Tripartite interaction of a patescibacterial epibiont, a methylotrophic gammaproteobacterial host and a jumbo phage

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]✉ and 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

✉email:david.moreira@universite-paris-saclay.fr; ludwig.jardillier@universite-paris-saclay.fr

ABSTRACT

Patescibacteria form a very diverse and widely distributed phylum of small bacteria inferred to have an episymbiotic lifestyle. However, the prevalence of this lifestyle within the phylum and their host specificity remain poorly known due to the scarcity of cultured representatives. Here, we describe a tripartite interaction of a patescibacterium, its gammaproteobacterial host, and a patescibacterial phage based on metagenomic data and enrichment cultures from two shallow, geographically close, freshwater ecosystems. The patescibacterium, *Strigamonas methylophilicida* sp. nov., defines a new genus within the family Absconditicoccaceae. It grows as epibiont on cells of methanotrophic species of the gammaproteobacterial family Methylophilaceae. *Strigamonas* cells are tightly attached to the host, sometimes forming stacks that connect two host cells. Despite a surprisingly large genome (1.9 Mb) compared to many other patescibacteria, *S. methylophilicida* lacks many essential biosynthetic pathways, including the complete biosynthesis of phospholipids, amino acids, and nucleic acids, implying a dependence on the host to obtain these molecules. From metagenomes of these ecosystems, we identified and assembled the genome of a jumbo phage likely infecting the patescibacterium based on a CRISPR spacer match. Its large genome (230 kb) contained a wide gene repertoire, including genes probably involved in the modification of the *Strigamonas* metabolism and cell surface. By manipulating the patescibacterial phenotype during infection, the phage likely influences in turn the interaction of the patescibacterial epibiont with its methylotrophic host in a parasitic *ménage-à-trois*. Our results confirm a prevalent episymbiotic lifestyle in Absconditicoccaceae and further suggest a clade-specific adaptation of these patescibacteria for gammaproteobacterial 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, 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, which consistently found reduced genomes coding for limited metabolic capabilities [2, 6, 7]. Notably, these bacteria lack the typical pathways for the synthesis of amino acids, nucleotides, and membrane phospholipids. Most likely, they obtain these essential molecules from other cells, which suggests that a symbiotic, likely parasitic, lifestyle is widespread within this phylum [2, 7, 8]. In agreement with these inferences, microscopy observations have shown that patescibacterial cells are very small and 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 [8]), and it cannot be discarded that part of them provide alternative ways to fulfill some of the apparently 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 from 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 subsequently cultured Saccharimonadia were also consistently found to be parasites of actinobacterial hosts [10–12]. Recently, two other patescibacterial representatives have been described, *Candidatus* Vampirococcus lugosii and *Candidatus* Absconditicoccus praedator, representing two distant genera in the family Absconditicoccaceae [13, 14]. They were both found in saline lakes, as episymbionts of anoxygenic photosynthetic bacteria belonging to the Chromatiales (Gammaproteobacteria, within the Pseudomonadota). 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 patescibacterial groups with cultured members appear to be specialized in interacting with hosts from specific lineages: Nanosynbacteraceae (class Saccharimonadia) parasitize Actinobacteria and Absconditicoccaceae parasitize photosynthetic Gammaproteobacteria (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 hampered 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 ones [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 gammaproteobacterial host. *Strigamonas* exhibits an S-layer and produces unique extracellular vesicles. Although it lacks many classical pathways for the biosynthesis of essential molecules, its 1.9 Mb genome is substantially larger than those of other Absconditicoccaceae representatives. By analyzing CRISPR spacers, we identified a phage likely targeting *S. methylophilicida*. The recovered phage genome 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.

MATERIALS AND METHODS

Sampling and enrichment of patescibacteria-host consortia

Water samples were collected from two shallow freshwater ecosystems in the Regional Parc de la Haute Vallée de Chevreuse (south of Paris), including the small stream Ru Sainte Anne (RSA) and the 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 (see below). After approximately one month of incubation, we observed small cells attached to bigger, 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 concentrated 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 488 nm, 405 nm, and 375 nm lasers. Side and Forward Scatter Areas (SSC-A and FSC-A) were used to discriminate the epibiont-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 applying voltages of 100V and 300V, respectively. The trigger was conjointly set to the FSC and SSC with a threshold of 500 and the fluidic system was run at 20 psi (3.102 bar) with a 100 µm nozzle. Samples were sorted at a speed of 3000–3500 events s^{-1} using sterile 1X PBS buffer as sheath fluid. We collected eight samples in 1.5 ml individual tubes (200 sorted events per tube) using the accurate “single-cell purity” sorting mode. The purity of the sorted events was visually inspected 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 (TEM), cells were fixed with 2.5%

glutaraldehyde in 0.2M cacodylate buffer for 2 hours on ice. Then, cells were washed with 0.2M cacodylate, post-fixed in 1% osmium for 1 h at RT, and subsequently dehydrated with successive baths of increasing ethanol concentrations (30%, 50%, 70%, 80%, 90%, 95%, 2 x 100%, for 10 min each) and, finally, in a 1:1 100% ethanol:acetone mix for 10 min and twice in pure acetone for 10 min. The dehydrated cells were then embedded in a 1:1 acetone: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 REPLI-g Single Cell kit (Qiagen) was used to extract DNA from the FACS-sorted cells and amplify their whole genomes by multiple displacement amplification (MDA) following the manufacturer's instructions. The amplified DNAs were diluted to 1/50 and 16S rRNA genes were amplified by PCR with the Platinum Taq DNA polymerase (Invitrogen) 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 were U515F (5'-GTGCCAGCMGCCGCGGTAA-3') and 926R (5′-CCGYCAATTYMTTTRAGTTT-3′). 16S rRNA gene clone libraries were produced with the Topo TA cloning kit following the manufacturer's instructions (Invitrogen). Individual clone inserts were amplified with 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 produced 16S rRNA gene aplicons affiliating to Patescibacteria and sequenced them (Eurofins Genomics, Ebersberg, Germany) using Illumina HiSeq (2×125 bp paired-end reads). Quality check and trimming of the reads were done with FastQC and Trimmomatic (MAXINFO:40:0.8, MINLEN:30 LEADING:20, TRAILING:20) [21, 22], respectively. Sequence 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. Three co-assemblies were performed: i) merging the three WGA datasets from the RSA ecosystem, ii) merging the two WGA datasets from the MG ecosystem, and iii) merging all the RSA and MG datasets. We used the Anvi'o v6 metagenomic workflow [24] to reconstruct MAGs. 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. We avoided using automatic binning to minimize the use of differential sequence coverage, which can be biased by the WGA. Contigs with a minimum length of 1000 bp were clustered hierarchically for manual binning via the Anvi'o v6 interactive interface [24]. 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]. 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
170 them with the GhostKOALA web server [35]. Protein structures were predicted with Alphafold [36]. Structural comparisons to the Protein Data Bank (PDB) were done on the DALI 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
175 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 MAGs using CRISPR-Cas finder [40]. A sequence logo of CRISPR spacer sequences was created using WebLogo 3 (https://weblogo.threeplusone.com/create.cgi) [41]. The identified CRISPR spacers were used
180 as query sequences to BLAST the metagenomes from filtered plankton samples collected in previous years from the same small freshwater ecosystems as described [19]. Total DNA was extracted from the filtered cell fraction-size 0.2–5 µm with the Power Soil kit (Qiagen) as described previously [19, 42], and Illumina-sequenced as described above (Eurofins Genomics). We identified in this way a phage genome from the MG metagenome. Its
185 completeness was estimated using checkV [43]. We classified it using the VPF-class tool [44] and the result was subsequently assigned to the corresponding ICTV classification [45]. The gene prediction of the viral genome was done 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 [46], the Prost online tool [47], and
190 HHsearch [48]. Visualizations were carried out using Kegg mapper [49] and Clinker [50].

Phylogenetic analyses

16S rRNA gene sequences were aligned against the Silva 138.1 database with SINA [51]. The multiple sequence alignment of the ~100 closest sequences was trimmed with TrimAl v1.4.rev22 -automated1 [52], and the corresponding maximum likelihood (ML) phylogenetic
195 tree was reconstructed with IQ-tree 1.6.11 [53] with the GTR+G model. Protein sequence alignments were produced using MAFFT v7.450 -auto [54] and trimmed with trimAl v1.4.rev22 -automated1 [52]. Single-protein ML phylogenetic trees were constructed using IQ-tree 1.6.11 [53] 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
200 IQ-tree 1.6.11 with the LG+C20+F+G model for the Patescibacteria and the LG+C60+F+G model for the Methylophilaceae. The gene markers used for the phylogenomic placement of the Patescibacteria were extracted using GTDB-tk, option identify [1, 32]. For all phylogenetic trees, branch supports were estimated using the ultrafast bootstrap method (1000 replicates) implemented in IQ-tree [53]. Trees were visualized using iTOL [55].

205

RESULTS AND DISCUSSION

Identification of an Absconditicoccaceae-Methylophilaceae consortium

A previous multiannual study of geographically-close small freshwater ecosystems in a forest area near Paris showed that Patescibacteria constituted a large part of the prokaryotic
210 community [19]. In two of these ecosystems, the Ru Sainte Anne brook (RSA) and the Mare Gabard pond (MG), the methylotrophic bacterial families Methylophilaceae and

Methylococcaceae (Beta- and Gammaproteobacteria, respectively, both now classified in the phylum Pseudomonadota [56]) 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 [57]. 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 [58]. Many of these cells presented smaller epibiotic cells strongly attached to them, often forming short piles of two to four cells (Fig. 1A-D and Supplementary Video S1), which sometimes bridged two different 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 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, close attachment, and internal cell organization of the epibionts (Fig. 1E-I) were coherent with those described for Patescibacteria such as *Ca.* Absconditicoccus praedator [14]. The ribosomes were positioned at the periphery of the cells, and a dense central area probably corresponded to highly compacted DNA [6, 14]. The cell envelope structure also resembled that of other Patescibacteria [6, 14]. In contrast with previous descriptions, 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.

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 sort cell populations highly enriched in these consortia (Fig. S2), as determined by microscopy observation. We carried out 16S rRNA gene amplification and sequencing for eight of these populations (four from MG and four from RSA) to identify the bacteria forming these consortia (Fig. S3). Three of them 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 five remaining samples. One RSA sample contained sequences exclusively of these two groups, whereas relatively small proportions (I<30%) of other bacteria were also found in the other samples. Combined with microscopy observations, these results strongly indicated a novel type of Patescibacteria-host interaction established between Absconditicoccaceae and Methylophilaceae.

To better understand this interaction using genomic data, we carried out whole genome amplification (WGA) from five FACS-sorted samples (three from RSA and two from MG) and Illumina-sequenced the amplified DNA. These WGA samples contained DNA from several species (Absconditicoccaceae, Methylophilaceae, and minor contaminants) and hence were treated 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 RSA MAG was 1.94 Mb in size with 81.1% completeness and 2.1% redundancy, while the MG MAG was 1.93 Mb with 83.1% completeness and 1.41% redundancy (Table S2). These values are typically calculated via the identification of single copy genes present in most bacteria [59]. Because Patescibacteria commonly exhibit genome reduction and multiple gene losses, a completeness higher than 70% is considered to correspond to complete and high quality genome sequences [60, 61]. The GC content of the RSA and MG MAGs was 28.47% and 27.01%, respectively. Our samples appeared to contain a mix of several Methylophilaceae species, which made sequence binning and MAG reconstruction for the host more challenging. We therefore co-assembled all samples from both ecosystems to increase genome coverage, which yielded four good-quality Methylophilaceae MAGs after two rounds of manual refinement (completeness between 90.14% and 95.77% and redundancy between 2.82% and 5.63%) (Table S2).

A maximum likelihood (ML) phylogenetic tree based on 14 concatenated ribosomal proteins revealed that the four MAGs 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]. These 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 family (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] did not assign our MAGs to any known genus. Altogether, these metrics supported that our freshwater patescibacterial MAGs defined a novel genus and species [1, 62–64], which we propose to name *Strigamonas methylophilicida* (see Taxonomic appendix below).

Gene content in *Strigamonas methylophilicida*

The *S. methylophilicida* RSA and MG MAGs had 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 [59]. We identified 1,655 and 1,629 coding sequences (CDS) in the two MAGs, notably 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 beyond sequence similarity recognition, *S. methylophilicida* seemed 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 was even more pronounced when *S. methylophilicida* was included, revealing a mere 236 genes shared among the three genera, *Strigamonas*, *Ca.* Absconditicoccus, and *Ca.* Vampirococcus (Fig. 3a). As expected, the majority of these core genes were involved in house-keeping functions such as translation, transcription, and replication (Fig. 3b). By contrast, more than 50% of the genus-specific genes had unknown functions, and the rest were mostly involved in cell wall, membrane, and envelope biogenesis. Among the three Absconditicoccaceae genera, *Strigamonas* stood out for its significant divergence in gene content. In addition to the 236

core genes, it only shared 34 genes with *Ca*. Absconditicoccus and 18 with *Ca*. Vampirococcus. 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. 3a). Many of these genes were organized in clusters in the two MAGs (Fig. S5) and had unknown functions (Fig. 3b, Fig. S6). However, 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 (e.g., MutT, see below), and ABC-type multidrug transport systems (Fig. 3b, Fig. S6, Table S3).

Genes coding for most ribosomal proteins were present (n=51), with only a few missing, such as *rpL30*, 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 membrane lipids. Similarly, biosynthetic pathways for nucleic acids, cofactors, and many canonical amino acids were incomplete. These absences suggest a strong dependence on the host for cell integrity maintenance and growth [7].

Genome-based inference of the metabolism and symbiotic lifestyle

The genes found in the *S. methylophilicida* MAGs suggest reduced carbon and energy metabolisms. As observed previously in *Ca.* V. lugosii and many other Patescibacteria [7, 13], *S. methylophilicida* possessed an incomplete glycolysis pathway, having only the genes involved in the steps from 3-phosphoglycerate to pyruvate (Fig. 4). Therefore, these species most likely require the uptake of 3-phosphoglycerate for ATP production. *Ca.* V. lugosii was hypothesized to obtain it from its photosynthetic host, which produces it abundantly thanks to CO_2 fixation by the ribulose 1,5-bisphosphate carboxylase/oxygenase (RuBisCO) [13]. The Methylophilaceae host of *S. methylophilicida* probably provides 3-phosphoglycerate resulting 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). The genes encoding these two enzymes shared the same genomic context in these two host MAGs, being close to the CO_2-responsive transcriptional regulator CbbR (Fig. S9). Another Methylophilaceae MAG encoding a complete CBB cycle has been recently reported [65], supporting that Methylophilaceae can, at least facultatively, fix carbon and synthesize 3-phosphoglycerate using various pathways. Interestingly, all hosts of Absconditicoccaceae epibionts identified to date possess RuBisCO [13, 14].

The *S. methylophilicida* MAGs encoded several additional carbon metabolism enzymes that are absent in the other characterized Absconditicoccaceae, including malate dehydrogenase, alcohol dehydrogenases, and aldehyde dehydrogenases. Thus, *S. methylophilicida* has the potential to produce or use malate, lactate, ethanol, and acetate from/to pyruvate (Fig. S10). These compounds have been hypothesized to be exported by Patescibacteria and used by other organisms, including their hosts [8]. Additionally, the *S. methylophilicida* MAGs encoded three enzymes of the pentose phosphate pathway: glucose-6-phosphate dehydrogenase (NADP+) and glucose-6-phosphate dehydrogenase [NAD(P)+], which both transform beta-D-glucose-6P into D-glucono-1,5-lactone-6P, and 6-phosphogluconolactonase, 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.

The patescibacterial MAGs also encoded a complete F-type ATP synthase which, based on the sequence characteristics of its subunit c, probably translocates H^+ to synthesize ATP, as suggested in *Ca.* Vampirococcus [13]. However, similarly to many Patescibacteria [8], no classical electron transport chain was present. Alternative proteins to generate proton motive force (PMF) proposed in the other two Absconditicoccaceae representatives [13, 14] were missing 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 (~20 nm), opening the possibility of direct salvage of protons from the host as previously hypothesized for other Patescibacteria [8].

The *S. methylophilicida* MAGs encoded a relatively complex repertoire of cell surface components, a common trait in the family Absconditicoccaceae [13, 14]. They contained genes for the type IV pilus (*pilBCDT*), probably involved in the attachment to their host, for the natural competence system (*comEC*), probably related to DNA uptake from the host, and for the secretion system Sec-SRP (*secDEFGYA* and *yidC*). As in *Ca.* A. praedator [14], 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. The S-layer is likely involved in the physical contact with the host cells. The complete peptidoglycan biosynthesis pathway 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^{2+}-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 *tetA* transporter of the Major Facilitator Superfamily (MFS), involved in the tetracycline efflux resistance mechanism.

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*, involved in the hydrolysis of 8-oxoguanine formed under oxidative stress [66], was present in eight copies, suggesting its particular importance in maintaining genome integrity. Other genes related to oxidative stress response were also detected, including those encoding glutathione peroxidase, glutathione S-transferase, thioredoxin reductase, thioredoxin-dependent peroxiredoxin, and superoxide dismutase Fe-Mn family members. Type I restriction-modification defense systems (*hsdRSM*) were present in four and two copies in RSA and MG MAGs, respectively. Both MAGs also encoded the type IV restriction enzyme Mrr but not the corresponding modification enzyme, whereas in the RSA MAG, a 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 [67, 68]. The toxin YafQ, an endoribonuclease that associates with the ribosome and blocks translation elongation through sequence-specific mRNA cleavage [69], 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 [70]. No other *cas* genes nor CRISPR array could be found in the surrounding genomic context. Additionally, a gene in the MG MAG encoded a Cas12a protein, a signature of the class II type V CRISPR-Cas systems [71, 72]. Similar CRISPR-Cas systems composed only of a CRISPR array and the *cas12a* gene have been previously reported [72]. 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 [72]. The MG MAG Cas12a 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).

Despite the differences mentioned above, the genomic region encoding the CRISPR-Cas systems presented conserved synteny in both *S. methylophilicida* MAGs. Within this region, we identified CRISPR arrays with 4 spacers in the RSA MAG and 6 specers in the MG MAG (Fig. 5a). These spacers showed no sequence similarity between the two MAGs, indicating that the two strains have experienced 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. The absence of the adaptation module suggests that this system is incapable of adding new spacers, which is consistent 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 but also large CRISPR arrays with 20-35 spacers. Interestingly, the direct repeat sequences of our MAGs were similar to those of *Ca.* V. lugosii and *Ca.* A. praedator (Fig. 5b), suggesting that these CRISPR-Cas systems derive from a common ancestral Absconditicoccaceae. The limited number of CRISPR spacers, the absence of the *cas1-cas2* adaptation module in both *S. methylophilicida* MAGs, and the loss of *cas12a* in the RSA MAG, collectively suggest that this ancestral Absconditicoccaceae CRISPR-Cas system is undergoing degradation in *S. methylophilicida*. To compensate for such loss, *S. methylophilicida* seems to have gathered in the nearby genomic context other defense systems, including restriction modification enzymes and an abortive infection (Abi) toxin-antitoxin system. This clustered organization in 'defense islands' is well known in bacteria [73].

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

We used the CRISPR spacers of the *S. methylophilicida* MAGs to search for potential phage genomes in metagenomes from samples collected in previous years from the same MG and RSA ecosystems. One 30 bp spacer from the RSA MAG matched a protospacer in a phage-like contig from the MG ecosystem with 96.6% identity (29 out of 30 positions; 100% coverage; see Fig. 5b). Supporting this CRISPR-based identification, the targeted phage contig was compatible with the Absconditicoccaceae genetic code 25 (UGA translated into glycine), indicating that it can use the *S. methylophilicida* cell machinery to replicate. We observed only

a few differences in the sets of phage proteins predicted using this code or the general prokaryotic code 11 (Fig. 5c). The phage contig was very similar to a phage found in European freshwater lakes and reservoirs (GenBank accession LR796420.1; [74]). However, our phage genome was estimated to be only 57% complete compared to LR796420.1 (189 kb), notably lacking capsid protein genes. We used the LR796420.1 sequence to recruit potential missing parts of our phage from the MG metagenomic assembly. We identified another 100 kb contig containing the genes 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. Its large size (230 kb) qualified it as a jumbo phage [75]. Only around 160 jumbo phage genomes have been reported so far, with sizes between 200 and 500 kb [76]**.** They belong to the class *Caudoviricetes* (tailed double-stranded DNA phages) in the new ICTV classification [45]. Given the small cell size of *S. methylophilicida*, evidence of its infection by a jumbo phage was most surprising.

The jumbo phage genome contained 292 genes. Only six genes were selectively translatable by using either genetic code 11 or 25 (Fig. 5b and Table S4):four exclusive to code 25 and two to code 11. The overall compatibility with both genetic codes suggested that this phage probably has a wide host range. Indeed, other phages that can also employ both genetic codes have been reported and proposed to potentially infect both Patescibacteria and other bacteria [18]. This differential translation of some genes could constitute a host-dependent regulation mechanism. Stop codon-recoding has been hypothesized to participate in regulating the expression of early- and late-stage infection proteins in phages [77]. We were able to functionally annotate three of the six code-specific genes. The first one (code 25-specific) encoded a protein involved in the propagation of phages between cells of multicellular hosts (GO:0046740). It could have a potential role in facilitating phage propagation within the stacks of *S. methylophilicida* cells or between *S. methylophilicida* and its host. The second gene (code 11-specific) coded for a SGNH hydrolase with esterase and lipase activities. Finally, the third gene (code 25-specific) encoded a protein with a transmembrane domain, suggesting a cell surface localisation**.**

Only 58% of the rest of genes of this jumbo phage could be functionally annotated (155 genes; Figs. 5c and 6). Nevertheless, this annotation was more complete than for other jumbo phages, which only have ~20% of genes annotated [76]. 25 genes encoded typical phage structural proteins, including various tail proteins, large terminase subunit, portal protein, HK97-like major capsid protein, and prohead maturation protease, among others. Many of them had homologs in T4-like phages, including the tail sheath protein, strongly suggesting that the *S. methylophilicida* jumbo phage presents a caudoviricete-like morphology with icosahedral head and contractile tail. The phage genome also encoded nine nucleases likely involved in DNA repair and replication, and 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 purine and pyrimidine nucleotide metabolism, including ribonucleotide reductase alpha and beta subunits, dUTP pyrophosphatase, thymidylate kinase, thymidylate synthase, adenylyl-sulfate kinase, and adenine phosphoribosyltransferase, which can provide the DNA precursors needed to replicate this 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. The widespread presence of these genes in jumbo phages has been suggested to represent a way to compensate for their smaller burst size compared to phages with smaller genomes [76]. The *S. methylophilicida* phage encoded a superinfection exclusion lipoprotein related to the Llp protein of phage T5, which precludes subsequent phage adsorption by locking the phage receptor in an inactive conformation [78]. Other potential counterdefense genes could target the host defenses. For example, a S-adenosyl-L-methionine (SAM) hydrolase gp73 can inhibit the host restriction-modification and other methylation-dependent defense systems by hydrolyzing the essential SAM cofactor; an HD superfamily phosphohydrolase/phosphodiesterase can probably clear out cyclic nucleotide second messengers that are common activators of diverse defense systems; and a Cas4-like nuclease, which in some viruses inhibits the adaptation stage of the host CRISPR-Cas immunity [79].

Auxiliary metabolic genes (AMGs), common in phages with large genomes, are likely involved in facilitating the hijacking and rewiring of the host metabolism [80]. We detected an array of AMGs that may influence the tripartite Methylophilaceae-*Strigamonas*-phage interaction. Remarkably, the *S. methylophilicida* phage genome contained a cluster of 25 co-localized tRNA genes covering all 20 proteinogenic amino acids (Fig. 5b, Table S5). The number of tRNA genes is correlated with the size of phage genomes and, therefore, jumbo and megaphages often contain multiple tRNA genes [75, 81, 82]. These genes decrease the dependency of the phage on the host gene expression machinery and allow to infect multiple hosts with different codon usage biases. Phage-encoded tRNA genes can also help overcoming host defensive tRNA-degrading ribonucleases deployed during phage infection [83]. 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, the acetoin:2,6-dichlorophenolindophenol oxidoreductase AcoA, and the pyruvate dehydrogenase E1 alpha and beta subunits. Given the central place of pyruvate in the metabolism of *S. methylophilicida*, the phage-encoded pyruvate dehydrogenase, which converts pyruvate to acetyl-CoA, can rewire the host carbon metabolism towards lipid and carbohydrate synthesis. Notably, we identified eight SGNH-hydrolase superfamily genes encoding acyl-CoA thioesterases/lipases that can 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 can participate in the degradation of host proteins and peptides or in the modulation of the amino acid pool for phage protein synthesis. The phage also appears to control the proper folding of its proteins with a GroEL/GroES chaperone-chaperonin system and a small heat shock protein.

If the jumbo phage can also infect the Methylophilaceae host of *S. methylophilicida*, as its compatibility with the two genetic codes 11 and 25 suggests, it also contains genes that may affect the Methylophilaceae metabolism. In particular, it encoded a RuBisCO activase CbbQ, which could modulate the activity of the Methylophilaceae-encoded RuBisCO. Besides, the most prominent category of AMGs in the phage 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 S-layer synthesis, 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 participate in the biosynthesis and modification of precursors of the lipopolysaccharide (LPS) O-antigen and the glycoproteins of the bacterial S-layer [84]. The phage also encoded an UDP-2,3-diacylglucosamine hydrolase, involved 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 [85, 86]. Additionally, 16 glycosyltransferase genes were also present, which can induce modifications in the cell envelope of the host [87]. In fact, jumbo phages have been reported to present multiple cell-surface modification genes, including genes related to LPS biosynthesis and modification [76, 87]. These cell envelope modifications are hypothesized to prevent viral super-infections, particularly by small phages that can replicate faster than the much bigger jumbo phages [76]. They are also suggested to impact the communication and interaction of the phage host with other cells [87]. In the context of the *Strigamonas*-Methylophilaceae interaction, our TEM observations and the presence of many genes encoding proteins with S-layer domains suggested that the *S. methylophilicida* cells most likely present an S-layer. Consistent with the genes found in the phage, the phage infection probably results in the modification of the cell surface (particularly the S-layer) of *S. methylophilicida*, which may impact the obligatory interaction with the Methylophilaceae host. Surprisingly, the closely related phage identified in GenBank had almost no genes involved in cell-surface modification.

Finally, the phage encoded several proteins targeted to the outer membrane. These included a TonB-dependent oligosaccharide transporter SusC and the secretion channel CsgG, a key component of the curli biogenesis system. Notably, curli represent the major proteinaceous components of a complex extra-cellular matrix in many bacteria and are involved in cell adhesion, cell aggregation, and biofilm formation [88]. Given that in the tripartite system analyzed here, an outer membrane is only present in the Methylophilaceae host, these 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. This is the third genus in this family to undergo comprehensive characterization. The interaction between *S. methylophilicida* and Methylophilaceae hosts broadens the spectrum of known hosts of Patescibacteria. Like other Absconditicoccaceae, *S. methylophilicida* exhibits an ectoparasitic lifestyle, evident in TEM observations illustrating its small cell size and close attachment to the host. In addition, its genome sequence also indicates a strong reliance on the host to obtain essential components, including fatty acids, amino acids, and nucleotides. It also displays a substantial presence of strain- and genus-specific genes, supporting the previously reported dynamic evolution of gene content in the Absconditicoccaceae [13]. *Strigamonas* stands out due to its large genome size, ~1.9 Mb, and the production of extracellular vesicles, structures previously unseen in Patescibacteria and that could potentially be involved in the interaction with its host.

We unveiled a novel jumbo phage that appears to target *S. methylophilicida*. This phage is compatible with genetic codes 11 and 25, suggesting a broad host range. It exhibits a large genome, containing genes necessary for its replication but also genes altering the metabolism and cell surface of *S. methylophilicida* (Fig. 6), a trait common in other phages with large genomes to prevent superinfections by smaller phages. Additionally, it could target the central metabolism of the Methylophilaceae cells, in particular the ability to fix carbon. It appears that this phage also impacts the cell-to-cell contact between *S. methylophilicida* and the Methylophilaceae cells, suggesting a synergistic patescibacterial-viral interaction in a peculiar *ménage-à-trois*. Further research is needed to unravel the precise mechanisms of this intricate tripartite 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 [89].

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 Gammaproteobacteria 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 in 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 of the 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.

DATA AVAILABILITY

The sequence datasets generated and analyzed for this study can be found at the Sequence Read Archive under Bioproject number PRJNA1086478.

REFERENCES

1. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P-A, et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat Biotechnol. 2018;36:996–1004.
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;523:208–211.
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;1:1–6.
4. Parks DH, Rinke C, Chuvochina M, Chaumeil P-A, Woodcroft BJ, Evans PN, et al. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. Nat Microbiol. 2017;2:1533–1542.
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;5:1–6.
6. Luef B, Frischkorn KR, Wrighton KC, Holman H-YN, Birarda G, Thomas BC, et al. Diverse uncultivated ultra-small bacterial cells in groundwater. Nat Commun. 2015;6:6372.
7. Castelle CJ, Banfield JF. Major New Microbial Groups Expand Diversity and Alter our Understanding of the Tree of Life. Cell 2018;172:1181–1197.
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;16:629–645.
9. He X, McLean JS, Edlund A, Yooseph S, Hall AP, Liu S-Y, et al. Cultivation of a human-associated TM7 phylotype reveals a reduced genome and epibiotic parasitic lifestyle. Proc Natl Acad Sci USA 2015;112:244–249.
10. Bor B, Bedree JK, Shi W, McLean JS, He X. Saccharibacteria (TM7) in the Human Oral Microbiome. J Dent Res. 2019;98:500–509.
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;37:1314.
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;12:1814666.
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;12:2454.
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;24:30–49.
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;37:35–41.
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;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;536:425–430.
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. 2022. bioRxiv. , 2022.11.22.517103
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;30:2162–2177.
20. Whittenbury R, Phillips KC, Wilkinson JF. Enrichment, Isolation and Some Properties of Methane-utilizing Bacteria. Microbiology 1970;61:205–218.
21. Andrews, S. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed 12 Dec 2023.
22. Bolger AM, Lohse M, Usadel B. Trimmomatic:a flexible trimmer for Illumina sequence data. Bioinformatics 2014;30:2114–2120.
23. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes:a new versatile metagenomic assembler. Genome Res. 2017;27:824–834.
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;3:e1319.
25. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal:prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 2010;11:119.
26. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–410.
27. Menzel P, Ng KL, Krogh A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat Commun. 2016;7:11257.
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;43:D261–D269.
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;44:D733-745.
30. Tatusov RL, Koonin EV, Lipman DJ. A Genomic Perspective on Protein Families. Science 1997;278:631–637.
31. Rodriguez-R LM, Konstantinidis KT. The enveomics collection:a toolbox for specialized analyses of microbial genomes and metagenomes. PeerJ. 2016;4:e1900v1.
32. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk v2:memory friendly classification with the genome taxonomy database. Bioinformatics 2022;38:5315–5316.
33. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk:a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics 2020;36:1925–1927.
34. Seemann T. Prokka:rapid prokaryotic genome annotation. Bioinformatics 2014;30:2068–2069.
35. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA:KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. J Mol Biol. 2016;428:726–731.
36. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature 2021;596:583–589.
37. Holm L, Laiho A, Törönen P, Salgado M. DALI shines a light on remote homologs:One hundred discoveries. Protein Sci. 2023;32:e4519.
38. Delmont TO, Eren AM. Linking pangenomes and metagenomes: the *Prochlorococcus* metapangenome. PeerJ. 2018;6:e4320.
39. Van Dongen S. Graph Clustering Via a Discrete Uncoupling Process. SIAM J Matrix Anal Appl. 2008;30:121–141.
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;46:W246–W251.
41. Crooks GE, Hon G, Chandonia J-M, Brenner SE. WebLogo:A Sequence Logo Generator. Genome Res. 2004;14:1188–1190.
42. 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;17:3610–3627.

43. 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;39:578–585.
44. 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;37:1805–1813.
45. 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;168:74.
46. 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;38:5825–5829.
47. Kilinc M, Jia K, Jernigan RL. Protein Language Model Performs Efficient Homology Detection. 2022. bioRxiv. 2022.03.10.483778
48. Söding J. Protein homology detection by HMM-HMM comparison. Bioinforma Oxf Engl. 2005;21:951–960.
49. Kanehisa M, Sato Y. KEGG Mapper for inferring cellular functions from protein sequences. Protein Sci Publ Protein Soc. 2020;29:28–35.
50. Gilchrist CLM, Chooi Y-H. clinker & clustermap.js:automatic generation of gene cluster comparison figures. Bioinformatics 2021;37:2473–2475.
51. Pruesse E, Peplies J, Glöckner FO. SINA:Accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics 2012;28:1823–1829.
52. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl:a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinforma Oxf Engl. 2009;25:1972–1973.
53. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE:A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol Biol Evol. 2015;32:268–274.
54. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT:a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002;30:3059–3066.
55. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5:an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021;49:W293–W296.
56. Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil P-A, 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;50:D785–D794.
57. Kelly DP, Ardley JK, Wood AP. Cultivation of Methylotrophs. In:McGenity TJ, Timmis KN, Nogales B (eds). *Hydrocarbon and Lipid Microbiology Protocols:Isolation and Cultivation*. 2017. Springer, Berlin, Heidelberg, pp 197–229.
58. 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;62:106–111.
59. 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 USA 2013;110:5540–5545.
60. Méheust R, Burstein D, Castelle C, Banfield J. The distinction of CPR bacteria from other bacteria based on protein family content. Nat Commun. 2019;10.
61. Shaiber A, Willis AD, Delmont TO, Roux S, Chen L-X, Schmid AC, et al. Functional and genetic markers of niche partitioning among enigmatic members of the human oral microbiome. Genome Biol. 2020;21:292.
62. Konstantinidis KT, Rosselló-Móra R, Amann R. Uncultivated microbes in need of their own taxonomy. ISME J. 2017;11:2399–2406.
63. Rosselló-Móra R, Amann R. Past and future species definitions for Bacteria and Archaea. Syst Appl Microbiol. 2015;38:209–216.
64. Konstantinidis KT, Tiedje JM. Prokaryotic taxonomy and phylogeny in the genomic era:

advancements and challenges ahead. Curr Opin Microbiol. 2007;10:504–509.
65. Garritano AN, Song W, Thomas T. Carbon fixation pathways across the bacterial and archaeal tree of life. PNAS Nexus 2022;1:pgac226.
66. David SS, O'Shea VL, Kundu S. Base Excision Repair of Oxidative DNA Damage. Nature 2007;447:941–950.
67. 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;42:4590–4605.
68. Wood TL, Wood TK. The HigB/HigA toxin/antitoxin system of *Pseudomonas aeruginosa* influences the virulence factors pyochelin, pyocyanin, and biofilm formation. MicrobiologyOpen 2016;5:499–511.
69. 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;71:1071–1087.
70. 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;13:722–736.
71. 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;163:759–771.
72. 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;18:67–83.
73. Makarova KS, Wolf YI, Snir S, Koonin EV. Defense Islands in Bacterial and Archaeal Genomes and Prediction of Novel Defense Systems. *J Bacteriol* 2011;193:6039–6056.
74. Kavagutti VS, Andrei A-Ş, Mehrshad M, Salcher MM, Ghai R. Phage-centric ecological interactions in aquatic ecosystems revealed through ultra-deep metagenomics. Microbiome 2019;7:1–15.
75. Yuan Y, Gao M. Jumbo Bacteriophages:An Overview. *Front Microbiol* 2017;8.
76. Younker IT, Duffy C. Jumbo Phages. In:Bamford DH, Zuckerman M (eds). *Encyclopedia of Virology (Fourth Edition)*. 2021. Academic Press, Oxford, pp 229–241.
77. 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;7:918–927.
78. 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;119:e2211672119.
79. Zhang Z, Pan S, Liu T, Li Y, Peng N. Cas4 Nucleases Can Effect Specific Integration of CRISPR Spacers. J Bacteriol. 2019;201:e00747-18.
80. Naknaen A, Suttinun O, Surachat K, Khan E, Pomwised R. A Novel Jumbo Phage PhiMa05 Inhibits Harmful *Microcystis* sp. Front Microbiol. 2021;12.
81. 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;4:693–700.
82. Al-Shayeb B, Sachdeva R, Chen L-X, Ward F, Munk P, Devoto A, et al. Clades of huge phages from across Earth's ecosystems. Nature 2020;578:425–431.
83. 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;12:e85183.
84. Messner P, Steiner K, Zarschler K, Schäffer C. S-layer nanoglycobiology of bacteria. Carbohydr Res. 2008;343:1934–1951.
85. Babinski KJ, Ribeiro AA, Raetz CRH. The *Escherichia coli* Gene Encoding the UDP-2,3-diacylglucosamine Pyrophosphatase of Lipid A Biosynthesis. J Biol Chem. 2002;277:25937–25946.
86. Park YS, Sweitzer TD, Dixon JE, Kent C. Expression, purification, and characterization of

CTP:glycerol-3-phosphate cytidylyltransferase from *Bacillus subtilis*. J Biol Chem. 1993;268:16648–16654.
87. 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;10:e02367-22.
88. Barnhart MM, Chapman MR. Curli biogenesis and function. Annu Rev Microbiol. 2006;60:131–147.
89. 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;7:1702–1708.

ACKNOWLEDGEMENTS

FB was supported by a PhD fellowship from the French Ministère de l'Enseignement Supérieur et de la Recherche. DM and PL-G were supported by grants from the European Research Council (ERC Advanced grants 787904 and 101141745, and ERC Starting grant 803151, respectively) and by the ANR project DArchFolds (ANR-22-CE02-0012-02). LJ was supported by grants from CNRS-INSU-EC2CO and AGROPARISTECH. We thank the UNICELL platform (https://www.deemteam.fr/en/unicell) for help with FACS and single cell manipulation.

AUTHOR CONTRIBUTIONS

FB, PL-G, PD, DM, and LJ designed the study. FB, PL-G, AG-P, GD, DM, and LJ collected samples and data at the study sites. FB performed the optical and electron microscopy. FB, MC, PB, and GD conducted flow cytometry and molecular biology experiments. FB and AG-P assembled and annotated the patescibacterial and host genome sequences;MK annotated the jumbo phage genome sequence. FB, PL-G, PD, MK, AG-P, DM, and LJ interpreted the genome annotation results. FB, PL-G, MK, DM, and LJ wrote the first draft of the manuscript. All authors edited the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://xxxxxxxxxxxxxxxxxxx.

Correspondence and requests for materials should be addressed to David Moreira or Ludwig Jardillier.

Figure Legends

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. White arrows indicate epibiotic patescibacterial cells. **E-I** Transmission electron microscopy (TEM) pictures of thin sections of the patescibacterial cells and their hosts. Asterisks indicate extracellular vesicles, CC: cell-to-cell contact, r: ribosomes. The white dashed line in panels G and I highlights the central part of the cell with dense material that probably corresponds to DNA.

Fig. 2 Phylogenetic placement of *Strigamonas methylophilicida* within the Patescibacteria. Maximum likelihood tree based on the concatenation of 67 single-copy protein markers (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.

Fig. 3 Comparison of gene content in characterized Absconditicoccaceae species. A Venn diagram showing the number of genes shared by different well-characterized Absconditicoccaceae species (absolute numbers and % of the total genes in the four Absconditicoccaceae species). **B** Functional annotation (COG categories) of the genes shared by these Absconditicoccaceae species. NA: not annotated.

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 gray 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*.

Fig. 5 The CRISPR-Cas systems of *S. methylophilicida* and CRISPR-based identification of a jumbo phage. **A** CRISPR-Cas gene organization in the *S. methylophilicida* MAGs and two patescibacterial genomes with systems encoding similar Cas12a proteins. Spacers are represented as colored squares separated by direct repeats (black diamonds). ORFs shared by different taxa are colored. **B** Sequence alignment of the direct repeats of *S. methylophilicida* and other Absconditicoccaceae;a sequence logo under the alignments shows the conservation of each position. **C** Sequence alignment of a *S. methylophilicida* CRISPR spacer with a phage genome. **D** Genome organization of the *S. methylophilicida* jumbo phage. The ORFs inferred with genetic codes 25 and 11 are color-coded according to their functions. ORFs highlighted in red are those differing in the two alternative translations.

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.

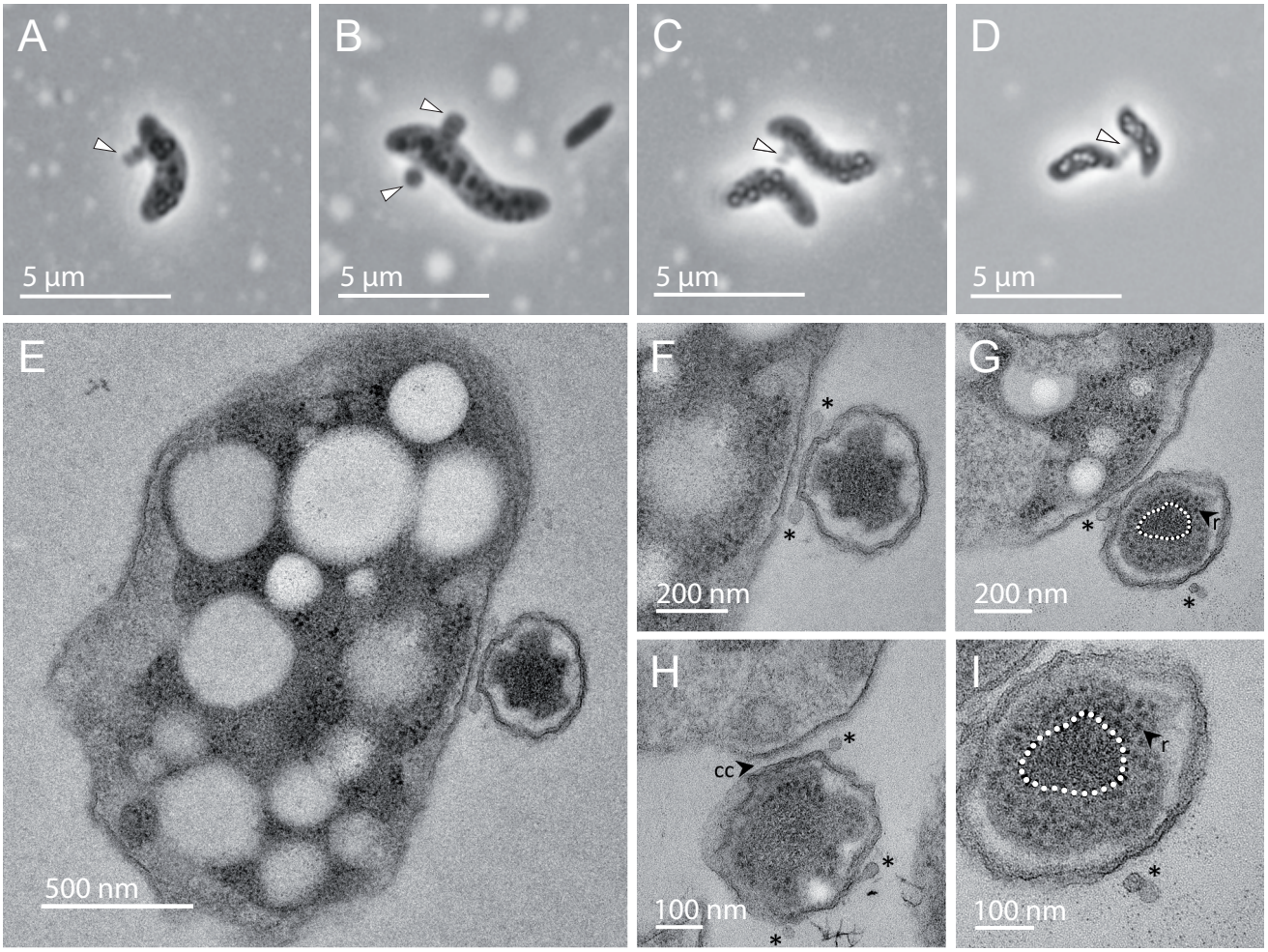

Fig. 1 Microscopy observations of patescibacterial-like cells and their hosts. **A-D** Phase contrast light microscopy pictures, notice the intracellular refringent inclusions in the host cells in C and D. White arrows indicate epibiotic patescibacterial cells. **E-I** Transmission electron microscopy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

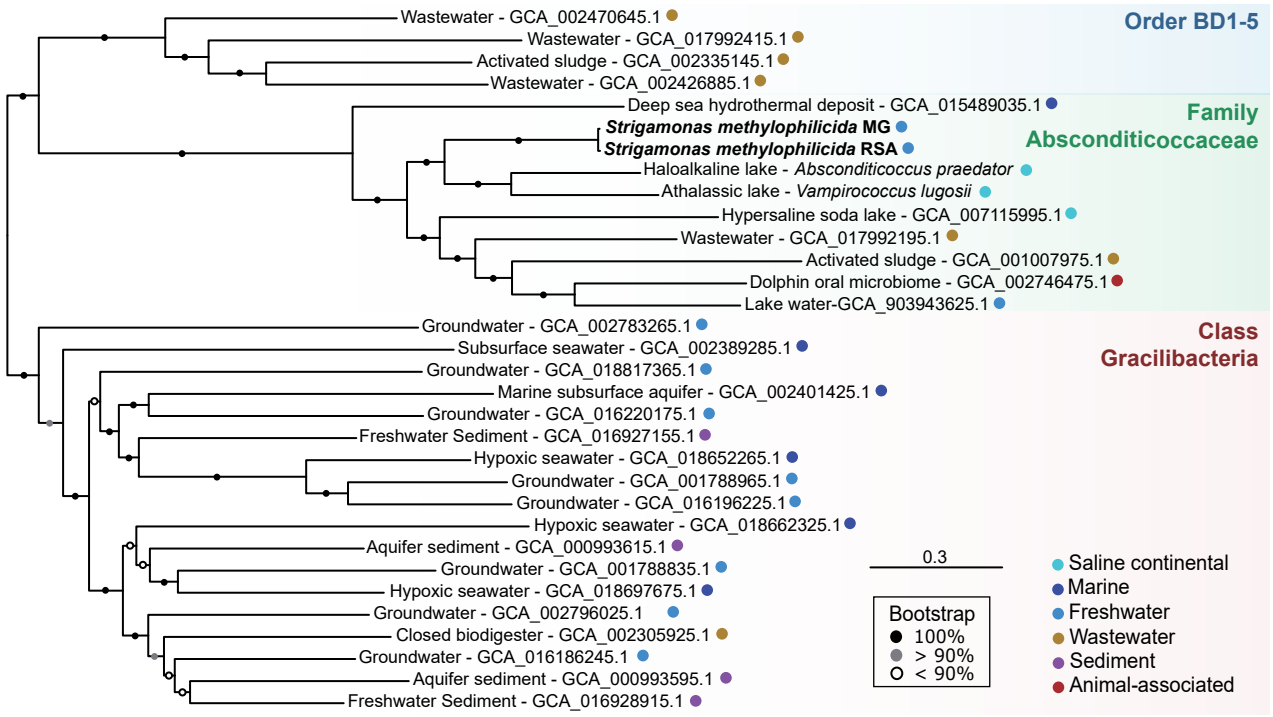

Fig. 2 Phylogenetic placement of *Strigamonas methylophilicida* within the Patescibacteria. Maximum likelihood tree based on the concatenation of 67 single-copy protein markers (20,766 sites). The scale bar indicates the number of substitutions per site. Support values are based on ultrafast bootstraps.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

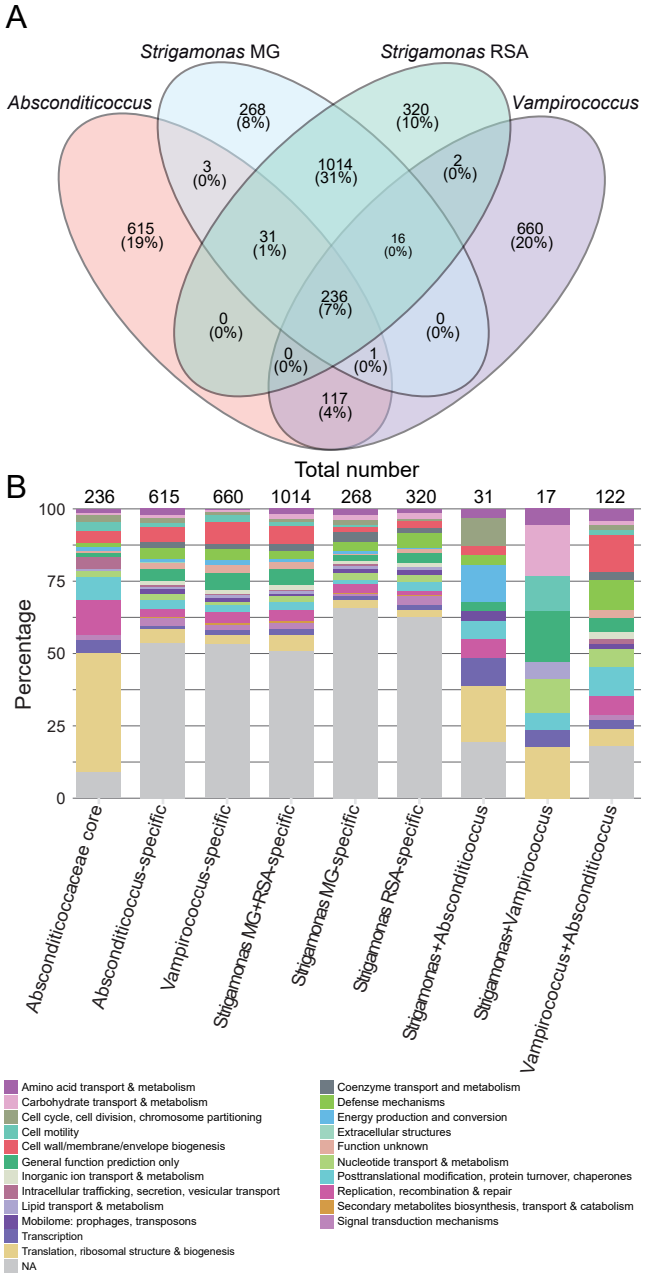

Fig. 3 Comparison of gene content in characterized Absconditicoccaceae species. A Venn diagram showing the number of genes shared by different well-characterized Absconditicoccaceae species (absolute numbers and % of the total genes in the four Absconditicoccaceae species). **B** Functional annotation (COG categories) of the genes shared by these Absconditicoccaceae species. NA: not annotated.

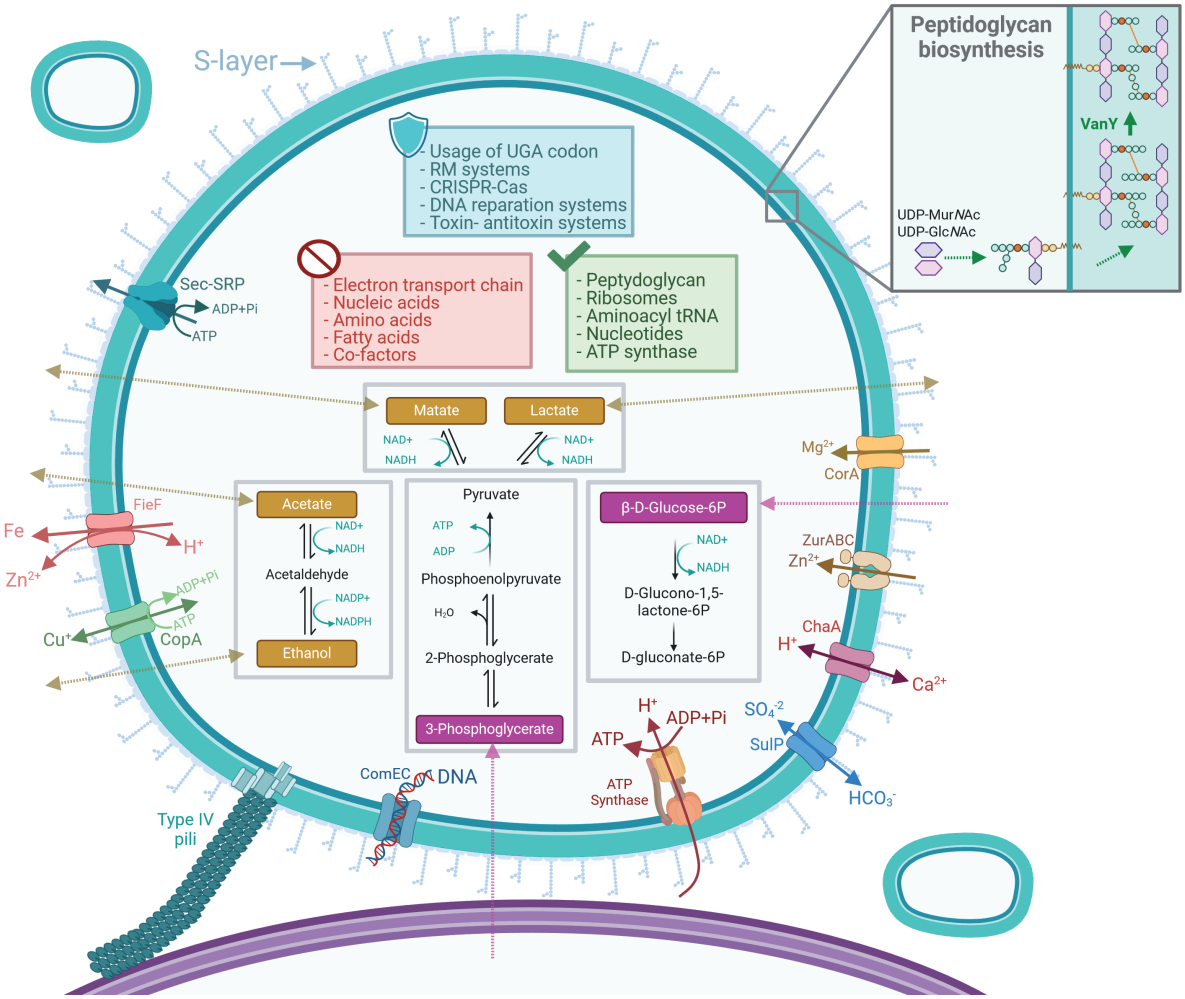

Fig. 4 Reconstruction of metabolic capacities and structural features of *Strigamonas methylophilicida*. The red and the green boxes indicate the incomplete/absent and present pathways, respectively. The blue box indicates the defense mechanisms detected. The gray boxes detail the carbon metabolism,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

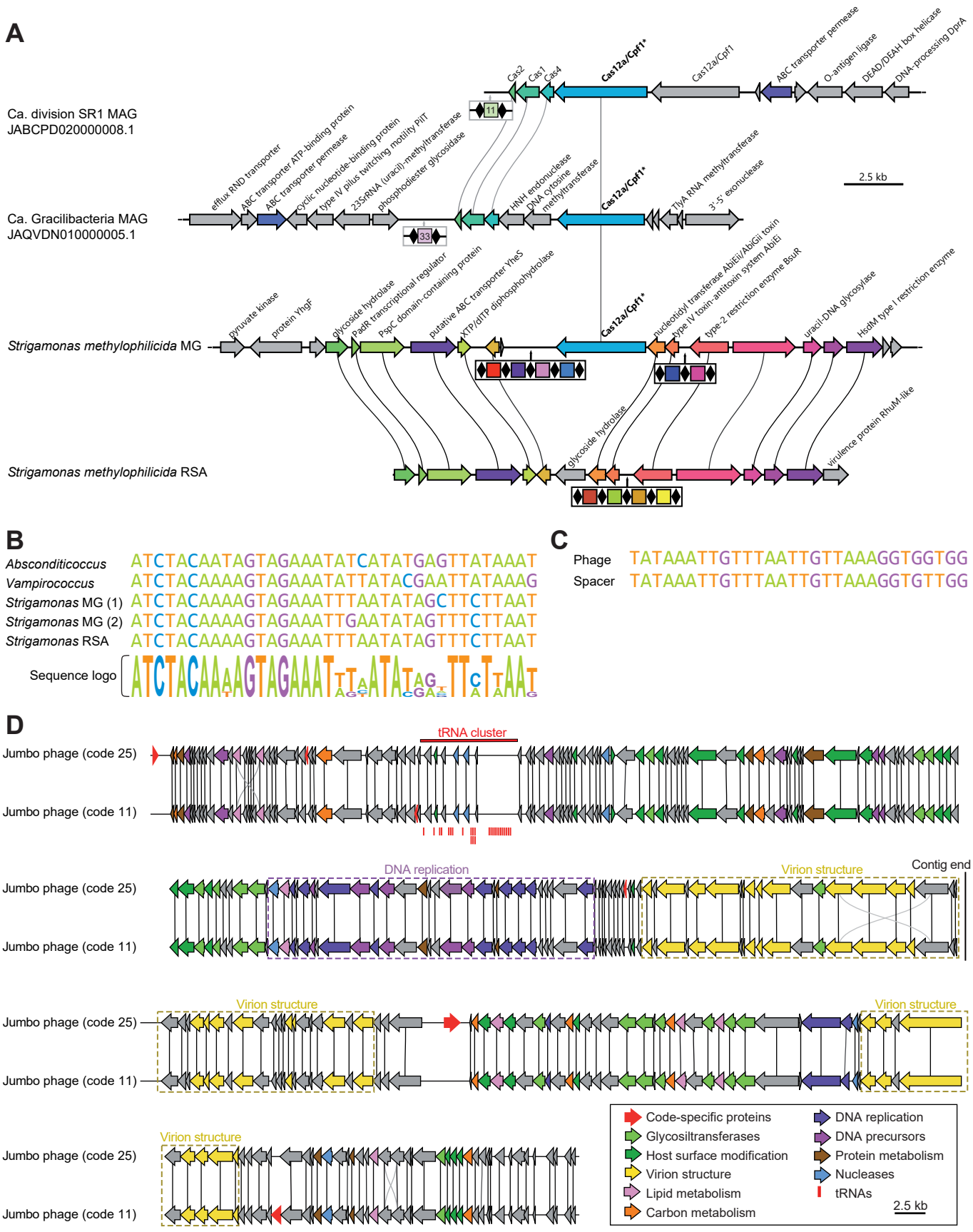

Fig. 5 The CRISPR-Cas systems of *S. methylophilicida* and CRISPR-based identification of a jumbo phage. **A** CRISPR-Cas gene organization in the *S. methylophilicida* MAGs and two patescibacterial genomes with systems encoding similar Cas12a proteins. Spacers are represented as colored squares separated by direct repeats (black diamonds). ORFs shared by different taxa are colored. **B** Sequence alignment of the direct repeats of *S. methylophilicida* and other Absconditicoccaceae;a sequence logo under the alignments shows the conservation of each position. **C** Sequence alignment of a *S. methylophilicida* CRISPR spacer with a phage genome. **D** Genome organization of the *S. methylophilicida* jumbo phage. The ORFs inferred with genetic codes 25 and 11 are color-coded according to their functions. ORFs highlighted in red are those differing in the two alternative translations.

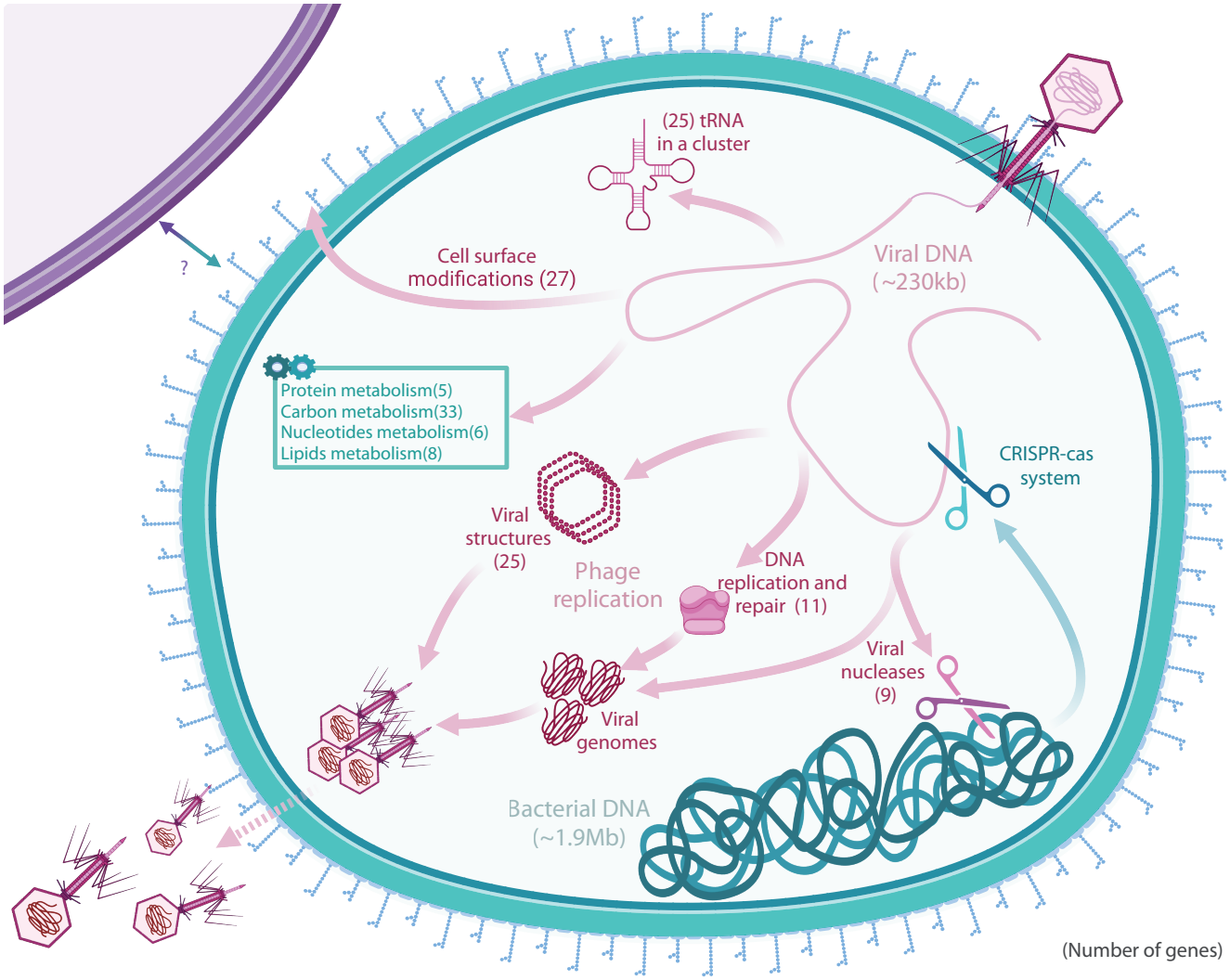

Fig. 6 Genome-based inference of the interaction between *S. methylophilicida* and its jumbo phage. *S. methylophilicida* systems are indicated in blue and those of the phage in pink. Numbers in parentheses indicate the number of genes involved in each system.