Culture- and genome-based characterization of a tripartite interaction between patescibacterial epibionts, methylotrophic proteobacteria, and a jumbo phage in freshwater ecosystems

Feriel Bouderka[1], Purificación López-García[1], Philippe Deschamps[1], Mart Krupovic[2], Ana Gutiérrez-Preciado[1], Maria Ciobanu[1], Paola Bertolino[1], Gwendoline David[1,3], David Moreira[1*], Ludwig Jardillier[1*]

[1]Unité d'Ecologie Systématique et Evolution, CNRS, Université Paris-Saclay, AgroParisTech, Gif-sur-Yvette, France, [2]Institut Pasteur, Université Paris Cité, Archaeal Virology Unit, Paris, France, [3]Institute of Freshwater Ecology and Inland Fisheries, Leibniz, Germany

* co-corresponding authors
ludwig.jardillier@universite –paris-saclay.fr
david.moreira@universite –paris-saclay.fr

ABSTRACT

Patescibacteria form a very diverse and widely distributed phylum of small bacteria that are inferred to have an episymbiotic lifestyle. However, our knowledge about the prevalence of this lifestyle within the phylum and the specificity of the hosts they interact with remains very limited due to the scarcity of cultured patescibacterial representatives. Thanks to enrichment cultures, we describe here a new tripartite interaction in freshwater ecosystems between a patescibacterium, a gammaproteobacterial host, and a phage. The patescibacterium, *Strigamonas methylophilicida*, belongs to a new genus within the family Absconditicoccaceae. It grows associated with species of the proteobacterial family Methylophilaceae, so far unknown to host patescibacterial epibionts. *Strigamonas* cells are closely attached to the host, sometimes forming stacks that connect two host cells. Despite having a surprisingly larger genome (1,9 Mb) than those of many other Patescibacteria, *S. methylophilicida* does not encode many essential biosynthetic pathways. This is the case of the complete biosynthesis of phospholipids, amino acids, and nucleic acids, suggesting a dependence on the host to obtain these molecules. The third player in this trio, a jumbo phage, was identified through a CRISPR spacer match in the patescibacterial genome. It has a very large genome (230 kb) that contains a wide repertoire of genes, including many that probably modify the metabolism and the cell surface of *S. methylophilicida* and impact the interaction with its methylotrophic host. Our results confirm a prevalent episymbiotic lifestyle in the Absconditicoccaceae and the selectivity to interact with gammaproteobacterial hosts, supporting a clade-specific adaptation of Patescibacteria to particular hosts.

INTRODUCTION

The recently described Candidate Phyla Radiation (CPR), now formally classified as the phylum Patescibacteria (1), represents an extremely phylogenetically diverse group of bacteria. It is estimated to encompass between 15 and 50% of all bacterial diversity and to be present in all kinds of environments, especially suboxic ones (1–5). Patescibacteria have mostly been characterized using genome-resolved metagenomic approaches, consistently finding reduced genomes coding for limited metabolic capabilities (2,6,7). Notably, they lack the typical pathways for the synthesis of amino acids, nucleotides, and membrane phospholipids. Most likely, they obtain these essential molecules from their hosts, which suggests that a symbiotic, likely parasitic, lifestyle is widespread within the phylum (2,7,8). In agreement with these inferences, microscopy observations have shown that patescibacterial cells are very small, usually attached to larger cells, which supports an episymbiotic lifestyle (2,6,7). Nevertheless, patescibacterial genomes contain many genes of unknown function (up to 50% of all predicted protein-coding genes), and it cannot be discarded that part of them provide alternative ways to fulfill some of the missing biosynthetic reactions. Cultivation of patescibacterial representatives is therefore crucial to determine the lifestyle of these bacteria confidently.

Only a few cultured patescibacterial species have been characterized, and they all seem to confirm the parasitic lifestyle inferred by genome-based approaches. The first patescibacterium ever cultured, *Candidatus* Nanosynbacter lyticus, belongs to the class Saccharimonadia and is an epiparasite of *Schaalia odontolytica* (formerly *Actinomyces odontolyticus*) in the human oral microbiome (9). Other Saccharimonadia have been cultured afterwards and consistently found to be parasites of actinobacterial hosts (10–12). Recently, two other patescibacterial representatives have been characterized, *Candidatus* Vampirococcus lugosii and *Candidatus* Absconditicoccus praedator, representing two distant genera in the Absconditicoccaceae family (13,14). They were both found in saline lakes, as episymbionts of anoxygenic photosynthetic bacteria belonging to the Chromatiales (Pseudomonadota, formerly Gammaproteobacteria). These species appear to exhibit a predatory lifestyle, leading to the rapid death of their hosts (13,14). Therefore, all patescibacterial representatives characterized so far appear to be obligatory episymbionts that require a host to survive. Although they are able to establish interactions with a broad range of bacteria, including Gram-negative and Gram-positive ones, the two cultured patescibacterial groups appear to be specialized in interacting with one host type: Nanosynbacteraceae (Saccharimonadia class) with Actinobacteria, and Absconditicoccaceae with photosynthetic Pseudomonadota. All reported hosts seem negatively impacted by their patescibacterial episymbionts, which exhibit parasitic and predatory interactions. Some microscopy studies have suggested that other Patescibacteria can establish interactions with eukaryotic and archaeal hosts (15,16) but further confirmation is needed. The limited availability of cultured representatives has also made difficult the identification of patescibacterial phages. A few studies have established links with phage genomes using CRISPR spacers found in patescibacterial genomes. For example, Paez-Espino et al. identified 17 putative patescibacterial phages (17), and Liu et al. recently reported the identification of 391 additional putative patescibacterial phages (18). Further efforts are required to comprehend the diversity and impact of phages on Patescibacteria.

In this study, we have characterized the new patescibacterial species *Strigamonas methylophilicida*, which belongs to the family Absconditicoccaceae. It was retrieved from small freshwater ecosystems in southern Paris, where Patescibacteria were previously detected to be highly abundant and diverse (19), and constitutes the first well-characterized freshwater patescibacterial representative. It grows attached to the surface of host cells of the family Methylophilaceae, and therefore represents the first reported association between a patescibacterium and a non-photosynthetic Pseudomonadota (Gammaproteobacteria) host. *Strigamonas* exhibits an S-layer and produces unique extracellular vesicles. Although it lacks many classical pathways for the biosynthesis of essential bacterial metabolites, its 1.9 Mb genome is remarkably larger than those of other Absconditicoccaceae representatives. By analyzing CRISPR spacers, we identified a phage likely targeting *S. methylophilicida*. We recovered the phage genome, which is surprisingly large (230 kb) and harbors many genes likely involved in modifying the metabolism and cell surface of its host. These phage-mediated modifications could also impact the interaction between *Strigamonas* and its Methylophilaceae host. To date, this is the first identified tripartite Patescibacteria-host-phage association.

MATERIAL AND METHODS

Sampling and enrichment of patescibacteria-host consortia

Water samples were collected from two small freshwater ecosystems in the regional Parc de la Haute Vallée de Chevreuse (south of Paris), including a small stream, Ru Sainte Anne (RSA), and a pond, Mare Gabard (MG) (19). We incubated water samples in 2 L glass bottles supplemented with 100 mM methanol at room temperature (RT). We regularly assessed the composition of the bacterial community using phase contrast light microscopy. After approximately one month of incubation, we started observing small-size cells attached to motile bacterial cells. We further enriched these consortia by inoculating 1 ml of the first enrichment in 5 ml of modified nitrate mineral salts (NMS) medium, known to promote the growth of methanotrophic and methylotrophic species (20). After one week of incubation at RT, we further enriched these consortia using fluorescence-activated cell sorting with a FACS Aria III flow cytometer (BD Biosciences, San Jose, CA, USA) equipped with one gas-state 633 nm laser and three solid-state lasers at 488 nm, 405 nm and 375 nm. Side and Forward Scatter Areas (SSC-A and FSC-A) were used to discriminate the patescibacteria-host consortia from other microbial cells. The 488 nm laser was used for the analysis of both FSC (488/10, 1.0 ND filter) and SSC (488/10) parameters and the applied voltages were 100V and 300V respectively. The trigger was conjointly set to the FSC and SSC with a threshold of 500. The fluidic system was run at 20 psi (3.102bar) with a 100 µm nozzle. Samples were sorted at a speed of 3000–3500 events s^{-1} and the sheath fluid consisted of sterile 1X PBS buffer. In addition to the excitation lasers, an additional BD Accudrop laser of 660nm was used for the drop delay setup. Cell suspensions were pre-filtered on a 30µm mesh just before sorting to avoid clogs. We collected eight consortia in 1.5 ml individual tubes (200 sorted events per tube) using the most accurate sorting mode “single-cell purity”. The purity of the sorted events was visually checked with a phase contrast light microscope (see below).

Microscopy

We performed light microscopy observations of living cells with a Zeiss Axioplan 2 microscope equipped with a 100X oil-immersion phase contrast objective and a Nikon Coolpix B500 color camera for recording. For transmission electron microscopy, cells were fixed with 2.5% glutaraldehyde in 0.2M cacodylate buffer for 2 hours on ice. Then, we washed the cells with 0.2M cacodylate and post-fixed them in 1% osmium for 1 h at RT. We subsequently dehydrated the cells with successive baths of increasing ethanol concentrations (30%, 50%, 70%, 80%, 90%, 95%, 2 x 100%, for 10 min each), in a 1:1 ethanol 100% and acetone mix for 10, and twice in pure acetone for 10 min. The dehydrated cells were then embedded in a 1:1 acetone and epoxy resin mix overnight. Then, they were incubated in pure epoxy resin twice for 3 hours. Finally, the dehydrated cells were incubated overnight in pure epoxy resin at 70°C. Polymerized resin blocks were cut using a Leica EM UC6 ultramicrotome and sections were observed with a JEOL1400 microscope with an acceleration voltage of 80 KeV at the Imagerie-Gif facility (https://www.i2bc.paris-saclay.fr/bioimaging/).

Whole genome and 16S rRNA gene amplifications

The DNA was extracted from the FACS sorted cells and whole genomes were amplified using multiple displacement amplification technique (MDA) with the REPLI-g Single Cell kit (Qiagen) by following the manufacturer instructions. The amplified DNAs were diluted to 1/50 and the 16S rRNA genes were amplified by PCR with the Platinum Taq DNA polymerase (Invitrogen) using the manufacturer's instructions and the following program : (i) initial denaturation for 2 min at 94 °C, (ii) 35 cycles of denaturation (94°C for 15 s), annealing (55 °C for 30 s), and extension (72 °C for 1 min and 30 s), (iii) final extension for 5 min at 72 °C. The primers used for this PCR amplification were the U515F (5'-GTGCCAGCMGCCGCGGTAA-3') and 926R (5′-CCGYCAATTYMTTTRAGTTT-3′). 16S rRNA clone libraries were produced with the Topo TA cloning kit following the manufacturer's instructions (Invitrogen). Individual clone inserts were amplified with the vector primers T7 (5'-TAATACGACTCACTATAGGG-3') and M13 reverse (5'-CAGGAAACAGCTATGAC-3') and Sanger-sequenced by Genewiz (Essex, UK).

Genome sequencing, assembly, and annotation

We selected the five WGA (Whole Genome Amplified) products that have produced 16S rRNA sequences of Patescibacteria (see above) and sequenced them (Eurofins Genomics, Ebersberg, Germany) using Illumina HiSeq (2×125 bp paired-end reads). The quality check and trimming of the reads were done using FastQC and Trimmomatic (MAXINFO:40:0.8, MINLEN:30 LEADING:20, TRAILING:20) (21,22), respectively. Two co-assemblies were performed, merging, on the one hand, the three WGA sequences of the sortings from the RSA ecosystem, and on the other hand, the two WGA sequences from the MG ecosystem. The assembly was done with SPAdes v3.15.3 (23) in MetaSPAdes mode with a combination of k-mer lengths of 21,25,31,35,41,45,51,55. We used the Anvi'o v6 metagenomic workflow (24) to reconstruct the MAGs, as follows. Open reading frames (ORFs) were predicted *de novo* using Prodigal V2.6.3 (25). Functional and taxonomic annotation of the ORFs were done with BLAST (26) and Kaiju v1.7.1 (27) against the NCBI COG (2014 release) and RefSeq (28–30) databases, respectively. To minimize the use of the differential coverage, which can be biased by the WGA, we avoided using automatic binning. Contigs with a minimum length of 1000 bp were clustered hierarchically for a manual binning via the Anvi'o v6 interactive interface (24).

We used the same approach to recover the host MAGs but we instead co-assembled all the five samples, regardless of the ecosystem to increase the MAG completeness.

We calculated the completeness and redundancy of all MAGs with Anvi'o v6 (24) and determined their phylogenetic novelty by calculating their average amino acid identities with the online AAI calculator (enve-omics.ce.gatech.edu/aai/) (31). Additionally, we used the GTDB-tk v2.3.0 pipeline for the taxonomic classification of the MAGs (32,33).

We predicted genes in all MAGs using Prokka 1.14.5 (34) and functionally annotated them using the GhostKOALA web server (35). Protein structures were predicted with Alphafold (36). The structural comparison to the Protein Data Bank (PDB) was done on the DALI protein structure comparison server (37). We performed a gene orthology comparison among all the well-characterized Absconditabacterales MAGs (*Ca*. A. predator, *Ca*. V. lugosii, and our MAGs) with the Anvi'o v7 pipeline (24,28,38). For that, genes were compared all *vs*. all with BLAST, clustered with MCL (39) (inflation parameter 2), annotated against the NCBI COG database (2014 release), and visualized using Anvi'o v7.

We searched for CRISPR-Cas systems in the *Strigamonas* MAGs using CRISPR-Cas finder (40). The identified CRISPR spacers were used as query sequences to explore via BLAST the metagenomes of the freshwater samples collected in previous years from the water column of the same small freshwater ecosystems, and filtered as described by (19). The DNA was extracted from the filtered fraction-size between 0.2–5 µm, with Power Soil kits (Qiagen) as described previously (19,41), and sequenced as described above by (Eurofins Genomics). We identified in this way a phage genome from the MG metagenome. Its completeness was estimated using checkV (42). We classified it using the VPF-class tool (43) and the result was subsequently assigned to the corresponding ICTV classification. The gene annotation of the viral genome was performed using Prokka (34), and functional annotation with a combination of tools: BLAST searches against the NCBI non-redundant database, the GhostKOALA online tool (35), eggNOG-mapper 2.1.9 (44), the Prost online tool (45), and HHsearch (46). Visualizations were carried out using Kegg mapper (47), and Clinker (48).

Phylogenetic analyses

16S rRNA sequences were aligned against the Silva 138.1 database with SINA (49). The multiple sequence alignment of the ~100 closest sequences was trimmed with TrimAl v1.4.rev22 -automated1 (50), and the corresponding maximum likelihood (ML) phylogenetic tree was reconstructed with IQtree 1.6.11 (51) with the GTR model.

Protein sequence alignments were produced using MAFFT v7.450 -auto (52), trimmed using trimAl v1.4.rev22 -automated1 (50). Single-protein ML phylogenetic trees were constructed using IQ-tree 1.6.11 (51) with the LG+C60+F+G model. For phylogenomic analyses, multiple sequence alignments were concatenated, and multi-protein ML phylogenetic trees were then reconstructed using IQ-tree 1.6.11 with the LG+C20+F+G model for Patescibacteria and the LG+C60+F+G model for the Methylophilaceae. The gene markers used for the phylogenomic placement of the Patescibacteria was extracted using GTDB-tk, option identify (1,32). For all phylogenetic trees, branch supports were estimated using the IQtree ultrafast bootstrap method (1000 replicates) implemented in IQ-tree. Trees were visualized using iTOL (53).

RESULTS AND DISCUSSION

Identification of a new Absconditicoccaceae-Methylophilaceae consortium

In a previous multiannual study of geographically-close small freshwater ecosystems near Paris, Patescibacteria were shown to constitute a large part of the prokaryotic diversity (19). In two of these ecosystems, the small stream Ru Sainte Anne (RSA) and the forest pond Mare Gabard (MG), the methylotrophic bacterial families Methylophilaceae and Methylococcaceae (Beta- and Gammaproteobacteria, respectively, both now classified in the phylum Pseudomonadota (54)) were also highly abundant (Fig. S1). Therefore, they represented potential hosts for some of the co-occurring Patescibacteria. To enrich them, we collected water samples from these ecosystems that we supplemented with 100 mM methanol. After one month of incubation, we observed motile bacterial cells carrying much smaller cells attached to them. We further enriched these consortia by inoculating tubes of modified nitrate mineral salts (NMS) medium with this initial enrichment to promote the growth of methylotrophic bacteria (55). After one week, we observed intense growth of motile cells with curved-rod cell shape that contained numerous refringent inclusions (Fig. 1A-D), similar to previously described Methylophilaceae species (56). These cells presented small epibiotic cells strongly attached to them, often forming short piles of two to four cells (Fig. 1A-D). They sometimes formed a bridge of cells between two host cells (Fig. 1C-D). These characteristics were similar to those previously described for Absconditicoccaceae species within the Patescibacteria (13,14). We collected and fixed cells to prepare thin sections that we examined using transmission electron microscopy (TEM). The host cells were ~2.5 µm long and ~1.25 µm in diameter, whereas the epibiont cells were ~300 nm in diameter and slightly flattened (n=20). The host and epibiont cells appeared tightly connected, with a gap of less than ~20 nm separating them. This cell size and close attachment were coherent with those described for Patescibacteria such as *Ca.* Absconditicoccus praedator (14). The internal cell organization of the epibionts also resembled previous observations of Patescibacteria (Fig. 1E-I). The ribosomes were positioned at the periphery of the cells, and a dense area in the center probably corresponded to highly compacted DNA (6,14). The cell envelope structure also resembled that described for other Patescibacteria (6,14). By contrast, we observed numerous extracellular vesicles of ~20 nm in diameter around the epibiotic cells which, to our knowledge, have not been previously reported in Patescibacteria.

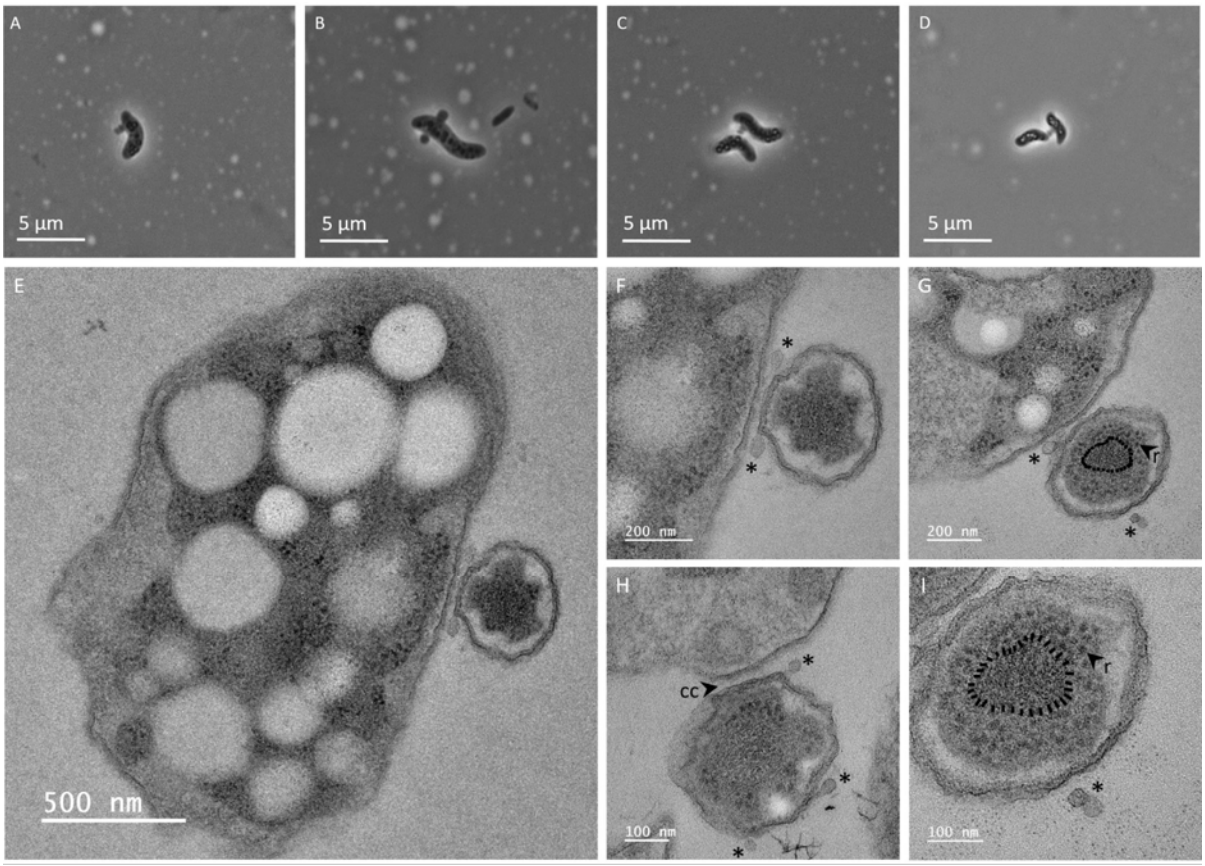

Fig. 1. Microscopy observations of patescibacterial-like cells and their hosts. **A-D** Phase contrast light microscopy pictures, notice the intracellular refringent inclusions in the host cells in C and D; **E-I** Transmission electron microscopy (TEM) pictures of thin sections of the epibiotic cells and their hosts. Asterisks indicate extracellular vesicles, CC: cell-to-cell contact, r: ribosomes. The dashed line in panel I highlights the central part of the cell with dense material that probably corresponds to the DNA.

Our microscopy observations revealed the presence of other bacteria in our enrichments. We used fluorescence-activated cell sorting (FACS) to further enrich the consortia from both MG and RSA samples. FACS allowed us to identify populations highly enriched in these consortia (Fig. S2), as determined by microscopy observation. We sorted and carried out 16S rRNA gene amplification and sequencing for eight of these cell populations (four from MG and four from RSA) to identify the bacteria in these consortia (Fig. S3). Three of the samples did not yield any patescibacterial sequences and were excluded from subsequent analyses (Table S1). Sequences from both Absconditicoccaceae and Methylophilaceae were consistently present in the eight remaining samples. One RSA sample contained sequences exclusively of these two groups, whereas small proportions of other bacteria were also found in the other samples (less than 30% of all the 16S rRNA gene sequences from these samples). Combined with microscopy observations, these results strongly indicated a new Patescibacteria-host interaction established between Absconditicoccaceae and Methylophilaceae species.

To get genomic data to better understand this novel interaction, we carried out whole genome amplification (WGA) from five FACS-sorted samples (three from RSA and two from MG) and sequenced the amplified DNA. These WGA samples contained DNA from several species (Absconditicoccaceae, Methylophilaceae, and minor contaminants), so they behaved as mini-metagenomes. We obtained two high-quality Absconditicoccaceae metagenome-

assembled genomes (MAGs) by co-assembling and manually binning the sequences from each ecosystem, RSA and MG (see Methods). These two MAGs showed an average nucleotide identity (ANI) of 94.82% and an average amino acid identity (AAI) of 93.51%. The MAG obtained from RSA was 1.94 Mb in size with 81.1% completeness and 2.1% redundancy, while the MAG from MG was 1.93 Mb with 83.1% completeness and 1.41% redundancy (Table S2). These values are calculated via identification of single copy genes that are universal for most bacteria (57). Because Patescibacteria commonly exhibit genome reduction and multiple gene losses, a completeness of 70% is considered a complete and high quality genome (58,59). The GC content of the RSA and MG MAGs was 28.47% and 27.01%, respectively. Reconstructing MAGs for the host was more challenging because our samples appeared to contain a mix of several Methylophilaceae species, which made sequence binning more difficult. To increase genome coverage, we co-assembled all samples from both ecosystems and, after two rounds of manual refinement, we obtained four good-quality Methylophilaceae MAGs (completeness between 90.14% and 95.77% and redundancy between 2.82% and 5.63%) (Table S2).

To confirm the phylogenetic placement of the four host MAGs, we built a maximum likelihood (ML) tree using 14 concatenated ribosomal proteins. All four of them belonged to different genera within the family Methylophilaceae (Fig. S4). For the RSA and MG epibiont MAGs, we reconstructed an ML tree using a subset of 67 markers well represented in Patescibacteria out of the 120 bacterial markers described in (1). The epibiont MAGs branched together within the family Absconditicoccaceae, forming a sister group to *Ca.* V. lugosii and *Ca.* A. praedator, the two previously characterized representatives of this clade (Fig. 2). The 16S rRNA gene sequence identity of our new epibionts with *Ca.* V. lugosii and *Ca.* A. praedator was 81% and 84%, respectively, and their average amino-acid identities (AAI) were 45 and 46%, respectively. Because of low Relative Evolutionary Divergence (RED) values (0.71519 for RSA MAG and 0.71601 for the MG MAG), the GTDB-tk pipeline (33) failed to assign our MAGs confidently to any known genus. Altogether, these metrics supported that our freshwater patescibacterial MAGs defined a novel genus and species (1,60–62), which we propose to name *Strigamonas methylophilicida* (see Taxonomic appendix below).

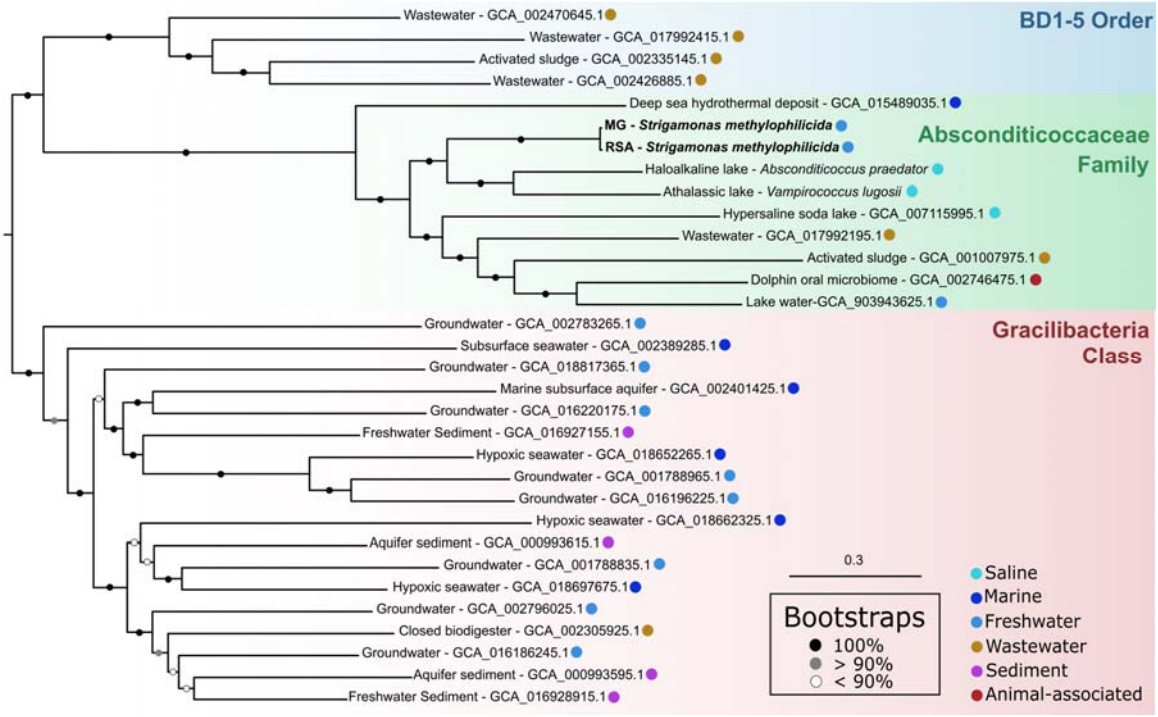

Fig. 2. Phylogenetic placement of *Strigamonas methylophilicida*. Maximum likelihood tree based on the concatenation of 67 single-copy protein alignments (20,766 sites). The scale bar indicates the number of substitutions per site. Support values are based on ultrafast bootstraps. Taxonomy corresponds to the Genome Taxonomy Database release 08-RS214.

Gene content in *Strigamonas methylophilicida*

MAGs of *S. methylophilicida* from RSA and MG presented a size of approximately 1.9 Mb, almost twice the average genome size of most Patescibacteria (~1 Mb, (8)). As described for other Absconditibacteraceae, *S. methylophilicida* uses a modified genetic code (code 25) where the stop codon UGA is translated to glycine (57). We identified 1,655 and 1,629 coding sequences (CDS) in the RSA and MG MAGs, respectively, which were many more than in other well-characterized Absconditicoccaceae representatives (1,054 in *Ca.* A. praedator and 1,124 in *Ca.* V. lugosii). However, we could only functionally annotate around one third of these CDS (565 for RSA and 543 for MG), compared to the ~50% functionally annotated in *Ca.* V. lugosii *and Ca.* A. praedator (13,14). Although a fraction of them may correspond to known genes that have diverged so much that they can no longer be recognized by sequence similarity, *S. methylophilicida* seems to bear a substantial amount of gene novelty. Whereas the pangenome of the Absconditabacteraceae has been reported to be highly dynamic, the core genome is very small, of only 390 genes (13). This seems to be even more pronounced when *S. methylophilicida* is included, revealing a mere 236 genes shared among the three genera, *Ca.* Absconditicoccus, *Ca.* Vampirococcus, and *Strigamonas* (Fig. 3). As expected, the majority of these core genes are involved in house-keeping functions such as translation, transcription, and replication (Fig. 3). By contrast, more than 50% of the genus-specific genes have unknown functions, and the rest are mostly involved in cell wall, membrane, and envelope biogenesis. Among the three Absconditicoccaceae genera, *Strigamonas* stood out for its significant divergence in gene content. By contrast, the two *Strigamonas* MAGs shared 1,014 genes that were not found in other Absconditicoccaceae, as well as 320 (RSA) and 268

(MG) strain-specific genes also absent in other Absconditicoccaceae (Fig. 3). Many of these genes were organized in clusters in the two MAGs (see Figs. S4b and S4c). Most of these genes had unknown functions (Fig. 3b, Fig. S5) but, given their very narrow distribution, it is tempting to speculate that some may be involved in particular ecological niche adaptations. In both MAGs, the most represented categories included the defense mechanisms, in particular restriction-modification systems, the protection against oxidative stress MutT (see text below), and ABC-type multidrug transport systems (Fig. 3b, Fig. S6, Table S3).

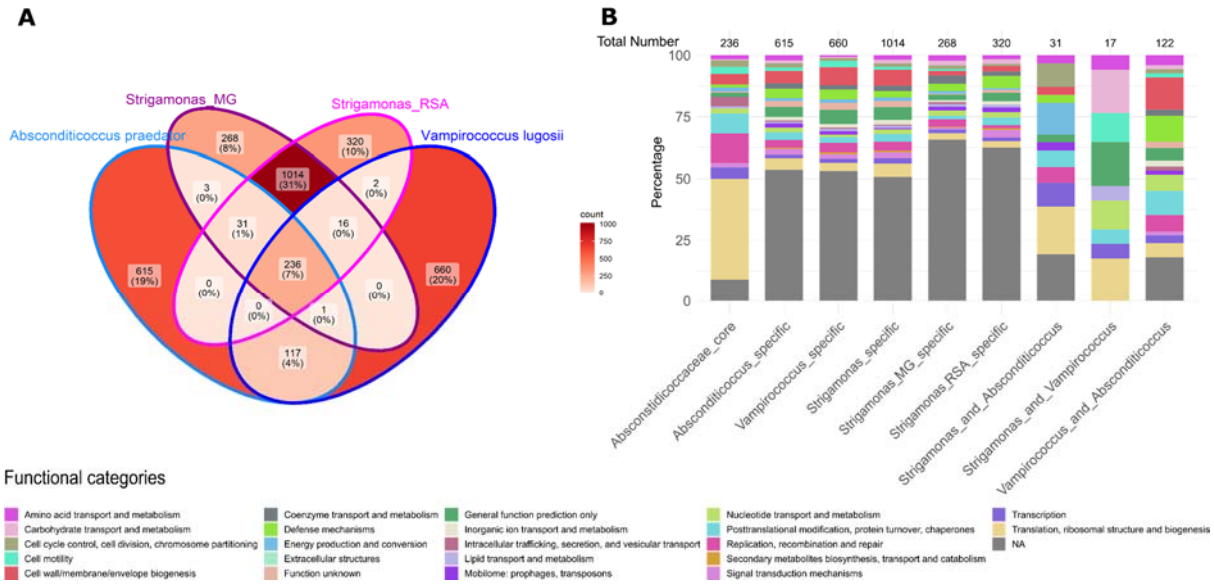

Fig. 3: Comparison of gene content in characterized Absconditicoccaceae species. A) Venn diagram showing the number of genes shared by different Absconditicoccaceae species. **B)** Functional annotation (COG categories) of the genes shared by different Absconditicoccaceae species.

Genes coding for most ribosomal proteins were present (n=51), with only a few missing, such as L30, which is often absent in other symbiotic bacteria (2,8). As previously described for many other Patescibacteria, *S. methylophilicida* lacked classical phospholipid biosynthesis genes, indicating its inability to synthesize its own membrane lipids. Similarly, biosynthetic pathways for nucleic acids, cofactors, and many canonical amino acids were missing. These absences suggest a strong dependence on the host for the maintenance of the cell integrity and growth (7).

Genome-based inference of the metabolism and symbiotic lifestyle

The MAGs of *S. methylophilicida* also encoded reduced carbon and energy metabolisms. As observed previously in *Ca.* V. lugosii and many other Patescibacteria (7,13), *S. methylophilicida* possesses an incomplete glycolysis pathway where only the genes involved in the steps from 3-phosphoglycerate to pyruvate could be detected (Fig. 4). Therefore, these species most likely require the uptake of 3-phosphoglycerate for ATP production. *Ca.* V. lugosii was hypothesized to obtain 3-phosphoglycerate from its photosynthetic host that possesses the enzyme ribulose 1,5-bisphosphate carboxylase/oxygenase (RuBisCO), which largely produces this metabolite (13). The Methylophilaceae host of *S. methylophilicida* can probably

provide 3-phosphoglycerate from the degradation of methanol (Fig. S7). Unexpectedly, akin to the photosynthetic host of *Ca.* V. lugosii, two of our Methylophilaceae MAGs encoded the enzymes of the Calvin-Benson-Bassham (CBB) cycle, including phosphoribulokinase (PRK) and RuBisCO (Fig. S8). These two host MAGs shared the same genomic context for the genes encoding these two enzymes, being close to the CO_2-responsive transcriptional regulator CbbR (Fig. S9). Another Methylophilaceae MAG encoding a complete CBB cycle has been recently reported (63), which supports that the presence of these genes in our host MAGs is not artifactual and that these Methylophilaceae might have the capacity to synthesize 3-phosphoglycerate using various pathways. An interesting point to notice is that, to date, all identified hosts of Absconditicoccaceae epibionts possess RuBisCO (13,14), suggesting that carbon fixation may be advantageous for the establishment of this symbiotic interaction.

The MAGs of *S. methylophilicida* encoded a few additional enzymes involved in carbon metabolism that are absent in the other characterized Absconditicoccaceae, including malate dehydrogenase (EC 1.1.1.38), alcohol dehydrogenases (EC 1.1.1.2), and aldehyde dehydrogenases (EC 1.2.1.3). Thus, *S. methylophilicida* has the potential to produce or use malate, lactate, ethanol, and acetate from/to pyruvate (Fig. S10). These compounds have been suggested to be exported by the Patescibacteria and used by other organisms, possibly including their hosts (8). Additionally, the MAGs of *S. methylophilicida* encoded three enzymes of the pentose phosphate pathway: glucose-6-phosphate dehydrogenase (NADP+) (EC 1.1.1.49) and glucose-6-phosphate dehydrogenase [NAD(P)+] (EC 1.1.1.363), which both transform beta-D-glucose-6P into D-glucono-1,5-lactone-6P, and 6-phosphogluconolactonase (EC 3.1.1.31), which further converts the latter product into D-gluconate-6P (Fig. 4 and Fig. S11). These reactions can generate NADPH, supplying the cell with reducing power.

These MAGs also contained the genes coding for a complete F-type ATP synthase and, based on the sequence characteristics of its subunit c, it probably transports H^+ to synthesize ATP, as in *Ca.* Vampirococcus (13). However, as described previously in many Patescibacteria (8), no classical electron transport chain was present. Alternative proteins for generating proton motive force (PMF) have been proposed in the other two Absconditicoccaceae representatives (13,14), but none of them could be detected in *S. methylophilicida.* In contrast with the relatively large distance between cells of *Ca.* Vampirococcus and its host (~100 nm), the cells of *S. methylophilicida* were tightly attached to their host, opening the possibility of direct salvage of protons from the host, as hypothesized for other Patescibacteria in previous studies (8).

The MAGs of *S. methylophilicida* encode a relatively complex cell surface, which is a common trait of the family Absconditicoccaceae (13,14). They contained genes for the type IV pilus (*pilBCDT*), probably involved in attachment to their host. They also contained genes coding for the natural competence system (*comEC*), probably related to DNA uptake from the host, as well as several genes involved in the secretion system Sec-SRP (*secDEFGYA* and *yidC*). We found a diverse set of genes (10 in RSA MAG and 8 in MG MAG) coding for proteins with S-layer homology (SLH) domains, a finding consistent with previous observations in *Ca.* A. praedator (14). The S-layer is likely involved in the physical contact between these epibionts with their hosts. The complete peptidoglycan biosynthesis pathway of peptidoglycan biosynthesis was also present, as well as many ion channels and electrochemical potential-driven transporters. These transporters included Ca^{2+}/H^+ antiporters, Mg^{2+} transporters, Ca-activated chloride channels, sulfate permeases, and P-type Cu^+ transporters. In addition, our

MAGs encoded many ABC transporters (28 in RSA and 31 in MG). While many of these transporters were not related to a particular function, others are known to participate in cell division (*ftsX* and *ftsE*), DNA metabolism regulation (*upp*), or transport of metals such as Zn (*znuABC*). They also encoded efflux systems for cobalt-zinc-cadmium and ferrous iron, and a transporter *tetA* of the Major Facilitator Superfamily (MFS), involved in the tetracycline efflux resistance mechanism.

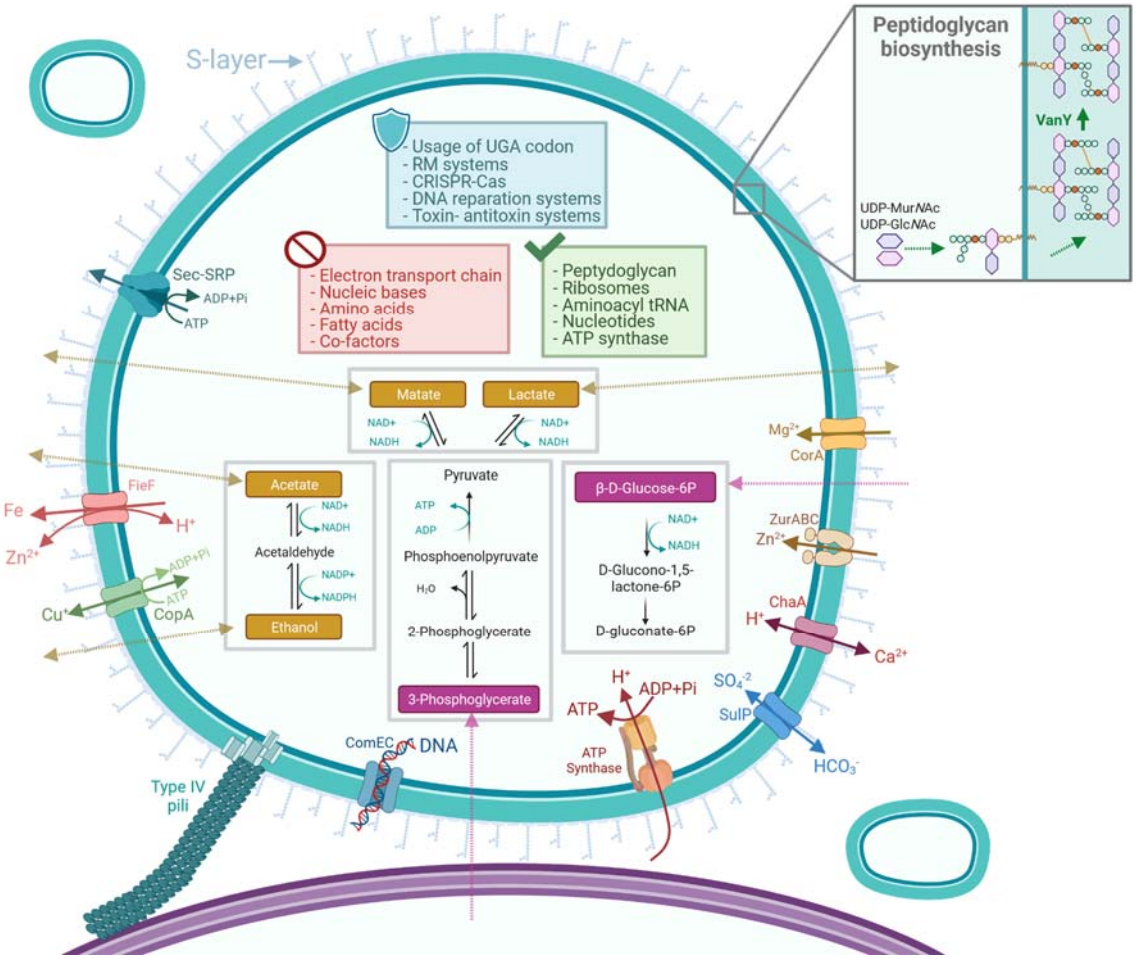

Fig. 4. Reconstruction of metabolic capacities and structural features of *Strigamonas methylophilicida*. The red and the green boxes indicate the incomplete/absent and present pathways, respectively. The blue box indicates the defense mechanisms detected. The grey boxes detail the carbon metabolism, highlighting in pink the compounds used and in yellow the compounds that can either be produced or used by *S. methylophilicida*.

Cell defense mechanisms

Both *S. methylophilicida* MAGs encoded several DNA repair systems, including the base excision, nucleotide excision, and mismatch repair systems (Fig. S12). Interestingly, the gene *mutT*, related to the hydrolysis of 8-oxoguanine formed due to oxidative stress (64), was present in eight copies in the MAGs, which probably reflects the particular importance of this gene in maintaining genome integrity. Other genes related to response to oxidative stress were also detected, including those coding for glutathione peroxidase, glutathione S-transferase, thioredoxin reductase, thioredoxin-dependent peroxiredoxin, and superoxide dismutase Fe-Mn family members. Patescibacteria usually employ the *recFOR* system for

homologous recombination instead of the *recBCD* pathway (8), but we could not detect any of them in *S. methylophilicida* . However, *ruvABCD* and *recG* also involved in homologous recombination were present in our MAGs. The type I restriction-modification defense system (*hsdRSM*) was present in four and two copies in RSA and MG MAGs, respectively. Both MAGs also encoded the restriction enzyme Mrr of the type IV system but not the corresponding modification enzyme, whereas in the RSA MAG, the cytosine-specific DNA methyltransferase was present. The MAGs also encoded the virulence-related HigA-HigB and AbiEii-AbiEi toxin-antitoxin systems, which can be involved in the interaction of *S. methylophilicida* with its host and/or participate in antiviral defense through programmed cell death (65,66). In addition, the toxin YafQ, an endoribonuclease that associates with the ribosome and blocks translation elongation through sequence-specific mRNA cleavage (67), was also detected in the MAGs.

We identified class I and class II CRISPR-Cas systems in *S. methylophilicida* (Fig. 5a) but, unexpectedly, neither of them included a complete adaptation module, which typically consists of *cas1*, *cas2*, and *cas4* genes. We only detected *cas4*, and only in the RSA MAG. The class I CRISPR-Cas system was very partial, only represented by *cas6*, found in both *S. methylophilicida* MAGs, and *cas3*, found only in the RSA MAG (68). No other *cas* genes nor CRISPR array could be found in the surrounding genomic context, suggesting that this system might be inactive and likely becoming reduced. Additionally, we identified a gene in the MG MAG coding a Cas12a protein, a signature of the class II type V CRISPR-Cas systems (69,70). Similar CRISPR-Cas systems composed only of a CRISPR array and the *cas12a* gene have been reported in previous studies (70). Cas12a is a multidomain enzyme that carries out the maturation of the pre-CRISPR RNA and the complete interference stage leading to the cleavage of the target DNA (70). The Cas12a protein encoded in the MG MAG was very divergent from those available in public databases. In particular, we only detected a fraction of the RuvC-like nuclease domain using the predicted 3D structure of the protein (Fig. S13). We identified two sequences similar to our divergent Cas12a encoded in two patescibacterial MAGs available at Genbank (Fig. S14). In these MAGs, the CRISPR-Cas systems appear complete, including the adaptation module (Fig. 5a).

The genomic region encoding the CRISPR-Cas systems in both *S. methylophilicida* MAGs presented conserved synteny despite the abovementioned differences. Within this region, we identified CRISPR arrays containing 4 spacers in the RSA MAG and 6 in the MG MAG (Fig. 5a). The direct repeats of these arrays were similar in both MAGs. By contrast, no sequence similarity was detected between the spacers, suggesting that the two strains have been subjected to different phage infection histories in their respective ecosystems. The CRISPR array of the MG MAG was located in the middle of a 19 kb contig, which reduced the chance of missing adjacent *cas* genes due to MAG incompleteness. This absence of *cas* genes suggests that this system is incapable of adding new spacers, which is coherent with the small number of spacers in these CRISPR arrays. This contrasts with the genomes of *Ca.* V. lugosii and *Ca.* A. praedator, which not only contain several complete CRISPR-Cas systems as previously mentioned, but also large CRISPR arrays with 20-35 spacers. Interestingly, the direct repeat sequences of our MAGs and those of *Ca.* V. lugosii and *Ca.* A. praedator presented some sequence similarity (Fig. 5b), suggesting that these CRISPR-Cas systems derive from a common ancestral Absconditicoccaceae. The limited number of CRISPR spacers and the absence of the Cas1-Cas2 adaptation module in both *S. methylophilicida* MAGs, along with the loss of Cas12a in the RSA MAG, collectively suggest that this ancestral Absconditicoccaceae CRISPR-Cas system is undergoing a process of

degradation in *S. methylophilicida*. To compensate for such an important loss, *S. methylophilicida* seems to have gathered in the nearby genomic context other systems to protect against phage infections, including restriction modification enzymes and an abortive infection (Abi) toxin-antitoxin system. This clustered organization in “defense islands” has been previously reported in bacteria (71).

CRISPR-based identification of a jumbo phage targeting *S. methylophilicida*

We used the spacers of the CRISPR arrays of both *S. methylophilicida* MAGs to search for potential phage genomes in metagenomic assemblies from samples collected from the same MG and RSA freshwater ecosystems in previous years. One 30 bp spacer of the RSA MAG recruited a potential phage contig from the MG ecosystem with 96.6% identity (29 out of 30 positions) and 100% coverage (see Fig. 5b). Supporting this CRISPR-based identification, the phage contig was compatible with the genetic code 25 (UGA translated into glycine) used by Absconditicoccaceae, indicating that it can use the cell machinery of *S. methylophilicida* to replicate. We observed only a few differences in the predicted phage proteins when inferring them using either the general prokaryotic code 11 or the code 25 (Fig. 5c). Our phage contig was very similar to a phage genome from european freshwater lakes and reservoirs (GenBank accession LR796420.1; (72)). However, our phage genome was estimated to be only 57% complete compared to LR796420.1, and it notably lacked capsid protein genes. We used the LR796420.1 sequence to recruit potential missing parts of our phage from the MG metagenomic assembly. In this way, we identified another 100 kb contig containing the gene coding for the capsid and other structural proteins. The two MG phage contigs showed very similar sequence coverage (34x and 35x, respectively), were compatible with genetic codes 11 and 25, and shared synteny with the reference phage genome (Fig. S15). Therefore, we concluded that the two contigs were part of the same phage genome. The large genome size of this phage (230 kb) qualified it as a jumbo phage (73). Only around 160 jumbo phage genomes have been reported so far, with sizes between 200 and 500 kb (74). According to the new ICTV classification (75), these phages are classified within the class *Caudoviricetes* (tailed double-stranded DNA phages). Given the small cell size of *S. methylophilicida*, it was surprising to find that it can be parasitized by a jumbo phage.

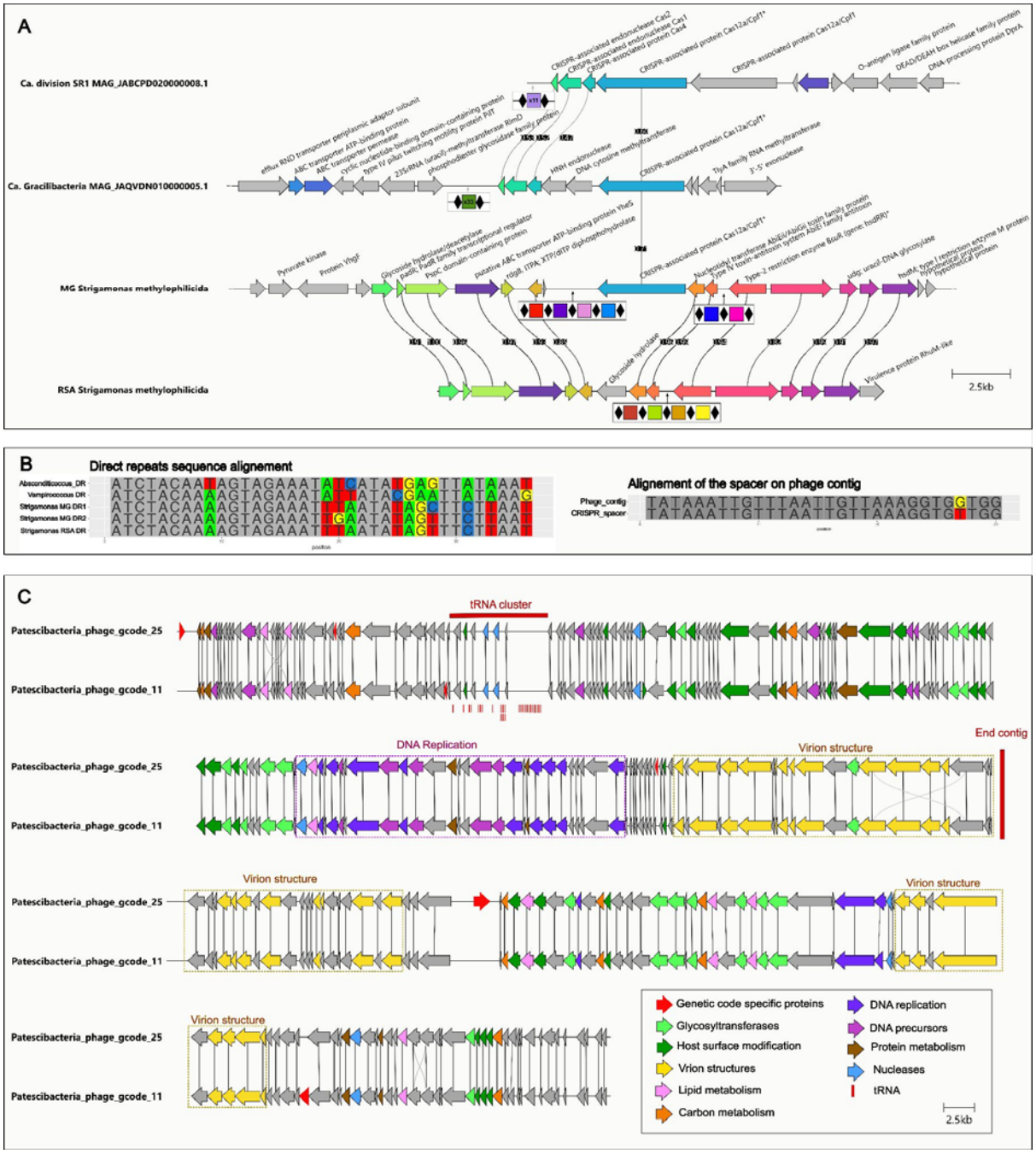

Fig. 5. The CRISPR-Cas systems of *S. methylophilicida*. A. CRISPR-Cas gene organization in both *S. methylophilicida* MAGs and two patescibacterial genomes with systems encoding similar Cas12a proteins. The spacers are represented as colored squares separated by direct repeats (black diamonds). **B.** The sequence alignment of the direct repeats of *S. methylophilicida* is shown and the sequence alignment of the spacer to phage genome. **C.** Genome organization of the *S. methylophilicida* jumbo phage. The ORFs from alternative translations with genetic codes 25 and 11 are presented. ORFs are color-coded according to their functions. ORFs highlighted in red are those differing in the two alternative translations.

The jumbo phage genome contained 292 genes, among which only six were selectively translated by using either genetic code 11 or 25 (Fig. 5b and Table S4)—four genes exclusive to code 25 and two to code 11. The simultaneous compatibility with both genetic codes suggested that this phage probably has a larger host range. Liu *et al.* recently reported

other phages that can also employ both genetic codes, suggesting a potential to infect both Patescibacteria and other bacteria (18). The differential translation of some genes could constitute a host-dependent regulation mechanism. Stop codon-recoding has been hypothesized to participate in the regulation of the expression of early and late infection-stage proteins in phages (76). We were able to functionally annotate three of the six code-specific genes. The first one, expressed under the code 25, encoded a protein that allows phages to propagate between cells in multicellular hosts (GO:0046740). The presence of this gene suggests a potential role in facilitating phage propagation within the stacks of cells of *S. methylophilicida* or between *S. methylophilicida* and its methylotrophic host. The second gene, expressed under the genetic code 11, coded for a SGNH hydrolase that acts as esterase and lipase. Finally, the third gene, expressed under the code 25, encoded a protein with a transmembrane domain, suggesting a cell surface localisation.

We explored the rest of genes present in the phage genome and could only assign a function to 58% of them (155 genes; Figs. 5c and 6). This annotation was more complete than for other jumbo phages, which only have around 20% of genes annotated (74). We identified 25 genes encoding typical phage structural proteins, including various tail proteins, large terminase subunit, portal protein, the HK97-like major capsid protein, prohead maturation, and serine protease, among others. Many of the structural proteins had homologs in T4-like phages, including the tail sheath protein, strongly suggesting that our jumbo phage presents a myovirus-like morphology with an icosahedral head and a contractile tail. The phage genome also encoded nine nucleases, likely involved in phage DNA repair and replication, but potentially also in counterdefense and cleavage of the host genome to generate nucleotides for phage genome replication. Notably, unlike its patescibacterial host, the phage also encoded several enzymes involved in nucleotide metabolism, both purines and pyrimidines, including ribonucleotide reductase alpha and beta subunits, dUTP pyrophosphatase, thymidylate kinase, thymidylate synthase, adenylyl-sulfate kinase, and adenine phosphoribosyltransferase, which can help to provide the DNA precursors needed for the replication of the large viral genome. Similar to other jumbo phages, it contained genes coding for an elaborate, near-complete suite of DNA replication proteins, including DNA clamp and clamp loader, DnaG-like primase, family A and B DNA polymerases, single-stranded DNA-binding protein, flap endonuclease, and DnaB-like replicative helicase. The phage also encoded several DNA repair proteins, including the SbcCD-like nuclease-helicase involved in double-strand break repair, the UV DNA damage repair endonuclease UvsE, and a RecA-like homologous recombination ATPase. It has been suggested that the widespread presence of these genes in jumbo phages represents a way for phages with large genomes to compensate for their smaller burst size compared to their relatives with smaller genomes (74). The *S. methylophilicida* phage encoded a putative superinfection exclusion lipoprotein related to the Llp protein of phage T5, which locks the phage receptor in an inactive conformation to preclude subsequent phage adsorption (77). We also identified several potential counterdefense genes, which could inhibit the host defenses. In particular, a S-adenosyl-L-methionine (SAM) hydrolase gp73, which can inhibit the host restriction-modification and other methylation-dependent defense systems by hydrolyzing the essential SAM cofactor; a HD superfamily phosphohydrolase/phosphodiesterase, likely capable of clearing out cyclic nucleotide second messengers that commonly act as activators of diverse defense systems; and a Cas4-like nuclease, which in some viruses functions as an inhibitor of the adaptation stage of the CRISPR-Cas immunity (78).

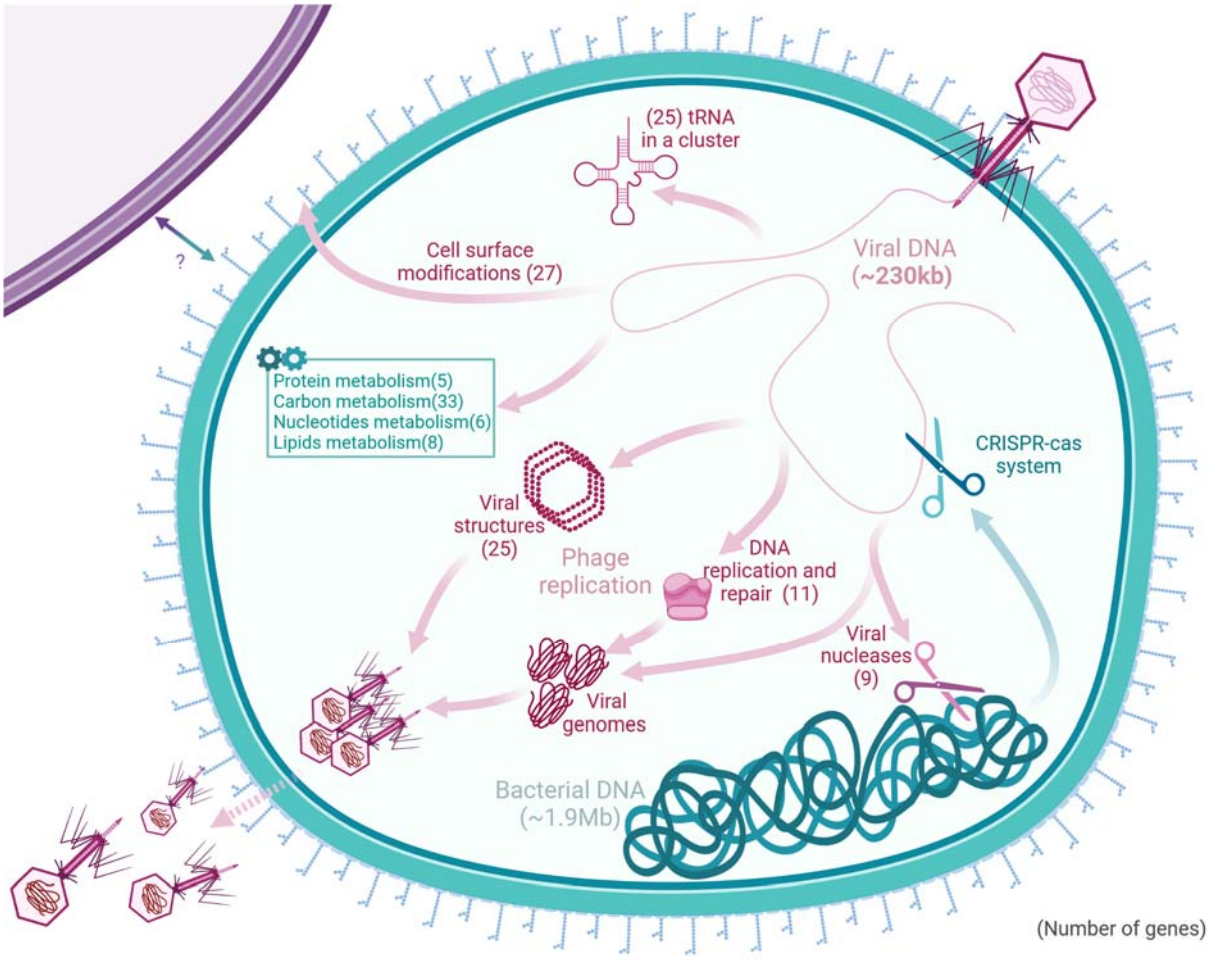

Fig. 6. Genome-based inference of the interaction between *S. methylophilicida* and its jumbo phage. *S. methylophilicida* systems are indicated in blue and those of the phage in pink. Numbers in parentheses indicate the number of genes involved in each system.

Auxiliary metabolic genes (AMGs) are common in phages with large genomes, likely involved in facilitating the hijacking and rewiring of the host metabolism (79). We detected an array of AMGs that may have an impact on the tripartite Methylophilaceae-*Strigamonas*-phage interaction. A remarkable feature of our phage genome was a cluster of 25 co-localized tRNA genes covering all 20 proteinogenic amino acids (Fig. 5b, Table S5). The number of tRNA genes has been reported to increase with the size of phage genomes. Thus, jumbo and megaphages may contain multiple tRNA genes (73,80,81). These tRNAs can help to decrease the dependency of the phage on the host gene expression machinery and to move between multiple hosts with different codon usage biases. Alternatively, phage-encoded tRNA genes could help to overcome host defenses mediated by tRNA-degrading ribonucleases deployed during phage infection (82). We also detected several genes that may modulate the carbon metabolism of the host, such as the 2-oxoglutarate/2-oxoacid ferredoxin oxidoreductase subunit alpha, and the acetoin:2,6-dichlorophenolindophenol oxidoreductase AcoA as well as alpha and beta subunits of the pyruvate dehydrogenase E1. Given the central place of pyruvate in the metabolism of *S. methylophilicida*, the phage-encoded pyruvate

dehydrogenase, which converts pyruvate to acetyl-CoA, could rewire the host carbon metabolism towards lipid and carbohydrate synthesis. Notably, we identified eight genes encoding acyl-CoA thioesterases/lipases of the SGNH-hydrolase superfamily, which could participate in lipid biosynthesis or degradation. The phage also contained genes involved in amino acid metabolism, including methionyl aminopeptidase, adenosylhomocysteine nucleosidase, threonine/serine dehydratases, prolyl 4-hydroxylase, and peptide deformylase, which could participate in the degradation of host proteins and peptides or in the modulation of the amino acid pool for phage protein synthesis. Notably, the phage also appears to control the proper folding of its proteins by supplementing the host with the GroEL/GroES chaperone-chaperonin system and with a small heat shock protein.

If the jumbo phage had also the capacity to infect the Methylophilaceae host of *S. methylophilicida*, as suggested by its compatibility with the two genetic codes 11 and 25, it also contained genes that may affect the Methylophilaceae metabolism. The phage encoded a putative RuBisCO activase CbbQ, which could modulate the activity of the Methylophilaceae-encoded RuBisCO. Besides, the most prominent category of AMGs in the phage genome were those related to carbohydrate synthesis and modification (n=33), potentially involved in cell envelope biosynthesis. These included genes coding for D-sedoheptulose-7-phosphate isomerase and D-glycero-alpha-D-manno-heptose-7-phosphate kinase, involved in the synthesis of the S-layer, two copies of the GDP-L-fucose synthase gene, and genes coding for the dTDP-4-dehydrorhamnose reductase, dTDP-4-dehydrorhamnose 3,5-epimerase, and CDP-4-dehydro-6-deoxyglucose reductase. All these genes are involved in the biosynthesis and modification of precursors of the lipopolysaccharide (LPS) O-antigen and the glycoproteins of the bacterial S-layer (83). The phage also encoded an UDP-2,3-diacylglucosamine hydrolase, which participates in the synthesis of lipids of Gram-negative bacteria, and two glycerol-3-phosphate cytidylyltransferases involved in the biosynthesis of teichoic acid in Gram-positive bacteria. Additionally, 16 glycosyltransferase genes were also present in the phage genome, which can induce modifications in the cell envelope of the host (84). In fact, jumbo phages have been reported to present multiple cell-surface modification genes. For example, genes related to LPS biosynthesis and modification have been found in phages isolated from distinct geographical locations (74,84). These cell envelope modifications are hypothesized to prevent viral super-infections, particularly from small phages that will replicate faster than the much bigger jumbo phages (74). They are also suggested to impact the communication and interaction of the phage host with other cells (84). In the context of the Patescibacteria-Methylophilaceae interaction, our TEM observations and the presence of many genes coding for proteins with S-layer domains suggested that the cells of *S. methylophilicida* most likely present an S-layer. Consistent with the genes found in the phage, the phage infection most likely results in the modification of the cell surface (particularly the S-layer) of *S. methylophilicida*, which may impact, potentially enhancing, the obligatory interaction with the Methylophilaceae host. Surprisingly, the closely related phage identified in GenBank has almost no genes involved in cell-surface modification.

Finally, the phage encoded several proteins targeted to the outer membrane. These included the putative TonB-dependent oligosaccharide transporter SusC and secretion channel CsgG, a key component of the curli biogenesis system. Notably, curli represent the major proteinaceous components of a complex extra-cellular matrix produced by many bacteria and are involved in adhesion to surfaces, cell aggregation, and biofilm formation (85). Given that in the tripartite system analyzed here, the outer membrane is only present in the

Methylophilaceae host of *S. methylophilicida*, the phage-encoded proteins could be directed to the Methylophilaceae partner and directly impact the intercellular interactions.

Conclusions

The newly characterized species *Strigamonas methylophilicida* belongs to a novel genus in the patescibacterial family Absconditicoccaceae. To date, this is the third genus in this family to undergo comprehensive analysis, allowing for more precise identification of shared family traits and enriching them with novel, unreported features. Like other Absconditicoccaceae, *S. methylophilicida* exhibits an ectoparasitic lifestyle, evident in TEM observations that depict its small cell size and close attachment to the host, forming occasionally cell bridges between host cells. In addition, its genome sequence indicates a strong reliance on the host to obtain essential components such as fatty acids, amino acids, and nucleic bases. It also displays a substantial presence of strain/genus-specific genes, supporting the previously reported dynamic evolution of gene content in the Absconditibacteraceae family (Moreira et al., 2021). *Strigamonas* stands out due to its large genome size, approximately 1.9 Mb, and the production of extracellular vesicles, structures previously unseen in Patescibacteria and that could be involved in the interaction with its host. The interaction between *S. methylophilicida* and a Methylophilaceae host broadens the spectrum of known hosts of Patescibacteria.

We have unveiled a novel phage that appears to target *S. methylophilicida*. Our analyses indicate that this phage is compatible with cell machinery operating on genetic codes 11 and 25, suggesting a broad host range. It exhibits a large genome and an extensive coding repertoire, including genes necessary for its replication but also genes altering the metabolism and cell surface of *S. methylophilicida* (Fig. 6), a trait commonly found in other phages with large genomes to prevent superinfections by other smaller phages. It is also possible that this phage impacts the cell-to-cell contact between *S. methylophilicida* and the Methylophilaceae cells. Alternatively, it could target the metabolism of *S. methylophilicida*, for example by using its machinery to synthesize nucleotides lipids, and some amino acids, or even the Methylophilaceae central metabolism, in particular its ability to fix carbon. Further research is needed to unravel the precise mechanisms through which this phage operates within this intricate ecological interaction.

Taxonomy

Based on the genetic distance with other patescibacterial genomes, we propose that our new MAGs define a novel genus and species within the Absconditicoccaceae, which we have described under the SeqCode (86).

Description of *Strigamonas* gen. nov.

Strigamonas (Stri.ga.mo'nas. N.L. fem. n. *striga*, a female evil spirit or vampire; L. fem. n. *monas*, a unit, monad; N.L. fem. n. *Strigamonas*, a female vampire monad).

Epibiotic bacteria that prey on methylotrophic proteobacteria of the family Methylophilaceae. Non-flagellated cells attached to the surface of the host. Gram-positive cell wall structure.

The type species is *Strigamonas methylophilicida*.

Description of *Strigamonas methylophilicida* sp. nov.

Strigamonas methylophilicida (me.thy.lo.phi.li.ci'da. N.L. neut. n. *methylum*, the methyl group; from French masc. n. *méthyle*; from French masc. n. *méthylène*; from Gr. neut. n. *methy*, wine; from Gr. fem. n. *hylê*, wood; N.L. neut. n. *methyl*, pertaining to the methyl radical; N.L. masc. adj. suff. *-philus*, friend, loving; from Gr. masc. adj. *philos*, on; L. masc. n. suff. *-cida*, killer; from L. v. *caedo*, to cut, kill; N.L. masc. n. *methylophilicida*, a killer of the methyl radical lover).

Displays the following properties in addition to those given by the genus description. The cells are rounded and slightly flattened, approximately 300 nm diameter, and form piles of up to 4 cells attached to the surface of their host. Found in freshwater environments. It has a 1.9 Mb genome, with a G+C content of 27-28%. It lacks most biosynthetic pathways, most likely growing as an epibiotic parasite. It is known from environmental sequencing and microscopy observation only.

*Strigamonas*_MG is the designated type MAG.

References

1. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil PA, et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat Biotechnol. 2018 Nov;36(10):996–1004.
2. Brown CT, Hug LA, Thomas BC, Sharon I, Castelle CJ, Singh A, et al. Unusual biology across a group comprising more than 15% of domain Bacteria. Nature. 2015 Jul;523(7559):208–11.
3. Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, et al. A new view of the tree of life. Nat Microbiol. 2016 Apr 11;1(5):1–6.
4. Parks DH, Rinke C, Chuvochina M, Chaumeil PA, Woodcroft BJ, Evans PN, et al. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. Nat Microbiol. 2017 Nov;2(11):1533–42.
5. Schulz F, Eloe-Fadrosh EA, Bowers RM, Jarett J, Nielsen T, Ivanova NN, et al. Towards a balanced view of the bacterial tree of life. Microbiome. 2017 Dec;5(1):1–6.
6. Luef B, Frischkorn KR, Wrighton KC, Holman HYN, Birarda G, Thomas BC, et al. Diverse uncultivated ultra-small bacterial cells in groundwater. Nat Commun. 2015 Feb 27;6(1):6372.
7. Castelle CJ, Banfield JF. Major New Microbial Groups Expand Diversity and Alter our Understanding of the Tree of Life. Cell. 2018 08;172(6):1181–97.
8. Castelle CJ, Brown CT, Anantharaman K, Probst AJ, Huang RH, Banfield JF. Biosynthetic capacity, metabolic variety and unusual biology in the CPR and DPANN radiations. Nat Rev Microbiol. 2018 Oct;16(10):629–45.
9. He X, McLean JS, Edlund A, Yooseph S, Hall AP, Liu SY, et al. Cultivation of a human-associated TM7 phylotype reveals a reduced genome and epibiotic parasitic lifestyle. Proc Natl Acad Sci. 2015 Jan 6;112(1):244–9.
10. Bor B, Bedree JK, Shi W, McLean JS, He X. Saccharibacteria (TM7) in the Human Oral Microbiome. J Dent Res. 2019 May;98(5):500–9.
11. Cross KL, Campbell JH, Balachandran M, Campbell AG, Cooper SJ, Griffen A, et al. Targeted isolation and cultivation of uncultivated bacteria by reverse genomics. Nat Biotechnol. 2019 Nov;37(11):1314.
12. Murugkar PP, Collins AJ, Chen T, Dewhirst FE. Isolation and cultivation of candidate phyla radiation Saccharibacteria (TM7) bacteria in coculture with bacterial hosts. J Oral Microbiol. 2020 Jan 1;12(1):1814666.
13. Moreira D, Zivanovic Y, López-Archilla AI, Iniesto M, López-García P. Reductive evolution and unique predatory mode in the CPR bacterium Vampirococcus lugosii. Nat Commun. 2021 Dec;12(1):2454.
14. Yakimov MM, Merkel AY, Gaisin VA, Pilhofer M, Messina E, Hallsworth JE, et al. Cultivation of a vampire: 'Candidatus Absconditicoccus praedator'. Environ Microbiol. 2022;24(1):30–49.
15. Gong J, Qing Y, Guo X, Warren A. "Candidatus Sonnebornia yantaiensis", a member of candidate division OD1, as intracellular bacteria of the ciliated protist Paramecium bursaria (Ciliophora, Oligohymenophorea). Syst Appl Microbiol. 2014 Feb 1;37(1):35–41.
16. Kuroda K, Yamamoto K, Nakai R, Hirakata Y, Kubota K, Nobu MK, et al. Symbiosis between Candidatus Patescibacteria and Archaea Discovered in Wastewater-Treating Bioreactors. mBio. 2022 Aug 31;0(0):e01711-22.
17. Paez-Espino D, Eloe-Fadrosh EA, Pavlopoulos GA, Thomas AD, Huntemann M, Mikhailova N, et al. Uncovering Earth's virome. Nature. 2016 Aug;536(7617):425–30.
18. Liu J, Jaffe AL, Chen L, Bor B, Banfield JF. Host translation machinery is not a barrier to phages that infect both CPR and non-CPR bacteria [Internet]. bioRxiv; 2022 [cited 2023 Feb 13]. p. 2022.11.22.517103. Available from: https://www.biorxiv.org/content/10.1101/2022.11.22.517103v2
19. David GM, López-García P, Moreira D, Alric B, Deschamps P, Bertolino P, et al. Small freshwater ecosystems with dissimilar microbial communities exhibit similar temporal patterns. Mol Ecol. 2021;30(9):2162–77.

20. Whittenbury R, Phillips KC, Wilkinson JF. Enrichment, Isolation and Some Properties of Methane-utilizing Bacteria. Microbiology. 1970;61(2):205–18.
21. Andrews, S. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data [Internet]. 2010 [cited 2023 Dec 12]. Available from: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
22. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014 Aug 1;30(15):2114–20.
23. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. Genome Res. 2017 May;27(5):824–34.
24. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, et al. Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ. 2015 Oct 8;3:e1319.
25. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010 Mar 8;11(1):119.
26. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990 Oct 5;215(3):403–10.
27. Menzel P, Ng KL, Krogh A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat Commun. 2016 Apr 13;7(1):11257.
28. Galperin MY, Makarova KS, Wolf YI, Koonin EV. Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Res. 2015 Jan 28;43(Database issue):D261–9.
29. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 2016 Jan 4;44(D1):D733-745.
30. Tatusov RL, Koonin EV, Lipman DJ. A Genomic Perspective on Protein Families. Science. 1997 Oct 24;278(5338):631–7.
31. Rodriguez-R LM, Konstantinidis KT. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes [Internet]. PeerJ Inc.; 2016 Mar [cited 2023 Nov 9]. Report No.: e1900v1. Available from: https://peerj.com/preprints/1900
32. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. Bioinformatics. 2022 Dec 1;38(23):5315–6.
33. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics. 2020 Mar 1;36(6):1925–7.
34. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014 Jul 15;30(14):2068–9.
35. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. J Mol Biol. 2016 Feb 22;428(4):726–31.
36. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021 Aug;596(7873):583–9.
37. Holm L, Laiho A, Törönen P, Salgado M. DALI shines a light on remote homologs: One hundred discoveries. Protein Sci. 2023;32(1):e4519.
38. Delmont TO, Eren AM. Linking pangenomes and metagenomes: the Prochlorococcus metapangenome. PeerJ. 2018 Jan 25;6:e4320.
39. Van Dongen S. Graph Clustering Via a Discrete Uncoupling Process. SIAM J Matrix Anal Appl. 2008 Jan;30(1):121–41.
40. Couvin D, Bernheim A, Toffano-Nioche C, Touchon M, Michalik J, Néron B, et al. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. Nucleic Acids Res. 2018 Jul 2;46(W1):W246–51.
41. Simon M, Jardillier L, Deschamps P, Moreira D, Restoux G, Bertolino P, et al. Complex communities of small protists and unexpected occurrence of typical marine lineages in shallow freshwater systems. Environ Microbiol. 2015 Oct;17(10):3610–27.

42. Nayfach S, Camargo AP, Schulz F, Eloe-Fadrosh E, Roux S, Kyrpides NC. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. Nat Biotechnol. 2021 May;39(5):578–85.
43. Pons JC, Paez-Espino D, Riera G, Ivanova N, Kyrpides NC, Llabrés M. VPF-Class: taxonomic assignment and host prediction of uncultivated viruses based on viral protein families. Bioinforma Oxf Engl. 2021 Jul 27;37(13):1805–13.
44. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. Mol Biol Evol. 2021 Dec 1;38(12):5825–9.
45. Kilinc M, Jia K, Jernigan RL. Protein Language Model Performs Efficient Homology Detection [Internet]. bioRxiv; 2022 [cited 2023 Nov 9]. p. 2022.03.10.483778. Available from: https://www.biorxiv.org/content/10.1101/2022.03.10.483778v1
46. Söding J. Protein homology detection by HMM-HMM comparison. Bioinforma Oxf Engl. 2005 Apr 1;21(7):951–60.
47. Kanehisa M, Sato Y. KEGG Mapper for inferring cellular functions from protein sequences. Protein Sci Publ Protein Soc. 2020 Jan;29(1):28–35.
48. Gilchrist CLM, Chooi YH. clinker & clustermap.js: automatic generation of gene cluster comparison figures. Bioinformatics. 2021 Aug 25;37(16):2473–5.
49. Pruesse E, Peplies J, Glöckner FO. SINA: Accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics. 2012 Jul 15;28(14):1823–9.
50. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinforma Oxf Engl. 2009 Aug 1;25(15):1972–3.
51. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol Biol Evol. 2015 Jan;32(1):268–74.
52. Katoh K, Misawa K, Kuma K ichi, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002 Jul 15;30(14):3059–66.
53. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021 Jul 2;49(W1):W293–6.
54. Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil PA, Hugenholtz P. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. Nucleic Acids Res. 2022 Jan 7;50(D1):D785–94.
55. Kelly DP, Ardley JK, Wood AP. Cultivation of Methylotrophs. In: McGenity TJ, Timmis KN, Nogales B, editors. Hydrocarbon and Lipid Microbiology Protocols: Isolation and Cultivation [Internet]. Berlin, Heidelberg: Springer; 2017 [cited 2023 Nov 2]. p. 197–229. (Springer Protocols Handbooks). Available from: https://doi.org/10.1007/8623_2014_4
56. Kalyuzhnaya MG, Beck DAC, Vorobev A, Smalley N, Kunkel DD, Lidstrom ME, et al. Novel methylotrophic isolates from lake sediment, description of Methylotenera versatilis sp. nov. and emended description of the genus Methylotenera. Int J Syst Evol Microbiol. 2012 Jan;62(Pt 1):106–11.
57. Campbell JH, O'Donoghue P, Campbell AG, Schwientek P, Sczyrba A, Woyke T, et al. UGA is an additional glycine codon in uncultured SR1 bacteria from the human microbiota. Proc Natl Acad Sci. 2013 Apr 2;110(14):5540–5.
58. Méheust R, Burstein D, Castelle C, Banfield J. The distinction of CPR bacteria from other bacteria based on protein family content. Nat Commun. 2019 Sep 13;10.
59. Shaiber A, Willis AD, Delmont TO, Roux S, Chen LX, Schmid AC, et al. Functional and genetic markers of niche partitioning among enigmatic members of the human oral microbiome. Genome Biol. 2020 Dec 16;21(1):292.
60. Konstantinidis KT, Rosselló-Móra R, Amann R. Uncultivated microbes in need of their own taxonomy. ISME J. 2017 Nov;11(11):2399–406.
61. Rosselló-Móra R, Amann R. Past and future species definitions for Bacteria and Archaea. Syst Appl Microbiol. 2015 Jun 1;38(4):209–16.

62. Konstantinidis KT, Tiedje JM. Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead. Curr Opin Microbiol. 2007 Oct 1;10(5):504–9.
63. Garritano AN, Song W, Thomas T. Carbon fixation pathways across the bacterial and archaeal tree of life. PNAS Nexus. 2022 Nov 1;1(5):pgac226.
64. David SS, O'Shea VL, Kundu S. Base Excision Repair of Oxidative DNA Damage. Nature. 2007 Jun 21;447(7147):941–50.
65. Dy RL, Przybilski R, Semeijn K, Salmond GPC, Fineran PC. A widespread bacteriophage abortive infection system functions through a Type IV toxin–antitoxin mechanism. Nucleic Acids Res. 2014 Apr 1;42(7):4590–605.
66. Wood TL, Wood TK. The HigB/HigA toxin/antitoxin system of Pseudomonas aeruginosa influences the virulence factors pyochelin, pyocyanin, and biofilm formation. MicrobiologyOpen. 2016 Mar 14;5(3):499–511.
67. Prysak MH, Mozdzierz CJ, Cook AM, Zhu L, Zhang Y, Inouye M, et al. Bacterial toxin YafQ is an endoribonuclease that associates with the ribosome and blocks translation elongation through sequence-specific and frame-dependent mRNA cleavage. Mol Microbiol. 2009 Mar;71(5):1071–87.
68. Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, et al. An updated evolutionary classification of CRISPR–Cas systems. Nat Rev Microbiol. 2015 Nov;13(11):722–36.
69. Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, et al. Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. Cell. 2015 Oct 22;163(3):759–71.
70. Makarova KS, Wolf YI, Iranzo J, Shmakov SA, Alkhnbashi OS, Brouns SJJ, et al. Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. Nat Rev Microbiol. 2020 Feb;18(2):67–83.
71. Makarova KS, Wolf YI, Snir S, Koonin EV. Defense Islands in Bacterial and Archaeal Genomes and Prediction of Novel Defense Systems. J Bacteriol. 2011 Oct 12;193(21):6039–56.
72. Kavagutti VS, Andrei AŞ, Mehrshad M, Salcher MM, Ghai R. Phage-centric ecological interactions in aquatic ecosystems revealed through ultra-deep metagenomics. Microbiome. 2019 Dec;7(1):1–15.
73. Yuan Y, Gao M. Jumbo Bacteriophages: An Overview. Front Microbiol [Internet]. 2017 [cited 2023 Apr 14];8. Available from: https://www.frontiersin.org/articles/10.3389/fmicb.2017.00403
74. Younker IT, Duffy C. Jumbo Phages. In: Bamford DH, Zuckerman M, editors. Encyclopedia of Virology (Fourth Edition) [Internet]. Oxford: Academic Press; 2021 [cited 2023 Jun 2]. p. 229–41. Available from: https://www.sciencedirect.com/science/article/pii/B9780128096338215227
75. Turner D, Shkoporov AN, Lood C, Millard AD, Dutilh BE, Alfenas-Zerbini P, et al. Abolishment of morphology-based taxa and change to binomial species names: 2022 taxonomy update of the ICTV bacterial viruses subcommittee. Arch Virol. 2023 Jan 23;168(2):74.
76. Borges AL, Lou YC, Sachdeva R, Al-Shayeb B, Penev PI, Jaffe AL, et al. Widespread stop-codon recoding in bacteriophages may regulate translation of lytic genes. Nat Microbiol. 2022 Jun;7(6):918–27.
77. van den Berg B, Silale A, Baslé A, Brandner AF, Mader SL, Khalid S. Structural basis for host recognition and superinfection exclusion by bacteriophage T5. Proc Natl Acad Sci U S A. 2022 Oct 18;119(42):e2211672119.
78. Zhang Z, Pan S, Liu T, Li Y, Peng N. Cas4 Nucleases Can Effect Specific Integration of CRISPR Spacers. J Bacteriol. 2019 Jun 15;201(12):e00747-18.
79. Naknaen A, Suttinun O, Surachat K, Khan E, Pomwised R. A Novel Jumbo Phage PhiMa05 Inhibits Harmful Microcystis sp. Front Microbiol [Internet]. 2021 [cited 2023 May 18];12. Available from: https://www.frontiersin.org/articles/10.3389/fmicb.2021.660351
80. Devoto AE, Santini JM, Olm MR, Anantharaman K, Munk P, Tung J, et al. Megaphages infect Prevotella and variants are widespread in gut microbiomes. Nat Microbiol. 2019

Apr;4(4):693–700.
81. Al-Shayeb B, Sachdeva R, Chen LX, Ward F, Munk P, Devoto A, et al. Clades of huge phages from across Earth's ecosystems. Nature. 2020 Feb;578(7795):425–31.
82. van den Berg DF, van der Steen BA, Costa AR, Brouns SJJ. Phage tRNAs evade tRNA-targeting host defenses through anticodon loop mutations. eLife. 2023 Jun 2;12:e85183.
83. Messner P, Steiner K, Zarschler K, Schäffer C. S-layer nanoglycobiology of bacteria. Carbohydr Res. 2008 Aug 11;343(12):1934–51.
84. Wang Q, Cai L, Zhang R, Wei S, Li F, Liu Y, et al. A Unique Set of Auxiliary Metabolic Genes Found in an Isolated Cyanophage Sheds New Light on Marine Phage-Host Interactions. Microbiol Spectr. 2022 Oct 3;10(5):e02367-22.
85. Barnhart MM, Chapman MR. Curli biogenesis and function. Annu Rev Microbiol. 2006;60:131–47.
86. Hedlund BP, Chuvochina M, Hugenholtz P, Konstantinidis KT, Murray AE, Palmer M, et al. SeqCode: a nomenclatural code for prokaryotes described from sequence data. Nat Microbiol. 2022 Oct;7(10):1702–8.