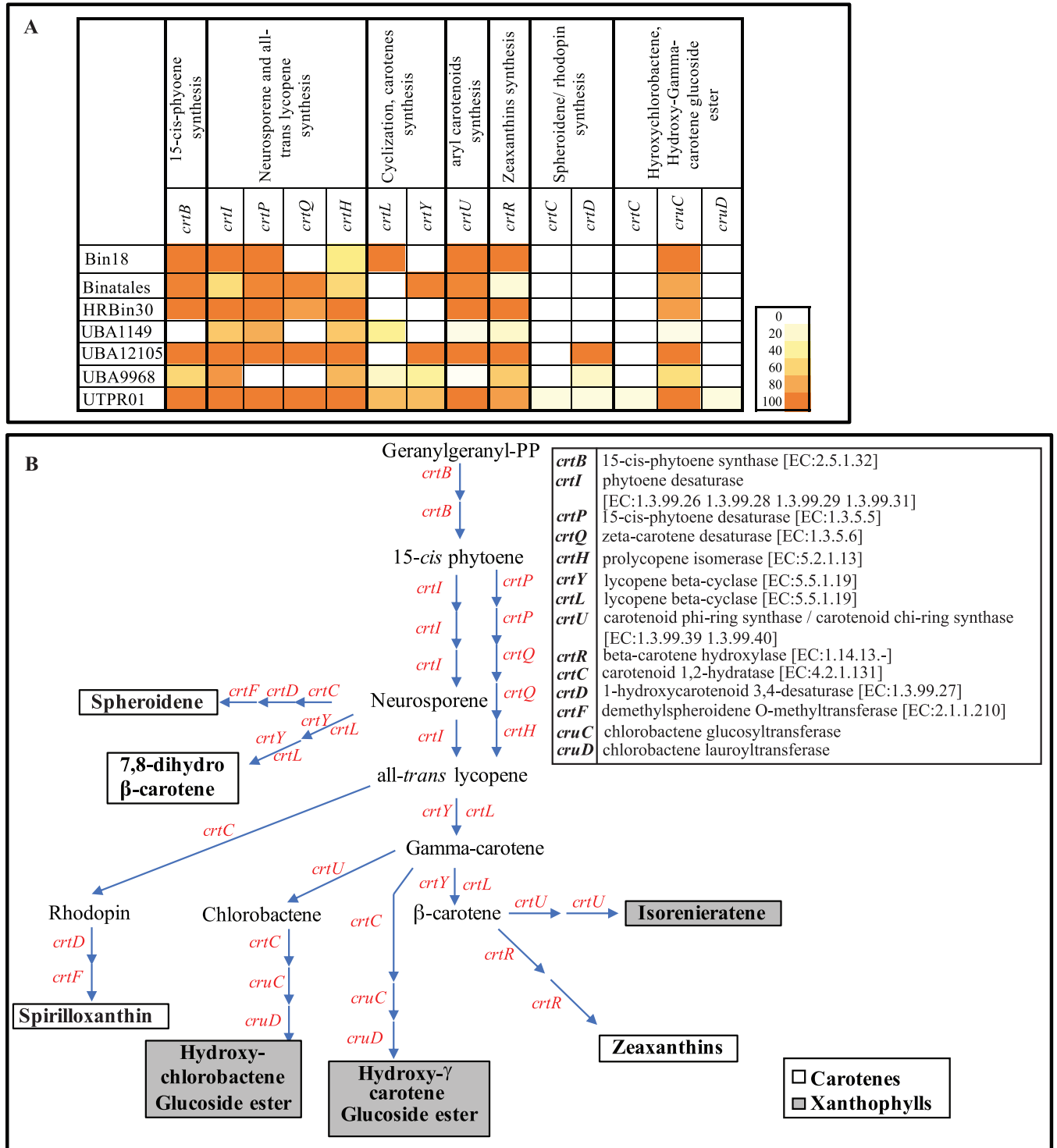
**Pigment production genes in the Binatota. Carotenoids.** Analysis of the Binatota genomes demonstrated a wide range of hydrocarbon (carotenes) and oxygenated (xanthophyll) carotenoid biosynthesis capabilities. Carotenoids biosynthetic machinery in the Binatota included *crtB* for 15-*cis*-phytylene synthesis from geranylgeranyl pyrophosphate (PP), *crtI*, *crtP*, *crtQ*, and *crtH* for neurosporene and all-*trans* lycopene formation from 15-*cis*-phytylene, *crtY* or *crtL* for gamma- and beta-carotene formation from all-*trans* lycopene, and a wide range of genes encoding enzymes for the conversion of neurosporene to spheroidene and 7,8-dihydro  $\beta$ -carotene, as well as the conversion of all-*trans* lycopene to spirilloxanthin, gamma-carotene to hydroxy-chlorobactene glucoside ester and hydroxy- $\gamma$ -carotene glucoside ester, and beta-carotene to isorenieratene and zeaxanthins (Fig. 5a and b; Extended Data set 1). Gene distribution pattern (Figure 5a; Extended Data set 1) predicts that all Binatota orders are capable of neurosporene and all-*trans* lycopene biosynthesis, and all but the order HRBin30 are capable of isorenieratene, zeaxanthin,  $\beta$ -carotene, and dihydro  $\beta$ -carotene biosynthesis and with specialization of order UTPRO1 in spirilloxanthin, spheroidene, hydroxy-chlorobactene, and hydroxy- $\gamma$ -carotene biosynthesis.

**Bacteriochlorophylls.** Surprisingly, homologues of multiple genes involved in bacteriochlorophyll biosynthesis were ubiquitous in Binatota genomes (Figure 6a to c). Bacteriochlorophyll biosynthesis starts with the formation of chlorophyllide *a* from protoporphyrin IX (Figure 6b). Within this pathway, genes encoding the first *bchl* (Mg-chelatase [EC: 6.6.1.1]), third *bchE* (magnesium-protoporphyrin IX monomethyl ester cyclase [EC: 1.21.98.3]), and fourth *bchLNB* (3,8-divinyl protochlorophyllide reductase [EC: 1.3.7.7]) steps were identified in the Binatota genomes (Fig. 6a and b; Extended Data set 1). However, homologues of genes encoding the second *bchM* (magnesium-protoporphyrin O-methyltransferase



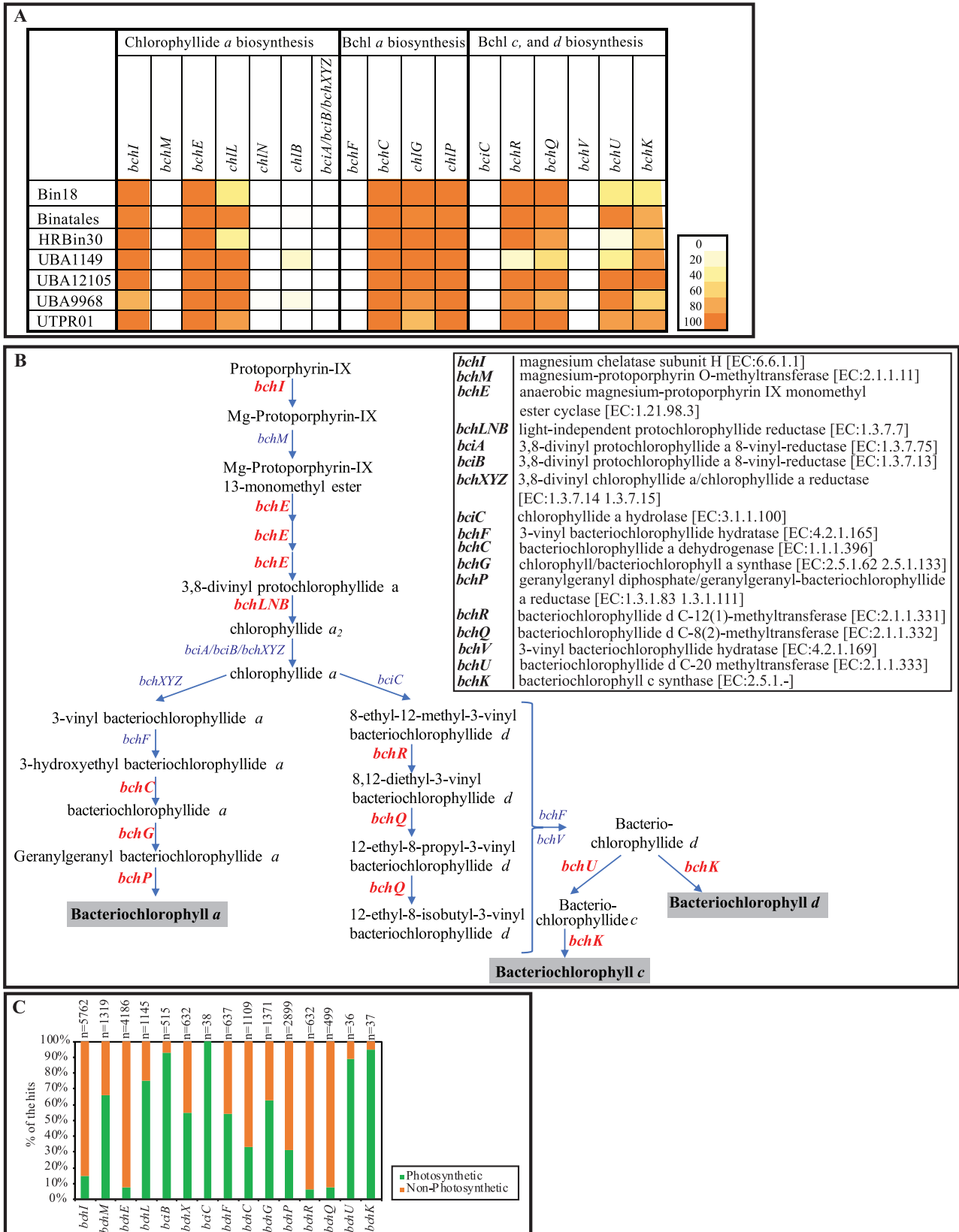
**FIG 4** (A) Cartoon depicting different metabolic capabilities encoded in the Binatota genomes with capabilities predicted for different orders shown as colored circles as shown in the legend. Enzymes for  $C_1$  metabolism are shown in blue and include the copper membrane monooxygenases (CuMMOs), methanol dehydrogenase (*xoxFG*), and methylamine dehydrogenase (*mauABC*), as well as the cytoplasmic formaldehyde dehydrogenase (FalDH) and formate dehydrogenase (FDH). Electron transport chain is shown as a green rectangle. Electron transfer from periplasmic enzymes to the ETC is shown as dotted green lines. The sites of proton extrusion to the periplasm are shown as black arrows, as is the F-type ATP synthase. Carbon dissimilation routes are shown as red arrows, while assimilatory routes are shown as purple arrows. Details of the assimilatory pathways are shown in Fig. S2 and S3. Reducing equivalents potentially fueling the ETC [NAD(P)H and FADH<sub>2</sub>] are shown in boldface. All substrates predicted to support growth are shown in boldface within gray boxes. A flagellum is depicted, the biosynthetic genes of which were identified in genomes belonging to all orders except Bin18, HRBin30, and UBA1149. The cell is depicted as rod-shaped based on the identification of the rod shape determining gene *rodA* in all genomes and the rod-shape determining genes *mreB* and *mreC* in genomes from all orders except UBA1149. The inset on top (B) details the electron transport chain in the Binatota with all electron transfer complexes (I, II, ACIII, IV) embedded in the inner membrane, along with the particulate methane monooxygenase (pMMO) and the H<sub>2</sub>-uptake [NiFe] hydrogenase (HyaABC). All genomes also encoded an F-type ATP synthase complex (V). Substrates potentially supporting growth are shown in blue with predicted entry points to the ETC shown as dotted black arrows. Sites of proton extrusion to the periplasm and proton motive force (PMF) creation are shown as solid black lines, while sites of electron (e<sup>-</sup>) transfer are shown as dotted green lines. Three possible physiological reductants are shown for pMMO (as dotted green arrows): the quinone pool coupled to ACIII, NADH, and/or some of the reduced quinones generated through H<sub>2</sub> oxidation by HyaABC. Abbreviations: CBB, Calvin Benson Bassham cycle; FalDH, NAD-linked glutathione-independent formaldehyde dehydrogenase, *fdhA*; FDH, NAD-dependent formate dehydrogenase (EC: 1.17.1.9); Fum, fumarate; GS, glyoxylate shunt; H<sub>4</sub>F, tetrahydrofolate; HyaABC, type I respiratory O<sub>2</sub>-tolerant H<sub>2</sub>-uptake [NiFe] hydrogenase; *mauABC*, methylamine dehydrogenase; CuMMO, copper membrane monooxygenases; *xoxFG*, *xoxF*-type methanol dehydrogenase; succ, succinate; TCA, tricarboxylic acid cycle; V, F-type ATP synthase (EC: 7.1.2.2 7.2.2.1).





**FIG 5** Carotenoids biosynthesis capabilities in Binatota genomes. (A) Distribution of carotenoid biosynthesis genes in the Binatota genomes. The heatmap colors (as explained in the key) correspond to the percentage of genomes in each order encoding a homologue of the gene in the column header. (B) Carotenoid biosynthesis scheme in Binatota based on the identified genes. Genes encoding enzymes catalyzing each step are shown in red and their descriptions with EC numbers are shown to the right. Binatota genomes encode the capability to biosynthesize both exclusively hydrocarbon carotenes (white boxes) or the oxygenated xanthophylls (gray boxes).

[EC: 2.1.1.11]) and the fifth *bciA* or *bicB* (3,8-divinyl protochlorophyllide *a* 8-vinyl-reductase) or *bchXYZ* (chlorophyllide *a* reductase [EC 1.3.7.15]) steps were absent (Figure 6a and b). A similar patchy distribution was observed in the pathway for bacteriochlorophyll *a* (BChl *a*) formation from chlorophyllide *a* (Figure 6b), where genes encoding *bchXYZ* (chlorophyllide *a* reductase



**FIG 6** Bacteriochlorophylls biosynthesis genes encountered in Binatota genomes studied suggesting an incomplete pathway for bacteriochlorophyll *a*, *c*, and/or *d* biosynthesis. (A) Distribution of chlorophyll biosynthesis genes in Binatota genomes. The heatmap colors (as explained in the key) correspond to the percentage of genomes in each order carrying a homologue of the gene in the column header. (B) Bacteriochlorophylls biosynthesis pathway. Genes identified in at least one Binatota genome are shown in red boldface text, while these with no homologues in the

(Continued on next page)

[EC 1.3.7.15]) and *bchF* (chlorophyllide *a* 3'-hydratase [EC 4.2.1.165]) were not identified, while genes encoding *bchC* (bacteriochlorophyllide *a* dehydrogenase [EC 1.1.1.396]), *bchG* (bacteriochlorophyll *a* synthase [EC: EC: 2.5.1.133]), and *bchP* (geranylgeranyl-bacteriochlorophyllide *a* reductase [EC 1.3.1.111]) were present in most genomes (Figure 6a; Extended Data set 1). Finally, within the pathway for bacteriochlorophylls *c* (BChl *c*) and *d* (BChl *d*) formation from chlorophyllide *a* (Figure 6b), genes for *bciC* (chlorophyllide *a* hydrolase [EC: 3.1.1.100]) and *bchF* (chlorophyllide *a* 3'-hydratase [EC: 4.2.1.165]) or *bchV* (3-vinyl bacteriochlorophyllide hydratase [EC: 4.2.1.169]) were not identified, while genes for *bchR* [bacteriochlorophyllide *d* C-12(1)-methyltransferase (EC: 2.1.1.331)], *bchQ* [bacteriochlorophyllide *d* C-8(2)-methyltransferase (EC: 2.1.1.332)], *bchU* (bacteriochlorophyllide *d* C-20 methyltransferase [EC: 2.1.1.333]), and *bchK* (bacteriochlorophyll *c* synthase [EC: 2.5.1.-]) were identified (Figure 6b; Extended Data set 1).

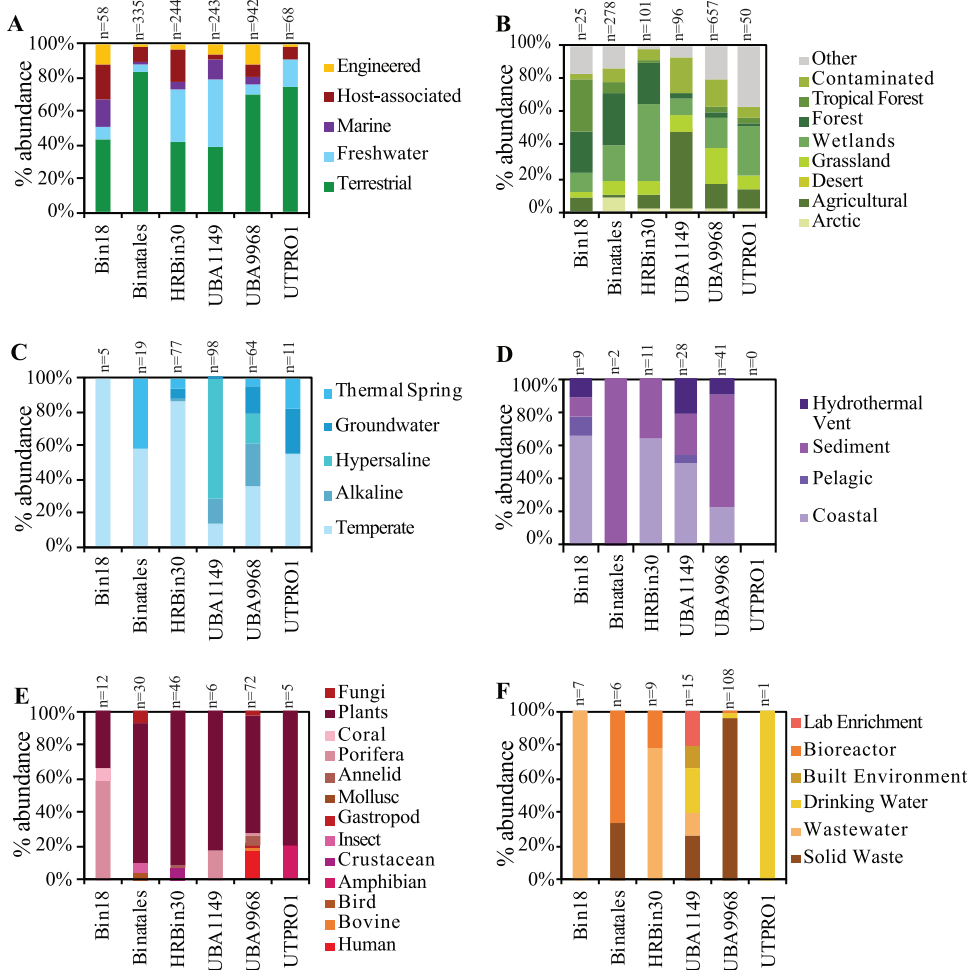
**Ecological distribution of the Binatota.** A total of 1,889 (GenBank nucleotide database [GenBank nt]) and 1,213 (IMG/M) 16S rRNA genes affiliated with the Binatota orders were identified (Extended Data set 2; Fig. 7; Fig. S5a). Analyzing their environmental distribution showed preference of Binatota to terrestrial soil habitats (39.5 to 83.0% of GenBank, 31.7 to 91.6% of IMG/M 16S rRNA gene sequences in various orders), as well as plant-associated (particularly rhizosphere) environments, although this could partly be attributed to sampling bias of these globally distributed and immensely important ecosystems (Figure 7a). On the other hand, a paucity of Binatota-affiliated sequences was observed in marine settings, with sequences absent or minimally present for Binatales, HRBin30, UBA9968, and UTPRO1 data sets (Figure 7a). The majority of sequences from marine origin were sediment-associated, being encountered in hydrothermal vents, deep marine sediments, and coastal sediments, with only the Bin18 sequences sampled from IMG/M showing representation in the vast, relatively well-sampled pelagic waters (Figure 7d).

In addition to the 16S rRNA-based analysis, we queried the data sets from which a Binatota MAG was binned using the sequence of their ribosomal protein S3 and estimating the Binatota relative abundance as the number of reads mapped to contigs with a Binatota ribosomal protein S3 as a percentage of the number of reads mapped to all contigs encoding a ribosomal protein S3 gene. Results showed relative abundances ranging between 0.1 and 10.21% (average  $3.84 \pm 3.21\%$ ) (Table S1).

In addition to phylum-wide patterns, order-specific environmental preferences were also observed. For example, in order Bin18, one of the two available genomes originated from the Mediterranean sponge *Aplysina aerophoba*. Analysis of the 16S rRNA data set suggests a notable association between Bin18 and sponges, with relatively high host-associated sequences (Figure 7a), the majority of which (58.3% NCBI-nt, 25.0% IMG/M) were recovered from the Porifera microbiome (Fig. 7e; Fig. S5f). Bin18-affiliated 16S rRNA gene sequences were identified in a wide range of sponges from 10 genera and 5 global habitat ranges (the Mediterranean genera *Ircinia*, *Petrosia*, *Chondrosia*, and *Aplysina*, the Caribbean genera *Agelas*, *Xestospongia*, and *Aaptos*, the Indo-West Pacific genus *Theonella*, the Pacific Dysideidae family, and the Great Barrier Reef genus *Rhopaloeides*), suggesting its widespread distribution beyond a single sponge species. The absolute majority of order Binatales sequences (83.0% NCBI-nt, 91.6% IMG/M) were of a terrestrial origin (Figure 7a; Fig. S5c), in addition to multiple rhizosphere-associated samples (7.5% NCBI-nt and 2.8% IMG/M, respectively) (Figure 7a; Fig. S5f). Notably, a relatively large proportion of Binatales soil sequences originated from either wetlands (peats, bogs) or forest soils (Fig. 7b; Fig. S5c), strongly suggesting the preference of the order Binatales to acidic and organic/methane-rich terrestrial habitats. This corresponds with the fact that 42 out of 48 Binatales genomes were recovered

#### FIG 6 Legend (Continued)

Binatota genomes are shown in blue text. Gene descriptions with EC numbers are shown to the right of the figure. (C) Distribution patterns of bacteriochlorophyll biosynthesis genes. The search was conducted in the functionally annotated bacterial tree of life AnnoTree (75) using single KEGG orthologies implicated in chlorophyll biosynthesis. Gene names are shown on the x axis, total number of hits is shown above the bar for each gene, and the percentage of hits in genomes from photosynthetic (green) versus nonphotosynthetic (orange) genera are in the stacked bars.



**FIG 7** Ecological distribution of Binatota-affiliated 16S rRNA sequences in GenBank nt database. Binatota orders are shown on the x axis, while percent abundance in different environments (classified based on the GOLD ecosystem classification scheme) is shown on the y axis (A). Further subclassifications for each environment are shown for (B) terrestrial, (C) freshwater, (D) marine, (E) host-associated, and (F) engineered environments. The total number of hit sequences for each order is shown above the bar graphs. Details, including GenBank accession number of hit sequences, are shown in Extended Data 2. Order UBA12105 genome assembly did not contain a16S rRNA gene, so this order is not included in the analysis.

from soil, 38 of which were from acidic wetland or forest soils (Fig. 1; Table S1). Genomes of UBA9968 were recovered from a wide range of terrestrial and nonmarine aquatic environments, and the observed 16S rRNA gene distribution verifies their ubiquity in all but marine habitats (Fig. 7a; Fig. S5b to g). Finally, while genomes from orders HRBin30, UBA1149, and UTPRO1 were recovered from limited environmental settings (thermal springs for HRBin30, gaseous hydrocarbon impacted habitats, e.g., marine hydrothermal vents and gas-saturated Lake Kivu, for UBA1149, and soil and hydrothermal environments for UTPRO1) (Fig. 1; Table S1), 16S rRNA gene analysis suggested their presence in a wide range of environments from each macro-scale environment classification (Fig. 7a; Fig. S5b to g).

**DISCUSSION**

**Expanding the world of methylotrophy.** The current study expands the list of lineages potentially capable of methylotrophy. An extensive repertoire of genes and pathways mediating the oxidation of multiple C<sub>1</sub> compounds to formaldehyde (Fig. 2 and 4), formaldehyde oxidation to CO<sub>2</sub> (Fig. S2), and formaldehyde assimilation pathways (Fig. S2) was identified, indicating that such capacity is a defining metabolic trait



in the Binatota. A certain degree of order-level substrate preference was observed, with potential utilization of methanol in all orders except HRBin30, methylamine in all orders except UBA9968, S-containing C<sub>1</sub> compound in Bin18, Binatales, and UBA9968, halogenated methane in Bin18, and possible methane utilization (methanotrophy) in Bin18 and Binatales (Figure 2a).

Aerobic methylotrophy has been documented in members of the alpha, beta, and gamma *Proteobacteria* (39), *Bacteroidetes* (40), *Actinobacteria* (e.g., genera *Arthrobacter* and *Mycobacterium*), *Firmicutes* (e.g., *Bacillus methanolicus*) (41), *Verrucomicrobia* (42), and “*Candidatus Methylomirabilis*” (NC10) (43). Further, studies employing genome-resolved metagenomics identified some signatures of methylotrophy, e.g., methanol oxidation (7, 44), formaldehyde oxidation/assimilation (45), and methylamine oxidation (7), in the *Gemmatimonadetes*, “*Candidatus Rokubacteria*,” *Chloroflexi*, *Actinobacteria*, *Acidobacteria*, and “*Ca. Lambdaproteobacteria*.” The possible contribution of Binatota to methane oxidation (methanotrophy) is especially notable, given the global magnitude of methane emissions and the relatively narrower range of organisms (*Proteobacteria*, *Verrucomicrobia*, and “*Candidatus Methylomirabilis*” [NC10]) (46) capable of this special type of methylotrophy. As described above, indirect evidence exists for the involvement of Binatota harboring TUSC-type CuMMO sequences in methane oxidation, while it is currently uncertain whether Binatota harboring SAR324/*Actinobacteria*-type CuMMO sequences are involved in oxidation of methane, gaseous alkanes, or both. pMMO of methanotrophs is also capable of oxidizing ammonia to hydroxylamine, which necessitates methanotrophs to employ hydroxylamine detoxification mechanisms (47). All 11 Binatota genomes encoding CuMMO also carried at least one homologue of *nir*, *nor*, and/or *nos* genes that could potentially convert harmful N-oxide byproducts to dinitrogen.

As previously noted (19), methylotrophy requires the possession of three metabolic modules: C<sub>1</sub> oxidation to formaldehyde, formaldehyde oxidation to CO<sub>2</sub>, and formaldehyde assimilation. Within the world of methylotrophs, a wide array of functionally redundant enzymes and pathways have been characterized that mediate various reactions and transformations in such modules. In addition, multiple combinations of different modules have been observed in methylotrophs, with significant variations existing even in phylogenetically related organisms. Our analysis demonstrates that such metabolic versatility indeed occurs within the methylotrophic modules of Binatota. While few phylum-wide characteristics emerged, e.g., utilization of serine pathway for formaldehyde assimilation, absence of H<sub>4</sub>MPT-linked formaldehyde oxidation, and potential utilization of PEP carboxykinase (*pckA*) rather than PEP carboxylase (*ppc*) for CO<sub>2</sub> entry to the serine cycle, multiple order-specific differences were observed, e.g., XoxF-type methanol dehydrogenase encoded by Bin18 and Binatales genomes, MDH2-type methanol dehydrogenase encoded by UBA1149 genomes, absence of methanol dehydrogenase homologues in HRBin30 genomes, absence of methylamine oxidation in order UBA9968, and potential utilization of the ethylmalonyl-CoA pathway for glyoxylate regeneration by the majority of the orders versus the glyoxylate shunt by UBA9968.

**Alkane degradation in the Binatota.** A second defining feature of the phylum Binatota, in addition to methylotrophy, is the widespread capacity for aerobic alkane degradation, as evident by the extensive arsenal of genes mediating aerobic degradation of short- (*prmABC*, propane monooxygenase), medium- (*alkB*, *cyp153*), and long-chain alkanes (*ladA*) identified (Fig. 3), in addition to complete pathways for odd- and even-numbered fatty acids oxidation (Fig. S3). Hydrocarbons, including alkanes, have been an integral part of the earth biosphere for eons, and a fraction of microorganisms have evolved specific mechanisms (O<sub>2</sub>-dependent hydroxylases and monooxygenases, anaerobic addition of fumarate) for their activation and conversion to central metabolites (48). Aerobic alkane-degradation capacity has so far been encountered in the *Actinobacteria*, *Proteobacteria*, *Firmicutes*, and *Bacteroidetes*, as well as in a few *Cyanobacteria* (48). As such, this study adds to the expanding list of phyla capable of aerobic alkane degradation.

**Metabolic traits explaining niche preferences in the Binatota.** Analysis of 16S rRNA gene data sets indicated that the Binatota display phylum-wide (preference to terrestrial habitats and methane/hydrocarbon-impacted habitats and rarity in pelagic marine environments) as well as order-specific (Bin18 in sponges, HRBin30 and UBA1149 in geothermal settings, Binatales in peats, bogs, and forest soils) habitat preferences (Fig. 7; Fig. S5). Such distribution patterns could best be understood in light of the phylum's predicted metabolic capabilities. Soils represent an important source of methane, generated through microoxic and anoxic niches within soil's complex architecture (49). Methane emission from soil is especially prevalent in peatlands, bogs, and wetlands, where incomplete aeration and net carbon deposition occurs. Indeed, anaerobic (50), fluctuating (51), and even oxic (52) wetlands represent one of the largest sources of methane emissions to the atmosphere. As well, terrestrial ecosystems represent a major source of global methanol emissions (53), with the release of methanol mediated mostly by demethylation reactions associated with pectin and other plant polysaccharides degradation.  $C_1$ -metabolizing microorganisms significantly mitigate methane and methanol release to the atmosphere from terrestrial ecosystems (54), and we posit that members of the Binatota identified in soils, rhizosphere, and wetlands contribute to such a process. The special preference of order Binatales to acidic peats, bogs, forests, and wetlands could reflect a moderate acidophilic specialization for this order and suggest their contribution to the process in these habitats.

Within the phylum Binatota, it appears that orders HRBin30 and UBA1149 are abundant in thermal vents, thermal springs, and thermal soils, suggesting a specialization to high-temperature habitats (Fig. 7). The presence of Binatota in such habitats could be attributed to high concentrations of alkanes typically encountered in such habitats. Hydrothermal vents display steep gradients of oxygen in their vicinity, emission of high levels of methane and other gaseous alkanes, and thermogenic generation of medium- and long-chain alkanes (55). Indeed, the presence and activity of aerobic hydrocarbon degraders in the vicinity of hydrothermal vents have been well established (27, 28, 56).

The recovery of Binatota genomes from certain lakes could be a reflection of the high gaseous load in such lakes. Multiple genomes and a large number of Binatota-affiliated 16S rRNA sequences were binned and identified from Lake Kivu, a meromictic lake characterized by unusually high concentrations of methane (57). Biotically, methane evolving from Lake Kivu is primarily oxidized by aerobic methanotrophs in surface waters (57–59), and members of the Binatota could contribute to this process. Binatota genomes were also recovered from sediments in Lake Washington, a location that has long served as a model for studying methylotrophy (60, 61). Steep counter gradients of methane and oxygen occurring in the lake's sediments enable aerobic methanotrophy to play a major role in controlling methane flux through the water column (62–65).

Finally, the occurrence and apparent wide distribution of members of the Binatota in sponges, particularly order Bin18, are notable and could possibly be viewed in terms of the wider symbiotic relationship between sponges and their microbiome. Presence of hydrocarbon degraders (66, 67), including methanotrophs (68), in the sponge microbiome has previously been noted, especially in deep-water sponges, where low levels of planktonic biomass restrict the amount of food readily acquired via filter feeding and hence biomass acquisition via methane and alkane oxidation is especially valuable.

**Carotenoid pigmentation: occurrence and significance.** The third defining feature of the Binatota, in addition to aerobic methylotrophy and alkane degradation, is the predicted capacity for carotenoid production. In photosynthetic organisms, carotenoids increase the efficiency of photosynthesis by absorbing in the blue-green region and then transferring the absorbed energy to the light-harvesting pigments (69). Carotenoid production also occurs in a wide range of nonphotosynthetic bacteria belonging to the *Alphaproteobacteria*, *Betaproteobacteria*, and *Gammaproteobacteria* (including methano- and methylotrophs) and *Bacteroidetes*, *Deinococcus*, *Thermus*,

*Deltaproteobacteria*, *Firmicutes*, *Actinobacteria*, *Planctomycetes*, and *Archaea*, e.g., *Halobacteriaceae* and *Sulfolobus*. Here, carotenoids could serve as antioxidants (70) and aid in radiation, UV, and desiccation resistance (71, 72). The link between carotenoid pigmentation and methylo/methanotrophy has long been observed (73), with the majority of known model *Alphaproteobacteria* and *Gammaproteobacteria* methano- and methylotrophs being carotenoid producers, although several Gram-positive methylotrophs (*Mycobacterium*, *Arthrobacter*, and *Bacillus*) are not pigmented. Indeed, root-associated facultative methylotrophs of the genus *Methylobacterium* have traditionally been referred to as “pink-pigmented facultative methylotrophs” and are seen as an integral part of root ecosystems (74). The exact reason for this correlation is currently unclear and could be related to the soil environment where they are prevalent, where periodic dryness and desiccation could occur, or to the continuous exposure of these aerobes in some habitats to light (e.g., in shallow sediments), necessitating protection from UV exposure.

**Chlorophyll biosynthesis genes in the Binatota.** Perhaps the most intriguing finding in this study is the identification of the majority of genes required for the biosynthesis of bacteriochlorophylls from protoporphyrin IX (6 out of 10 genes for bacteriochlorophyll *a* and 7 out of 11 genes for bacteriochlorophyll *c* and *d*). While such a pattern is tempting to propose phototrophic capacities in the Binatota, the consistent absence of critical genes (*bchM* methyltransferase, *bciA/bciB/bchXYZ* reductases, *bciC* hydrolase, and *bchF/V* hydratases), coupled with our inability to detect reaction center-encoding genes, prevents such a proclamation. Identification of a single or few gene shrapnel from the chlorophyll biosynthesis pathway in microbial genomes is not unique. Indeed, searching the functionally annotated bacterial tree of life AnnoTree (75) using single KEGG orthologies implicated in chlorophyll biosynthesis identifies multiple hits (in some cases thousands) in genomes from nonphotosynthetic organisms (Figure 6c). This is consistent with the identification of a *bchG* gene in a “*Candidatus* Bathyarchaeota” fosmid clone (76) and, more recently, a few bacteriochlorophyll synthesis genes in an Asgard genome (77). However, it should be noted that the high proportion of genes in the bacteriochlorophyll biosynthetic pathway identified in the Binatota genomes has never previously been encountered in nonphotosynthetic microbial genomes. Indeed, a search in AnnoTree for the combined occurrence of all seven bacteriochlorophyll synthesis genes identified in Binatota genomes yielded only photosynthetic organisms.

Accordingly, we put forward three scenarios to explain the proposed relationship between Binatota and phototrophy. The most plausible scenario, in our opinion, is that members of the Binatota are pigmented nonphotosynthetic organisms capable of carotenoid production but incapable of chlorophyll production and lack a photosynthetic reaction center. The second scenario posits that members of the Binatota are indeed phototrophs, possessing a complete pathway for chlorophyll biosynthesis and a novel type of reaction center that is bioinformatically unrecognizable. A minimal photosynthetic electron transport chain, similar to that of *Chloroflexus aurantiacus* (78), with the yet-unidentified reaction center, quinone, alternate complex III (or complex III), and some type of cytochrome *c* would possibly be functional. Under such scenario, members of the Binatota would be an extremely versatile photoheterotrophic facultative methylotrophic lineage. While such versatility, especially coupling methylotrophy to phototrophy, is rare (79), it has previously been observed in some *Rhodospirillaceae* species (80). A third scenario is that Binatota are capable of chlorophyll production but still incapable of conducting photosynthesis. Under this scenario, genes being missed in the pathway is due to shortcomings associated with *in silico* prediction and conservative gene annotation. For example, the missing *bchM* (EC: 2.1.1.11) could possibly be encoded by general methyltransferases (EC: 2.1.1.-), the missing *bciC* (EC: 3.1.1.100) could possibly be encoded by general hydrolases (EC: 3.1.1.-), and the missing *bchF* (EC: 4.2.1.165) or *bchV* (EC: 4.2.1.169) could possibly be encoded by general hydratases (EC: 4.2.1.-).

Encountering incomplete pathways in genomes of uncultured lineages is an exceedingly common occurrence in SAG and MAG analysis (81, 82). In many cases, this could plausibly indicate an incomplete contribution to a specific biogeochemical process, e.g., incomplete denitrification of nitrate to nitrite but not ammonia (82) or reduction of sulfite but not sulfate to sulfide (83), provided the thermodynamic feasibility of the proposed partial pathway and, preferably, prior precedence in pure cultures. In other cases, a pattern of absence of peripheral steps could demonstrate the capability for synthesis of a common precursor, e.g., synthesis of precorrin-2 from uroporphyrinogen but lack of the peripheral pathway for corrin ring biosynthesis leading to an auxotrophy for vitamin B<sub>12</sub>. Such auxotrophies are common in the microbial world and could be alleviated by nutrient uptake from the outside environment (84) or engagement in a symbiotic lifestyle (85). However, arguments for metabolic interdependencies, syntrophy, or auxotrophy could not be invoked to explain the consistent absence of specific genes in a dedicated pathway, such as bacteriochlorophyll biosynthesis, especially when analyzing a large number of genomes from multiple habitats. As such, we here raise awareness that using a certain occurrence threshold to judge a pathway's putative functionality could lead to misinterpretations of organismal metabolic capacities due to the frequent occurrence of partial, nonfunctional pathways and “gene shrapnel” in microbial genomes.

In conclusion, our work provides a comprehensive assessment of the yet-uncultured phylum Binatota and highlights its aerobic methylotrophic and alkane-degradation capacities, as well as its carotenoid production and abundance of bacteriochlorophyll synthesis genes in its genomes. Future efforts should focus on confirming these *in silico* predicted capabilities and characteristics through targeted enrichment and isolation efforts as well as functional genomics approaches. We also propose a role for this lineage in mitigating methanol and perhaps even methane emissions from terrestrial and freshwater ecosystems, alkanes degradation in hydrocarbon-rich habitats, and nutritional symbiosis with marine sponges. We present specific scenarios that could explain the unique pattern of chlorophyll biosynthesis gene occurrence and stress the importance of detailed analysis of pathways completion patterns for appropriate functional assignments in genomes of uncultured taxa.

## MATERIALS AND METHODS

**Genomes.** All genomes classified as belonging to the Binatota in the Genome Taxonomy Database (GTDB) database ( $n=22$  MAGs, April 2020) were downloaded as assemblies from NCBI. In addition, 128 metagenome-assembled genomes with the classification “Bacteria;UBP10” were downloaded from the IMG/M database (April 2020). These genomes were recently assembled from public metagenomes as part of a wider effort to generate a genomic catalogue of Earth's microbiome (13). Finally, 6 metagenome-assembled genomes were obtained as part of the Microbial Dark Matter MDM-II project. CheckM (86) was utilized for estimation of genome completeness, strain heterogeneity, and genome contamination. Only genomes with  $>70\%$  completion and  $<10\%$  contamination ( $n=108$ ) were retained for further analysis (Tables S1 and S2). MAGs were classified as high- or medium-quality drafts based on the criteria set forth by reference 15. The utilization of all publicly available genomes through prior individual efforts, as well as the global comprehensive Earth Microbiome collection, ensures the global scope of the survey conducted. Continuous addition of new data sets would certainly increase the number of available high-quality Binatota MAGs in the future.

**Phylogenetic analysis.** Taxonomic classifications followed the Genome Taxonomy Database (GTDB) release r89 (11, 87) and were carried out using the `classify_workflow` in GTDB-Tk (88) (v1.1.0). Phylogenomic analysis utilized the concatenated alignment of a set of 120 single-copy marker genes (11, 87) generated by the GTDB-Tk. Maximum-likelihood phylogenomic tree was constructed in RAxML (89) (with a cultured representative of the phylum Deferrisomatota as the outgroup). Small-subunit (SSU) rRNA gene-based phylogenetic analysis was also conducted using 16S rRNA gene sequences extracted from genomes using RNAmmer (90). Putative taxonomic ranks were deduced using average amino acid identity (AAI; calculated using AAI calculator [<http://enve-omics.ce.gatech.edu/>]), with the arbitrary cutoffs 56% and 68% for family and genus, respectively.

**Annotation.** Protein-coding genes in genomic bins were predicted using Prodigal (91). For initial prediction of function, pangenomes were constructed for each order in the phylum Binatota separately using PIRATE (92) with percent identity thresholds of 40, 45, 50, 55, 60, 65, 70, 75, 80, 90, a cd-hit step size of 1, and CD-HIT lowest percent identity of 90. The longest sequence for each PIRATE-identified allele was chosen as a representative and assembled into a pangenome. These pangenomes were utilized to gain preliminary insights on the metabolic capacities and structural features of different orders. BlastKOALA (93) was used to assign protein-coding genes in each of the pangenomes constructed to KEGG orthologies (KO), which were subsequently visualized using KEGG mapper (94). Analysis of specific capabilities and functions of interest was conducted on individual genomic bins by building and



scanning hidden Markov model (HMM) profiles. All predicted protein-coding genes in individual genomes were searched against custom-built HMM profiles for genes encoding  $C_1$ , alkanes, and fatty acids metabolism,  $C_1$  assimilation, [NiFe] hydrogenases, electron transport chain complexes, and carotenoid and chlorophyll biosynthesis. To build the HMM profiles, Uniprot reference sequences for all genes with an assigned KO number were downloaded and aligned using Clustal-omega (95), and the alignment was used to build an HMM profile using hmmbuild (HMMER 3.1b2). For genes not assigned a KO number (e.g., alternative complex III genes, different classes of cytochrome c family, cytochrome P450 medium-chain alkane hydroxylase cyp153, methanol dehydrogenase MNO/MDO family), a representative protein was compared against the KEGG Genes database using BLASTP and significant hits (those with E values  $< e^{-80}$ ) were downloaded and used to build HMM profiles as explained above. The custom-built HMM profiles were then used to scan the analyzed genomes for significant hits using hmmscan (HMMER 3.1b2) with the option -T 100 to limit the results to only those profiles with an alignment score of at least 100. Further confirmation was achieved through phylogenetic assessment and tree building procedures, in which potential candidates identified by hmmscan were aligned to the reference sequences used to build the custom HMM profiles using Clustal-omega (95), followed by maximum-likelihood phylogenetic tree construction using FastTree (96). Only candidates clustering with reference sequences were deemed true hits and were assigned to the corresponding KO.

**Search for photosynthetic reaction center.** Identification of genes involved in chlorophyll biosynthesis in Binatota genomes prompted us to search the genomes for photosynthetic reaction center genes. HMM profiles for reaction center type 1 (RC1; PsaAB) and reaction center type 2 (RC2; PufLM and PsaD<sub>1</sub>D<sub>2</sub>) were obtained from the pfam database (pfam00223 and pfam00124, respectively). Additionally, HMM profiles were built for PscABCD (*Chlorobia*-specific), PshA/B (*Heliobacteria*-specific) (97), and the newly identified Psa-like genes from *Chloroflexota* (98). The HMM profiles were used to search Binatota genomes for potential hits using hmmscan. To guard against overlooking a distantly related reaction center, we relaxed our homology criteria (by not including -T or -E options during the hmmscan). An additional search using a structurally informed reaction center alignment (97, 99) was also performed. The best potential hits were modeled using the SWISS-MODEL homology modeler (100) to check for veracity. Since the core subunits of type 1 RC proteins are predicted to have 11 transmembrane  $\alpha$ -helices, while type 2 RC are known to contain 5 transmembrane helices (101, 102), we also searched for all predicted proteins harboring either 5 or 11 transmembrane domains using TMHMM (103). All identified 5- or 11-helix-containing protein-coding sequences were searched against GenBank protein nr database to identify and exclude all sequences with a predicted function. All remaining 5- or 11-helix-containing proteins with no predicted function were then submitted to SWISS-MODEL homology modeler using the automated mode to predict homology models.

**Classification of [NiFe] hydrogenase sequences.** All sequences identified as belonging to the respiratory  $O_2$ -tolerant  $H_2$ -uptake [NiFe] hydrogenase large subunit (HyaA) were classified using the HydDB web tool (104).

**Particulate methane monooxygenase 3D model prediction and visualization.** SWISS-MODEL (100) was used to construct pairwise sequence alignments of predicted Binatota particulate methane monooxygenase with templates from *Methylococcus capsulatus* strain Bath (PDB: 3RGB) and to predict tertiary structure models. Predicted models were superimposed on the template enzyme in PyMol (version 2.0, Schrödinger, LLC). Modeling of the active site was conducted similarly. The dicopper-binding site proposed for *Methylococcus capsulatus* strain Bath pMMO (105) (PDB: 3RGB) was used. Alignment of Binatota PmoB sequences with reference *Methylococcus capsulatus* strain Bath PmoB was performed with Clustal-omega (95) and visualized using the ENDScript server (106).

**Ecological distribution of Binatota.** We queried 16S rRNA sequence databases using representative 16S rRNA gene sequences from six out of the seven Binatota orders (order UBA12105 genome assembly did not contain a 16S rRNA gene). Two databases were searched: (i) GenBank nucleotide (nt) database (accessed in July 2020) using a minimum identity threshold of 90%,  $\geq 80\%$  subject length alignment for near full-length query sequences or  $\geq 80\%$  query length for non-full-length query sequences, and a minimum alignment length of 100 bp and (ii) The IMG/M 16S rRNA public assembled metagenomes using a cutoff E value of  $1e^{-10}$ , percentage similarity of  $\geq 90\%$ , and either  $\geq 80\%$  subject length for full-length query sequences or  $\geq 80\%$  query length for non-full-length query sequences. Hits satisfying the above criteria were further trimmed after alignment to the reference sequences from each order using Clustal-omega and inserted into maximum-likelihood phylogenetic trees in FastTree (v 2.1.10, default settings). The ecological distribution for each of the Binatota orders was then deduced from the environmental sources of its hits. All environmental sources were classified according to the GOLD ecosystem classification scheme (107). We also queried the data sets from which the Binatota MAGs were binned using the sequence of their ribosomal protein S3. We estimated their relative abundance as the number of reads mapped to contigs with a Binatota ribosomal protein S3 as a percentage to the number of reads mapped to all contigs carrying a ribosomal protein S3 gene. More details on the specifics of the search are in Table S1 footnotes.

**Data availability.** Genomic bins, predicted proteins, and extended data for Fig. 2, 3, 5, and 6, Fig. S2 to S4, Fig. 7a to f, and Fig. S5b to g are available at <https://github.com/ChelseaMurphy/Binatota>. Maximum-likelihood trees (Fig. 1 and Fig. S5a) can be accessed at <https://itol.embl.de/shared/1WgxEjrQfEYWk>. Maximum-likelihood trees for chlorophyll biosynthesis genes are available at <https://itol.embl.de/shared/34y3BUHcQd7Lh>.

## SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

**TEXT S1**, PDF file, 0.3 MB.

**FIG S1**, PDF file, 1.4 MB.

**FIG S2**, PDF file, 0.2 MB.

**FIG S3**, PDF file, 0.3 MB.

**FIG S4**, PDF file, 0.3 MB.

**FIG S5**, PDF file, 0.4 MB.

**TABLE S1**, XLSX file, 0.02 MB.

**TABLE S2**, XLSX file, 0.02 MB.

**TABLE S3**, XLSX file, 0.02 MB.

## ACKNOWLEDGMENTS

This work has been supported by NSF grants 2016423 (to N.H.Y. and M.S.E.), 1441717, and 1826734 (to R.S.). We thank Kevin Redding (Arizona State University) for helpful discussions. Work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported under Contract number DE-AC02-05CH11231. J.R.S. is supported by NASA Astrobiology Rock Powered Life and was granted U.S. Forest Service permit number MLD15053 to conduct field work on Cone Pool and the Little Hot Creek, Mammoth Lakes, California. Work on the Paint Pots and Dewar Creek sites was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC DG 2019-06265) and facilitated by Parks Canada, BC Parks, and the Ktunaxa Nation. Thanks to students and participants of the 2014 to 2016 International Geobiology Course for research works on Cone Pool.

## REFERENCES

- Doud DFR, Bowers RM, Schulz F, Raad MD, Deng K, Tarver A, Glasgow E, Meulen KV, Fox B, Deutsch S, Yoshikuni Y, Northen T, Hedlund BP, Singer SW, Ivanova N, Woyke T. 2020. Function-driven single-cell genomics uncovers cellulose-degrading bacteria from the rare biosphere. *ISME J* 14:659–675. <https://doi.org/10.1038/s41396-019-0557-y>.
- Anantharaman K, Hausmann B, Jungbluth SP, Kantor RS, Lavy A, Warren LA, Rappe MS, Pester M, Loy A, Thomas BC, Banfield JF. 2018. Expanded diversity of microbial groups that shape the dissimilatory sulfur cycle. *ISME J* 12:1715–1728. <https://doi.org/10.1038/s41396-018-0078-0>.
- Becraft ED, Woyke T, Jarett J, Ivanova N, Godoy-Vitorino F, Poulton N, Brown JM, Brown J, Lau MCY, Onstott T, Eisen JA, Moser D, Stepanauskas R. 2017. Rokubacteria: genomic giants among the uncultured bacterial phyla. *Front Microbiol* 8:2264. <https://doi.org/10.3389/fmicb.2017.02264>.
- Rinke R, Rubino F, Messer LF, Youssef N, Parks DH, Chuvochina M, Brown M, Jeffries T, Tyson GW, Seymour JR, Hugenholtz P. 2019. A phylogenomic and ecological analysis of the globally abundant Marine Group II archaea (*Ca. Poseidoniales* ord. nov.). *ISME J* 13:663–675. <https://doi.org/10.1038/s41396-018-0282-y>.
- Farag IF, Davis JP, Youssef NH, Elshahed MS. 2014. Global patterns of abundance, diversity and community structure of the Aminicenantes (candidate phylum OP8). *PLoS One* 9:e92139. <https://doi.org/10.1371/journal.pone.0092139>.
- Hu P, Dubinsky EA, Probst AJ, Wang J, Sieber CMK, Tom LM, Gardinali PR, Banfield JF, Atlas RM, Andersen GL. 2017. Simulation of Deepwater Horizon oil plume reveals substrate specialization within a complex community of hydrocarbon degraders. *Proc Natl Acad Sci U S A* 114:7432–7437. <https://doi.org/10.1073/pnas.1703424114>.
- Zhou Z, Tran PQ, Kieft K, Anantharaman K. 2020. Genome diversification in globally distributed novel marine Proteobacteria is linked to environmental adaptation. *ISME J* 14:2060–2077. <https://doi.org/10.1038/s41396-020-0669-4>.
- Youssef NH, Blainey PC, Quake SR, Elshahed MS. 2011. Partial genome assembly for a candidate division OP11 single cell from an anoxic spring (Zodletone Spring, Oklahoma). *Appl Environ Microbiol* 77:7804–7814. <https://doi.org/10.1128/AEM.06059-11>.
- Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng JF, Darling A, Malfatti S, Swan BK, Gies EA, Dodsworth JA, Hedlund BP, Tsiamis G, Sievert SM, Liu WT, Eisen JA, Hallam SJ, Kyrpides NC, Stepanauskas R, Rubin EM, Hugenholtz P, Woyke T. 2013. Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 499:431–437. <https://doi.org/10.1038/nature12352>.
- Beam JP, Becraft ED, Brown KM, Schulz F, Jarett JK. 2020. Ancestral absence of electron transport chains in Patescibacteria and DPANN. *Front Microbiol* 11:1848. <https://doi.org/10.3389/fmicb.2020.01848>.
- Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil PA, Hugenholtz P. 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* 36:996–1004. <https://doi.org/10.1038/nbt.4229>.
- Parks DH, Rinke C, Chuvochina M, Chaumeil PA, Woodcroft BJ, Evans PN, Hugenholtz P, Tyson GW. 2017. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat Microbiol* 2:1533–1542. <https://doi.org/10.1038/s41564-017-0012-7>.
- Nayfach S, Roux S, Seshadri R, Udwy D, Varghese N, Schulz F, Wu D, Paez-Espino D, Chen I-M, Huntemann M, Palaniappan K, Ladau J, Mukherjee S, Reddy TBK, Nielsen T, Kirton E, Faria JP, Edirisinghe JN, Henry CS, Jungbluth SP, Chivian D, Dehal P, Wood-Charlson EM, Arkin AP, Tringe SG, Visel A, Woyke T, Mouncey NJ, Ivanova NN, Kyrpides NC, Eloe-Fadrosh EA, IMG/M Data Consortium. 2021. A genomic catalogue of Earth's microbiomes. *Nat Biotechnol* 39:499–509. <https://doi.org/10.1038/s41587-020-0718-6>.
- Chuvochina M, Rinke C, Parks DH, Rapp MS, Tyson GW, Yilmaz P, Whitman WB, Hugenholtz P. 2019. The importance of designating type material for uncultured taxa. *Syst Appl Microbiol* 42:15–21. <https://doi.org/10.1016/j.syapm.2018.07.003>.
- Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, Schulz F, Jarett J, Rivers AR, Eloe-Fadrosh EA, Tringe SG, Ivanova NN, Copeland A, Clum A, Becraft ED, Malmstrom RR, Birren B, Podar M, Bork P, Weinstock GM, Garrity GM, Dodsworth JA, Yooseph S, Sutton G, Glöckner FO, Gilbert JA, Nelson WC, Hallam SJ, Jungbluth SP, Ettema TJG, Tighe S, Konstantinidis KT, Liu W-T, Baker BJ, Rattei T, Eisen JA, Hedlund B, McMahon KD, Fierer N, Knight R, Finn R, Cochrane G, Karsch-Mizrachi I, Tyson GW, Rinke C, Lapidus A, Meyer F, Yilmaz P, Parks DH, Murat Eren A, The Genome Standards Consortium, et al. 2017. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol* 35:725–731. <https://doi.org/10.1038/nbt.3893>.
- Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2012. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 41:D590–D596. <https://doi.org/10.1093/nar/gks1219>.
- Hektor HJ, Kloosterman H, Dijkhuizen L. 2000. Nicotinoprotein methanol dehydrogenase enzymes in Gram-positive methylotrophic bacteria. *J Mol Cat B* 8:103–109. [https://doi.org/10.1016/S1381-1177\(99\)00073-9](https://doi.org/10.1016/S1381-1177(99)00073-9).

18. Kalyuzhnaya MG, Hristova KR, Lidstrom ME, Chistoserdova L. 2008. Characterization of a novel methanol dehydrogenase in representatives of Burkholderiales: implications for environmental detection of methylotrophy and evidence for convergent evolution. *J Bacteriol* 190:3817–3823. <https://doi.org/10.1128/JB.00180-08>.
19. Chistoserdova L. 2011. Modularity of methylotrophy, revisited. *Environ Microbiol* 13:2603–2622. <https://doi.org/10.1111/j.1462-2920.2011.02464.x>.
20. Erikstad HA, Jensen S, Keen TJ, Birkeland NK. 2012. Differential expression of particulate methane monooxygenase genes in the verrucomicrobial methanotroph 'Methylacidiphilum kamchatkense' Kam1. *Extremophiles* 16:405–409. <https://doi.org/10.1007/s00792-012-0439-y>.
21. Ettwig KF, Butler MK, Le Paslier D, Pelletier E, Mangenot S, Kuypers MM, Schreiber F, Dutilh BE, Zedelius J, de Beer D, Gloerich J, Wessels HJ, van Alen T, Luesken F, Wu ML, van de Pas-Schoonen KT, Op den Camp HJ, Janssen-Megens EM, Francoijs KJ, Stunnenberg H, Weissenbach J, Jetten MS, Strous M. 2010. Nitrite-driven anaerobic methane oxidation by oxygenic bacteria. *Nature* 464:543–548. <https://doi.org/10.1038/nature08883>.
22. Iguchi H, Yurimoto H, Sakai Y. 2010. Soluble and particulate methane monooxygenase gene clusters of the type I methanotroph *Methylovulum miyakonense* HT12. *FEMS Microbiol Lett* 312:71–76. <https://doi.org/10.1111/j.1574-6968.2010.02101.x>.
23. Ricke P, Erkel C, Kube M, Reinhardt R, Liesack W. 2004. Comparative analysis of the conventional and novel pmo (particulate methane monooxygenase) operons from *Methylocystis* strain SC2. *Appl Environ Microbiol* 70:3055–3063. <https://doi.org/10.1128/AEM.70.5.3055-3063.2004>.
24. Fisher OS, Kenney GE, Ross MO, Ro SY, Lemma BE, Batelu S, Thomas PM, Sosnowski VC, DeHart CJ, Kelleher NL, Stemmler TL, Hoffman BM, Rosenzweig AC. 2018. Characterization of a long overlooked copper protein from methane- and ammonia-oxidizing bacteria. *Nat Commun* 9:4276. <https://doi.org/10.1038/s41467-018-06681-5>.
25. Rochman FF, Kwon M, Khadka R, Tamas I, Lopez-Jauregui AA, Sheremet A, V Smirnova A, Malmstrom RR, Yoon S, Woyke T, Dunfield PF, Verbeke TJ. 2020. Novel copper-containing membrane monooxygenases (CuMMOs) encoded by alkane-utilizing Betaproteobacteria. *ISME J* 14:714–726. <https://doi.org/10.1038/s41396-019-0561-2>.
26. Knief C. 2015. Diversity and habitat preferences of cultivated and uncultivated aerobic methanotrophic bacteria evaluated based on pmoA as molecular marker. *Front Microbiol* 6:1346.
27. Li M, Jain S, Baker BJ, Taylor C, Dick GJ. 2014. Novel hydrocarbon monooxygenase genes in the metatranscriptome of a natural deep-sea hydrocarbon plume. *Environ Microbiol* 16:60–71. <https://doi.org/10.1111/1462-2920.12182>.
28. Sheik CS, Jain S, Dick GJ. 2014. Metabolic flexibility of enigmatic SAR324 revealed through metagenomics and metatranscriptomics. *Environ Microbiol* 16:304–317. <https://doi.org/10.1111/1462-2920.12165>.
29. Kalyuzhnaya MG, Zabinsky R, Bowerman S, Baker DR, Lidstrom ME, Chistoserdova L. 2006. Fluorescence in situ hybridization-flow cytometry-cell sorting-based method for separation and enrichment of type I and type II methanotroph populations. *Appl Environ Microbiol* 72:4293–4301. <https://doi.org/10.1128/AEM.00161-06>.
30. Hamamura N, Yeager CM, Arp DJ. 2001. Two distinct monooxygenases for alkane oxidation in *Nocardioides* sp. strain CF8. *Appl Environ Microbiol* 67:4992–4998. <https://doi.org/10.1128/AEM.67.11.4992-4998.2001>.
31. Balasubramanian R, Smith SM, Rawat S, Yatsunyk LA, Stemmler TL, Rosenzweig AC. 2010. Oxidation of methane by a biological dicopper centre. *Nature* 465:115–119. <https://doi.org/10.1038/nature08992>.
32. Cao L, Caldararu O, Rosenzweig AC, Ryde U. 2018. Quantum refinement does not support dinuclear copper sites in crystal structures of particulate methane monooxygenase. *Angew Chem Int Ed* 57:162–166. <https://doi.org/10.1002/anie.201708977>.
33. Ross MO, MacMillan F, Wang J, Nisthal A, Lawton TJ, Olafson BD, Mayo SL, Rosenzweig AC, Hoffman BM. 2019. Particulate methane monooxygenase contains only mononuclear copper centers. *Science* 364:566–570. <https://doi.org/10.1126/science.aav2572>.
34. Chen Q, Janssen DB, Witholt B. 1995. Growth on octane alters the membrane lipid fatty acids of *Pseudomonas oleovorans* due to the induction of alkB and synthesis of octanol. *J Bacteriol* 177:6894–6901. <https://doi.org/10.1128/JB.177.23.6894-6901.1995>.
35. van Beilen JB, Funhoff EG. 2007. Alkane hydroxylases involved in microbial alkane degradation. *Appl Microbiol Biotechnol* 74:13–21. <https://doi.org/10.1007/s00253-006-0748-0>.
36. Li L, Liu X, Yang W, Xu F, Wang W, Feng L, Bartlam M, Wang L, Rao Z. 2008. Crystal structure of long-chain alkane monooxygenase (LadA) in complex with coenzyme FMN: unveiling the long-chain alkane hydroxylase. *J Mol Biol* 376:453–465. <https://doi.org/10.1016/j.jmb.2007.11.069>.
37. Nagata Y, Miyauchi K, Damborsky J, Manova K, Anvorgova A, Takagi M. 1997. Purification and characterization of a haloalkane dehalogenase of a new substrate class from a gamma-hexachlorocyclohexane-degrading bacterium, *Sphingomonas paucimobilis* UT26. *Appl Environ Microbiol* 63:3707–3710. <https://doi.org/10.1128/AEM.63.9.3707-3710.1997>.
38. Carere CR, Hards K, Houghton KM, Power JF, McDonald B, Collet C, Gapes DJ, Sparling R, Boyd ES, Cook GM, Greening C, Stott MB. 2017. Mixotrophy drives niche expansion of verrucomicrobial methanotrophs. *ISME J* 11:2599–2610. <https://doi.org/10.1038/ismej.2017.112>.
39. Chistoserdova L, Kalyuzhnaya MG, Lidstrom ME. 2009. The expanding world of methylotrophic metabolism. *Annu Rev Microbiol* 63:477–499. <https://doi.org/10.1146/annurev.micro.091208.073600>.
40. Boden R, Thomas E, Savani P, Kelly DP, Wood AP. 2008. Novel methylotrophic bacteria isolated from the River Thames (London, UK). *Environ Microbiol* 10:3225–3236. <https://doi.org/10.1111/j.1462-2920.2008.01711.x>.
41. McTaggart TL, Beck DA, Setboonsarng U, Shapiro N, Woyke T, Lidstrom ME, Kalyuzhnaya MG, Chistoserdova L. 2015. Genomics of methylotrophy in Gram-positive methylamine-utilizing bacteria. *Microorganisms* 3:94–112. <https://doi.org/10.3390/microorganisms3010094>.
42. Pol A, Heijmans K, Harhangi HR, Tedesco D, Jetten MSM, Op den Camp HJM. 2007. Methanotrophy below pH 1 by a new Verrucomicrobia species. *Nature* 450:874–878. <https://doi.org/10.1038/nature06222>.
43. Ettwig KF, van Alen T, van de Pas-Schoonen KT, Jetten MS, Strous M. 2009. Enrichment and molecular detection of denitrifying methanotrophic bacteria of the NC10 phylum. *Appl Environ Microbiol* 75:3656–3662. <https://doi.org/10.1128/AEM.00067-09>.
44. Diamond S, Andeer PF, Li Z, Crits-Christoph A, Burstein D, Anantharaman K, Lane KR, Thomas BC, Pan C, Northen TR, Banfield JF. 2019. Mediterranean grassland soil C-N compound turnover is dependent on rainfall and depth, and is mediated by genomically divergent microorganisms. *Nat Microbiol* 4:1356–1367. <https://doi.org/10.1038/s41564-019-0449-y>.
45. Butterfield CN, Li Z, Andeer PF, Spaulding S, Thomas BC, Singh A, Hettich RL, Suttle KB, Probst AJ, Tringe SG, Northen T, Pan C, Banfield JF. 2016. Proteogenomic analyses indicate bacterial methylotrophy and archaeal heterotrophy are prevalent below the grass root zone. *PeerJ* 4:e2687. <https://doi.org/10.7717/peerj.2687>.
46. Khmelenina VN, Colin Murrell J, Smith TJ, Trotsenko YA. 2018. Physiology and biochemistry of the aerobic methanotrophs, p 1–25. *In* Rojo F (ed), *Aerobic utilization of hydrocarbons, oils and lipids*. Springer International Publishing, Cham. [https://doi.org/10.1007/978-3-319-39782-5\\_4-1](https://doi.org/10.1007/978-3-319-39782-5_4-1).
47. Mohammadi SS, Pol A, van Alen T, Jetten MSM, Op den Camp HJM. 2017. Ammonia oxidation and nitrite reduction in the verrucomicrobial methanotroph *Methylacidiphilum fumarolicum* SolV. *Front Microbiol* 8:1901. <https://doi.org/10.3389/fmicb.2017.01901>.
48. Prince RC, Amade TJ, McGenity TJ. 2019. Prokaryotic hydrocarbon degraders, p 1–39. *In* McGenity TJ (ed), *Taxonomy, genomics and ecophysiology of hydrocarbon-degrading microbes*. Springer International Publishing, Cham. [https://doi.org/10.1007/978-3-030-14796-9\\_15](https://doi.org/10.1007/978-3-030-14796-9_15).
49. Le Mer J, Roger P. 2001. Production, oxidation, emission and consumption of methane by soils: a review. *Eur J Soil Biol* 37:25–50. [https://doi.org/10.1016/S1164-5563\(01\)01067-6](https://doi.org/10.1016/S1164-5563(01)01067-6).
50. Wang Z, Zeng D, Patrick WH. 1996. Methane emissions from natural wetlands. *Environ Monit Assess* 42:143–161. <https://doi.org/10.1007/BF00394047>.
51. He S, Malfatti SA, McFarland JW, Anderson FE, Pati A, Huntemann M, Tremblay J, Glavina del Rio T, Waldrop MP, Windham-Myers L, Tringe SG. 2015. Patterns in wetland microbial community composition and functional gene repertoire associated with methane emissions. *mBio* 6:e00066-15. <https://doi.org/10.1128/mBio.00066-15>.
52. Angle JC, Morin TH, Solden LM, Narowe AB, Smith GJ, Borton MA, Rey-Sanchez C, Daly RA, Mirfenderesgi G, Hoyt DW, Riley WJ, Miller CS, Bohrer G, Wrighton KC. 2017. Methanogenesis in oxygenated soils is a substantial fraction of wetland methane emissions. *Nat Commun* 8:1567. <https://doi.org/10.1038/s41467-017-01753-4>.
53. Kolb S. 2009. Aerobic methanol-oxidizing bacteria in soil. *FEMS Microbiol Lett* 300:1–10. <https://doi.org/10.1111/j.1574-6968.2009.01681.x>.
54. Conrad R. 2009. The global methane cycle: recent advances in understanding the microbial processes involved. *Environ Microbiol Rep* 1:285–292. <https://doi.org/10.1111/j.1758-2229.2009.00038.x>.
55. McCollom TM. 2013. Laboratory simulations of abiotic hydrocarbon formation in Earth's deep subsurface. *Rev Mineral Geochem* 75:467–494. <https://doi.org/10.2138/rmg.2013.75.15>.



56. Wang W, Li Z, Zeng L, Dong C, Shao Z. 2020. The oxidation of hydrocarbons by diverse heterotrophic and mixotrophic bacteria that inhabit deep-sea hydrothermal ecosystems. *ISME J* 14:1994–2006. <https://doi.org/10.1038/s41396-020-0662-y>.
57. Pasche N, Schmid M, Vazquez F, Schubert CJ, Wüest A, Kessler JD, Pack MA, Reeburgh WS, Bürgmann H. 2011. Methane sources and sinks in Lake Kivu. *J Geophys Res Biogeosciences* 116:G03006.
58. Borges AV, Abril G, Delille B, Descy J-P, Darchambeau F. 2011. Diffusive methane emissions to the atmosphere from Lake Kivu (Eastern Africa). *J Geophys Res Biogeosciences* 116:G03032.
59. Llíros M, Descy J-P, Libert X, Morana C, Schmitz M, Wimba L, Nzavugazere A, García-Armisen T, Borrego C, Servais P, Darchambeau F. 2012. Microbial ecology of Lake Kivu, p 85–105. *In* Descy J-P, Darchambeau F, Schmid M (ed), *Lake Kivu: limnology and biogeochemistry of a tropical great lake*. Springer Netherlands, Dordrecht. [https://doi.org/10.1007/978-94-007-4243-7\\_6](https://doi.org/10.1007/978-94-007-4243-7_6).
60. Chistoserdova L. 2011. Methylophily in a lake: from metagenomics to single-organism physiology. *Appl Environ Microbiol* 77:4705–4711. <https://doi.org/10.1128/AEM.00314-11>.
61. Chistoserdova L. 2013. The distribution and evolution of C1 transfer enzymes and evolution of the Planctomycetes, p 195–209. *In* Fuerst JA (ed), *Planctomycetes: cell structure, origins and biology* [https://doi.org/10.1007/978-1-62703-502-6\\_8](https://doi.org/10.1007/978-1-62703-502-6_8). Humana Press, Totowa, NJ.
62. Auman AJ, Lidstrom ME. 2002. Analysis of sMMO-containing type I methanotrophs in Lake Washington sediment. *Environ Microbiol* 4:517–524. <https://doi.org/10.1046/j.1462-2920.2002.00323.x>.
63. Auman AJ, Stolyar S, Costello AM, Lidstrom ME. 2000. Molecular characterization of methanotrophic isolates from freshwater lake sediment. *Appl Environ Microbiol* 66:5259–5266. <https://doi.org/10.1128/AEM.66.12.5259-5266.2000>.
64. Chistoserdova L. 2015. Methylophily in natural habitats: current insights through metagenomics. *Appl Microbiol Biotechnol* 99:5763–5779. <https://doi.org/10.1007/s00253-015-6713-z>.
65. Kuivila KM, Murray JW, Devol AH, Lidstrom ME, Reimers CE. 1988. Methane cycling in the sediments of Lake Washington. *Limnol Oceanogr* 33:571–581. <https://doi.org/10.4319/lo.1988.33.4.0571>.
66. Arellano SM, Lee OO, Lafi FF, Yang J, Wang Y, Young CM, Qian PY. 2013. Deep sequencing of Myxilla (Ectomyxilla) methanophila, an epibiotic sponge on cold-seep tubeworms, reveals methylophily, thiotrophic, and putative hydrocarbon-degrading microbial associations. *Microb Ecol* 65:450–461. <https://doi.org/10.1007/s00248-012-0130-y>.
67. Tian RM, Zhang W, Cai L, Wong YH, Ding W, Qian PY. 2017. Genome reduction and microbe-host interactions drive adaptation of a sulfur-oxidizing bacterium associated with a cold seep sponge. *mSystems* 2:e00184. <https://doi.org/10.1128/mSystems.00184-16>.
68. Rubin-Blum M, Antony CP, Sayavedra L, Martínez-Pérez C, Birgel D, Peckmann J, Wu Y-C, Cardenas P, MacDonald I, Marcon Y, Sahling H, Hentschel U, Dubilier N. 2019. Fueled by methane: deep-sea sponges from asphalt seeps gain their nutrition from methane-oxidizing symbionts. *ISME J* 13:1209–1225. <https://doi.org/10.1038/s41396-019-0346-7>.
69. Hashimoto H, Uragami C, Cogdell RJ. 2016. Carotenoids and photosynthesis. *Subcell Biochem* 79:111–139. [https://doi.org/10.1007/978-3-319-39126-7\\_4](https://doi.org/10.1007/978-3-319-39126-7_4).
70. Fiedor J, Sulikowska A, Orzechowska A, Fiedor L, Burda K. 2012. Antioxidant effects of carotenoids in a model pigment-protein complex. *Acta Biochim Pol* 59:61–64. [https://doi.org/10.18388/abp.2012\\_2172](https://doi.org/10.18388/abp.2012_2172).
71. Krisko A, Radman M. 2013. Biology of extreme radiation resistance: the way of *Deinococcus radiodurans*. *Cold Spring Harb Perspect Biol* 5:a012765. <https://doi.org/10.1101/cshperspect.a012765>.
72. Du X-j, Wang X-y, Dong X, Li P, Wang S. 2018. Characterization of the desiccation tolerance of *Cronobacter sakazakii* strains. *Front Microbiol* 9:2867. <https://doi.org/10.3389/fmicb.2018.02867>.
73. Bowman JP, Sly LI, Nichols PD, Hayward AC. 1993. Revised taxonomy of the methanotrophs: description of *Methylobacter* gen. nov., emendation of *Methylococcus*, validation of *Methylosinus* and *Methylocystis* species, and a proposal that the family Methylococcaceae includes only the group I methanotrophs. *Int J Syst Evol Microbiol* 43:735–753.
74. Irvine IC, Brigham CA, Suding KN, Martiny JB. 2012. The abundance of pink-pigmented facultative methylophily in the root zone of plant species in invaded coastal sage scrub habitat. *PLoS One* 7:e31026. <https://doi.org/10.1371/journal.pone.0031026>.
75. Mandler K, Chen H, Parks DH, Lobb B, Hug LA, Doxey AC. 2019. AnnoTree: visualization and exploration of a functionally annotated microbial tree of life. *Nucleic Acids Res* 47:4442–4448. <https://doi.org/10.1093/nar/gkz246>.
76. Meng J, Wang F, Wang F, Zheng Y, Peng X, Zhou H, Xiao X. 2009. An uncultivated crenarchaeota contains functional bacteriochlorophyll a synthase. *ISME J* 3:106–116. <https://doi.org/10.1038/ismej.2008.85>.
77. Liu R, Cai R, Zhang J, Sun C. 2020. Heimdallarchaeota harness light energy through photosynthesis. *bioRxiv* <https://doi.org/10.1101/2020.02.20.957134>.
78. Gao X, Xin Y, Bell PD, Wen J, Blankenship RE. 2010. Structural analysis of alternative complex III in the photosynthetic electron transfer chain of *Chloroflexus aurantiacus*. *Biochemistry* 49:6670–6679. <https://doi.org/10.1021/bi100858k>.
79. Chistoserdova L, Lidstrom ME. 2013. Aerobic methylophily prokaryotes, p 267–285. *In* Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F (ed), *The prokaryotes: prokaryotic physiology and biochemistry*. Springer Berlin Heidelberg, Berlin, Heidelberg. [https://doi.org/10.1007/978-3-642-30141-4\\_68](https://doi.org/10.1007/978-3-642-30141-4_68).
80. Quayle JR, Pfennig N. 1975. Utilization of methanol by Rhodospirillaceae. *Arch Microbiol* 102:193–198. <https://doi.org/10.1007/BF00428368>.
81. Anantharaman K, Brown CT, Hug LA, Sharon I, Castelle CJ, Probst AJ, Thomas BC, Singh A, Wilkins MJ, Karaoz U, Brodie EL, Williams KH, Hubbard SS, Banfield JF. 2016. Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat Commun* 7:13219. <https://doi.org/10.1038/ncomms13219>.
82. Hug LA, Co R. 2018. It takes a village: microbial communities thrive through interactions and metabolic handoffs. *mSystems* 3:e00152-17. <https://doi.org/10.1128/mSystems.00152-17>.
83. Colman DR, Lindsay MR, Amenabar MJ, Fernandes-Martins MC, Roden ER, Boyd ES. 2020. Phylogenomic analysis of novel Diaporarchaea is consistent with sulfite but not sulfate reduction in volcanic environments on early Earth. *ISME J* 14:1316–1331. <https://doi.org/10.1038/s41396-020-0611-9>.
84. Garcia SL, Buck M, McMahon KD, Grossart H-P, Eiler A, Warnecke F. 2015. Auxotrophy and intrapopulation complementarity in the ‘interactome’ of a cultivated freshwater model community. *Mol Ecol* 24:4449–4459. <https://doi.org/10.1111/mec.13319>.
85. Croft MT, Lawrence AD, Raux-Deery E, Warren MJ, Smith AG. 2005. Algae acquire vitamin B12 through a symbiotic relationship with bacteria. *Nature* 438:90–93. <https://doi.org/10.1038/nature04056>.
86. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25:1043–1055. <https://doi.org/10.1101/gr.186072.114>.
87. Parks DH, Chuvochina M, Chaumeil PA, Rinke C, Mussig AJ, Hugenholtz P. 2020. A complete domain-to-species taxonomy for Bacteria and Archaea. *Nat Biotechnol* 38:1079–1086. <https://doi.org/10.1038/s41587-020-0501-8>.
88. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. 2019. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 36:1925–1927. <https://doi.org/10.1093/bioinformatics/btz848>.
89. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>.
90. Lagesen K, Hallin P, Rødland EA, Stærfeldt H-H, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* 35:3100–3108. <https://doi.org/10.1093/nar/gkm160>.
91. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. <https://doi.org/10.1186/1471-2105-11-119>.
92. Bayliss SC, Thorpe HA, Coyle NM, Sheppard SK, Feil EJ. 2019. PIRATE: a fast and scalable pangenomics toolbox for clustering diverged orthologues in bacteria. *Gigascience* 8:giz119. <https://doi.org/10.1093/gigascience/giz119>.
93. Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol* 428:726–731. <https://doi.org/10.1016/j.jmb.2015.11.006>.
94. Kanehisa M, Sato Y. 2020. KEGG mapper for inferring cellular functions from protein sequences. *Protein Sci* 29:28–35. <https://doi.org/10.1002/pro.3711>.
95. Sievers F, Higgins DG. 2018. Clustal omega for making accurate alignments of many protein sequences. *Protein Sci* 27:135–145. <https://doi.org/10.1002/pro.3290>.



96. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490. <https://doi.org/10.1371/journal.pone.0009490>.
97. Orf GS, Gisriel C, Redding KE. 2018. Evolution of photosynthetic reaction centers: insights from the structure of the heliobacterial reaction center. *Photosynth Res* 138:11–37. <https://doi.org/10.1007/s11120-018-0503-2>.
98. Tsuji J, Shaw N, Nagashima S, Venkiteswaran J, Schiff S, Hanada S, Tank M, Neufeld J. 2020. Anoxygenic phototrophic Chloroflexota member uses a type I reaction center. *bioRxiv* <https://doi.org/10.1101/2020.07.07.190934>.
99. Sadekar S, Raymond J, Blankenship RE. 2006. Conservation of distantly related membrane proteins: photosynthetic reaction centers share a common structural core. *Mol Biol Evol* 23:2001–2007. <https://doi.org/10.1093/molbev/msl079>.
100. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T. 2018. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* 46:W296–W303. <https://doi.org/10.1093/nar/gky427>.
101. Hohmann-Marriott MF, Blankenship RE. 2011. Evolution of photosynthesis. *Annu Rev Plant Biol* 62:515–548. <https://doi.org/10.1146/annurev-arplant-042110-103811>.
102. Allen JP, Williams JC. 2008. Reaction centers from purple bacteria, p 275–293. *In* Fromme P (ed), *Photosynthetic protein complexes*. Wiley-Blackwell, Weinheim, Germany. <https://doi.org/10.1002/9783527623464.ch12>.
103. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 305:567–580. <https://doi.org/10.1006/jmbi.2000.4315>.
104. Søndergaard D, Pedersen CN, Greening C. 2016. HydDB: a web tool for hydrogenase classification and analysis. *Sci Rep* 6:34212. <https://doi.org/10.1038/srep34212>.
105. Smith SM, Rawat S, Telser J, Hoffman BM, Stemmler TL, Rosenzweig AC. 2011. Crystal structure and characterization of particulate methane monooxygenase from *Methylocystis* species strain M. *Biochemistry* 50:10231–10240. <https://doi.org/10.1021/bi200801z>.
106. Robert X, Gouet P. 2014. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* 42:W320–W324. <https://doi.org/10.1093/nar/gku316>.
107. Mukherjee S, Stamatis D, Bertsch J, Ovchinnikova G, Katta HY, Mojica A, Chen IA, Kyrpides NC, Reddy T. 2019. Genomes OnLine database (GOLD) v.7: updates and new features. *Nucleic Acids Res* 47:D649–D659. <https://doi.org/10.1093/nar/gky977>.