Supplementary Materials for

A centimeter-long bacterium with DNA contained in metabolically active membrane-bound organelles

Jean-Marie Volland, Silvina Gonzalez-Rizzo, Olivier Gros, Tomáš Tyml, Natalia Ivanova, Frederik Schulz, Danielle Goudeau, Nathalie H Elisabeth, Nandita Nath, Daniel Udwary, Rex R Malmstrom, Chantal Guidi-Rontani, Susanne Bolte-Kluge, Karen M Davies, Maïtena R Jean, Jean-Louis Mansot, Nigel J Mouncey, Esther R Angert, Tanja Woyke, Shailesh V Date

This PDF file includes:

Other Supplementary Materials for this manuscript include the following:

Materials and Methods

Sampling

Samples of Large Sulfur Bacteria (LSB) were collected from a marine mangrove environment (ambient temperature 28°C) in "La manche à eau" in Guadeloupe, Lesser Antilles, at one site (16°16'40"N, 61°33'28"W) (*1*). Sunken leaves of *Rhizophora mangle* containing LSB were sampled by hand from the surface layer of the sediment (*c.* 1 m depth). Living LSB samples were processed within 1 h after collection under a dissecting microscope. The samples were then washed with 0.22 µm filtered seawater prior to use for molecular experiments. Individual bacteria were fixed at 4°C either for 4 hours in 4% paraformaldehyde in sterile seawater, or in 2.5% glutaraldehyde in 0.1 M cacodylate buffer (pH 7.2), which was made iso-osmotic (900 mOsmoles) with sea water by the addition of sodium chloride and calcium chloride. Samples were stored at 4°C until analysis.

Light microscopy

Samples were observed live or fixed under standard stereomicroscopes. If applicable, a series of images captured at different focus distances was merged using the focus stacking software Helicon Focus® (Fig. S1). Fixed samples were also observed using a light microscope Axio Observer.D1 (Carl Zeiss, Jena, Germany) equipped with a black-and-white high-resolution camera (AxioCam MRm, Carl Zeiss, Jena, Germany).

Ultrastructural analysis

For conventional SEM analysis, samples were briefly rinsed in the cacodylate buffer and then dehydrated through a graded acetone series before drying under CO_2 using a critical point dryer machine (EM CPD300, Leica). The samples were then sputter-coated with gold (Sputter Coater SC500, BioRad) before observation at 20 kV with an FEI Quanta 250 scanning electron microscope.

For scanning transmission electron microscopy (STEM) analysis, prefixed bacterial filaments were washed twice in 0.1 M sodium cacodylate buffer to remove aldehydes before fixation in 1% osmium tetroxide for 45 min at room temperature. Samples were then rinsed in distilled water and post-fixed with 2% aqueous uranyl acetate for 1 h at room temperature. After three washes with distilled water, each sample was dehydrated through a graded acetone series and embedded in Epon-Araldite according to Glauert (1975) (*2*). Thin sections (60 nm thick) were stained for contrast for 30 min in 2% aqueous uranyl acetate before examination with a Quanta 250 (FEI - STEM Mode).

For Transmission Electron Microscopy (TEM) observations, glutaraldehyde fixed samples were washed in a cacodylate buffer and post-fixed with osmium tetroxide as described above. They were then cryo-immobilized using a BAL-TEC HPM 010 high-pressure freezer and further placed in a freeze-substitution medium made of 1% osmium tetroxide, 0.1% uranyl acetate, and 5% ddH_2O in acetone. The samples were freeze-substituted following the super quick procedure described in McDonald (2011) (*3*). The substitution medium was washed away with pure acetone and the samples were infiltrated and flat-embedded in Epon-Araldite resin as described in Müller-Reichert et al. (2003) (*4*). Thin sections of 70 nm were mounted on formvar coated slot grids and stained for 4 min in 2 % uranyl acetate followed by 4 min in Reynolds Lead Citrate. Slot grids were observed with a FEI Tecnai 12 or a Jeol 1400FLASH TEM. Montages were acquired either manually or automatically using the SerialEM software. Tile images were assembled either manually using GIMP, or with the Etomo program from the Imod suite (*5*), or with the software Image Composite Editor (Microsoft), or with the FIJI package on ImageJ (*6*). Morphometric measurements were performed using FIJI. The average thickness of the cytoplasm was obtained from 118 measurements realized on

sections from three different cells (cell 1: n = 51; cell 2: n = 17; cell 3: n = 50). The average diameter of the elemental sulfur granules was obtained from 338 measurements realized on sections from three different cells (cell 1: n = 103; cell 2: n = 31; cell 3: n = 204). The average diameter of pepins was obtained from 92 measurements realized on sections from five different cells (cell 1: n = 8; cell 2: n = 29; cell 3: n = 20; cell 4: n = 5; cell 5: n = 30).

Elemental analysis

The elemental composition of fully hydrated samples was analyzed using an Environmental Scanning Electron Microscope (ESEM). The dehydration step was avoided because the elemental sulfur S_8 is soluble in alcohol and acetone. The samples were fixed in 2.5% glutaraldehyde in seawater and kept in the same solution until examination. Samples were simply washed quickly in distilled water to remove salts and then introduced to an ESEM (FEI Quanta FEG) under a pressure of 650 Pa. ESEM studies were carried out using various acceleration voltages to reveal the presence of sulfur in the samples, as well as their morphology. We used 1) 15 kV for back scattered electron images (Z contrast) and energy dispersive X-ray spectroscopy (analyses and elemental mapping) and 2) 3 kV for secondary electron images (sample morphology).

Confocal Laser Scanning Microscopy (CLSM)

Paraformaldehyde-fixed *Ca.* T. magnifica cells (n=8) were washed three times in sterile seawater, dehydrated through an ascending ethanol series - to dissolve elemental sulfur granules - and rehydrated through a descending ethanol series. Individual cells were transferred onto glass slides equipped with Gene Frame (Thermofisher) to avoid crushing the cells with the coverslips. Membrane labeling of cells was carried out using the lipophilic dye FM 1-43x (Molecular Probes, Eugene, OR, USA) according to the manufacturer's protocol. DNA was labelled with DAPI (4′,6-Diamidino- 2-phenylindole dihydrochloride; Millipore Sigma) following the manufacturer's protocol. We imaged 8 cells on a Zeiss LSM 710 microscope by acquiring multiple overlapping Z-stack images (tiles) with the Zen software (Zeiss). From these 8 cells, 6 were imaged in their entire length (cells F to K, see Table S2) and were used to confirm the single cell nature of the filaments, measure precise morphometric parameters (length, minimum and maximum diameters), and assess polyploidy level for one of them. The remaining 2 cells were imaged only partially (cells L and M, see Table S2) and used exclusively to assess the polyploidy level by counting DAPI signal clusters. As positive controls, we followed the same protocol and prepared multicellular filaments of the Cyanobacteria *Microcoleus vaginatus* from a pure culture, *Beggiatoa*-like filaments collected in a Guadeloupe mangrove, and *Marithrix*-like filaments collected at the White Point Beach hydrothermal vent (*7, 8*) (Fig. S4; Movie S4). All 3D data sets were imported into the ORS Dragonfly software for stitching, 3D rendering and morphometric analyses and/or polyploidy analysis. Three cells observed in their entirety with high lateral resolution (cells F, G and H) were segmented using the ORS Dragonfly deep learning tool and manual segmentation tools. Precise volumes of the central vacuole and the cytoplasm were measured (see Table S2).

Hard x-ray computed tomography

We used hard x-ray computed tomography to visualize six cells in three dimensions with isotropic resolution. Four cells were observed in their entirety and two were observed only partially (see Table S2). *Ca.* T. magnifica cells fixed with glutaraldehyde and post-fixed with osmium tetroxide, as described above, were washed three times in sterile seawater, dehydrated through ascending ethanol series and stored in 70 % ethanol at 4° C. Before analysis, cells were re-hydrated and immobilized inside plastic capillaries with 1 % low melting point agarose. After sealing the capillaries on both ends, we glued them onto the head of a sewing pin and imaged them in a Zeiss Xradia 520 Versa x-rays microscope. The technical details of each scan are provided in Table S1. The CT scans were reconstructed using the Zeiss software and further

imported in the ORS Dragonfly software for stitching of the tiles, segmentation, and analysis. Four cells were observed in their entirety (cells A, B, D and E), three of which were segmented using the ORS Dragonfly deep learning tool and manual segmentation tools. Precise total volumes of the cells were measured and for one of them (cell D) the volumes of the cytoplasm and central vacuole were measured as well (see Table S2).

Fluorescent in situ hybridization (FISH) and TEM correlation

Paraformaldehyde fixed cells (n=4) were washed in seawater before dehydration through an ascending ethanol series. We then infiltrated the cells with medium grade LR-White resin. LR-White embedded cells were polymerized at 40 °C under anaerobic conditions for three days. We analyzed with FISH a total of 84 semi-thin sections (500 nm), coming from 7 different areas analyzed in triplicates within each of the four cells. Sections were mounted on PTFE coated microscope glass slides (Electron Microscopy Sciences, Hatfield, PA) and analyzed with FISH. Some consecutive thin sections (70 nm) were prepared for correlative TEM as described in the ultrastructural analysis section. The FISH hybridization solution (0.9 mol L-1 NaCl, 20 mmol L-1 Tris/HCl pH 8.0, 0.01% SDS, 10% formamide) was applied onto the section in a 20 µL drop containing 0.5 µM of each oligonucleotide probe. Hybridization was performed in a humid chamber at 46°C for 3 h. Washing was performed under stringent conditions at 48°C for 15 min (*9*). We used a combination of a eubacterial probe mixture (EUB338 Alexa Fluor® 488 single labeled) (*10*), a general gammaproteobacteria probe (Gam42a Cy3 double-labeled) (*9*), an unlabeled competitor probe (BET42a) (*9*) and a genus-specific *Thiomargarita* probe (Thm482, Cy5 double-labeled) (*11*). Nonsense probes (Non-EUB) (*12*) labeled in Cy3 and Cy5 were also applied to all slides to control for false positive signals due to autofluorescence or nonspecific probe binding, but no signals were observed in these controls. We mounted the FISH slides with antifadent solution CitiFluor AF1 Plus DAPI (Electron Microscopy Sciences, Hatfield, PA). Micrographs were taken using a 63x oil-immersion objective on an Inverted epifluorescence microscope (Axio Observer.D1, Carl Zeiss, Jena, Germany) equipped with a black-and-white high-resolution camera (AxioCam MRm, Carl Zeiss, Jena, Germany). FISH images were overlaid with their corresponding TEM observations in the GIMP software.

Immunohistochemistry

Paraformaldehyde fixed cells (n=4) embedded in LR-White resin as described for FISH above were used for immunolabelling of ATP synthase. Semi-thin sections (500 nm) were submerged in a blocking solution (PBS1X, 3% bovine serum albumin) prior to incubation with primary antibody (anti-ATP synthase subunit β, mouse IgG1clone 3D5, 1:10 dilution, Invitrogen) at 37 °C for 2 hours. Following the washing procedure, sections were incubated with Cy5 labeled secondary antibody (Goat anti-mouse IgG, 1:20 dilution, Invitrogen) for 2 h at 37 °C, and counterstained with DAPI (Electron Microscopy Sciences, Hatfield, PA). Negative controls were performed by omitting the primary antibody and showed no signal. Three independent experiments were conducted for which representative images are shown in Figs. 2K-M and S20.

Bioorthogonal non-canonical amino acid tagging (BONCAT)

Live cells were freshly collected from mangrove and isolated from their dead leaf substrate. They were then transferred into closed glass vials and incubated 24 hours in sulfidic seawater (100 µM H_2S in 0.2 µm-filtered natural seawater) supplemented with 50 µM l-homopropargylglycine (HPG), a clickable analogue of l-methionine. Negative controls were incubated in the same manner but omitting the HPG. The incubation was halted by three washes in sterile seawater. The click-reaction mixture was prepared fresh each time by mixing 18.23 µl of the dye premix (CuSo4 100 µM, tris-hydroxypropyltriazolylmethylamine 500 µM, AZDye 488 Azide 5 µM) with 980 µl of the reaction buffer (sodium ascorbate 5 mM, aminoguanidine HCl 5 mM in PBS 2.5X). All reagents were purchased from Click Chemistry Tools (Click

Chemistry Tools, Scottsdale, AZ, USA). The cells were incubated in the click reaction mixture for 1 hour and washed (3x5 min) in PBS 2.5X before observation under a LSM TCS SPE Leica Microscope. A total of 12 cells were observed from three different experiments following three samplings, showing consistent fluorescent patterns. Negative controls were included for each experiment and did not show any labelling.

Sulfide measurements

Sunken leaves with attached elongated cells were brought to the laboratory and placed in a mesocosm mounted the day before measurement to simulate their mangrove environment. The samples were collected from the field on the day of measurement. Sulfide measurements were carried out using H_2S100 microsensors (Unisense®) attached to a micromanipulator (type MD4 Rechts, Märzhäuser®). One microsensor was placed into the water column 5 cm above the bacterial cells and the other positioned in the middle of the LSB "bouquet" attached onto a submerged leaf of *Rhizophora mangle*. The measurements were recorded every 30 s using SensorBasic® software. Calibrations were performed according to the Unisense® instructions. The pH was measured with an autonomous probe (NKE), similar to that described by Le Bris *et al.*, (2001) (*13*), fixed to the micromanipulator. Total sulfide concentrations (S^{2-}_{tot}= H_2S+HS^- +S^{2-}) were calculated, accounting for the measured pH and salinity using a pK of 6.51 (*14*).

Genome sequencing, assembly, binning, and annotation

We processed five *Ca.* T. magnifica filaments for single-cell genomic sequencing. Within one hour after sampling, we dissected individual cells out of the decaying leaf and washed them three times in sterile seawater before storing them at -80°C. We thawed each individual filament and immediately amplified the genomic DNA by multiple displacement amplification using the REPLI-g kit (Qiagen). DNA libraries were created from 200pg of DNA from each of the amplified products using Nextera XT DNA library creation kit (Illumina). We sequenced the DNA libraries on an Illumina Nextseq High Output platform. We then imported pair-end reads (2x150 bp) into the KBase platform (www.kbase.us) (*15*). In KBase, we used SPAdes (v3.13.0) to assemble reads into contigs of at least 500 bp (using kmers of 33, 67, 99, 125 bp). We then binned contigs over 2000 bp using MetaBAT2 resulting in 2 to 6 bins per filament (see Supplementary Table 3). Only one bin per filament was taxonomically identified as *Thiomargarita* by the GTDB-Tk classify app (v0.1.4). For each of the five assemblies we extracted the contigs from the *Thiomargarita* bin and treated them as genome assemblies for further analyses (referred to as: filament 1 genome, filament 2 genome, filament 3 genome, filament 4 genome and filament 5 genome). We assessed genome qualities with CheckM (v1.0.18) (Table S4).

Genome analysis

Average Nucleotide Identities and within filament clonality assessment

We choose filament #5 as a reference genome based on its assembly statistics. We computed pairwise Average Nucleotide Identities (ANIs) with FastANI. We assessed within filament clonality by mapping the reads from filament #1, #2, #3 and #4 onto filament #5 genome using BBMap (v38.79) (https://sourceforge.net/projects/bbmap/) with the flags minid=0.95 minaveragequality=30, and called variants (or single nucleotide polymorphisms, SNPs), with the BBTools scripts pileup.sh and callvariants.sh and the flags minreads=2 minquality=30 minscore=30 minavgmapq=20 minallelefraction=0.1.

Phylogenomics

A set of 56 universal single copy marker proteins (*16, 17*) was used to build a phylogenetic tree of the filament assemblies and related gammaproteobacterial genomes available in the IMG/M database (*18*). Marker proteins were identified with hmmsearch (version 3.1b2, hmmer.org) using a specific HMM for

each of the markers. For every marker protein, alignments were built with MAFFT (v7.294b) (*19*) and subsequently trimmed with BMGE (*20*) using BLOSUM30. Single protein alignments were then concatenated and maximum likelihood phylogenies were inferred with iq-tree v2.0.3 (*21*) using the LG4X+F model. The tree was visualized in itol (*22*).

Secondary metabolism

Bacterial genomes were analyzed for secondary metabolite Biosynthetic Gene Clusters (BGCs) with antiSMASH v5.1.2 (*23*). Fungal BGC data for the fungus model system *Aspergillus nidulans* were retrieved from antismash-db. %BGC was calculated by summing the nucleotide sequence length of each antiSMASH BGC region and dividing by total genome size.

Genome annotation and functional analysis

Ca. T. magnifica genomes were annotated using a JGI prokaryotic structural and functional genome annotation pipeline (https://img.jgi.doe.gov/docs/pipelineV5/) and loaded into IMG/MER database (*18*). Assignments of proteins to protein families, such as Pfam v.30 (*24*) and KEGG v.77.1 (*25*) in conjunction with the tools provided by IMG user interface were used to infer functional capabilities encoded by the genomes and to visualize chromosomal neighborhoods of genes of interest. Protein sequences of interest were exported from IMG/MER and alignments were built with MAFFT (v7.294b) (*19*).

Supplementary Text

Evidence of thiotrophy in *Ca.* T. magnifica

Mangrove swamps accumulate fine sediment with high organic content. Under anoxic conditions sulfate reducing bacteria degrade organic matter producing large amounts of sulfide and sustaining sulfur oxidizing chemoautotrophic (thiotrophic) microbial communities. In Guadeloupean mangroves, sulfide produced by the reduced sediment ranges from 0.19 mM to 2.40 mM (*26*). It passively diffuses to the overlaying water where it gets rapidly oxidized by oxygen creating a steep redox gradient at the sediment/water interface. In order to characterize the microenvironment of *Ca.* T. magnifica, we brought to the laboratory filaments still attached to sunken leaves and immersed the leaves in a mesocosm-like setup on top of reduced mangrove sediment. Using microsensors we measured high and relatively stable concentrations of reduced sulfur species in the microenvironment of the filaments with concentrations fluctuating between 1.20 to 1.79 mM (Fig. S13) while no sulfide was detectable in the overlaying water.

Genomics and oxidoreductase activity experiments have shown that other *Thiomargarita* spp. can use reduced sulfur species as electron donors (*27-29*). Thiotrophic gammaproteobacteria from sulfidic environments are known to store elemental sulfur in membrane bound vesicles (*30, 31*). In order to determine if *Ca.* T. magnifica filaments display an active thiotrophic metabolism, we placed lightly fixed fully hydrated cells in an environmental scanning electron microscope and interrogated the presence of internal sulfur granules using Energy Dispersive X-ray Spectroscopy (EDXS). Images collected by secondary and backscattered electron detectors clearly showed the presence of bright round-shaped areas of on average 2.33 ± 0.46 µm in diameter (Fig. S13B and C). EDXS microanalysis (Fig. S13A) and sulfur mapping (Fig. S13F) clearly show that these areas are sulfur-rich and therefore correspond to sulfur globules (also visible on the TEM as numerous electron lucent vesicles 2.40 ± 1.03 µm in diameter).

Estimation of the genome copy number

We analyzed in 3D three datasets coming from three different cells after DNA was fluorescently labeled with DAPI. We visualized an entire 2.39 mm long cell (cell H, Table S2), as well as two cell portions of 195 µm (cell L, Table S2) and 660 µm (cell M, Table S2). After segmentation of the DNA clusters labeled with DAPI, we used the ORS Dragonfly© multi-ROIs analysis tool and counted 7647, 1518 and 3910 individual objects for the three cells respectively. The volume of individual DNA clusters ranged from 0.48 to 3511 μm^3. We assumed that the smallest DAPI stained DNA clusters, with a volume of 0.48 μm^3, represented a single genome copy and observed that larger clusters represented multiple chromosomes too close to one another to be segmented as individual objects. Therefore, we divided larger DNA clusters by the volume of the smallest ones (0.48 μm^3) and estimated the total number of genome copies for each data set. We then divided the total number of genome copies by the length of the cell analyzed allowing us to express the polyploidy level per millimeter of filament and extrapolate to a fully grown 20 mm long cell (see Table S2). We estimated that *Ca.* T. magnifica cells present on average 36,880 ± 7,956 genome copies per millimeter of filament and extrapolated that fully grown 20 mm cells would therefore present 737,598 ± 159,115 genome copies.

FISH investigations

The specificity of the FISH probe Thm482 was confirmed by the TestProbe tool from the SILVA database. In addition, we extracted all 16S rRNA gene sequences from the published *Thiomargarita*

metagenome as well as from our five *Ca.* T. magnifica single-cell assemblies and further confirmed that the Thm482 probe only matched the *Thiomargarita* 16S rRNA sequence.

The FISH investigations were conducted on 4 different cells. From each cell, triplicate sections from 7 different locations consistently showed the same results, representative exemplars of which are shown in Figs. 2 and S7 to S9.

Extended description of *Ca.* T. magnifica metabolism based on the genome analysis

The large genomes of *Ca*. T. magnifica filaments encoded a wide range of metabolic capabilities, some unique and some shared with other large sulfur-oxidizing bacteria. A complete glycolysis pathway was present, which included a gene for glucose-6-phosphate isomerase; this gene was reported to be absent from previously sequenced single cells of *Ca*. T. nelsonii (Table S9) (*29*). The tricarboxylic acid cycle is also complete and can partially function in both directions due to the presence of 2-oxoglutarate dehydrogenase complex and 2-oxoglutarate-ferredoxin oxidoreductase, as well as succinate dehydrogenase and fumarate reductase. Multiple anaplerotic enzymes are also encoded (Table S9). *Ca*. T. magnifica genomes encoded RuBisCO form I (Fig.S15) and other enzymes for carbon fixation via the Calvin–Benson–Bassham (CBB) cycle. Similar to *Ca.* T. nelsonii, sedoheptulose-1,7-bisphophatase and the fructose-1,6-bisphosphatase were missing and may have been replaced by a polyphosphate-dependent 6-phosphofructokinase (*29*).

While *Ca.* T. nelsonii were found to perform denitrification or dissimilatory and assimilatory reduction of nitrate to ammonia, the only enzymes found in *Ca.* T. magnifica genomes were dissimilatory/respiratory nitrate reductases, Nar and Nap. Although there were genes encoding proteins similar to NorD and NorQ, accessory proteins of nitric oxide reductase, its catalytic subunits NorCB were missing, and so were genes encoding assimilatory nitrate reductase, both types of nitrite reductase (NirBD and NirS), and nitrous oxide reductase. Instead, a complete set of urease subunits was present (Table S9). This suggests that *Ca.* T. magnifica likely obtains ammonia for growth from organic sources, whereas nitrate is used exclusively as an electron acceptor.

Like other large sulfur-oxidizing bacteria, *Ca.* T. magnifica genomes encoded a set of genes for sulfur and hydrogen oxidation. Genes encoding two Ni-Fe hydrogenases with a complement of accessory proteins were found. Both sulfide-quinone-oxidoreductase Sqr and flavocytochrome c sulfide dehydrogenase FccAB were present, reverse dissimilatory sulfite reductase pathway, as well as thiosulfate-oxidizing sox genes. As in *Ca.* T. nelsonii, a group I catalytic intron was inserted into the *Ca.* T. magnifica *dsrA* gene encoding alpha subunit of dissimilatory sulfite reductase (*28*). *Ca.* T. magnifica genomes encoded a branched respiratory electron transport chain, which included NADH dehydrogenase, succinate dehydrogenase, ubiquinol-cytochrome c reductase, cbb3-type cytochrome c oxidase, and cytochrome d ubiquinol oxidase, as well as heterodisulfide reductase and sodium-transporting Rnf complex. The genomes also encoded both F-type and V-type ATPases.

In addition to the common bacterial secretion systems, a type VI secretion system (T6SS) was present in all *Ca.* T. magnifica genomes (Table S10), and some of its components were found in *Ca.* T. nelsonii. Often found in bacterial pathogens, this secretion system is analogous to contractile tails of bacteriophages and is known to mediate contact-dependent killing of neighboring cells via intracellular delivery of toxic effectors (*32*). T6SS consists of an inner tube composed of Hcp protein capped with VgrG and proline-alanine-alanine-arginine (PAAR) repeat-containing proteins surrounded by a sheath made of TssB/TssC heterodimers (also known as ImpB/ImpC or VipB/VipC). Other components of T6SS include a transmembrane complex composed of three subunits, TssJ/TssL/TssM (also known as ImpJ/ImpK/ImpL) and a baseplate, which is connected to the transmembrane complex and serves as a

platform for inner tube and sheath polymerization. The injection process presumably starts with rearrangement of baseplate components leading to sheath contraction, opening of the baseplate and release of the inner tube and its payload into the target cell. Known T6SS effectors include peptidoglycan hydrolases, phospholipases, nucleases, ADP-ribosyltransferases, and pore-forming proteins (*33*); however, the vast majority of T6SS effectors remain unknown. Some of them can be identified based on the presence of VgrG or PAAR domain in their N-terminus, while others are characterized by the presence of other signature motifs, such as rearrangement hotspots (RHS), YD repeats, MIX motifs and FIX motifs (*34*). None of the known T6SS effectors were immediately recognizable in *Ca.* T. magnifica genomes; however, these genomes did encode between 8 and 22 RHS/YD repeat proteins (Table S11), as compared to 0 to 5 in closely related genomes. Three genes encoding RHS proteins were located near the T6SS component *vgrG*. The large variability of RHS protein count in *Ca.* T. magnifica may represent natural variation in the population or alternatively be an artifact of assembly due to the length and repetitive nature of these proteins. It is likely that T6SS found in *Ca.* Thiomargarita genomes confers a competitive advantage in the course of inter- and intraspecific conflict.

Among protein families, which have undergone significant expansion in *Ca.* T. magnifica, there were multiple families associated with genome rearrangements including mobile genetic elements, introns, and site-specific recombinases. *Ca.* T. nelsonii single cells possess the same genomic features (*28, 29*). In addition, *Ca.* T. magnifica genomes were riddled with a myriad of toxin/antitoxin systems (Table S11). While some of these systems may be involved in maintaining genetic material, such as mobile genetic elements, others may play regulatory roles in stress response, adaptation, as well as contribute to the complex lifestyle (*35*).

Expansion of other protein families have also been attributed to complex lifestyle and morphological changes: in addition to metacaspases (PF00656) described by Flood et al., 2016 (*28*), these include multiple copies of protein kinase domain (PF00069) and another family of peptidases and inactivated derivatives called CHAT (PF12770) (*36*). Curiously, nearly a third of caspase domains are associated with another highly overrepresented domain, called FGE-sulfatase (PF03781). Similarly, nearly a third of protein kinase domains are associated with FGE-sulfatase. This domain is known to generate formylglycine, which is the catalytic residue of sulfatases, and inactivation of a human formylglycine-generating enzyme SUMF1 leads to a multiple sulfatase deficiency (*37*). In bacteria characterized representatives of this family are iron(II)-dependent oxidoreductases EgtB and OvoA catalyzing C-S bond formation in the biosynthesis of ergothioneine and ovothiols (*38*). One of the FGE-sulfatase family proteins in *Ca.* T. magnifica is a likely bifunctional OvoA protein with 5-histidylcysteine sulfoxide synthase and mercaptohistidine methyltransferase activities. However, the function of more than a hundred copies of a gene encoding FGE-sulfatase in these genomes is unclear. A few copies of a gene encoding a predicted sulfatase (PF00884) were present in *Ca.* Thiomargarita genomes, but its abundance was not nearly enough to explain the expansion of the FGE-sulfatase family. Given its association with other domains implicated in complex life cycle and cell morphology, we hypothesize that the FGE-sulfatase family in *Ca.* Thiomargarita is also involved in these processes.

Another remarkable feature of *Ca.* T. magnifica was a significant rearrangement of the core genes involved in cell division and morphogenesis. Whereas all genes necessary for lipid-linked peptidoglycan monomer biosynthesis and export were present (Table S8), many core cell division proteins involved in Z ring assembly and regulation were missing. Of the main Z ring components, only a cytoskeletal protein FtsZ was found; protein FtsA, which tethers FtsZ to the membrane, and protein ZipA interacting with FtsZ and essential division peptidoglycan synthases were absent. FtsZ is part of well-conserved dcw ("division and cell wall") operon, which in most other large sulfur-oxidizing bacteria also includes FtsA and a key late divisome protein FtsQ along with peptidoglycan monomer synthesis enzymes MurC, MurG, and Ddl. The overall structure of dcw operon was preserved in *Ca.* T. magnifica and *Ca.* T. nelsonii Bud S10, but FtsA and FtsQ were conspicuously missing (Fig. 3B). Consistent with the absence

of FtsA, a widely conserved complex FtsE-FtsX, which plays many roles in promoting Z ring assembly including regulation of FtsA dynamics, was also missing from *Ca.* Thiomargarita genomes. In contrast, ZapA and ZapD proteins, which interact with FtsZ to promote Z ring assembly and stability, were present, and so was ZapA-interacting protein ZapB.

Even more remarkably, none of the late divisome proteins were found in *Ca.* Thiomargarita genomes. These include peptidoglycan polymerizing glycosyltransferase FtsW, peptidoglycan transpeptidase FtsI (peptidoglycan-binding protein 3), a key divisome complex FtsQLB, which recruits and regulates peptidoglycan synthesis activities, and FtsK protein, which bridges Z ring with late divisome components (*39*). We cannot exclude the possibility that the lack of these proteins was due to the assembly gaps, but the fact that not a single one of them was found in any of *Ca.* T. magnifica or *Ca.* T. nelsonii genomes suggests that they are truly missing and not merely an artifact of incomplete sequences.

While the complement of divisome proteins encoded by *Ca.* T. magnifica genomes was greatly reduced, components of cell elongation complex have been duplicated, which likely happened in the ancestor of *Thiomargarita* spp. (Fig. S16-18). Elongasome components, which include cytoskeletal protein MreB required for maintenance of cylindrical cell shape, its interacting partners MreC, MreD and regulator of polymerization RodZ (Fig. S16), as well as peptidoglycan polymerizing glycosyltransferase RodA and peptidoglycan transpeptidase MrdA (peptidoglycan-binding protein 2), were organized in *Ca.* Thiomargarita into 2 chromosomal clusters. The first included MreBCD proteins, RodA and MrdA, as well as membrane-bound lytic murein transglycosylase MltB and rare lipoprotein A (Fig. 3B). The second included RodZ and genes unrelated to elongasome, such as 4-hydroxy-3-methylbut-2-enyl-diphosphate synthase ispG (Fig. 3B). Comparison of *mre* and *rodZ* regions of *Ca.* Thiomargarita to those of other large sulfur-oxidizing bacteria revealed duplications of *rodZ*, *mreD* and *mrdA*, with both copies adjacent on the chromosome. While the physiological effect of this increase of gene dosage is unclear, it is likely related to the complex morphology and life cycle of *Ca.* T. magnifica. Alternatively, they may compensate for the lack of peptidoglycan synthesis components of the divisome.

Developmental cycle of *Ca.* T. magnifica

Because we observed fully grown filaments releasing their most apical segment live in the lab, we hypothesized that the released terminal segments represent a dispersive stage of the developmental cycle (Fig. 1C, S1). To test this hypothesis, we quantified the biovolume of the apical buds and compared it to the biovolume of a small filament (since the bacterium is uncultured). The smallest filament analyzed was observed with hard x-ray tomography and had a volume of 2.37×10^{-13} m^3 which corresponds almost exactly to the volume of terminal buds from the apical pole of the fully-grown filament (2.1 and 2.4×10^{-13} m^3). It is likely that the rod-shaped terminal segment of filaments represents dispersive daughter cells which get released to the environment and eventually attach and grow on a new substrate. The newly attached daughter cell apparently first reshapes itself into a thin filamentous cell while conserving the same volume. Young filaments then grow in the vertical direction with the apical pole starting to constrict to form new buds (Fig. 1C, S1).

Based on field observations (by O. Gros, co-author), the *Ca.* T. magnifica developmental cycle may be somewhat analogous to *Zoothamnium niveum,* a giant colonial ciliate that sometimes co-occurs on the same substrate. Like *Ca*. T. magnifica, elongated colonies of *Z. niveum* grow up to 1.5 cm high, emerging above the competitive biofilm to provide an optimal access to hydrogen sulfide and oxygen to its sulfur-oxidizing gammaproteobacterial symbiont (*40*). Like *Ca.* T. magnifica, specialized cells eventually detach from the colony, disperse, settle in favorable environments, and grow into colonies. We observed co-occurring *Ca.* T. magnifica cells and *Z. niveum*, sometimes on the same decaying leaf. The gigantism and life cycle of *Ca.* T. magnifica may therefore allow them to exploit a niche so far known to be occupied

only by Eukaryotes such as *Z. niveum.* Just like the sulfur-oxidizing ectosymbionts escape the biofilm competition by growing vertically through their association with the ciliate, *Ca.* T. magnifica grows above the biofilm where it is still exposed to high sulfide concentrations (Fig. S14).

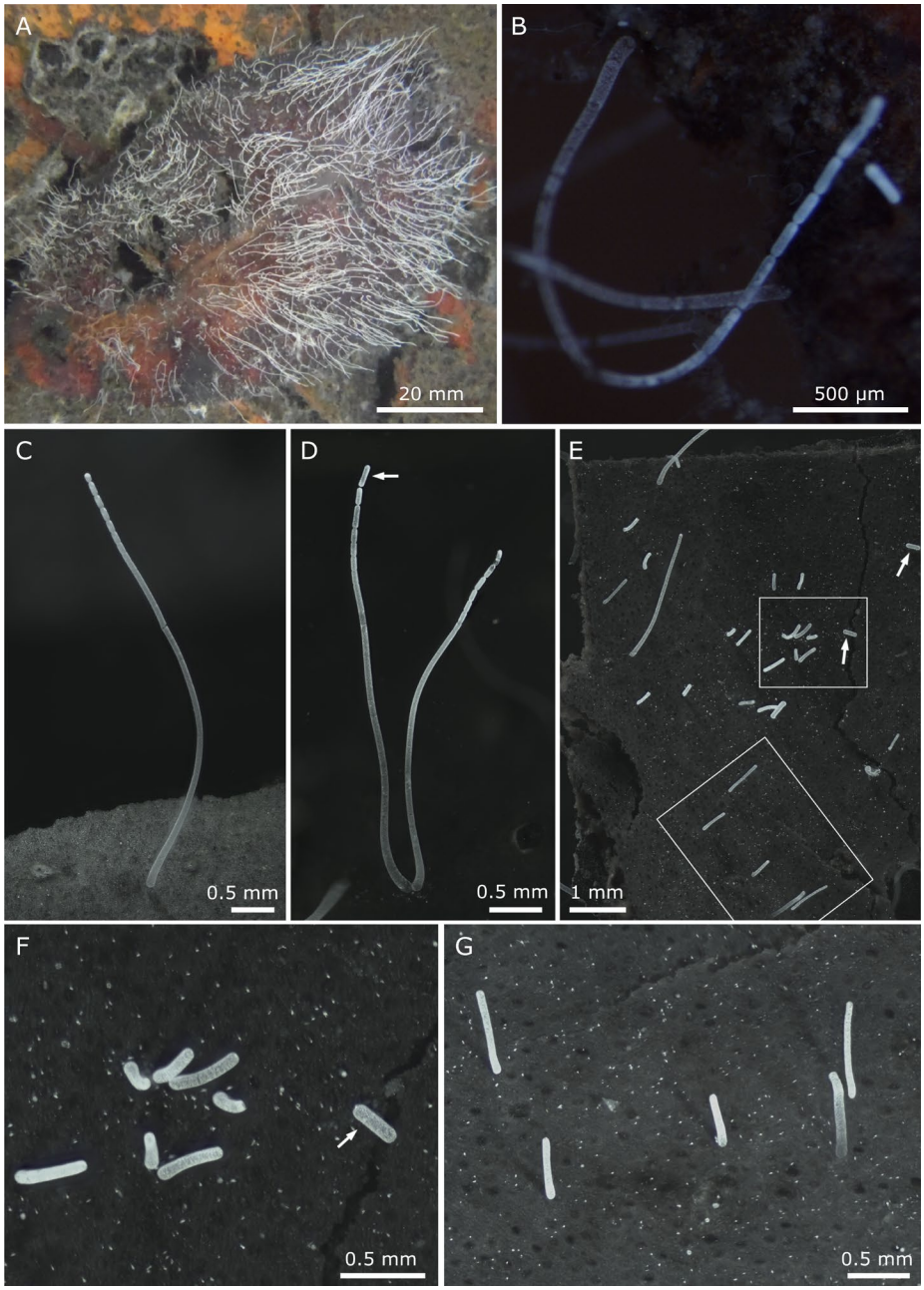

Fig. S1. Light micrographs of *Ca*. Thiomargarita magnifica attached to sunken leaves of *Rhizophora mangle*. A. Hundreds of filaments develop on sunken dead leaves partially buried in the sulfidic sediment of the mangrove. Similar patches were occasionally observed growing on partially buried plastic and oyster shell debris. The cells appeared bright white due to their intracellular elemental sulfur stores. B-D. Detail of budding cells showing a gradual constriction at the filament apex. Note that manipulation of these fragile cells sometimes folded the filaments, leaving a mark as seen at mid-cell on B. The terminal segment (arrow) is still attached to the mother cell on D and has just been released into the water column for dispersion on B. E. Area of a sunken leaf with recently released terminal segments (arrows) and newly settled filaments of various sizes. Note that these young filaments (< 3 mm) are not displaying apical constrictions yet. The two rectangles' areas are shown at higher magnification on F and G.

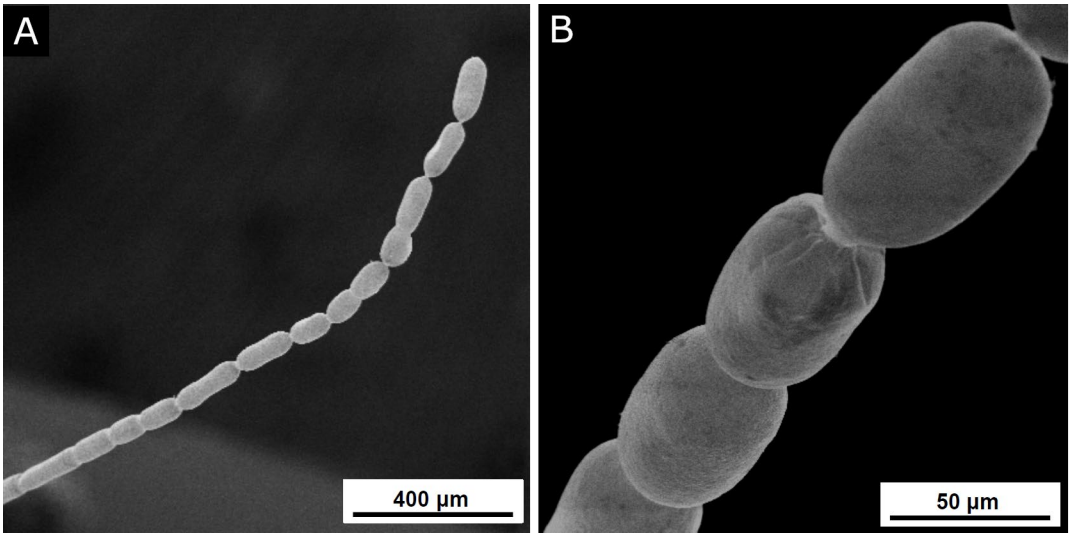

Fig. S2. Scanning electron microscopy observation of an individual *Ca*. Thiomargarita magnifica. A. Detail of the last 2 mm of the cell apex characterized by the typical multiple budding daughter cells. B. Higher magnification showing the smooth surface of the cell wall and the absence of epibiotic bacteria or biofilm.

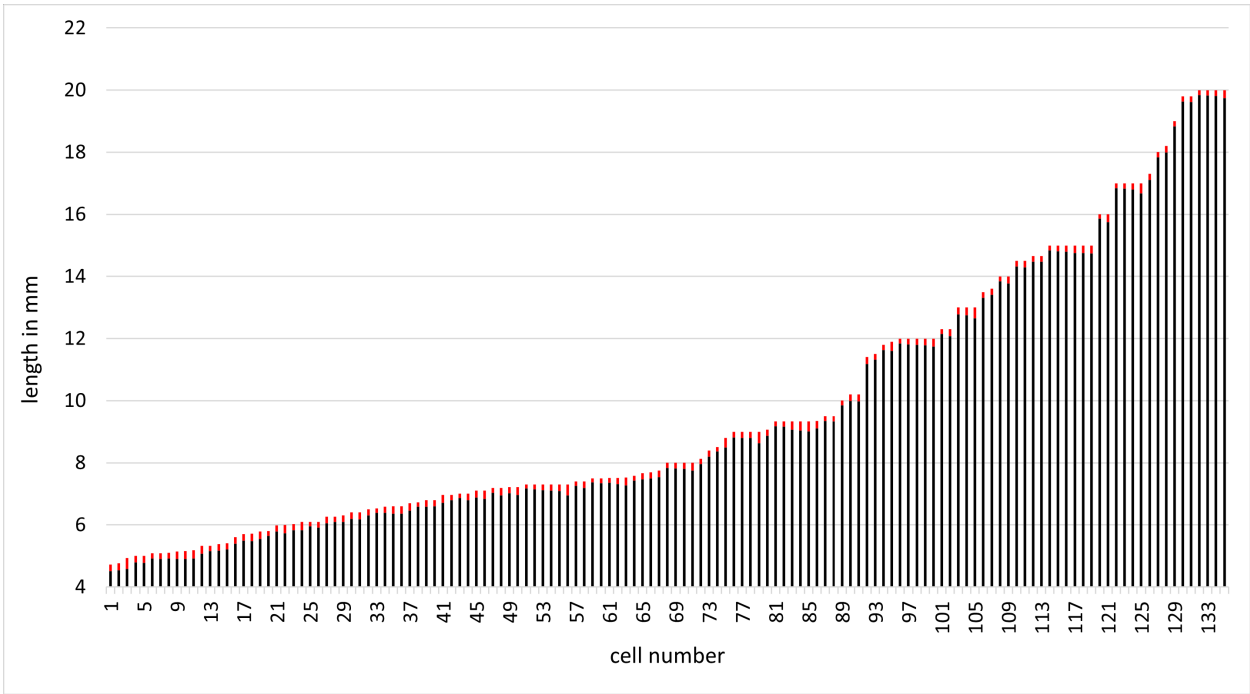

Fig. S3. Lengths of 135 *Ca.* Thiomargarita magnifica budding cells measured under a stereomicroscope. The smallest filament observed with a terminal segment here was 4.73 mm and the largest filaments were 20 mm. The length of the most apical segment is displayed in red.

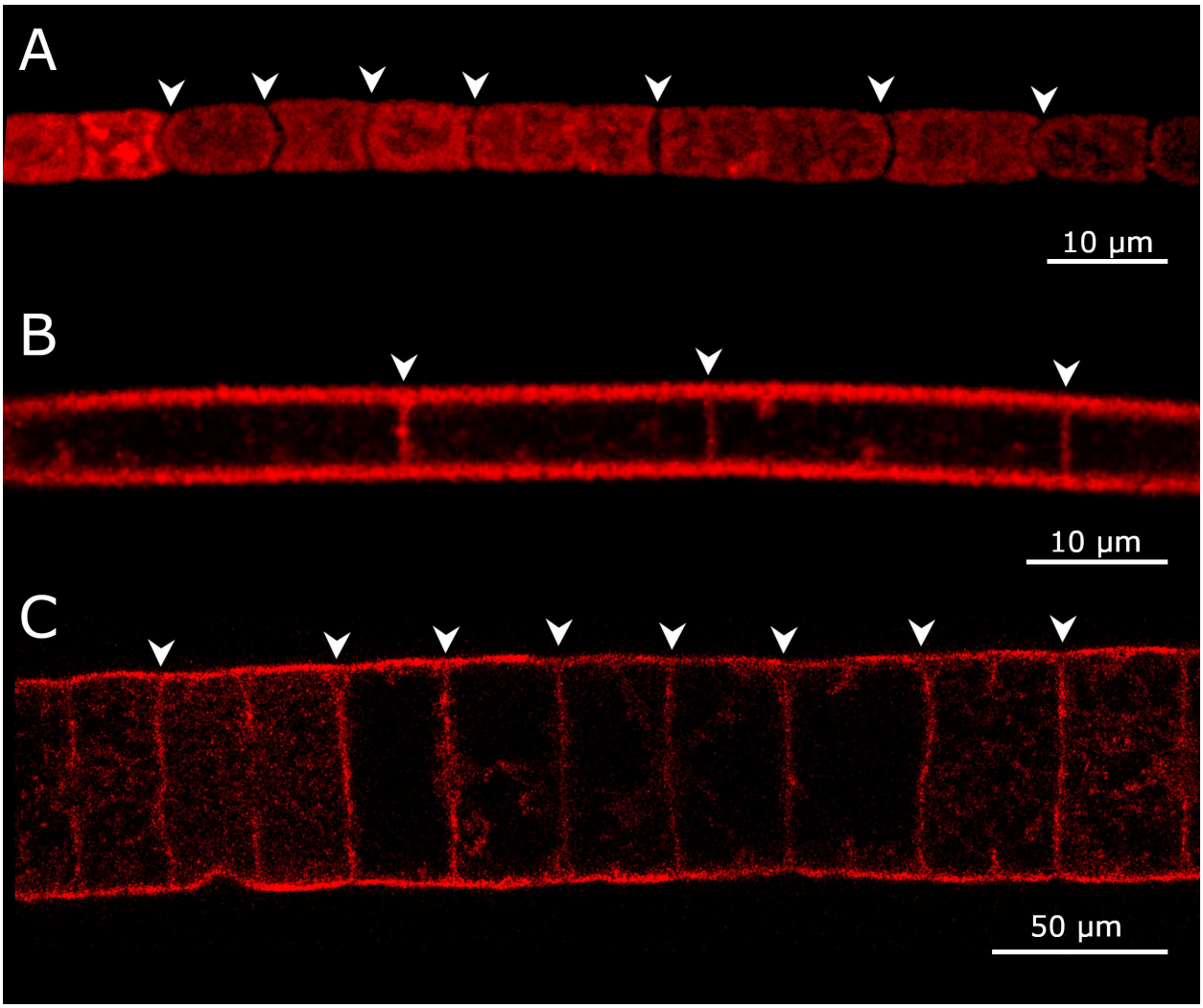

Fig. S4. Confocal laser scanning microscopy observations of three multicellular large filamentous bacteria after fluorescent labeling of membranes using FM 1-43x dye. A. The separations between the cells (white arrowheads) of the Cyanobacterium *Microcoleus vaginatus* are clearly visible after FM 1-43x staining. B. The membrane septum separating the vacuolated cells of a *Beggiatoa*-like filament are clearly stained by the membrane dye. C. The membrane septum separating the large vacuolated cells of the *Marithrix*-like bacterial filaments are clearly stained by the membrane dye.

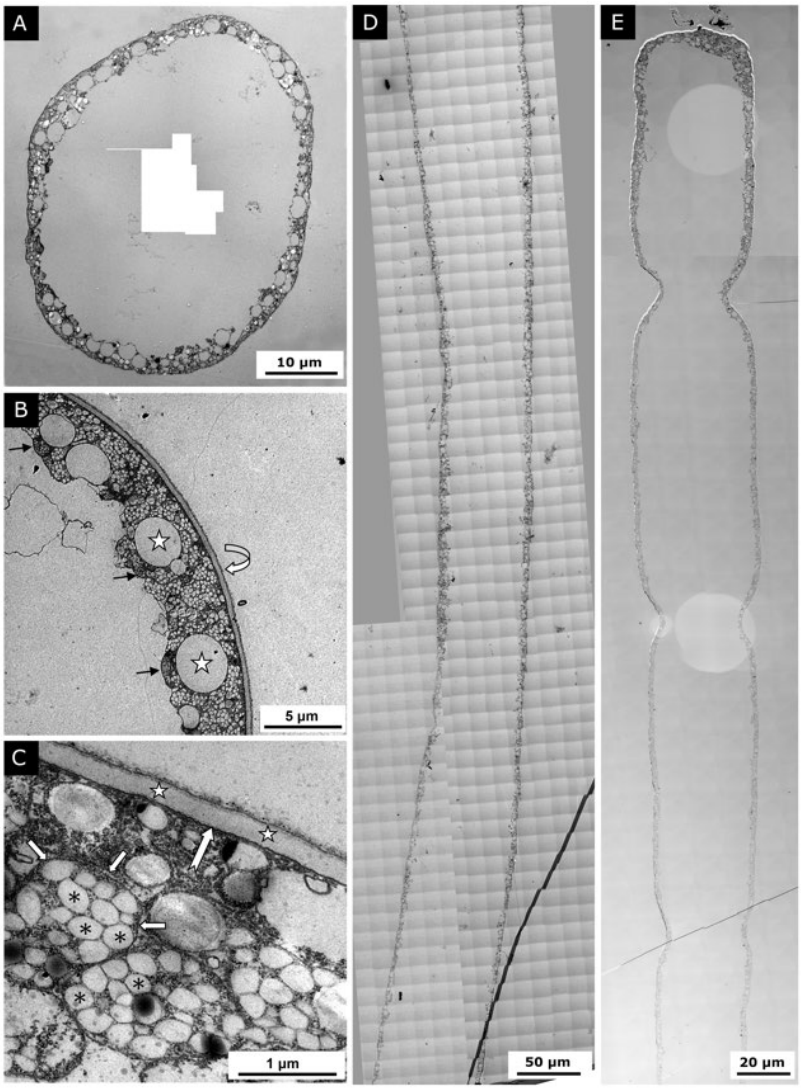

Fig. S5. Transmission electron microscopy observation of *Ca.* T. magnifica. A. Photo montage of a cell cross section. The cytoplasm is organized at the periphery of the cell around a large central vacuole. B. The cell envelope is represented by a thick outer layer covering the cytoplasmic membrane (curved white arrow). Electron lucent vesicles correspond to elemental sulfur granules (stars) dissolved by ethanol during sample dehydration. Pepins are located adjacent to the central vacuole (black arrows). C. Detailed view of the cytoplasmic membrane (big arrow) covered by a cell wall (stars). The cytoplasm is filled with organelles, each delimited by a membrane (small arrows), that sometimes contain multiple electron lucent vesicles (asterisks). D. Photomontage (764 images taken at magnification 3000x) of a cell longitudinal section 850.6 µm long. The section corresponds to the middle part of the filament. No septum was observed and the cytoplasm, cell membrane and central vacuole appear continuous throughout the whole section. E. Photomontage (349 images taken at magnification 1400x) of a cell longitudinal section 374 µm long. The montage shows the continuity of the cell and the absence of any membrane septum, even at the constriction sites.

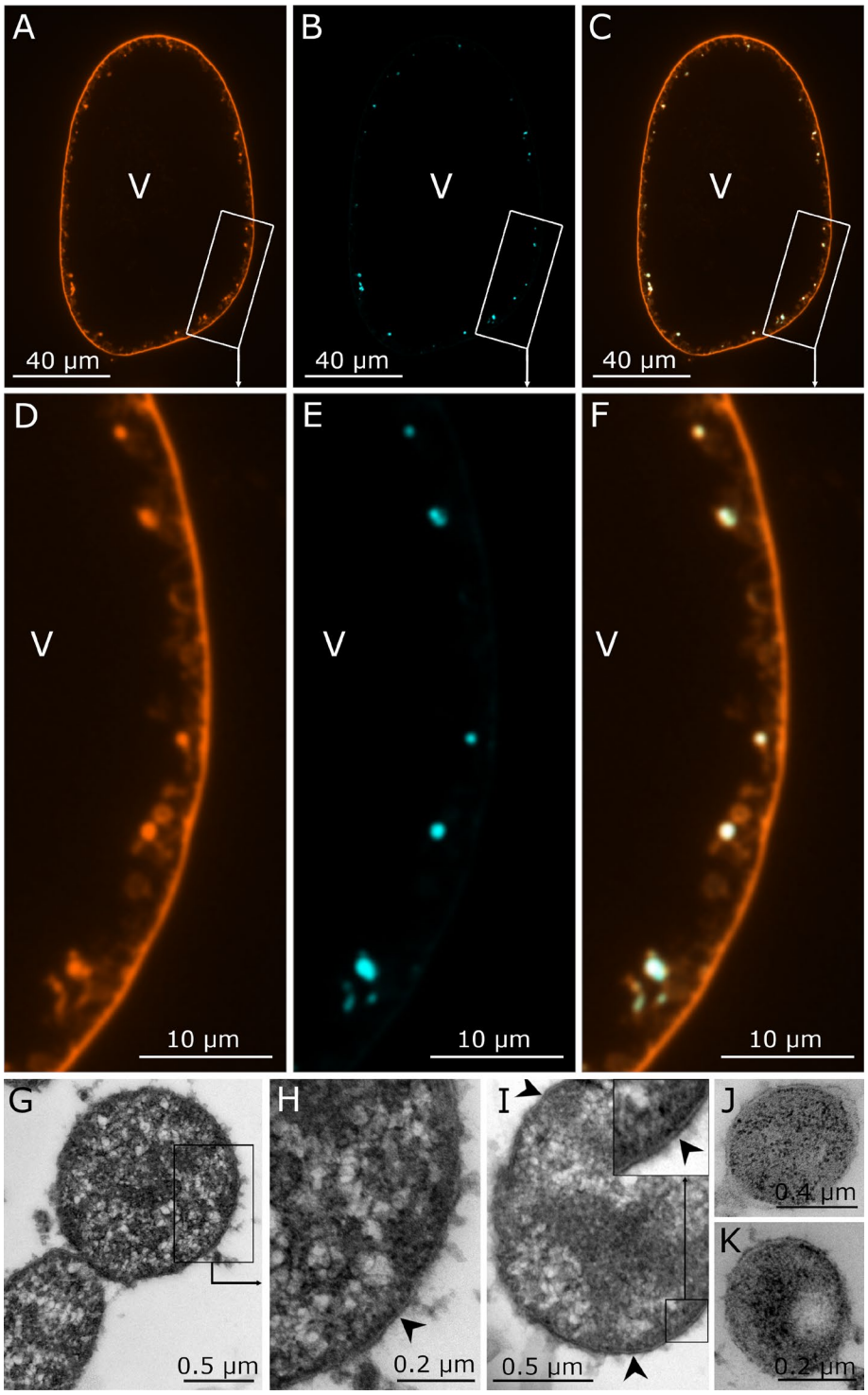

Fig. S6: Visualization of membrane and DNA on cross sections of a *Ca*. Thiomargarita magnifica cell and TEM detail of the pepins. A, overview of a cell's cross section showing the membrane in orange after staining with FM 1-43x. B, visualization of the DNA in blue on the same section after DAPI staining. C, Overlay from A and B. D-F, higher magnification of the rectangle area from A, B and C. G-K, Electron microscopy observations of the same cell showing pepins surrounded by a membrane (arrowheads). V, vacuole.

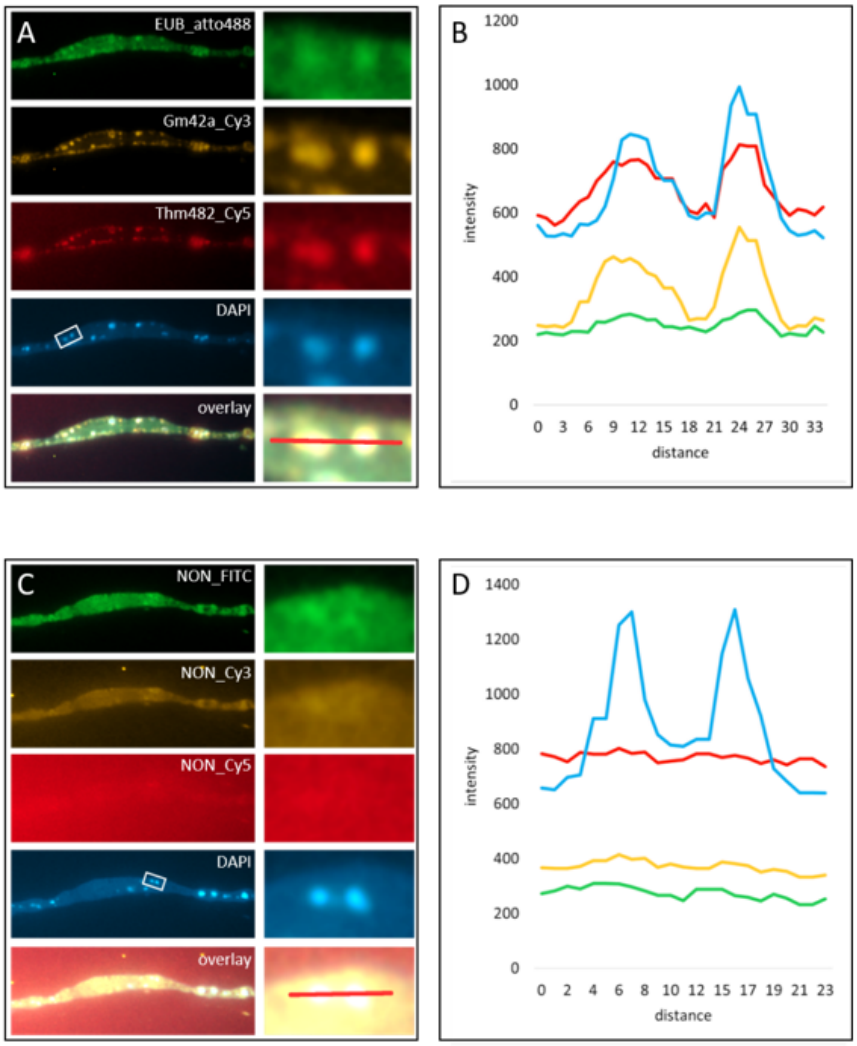

Fig. S7. Fluorescence *In Situ* Hybridization (FISH) with general, specific and nonsense probes and DNA staining with DAPI on sections of resin-embedded *Ca.* Thiomargarita magnifica cells. A. Pepins are labeled with the general bacterial probe (green), a gammaproteobacteria-specific probe (yellow), the *Thiomargarita* specific probe (red), or with DAPI (blue). The detail of the two pepins in the white rectangle on the DAPI image is shown at higher magnification on the right. B. The intensity profiles of a line crossing the two pepins from the insert is showing peaks in all four channels. C. FISH performed on sections of the same cell placed on the same slide show no hybridization of the nonsense probe. D. The intensity profiles of the line crossing the two pepins from the insert on B show a peak for the DAPI signal but no signal in any of the probe colors.

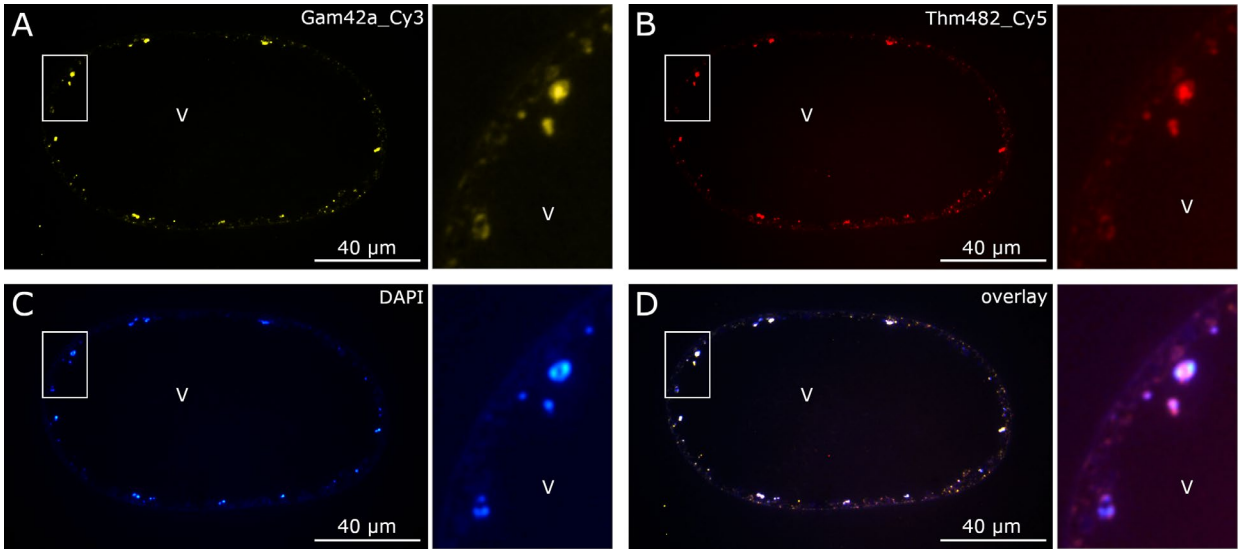

Fig. S8. Fluorescence *In Situ* Hybridization (FISH) with gammaproteobacteria (A) and *Thiomargarita* (B) probes and DNA staining with DAPI (C) on a cross section of the stalk area of a *Ca.* Thiomargarita magnifica cell. The pepins are labeled with both gammaproteobacteria and *Thiomargarita* probes - *Thiomargarita* is a genus within the gammaproteobacteria - as well as by DAPI as shown by the overlay of the three images (D). A higher magnification of the area in the white rectangle is provided on the right of each image. It shows that while the FISH and DAPI signals colocalize in the pepins, they do not display the same pattern. On the larger pepin for instance, the ribosomes' RNAs labeled in yellow and red are concentrated in the center of the organelle while the DNA labeled in blue shows higher signal on the sides. V, vacuole.

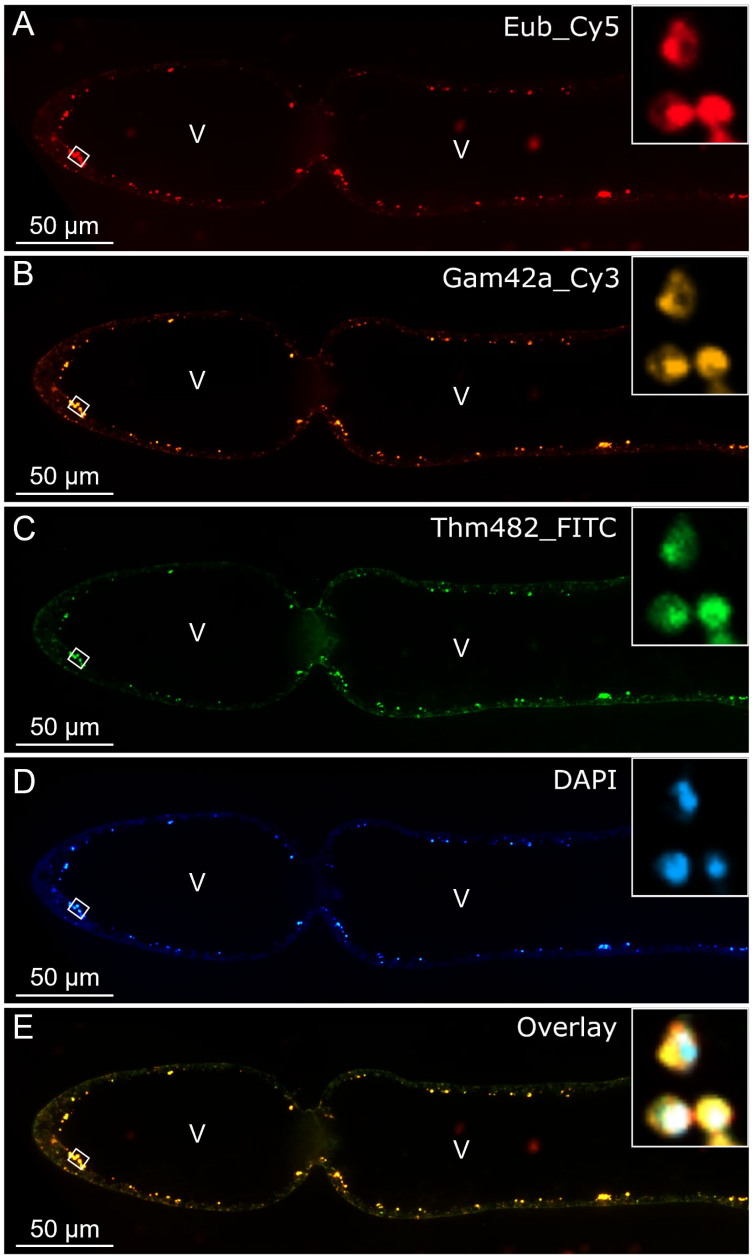

Fig. S9. Fluorescence *In Situ* Hybridization (FISH) with probes targeting general bacteria (A, probe EUB labeled with Cy5); gammaproteobacteria (B, Gam42a probe labeled with Cy3); and *Thiomargarita,* a genus within gammaproteobacteria (C, probe Thm482 labelled with FITC); and DNA staining with DAPI (D); on a longitudinal section of the apical end of a *Ca.* Thiomargarita magnifica cell. On this longitudinal section pepins are present in the cytoplasm of the terminal segment as well as in the stalk before the constriction. Note that the terminal segment, like the rest of cell, presents a large central vacuole which occupies most of the cell volume. The pepins labeled with all three probes and DAPI are shown by the overlay (E). The detail of the three pepins in the white rectangle is given in the top right insert. Note that no epibiotic bacteria are detected with FISH. V, vacuole.

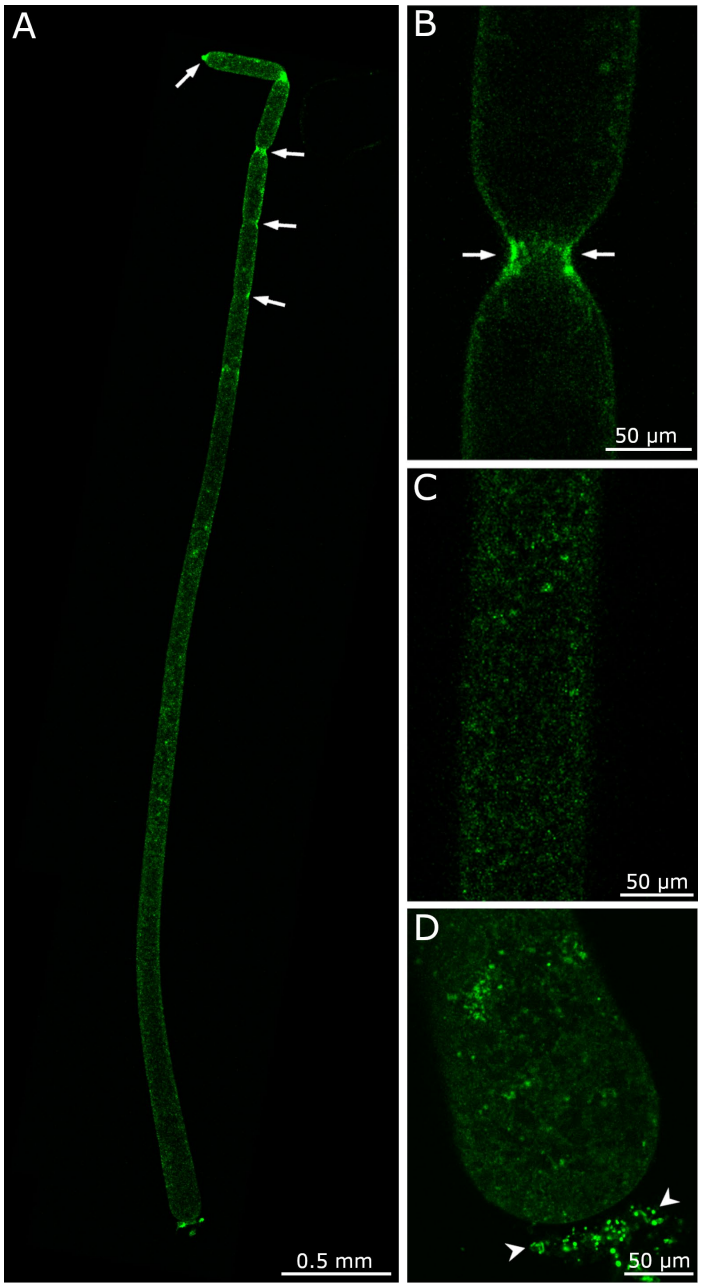

Figure S10: Visualization of translational activity in an entire *Ca*. Thiomargarita magnifica cell using bioorthogonal non-canonical amino acid tagging (BONCAT). Proteins synthesized during the 24h of incubation with a clickable methionine analogue are revealed by green fluorescence. A. Photomontage of an entire filament showing that the translational activity is detected throughout the cell. Activity is, however, not homogeneously distributed, with label being more prevalent at the constriction sites (arrows) towards the apical pole of the cell. Higher magnification of the top, middle and basal part of the cell is shown in B, C and D respectively. Newly synthesized proteins are primarily localized in small round-shaped areas akin to pepins in shape, size, and localization. Note the presence of numerous BONCAT-positive bacteria (arrowheads), outside the *Ca*. T. magnifica cell, in the biofilm covering the dead leaf substrate in D. A total of 12 cells were analyzed after three different incubation experiments following three different samplings and showed consistent labeling patterns.

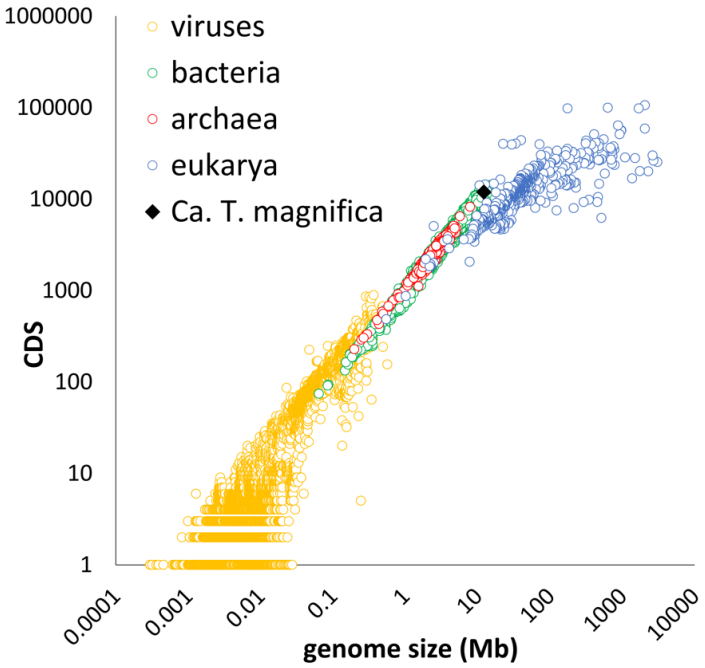

Fig. S11. Total number of coding sequences (CDS) compared to the total genome size for all genomes retrieved from IMG/M. *Ca.* Thiomargarita magnifica's genome is among the largest bacterial genomes with one of the highest numbers of coding sequences.

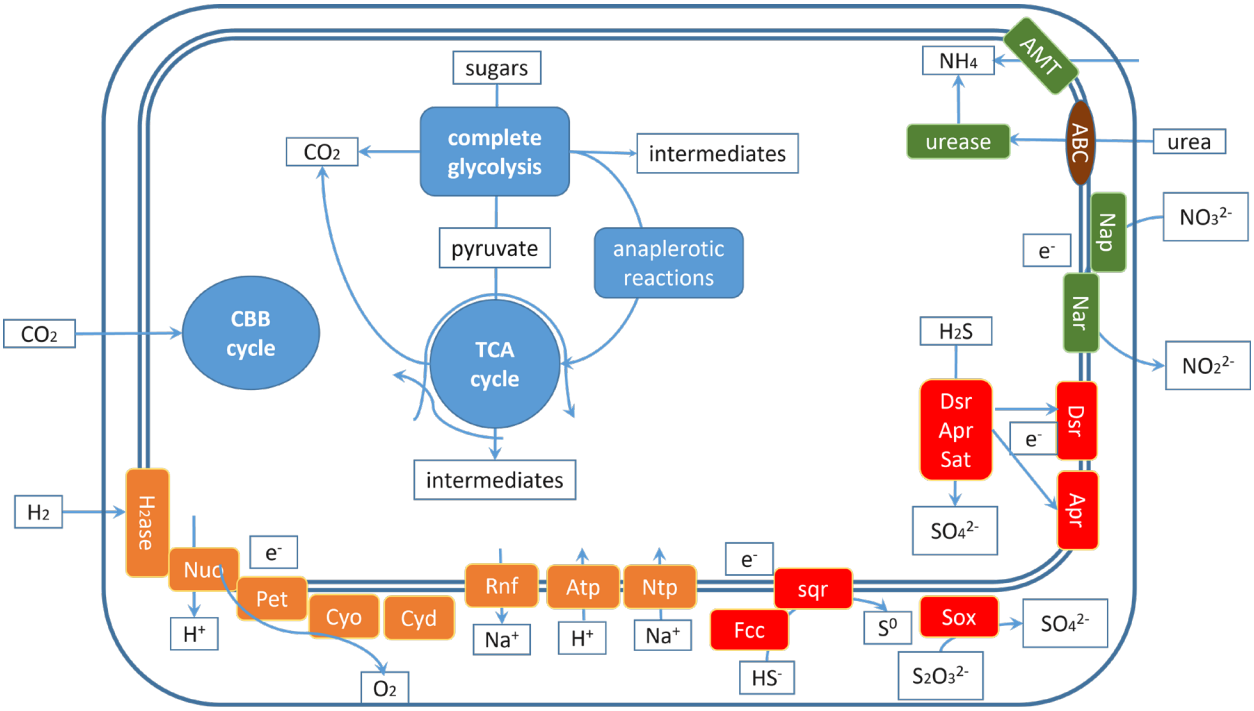

Fig. S12: Schematic representation of the carbon (blue), nitrogen (green) and sulfur (red) metabolisms as well as electron transport chain (orange) based on the *Ca*. Thiomargarita magnifica draft genomes. TCA, tricarboxylic acid; CBB, Calvin–Benson–Bassham; ABC, ABC transporter; Cyo - cbb3 type cytochrome c oxidase; H_2ase, Ni-Fe hydrogenase; Apr, adenylylsulfate reductase; Atp, F-type ATP synthase; Cyd, cytochrome d ubiquinol oxidase; Cyo, cbb3 type cytochrome c oxidase; Dsr, sulfite reductase; H_2ase, Ni-Fe hydrogenase; Fcc, sulfide dehydrogenase; Nap, periplasmic nitrate reductase; Nar, nitrate reductase; Ntp, V-type ATP synthase; Nuo, NADH-quinone oxidoreductase; Pet, ubiquinol-cytochrome c reductase; Rnf, electron transport complex protein Rnf; Sat, sulfate adenylyltransferase; Sox, sulfur-oxidizing protein Sox; Sqr, sulfide:quinone oxidoreductase.

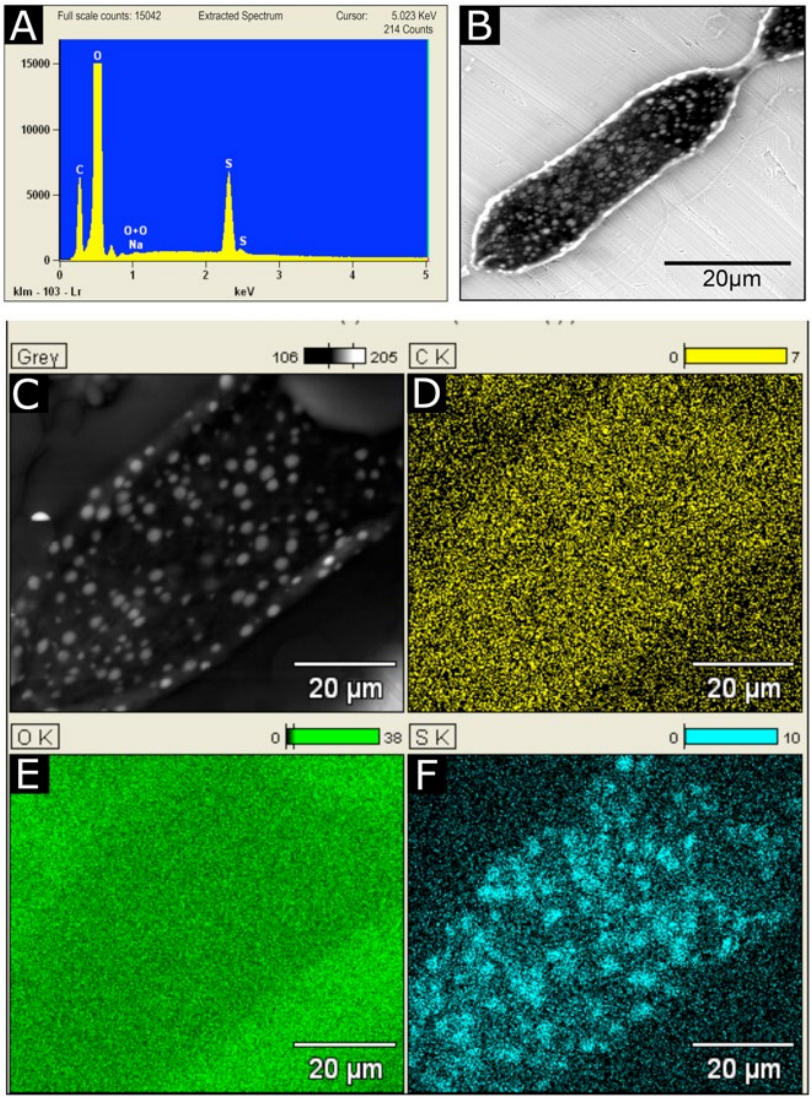

Fig. S13. Environmental Scanning Electron Microscopy images and analyses recorded on slightly fixed *Ca.* Thiomargarita magnifica under 650 mPa water vapor atmosphere. A: EDXS spectrum acquired on the area corresponding to Figure S13B. Sulfur is clearly identified. B: Image collected with secondary electron detector with 15 kV accelerating voltage. The image is dominated by backscattered electrons due to the high penetration power of the incident electrons in the specimen. The resulting atomic number (Z) contrast reveals high Z number granules (sulfur) as bright areas embedded in the light (C, N, O, H) matrix in black. Secondary contrast is present but only visible at the periphery of the cell. C: Electron back scattered images obtained at 15 kV highlighting the sulfur granules as in B. Each empty granule appears white in a back-scattered electron image due to the intensity of the BSE signal which is strongly related to the atomic number of the chemical element. D-F: X-ray maps characterizing distributions of carbon (D), oxygen (E) and sulfur (F) in the sample. The sulfur map (F) clearly confirms that the bright areas appearing in the BSE images correspond to sulfur locations.

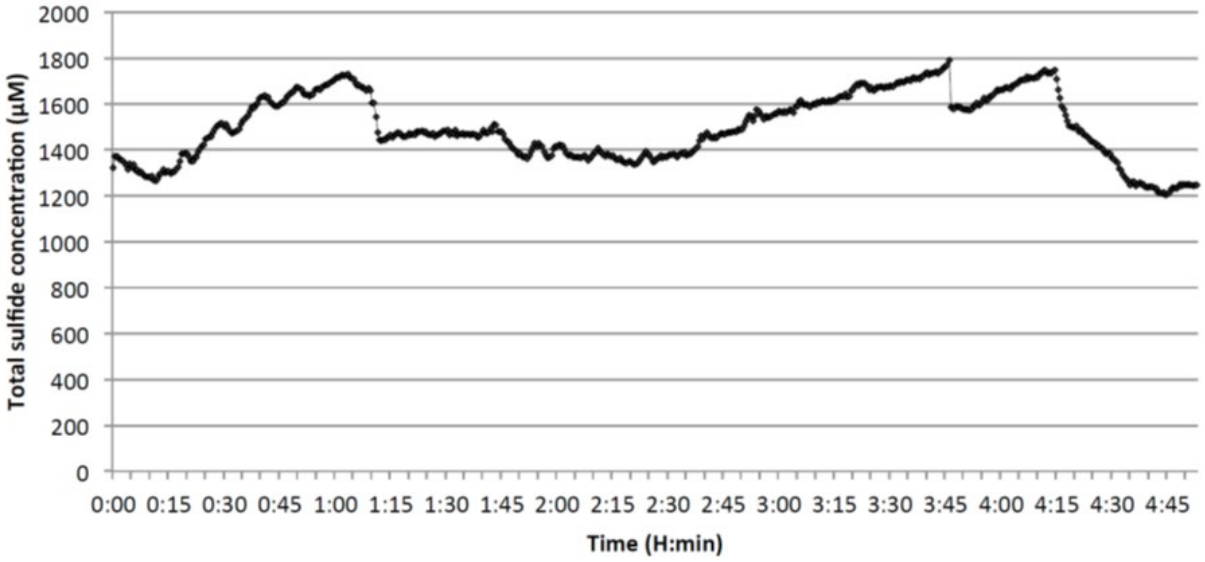

Fig. S14. Variation of total sulfide concentration measured in the middle of a "bouquet" of *Ca*. Thiomargarita magnifica attached to sunken leaves of *Rhizophora mangle* over 5 hours.

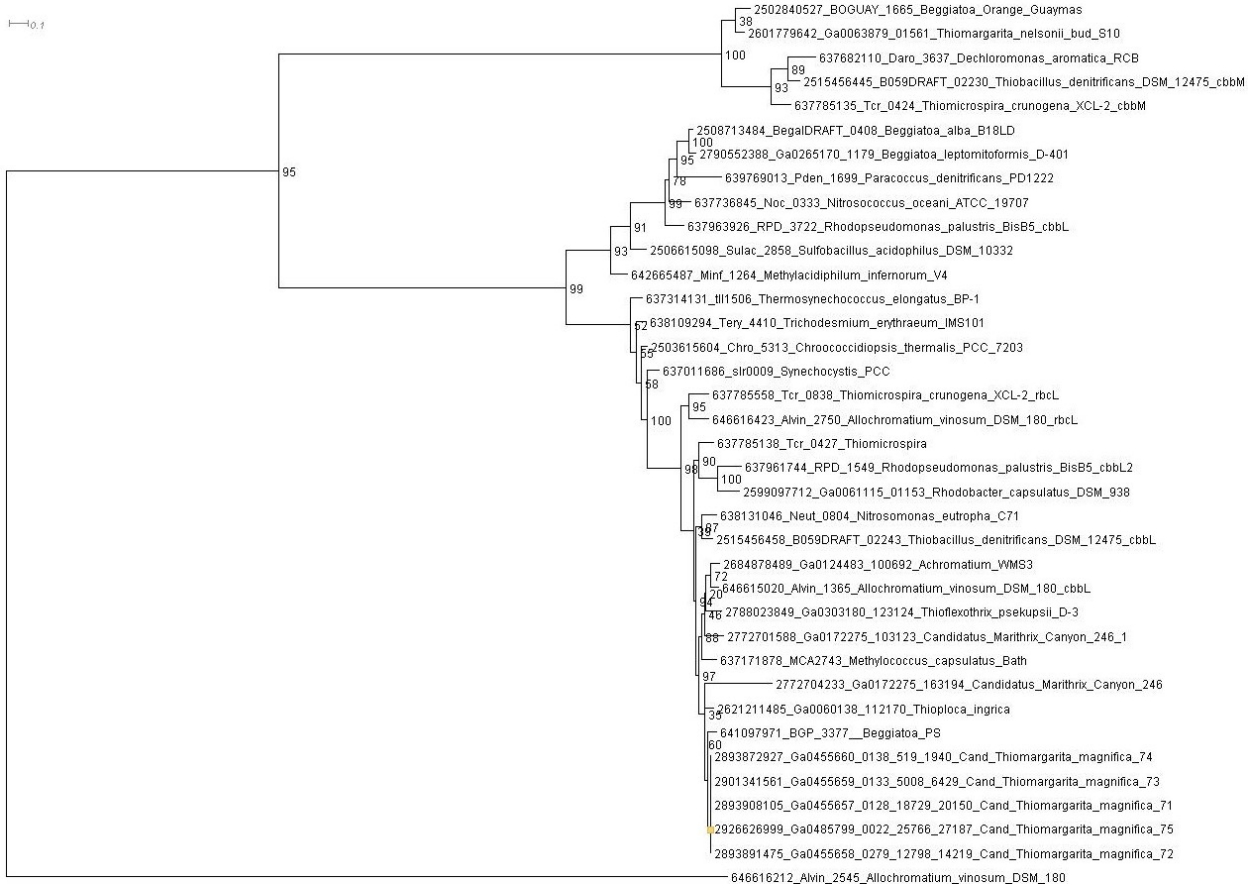

Fig. S15. Phylogenetic tree of RuBisCo large subunit proteins encoded by *Ca*. T. magnifica and representative sequences from other genomes. *Ca*. T. magnifica *rbcL* sequences cluster within the type I clade, separately from type II RuBisCO encoded by *Ca*. T. nelsonii bud S10 genome.

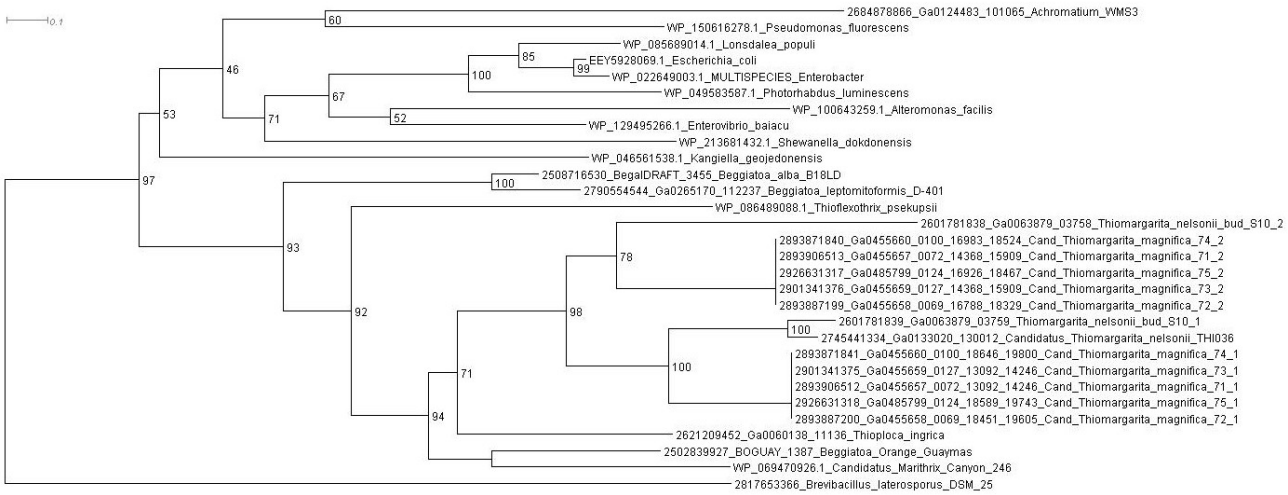

Fig. S16. Phylogenetic tree of RodZ proteins encoded by *Ca*. T. magnifica draft genomes and selected reference sequences. Two copies of *rodZ* found next to each other in *Ca*. T. magnifica and *Ca*. T. nelsonii bud S10 are more similar to each other than to any other sequence, suggesting a duplication in a common ancestor of *Ca*. Thiomargarita spp.

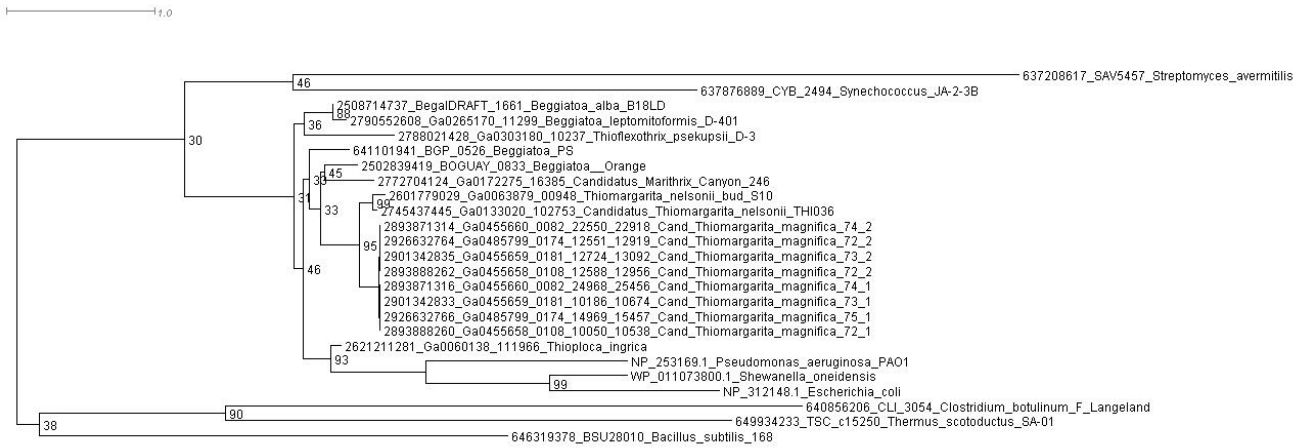

Fig. S17. Phylogenetic tree of MreD proteins encoded by *Ca*. T. magnifica draft genomes and selected reference sequences. Two copies of the *mreD* gene found close to each other in *Ca*. T. magnifica are more similar to each other than to any other sequence, suggesting a duplication in *Ca*. T. magnifica.

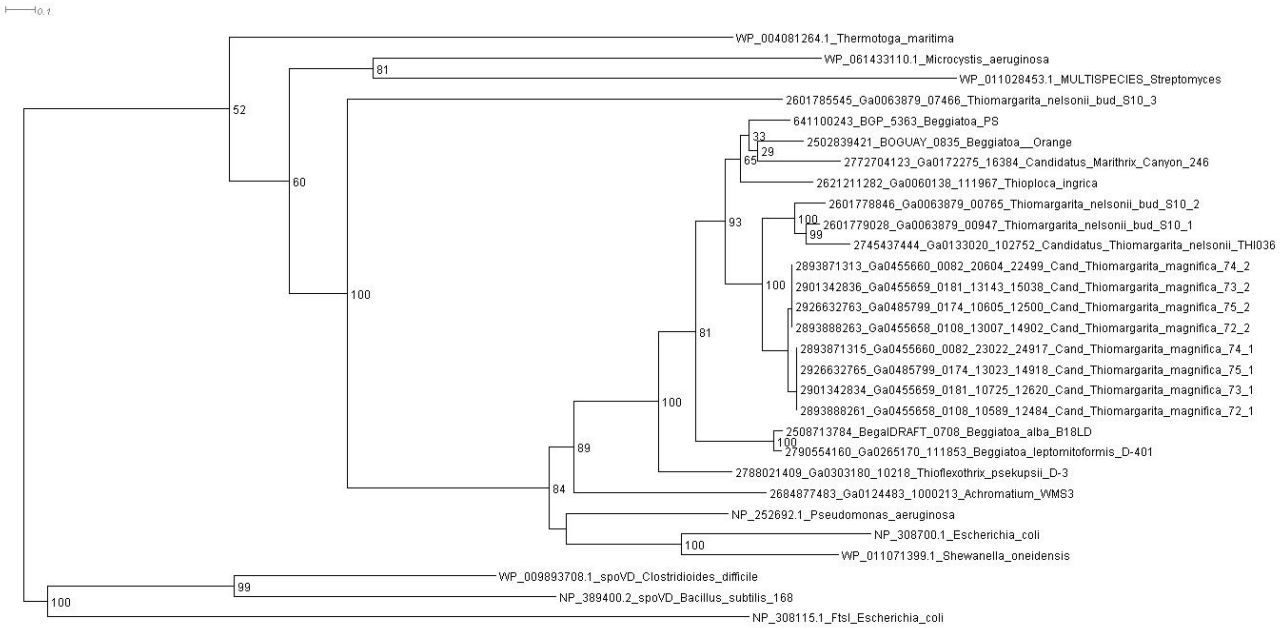

Fig. S18. Phylogenetic tree of MrdA proteins encoded by *Ca*. T. magnifica draft genomes and selected reference sequences. Two copies of the *mrdA* gene found close to each other in *Ca*. T. magnifica are more similar to each other than to any other sequence suggesting a duplication in *Ca*. T. magnifica.

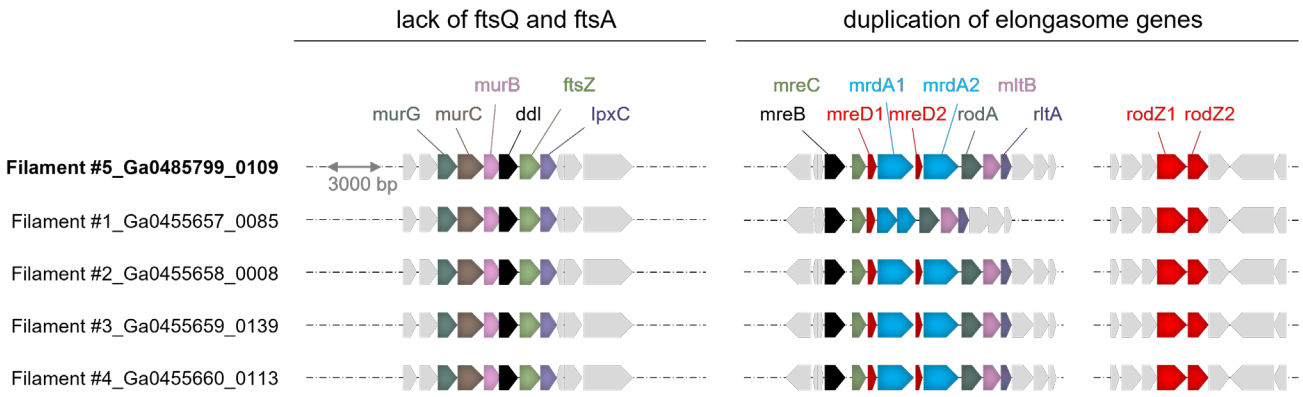

Fig. S19. Detail of the *ddl*, *mreB* and *rodZ* gene neighborhood for the five *Ca*. T. magnifica draft genomes (see also Fig. 3B).

5

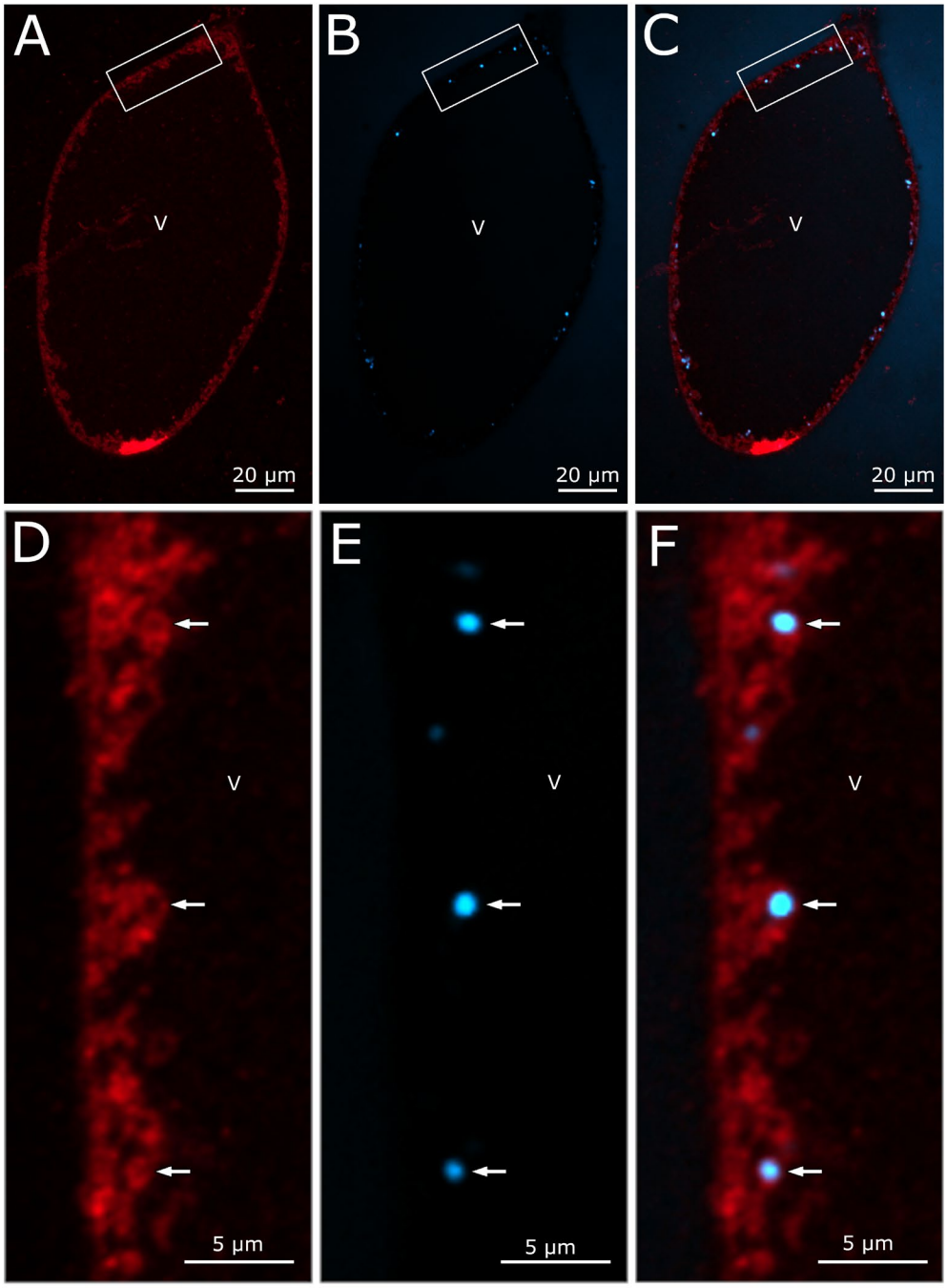

Figure S20: Localization of bioenergetic membranes in *Ca*. Thiomargarita magnifica with ATP synthase immunohistochemistry. A. Infrared fluorescence of the Cy5-labeled secondary antibody (red). ATP synthase positive signal is visible throughout the whole cytoplasm of the cell which is constrained at the periphery of the central vacuole (V). B. Fluorescent labeling of DNA containing pepins using DAPI (blue). C. Overlay of the red and blue channels. The detail of the rectangle area is provided below each image (D-F). The pepin membranes are positive for the detection of ATP synthase suggesting the organelle's role in the production of bioenergetic membrane. The complex cytoplasmic membrane network outside the pepins is also labeled but the outermost membrane, delimiting the cell, is negative for ATP synthase.

	cells A, B, C, E & N	cell D
scan date	8/30/2019	8/27/2019
total number of cells in sample	5	1
number of entire cells in the scan	3	1
dataset width (μm)	1755	984
dataset height (μm)	1843	1008
dataset depth (μm)	4668	6270
number of tiles	3	8
pixel size (μm)	1.00	1.00
number of projections per tile	1601	801
camera binning	1	2
source filter	air	air
source setting (kV)	40	40
source setting (μA)	74	74
source-RA distance (mm)	-9.2	0.0
optical magnification	4.0	4.0
exposure time (s)	15	10

Table S1. Summary of hard x-ray computed tomography scans parameters.

cell ID	analyzed with	entire cell	cell length (mm)	min diam (µm)	max diam (µm)	total cell volume (m^3)	cytoplasm volume (m^3)	vacuole volume (m^3)	% of vacuole	est. genome copies	est. genome copies/mm	est. genome copies in 2cm cell	resolutio X - Y (µm)	resolution Z (µm)	tiles no.
A	HXT	Y	1.19	17	33	2.37E-13	-	-	-	-	-	-	1	1	1
B	HXT	Y	1.54	17	29	3.50E-13	-	-	-	-	-	-	1	1	1
C	HXT	N	1.66	70	119	-	-	-	-	-	-	-	1	1	1
D	HXT	Y	6.37	43	83	1.42E-11	5.30E-12	8.89E-12	63	-	-	-	1	1	8
E	HXT	Y	9.66	47	96	-	-	-	-	-	-	-	1	1	3
F	CLSM	Y	1.92	54	96	6.59E-12	1.40E-12	5.19E-12	79	-	-	-	0.42	2	12
G	CLSM	Y	4.27	61	147	2.20E-11	5.91E-12	1.61E-11	73	-	-	-	0.63	2.73	5
H	CLSM	Y	2.39	39	62	4.29E-12	9.26E-13	3.36E-12	78	72419	30263	605256	0.23	1	14
I	CLSM	Y	3.35	66	112	-	-	-	-	-	-	-	1.27	2.73	6
J	CLSM	Y	2.32	52	87	-	-	-	-	-	-	-	1.24	2.73	3
K	CLSM	Y	2.83	67	71	-	-	-	-	-	-	-	0.92	2.95	8
L	CLSM	N	0.20	-	-	-	-	-	-	8931	45707	914140	0.1	1	1
M	CLSM	N	0.66	-	-	-	-	-	-	22882	34670	693397	0.23	1	1
N	HXT	N	1.97	58	101	-	-	-	-	-	-	-	1	1	2

Table S2. Summary of the *Ca.* T. magnifica observations with Hard X-ray Tomography (HXT) and Confocal Laser Scanning Microscopy (CLSM). A total of 14 cells were analyzed in 3D, ten of which were analyzed in their entire length. The table provides the detail of the length of the observed cell. When applicable the table also shows the cell minimum and maximum diameters, its total volume, cytoplasm volume and central vacuole volume as well as the percentage of vacuole volume relative to the whole cell. For three cells, the numbers of genome copies were estimated. This estimation is also provided as a number of genome copies per millimeter of cell and as an extrapolation for a fully grown 2 cm cell. Finally, the lateral resolution of the image as well as the z resolution and the number of tiles assembled in the dataset is provided.

	Filament 1	Filament 2	Filament 3	Filament 4	Filament 5
IMG taxon ID	2893903249	2893884007	2956128196	2893867023	2926625205
SRA run ID	SRR18723970	SRR18723883	SRR18723969	SRR18725150	SRR18724479
Library total bases (Mb)	816.5	1007.5	1659.2	1644.6	3057.3
Read counts (x10^6)	5.5	6.7	11.1	11.0	20.4
Assembly size (Mb)	18.87	16.88	22.48	16.26	16.48
Number of contigs	6613	4720	2713	2383	2196
Number of bins	3	2	6	2	2

Table S3. Accession numbers, reads, assembly and binning statistics for the five single amplified genomes of *Ca*. Thiomargarita magnifica. A single genome bin per each filament was identified as *Ca*. Thiomargarita magnifica and used for further genomic analysis.

	Ca. T. magnifica bin from:					*Ca*. T nelsonii	
	Filament 1	Filament 2	Filament 3	Filament 4	Filament 5	Thio36	Bud S10
IMG taxon ID	2955952073	2956002012	2955989171	2955976453	2955963837	2236661048	2600255314
bases (Mb)	11.5	11.6	12.1	12.2	12	5.3	6.2
contigs	787	795	709	647	534	3613	439
gene count	11196	11401	11068	11788	11742	7596	7525
coding %	78.3	78.2	77.7	77.6	77.4	72	82
max contig length	183 kb	201 kb	167 kb	150 kb	202 kb	14 kb	190 kb
GC %	42.37	42.34	42.31	42.41	42.39	42	41.3
prot. of known function	5706	5804	5824	5925	5887	3486	4310
genome completeness %	91.01	92.15	93.03	92.86	93.74	70	89.8
contamination %	7.47	8.31	9.54	9.17	8.83	-	-
strain heterogeneity %	5.56	6.67	7.26	8.77	9.09	-	-
Reference	this study	this study	this study	this study	this study	Winkel et al 2016	Flood et al 2016

Table S4. Statistics of the five single cell draft genomes of *Ca*. Thiomargarita magnifica and the two published draft genomes of *Ca*. T. nelsonii.

	Ca. T. magnifica 1	Ca. T. magnifica 2	Ca. T. magnifica 3	Ca. T. magnifica 4	Ca. T. magnifica 5	Ca. T. nelsonii Thio36	Ca. T. nelsonii Bud S10
Ca. T. magnifica 1		99.63	99.84	99.85	99.84	85.23	85.33
Ca. T. magnifica 2	99.57		99.68	99.68	99.63	85.30	85.35
Ca. T. magnifica 3	99.78	99.57		99.87	99.83	85.17	85.32
Ca. T. magnifica 4	99.81	99.64	99.86		99.86	85.22	85.33
Ca. T. magnifica 5	99.77	99.64	99.85	99.90		85.39	85.46
Ca. T. nelsonii Thio36	85.55	85.50	85.88	85.66	85.80		87.07
Ca. T. nelsonii Bud S10	85.55	85.54	85.69	85.61	85.63	86.39	

Table S5. Pairwise Average Nucleotide Identity (ANI) between all available *Thiomargarita* genomes. ANI values are colored from green (highest ANIs) to red (lowest ANIs).

	Filament 1	Filament 2	Filament 3	Filament 4	Filament 5
Total SNPs	153	154	137	116	140
Substitutions	119	116	93	79	86
Deletions	32	35	42	35	53
Insertions	2	3	2	2	1
Assembly Size	10,913,008	11,052,258	11,830,191	11,817,984	11,769,126
SNPs/100Kb	1.40	1.39	1.16	0.98	1.19

Table S6. Summary of the variant calling analysis. Detailed information on SNPs in single amplified genomes of five sorted filaments (total number of SNPs and SNPs/100kb) and the breakdown of SNPs by type (substitutions, deletions, insertions).

genome id	species	large sulfur bacteria	genome size (Mbp)	BGC count	total length of BGCs (Mbp)	percentage of BGCs (%)
IMG2955963837	*Ca*. Thiomargarita magnifica fil. 5	+	12.01	143	3.11	25.86
IMG2619619276	*Thioploca ingrica* Lake Okotanpe	+	4.81	50	1.01	21.04
IMG2600255314	*Ca.* Thiomargarita nelsonii bud S10	+	7.71	85	1.41	18.33
NC_009380.1	*Salinispora tropica* CNB-440	-	5.18	16	0.90	17.30
IMG2772190729	*Ca.* Marithrix sp. Canyon 246	+	3.22	30	0.54	16.70
IMG2788500497	*Cycloclasticus* sp. SP43	-	2.35	28	0.35	15.04
NC_017765.1	*Streptomyces hygroscopicus*	-	10.15	38	1.53	15.04
IMG2786546773	*Thioflexothrix psekupsii* D-3	+	3.97	22	0.59	14.87
IMG2585428147	*Thiothrix eikelboomii* ATCC 49788	+	4.16	40	0.51	12.23
IMG2506520049	*Thiothrix nivea* JP2, DSM 5205	+	4.69	45	0.57	12.13
NZ_CP042324.1	*Streptomyces coelicolor* A3(2)	-	8.67	27	0.98	11.35
IMG2515154021	*Thiolinea disciformis* DSM 14473	+	3.92	24	0.35	9.05
GCF_000149205.2	*Aspergillus nidulans* FGSC A4	-	30.28	53	2.43	8.03
IMG2599185268	*Thiothrix caldifontis* DSM 21228	+	3.94	24	0.30	7.71
IMG2619618925	*Thiothrix* sp. EBPR_Bin_364	-	3.75	27	0.28	7.45
IMG2802428810	*Cocleimonas flava* DSM 24830	-	4.49	18	0.28	6.34
IMG2740892603	*Methylophaga* sp. SM14	-	2.89	19	0.16	5.69
IMG2517093012	*Leucothrix mucor* DSM 2157	+	5.19	30	0.29	5.53
IMG2502790011	*Beggiatoa* sp. Orange Guaymas	+	4.77	26	0.26	5.38
IMG2556921650	*Thiothrix lacustris* DSM 21227	+	3.72	29	0.20	5.28
IMG2788500402	*Beggiatoa leptomitoformis* D-401	+	4.27	18	0.18	4.17
IMG2515154111	*Thiofilum flexile* DSM 14609	+	3.79	23	0.16	4.15
IMG2508501047	*Beggiatoa alba* B18LD	+	4.27	16	0.15	3.60
IMG2684622566	*Achromatium* sp. WMS3	+	3.61	10	0.11	3.17
IMG2236661048	*Ca.* Thiomargarita nelsonii Thio36	+	5.26	31	0.11	2.10
IMG640963011	*Beggiatoa* sp. PS	+	6.06	13	0.07	1.12

Table S7. Biosynthetic Gene Cluster (BGC) analysis in *Thiomargarita* species, other bacteria (including large sulfur bacteria) and one fungus model system (*Aspergillus nidulans*) famously rich in secondary metabolism. Only bacterial MAGs with CheckM-based contamination estimates not exceeding 10% were included.

See attached excel sheets for the tables S8 to S11:

Table S8. Summary table of the divisome and elongasome genes detected in the genomes of *Ca.* T. magnifica, *Ca.* T. nelsonii and eight other LSBs.

Table S9. Summary table of the metabolic capabilities of *Ca.* T. magnifica, *Ca.* T. nelsonii and height other LSBs based on their genome analysis.

Table S10. Summary table of type VI secretion genes found in the genomes of *Ca*. T. magnifica, *Ca*. T. nelsonii and eight other LSBs.

Table S11. Summary table of Pfam families overrepresented in *Ca*. T. magnifica genomes in comparison to other LSBs. See the excel file for better quality.

Movie S1.
3D rendering of 6.37 mm long *Ca.* T. magnifica cell observed with hard x-ray tomography (cell *D*, see raw data in Movie S2). The cell cytoplasm has been segmented out and appears continuous along most of the cell length until the most apical constrictions close off completely. See 2D virtual slices in Movie S2.

Movie S2.
Fly-through animation of the 6271 virtual slices from cell *D* dataset acquired with hard x-ray tomography with an isotropic resolution of 1 µm. The cell cytoplasm appears as a white ring. Note that samples for HXT were stained with osmium tetroxide to increase contrast on cell membranes (*41, 42*). The cell cytoplasm was segmented on each slice to produce the 3D rendering of the entire cell presented in Movie S1.

Movie S3.
Fly-through animation of the virtual slices from the 4.27 mm long cell *G* observed with CLSM after fluorescent labeling of membranes (Table S2). The 3D mesh rendering shows the cell wall in green and the continuous central vacuole in red.

Movie S4.
The video shows portions of a *Ca.* T. magnifica and a *Marithrix*-like cell observed in 3D at the confocal laser scanning microscope after staining with the membrane dye FM 1-43x. The membrane septa delimiting the cells of the multicellular *Marithrix*-like filament are clearly visible while no membrane septum is visible inside the *Ca.* T. magnifica cell.

Movie S5.
The video shows a portion of the *Ca.* T. magnifica cell used for the estimation of the polyploidy level (cell *M*, Table S2). After DAPI staining, the cell was observed at the confocal laser scanning microscope in 3D by acquiring a z-stack of images. The 3D reconstruction clearly shows the numerous DNA clusters of various sizes spread throughout the cell cytoplasm at the periphery of the central vacuole.

Movie S6.
3D rendering of the two smallest cells observed with hard x-ray tomography (cells *A* in yellow and *B* in blue, Table S2). The basal parts of three larger cells (*C*, *E*, and *N*) are also visible. All *Ca.* T. magnifica cells are attached to a sunken leaf and are growing out of the biofilm covering the leaf.