

# Lawrence Berkeley National Laboratory

## Joint Genome Institute

### Title

Genome Analysis of a Verrucomicrobial Endosymbiont With a Tiny Genome Discovered in an Antarctic Lake

### Permalink

<https://escholarship.org/uc/item/5gk9q27p>

### Authors

Williams, Timothy J  
Allen, Michelle A  
Ivanova, Natalia  
[et al.](#)

### Publication Date

2021

### DOI

10.3389/fmicb.2021.674758

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed



# Genome Analysis of a Verrucomicrobial Endosymbiont With a Tiny Genome Discovered in an Antarctic Lake

Timothy J. Williams<sup>1</sup>, Michelle A. Allen<sup>1</sup>, Natalia Ivanova<sup>2</sup>, Marcel Huntemann<sup>2</sup>, Sabrina Haque<sup>1</sup>, Alyce M. Hancock<sup>1†</sup>, Sarah Brazendale<sup>1†</sup> and Ricardo Cavicchioli<sup>1\*</sup>

## OPEN ACCESS

### Edited by:

Anne D. Jungblut,  
Natural History Museum,  
United Kingdom

### Reviewed by:

Francisco Rodriguez-Valera,  
Miguel Hernández University of Elche,  
Spain

Stefano Campanaro,  
University of Padua, Italy

### \*Correspondence:

Ricardo Cavicchioli  
r.cavicchioli@unsw.edu.au

### †Present address:

Alyce M. Hancock,  
Institute for Marine and Antarctic  
Studies, University of Tasmania,  
Battery Point, TAS, Australia  
Sarah Brazendale,  
King Island Brewhouse, Pegarah,  
TAS, Australia

### Specialty section:

This article was submitted to  
Extreme Microbiology,  
a section of the journal  
Frontiers in Microbiology

Received: 01 March 2021

Accepted: 23 April 2021

Published: 01 June 2021

### Citation:

Williams TJ, Allen MA, Ivanova N,  
Huntemann M, Haque S,  
Hancock AM, Brazendale S and  
Cavicchioli R (2021) Genome Analysis  
of a Verrucomicrobial Endosymbiont  
With a Tiny Genome Discovered in an  
Antarctic Lake.

Front. Microbiol. 12:674758.  
doi: 10.3389/fmicb.2021.674758

<sup>1</sup> School of Biotechnology and Biomolecular Sciences, UNSW Sydney, Sydney, NSW, Australia, <sup>2</sup> U.S. Department of Energy Joint Genome Institute, Berkeley, CA, United States

Organic Lake in Antarctica is a marine-derived, cold ( $-13^{\circ}\text{C}$ ), stratified (oxic-anoxic), hypersaline ( $>200\text{ g l}^{-1}$ ) system with unusual chemistry (very high levels of dimethylsulfide) that supports the growth of phylogenetically and metabolically diverse microorganisms. Symbionts are not well characterized in Antarctica. However, unicellular eukaryotes are often present in Antarctic lakes and theoretically could harbor endosymbionts. Here, we describe *Candidatus* *Organicella extenuata*, a member of the Verrucomicrobia with a highly reduced genome, recovered as a metagenome-assembled genome with genetic code 4 (UGA-to-Trp recoding) from Organic Lake. It is closely related to *Candidatus* *Pinguicoccus supinus* (163,218 bp, 205 genes), a newly described cytoplasmic endosymbiont of the freshwater ciliate *Euplotes vanleeuwenhoekii* (Serra et al., 2020). At 158,228 bp (encoding 194 genes), the genome of *Ca. Organicella extenuata* is among the smallest known bacterial genomes and similar to the genome of *Ca. Pinguicoccus supinus* (163,218 bp, 205 genes). *Ca. Organicella extenuata* retains a capacity for replication, transcription, translation, and protein-folding while lacking any capacity for the biosynthesis of amino acids or vitamins. Notably, the endosymbiont retains a capacity for fatty acid synthesis (type II) and iron-sulfur (Fe-S) cluster assembly. Metagenomic analysis of 150 new metagenomes from Organic Lake and more than 70 other Antarctic aquatic locations revealed a strong correlation in abundance between *Ca. Organicella extenuata* and a novel ciliate of the genus *Euplotes*. Like *Ca. Pinguicoccus supinus*, we infer that *Ca. Organicella extenuata* is an endosymbiont of *Euplotes* and hypothesize that both *Ca. Organicella extenuata* and *Ca. Pinguicoccus supinus* provide fatty acids and Fe-S clusters to their *Euplotes* host as the foundation of a mutualistic symbiosis. The discovery of *Ca. Organicella extenuata* as possessing genetic code 4 illustrates that in addition to identifying endosymbionts by sequencing known symbiotic communities and searching metagenome data using reference endosymbiont genomes, the potential exists to identify novel endosymbionts by searching for unusual coding parameters.

**Keywords:** Antarctic microbiology, Bacterial endosymbionts, metagenome, extreme genome reduction, genetic code 4

## INTRODUCTION

Bacteria with highly reduced genome sizes are only found as host-restricted symbionts and pathogens (**Supplementary Table 1**; Moran and Bennett, 2014). The smallest bacterial genomes are only known to occur in symbionts that are required by a host (obligate symbionts), with those possessing genomes < 500 kbp being completely dependent on the host while also providing benefit to the host to be retained (mutualistic symbionts) (Moran and Bennett, 2014). Insects that feed on sap (phloem or xylem) rely on endosymbionts to supplement their restrictive or unbalanced diets; these bacteria, either individually or as “patchworks” of metabolically complementary co-symbionts or nested symbionts, provide essential amino acids and/or vitamins for their respective protist hosts (Nakabachi and Ishikawa, 1999; Zientz et al., 2004; Nakabachi et al., 2006; Pérez-Brocail et al., 2006; Bennett and Moran, 2013; Brown et al., 2015; Gil et al., 2018). The cellulolytic protists that reside in the hindguts of termites harbor cytoplasmic endosymbionts that belong to diverse bacterial clades (e.g., Endomicrobia, Deltaproteobacteria, Bacteroidetes, and Actinobacteria) and confer metabolic and nutritional benefits to their respective protist hosts (Stingl et al., 2005; Ohkuma et al., 2007; Hongoh et al., 2008a,b; Sato et al., 2009; Ikeda-Ohtsubo et al., 2016; Strassert et al., 2016; Kuwahara et al., 2017).

Verrucomicrobia is a diverse phylum of bacteria that has been found in a wide array of habitats, with free-living representatives isolated from soils, seawater, marine sediments, lakes, and hot springs (Wagner and Horn, 2006; Dunfield et al., 2007; Yoon et al., 2007a,b). Certain verrucomicrobia live in close association with eukaryotes, including marine sponges (Scheuermayer et al., 2006; Yoon et al., 2008) and tunicates (Lopera et al., 2017), as well as inside the intestinal tracts of humans (Derrien et al., 2004), termites (Wertz et al., 2012), and marine clam worms (Choo et al., 2007). Some verrucomicrobia have entered into very close symbiotic associations with eukaryotic hosts, including anti-predator ectosymbionts (epixenosomes) of the ciliate *Euplotidium* (Petroni et al., 2000) and various endosymbionts, such as inside the cytoplasm of intestinal and ovarian cells of nematode worms (Vandekerckhove et al., 2002), nuclei of cellulolytic protists (Sato et al., 2014), and the cytoplasm of the ciliate *Euplotes vanleeuwenhoekii* (Serra et al., 2020). *Candidatus* Xiphinematobacter, the verrucomicrobial endosymbiont of nematodes, has a 0.916-Mbp metagenome assembled genome (MAG) encoding 817 predicted protein-coding sequences (CDS); compared with free-living relatives, genes are retained for the biosynthesis of amino acids predicted to be required by their nematode hosts (Brown et al., 2015). The unpublished MAG of the intranuclear endosymbiont *Ca. Nucleococcus* (Sato et al., 2014) is ~1 Mbp and encodes ~700 CDS (Y. Hongoh, personal communication). The genome of the *Euplotes* endosymbiont *Ca. Pinguicoccus supinus* has an ‘extremely reduced genome’ at only 0.163 Mbp and encodes 168 CDS (Serra et al., 2020).

Organic Lake is a shallow (~7 m deep), marine-derived, Antarctic lake formed ~3,000 years ago (Gibson, 1999). The lake is characterized by a salinity gradient that reaches a maximum of

~230 g L<sup>-1</sup> (Gibson, 1999) and has unusual chemistry, with very high levels of dimethylsulfide (Gibson et al., 1991). Temperatures in the upper waters have been recorded as high as 15°C and as low as -14°C (Franzmann et al., 1987), whereas bottom waters (5 - ~7 m) have typically registered temperatures of -5 to -6°C (Franzmann et al., 1987; Gibson et al., 1991; Roberts et al., 1993; James et al., 1994) but as low as -13°C (Yau et al., 2013). Metaproteogenomic analyses have inferred important roles for viroplage-mediated control of algal primary production (Yau et al., 2011) and roles in nutrient cycling by phylogenetically and metabolically diverse bacteria (Yau et al., 2013). The lake is located in the Vestfold Hills, a ~ 400 km<sup>2</sup> region of East Antarctica that contains hundreds of water bodies, many of which are marine-derived, having been formed ~3,000–7,000 years ago as a result of the isostatic rebound of the continent (Gibson, 1999; Cavicchioli, 2015; **Supplementary Figure 1**). The water bodies in the Vestfold Hills range in salinity from freshwater to hypersaline, most of which have not been subject to metagenomic analysis of their biota (Cavicchioli, 2015).

During analyses of unusual coding parameters (genetic code 4) in metagenome contigs, we discovered a 158-kbp verrucomicrobial MAG that was assembled from new metagenome data derived from a complete seasonal cycle of Organic Lake. The MAG is comparable in size with the endosymbiont *Ca. Pinguicoccus supinus* (Serra et al., 2020), as well as to obligate mutualistic endosymbionts that belong to the phyla Proteobacteria and Bacteroidetes that also have extremely reduced genomes (Moran and Bennett, 2014). The environmental distribution and inferred *Euplotes* host of the Organic Lake endosymbiont was assessed by analyzing 337 Antarctic metagenomes, including 150 new metagenomes of unstudied Vestfold Hills lakes and neighboring marine locations. The Organic Lake endosymbiont is closely related to *Ca. Pinguicoccus*; herein, we describe the functional traits of this bacterial lineage that seem to underpin the endosymbiosis and discuss the value of searching for unusual coding parameters as a means of identifying endosymbionts.

## MATERIALS AND METHODS

### Sampling and DNA Extraction

Microbial biomass was obtained and field observations recorded from lakes in the Vestfold Hills, Antarctica (**Supplementary Figure 1**). Sampling at Organic Lake was performed by sequential size fractionation through a 20-μm prefilter onto 3.0-, 0.8-, and 0.1-μm large-format membrane-filters (293 mm diameter polyethersulfone), samples preserved and DNA extracted, as described previously (Yau et al., 2011, 2013; Tschitschko et al., 2018; Panwar et al., 2020).

For other lakes, including Unnamed Lake 18, “Portals” Lake, Unnamed Lake 13, Unnamed Lake 17, “Swamp” Lake, Unnamed Lake 12, and Unnamed Lake 7, biomass was captured using Sterivex cartridges (MilliporeSigma, Burlington, MA, United States) by pumping water from the lake through a 20-μm prefilter using a hand-driven peristaltic pump. After field collection, Sterivex cartridges were kept cold (e.g., in snow)

before transportation to Davis Research Station, where they were cryogenically preserved at  $-80^{\circ}\text{C}$  and shipped at  $-80^{\circ}\text{C}$  to Australia. To extract DNA, the Sterivex cartridge was removed from  $-80^{\circ}\text{C}$  storage and filled with 1.6 ml of freshly prepared “XS” buffer (1% potassium ethyl xanthogenate; 100-mM Tris-hydrochloride, pH 7.4; 20-mM ethylenediamine tetraacetic acid, pH 8; 1% sodium dodecyl sulfate; 800-mM ammonium acetate) (Tillett and Neilan, 2000). Both ends of the cartridge were sealed with parafilm, and the cartridge was placed into an empty 50-ml Falcon tube and incubated in a water bath at  $65^{\circ}\text{C}$  for 2 h. After incubation, 200  $\mu\text{l}$  of 10% sodium dodecyl sulfate and 50  $\mu\text{l}$  of 20  $\text{mg ml}^{-1}$  Proteinase K (Thermo Fisher Scientific, Waltham, MA, United States) was added through the Luer-lock end of the cartridge, re-sealed, and returned to the 50-ml Falcon tube for incubation in a water bath at  $55^{\circ}\text{C}$  for 2 h. After incubation, a syringe was attached to the Luer-lock end and air injected to recover the liquid in a 20-ml Falcon tube. The liquid was decanted, placing 500- $\mu\text{l}$  aliquots into 1.5-ml microfuge tubes, 60  $\mu\text{l}$  of phenol added, the tubes inverted several times to mix the solution, 500  $\mu\text{l}$  of chloroform: isoamyl alcohol (24:1) was added, and each tube mixed by inversion. The tubes were centrifuged at  $16,800 \times g$  for 10 min at room temperature, the aqueous phase of each sample was collected into a fresh 1.5-ml tube, 1.5  $\mu\text{l}$  of GlycoBlue (Thermo Fisher Scientific) was added to each tube, and tubes were left at room temperature for 1 h. Ammonium acetate (3 M, 500  $\mu\text{l}$ ) was added to each tube, mixed by inversion, left at room temperature for 30 min, tubes centrifuged at  $16,800 \times g$  for 15 min, and the supernatant placed into fresh 2-ml tubes. A total of 1 ml of 100% ethanol was added to each tube, and after storage overnight at  $4^{\circ}\text{C}$ , tubes were centrifuged at  $14,000 \times g$  for 30 min at room temperature and the supernatant carefully discarded. Pellets were washed by adding 500  $\mu\text{l}$  of 70% ethanol and tubes centrifuged at  $14,000 \times g$  for 5 min. Ethanol was removed, the pellets air-dried on a heating block at  $37^{\circ}\text{C}$ , pellets resuspended in Tris-hydrochloride-ethylenediaminetetraacetic acid buffer (10-mM Tris-hydrochloride, pH 7.4; 1-mM ethylenediamine tetraacetic acid, pH 8) and tubes stored at  $-80^{\circ}\text{C}$ . DNA yields were quantified using Qubit dsDNA BR Assay Kit (Thermo Fisher Scientific) and the quality of DNA assessed by agarose gel electrophoresis. DNA was sequenced at the Joint Genome Institute using Hi-Seq2500 ( $2 \times 151$  bp run) as described previously (Tschitschko et al., 2018; Panwar et al., 2020) or at the Australian Centre for Ecogenomics using NextSeq500 (on a  $2 \times 150$  bp run) and raw reads filtered using Trimmomatic (Trimmomatic manual: V0.32, no date). Assembly was performed with metaSpades and all contigs  $> 200$  bp uploaded and annotated by the IMG pipeline (Huntemann et al., 2015).

## Analyses of DNA Sequence Data

The *Ca. Organicella* MAG was identified using a pipeline to identify stop codon reassignments in metagenomic data (Ivanova et al., 2014). The set of contigs with potential UGA reassignment was identified in Organic Lake metagenomes based on the higher total coding potential as computed by Prodigal upon reannotation with genetic code 4. These contigs had an average GC content of 32%, and they appeared to have characteristics of

bacterial genomes, namely, high coding density, typical bacterial gene complement with translation, transcription, and replication machinery, but no multi-subunit NADH dehydrogenase and cytochrome oxidase complexes indicative of mitochondria and no photosynthesis genes indicative of chloroplasts. The longest of these contigs, which were  $\sim 158$  kb, turned out to be circular due to an overlap of 100 nt at the ends. No other putative bacterial contigs with UGA reassignment were found in the same metagenomes, suggesting that these circular contigs constituted the entire genome of a bacterium. Because the automated annotations initially performed by IMG used genetic code 11, in which UGA is a stop codon, manual inspection of these contigs identified genes interrupted by stop codons within open reading frames. Re-calling open reading frames and annotating the genome using PROKKA (Seemann, 2014) with codon chart 4, reassigned the opal stop codon (UGA) as tryptophan. This reduced the number of genes from 249 to 193 for the reference *Ca. Organicella extenuata* MAG (contigID Ga0307966\_1000010). Total coding density was calculated using all protein-coding genes (CDS), rRNA, tRNA, and tmRNA genes in the genome. Protein identities were determined using ExPASy BLAST for all CDS and, where necessary, InterProScan and HHPred. The isoelectric point (pI) of protein sequences was determined using the Isoelectric Point Calculator (Kozłowski, 2016). The genomic functional potential was assessed by considering cellular and metabolic traits based upon manual examination of genes and pathways performed in a similar way to previous assessments of the veracity of gene functional assignments (Allen et al., 2009; Panwar et al., 2020).

Mapping of reads from 340 Antarctic metagenomes to the *Ca. Organicella* MAG was performed using BWA v0.7.17 (Li and Durbin, 2009). FastANI (Jain et al., 2018) was used to calculate ANI between *Ca. Organicella* MAGs. Multiple alignments were constructed using Clustal (DNA sequences) (Thompson et al., 1994) or MUSCLE (protein sequences) (Edgar, 2004) and used to construct phylogenetic trees (for *Ca. Organicella* and for *Euplotes* sp. AntOrgLke) by the maximum-likelihood method (Tamura and Nei, 1993) in MEGA6 (Tamura et al., 2013) with 1,000 bootstraps. Marker genes predicted from the *Ca. Organicella* MAG were used to place the MAG into a concatenated 43-marker gene tree by CheckM (Parks et al., 2014) using the tree command.

To identify the potential host(s) of *Ca. Organicella*, six metagenomes where *Ca. Organicella* was abundant (Org-646, Org-46, Org-175, Org-784, Portals, and UnnamedLake18) were selected to create a co-assembly using Megahit v.1.2.2b (Li et al., 2016), with contigs binned into 188 MAGs by Metabat v.2.12.1 (Kang et al., 2019) with default settings (min contig length 2,500). During the co-assembly, a single contig representing the *Ca. Organicella* MAG was assembled (k141\_311079; 158,131 bp). As this *Ca. Organicella* MAG was not initially binned, due to falling below the default minimum bin size of 200 kb, it was manually assigned to bin189. The ANI between the original *Ca. Organicella* MAG (Ga0307966\_1000010) and the co-assembled MAG (k141\_311079) was 99.9924%. The *Ca. Organicella* MAG (bin189) and the 188 bins resulting from Metabat binning were screened for contamination, completion, and taxonomic identity using checkM (Parks et al., 2014) and refineM (Parks et al.,

2017). In addition, the abundance of each bin [calculated as the sum of (contig length  $\times$  contig coverage) for all contigs in the bin] was determined for each of the 29 metagenomes where *Ca. Organicella* was detected by mapping the metagenome reads to the bins with bmap v38.51 (Bushnell, 2014). These bin abundances were used as input for SparCC (Friedman and Alm, 2012) implemented in python 3<sup>1</sup> to estimate correlation values from the compositional data.

To identify the taxonomy of the bins that were highly correlated to *Ca. Organicella*, MetaEuk v. 20200908 (Levy Karin et al., 2020) was used to identify eukaryotic proteins and assign taxonomy *via* the 2bLCA lowest common ancestor approach. To identify the maximal number of proteins, the larger MERC\_MMETSP\_Uniclust50\_profiles database was used as the reference dataset for MetaEuk, whereas to assign contig taxonomy and putative protein function, the TaxDB\_uniclust90\_2018\_08 database was used. Both databases were obtained from [http://wwwuser.gwdg.de/\\$sim\\$compbiol/metaeuk/](http://wwwuser.gwdg.de/$sim$compbiol/metaeuk/). The rRNA gene contig missing from the *Euplotes* sp. AntOrgLke MAG was identified as contig k141\_859071 by blasting *Euplotes* spp. 18S rRNA genes against the co-assembled contigs used for Metabat. 18S/28S rRNA genes on k141\_859071 were predicted using the RNAMmer 1.2 Server at <http://www.cbs.dtu.dk/services/RNAMmer/> (Lagesen et al., 2007). The mitochondrial genome of *Euplotes* sp. AntOrgLke was identified by blasting the mitochondrial proteins of *E. vanleeuwenhoekii*, *Euplotes crassus*, and *Euplotes minuta* against the co-assembled contigs used for the Metabat binning, with the resulting contigs then blasted against the metagenome where *Ca. Organicella* was most abundant.

As few non-mitochondrial *Euplotes* proteins were available in the National Center for Biotechnology Information (NCBI) nr database, additional protein sequences were gathered from five reference *Euplotes* species. The data were obtained from genome-specific databases: *Euplotes octocarinatus*, <http://ciliates.ihb.ac.cn/database/home/#eo> (Wang et al., 2018); *Euplotes vannus*, <http://evan.ciliate.org/> (Chen et al., 2019); from proteins predicted from the Marine Microbial Eukaryote Meta/transcriptome sequencing project (MMETSP): *Euplotes harpa* FSP1.4, IMG ID 3300017294; *Euplotes focardii* TN1, IMG IDs 3300017169 and 3300016941; *E. crassus* CT5, IMG ID 3300017039; all accessed at <https://img.jgi.doe.gov/>; by manually running MetaEuk for protein prediction on genome sequences held in NCBI Genome: *E. focardii*, GCA\_001880345.1 ASM188034v1; *E. crassus* GCA\_001880385.1 ASM188038v1; or by manually running MetaEuk on genomic data held in a custom database: *E. vannus*, <http://evan.ciliate.org/>. Average amino acid identity (AAI) was calculated at <http://enve-omics.ce.gatech.edu/aai/index> (Rodriguez-R and Konstantinidis, 2016) between the *Euplotes* sp. AntOrgLke and the five reference *Euplotes* species, using the protein sequences downloaded or predicted for their respective genomes (*Euplotes* sp. AntOrgLke – 15328 proteins predicted in this study; *E. octocarinatus* – 29076 proteins obtained from <http://ciliates.ihb.ac.cn/database/home/#eo>; *E. focardii* TN1, *E. crassus* CT5, and *E. harpa* FSP1.4 –

12634, 12729, and 19386 proteins, respectively, predicted from the Marine Microbial Eukaryote Meta/transcriptome sequencing project (MMETSP) and accessed at <http://img.jgi.doe.gov/>; *E. vannus* – 43338 proteins obtained from <http://evan.ciliate.org/>).

To correctly identify the CDS in the *Ca. Pinguicoccus* genome, Prokka (Seemann, 2014) was used with genetic code 4. To investigate the relationship between *Ca. Organicella* and the newly released *Ca. Pinguicoccus* genome, nucleotide synteny was visualized with progressiveMauve (Darling et al., 2004), a tblastx plot was performed at NCBI<sup>2</sup>, and AAI was calculated at <http://enve-omics.ce.gatech.edu/aai/index> (Rodriguez-R and Konstantinidis, 2016), followed by manual inspection of protein identifications to identify shared and unique metabolic capacities.

## RESULTS AND DISCUSSION

### Organic Lake MAG Summary and Phylogeny

A MAG (Ga0307966\_1000010) representing a complete circular genome with a length of 158,228 bp was identified in new metagenome data from Organic Lake. The MAG encoded 194 bacterial genes, 156 of which were inferred to be CDS (Supplementary Table 2) with 145 assigned putative biological functions (Supplementary Table 3). Most (76 proteins) were assigned to translation (including tRNA modifications) (Supplementary Table 3). Other categories were fatty acid synthesis (including pyruvate oxidation) (18 proteins); cell wall biogenesis including lipopolysaccharides (17), iron–sulfur (Fe-S) cluster assembly (8), protein folding and stability (8), replication and repair (6), and transcription (6). A total of 16 CDS could not be assigned any function, and some or all of these could be pseudogenes. The MAG had one copy each of 23S, 16S, and 5S rRNA genes and 34 identifiable tRNA genes (Supplementary Table 2). The highly restricted genomic potential illustrates this bacterium would not be capable of autonomous growth, and we name it *Candidatus Organicella extenuata* gen. et. sp. nov.; the genus name derives from the locality from where the MAG sequence was originally recovered (Organic Lake, Antarctica) with the addition of the diminutive Latin suffix -ella; the species “extenuata” means reduced or diminished in Latin and is in reference to the highly reduced genome.

Additional MAGs for *Ca. Organicella* were generated from a number of Antarctic metagenomes (see section *Ca. Organicella Environmental Distribution and Host* below), enabling the analysis of 23 *Ca. Organicella* 16S rRNA genes (Supplementary Table 4). Phylogenetic analysis of these genes found *Ca. Organicella* to be most closely related to *Ca. Pinguicoccus* (Serra et al., 2020), with 85% 16S rRNA gene identity (see section *Comparison of Ca. Organicella and Ca. Pinguicoccus Genomes* below). Both *Ca. Organicella* and *Ca. Pinguicoccus* belong to a cluster of uncultured Verrucomicrobia that also includes *Ca. Nucleococcus* and related endosymbionts of certain amitochondriate protists (*Trichonympha*, *Caduceia*, and *Oxymonas*) present in termite hindguts (Yang et al., 2005;

<sup>1</sup><https://github.com/JCSzamosi/SparCC3>

<sup>2</sup><https://blast.ncbi.nlm.nih.gov/>

Hongoh et al., 2007; Ikeda-Ohtsubo et al., 2010; Sato et al., 2014; **Figure 1**). This cluster, previously termed the “termite cluster” (Sato et al., 2014), is not closely related to other known verrucomicrobial endosymbionts (Vandekerckhove et al., 2002) or ectosymbionts (Petroni et al., 2000). In view of the cluster now including *Ca. Organicella* and *Ca. Pinguicoccus*, and no longer containing species exclusive to the termite gut, we suggest the cluster be termed the “Nucleococcus cluster.” To date, known representatives of this “Nucleococcus cluster” of Verrucomicrobia include both nuclear and cytoplasmic endosymbionts of unicellular eukaryotes.

The *Ca. Organicella* + *Ca. Pinguicoccus* branch within the “Nucleococcus cluster” of the 16S rRNA gene tree was far longer than other branches (**Figure 1**), and similar topology occurred in trees constructed using conserved marker genes (**Supplementary Figure 2**). Such long branches were not evident for any other sequences, including the endosymbionts *Ca. Nucleococcus* and *Ca. Xiphinematobacter* (**Figure 1**), both of which had much larger genomes (~ 1 Mbp) than *Ca. Organicella* and *Ca. Pinguicoccus* (**Supplementary Table 1**). Long branches likely reflect rapid sequence evolution and are characteristic of degenerate genomes (McCutcheon and Moran, 2012), consistent with *Ca. Organicella* and *Ca. Pinguicoccus* being the only known representatives of Verrucomicrobia with extremely reduced genomes.

## Endosymbiont Features

The *Ca. Organicella* MAG exhibits a number of features typical of obligate symbionts that have highly reduced genomes (McCutcheon and Moran, 2010, 2012). The MAG has a high coding density (95% for all genes and 90% for CDS only), with shortened intergenic regions and 23 overlapping genes (**Supplementary Table 1**), which is characteristic of extreme genome reduction (Nakabachi et al., 2006; Moya et al., 2008). The MAG has genetic code 4<sup>3</sup> with UGA stop codons recoded to tryptophan. Of note is that a tRNA-Opal-TCA is also encoded (Ga0307966\_1000010189) that has the highest similarity to trnW (UGA) from mitochondria of *Paralemanea* sp. (GenBank accession MG787097.1). UGA-to-Trp recoding is known to occur rarely, having been found in mycoplasmas (Yamao et al., 1985); certain symbiotic bacteria (McCutcheon et al., 2009), including *Ca. Pinguicoccus* (Serra et al., 2020); and several mitochondrial lineages (Knight et al., 2001). The UGA-to-Trp conversion permits the loss of peptide chain release factor 2 (PrfB) (which recognizes UGA codons) through genome erosion (McCutcheon et al., 2009). UGA-to-Trp recoding is typically associated with low GC content (McCutcheon et al., 2009), although some insect endosymbionts with UGA-to-Trp have high GC content (e.g., *Ca. Hodgkinia cicadicola* 58%; *Ca. Tremblaya princeps* PCIT 59%) (McCutcheon and Moran, 2012). The GC content of the *Ca. Organicella* MAG is 32%, compared with 25% for *Ca. Pinguicoccus* (Serra et al., 2020) (also see *Comparison of Ca. Organicella and Ca. Pinguicoccus Genomes*, later). No mobile elements were identified in the *Ca. Organicella* MAG, which

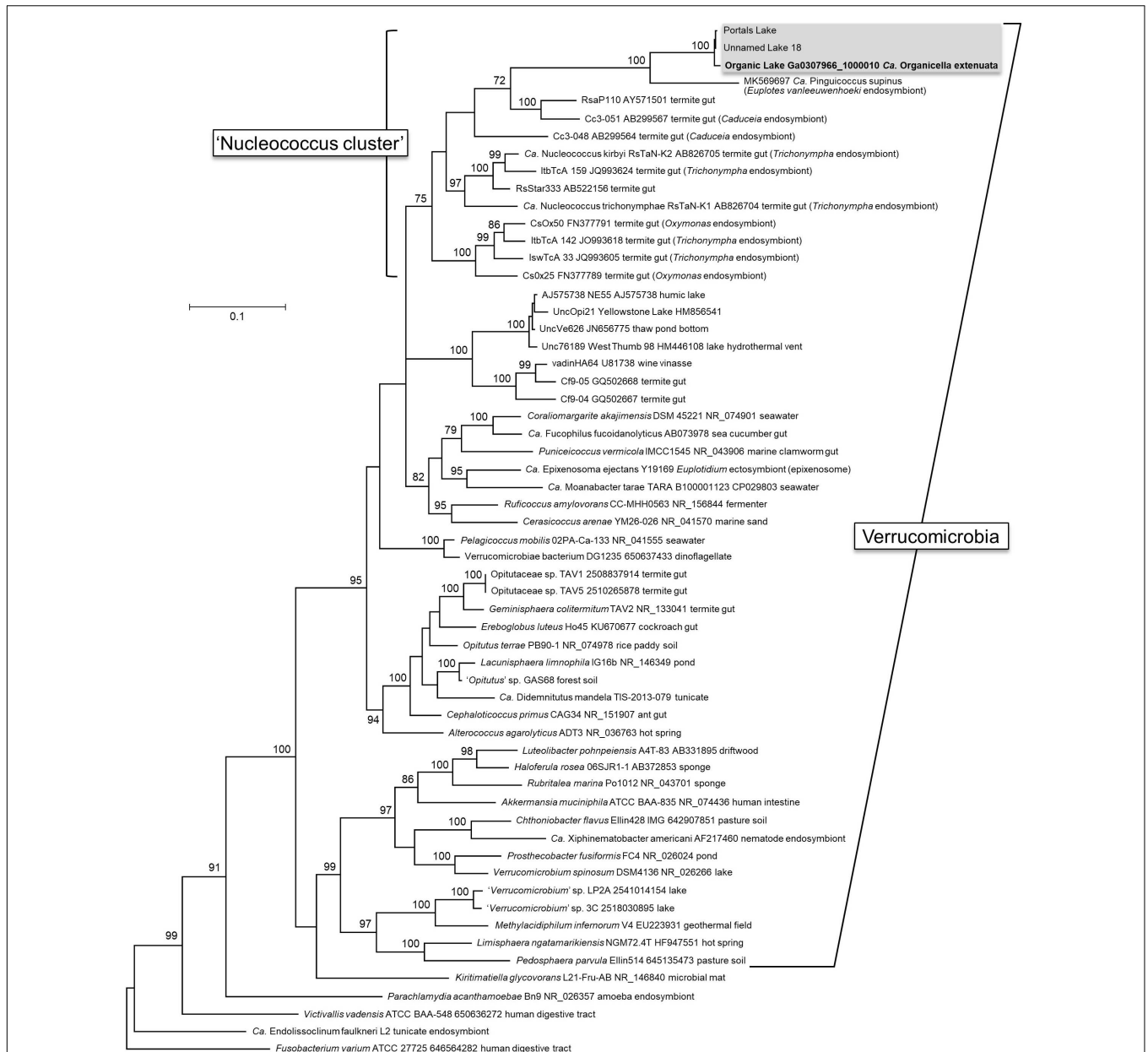
is another trait of symbionts with extremely reduced genomes (McCutcheon and Moran, 2012).

The possession of a minimal complement of genes required for transcription and translation (McCutcheon, 2010; McCutcheon and Moran, 2012), and some capacity to perform DNA replication, enables a level of autonomy over cellular processes that distinguishes endosymbiotic bacteria from organelles (McCutcheon and Moran, 2012). *Ca. Organicella* encodes some enzymes involved in DNA replication, including DNA gyrase (GyrAB), DNA primase (DnaG), and replicative DNA helicase (DnaB), but a dedicated DNA polymerase for DNA replication was not identifiable. Although certain insect endosymbionts lack the DNA polymerase III holoenzyme, they at least encode DNA polymerase  $\alpha$ -subunit (DnaE), responsible for 5' to 3' polymerization activity of DNA replication (McCutcheon, 2010; McCutcheon and Moran, 2012). In the absence of DnaE, genomic replication is presumably carried out by host proteins (Serra et al., 2020). As in many other reduced endosymbiont genomes, *Ca. Organicella* lacks the DnaA protein for initiation of DNA replication, and this function is presumably carried out by the host (Gil et al., 2003; López-Madrugal et al., 2013), possibly as a mechanism to exercise control over endosymbiont proliferation (e.g., Akman et al., 2002; Gil et al., 2003; Bennett et al., 2014; Bennett and Moran, 2015).

Three subunits of the DNA-directed RNA polymerase (RNAP) for transcription were identified (RpoA, RpoB, and RpoC) as well as a sigma factor (RpoD), components that are typical of endosymbionts (McCutcheon and Moran, 2012). Thus, the components of RNAP retained by *Ca. Organicella* parallel those of unrelated symbionts with genomes of comparable size (McCutcheon, 2010; McCutcheon and Moran, 2012). A total of 34 amino acyl tRNAs for all 20 proteinogenic amino acids were identified, plus aminoacyl tRNA synthetases (aaRS) for 13 of the amino acids (Met, Leu, Ile, Val, Lys, Gly, Ser, Cys, Arg, Tyr, Ala, Phe, and Glu) and a glutamyl/aspartyl-tRNA amidotransferase. The missing aaRS may be provided by the host (Van Leuven et al., 2019), or existing aaRS may catalyze multiple aminoacylation reactions (Moran and Bennett, 2014). *Ca. Organicella* encodes initiation factors IF-1 and IF-2 (but not IF-3); elongation factors EF-G, EF-Ts, and EF-4; translational release factor PrfA (but not PrfB); and ribosome recycling factor. Most, but not all, ribosomal subunits were identified. Known endosymbionts with highly reduced genomes typically do not encode a complete set of ribosomal proteins (McCutcheon, 2010; Moran and Bennett, 2014). Individual ribosomal subunits that could not be identified in the *Ca. Organicella* MAG are also missing from some obligate insect endosymbionts (e.g., RplA, RpmC, RpmD, RpsF, and RpmF) (Moran and Bennett, 2014). Certain tRNA modification enzymes were also evident in the *Ca. Organicella* MAG (e.g., Mnm complex and TsaD) that are usually retained in endosymbionts (McCutcheon and Moran, 2012; Van Leuven et al., 2019) (see **Supplementary Text – tRNA Modification**).

The only identifiable dedicated DNA repair enzyme in *Ca. Organicella* was a RecA homolog. Depleted DNA repair abilities are typical of bacteria with highly reduced genomes and contribute to the accumulation of deleterious substitutions,

<sup>3</sup><https://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi#SG4>



**FIGURE 1 |** Phylogeny of *Candidatus Organicella extenuata*. Phylogeny of Verrucomicrobia and related bacteria based on 16S rRNA sequences, showing *Ca. Organicella extenuata* nested inside the newly proposed "Nucleococcus cluster"; other than *Ca. Organicella extenuata*, this cluster comprises the cytoplasmic endosymbiont *Ca. Pinguicoccus supinus* from a freshwater ciliate and intranuclear endosymbionts of amitochondriate protists resident in the hindgut of termites. The maximum likelihood tree was constructed with 59 sequences, and positions with less than 80% site coverage were eliminated, resulting in 1,415 positions in the final dataset. Bootstrap values > 70 are shown next to individual nodes. *Fusobacterium varium* is the outgroup. Accessions are given as NCBI Nucleotide accessions or IMG Gene IDs: for *Ca. Organicella extenuata*, sequences were included for the original Organic Lake MAG (contig Ga0307966\_1000010, bases 107297..108828), Unnamed Lake 18 (contig Ga0400283\_000007, bases 52431..53966), and "Portals" Lake (contig Ga0400669\_009478, bases 1..1071 and contig Ga0400669\_039189, bases 1314..1821). Sequences identical to the 16S rRNA sequence from the original Organic Lake MAG were represented in metagenome data from 19 other Organic Lake metagenomes and also in Unnamed Lake 13 (**Supplementary Table 4**). Note that nine-digit accessions are IMG Gene IDs, and all others are NCBI Nucleotide accessions.

including in CDS (McCutcheon and Moran, 2012; Bennett and Moran, 2015). The average predicted pI of *Ca. Organicella* proteins was 9.2 (**Supplementary Table 3**). It has been proposed that high (alkaline) pI of the proteome of intracellular parasites and endosymbionts may result from the accumulation of

mutations (Kiraga et al., 2007). However, not all *Ca. Organicella* proteins were predicted to have a high pI. Notably, the two most acidic proteins are ferredoxin (pI 4.1) and acyl carrier protein (ACP) (pI 4.2), both of which are naturally acidic proteins (Knaff and Hirasawa, 1991; McAllister et al., 2006). If high pI

does arise from high rates of mutation, the acidic pI of ferredoxin and ACP may be indicative of a strong positive selection to preserve function.

Another trait that is shared between *Ca. Organicella* and known bacterial symbionts with highly reduced genomes is the retention of chaperone proteins (GroES-GroEL; DnaK); these chaperone proteins are thought to ameliorate the adverse effects of accumulated deleterious substitutions on correct protein-folding (Moran, 1996; McCutcheon and Moran, 2012). The bacteria that synthesize these chaperones are therefore heat-sensitive, limiting the thermal tolerance of their hosts (Burke et al., 2010; Fan and Wernegreen, 2013; Moran and Bennett, 2014). Thermal instability would not be expected to be a problem for *Ca. Organicella* in Antarctica (Franzmann et al., 1987; Gibson, 1999; Yau et al., 2013). Proteins that are damaged and cannot be correctly re-folded could be degraded to peptides by the encoded ClpXP (Sabree et al., 2013), although the fate of the peptides is unclear in the absence of identifiable peptidases.

## Ca. Organicella Environmental Distribution and Host

To examine the environmental distribution of *Ca. Organicella*, 337 Antarctic lake and marine metagenomes were analyzed, which encompass 77 different Antarctic aquatic locations, including a time (December 2006 to January 2015) and depth series of Organic Lake (Supplementary Figure 1 and Supplementary Table 5). Sequence coverage of *Ca. Organicella* MAGs from Organic Lake was higher at depth in the lake and higher in winter compared with spring or summer (Supplementary Table 5). Although the highest abundance of *Ca. Organicella* was from Organic Lake (up to a median read depth of 71), read coverage showed *Ca. Organicella* was also present in seven other lakes in the Vestfold Hills (Supplementary Figure 1), including a complete MAG from a small pond ~15 km away from Organic Lake (“Unnamed Lake 18”), which had a median read depth of 22 and coverage of the original *Ca. Organicella* MAG (Ga0307966\_1000010) of 99.97% (Supplementary Table 5). The MAGs from Organic Lake (11 close to full length) had an ANI of  $\geq 99.5\%$ , with the ANI across all MAGs from Organic Lake, Unnamed Lake 18, Portals Lake, and Unnamed Lake 13,  $\geq 98.1\%$ . Outside of these *Ca. Organicella* MAGs and *Ca. Pinguicoccus*, the best BLAST matches to the *Ca. Organicella* 16S rRNA gene in NCBI-nr and IMG databases were  $\leq 82\%$ . This indicates that a single species of *Ca. Organicella* is present in the Vestfold Hills, with *Ca. Pinguicoccus* being the only similar species identifiable elsewhere in the world.

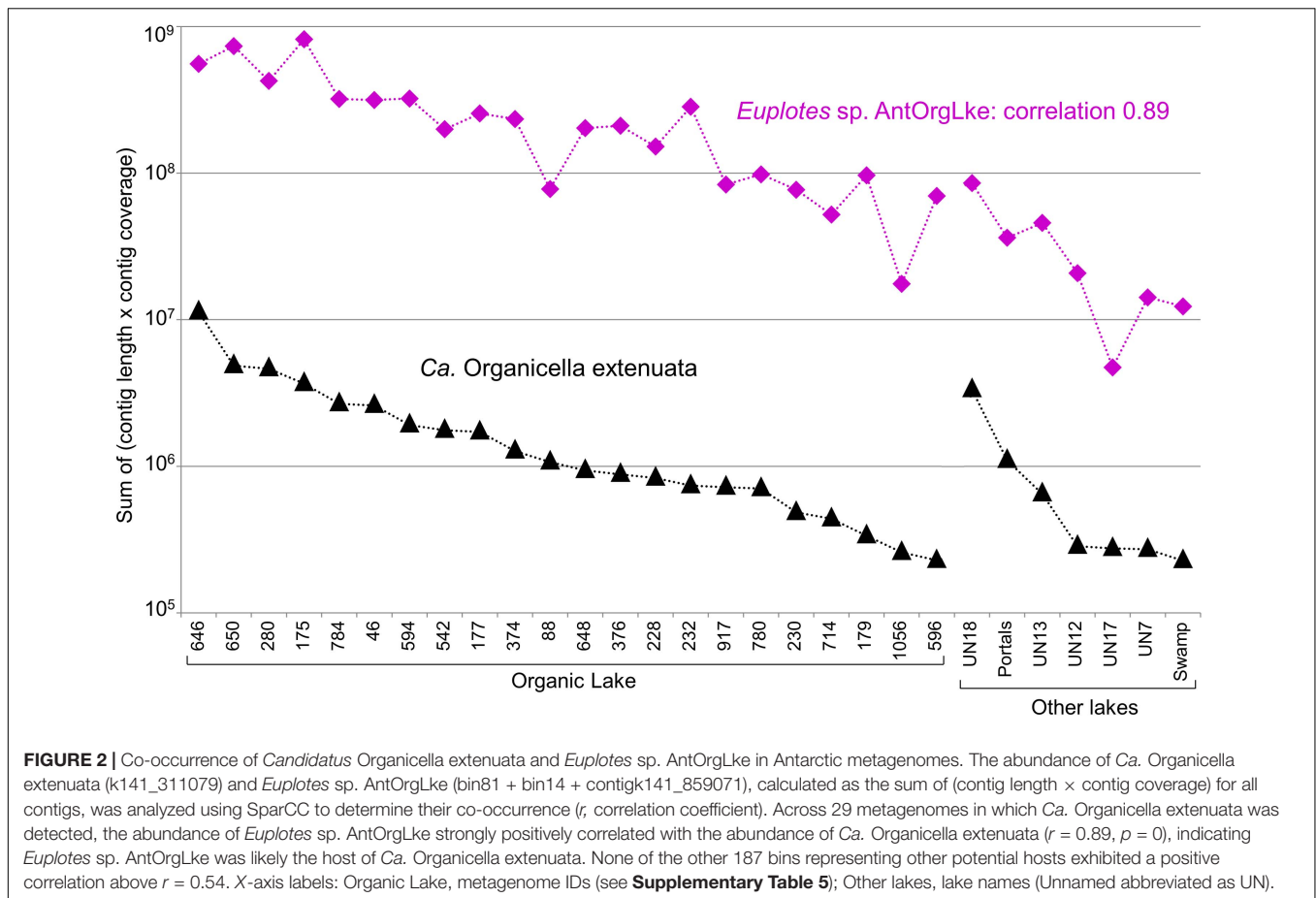
To identify the potential host(s) of *Ca. Organicella*, metagenomes were co-assembled using Metabat, generating a *Ca. Organicella* MAG (k141\_311079) plus 188 potential host bins. The abundance of each bin was determined for each of the 29 metagenomes where *Ca. Organicella* was detected by mapping the metagenome reads to the bins, and the correlation of bin abundances was calculated using SparCC. The abundance of *Ca. Organicella* was highly positively correlated with bin81 ( $r = 0.89$ ,  $p = 0$ ), bin149 ( $r = 0.95$ ,  $p = 0$ ), and contig k141\_859071 ( $r = 0.85$ ,  $p = 0$ ). The two bins and the contig were also highly positively

correlated to each other ( $r = 0.94 - 0.99$ ,  $p = 0$ ). Bin81 (12,580 contigs) and bin149 (18 contigs) were dominated by sequences assigned to the ciliate *Euplotes* (Euplotidae, Spirotrichea, and Ciliophora), and the 8.1-kb contig, k141\_859071 contained a 28S rRNA gene (4,455 bp), region of a 5.8S rRNA gene and 18S rRNA gene (1,895 bp) that matched to *Euplotes* (e.g., 28S rRNA, 84.2% identity to *Euplotes aediculatus* across 79% of query length). We infer that bins 81 and 149 plus the rRNA contig represent a MAG that pertains to a single OTU that we refer to as “*Euplotes* sp. AntOrgLke” (Supplementary Table 6). The *Euplotes* sp. AntOrgLke MAG (Supplementary Dataset 1) comprises 29.98 Mbp across 12,599 contigs (longest contig 19,935 bp, N50 = 2,645, L50 = 3,806, GC = 38.15%), with 6,451 proteins predicted against the TaxDB\_uniClust90\_2018\_08 database (Supplementary Dataset 2) and 15,328 proteins predicted against the MERC\_MMETSP\_UniClust50\_profiles database (Supplementary Dataset 3). Of relevance, the abundance of the *Ca. Organicella* MAG was highly positively correlated with the *Euplotes* sp. AntOrgLke MAG ( $r = 0.89$ ,  $p = 0$ ) (Figure 2), consistent with this ciliate being the host. Moreover, contigs belonging to the *Euplotes* sp. AntOrgLke mitochondrial genome were also detected (Supplementary Table 7; Supplementary Dataset 4).

*Euplotes* sp. AntOrgLke had 97% 18S rRNA identity to *Euplotes cf. antarcticus* and *E. vanleeuwenhoekii*. Tree topology was consistent for all three RNA polymerase sequences (Figure 3) and 18S rRNA sequence (Supplementary Figure 3), and *Euplotes* sp. AntOrgLke seems to be a member of *Euplotes* Clade A (Syberg-Olsen et al., 2016; Boscaro et al., 2018; Serra et al., 2020). The AAI calculated from available *Euplotes* genomic data (six species, including *Euplotes* sp. AntOrgLke) ranged from 49 to 91%, with *Euplotes* sp. AntOrgLke sharing 53–57% with the other five species (Supplementary Table 8). Thus, our data indicate *Euplotes* sp. AntOrgLke is likely a novel Antarctic member of the genus *Euplotes*, and *Ca. Organicella* is a verrucomicrobial endosymbiont of a ciliate species known as *Ca. Pinguicoccus* (Serra et al., 2020). *E. vanleeuwenhoekii*, the host of *Ca. Pinguicoccus*, is a freshwater ciliate (Serra et al., 2020), whereas Organic Lake is hypersaline (Franzmann et al., 1987; Yau et al., 2013).

*Euplotes* is a speciose genus of motile, unicellular ciliate found in many aquatic environments (Boscaro et al., 2019), including Organic Lake, where it was previously detected based on SSU rRNA sequences (Yau et al., 2013). *Euplotes* species have a propensity to harbor one or multiple endosymbiotic bacteria, with at least six genera and 21 species known to date, all of which reside in the cytoplasm (Boscaro et al., 2019; Serra et al., 2020). The majority of reported *Euplotes* endosymbiont species belong to Proteobacteria and are predominantly members of Burkholderiaceae (e.g., *Polynucleobacter*) and the specialized intracellular clades Rickettsiales and Holosporales (Boscaro et al., 2019). The exception is *Ca. Pinguicoccus*, a member of Verrucomicrobia, and the sole known endosymbiont of *E. vanleeuwenhoekii* (Serra et al., 2020). In *E. vanleeuwenhoekii*, *Ca. Pinguicoccus* cells are located free in the cytoplasm and were frequently observed to be in contact with mitochondria and lipid droplets (Serra et al., 2020). The exact benefit of *Ca. Pinguicoccus*





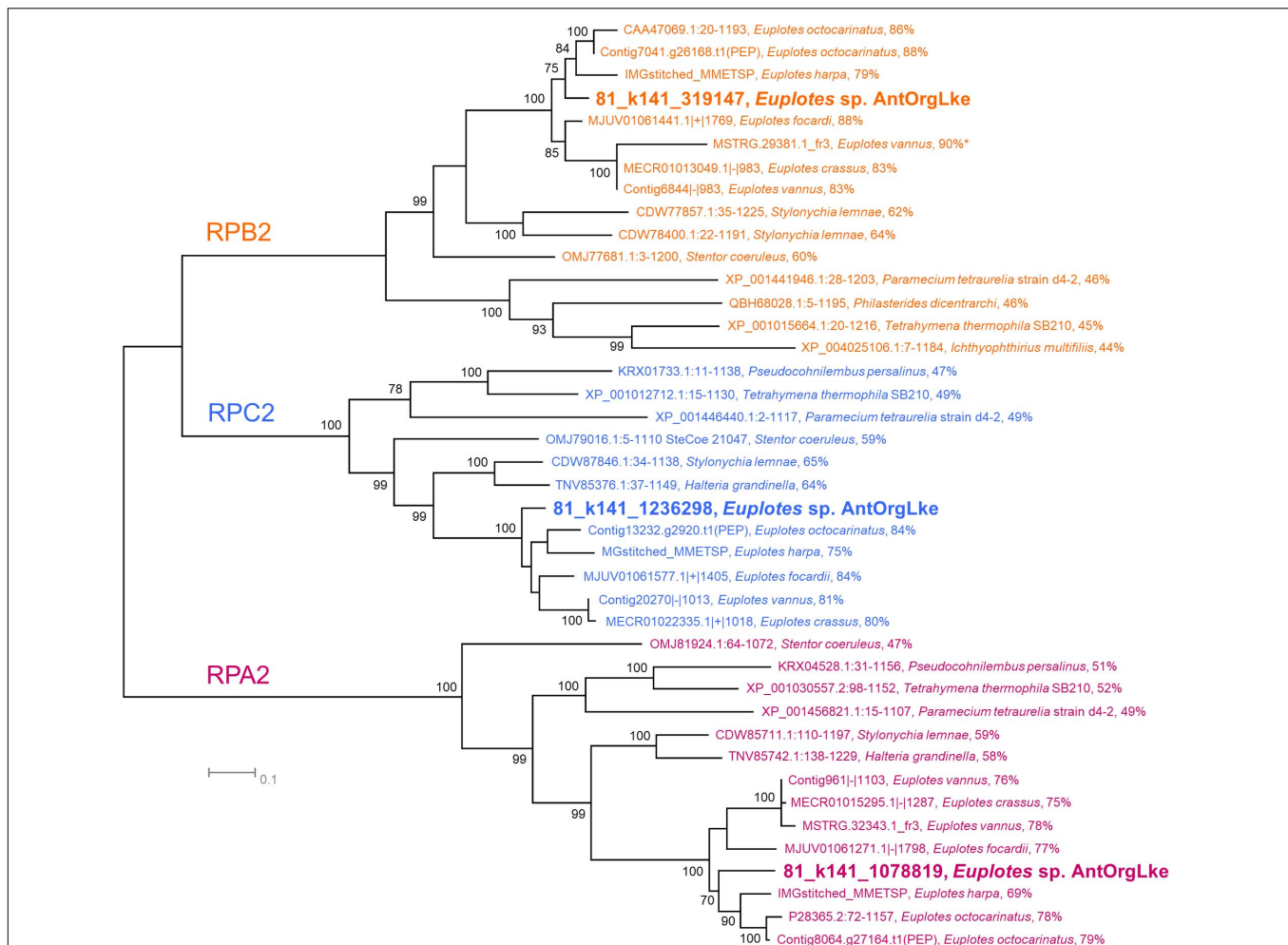
to its ciliate host is unclear, although it is unlikely to be nutritional (see *Ca. Organiccella–Euplotes Interactions*, later) (Serra et al., 2020). Similarly, the foundations of the symbiotic relationship between proteobacterial endosymbionts and *Euplotes* are unclear, including those that are essential symbionts (*Polynucleobacter*, *Ca. Protistobacter*, and *Ca. Devosia*) and accessory symbionts, with the latter possibly being parasitic (Boscaro et al., 2013, 2019).

### Ca. Organiccella–Euplotes Interactions

One possibility is that *Ca. Organiccella* provides Fe-S clusters and fatty acids to its host as the foundation for a mutualistic symbiosis (Figure 4). This is pertinent to *Euplotes*, in which, as in other ciliates, the mitochondrial genome does not encode these functions. We identified 41.8 kb of *Euplotes* sp. AntOrgLke mitochondrial genome sequence—a comparable length to the mitochondrial genome sequences reported for other *Euplotes* species (de Graaf et al., 2009; Serra et al., 2020). Like the mitochondrial genomes of *E. crassus*, *E. minuta*, and *E. vanleewenhoeki*, that of *Euplotes* sp. AntOrgLke has genes that encode electron transport chain proteins, ribosomal proteins, rRNA, tRNA, and a cytochrome *c* assembly protein, along with multiple genes that have no known function, but no identifiable Fe-S cluster or fatty acid synthesis genes (Supplementary Table 7; Pritchard et al., 1990; de Graaf et al., 2009; Swart et al., 2011; Johri et al., 2019; Serra et al., 2020). Within the genus *Euplotes*,

the mitochondrial genetic code includes a single stop codon (UAA), a single unused codon (UAG), and tryptophan-encoding UGA (Pritchard et al., 1990; Burger et al., 2000; Brunk et al., 2003; de Graaf et al., 2009; Swart et al., 2011). By comparison, in *Ca. Organiccella*, UGA is reassigned to Trp, whereas both UAA and UAG are stop codons.

The *Ca. Organiccella* MAG encodes ferredoxin and sulfur utilization factor (SUF) proteins involved in Fe-S cluster biogenesis (SufCBD, SufU, and SufT), including cysteine desulfurase (SufS) for the mobilization of sulfur from cysteine (Selbach et al., 2014; Supplementary Table 3). In eukaryotes, the iron–sulfur cluster (ISC) and SUF pathways are the dominant Fe-S cluster synthesis pathways, with ISC assembly proteins located in the mitochondria, whereas SUF assembly proteins are localized to plastid organelles (Kispal et al., 1999; Tsaousis, 2019), the latter including chloroplasts and apicoplasts (Takahashi et al., 1986; Lill and Mühlenhoff, 2005; Lim and McFadden, 2010; Gisselberg et al., 2013), although, in certain protists, SUF assembly proteins are located in the cytoplasm (Tsaousis et al., 2012; Karnkowska et al., 2016). Typical of eukaryotes, *Euplotes* sp. AntOrgLke encodes homologs of ISC proteins inferred to be present in the model ciliate *Tetrahymena thermophila*, including cysteine desulfurase (Nfs1), ferredoxin (Yah1), and ferredoxin reductase (Arh1) (Supplementary Table 9); ISC assembly would occur in the mitochondrion and depend on nuclear-encoded enzymes



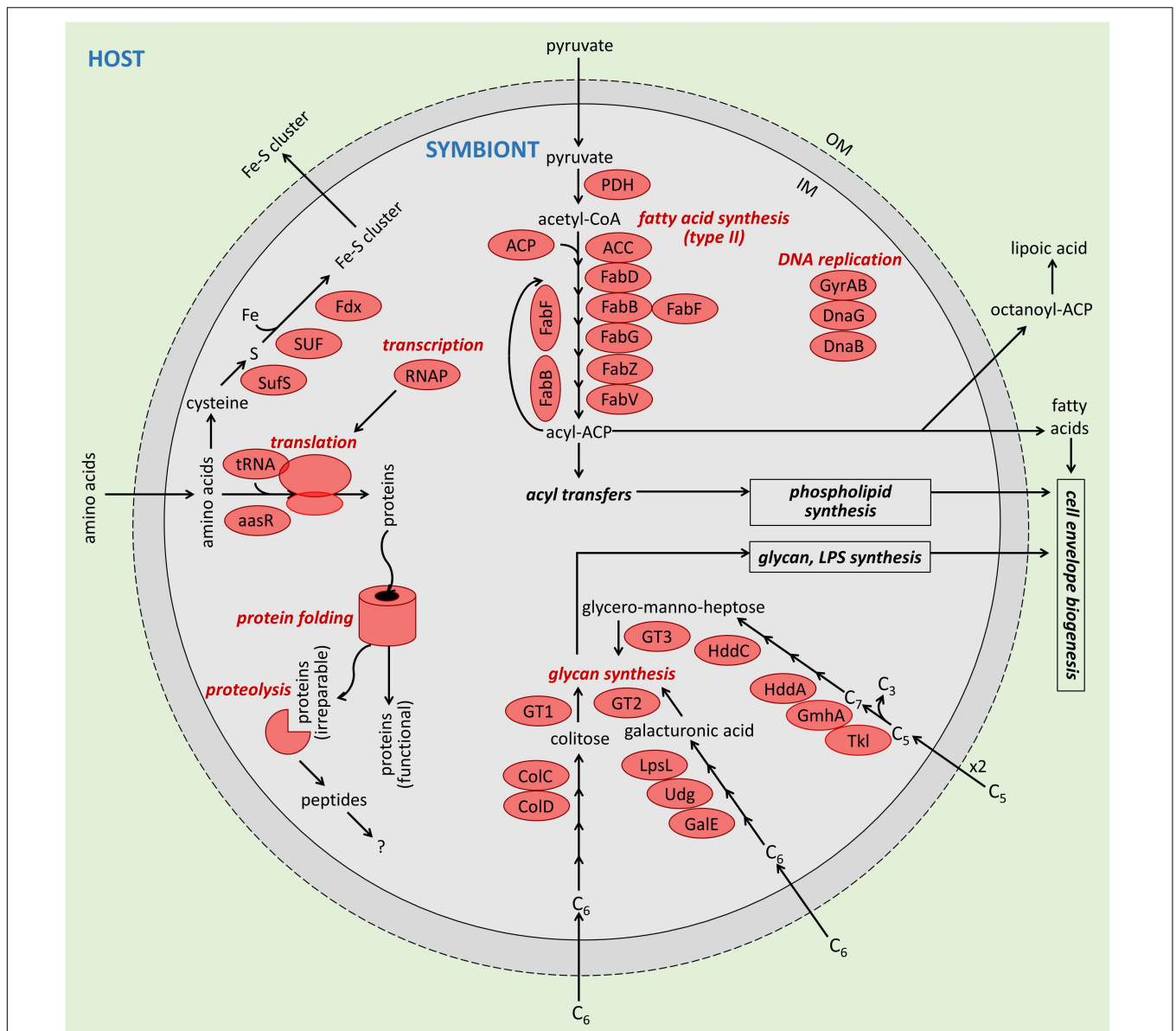
**FIGURE 3 |** Phylogeny of *Euplotes* sp. AntOrgLke. Unrooted maximum likelihood phylogeny of RNA polymerase subunit II proteins from members of Ciliophora showing *Euplotes* sp. AntOrgLke clustering with members of the *Euplotes* genus. Within the cluster for each RNA polymerase type (RPB, RPC, and RPA), the percent identity between *Euplotes* sp. AntOrgLke protein and an individual protein is shown after the species name. Bootstrap values  $\geq 70$  are shown next to individual nodes, and protein sequences are available in **Supplementary Dataset 5**. A total of 41 RNA polymerase subunit II amino acid sequences were used in analysis. Positions with less than 80% site coverage were eliminated, and 944 positions remained in the final dataset, with the exception of MSTRG.29381.1\_fr3, *Euplotes vannus* which was a partial sequence (283 aa) and is marked with an\*.

(Smith et al., 2007). The SUF system of *Ca. Organicella* could therefore function as a complementary Fe-S cluster assembly system to ISC. The SUF system is more resistant to reactive oxygen species than the ISC system (Santos-Garcia et al., 2014); thus, the SUF system encoded by *Ca. Organicella* may be especially important to the host under oxidative stress conditions in response to the degradation of Fe-S clusters of host proteins (Tsaousis, 2019). The SUF system may be especially relevant to *Euplotes* sp. AntOrgLke in Organic Lake and the other Vestfold Hills lakes due to the prevailing environmental conditions (high oxygen concentrations; freezing temperatures; enhanced UV irradiation; **Supplementary Figure 1**) that promote the production of reactive oxygen species (Ricci et al., 2017).

*Ca. Organicella* also encodes an almost complete suite of genes for bacterial type II fatty acid synthesis (FASII), except for FabH, an enzyme involved in fatty acid elongation (see **Supplementary**

**Text – Pyruvate Oxidation and Fatty Acid Synthesis**). It is likely that another condensing enzyme involved in fatty acid elongation encoded in *Ca. Organicella* (FabB or FabF) would substitute for FabH, as inferred for *Ca. Wigglesworthia*, which similarly lacks FabH but otherwise encodes a complete FASII pathway (Zientz et al., 2004; Parsons and Rock, 2013). In support of this, *Escherichia coli* and *Lactococcus lactis* mutants that lack *fabH* are still capable of synthesizing fatty acids (Morgan-Kiss and Cronan, 2008; Yao et al., 2012). For *Ca. Organicella*, the acetyl-CoA precursor for straight-chain fatty acid biosynthesis would be generated using a pyruvate dehydrogenase (PDH) complex, presumably using pyruvate acquired from the host (**Figure 4**).

Many protists depend on the fatty acids provided by mitochondrial FASII for processes such as lipoylation of essential enzymes or for incorporation into phospholipids; despite having their own cytoplasmic FAS apparatus (FAS type I), these



**FIGURE 4 |** Depiction of function of *Candidatus Organicella extenuata* within *Euplotes sp. AntOrgLke*. The potential metabolic capacities of *Ca. Organicella extenuata* are limited to pyruvate oxidation; type II fatty acid biosynthesis; iron-sulfur (Fe-S) cluster assembly; and conversion, activation, and transfer of hexose and heptose sugars. The *Ca. Organicella extenuata* MAG lacks any identifiable genes for glycolysis, tricarboxylic acid cycle (aside from pyruvate oxidation), pentose phosphate pathway, respiration, fermentation, ATP generation (either by oxidative phosphorylation and ATP synthase, or substrate-level phosphorylation), or synthesis of phospholipids (aside from fatty acids), amino acids, nucleic acids, or vitamins. There were no identifiable transporter genes. Processes, pathways, and enzymes that were inferred to be functional in *Ca. Organicella extenuata* are shaded in red. OM, outer membrane; IM, inner membrane; LPS, lipopolysaccharide. Fatty acid synthesis (type II): PDH, pyruvate dehydrogenase; CoA, coenzyme A; ACP, acyl carrier protein; ACC, acetyl-CoA carboxylase complex; FabD, malonyl-CoA-ACP-transacylase; FabB and FabF, 3-oxoacyl-ACP synthase; FabG, 3-oxoacyl-ACP reductase; FabZ, 3-hydroxyacyl-ACP dehydratase; FabV, enoyl-ACP reductase. Glycan synthesis: GT, glycosyltransferase (three different GT); ColD, GDP-4-keto-6-deoxy-D-mannose 3-dehydratase; ColC, GDP-L-colitose synthase; Udg, UDP-glucose 6-dehydrogenase; LpsL, UDP-glucuronate epimerase; Tkl, transketolase; GmhA, phosphoheptose isomerase; HddA, D-glycero- $\alpha$ -D-manno-heptose 7-phosphate kinase; HddC, D-glycero- $\alpha$ -D-manno-heptose 1-phosphate guanylyltransferase. DNA replication: GyrAB, DNA gyrase; DnaG, DNA primase; DnaB, replicative DNA helicase. Transcription: RNAP, RNA polymerase. Translation: tRNA, transfer RNA; aasR, aminoacyl tRNA synthetases. Fe-S cluster assembly: Fdx, ferredoxin; SufS, cysteine desulfurase; SUF, Fe-S cluster assembly complex (SufCBD, SufU, SufT). In this reconstruction, pyruvate is supplied by the host and the sole purpose of PDH is to provide the acetyl-CoA precursor for fatty acid synthesis. The fatty acid synthesis pathway is functionally complete in *Ca. Organicella extenuata*, with FabF or FabB substituting for missing FabH. There are three pathways involved in synthesis of heptose (glycero-manno-heptose) or hexose (galacturonic acid; colitose) subunits of lipopolysaccharide glycans in *Ca. Organicella extenuata*, but none of these pathways are complete, and all depend on exogenous precursors.

eukaryotes depend on the fatty acids provided by mitochondria (Stephens et al., 2007; Hiltunen et al., 2009). However, as in other ciliates, the mitochondrial genome of *Euplotes* lacks genes associated with FASII (Pritchard et al., 1990; Burger et al., 2000; Brunk et al., 2003; Swart et al., 2011; Johri et al., 2019). Thus, we propose the hypothesis that *Ca. Organicella* provides fatty acids to the host for these essential purposes.

Another possibility is that fatty acids are supplied to the host in a nutritional capacity. For example, there is evidence that fatty acids synthesized by *Ca. Blochmannia floridanus* form part of the symbiont's nutritional support to its host (carpenter ant *Camponotus chromaiodes*) during periods when the insect host is feeding on sugar-rich exudates (Zientz et al., 2004; Fan and Wernegreen, 2013). However, we regard this as unlikely, given that it has been predicted that nutritional symbioses are not likely to be necessary for heterotrophic algal and bacterial feeders such as *Euplotes* that can probably obtain all their required nutrients from their diets (Boscaro et al., 2013, 2019; Serra et al., 2020).

It is also possible that FASII in *Ca. Organicella* contributes to its own cellular requirements, including lipoylation of its own PDH and providing precursors for modification of its own cell envelope (Figure 4). In addition to encoding a functionally complete FASII pathway, 17 *Ca. Organicella* genes are predicted to be involved in the biosynthesis of precursors for lipopolysaccharide components: nine proteins are implicated in the biosynthesis of heptose and hexose units (although we could not reconstruct complete pathways), and eight proteins are glycosyltransferases that may be involved in the transfer of nucleotide-activated sugars to construct glycan chains (Supplementary Table 3; Supplementary Text – Glycan Synthesis). Obligate endosymbionts with genomes < 500 kbp typically have few if any genes for cell envelope biogenesis, with these pathways being especially prone to loss (McCutcheon and Moran, 2012; Moran and Bennett, 2014; Brown et al., 2015). *Ca. Organicella* lacks acyltransferases necessary for transferring acyl-ACP to glycerol-3-phosphate to produce phosphatidic acid, the phospholipid precursor in bacteria (Yao et al., 2012), and there are no identifiable genes for the biosynthesis of the glycerophosphate backbone or headgroups of phospholipids or for the 3-deoxy-D-manno-octulosonic acid-lipid A (Kdo<sub>2</sub>-lipid A) precursor of lipopolysaccharides (Wang et al., 2015).

Thus, *Ca. Organicella*, as in other endosymbionts with highly reduced genomes, is assumed to rely entirely on host-derived membranes (Baumann, 2005; McCutcheon and Moran, 2012; Husnik and McCutcheon, 2016). The presence of lipopolysaccharide- and other cell-wall-related genes is not unusual for symbiotic bacteria with larger genomes (Zientz et al., 2004; Nikoh et al., 2011); for example, the insect endosymbionts *Ca. Wigglesworthia* and *Ca. Blochmannia* (both between 615 and 706 kbp) encode the majority of genes necessary for the synthesis of a normal gram-negative cell wall, including phospholipids and lipopolysaccharides (Akman et al., 2002; Gil et al., 2003; Zientz et al., 2004). Additionally, certain obligately symbiotic bacteria with larger genomes (>600 kb) retain a complete set of FASII genes (Akman et al., 2002; Gil et al., 2003; Nikoh et al., 2011; Lamelas et al., 2011;

Chong and Moran, 2018). In these symbionts, the retention of genes necessary for the synthesis of a normal Gram-negative cell wall (including lipopolysaccharides) is possibly for protection against the host and/or reflects a relatively recent symbiotic association (Akman et al., 2002; Gil et al., 2003). The latter does not apply to *Ca. Organicella*, with the extreme reduction in genome size reflecting an ancient symbiosis (Serra et al., 2020).

Nevertheless, *Ca. Organicella* might contribute glycan components to its own cell envelope (including lipopolysaccharides). One possibility is that modifications of the endosymbiont cell wall confer some protection against the host, such as through variation of fatty acid length or altering the glycan moieties of lipopolysaccharides (core and/or O-specific polysaccharides) using modified sugars by the action of glycosyltransferases (Serra et al., 2020). *Ca. Pinguicoccus* has a very similar genome size and gene composition as *Ca. Organicella*, including retaining homologs of the same glycan/lipopolysaccharide-related genes (see *Comparison of Ca. Organicella and Ca. Pinguicoccus Genomes*, later). *Ca. Pinguicoccus* resides free in the cytoplasm of *E. vanleeuwenhoekii*, and it has been proposed that endosymbionts in the host cytoplasm of eukaryote cells face a less stable and possibly hostile environment compared with those endosymbionts that are enclosed within specialized bacteriocytes or host-derived vesicles (Gil et al., 2003; Wu et al., 2004; Serra et al., 2020). For this reason, *Ca. Pinguicoccus* may exercise some control over the composition of its cell envelope because it is in direct contact with the host cytoplasm (Serra et al., 2020). This might also be true of *Ca. Organicella*, which, based on its close phylogenetic affiliation with *Ca. Pinguicoccus* and having *Euplotes* as the putative host, likely lives in the host cytoplasm.

## Comparison of *Ca. Organicella* and *Ca. Pinguicoccus* Genomes

The genome sizes of *Ca. Organicella* (158,228 bp, 194 genes, 163 CDS) and *Ca. Pinguicoccus* (163,218 bp, 205 genes, 168 CDS; Serra et al., 2020) are similar; note that the protein sequences for the *Ca. Pinguicoccus* NCBI (Accession CP039370) genome sequence were auto-predicted with genetic code 11, but using genetic code 4, we predict a total of 200 genes [five less than reported in Serra et al. (2020)], consisting of 163 CDS, 34 tRNAs, and the 16S, 5S, and 23S rRNA genes (Supplementary Table 1). The two genomes share extensive synteny (Supplementary Figure 4). Although the genomic nucleotide sequences were too divergent to calculate ANI, the AAI between the two symbiont genomes was determined to be 46% (two-way AAI based on 134 proteins, all predicted with genetic code 4). Both genomes retain an almost identical small subset of genes represented across Verrucomicrobia (Serra et al., 2020; Supplementary Table 3). They also share homologous proteins required for DNA replication, transcription, and translation, in common with other endosymbionts, but both lack the catalytic subunit of DNA polymerase (DnaE), which is exceptional among endosymbionts (Serra et al., 2020).

*Ca. Pinguicoccus* encodes the same components of the SUF system and a functionally complete FASII pathway as *Ca. Organicella*, suggesting that *Ca. Pinguicoccus* confers the same benefits to its *Euplotes* host that we infer for *Ca. Organicella*. Of interest is that *Ca. Pinguicoccus* cells were often observed associated with lipid droplets in *E. vanleeuwenhoekii* cytoplasm, raising the possibility of a link between the retention of FASII genes and interaction with the host's lipids (Serra et al., 2020). *Ca. Pinguicoccus* also encodes homologs of the same glycosyltransferases and heptose- and hexose-related enzymes encoded in *Ca. Organicella* (**Supplementary Table 3**). Nevertheless, *Ca. Pinguicoccus* retains a putative phospholipid synthesis protein (CDP-diacylglycerol-glycerol-3-phosphate 3-phosphatidyltransferase homolog) not identifiable in *Ca. Organicella*. *Ca. Pinguicoccus* encodes a thioredoxin-thioredoxin reductase system (for maintaining thiol-disulfide redox balance) and NADP-dependent glutamate dehydrogenase (for the reversible oxidative deamination of glutamate), neither of which are identifiable in *Ca. Organicella*. There are also variations between the two genera in the exact complement of ribosomal subunits, aaRS, and initiation factor subunits (**Supplementary Table 3**), with these components being prone to loss among endosymbionts (Moran and Bennett, 2014). However, *Ca. Organicella* and *Ca. Pinguicoccus* possess the same 34 tRNA genes.

Overall, the data suggest that, as their divergence from a common ancestor had a highly reduced genome, further genomic erosion has occurred independently in *Ca. Organicella* and *Ca. Pinguicoccus*, with differential loss of certain genes, especially those involved in translation. By contrast, SUF, FASII, and certain lipopolysaccharide/glycan-related genes are conserved between the two genera. This suggests that these particular genes play important roles in the interactions of these endosymbionts with their ciliate hosts.

## CONCLUSION

Many of the smallest bacterial genomes are from insect symbionts that exist as metabolically complementary partnerships within the host (43) (**Supplementary Table 1**). For example, *Ca. Nasuia deltocephalinicola* (~112 kbp) and *Ca. Hodgkinia cicadicola* (~144 kbp) are each co-resident with *Ca. Sulcia* (Bennett and Moran, 2013; McCutcheon et al., 2009), whereas *Ca. Tremblaya princeps* (~139 kbp) contains *Ca. Moranella endobia* to constitute a nested symbiosis (McCutcheon and von Dohlen, 2011). However, *Ca. Carsonella ruddii* (~160 kbp) is a lone endosymbiont resident in sap-feeding psyllids (Thao et al., 2000; Nakabachi et al., 2006). Unlike known insect symbionts with highly reduced genomes (Nakabachi and Ishikawa, 1999; Zientz et al., 2004; Nakabachi et al., 2006; Pérez-Brocal et al., 2006; Bennett and Moran, 2013; Brown et al., 2015; Gil et al., 2018), *Ca. Organicella* and *Ca. Pinguicoccus* lack any capacity for the biosynthesis of amino acids or vitamins (Serra et al., 2020). Thus, as with the *Ca. Pinguicoccus-E. vanleeuwenhoekii* symbiosis, there is no reason to assume that *Ca. Organicella* exists as part of a co-symbiotic partnership, especially considering

that none of the genes encode enzymes involved in amino acid or vitamin biosynthesis, as is typical for such consortia (McCutcheon et al., 2009; McCutcheon and von Dohlen, 2011). The absence of a nutritional basis of a proposed *Ca. Organicella-Euplotes* symbiosis likely reflects the algivorous and bacterivorous nature of the ciliate host (Serra et al., 2020), in contrast to insects with specialized and nutritionally unbalanced diets. Instead, we propose that *Ca. Organicella* and *Ca. Pinguicoccus* provide SUF Fe-S clusters and FASII fatty acids as essential molecules to the host, with FASII replacing a lost mitochondrial function in *Euplotes*. Additionally, the ciliate host would possess dual Fe-S cluster biogenesis systems, with the SUF system provided by endosymbiotic Verrucomicrobia.

*Ca. Organicella* was identified as possibly being an endosymbiont by virtue of having unusual coding parameters (Ivanova et al., 2014) rather than by searching our metagenome data for symbionts. Previous metagenomic screening of multiple *Euplotes* strains and their resident bacteria did not recover symbionts that belong to phylum Verrucomicrobia (Boscaro et al., 2019). In the study, the identification of putative symbionts in *Euplotes*-based metagenome data was based on bacterial taxa that were referable to known clades of exclusively intracellular bacteria (e.g., Rickettsiales) or related to previously described protist symbionts (e.g., *Polynucleobacter*) (Boscaro et al., 2019); thus, any verrucomicrobial symbionts might have been overlooked, especially if they were present at low coverage. Targeted hosts and/or symbiont reference genomes have been used extensively for identifying both ecto- and endo-symbionts of a broad range of taxa, including magnetotactic bacteria of marine protists (Monteil et al., 2019), gut symbionts of hadal snailfish (Lian et al., 2020) and phytophagous stink bugs (Kashkouli et al., 2020), and symbionts of pea aphids (Guyomar et al., 2018) and scleractinian corals (Shinzato et al., 2014). The discovery of *Ca. Pinguicoccus* arose through the development of a “next-generation taxonomy” approach for assessing symbiont-host associations that combines “bio-taxonomy tools, classical morphology, ultrastructure, molecular phylogeny, genomics, and bioinformatics” (Serra et al., 2020). The study focused on *Euplotes* as a model protist “holobiont,” in the process of identifying *Ca. Pinguicoccus*. Being a host-based approach to endosymbiont discovery, the “next-generation taxonomy” approach is applicable to known symbiotic communities. Software (MinYS) has also recently been reported to specifically identify symbionts from genome assemblies of symbiotic communities by using reference genomes (Guyomar et al., 2020). Although genetic code 4 (UGA stop codons recoded to tryptophan) has been reported to only occur rarely (Yamao et al., 1985; Knight et al., 2001; McCutcheon et al., 2009), our findings raise the enticing prospect that searching contigs and MAGs for this recoding may reveal new symbionts, including members of the verrucomicrobial “Nucleococcus cluster” (**Figure 1**).

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The repository and accession numbers are: NCBI

(*Euplotes* sp. AntOrgLke MAG, accession: JAGXKF000000000 and PRJNA720161); IMG (<https://img.jgi.doe.gov/>) (*Ca. Organicella extenuata* MAG: Scaffold Ga0307966\_1000010; *Euplotes* sp. AntOrgLke mitochondrial genome: 2 Scaffolds, Ga0307966\_1001133 and Ga0307966\_1001206).

## ETHICS STATEMENT

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

## AUTHOR CONTRIBUTIONS

TW, MA, NI, MH, and RC conceived the study, analyzed the data, and conducted data interpretation. SH performed Sterivex filter DNA extractions. AH and SB spent 18 months in Antarctica running the 2013–2015 expedition. TW, MA, and RC wrote the manuscript with input from all other co-authors. All authors have read and approved the manuscript submission.

## REFERENCES

- Akman, L., Yamashita, A., Watanabe, H., Oshima, K., Shiba, T., Hattori, M., et al. (2002). Genome sequence of the endocellular obligate symbiont of tsetse flies. *Wigglesworthia glossinidia*. *Nat. Genet.* 32, 402–407. doi: 10.1038/ng986
- Allen, M. A., Lauro, F. M., Williams, T. J., Burg, D., Siddiqui, K. S., DeFrancisci, D., et al. (2009). The genome sequence of the psychrophilic archaeon, *Methanococcoides burtonii*: the role of genome evolution in cold adaptation. *ISME J.* 3, 1012–1035. doi: 10.1038/ismej.2009.45
- Baumann, P. (2005). Biology of bacteriocyte-associated endosymbionts of plant sap-sucking insects. *Annu. Rev. Microbiol.* 59, 155–189. doi: 10.1146/annurev.micro.59.030804.121041
- Bennett, G. M., McCutcheon, J. P., MacDonald, B. R., Romanovicz, D., and Moran, N. A. (2014). Differential genome evolution between companion symbionts in an insect-bacterial symbiosis. *mBio* 5:e01697-14.
- Bennett, G. M., and Moran, N. A. (2013). Small, smaller, smallest: the origins and evolution of ancient dual symbioses in a phloem-feeding insect. *Genome Biol. Evol.* 5, 1675–1688. doi: 10.1093/gbe/evt118
- Bennett, G. M., and Moran, N. A. (2015). Heritable symbiosis: the advantages and perils of an evolutionary rabbit hole. *Proc. Natl. Acad. Sci. U.S.A.* 112, 10169–10176. doi: 10.1073/pnas.1421388112
- Boscaro, V., Felletti, M., Vannini, C., Ackerman, M. S., Chain, P. S., Malfatti, S., et al. (2013). *Polynucleobacter necessarius*, a model for genome reduction in both free-living and symbiotic bacteria. *Proc Natl Acad Sci U.S.A.* 110, 18590–18595. doi: 10.1073/pnas.1316687110
- Boscaro, V., Husnik, F., Vannini, C., and Keeling, P. J. (2019). Symbionts of the ciliate *Euplotes*: diversity, patterns and potential as models for bacteria-eukaryote endosymbioses. *Proc. R. Soc. Lond. B* 286:20190693. doi: 10.1098/rspb.2019.0693
- Boscaro, V., Syberg-Olsen, M. J., Irwin, N. A. T., del Campo, J., and Keeling, P. J. (2018). What can environmental sequences tell us about the distribution of low-rank taxa? The case of *Euplotes* (Ciliophora, Spirotrichea), including a description of *Euplotes enigma* sp. nov. *J. Eukaryot. Microbiol.* 66, 281–293. doi: 10.1111/jeu.12669
- Brown, A. M., Howe, D. K., Wasala, S. K., Peetz, A. B., Zasada, I. A., and Denver, D. R. (2015). Comparative genomics of a plant-parasitic nematode endosymbiont suggest a role in nutritional symbiosis. *Genome Biol. Evol.* 7, 2727–2746. doi: 10.1093/gbe/evv176
- Brunk, C. F., Lee, L. C., Tran, A. B., and Li, J. (2003). Complete sequence of Mt genome of *Tetrahymena thermophila* and comparative methods for identifying

## FUNDING

This work was supported by the Australian Research Council (DP150100244) and the Australian Antarctic Science Program (project 4031). The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, was supported by the Office of Science of the U.S. Department of Energy under contract no. DE-AC02-05CH11231.

## ACKNOWLEDGMENTS

Computational analyses at UNSW Sydney were performed on the computational cluster Katana, supported by Research Technology Services at UNSW Sydney.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2021.674758/full#supplementary-material>

- highly divergent genes. *Nucl. Acid Res.* 31, 1673–1682. doi: 10.1093/nar/gkg270
- Burger, G., Zhu, Y., Littlejohn, T. G., Greenwood, S. J., Schnare, M. N., Lang, B. F., et al. (2000). Complete sequence of the mitochondrial genome of *Tetrahymena pyriformis* and comparison with *Paramecium aurelia* mitochondrial DNA. *J. Mol. Biol.* 297, 365–380. doi: 10.1006/jmbi.2000.3529
- Burke, G., Fiehn, O., and Moran, N. (2010). Effects of facultative symbionts and heat stress on the metabolome of pea aphids. *ISME J.* 4, 242–252. doi: 10.1038/ismej.2009.114
- Bushnell, B. (2014). *BBMap: A Fast, Accurate, Splice-Aware Aligner*. Lawrence Berkeley National Laboratory. LBNL Report #: LBNL-7065E. Berkeley, CA: LBNL.
- Cavicchioli, R. (2015). Microbial ecology of Antarctic aquatic systems. *Nat. Rev. Microbiol.* 13, 691–706. doi: 10.1038/nrmicro3549
- Chen, X., Jiang, Y., Gao, F., Zheng, W., Krock, T. J., Stover, N. A., et al. (2019). Genome analyses of the new model protist *Euplotes vannus* focusing on genome rearrangement and resistance to environmental stressors. *Mol. Ecol. Resour.* 19, 1292–1308. doi: 10.1111/1755-0998.13023
- Chong, R. A., and Moran, N. A. (2018). Evolutionary loss and replacement of *Buchnera*, the obligate endosymbiont of aphids. *ISME J.* 12, 898–908. doi: 10.1038/s41396-017-0024-6
- Choo, Y. J., Lee, K., Song, J., and Cho, J. C. (2007). *Puniceicoccus vermicola* gen. nov., sp. nov., a novel marine bacterium, and description of Puniceicocaceae fam. nov., Puniceicoccales ord. nov., Opitutaceae fam. nov., Opitutales ord. nov. and Opitutae classis nov. in the phylum ‘Verrucomicrobia’. *Int. J. Syst. Evol. Microbiol.* 57, 532–537. doi: 10.1099/ijs.0.64616-0
- Darling, A. C., Mau, B., Blattner, F. R., and Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14, 1394–1403. doi: 10.1101/gr.2289704
- de Graaf, R. M., van Alen, T. A., Dutilh, B. E., Kuiper, J. W., van Zoggel, H. J., Huynh, M. B., et al. (2009). The mitochondrial genomes of the ciliates *Euplotes minuta* and *Euplotes crassus*. *BMC Genomics* 10:514. doi: 10.1186/1471-2164-10-514
- Derrien, M., Vaughan, E. E., Plugge, C. M., and de Vos, W. M. (2004). *Akkermansia muciniphila* gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. *Int. J. Syst. Evol. Microbiol.* 54, 1469–1476. doi: 10.1099/ijs.0.02873-0
- Dunfield, P. F., Yuryev, A., Senin, P., Smirnova, A. V., Stott, M. B., Hou, S., et al. (2007). Methane oxidation by an extremely acidophilic bacterium of the phylum Verrucomicrobia. *Nature* 450, 879–882. doi: 10.1038/nature06411

- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340
- Fan, Y., and Wernegreen, J. J. (2013). Can't take the heat: high temperature depletes bacterial endosymbionts of ants. *Microb. Ecol.* 66, 727–733. doi: 10.1007/s00248-013-0264-6
- Franzmann, P. D., Deprez, P. P., Burton, H. R., and van den Hoff, J. (1987). Limnology of Organic Lake, Antarctica, a meromictic lake that contains high concentrations of dimethyl sulfide. *Aust. J. Mar. Freshw. Res.* 38, 409–417. doi: 10.1071/mf9870409
- Friedman, J., and Alm, E. J. (2012). Inferring correlation networks from genomic survey data. *PLoS Comput. Biol.* 8:e1002687. doi: 10.1371/journal.pcbi.1002687
- Gibson, J. A. E. (1999). The meromictic lakes and stratified marine basins of the Vestfold Hills, East Antarctica. *Antarct. Sci.* 11, 175–192. doi: 10.1017/s0954102099000243
- Gibson, J. A. E., Garrick, R. C., Franzmann, P. D., Deprez, P. P., and Burton, H. (1991). Reduced sulfur gases in saline lakes of the Vestfold Hills, Antarctica. *Palaeogeog. Palaeoclimatol. Palaeoecol.* 84, 131–140. doi: 10.1016/0031-0182(91)90040-x
- Gil, R., Silva, F. J., Zientz, E., Delmotte, F., Gonzalez-Candelas, F., Latorre, A., et al. (2003). The genome sequence of *Blochmannia floridanus*: comparative analysis of reduced genomes. *Proc. Natl. Acad. Sci. U.S.A.* 100, 9388–9393. doi: 10.1073/pnas.1533499100
- Gil, R., Vargas-Chavez, C., López-Madrugal, S., Santos-García, D., Latorre, A., and Moya, A. (2018). *Tremblaya phenacola* PPER: an evolutionary beta-gammaproteobacterium collage. *ISME J.* 12, 124–135. doi: 10.1038/ismej.2017.144
- Gisselberg, J. E., Dellibovi-Ragheb, T. A., Matthews, K. A., Bosch, G., and Prigge, S. T. (2013). The suf iron-sulfur cluster synthesis pathway is required for apicoplast maintenance in malaria parasites. *PLoS Pathog.* 9:e1003655. doi: 10.1371/journal.ppat.1003655
- Guyomar, C., Delage, W., Legeai, F., Mougél, C., Simon, J.-C., and Lemaitre, C. (2020). MinYS: mine your symbiont by targeted genome assembly in symbiotic communities. *NAR Genom. Bioinform.* 2:lqaa047.
- Guyomar, C., Legeai, F., Jousselin, E., Mougél, C., Lemaitre, C., and Simon, J. C. (2018). Multi-scale characterization of symbiont diversity in the pea aphid complex through metagenomic approaches. *Microbiome* 6:181.
- Hiltunen, J. K., Schonauer, M. S., Autio, K. J., Mittelmeier, T. M., Kastaniotis, A. J., and Dieckmann, C. L. (2009). Mitochondrial fatty acid synthesis type II: more than just fatty acids. *J. Biol. Chem.* 284, 9011–9015. doi: 10.1074/jbc.r800068200
- Hongoh, Y., Sato, T., Dolan, M. F., Noda, S., Ui, S., Kudo, T., et al. (2007). The motility symbiont of the termite gut flagellate *Caduceia versatilis* is a member of the "Synergistes" group. *Environ. Microbiol.* 73, 6270–6276. doi: 10.1128/aem.00750-07
- Hongoh, Y., Sharma, V. K., Prakash, T., Noda, S., Taylor, T. D., Kudo, T., et al. (2008a). Complete genome of the uncultured Termite Group 1 bacteria in a single host protist cell. *Proc. Natl. Acad. Sci. U.S.A.* 105, 5555–5560. doi: 10.1073/pnas.0801389105
- Hongoh, Y., Sharma, V. K., Prakash, T., Noda, S., Toh, H., Taylor, T. D., et al. (2008b). Genome of an endosymbiont coupling N<sub>2</sub> fixation to cellulolysis within protist cells in termite gut. *Science* 322, 1108–1109. doi: 10.1126/science.1165578
- Huntemann, M., Ivanova, N. N., Mavromatis, K., Tripp, H. J., Paez-Espino, D., Tennessen, K., et al. (2015). The standard operating procedure of the DOE-JGI Metagenome Annotation Pipeline (MAP v.4). *Stand. Genomic Sci.* 11:17.
- Husnik, F., and McCutcheon, J. P. (2016). Repeated replacement of an intrabacterial symbiont in the tripartite nested mealybug symbiosis. *Proc. Natl. Acad. Sci. U.S.A.* 113, E5416–E5424.
- Ikeda-Ohtsubo, W., Faivre, N., and Brune, A. (2010). Putatively free-living "Endomicrobia" – ancestors of the intracellular symbionts of termite gut flagellates? *Environ. Microbiol. Rep.* 2, 554–559. doi: 10.1111/j.1758-2229.2009.00124.x
- Ikeda-Ohtsubo, W., Strasser, J. F., Kohler, T., Mikaelyan, A., Gregor, I., McHardy, A. C., et al. (2016). 'Candidatus Adiatritrix intracellularis', an endosymbiont of termite gut flagellates, is the first representative of a deep-branching clade of Deltaproteobacteria and a putative homoacetogen. *Environ. Microbiol.* 18, 2548–2564. doi: 10.1111/1462-2920.13234
- Ivanova, N. N., Schwientek, P., Tripp, H. J., Rinke, C., Pati, A., Huntemann, M., et al. (2014). Stop codon reassignments in the wild. *Science* 344:909. doi: 10.1126/science.1250691
- Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., and Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* 9:5114.
- James, S. R., Burton, H. R., McMeekin, T. A., and Mancuso, C. A. (1994). Seasonal abundance of *Halomonas meridiana*, *Halomonas subglaciescola*, *Flavobacterium gondwanense* and *Flavobacterium salegens* in four Antarctic lakes. *Antarct. Sci.* 6, 325–332. doi: 10.1017/s0954102094000490
- Johri, P., Marinov, G. K., Doak, T. G., and Lynch, M. (2019). Population genetics of *Paramecium* mitochondrial genomes: recombination, mutation spectrum, and efficacy of selection. *Genome Biol. Evol.* 11, 1398–1416. doi: 10.1093/gbe/evz081
- Kang, D. D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., et al. (2019). MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* 7:e7359. doi: 10.7717/peerj.7359
- Karnkowska, A., Vacek, V., Zubáková, Z., Treitl, S. C., Petrželková, R., Eme, L., et al. (2016). A Eukaryote without a mitochondrial organelle. *Curr. Biol.* 26, 1274–1284. doi: 10.1016/j.cub.2016.03.053
- Kashkouli, M., Castelli, M., Floriano, A. M., Bandi, C., Epis, S., Fathipour, Y., et al. (2020). Characterization of a novel *Pantoea* symbiont allows inference of a pattern of convergent genome reduction in bacteria associated with Pentatomidae. *Environ. Microbiol.* 23, 36–50. doi: 10.1111/1462-2920.15169
- Kiraga, J., Mackiewicz, P., Mackiewicz, D., Kowalczyk, M., Bieček, P., Polak, N., et al. (2007). The relationships between the isoelectric point and length of proteins, taxonomy and ecology of organisms. *BMC Genomics* 8:163. doi: 10.1186/1471-2164-8-163
- Kispał, G., Csere, P., Prohl, C., and Lill, R. (1999). The mitochondrial proteins Atm1p and Nfs1p are required for biogenesis of cytosolic Fe/S proteins. *EMBO J.* 18, 3981–3989. doi: 10.1093/emboj/18.14.3981
- Knaff, D. B., and Hirasawa, M. (1991). Ferredoxin-dependent chloroplast enzymes. *Biochim. Biophys. Acta* 1056, 93–125. doi: 10.1016/s0005-2728(05)80277-4
- Knight, R. D., Freeland, S. J., and Landweber, L. F. (2001). Rewiring the keyboard: evolvability of the genetic code. *Nat. Rev. Genet.* 2, 49–58. doi: 10.1038/35047500
- Kozłowski, L. P. (2016). IPC – Isoelectric point calculator. *Biol. Direct* 11:55.
- Kuwahara, H., Yuki, M., Izawa, K., Ohkuma, M., and Hongoh, Y. (2017). Genome of "Ca.*Desulfovibrio trichonymphae*", an H<sub>2</sub>-oxidizing bacterium in a tripartite symbiotic system within a protist cell in the termite gut. *ISME J.* 11, 766–776. doi: 10.1038/ismej.2016.143
- Lagesen, K., Hallin, P., Rødland, E. A., Staerfeldt, H. H., Rognes, T., and Ussery, D. W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35, 3100–3108. doi: 10.1093/nar/gkm160
- Lamelas, A., Gosalbes, M. J., Moya, A., and Latorre, A. (2011). New clues about the evolutionary history of metabolic losses in bacterial endosymbionts, provided by the genome of *Buchnera aphidicola* from the aphid *Cinara tujafilina*. *Appl. Environ. Microbiol.* 77, 4446–4454. doi: 10.1128/aem.00141-11
- Levy Karin, E., Mirdita, M., and Söding, J. (2020). MetaEuk—sensitive, high-throughput gene discovery, and annotation for large-scale eukaryotic metagenomics. *Microbiome* 8:48.
- Li, D., Luo, R., Liu, C. M., Leung, C. M., Ting, H. F., Sadakane, K., et al. (2016). MEGAHIT v1.0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods* 102, 3–11. doi: 10.1016/j.ymeth.2016.02.020
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- Lian, C. A., Yan, G. Y., Huang, J. M., Danchin, A., Wang, Y., and He, L. S. (2020). Genomic characterization of a Novel gut symbiont from the Hadal Snailfish. *Front. Microbiol.* 202978. doi: 10.3389/fmicb.2019.02978
- Lill, R., and Mühlenhoff, U. (2005). Iron-sulfur-protein biogenesis in eukaryotes. *Trends Biochem. Sci.* 30, 133–141. doi: 10.1016/j.tibs.2005.01.006
- Lim, L., and McFadden, G. I. (2010). The evolution, metabolism and functions of the apicoplast. *Phil. Trans. R. Soc. Lond. B* 365, 749–763. doi: 10.1098/rstb.2009.0273

- Lopera, J., Miller, I. J., McPhail, K. L., and Kwan, J. C. (2017). Increased biosynthetic gene dosage in a genome-reduced defensive bacterial symbiont. *mSystems* 2:e00096-17. doi: 10.1128/mSystems.00096-17
- López-Madrugal, S., Latorre, A., Porcar, M., Moya, A., and Gil, R. (2013). Mealybugs nested endosymbiosis: going into the 'mattyoshka' system in *Planococcus citri* in depth. *BMC Microbiol.* 13:74. doi: 10.1186/1471-2180-13-74
- McAllister, K. A., Peery, R. B., and Zhao, G. (2006). Acyl carrier protein synthases from gram-negative, gram-positive, and atypical bacterial species: biochemical and structural properties and physiological implications. *J. Bacteriol.* 188, 4737–4748. doi: 10.1128/jb.01917-05
- McCutcheon, J. P. (2010). The bacterial essence of tiny symbiont genomes. *Curr. Opin. Microbiol.* 13:73. doi: 10.1016/j.mib.2009.12.002
- McCutcheon, J. P., McDonald, B. R., and Moran, N. A. (2009). Origin of an alternative genetic code in the extremely small and GC-rich genome of a bacterial symbiont. *PLoS Genet.* 5:e1000565. doi: 10.1371/journal.pgen.1000565
- McCutcheon, J. P., and Moran, N. A. (2010). Functional convergence in reduced genomes of bacterial symbionts spanning 200 My of evolution. *Genome Biol. Evol.* 2, 708–718. doi: 10.1093/gbe/evq055
- McCutcheon, J. P., and Moran, N. A. (2012). Extreme genome reduction in symbiotic bacteria. *Nat. Rev. Microbiol.* 10, 13–26. doi: 10.1038/nrmicro2670
- McCutcheon, J. P., and von Dohlen, C. D. (2011). An interdependent metabolic patchwork in the nested symbiosis of mealybugs. *Curr. Biol.* 21, 1366–1372. doi: 10.1016/j.cub.2011.06.051
- Monteil, C. L., Vallenet, D., Menguy, N., Benzerara, K., Barbe, V., Fouteau, S., et al. (2019). Ectosymbiotic bacteria at the origin of magnetoreception in a marine protist. *Nat. Microbiol.* 4, 1088–1095. doi: 10.1038/s41564-019-0432-7
- Moran, N. A. (1996). Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 93, 2873–2878. doi: 10.1073/pnas.93.7.2873
- Moran, N. A., and Bennett, G. M. (2014). The tiniest tiny genomes. *Annu. Rev. Microbiol.* 68, 195–215. doi: 10.1146/annurev-micro-091213-112901
- Morgan-Kiss, R. M., and Cronan, J. E. (2008). The *Lactococcus lactis* FabF fatty acid synthetic enzyme can functionally replace both the FabB and FabF proteins of *Escherichia coli* and the FabH protein of *Lactococcus lactis*. *Arch. Microbiol.* 190, 427–437. doi: 10.1007/s00203-008-0390-6
- Moya, A., Pereto, J., Gil, R., and Latorre, A. (2008). Learning how to live together: genomic insights into prokaryote-animal symbioses. *Nat. Rev. Genet.* 9, 218–229. doi: 10.1038/nrg2319
- Nakabachi, A., and Ishikawa, H. (1999). Provision of riboflavin to the host aphid, *Acyrothosiphon pisum*, by endosymbiotic bacteria, *Buchnera*. *J. Insect. Physiol.* 45, 1–6. doi: 10.1016/s0022-1910(98)00104-8
- Nakabachi, A., Yamashita, A., Toh, H., Ishikawa, H., Dunbar, H. E., Moran, N. A., et al. (2006). The 160-kilobase genome of the bacterial endosymbiont *Carsonella*. *Science* 314:267. doi: 10.1126/science.1134196
- Nikoh, N., Hosokawa, T., Oshima, K., Hattori, M., and Fukatsu, T. (2011). Reductive evolution of bacterial genome in insect gut environment. *Genome Biol. Evol.* 3, 702–714. doi: 10.1093/gbe/evr064
- Ohkuma, M., Sato, T., Noda, S., Ui, S., Kudo, T., and Hongoh, Y. (2007). The candidate phylum 'Termite Group 1' of bacteria: phylogenetic diversity, distribution, and endosymbiont members of various gut flagellated protists. *FEMS Microbiol. Ecol.* 60, 467–476. doi: 10.1111/j.1574-6941.2007.00311.x
- Panwar, P., Allen, M. A., Williams, T. J., Hancock, A. M., Brazendale, S., Bevington, J., et al. (2020). Influence of the polar light cycle on seasonal dynamics of an Antarctic lake microbial community. *Microbiome* 8:116.
- Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2014). Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055. doi: 10.1101/gr.186072.114
- Parks, D. H., Rinke, C., Chuvochina, M., Chaumeil, P.-A., Woodcroft, B. J., Evans, P. N., et al. (2017). Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat. Microbiol.* 2, 1533–1542. doi: 10.1038/s41564-017-0012-7
- Parsons, J. B., and Rock, C. O. (2013). Bacterial lipids: metabolism and membrane homeostasis. *Prog. Lipid Res.* 52, 249–276. doi: 10.1016/j.plipres.2013.02.002
- Pérez-Brocal, V., Gil, R., Ramos, S., Lamelas, A., Postigo, M., Michelena, J. M., et al. (2006). A small microbial genome: the end of a long symbiotic relationship? *Science* 314, 312–313. doi: 10.1126/science.1130441
- Petroni, G., Spring, S., Schleifer, K. H., Verni, F., and Rosati, G. (2000). Defensive extrusive ectosymbionts of *Euplotidium* (Ciliophora) that contain microtubule-like structures are bacteria related to Verrucomicrobia. *Proc. Natl. Acad. Sci. U.S.A.* 97, 1813–1817. doi: 10.1073/pnas.030438197
- Pritchard, A. E., Seilhamer, J. J., Mahalingam, R., Sable, C. L., Venuti, S. E., and Cummings, D. J. (1990). Nucleotide sequence of the mitochondrial genome of *Paramecium*. *Nucleic Acids Res.* 18, 173–180.
- Ricci, F., Lauro, F. M., Grzymalski, J. J., Read, R., Bakiu, R., Santovito, G., et al. (2017). The anti-oxidant defense system of the marine polar ciliate *Euplotes nobilii*: characterization of the MsrB gene family. *Biology* 6:4. doi: 10.3390/biology6010004
- Roberts, N. J., Burton, H. R., and Pitson, G. A. (1993). Volatile organic compounds from Organic Lake, an Antarctic, hypersaline, meromictic lake. *Antarctic Sci.* 5, 361–366. doi: 10.1017/s0954102093000483
- Rodriguez-R, L. M., and Konstantinidis, K. T. (2016). The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. *PeerJ* 4:e1900v1.
- Sabree, Z. L., Huang, C. Y., Okusu, A., Moran, N. A., and Normark, B. B. (2013). The nutrient supplying capabilities of *Uzinura*, an endosymbiont of armored scale insects. *Environ. Microbiol.* 15, 1988–1999. doi: 10.1111/1462-2920.12058
- Santos-García, D., Latorre, A., Moya, A., Gibbs, G., Hartung, V., Dettner, K., et al. (2014). Small but powerful, the primary endosymbiont of moss bugs, *Candidatus Evansia muelleri*, holds a reduced genome with large biosynthetic capabilities. *Genome Biol. Evol.* 6, 1875–1893. doi: 10.1093/gbe/evu149
- Sato, T., Hongoh, Y., Noda, S., Hattori, S., Ui, S., and Ohkuma, M. (2009). *Candidatus Desulfobivrio trichonymphae*, a novel intracellular symbiont of the flagellate *Trichonympha agilis* in termite gut. *Environ. Microbiol.* 11, 1007–1015. doi: 10.1111/j.1462-2920.2008.01827.x
- Sato, T., Kuwahara, H., Fujita, K., Noda, S., Kihara, K., Yamada, A., et al. (2014). Intranuclear verrucomicrobial symbionts and evidence of lateral gene transfer to the host protist in the termite gut. *ISME J.* 8, 1008–1019. doi: 10.1038/ismej.2013.222
- Scheuermayer, M., Gulder, T. A. M., Bringmann, G., and Hentschel, U. (2006). *Rubritalea marina* gen. nov., sp. nov., a marine representative of the phylum Verrucomicrobia, isolated from a sponge (Porifera). *Int. J. Syst. Bacteriol.* 56, 2119–2124. doi: 10.1099/ijs.0.64360-0
- Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153
- Selbach, B. P., Chung, A. H., Scott, A. D., George, S. J., Cramer, S. P., and Dos Santos, P. C. (2014). Fe-S cluster biogenesis in gram-positive bacteria: sufU is a zinc-dependent sulfur transfer protein. *Biochemistry* 53, 152–160. doi: 10.1021/bi4011978
- Serra, V., Gammuto, L., Nitla, V., Castelli, M., Lanzoni, O., Sasser, D., et al. (2020). Morphology, ultrastructure, genomics, and phylogeny of *Euplotes vanleeuwenhoekii* sp. nov. and its ultra-reduced endosymbiont "*Candidatus Pinguicoccus supinus*" sp. nov. *Sci. Rep.* 10:20311.
- Shinzato, C., Inoue, M., and Kusakabe, M. (2014). A snapshot of a coral "holobiont": a transcriptome assembly of the scleractinian coral, porites, captures a wide variety of genes from both the host and symbiotic zooxanthellae. *PLoS One* 9:e85182. doi: 10.1371/journal.pone.0085182
- Smith, D. G., Gawryluk, R. M., Spencer, D. F., Pearlman, R. E., Siu, K. W., and Gray, M. W. (2007). Exploring the mitochondrial proteome of the ciliate protozoan *Tetrahymena thermophila*: direct analysis by tandem mass spectrometry. *J. Mol. Biol.* 374, 837–863. doi: 10.1016/j.jmb.2007.09.051
- Stephens, J. L., Lee, S. H., Paul, K. S., and Englund, P. T. (2007). Mitochondrial fatty acid synthesis in *Trypanosoma brucei*. *J. Biol. Chem.* 282, 4427–4436. doi: 10.1074/jbc.m609037200
- Stingl, U., Radek, R., Yang, H., and Brune, A. (2005). "Endomicrobia": cytoplasmic symbionts of termite gut protozoa form a separate phylum of prokaryotes. *Appl. Environ. Microbiol.* 71, 1473–1479. doi: 10.1128/aem.71.3.1473-1479.2005
- Strassert, J. F., Mikaelian, A., Woyke, T., and Brune, A. (2016). Genome analysis of "*Candidatus Ancillula trichonymphae*", first representative of a deep-branching clade of Bifidobacteriales, strengthens evidence for convergent evolution in flagellate endosymbionts. *Environ. Microbiol. Rep.* 8, 865–873. doi: 10.1111/1758-2229.12451
- Swart, E. C., Nowacki, M., Shum, J., Stiles, H., Higgins, B. P., Doak, T. G., et al. (2011). The *Oxytricha trifallax* mitochondrial genome. *Genome Biol. Evol.* 4, 136–154. doi: 10.1093/gbe/evr136



- Syberg-Olsen, M. J., Irwin, N. A. T., Vannini, C., Erra, F., Di Giuseppe, G., Boscaro, V., et al. (2016). Biogeography and character evolution of the ciliate genus *Euplotes* (Spirotrichea, Euplotia), with description of *Euplotes curdsi* sp. nov. *PLoS One* 11:e0165442. doi: 10.1371/journal.pone.0165442
- Takahashi, Y., Mitsui, A., Hase, T., and Matsubara, H. (1986). Formation of the iron-sulfur cluster of ferredoxin in isolated chloroplasts. *Proc. Natl. Acad. Sci. U.S.A.* 83, 2434–2437. doi: 10.1073/pnas.83.8.2434
- Tamura, K., and Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* 10, 512–526.
- Tamura, K., Stecher, G., Peterson, D., Filipi, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197
- Thao, M. L., Moran, N. A., Abbot, P., Brennan, E. B., Burckhardt, D. H., and Baumann, P. (2000). Cospeciation of psyllids and their primary prokaryotic endosymbionts. *Appl. Environ. Microbiol.* 66, 2868–2905.
- Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680. doi: 10.1093/nar/22.22.4673
- Tillett, D., and Neilan, B. A. (2000). Xanthogenate nucleic acid isolation from cultured and environmental cyanobacteria. *J. Phycol.* 36, 251–258. doi: 10.1046/j.1529-8817.2000.99079.x
- Tsaousis, A. D. (2019). On the origin of iron/sulfur cluster biosynthesis in eukaryotes. *Front. Microbiol.* 10:2478. doi: 10.3389/fmicb.2019.02478
- Tsaousis, A. D., Ollagnier de Choudens, S., Gentekaki, E., Long, S., Gaston, D., Stechmann, A., et al. (2012). Evolution of Fe/S cluster biogenesis in the anaerobic parasite *Blastocystis*. *Proc. Natl. Acad. Sci. U.S.A.* 109, 10426–10431. doi: 10.1073/pnas.1116067109
- Tschitschko, B., Erdmann, S., DeMaere, M. Z., Roux, S., Panwar, P., Allen, M. A., et al. (2018). Genomic variation and biogeography of Antarctic haloarchaea. *Microbiome* 6:113.
- Van Leuven, J. T., Mao, M., Xing, D. D., Bennett, G. M., and McCutcheon, J. P. (2019). Cicada endosymbionts have tRNAs that are correctly processed despite having genomes that do not encode all of the tRNA processing machinery. *Mbio* 10:e01950-18.
- Vandekerckhove, T. T. M., Coomans, A., Cornelis, K., Baert, P., and Gillis, M. (2002). Use of the verrucomicrobia-specific probe EUB338-III and fluorescent *in situ* hybridization for detection of “*Candidatus* Xiphinematobacter” cells in nematode hosts. *Appl. Environ. Microbiol.* 68, 3121–3125. doi: 10.1128/aem.68.6.3121-3125.2002
- Wagner, M., and Horn, M. (2006). The Planctomycetes, Verrucomicrobia, Chlamydiae and sister phyla comprise a superphylum with biotechnological and medical relevance. *Curr. Opin. Biotechnol.* 17, 241–249. doi: 10.1016/j.copbio.2006.05.005
- Wang, R. L., Miao, W., Wang, W., Xiong, J., and Liang, A. H. (2018). EOGD: the *Euplotes octocarinatus* genome database. *BMC Genomics* 19:63. doi: 10.1186/s12864-018-4445-z
- Wang, X., Quinn, P. J., and Yan, A. (2015). Kdo2-lipid A: structural diversity and impact on immunopharmacology. *Biol. Rev. Camb. Philos. Soc.* 90, 408–427. doi: 10.1111/brv.12114
- Wertz, J. T., Kim, E., Breznak, J. A., Schmidt, T. M., and Rodrigues, J. L. (2012). Genomic and physiological characterization of the Verrucomicrobia isolate *Geminisphaera colitermitum* gen. nov., sp. nov., reveals microaerophily and nitrogen fixation genes. *Appl. Environ. Microbiol.* 78, 1544–1555. doi: 10.1128/aem.06466-11
- Wu, M., Sun, L. V., Vamathevan, J., Riegler, M., Deboy, R., Brownlie, J. C., et al. (2004). Phylogenomics of the reproductive parasite *Wolbachia pipientis* wMel: a streamlined genome overrun by mobile genetic elements. *PLoS Biol.* 2:E69. doi: 10.1371/journal.pbio.0020069
- Yamao, F., Muto, A., Kawauchi, Y., Iwami, M., Iwagami, S., Azumi, Y., et al. (1985). UGA is read as tryptophan in *Mycoplasma capricolum*. *Proc. Natl. Acad. Sci. U.S.A.* 82, 2306–2309. doi: 10.1073/pnas.82.8.2306
- Yang, H., Schmitt-Wagner, D., Stingl, U., and Brune, A. (2005). Niche heterogeneity determines bacterial community structure in the termite gut (*Reticulitermes santonensis*). *Environ. Microbiol.* 7, 916–932. doi: 10.1111/j.1462-2920.2005.00760.x
- Yao, Z., Davis, R. M., Kishony, R., Kahne, D., and Ruiz, N. (2012). Regulation of cell size in response to nutrient availability by fatty acid biosynthesis in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 109, E2561–E2568.
- Yau, S., Lauro, F. M., DeMaere, M. Z., Brown, M. V., Thomas, T., Raftery, M. J., et al. (2011). Virophage control of Antarctic algal host-virus dynamics. *Proc. Natl. Acad. Sci. U.S.A.* 108, 6163–6168. doi: 10.1073/pnas.1018221108
- Yau, S., Lauro, F. M., Williams, T. J., Demare, M. Z., Brown, M. V., Rich, J., et al. (2013). Metagenomic insights into strategies of carbon conservation and unusual sulfur biogeochemistry in a hypersaline Antarctic lake. *ISME J.* 7, 1944–1961. doi: 10.1038/ismej.2013.69
- Yoon, J., Matsuo, Y., Katsuta, A., Jang, J. H., Matsuda, S., Adachi, K., et al. (2008). *Haloferula rosea* gen. nov., sp. nov., *Haloferula harenae* sp. nov., *Haloferula phyci* sp. nov., *Haloferula helveola* sp. nov. and *Haloferula sargassicola* sp. nov., five marine representatives of the family Verrucomicrobiaceae within the phylum ‘Verrucomicrobia’. *Int. J. Syst. Evol. Microbiol.* 58, 2491–2500. doi: 10.1099/ijs.0.2008/000711-0
- Yoon, J., Yasumoto-Hirose, M., Katsuta, A., Sekiguchi, H., Matsuda, S., Kasai, H., et al. (2007a). *Coralimargarita akajimensis* gen. nov., sp. nov., a novel member of the phylum ‘Verrucomicrobia’ isolated from seawater in Japan. *Int. J. Syst. Evol. Microbiol.* 57, 959–963. doi: 10.1099/ijs.0.64755-0
- Yoon, J., Yasumoto-Hirose, S., Matsuda, S., Nozawa, M., Matsuda, S., Kasai, H., et al. (2007b). *Pelagicoccus mobilis* gen. nov., sp. nov., *Pelagicoccus albus* sp. nov. and *Pelagicoccus litoralis* sp. nov., three novel members of subdivision 4 within the phylum ‘Verrucomicrobia’, isolated from seawater by *in situ* cultivation. *Int. J. Syst. Evol. Microbiol.* 57, 1377–1385. doi: 10.1099/ijs.0.64970-0
- Zientz, E., Dandekar, T., and Gross, R. (2004). Metabolic interdependence of obligate intracellular bacteria and their insect hosts. *Microbiol. Mol. Biol. Rev.* 68, 745–770. doi: 10.1128/mmr.68.4.745-770.2004

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Williams, Allen, Ivanova, Huntemann, Haque, Hancock, Brazendale and Cavicchioli. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.