

Single-cell genomics-based analysis reveals a vital ecological role of *Thiocapsa* sp. LSW in the meromictic Lake Shunet, Siberia

Yu-Ting Wu¹, Pei-Wen Chiang², Kshitij Tandon², Denis Yu Rogozin^{3,4}, Andrey G. Degermendzhy³ and Sen-Lin Tang^{2,*}

Abstract

Meromictic lakes usually harbour certain prevailing anoxygenic phototrophic bacteria in their anoxic zone, such as the purple sulfur bacterium (PSB) *Thiocapsa* sp. LSW (hereafter LSW) in Lake Shunet, Siberia. PSBs have been suggested to play a vital role in carbon, nitrogen and sulfur cycling at the oxic–anoxic interface of stratified lakes; however, the ecological significance of PSBs in the lake remains poorly understood. In this study, we explored the potential ecological role of LSW using a deep-sequencing analysis of single-cell genomics associated with flow cytometry. An approximately 2.7 Mb draft genome was obtained based on the co-assembly of five single-cell genomes. LSW might grow photolithoautotrophically and could play putative roles not only as a carbon fixer and diazotroph, but also as a sulfate reducer/oxidizer in the lake. This study provides insights into the potential ecological role of *Thiocapsa* sp. in meromictic lakes.

DATA SUMMARY

FASTQ sequences have been deposited in the National Center for Biotechnology Information Sequence Read Archive under the BioProject accession number PRJNA252899.

INTRODUCTION

Researchers have recently shown an increased interest in meromictic lakes, which have layers of water that do not intermix and comprise specific zones with stratified and unique biogeochemical characteristics [1–5]. As a consequence of the layers, such lakes usually harbour uncommon microbial communities [2–4, 6, 7], specifically prevailing anoxygenic phototrophic bacteria (i.e. purple sulfur bacteria; PSBs), such as *Lamprocystis purpurea* in Lake Mahoney [8], *Chromatium okenii* in Lake Cadagno [9], *Rheinheimera* sp. in Lake Shira [6], *Halochromatium* sp. in Lake Oigon [6], *Thiocapsa* sp. in Lake Shunet [6, 7] and *Thiocapsa marina* in Lake Trekhtsvetnoe [10]. Many studies have suggested that PSBs play an important role in the cycling of various nutrients, including carbon, nitrogen and sulfur, specifically at the

oxic–anoxic interface of stratified lakes [9, 11–19]; however, the ecological role of PSBs in Lake Shunet remains unclear.

Lake Shunet (54° 25' N, 90° 13' E), located in the North-Minusinsk Depression of the central Altai-Sayan Mountain region in southern Siberia, is one of the only four meromictic lakes in the entire Asian region of Russia [20]. The lake is shaped by the local climate (evaporation) and water resource (mainly snow), which give it sharp gradients in salt content; parts of Lake Shunet have high salinity (~71 g l⁻¹) and hydrogen sulfide levels (~400 mg H₂S l⁻¹) in the monimolimnion, and a dense layer of PSB species known as *Thiocapsa* sp. in the interface zone [6, 7, 21], which is different from other meromictic lakes, such as Lakes Shira, Cadagno and Ace [22–24]. *Thiocapsa* has previously been described from brackish coastal, marine environments and freshwater [25–27]; the *Thiocapsa* sp. LSW (hereafter LSW, for Lake Shunet Water) from Lake Shunet was the first *Thiocapsa* described from an inland saline meromictic lake. *Thiocapsa* spp. are anoxygenic phototrophic PSBs that use hydrogen sulfide, thiosulfate, elemental sulfur and molecular hydrogen as electron donors during photolithotrophic growth [25]. *Thiocapsa* cells have been

Received 17 April 2021; Accepted 11 October 2021; Published 03 December 2021

Author affiliations: ¹Department of Forestry, National Pingtung University of Science and Technology, Pingtung 91201, Taiwan, ROC; ²Biodiversity Research Center, Academia Sinica, Taipei 115, Taiwan, ROC; ³Institute of Biophysics, Siberian Division of the Russian Academy of Sciences, Krasnoyarsk, Russia; ⁴Siberian Federal University, Krasnoyarsk, Russia.

*Correspondence: Sen-Lin Tang, sltang@gate.sinica.edu.tw

Keywords: flow cytometry; Lake Shunet; purple sulfur bacteria; single-cell genomics.

Abbreviations: ANI, average nucleotide identity; BChl, bacteriochlorophyll; NCBI, National Center for Biotechnology Information; PSB, purple sulfur bacterium.

Data statement: All supporting data, code and protocols have been provided within the article or through supplementary data files. Two supplementary tables and three supplementary figures are available with the online version of this article.

000712 © 2021 The Authors



This is an open-access article distributed under the terms of the Creative Commons Attribution License.

described as non-motile, cocci-shaped bacteria [25, 27, 28]. However, the ecological role of *Thiocapsa* in this specific lake ecosystem is not well described. As most microorganisms in the field are difficult to cultivate [29], the single-cell genomics approach based on whole-genome amplification has become widely used in microbial ecology. A single cell represents a distinct informational entity for an organism [30], making the approach particularly useful for investigating the ecology of microbial species or individual isolates that are difficult to cultivate [31–34].

The only genome-centric study of *Thiocapsa* was carried out by Hemp and colleagues [31], who isolated and cultured the genus from a sewage-treatment facility. Cultivation-based methods might obscure the microorganisms that are difficult to cultivate in a complex microbial community, so we aimed to isolate single cells directly from lake water using flow cytometry [35], then use the single-cell genomics technique to clarify the ecological importance of the prevalent *Thiocapsa* sp. of Lake Shunet in Siberia. Therefore, we provide comprehensive insights into the potential ecological role of *Thiocapsa* sp. in this lake based on a genome-centric analysis. Furthermore, based on the percentage of read counts in bins that were binned from metagenomes of Lake Shunet, *Enterobacter* sp. was identified as another prevailing microbial group in the chemocline, along with *Thiocapsa* sp. [7]; thus, we explored the potential interactions between *Enterobacter* sp. and *Thiocapsa* sp. in this specific niche using genome annotation.

METHODS

Sample collection

Water samples for single-cell sorting were collected at 5.0, 5.1 and 5.2 m deep with a multi-syringe stratification sampler [36] on July 21 2010 from Lake Shunet (54° 25' N, 90° 13' E) in the Republic of Khakasia, Siberia. The entire sampling time was ~2–2.5 h, including the time it took to transport the samples back to the field station at Shira. Water samples were immediately filtered through a 10 µm plankton net and then divided into Falcon 50 ml conical centrifuge tubes. The samples were then transported in ice bags back to Academia Sinica (Taipei, Taiwan). There, they were divided into aliquots and stored at –80 °C until further processing.

The sampling process for determining the bacterial community composition and metagenomes was different from single-cell sorting. Nearly 20 l of water samples was subsequently filtered through a 10 µm plankton net and placed in a Millipore-Pellicon TFF system (0.22 µm filter membrane) to collect microbes on the membranes in the local field station at Lake Shira [7].

Single-cell sorting, whole-genome amplification and PCR screening

Water samples were prepared for single-cell sorting as follows. Each sample (approximately 300 µl) was thawed on ice and centrifuged at 12000 g for 10 min, the supernatant was removed, and cells were washed three times with 300 µl

Impact Statement

Lake Shunet is one of the four meromictic lakes located in Siberia. Meromictic lakes usually harbour specific prevailing anoxygenic phototrophic bacteria in their anoxic zone, such as the purple sulfur bacterium (PSB) *Thiocapsa* sp. LSW (hereafter LSW) in Lake Shunet. PSBs have been indicated to play a key role in element cycling at the chemocline of stratified lakes; however, the ecological significance of PSBs in the lake remains poorly understood. A deep-sequencing analysis of single-cell genomics associated with flow cytometry was used to explore the potential ecological role of LSW. Based on the approximately 2.7 Mb draft genome, LSW might grow photolithoautotrophically and could play putative roles not only as a carbon fixer and diazotroph, but also as a sulfate reducer/oxidizer in the lake. This study provides insights into the potential ecological role of *Thiocapsa* sp. in meromictic lakes.

1× PBS. Then, cells were resuspended in 1 ml 1× PBS and immediately subjected to a fluorescence-activated MoFlo XDP cell sorter (Beckman Coulter) with a PE-Cy 7 detector as the main trigger requirement. Specifically, we first compared the scatter plot SSC (side scatter) to the FSC (forward scatter) based on the particle size and internal granularity, and then the emission in the infrared portion of the spectrum. The region that included the LSW population (R1) was selected and one cell was deposited into one PCR tube containing 5 µl TE buffer (Fig. S1, available with the online version of this article). Single cells were handled and downstream PCR was processed under the most stringent conditions to avoid contamination.

We used a procedure for lysing fluorescence-activated sorted single cells modified from that of Siegl *et al.* [37], including two cycles of thawing and freezing (70 and –80 °C, respectively; 10 min each), followed by the addition of 5 µl denaturation buffer (0.4 M KOH, 10 mM EDTA) and incubation at 65 °C for 3 min. Then, 0.5 µl neutralization buffer (1×0.8 M Tris-HCl, pH4) was added. Whole genome amplification was performed in a total volume of 50 µl based on phi29 polymerase-mediated multiple displacement amplification using the REPLI-g mini kit (Qiagen). Whole genome amplification was initially incubated at 30 °C for 10–16 h, then the phi 29 DNA polymerase was inactivated by heating the sample for 3 min at 65 °C. The amplified genomic DNA was stored at –20 °C until further processing.

The amplified single-cell genomic DNAs were screened using Sanger sequencing on the V1–V2 hypervariable regions of the 16S rRNA gene. PCR was carried out in a total volume of 50 µl containing 2.5 U *Superrun EX Taq* HS, 5 µl 10× *EX Taq* buffer, 200 µM dNTPs, 0.2 µM each primer (27F 5'-AGAGTTTGATCMTGGCTCAG-3' and 341R 5'-CTGCTGCCTCCCGTAGG-3') [38] and 2–5 µg single-cell multiple

Table 1. Properties of the single-cell genome of *Thiocapsa* sp. LSW

Assembly	<i>Thiocapsa</i> sp.
Raw reads (paired end)	
WGA1-9	1906273×2
WGA11-10	2024697×2
WGA12-4	2850582×2
WGA12-6	2121645×2
WGA12-16	1950558×2
QC filtered and trimmed reads (paired end)	
WGA1-9	1879220×2
WGA11-10	1969896×2
WGA12-4	2764148×2
WGA12-6	2058986×2
WGA12-16	1884446×2
Assembly statistics	
No. of contigs	563
Genome size (bases)	2707436
No. of genes	2656
No. of coding sequence	2626
No. of rRNA	4 (5S, 16S, 23S)
No. of tRNA	25

displacement amplification products. The PCR programme was as follows: an initial denaturation at 95 °C for 5 min; followed by 30 cycles at 95 °C for 20 s, 52 °C for 20 s and 72 °C for 20 s; and a final extension step at 72 °C for 5 min; then cooling at 4 °C. Six single-cell multiple displacement amplification products were identified as *Thiocapsa* sp. and sent for shotgun sequencing and analysis based on the clean Sanger sequencing electropherogram.

Genome sequencing, assembly and bioinformatics

We sorted and analysed more than 200 single cells, but only 6 single cells were classified as the same *Thiocapsa* sp. strain based on the 16S rRNA with 99% identity against the National Center for Biotechnology Information (NCBI) database and with 100% identity among the six 16S rRNA sequences. The genomes of these six single cells were sent for sequencing and individual single-cell assemblies were initially performed with SPAdes [39] using *k*-mer sizes of 21, 33 and 55. One sample was removed from the genome assembly due to contamination of ambiguous sequences. Furthermore, average nucleotide identity (ANI) was calculated one versus the other among the five single cells. Contigs ≥ 1000 bp were used for ANI calculations using <http://enve-omics.ce.gatech.edu/ani/index> [40]. All ANI values ranged from 99 to 100% (Table S1). Therefore, the current draft genome was co-assembled using reads from five single cells.

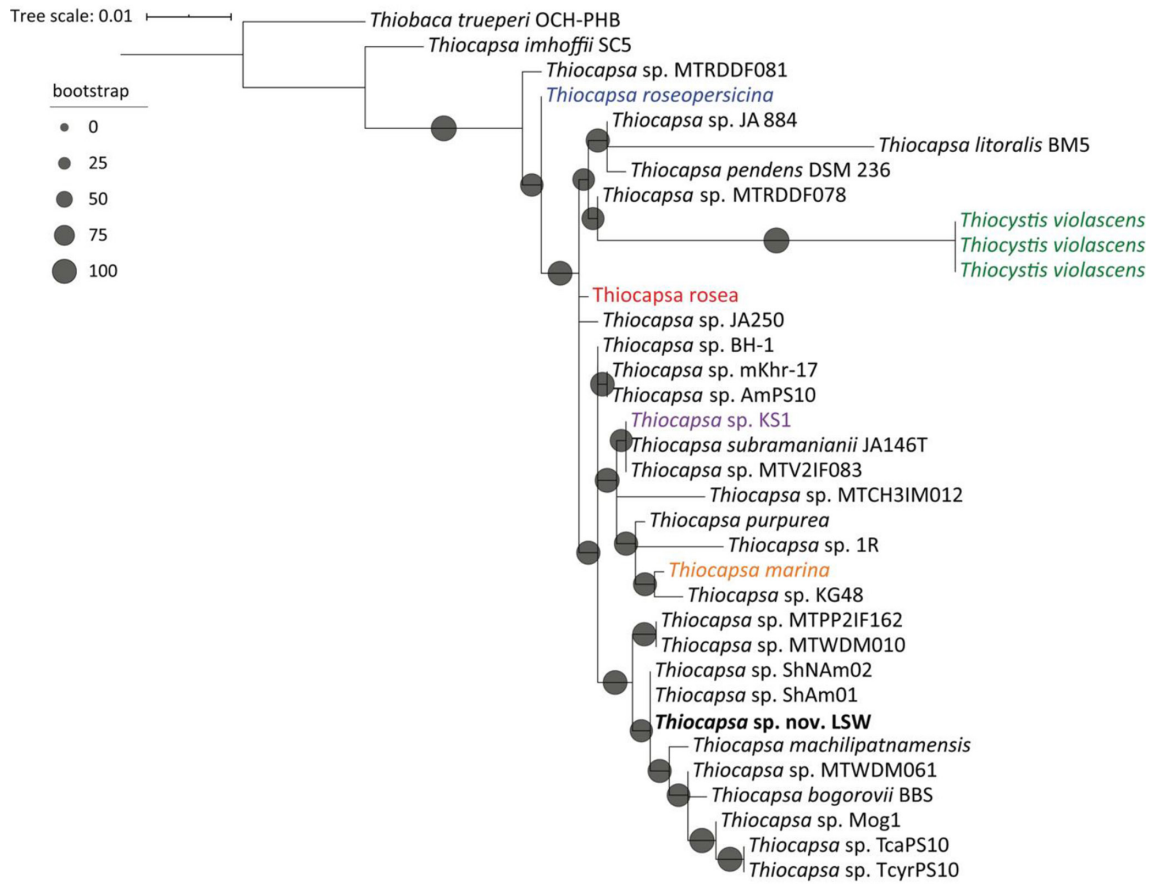
Briefly, 2 μ g of five individual single-cell amplified DNAs were sheared, then purification, end-repairing and ligation were performed. The library was sequenced using the Illumina HiSeq 2000 sequencing platform with the mode 2×150 bp at Yourgene Bioscience (Taiwan). The single-cell reads obtained were quality checked with FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), quality filtered with a minimum phred score of 20 and adapter trimmed using AdapterRemoval [41]. Reads with ambiguous bases were also removed. Quality filtered, non-ambiguous and trimmed reads were co-assembled with MEGAHIT [42] using *k*-mer sizes of 21, 33 and 55. Contigs obtained from co-assembly were subjected to automatic binning with CONCOCT [43] with minimum contig size set to 1000 bp. CONCOCT uses nucleotide *k*-mer frequencies and coverage data from multiple samples to perform an unsupervised binning. Binning was conducted to remove any contaminant sequences that the kit might have held from the data. Bins obtained from CONCOCT were then checked for their completeness and contamination using CheckM [44]. Bins with completeness $>50\%$ and contamination $\sim 1\%$ were selected for downstream analysis. Gene prediction was performed with Prokka, which uses Barrnap [45] (<https://github.com/tseemann/barrnap>) as the default for 16S rRNA gene prediction. ORFs were annotated by searching against the EggNOG (<http://eggnoг.embl.de>) and the Kyoto Encyclopedia of Genes and Genomes [46] databases using BLASTP (*E* value $\leq 10^{-5}$ and bit score ≥ 100).

Phylogenetic analysis

Available 16S rRNA gene sequences for *Thiocapsa* species were downloaded from the NCBI Taxonomy database (last accessed August 2020). We selected 35 sequences, including one 16S rRNA gene sequence of *Thiobaca trueperi* as the outgroup, for phylogenetic tree reconstruction. Sequence alignment was performed with cmalign from the infernal package [47], which performs a covariance model (CM)-guided small subunit rRNA alignment. The CM for the bacterial domain was downloaded from the rfam database (last accessed January 2020) [48]. The alignment was subjected to web-based Gblocks [49] to extract conserved blocks with default settings. A maximum-likelihood phylogenetic tree (with 914 bases) was computed using IQ-TREE v1.6.11 [50] with the TN+F+I model – selected based on Bayesian information criterion (BIC) values – and 1000 bootstraps. The tree obtained was visualized and edited in iTOL v4 [51].

In addition, a phylogenetic tree based on single-copy marker proteins was also reconstructed. Sequenced genomes of four *Thiocapsa* (*Thiocapsa* sp. UBA6158, *Thiocapsa roseopersicina* DSM 217, *Thiocapsa marina* 5811 and *Thiocapsa rosea* DSM235) and *Thiohalocapsa marina* DSM 19078 and *Thiocystis violascens* were downloaded from the NCBI Genomes database and a phylogeny was reconstructed using single-copy marker proteins and a two-step approach. Single-copy marker proteins were identified and aligned using ezTree [52]. The alignment obtained from ezTree was used to reconstruct a maximum-likelihood phylogeny with IQ-TREE with the JTTDCMut +F+I+G4 model – selected based on BIC values

(a)



(b)

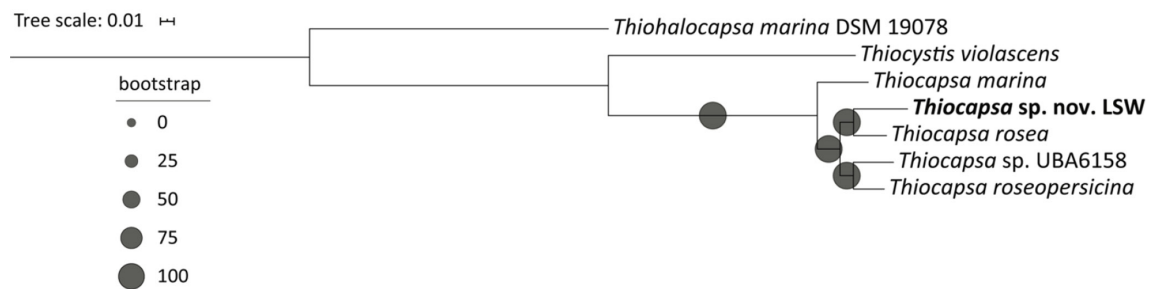


Fig. 1. Phylogenetic analyses of *Thiocapsa* sp. LSW based on (a) 16S rRNA genes, the sequences of which were downloaded from the NCBI Taxonomy database, with *Thiobaca trueperi* as the outgroup; (b) 248 single-copy marker genes. Sequenced genomes of *Thiocapsa* spp. were downloaded from the NCBI Genomes database. Both trees were reconstructed using the maximum-likelihood method with 1000 bootstraps in IQ-TREE, and visualized and formatted in iTOL. *Thiocapsa* with coloured text in (a) were also analysed for single-copy marker genes in (b).

– and 1000 bootstraps. A consensus tree was visualized and edited in iTOL v4.

RESULTS

Genome properties and phylogeny

More than 18 million high-quality reads were obtained and assembled into 563 contigs with 127× coverage. The assembly of the LSW single-cell genome resulted in a total size of 2.7 Mb (Table 1). The draft genome was about 50.39% complete with 1.12% contamination based on the CheckM results; therefore, the LSW was assumed to be approximately 5.4 Mb long.

One copy of the 16S rRNA gene was annotated in this draft genome, and a phylogenetic tree was built based on the 16S rRNA gene and the other three available *Thiocapsa*, three *Thiocystis violascens* and one *Thiobaca trueperi* 16S rRNA genes from the NCBI Taxonomy database (Fig. 1a). The phylogenetic tree shows that LSW is closely related to *Thiocapsa* sp. ShAm01 and ShNA02, which were isolated from Lake Shira and Lake Shunet, respectively. A heatmap was constructed based on the 16S rRNA gene identity (%) between the closest relatives of *Thiocapsa* (Figs 1a and S2). In addition, 248 single-copy marker proteins (Table S2) were identified in the five available genomes of *Thiocapsa* (including *Thiocapsa* sp. LSW from this study), and a phylogenetic tree based on single-copy marker proteins was reconstructed (Fig. 1b). LSW from this study was phylogenetically close to *Thiocapsa rosea*.

Phototrophy

Only one gene (*pufC*; cytochrome c subunit) involved in the RCII gene cluster of LSW was annotated, and it had 97% sequence identity to the closest sequenced strain, *Thiocapsa rosea* DSM 235. Multiple copies of genes (*pufA* and *pufB*) for the main antenna of purple bacteria (light-harvesting complex LHI) and photosynthetic reaction centre cytochrome c subunit (*pufC*) were also present. PucC is a subunit involved in the photosynthetic reaction centre that encodes RC-LHI auxiliary proteins, which are membrane-bound bacteriochlorophyll (BChl) synthesis enzymes. In addition, two copies of high-potential iron-sulfur proteins (HiPIPs), which generate an electrochemical potential coupled to ATP synthesis, were annotated. The genes related to the pigments that were detected in their entirety were required for the biosynthesis of spirilloxanthin, which might be the major carotenoid in LSW, with the exception of BChl (Fig. 2).

Carbon metabolism

LSW might undergo autotrophy due to CO₂ assimilation via the Calvin–Benson–Bassham cycle, which is nearly completely annotated in the draft genome. Two copies of *rbcL* and *rbcS* genes encoding ribulose-bisphosphate carboxylase (type I RuBisCO) large and small subunits were present. Moreover, genes encoding phosphoenolpyruvate carboxylase for phosphoenolpyruvate carboxylation involved in the C4-dicarboxylic acid cycle were also annotated. Carboxysome shell proteins and carbonic anhydrase for the conversion of

bicarbonate to carbon dioxide in carboxysomes were also detected.

Nitrogen metabolism

A complete set of *nif* genes for nitrogen fixation was annotated, including molybdenum-iron nitrogenase (NifDK), nitrogenase-stabilizing/protective protein (NifW) and nitrogenase reductase (NifH), suggesting that LSW acts as a diazotroph in Lake Shunet. Specifically, the complete *nifHDKT* operons of four *Thiocapsa* spp. including LSW were annotated, and the phylogenetic tree reconstructed based on *nifH* protein sequences (294 aa) indicated that LSW was functionally related to *Thiocapsa roseopersicina* in terms of nitrogen fixation (Fig. 3). Although no genes involved in assimilatory nitrate or nitrite reduction were annotated in the current draft genome, a gene encoding the nitrate/nitrite transporter (NRT) was detected. Therefore, LSW might utilize nitrate or nitrite as nitrogen sources. In addition, the presence of genes encoding urease (*ureABC*) and urea transport system permease (*urtB*) indicated another potential alternative mechanism by which the organism can obtain nitrogen. LSW had one candidate gene involved in the nitrification pathway: hydroxylamine dehydrogenase (*hao*); these and the potential nitrate/nitrite transporter suggest that the LSW might be able to perform complete nitrification. Four genes related to (NiFe) hydrogenases, which play a role in electron transport, were also present. Generally, LSW can utilize a range of organic and inorganic low-potential electron donors to conserve energy.

Sulfur metabolism

A wide variety of enzymes involved in sulfur metabolism were annotated in the draft genome of LSW, including those for assimilatory sulfate reduction, dissimilatory sulfate reduction and oxidation, and thiosulfate oxidation. In the oxidation systems, sulfur compounds are utilized as electron donors for anoxygenic photosynthesis. In the reduction systems, a nearly complete assimilatory sulfate reduction pathway via adenosine phosphosulfate, 3'-phosphoadenosine phosphosulfate was annotated, with elemental sulfur (S⁰) as the final product (Fig. 2), which might be stored in cellular sulfur globules. In terms of dissimilatory sulfate reduction, LSW may be capable of conducting the reduction via adenosine phosphosulfate and sulfite, and use sulfate as a terminal electron acceptor to produce H₂S. In further phylogenetic analyses based on key genes, the presence of *dsrA* for the sulfate reduction and *soxB* for the sulfate oxidation of five *Thiocapsa* spp. suggests that LSW is related to *Thiocapsa marina* and *Thiocapsa rosea*, respectively (Fig. S3).

DISCUSSION

Unique *Thiocapsa* sp. from the saline meromictic Lake Shunet

According to the NCBI Taxonomy database, about 31 *Thiocapsa* spp. have been described to date. They are ubiquitous in fresh, marine, alkaline, acidic and hot or cold waters, and

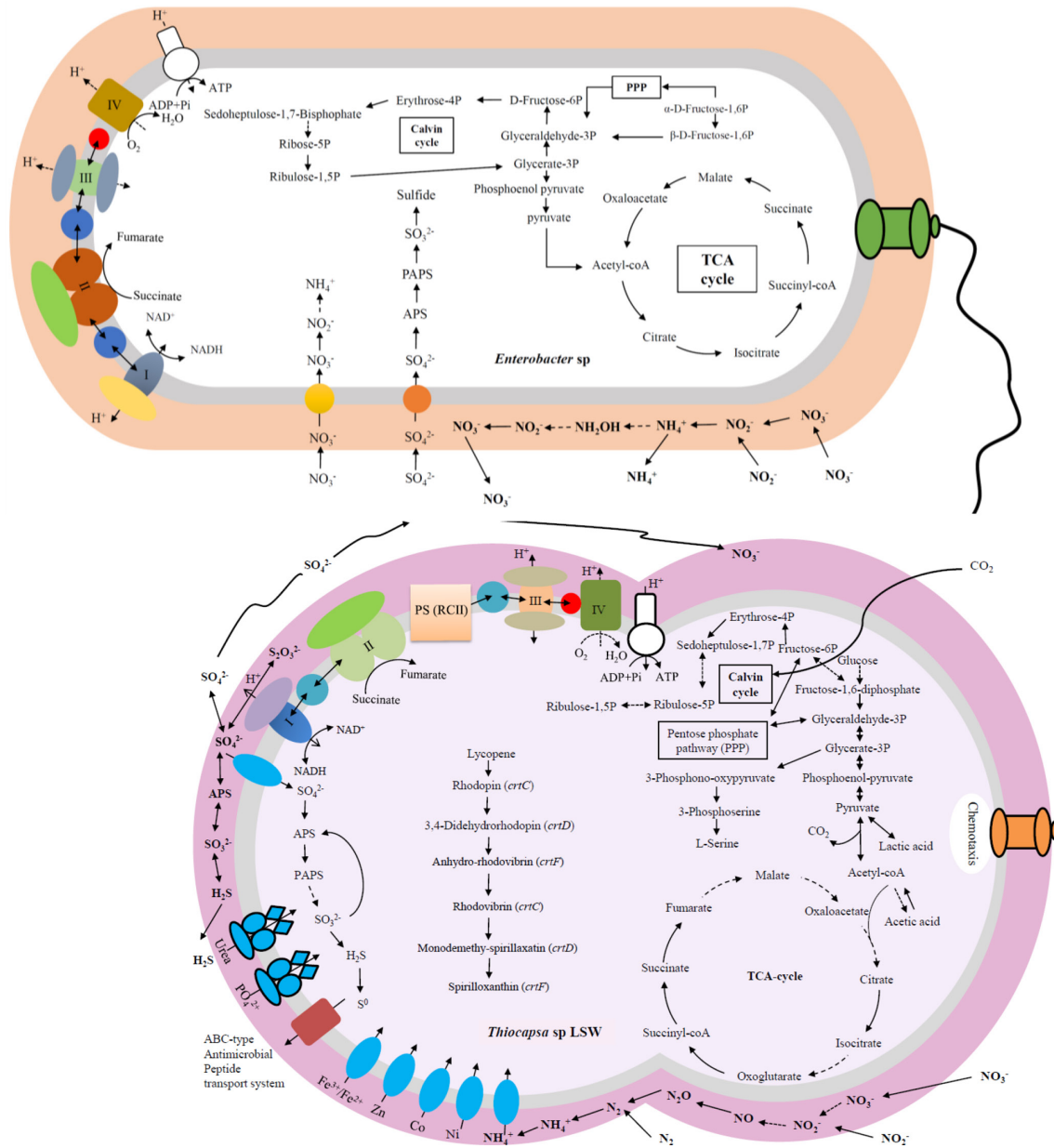


Fig. 2. Putative interactions between the two most prevalent groups in the chemocline layer in Lake Shunet: *Thiocapsa* sp. LSW (shown with a purple outermost layer) and *Enterobacter* sp. (shown with an orange outermost layer). The dotted line with the arrow denotes the genes that were not annotated in the draft genome.

include the isolates *Thiocapsa roseopersicina* BBS [53] from the White Sea, *Thiocapsa marina* 5811 from the Mediterranean Sea [25], *Thiocapsa imhoffii* from haloalkaline Soap Lake [54] and *Thiocapsa* KS1 from a sewage-treatment facility [27]. LSW from this study was isolated from the saline meromictic Lake Shunet and clustered independently with the isolates of *Thiocapsa* sp. ShNAM01 and ShAm01 (96% bootstrap) from Lake Shunet [28] and Lake Shira [55], respectively, in the reconstructed 16S rRNA gene phylogenetic tree. Isolates of *Thiocapsa* spp. from different geographical locations or ecosystems seem to differ based on specific environmental

factors and phylogenetics. Furthermore, a genome tree based on 248 single-copy marker genes was built from five available genomes of *Thiocapsa*: *Thiocapsa marina* 5811, *Thiocapsa roseopersicina* DSM217, *Thiocapsa* sp. UBA6158 [56], and *Thiocapsa rosea* DSM235 and LSW (current study). Among these, *Thiocapsa marina* 5811 was isolated from the Mediterranean Sea [25], while *Thiocapsa roseopersicina* DSM217, *Thiocapsa* sp. UBA6158 and *Thiocapsa rosea* DSM235 were isolated from sewage lagoons or wastewater ponds [56–58]. Based on the genome tree, the five *Thiocapsa* genomes represent specific roles in the ecosystems from which they were

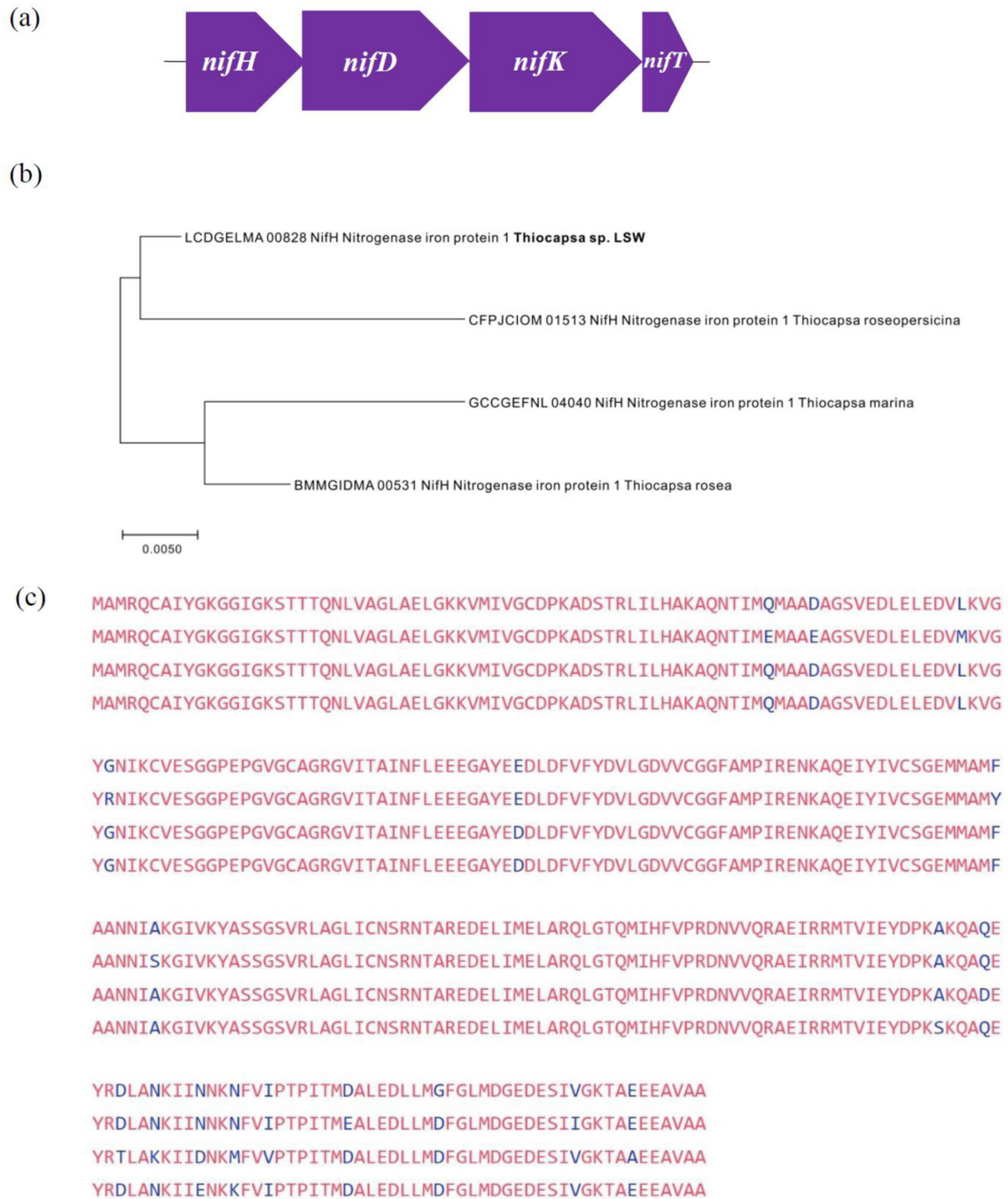


Fig. 3. Nitrogen fixation. (a) Schematic diagram of the complete *nifHDKT* operon of LSW and three *Thiocapsa* species. (b) Phylogenetic tree of *Thiocapsa* species based on the *nifH* gene and reconstructed using the maximum-likelihood method with 1000 bootstraps. (c) Alignment of *nifH* amino acid sequence of *Thiocapsa* species (from the top down, LSW, *Thiocapsa roseopersicina*, *Thiocapsa marina* and *Thiocapsa rosea*). Identical amino acids are highlighted in red; different amino acids are highlighted in blue.

isolated. LSW represents an independent cluster separate from the other four genomes, indicating that it may have an important ecological role in this saline meromictic lake.

More specifically, the complete *nifHDKT* operon in the nitrogen fixation gene cluster is annotated in LSW, along with three other *Thiocapsa* draft genomes: *Thiocapsa roseopersicina*,

Thiocapsa marina and *Thiocapsa rosea*. Unlike the phylogenetic analysis of 16S rRNA, LSW clustered instead closely to the type strain, *Thiocapsa roseopersicina*, on the basis of the *nifH* gene. LSW and *Thiocapsa roseopersicina* were from Lake Shunet and a wastewater pond from a sugar mill, respectively; both natural habitats may have common characteristics

specifically for nitrogen fixation. This also indicated that phylogenetic analysis based on 16S rRNA and *nifH* genes of *Thiocapsa* showed divergence. The finding was consistent with that of Zehr *et al.* [59], in which both 16S rRNA and *nifH* genes evolved at different rates.

Potential ecological role of LSW and its interaction with *Enterobacter* sp. in the chemocline

The highest number of phototrophic bacteria was recorded at a depth of 5.0 m; the bacteria included the green sulfur bacteria *Prosthecochloris vibrioformis*, PSBs *Thiocapsa* and *Halochromatium* species, and the purple nonsulfur bacteria *Rhodovulum euryhalinum* and *Pinkicyclus mahoneyensis*, among which *Thiocapsa* was the most dominant phototroph [28]. Our previous study [7] also indicated that the PSB LSW is a major anoxygenic phototrophic bacterium in the anoxic layers of Lake Shunet. Unlike *Thiocapsa marina*, which contains okenone as its main pigment [25], our annotation results suggest that the major carotenoid in LSW is spirilloxanthin, as it is in *Thiocapsa roseopersicina* [60, 61] and *Thiocapsa* KS1 [31].

The photosynthetic apparatus in many purple bacteria – including *Thiocapsa marina* DSM 5653 and *Thiocapsa* KS1 – consists of two light-harvesting antenna complexes (LHI and LHII) [12, 23]; however, only LHI was detected in the current draft genome of LSW. Although no genes responsible for the biosynthesis of BChl were annotated in the current draft genome, we suggest that LSW might utilize BChl α for phototrophic growth. This is because our results used PE-Cy7 for single-cell sorting and detected an emission wavelength in the near-infrared region at about 800 nm, which is within the spectrum characteristic of BChl α in purple bacteria, specifically *Chromatiaceae* [62], and matches the spectral characteristic fluorochromes of PE-Cy7, including near-infrared [63]. This was echoed in the specific absorption spectra of the pigments of *Thiocapsa* in the lake [28].

BChl α has distinct light absorbance from the chlorophyll of cyanobacteria, which might be one of the factors causing *Thiocapsa* sp. and cyanobacteria to colonize densely in the anoxic and oxic zones, respectively. Additionally, the genes encoding enzymes involved in the Calvin cycle were annotated in the draft genome, which supposed that *Thiocapsa* sp. grew photolithoautotrophically in Lake Shunet. This finding agrees with the work by Kondratieva [64], which showed that all *Chromatiaceae* employ the Calvin cycle when undergoing photolithoautotrophic growth. Therefore, LSW was specifically prevalent on the surface of the chemocline due to the light demands for phototrophic development, as there is little doubt that light is vital to LSW in this specific lake. Based on the gene annotation, which identified phosphoenolpyruvate carboxylase and carbonic anhydrases in the draft genome, LSW might enhance carbon fixation efficiency by phosphoenolpyruvate carboxylation and concentrating CO₂ [65, 66].

In terms of the sulfur cycle, LSW harbours genes involved in the oxidation of sulfide, sulfite and thiosulfate, which might provide electrons to carry out photosynthetic carbon dioxide

fixation to produce carbohydrates. This has been experimentally confirmed in anoxygenic phototrophic bacteria including *Chromatium okenii* and *Thiocapsa roseopersicina* BBS [53, 67].

In the metagenomics study, abundance was calculated as the read counts in the bin divided by the total read counts; based on this, *Thiocapsa* sp. and *Enterobacter* sp. (*Enterobacteriaceae*) were the two most abundant bacterial groups in the chemocline of Lake Shunet. The metabolisms conducted by *Enterobacter* sp. were reconstructed (Fig. 2) based on the gene annotation of the assembled genomic bin [7]. As the prevailing bacteria in the chemocline, they might play an important role in nitrogen cycling in the layer, considering that the complete dissimilatory nitrate reduction pathway and partial gene involved in nitrification were annotated (described by Bonin [68]). In terms of the sulfur cycle, only a complete assimilatory sulfate reduction pathway was annotated. The sulfur cycle was reconstructed based on the gene annotation predicted from metagenomes in the chemocline of Lake Shunet, and *Thiocapsa* sp. seemed to be the only bacterium that could undergo dissimilatory sulfate oxidation [7]. The genes involved in the production of SO₄²⁻ were detected in the draft genome of LSW from this study, which is in accordance with work by Trüper and Fischer [69]. Therefore, LSW might contribute SO₄²⁻ to other bacteria, including *Enterobacter* sp., which carries out assimilatory sulfate reduction (Fig. 2). In addition, complete genes that encode steps in the nitrification process were annotated in the draft genome of LSW, and LSW likely provides nitrate or nitrite for *Enterobacter* sp. to undergo the nitrate reduction.

A previous metagenomic analysis [7] suggested that *Thiocapsa* sp. oxidizes sulfide in the chemocline and generates sulfite, thiosulfate and sulfate for other microbes in the layer, and the draft genome from the present study confirms this. Therefore, LSW may play a role as both a sulfate reducer and oxidizer in Lake Shunet. In terms of nitrogen metabolism, LSW can undergo denitrification and nitrogen fixation, and preferentially utilizes ammonium as a nitrogen source; this echoes work by Brown and Herbert [70], which showed that ammonium is the preferred nitrogen source for PSBs. Consequently, LSW as a diazotroph in Lake Shunet might also contribute ammonium to other microbes.

Funding information

This research was funded by the Ministry of Science and Technology in Taiwan through the Taiwan–Russia Joint Project Grant NSC102-2923-B-001-004-MY3 and by the Russian Foundation for Basic Research Grant (project no. 21-54-52001) and MOST (105-2923-B-001-001-MY3).

Acknowledgements

We thank our Russian colleagues at the Institute of Biophysics, Siberian Division of the Russian Academy of Sciences, Russia, for their assistance with sampling, and Ms Su-Hsing Huang for her assistance operating the MoFlo XDP Cell Sorter at the Institute of Plant and Microbial Biology, Academia Sinica, Taiwan. We also thank Noah Last of Third Draft Editing for his English language editing.

Author contributions

Conceptualization, S.-L.T. and Y.-T.W.; field sampling, P.-W. C., D.Y.R. and A.G.D.; methodology, Y.-T.W.; bioinformatics, K.T.; software, Y.-T.W. and K.T.; formal analysis, Y.-T.W.; investigation, S.-L.T., Y.-T.W. and K.T.;

resources, S.-L.T.; data curation, Y.-T.W. and K.T.; writing – original draft preparation, Y.-T.W.; writing – review and comment, S.-L.T., K.T. and D.Y.R.; writing – editing and finalization, Y.-T.W.; visualization, Y.-T.W.; supervision, S.-L.T.; funding acquisition, S.-L.T.

Conflicts of interest

The authors declare that there are no conflicts of interest.

References

1. Lauro FM, DeMaere MZ, Yau S, Brown MV, Ng C, et al. An integrative study of a meromictic lake ecosystem in Antarctica. *ISME J* 2011;5:879–895.
2. Comeau AM, Harding T, Galand PE, Vincent WF, Lovejoy C. Vertical distribution of microbial communities in a perennially stratified Arctic lake with saline, anoxic bottom waters. *Sci Rep* 2012;2:604.
3. Andrei A-Ş, Robeson MS 2nd, Baricz A, Coman C, Muntean V, et al. Contrasting taxonomic stratification of microbial communities in two hypersaline meromictic lakes. *ISME J* 2015;9:2642–2656.
4. Zimmermann M, Escrig S, Hübschmann T, Kirf MK, Brand A, et al. Phenotypic heterogeneity in metabolic traits among single cells of a rare bacterial species in its natural environment quantified with a combination of flow cell sorting and NanoSIMS. *Front Microbiol* 2015;6:243.
5. Nobu MK, Dodsworth JA, Murugapiran SK, Rinke C, Gies EA, et al. Phylogeny and physiology of candidate phylum 'Atribacteria' (OP9/JS1) inferred from cultivation-independent genomics. *ISME J* 2015;10:273–286.
6. Baatar B, Chiang P-W, Rogozin DY, Wu Y-T, Tseng C-H, et al. Bacterial communities of three saline meromictic lakes in central Asia. *PLoS One* 2016;11:e0150847.
7. Wu Y-T, Yang C-Y, Chiang P-W, Tseng C-H, Chiu H-H, et al. Comprehensive insights into composition, metabolic potentials, and interactions among archaeal, bacterial, and viral assemblages in meromictic Lake Shunet in Siberia. *Front Microbiol* 2018;9:1763.
8. Klepac-Ceraj V, Hayes CA, Gilhooly WP, Lyons TW, Kolter R, et al. Microbial diversity under extreme euxinia: Mahoney Lake, Canada. *Geobiology* 2012;10:223–235.
9. Danza F, Ravasi D, Storelli N, Roman S, Lüdin S, et al. Bacterial diversity in the water column of meromictic Lake Cadagno and evidence for seasonal dynamics. *PLoS One* 2018;13:e0209743.
10. Lunina ON, Savvichev AS, Babenko VV, Boldyreva DI, Kuznetsov BB, et al. Seasonal variations in the structure of an anoxygenic phototrophic bacterial community from the meromictic Lake Trekhtsvetnoe (Kandalaksha Bay, White Sea). *Microbiology* 2019;88:100–114.
11. Overmann J, Beatty JT, Hall KJ. Purple sulfur bacteria control the growth of aerobic heterotrophic bacterioplankton in a meromictic salt lake. *Appl Environ Microbiol* 1996;62:3251–3258.
12. Imhoff JF. The chromatiaceae. In: Dworkin M, Falkow S, Rosenberg E, Schleifer KH and Stackebrandt E (eds). *The Prokaryotes*. New York: Springer; 2006. pp. 846–873.
13. Casamayor EO, García-Cantizano J, Pedrós-Alió C. Carbon dioxide fixation in the dark by photosynthetic bacteria in sulfide-rich stratified lakes with oxic-anoxic interfaces. *Limnol Oceanogr* 2008;53:1193–1203.
14. Casamayor EO, Llírs M, Picazo A, Barberán A, Borrego CM, et al. Contribution of deep dark fixation processes to overall CO₂ incorporation and large vertical changes of microbial populations in stratified karstic lakes. *Aquat Sci* 2011;74:61–75.
15. Berg JS, Pjevac P, Sommer T, Buckner CRT, Philippi M, et al. Dark aerobic sulfide oxidation by anoxygenic phototrophs in anoxic waters. *Environ Microbiol* 2019;21:1611–1626.
16. Pjevac P, Dyksma S, Goldhammer T, Mujakić I, Koblížek M, et al. In situ abundance and carbon fixation activity of distinct anoxygenic phototrophs in the stratified seawater Lake Rogoznica. *Environ Microbiol* 2019;21:3896–3908.
17. Savvichev AS, Kadnikov VV, Rusanov II, Beletsky AV, Krasnova ED, et al. Microbial processes and microbial communities in the water column of the polar meromictic Lake Bol'shie Khruslomeny at the White Sea Coast. *Front Microbiol* 2020;11:1945.
18. Chan Y-F, Chiang P-W, Tandon K, Rogozin D, Degermendzhi A, et al. Spatiotemporal changes in the bacterial community of the meromictic Lake Uchum, Siberia. *Microb Ecol* 2021;81:357–369.
19. Di Nezio F, Beney C, Roman S, Danza F, Buetti-Dinh A, et al. Anoxygenic photo- and chemo-synthesis of phototrophic sulfur bacteria from an alpine meromictic lake. *FEMS Microbiol Ecol* 2021;97:fiab010.
20. Rogozin DY, Zykov VV, Ivanova EA, Anufrieva TN, Barkhatov YV, et al. Meromixis and seasonal dynamics of vertical structure of Lake Uchum (South Siberia). *Contemp Probl Ecol* 2018;11:195–206.
21. Rogozin DY, Zykov VV, Degermendzhi AG. Ecology of purple sulfur bacteria in the highly stratified meromictic Lake Shunet (Siberia, Khakassia) in 2002–2009. *Microbiology* 2012;81:727–735.
22. Kallistova AYU, Kevbrina MV, Pimenov NV, Rusanov II, Rogozin DYU, et al. Sulfate reduction and methanogenesis in the Shira and Shunet meromictic lakes (Khakassia, Russia). *Microbiology* 2006;75:720–726.
23. Lauro FM, DeMaere MZ, Yau S, Brown MV, Ng C, et al. An integrative study of a meromictic lake ecosystem in Antarctica. *ISME J* 2011;5:879–895.
24. Sommer T, Danza F, Berg J, Sengupta A, Constantinescu G, et al. Bacteria-induced mixing in natural waters. *Geophys Res Lett* 2017;44:9424–9432.
25. Caumette P, Guyoneaud R, Imhoff JF, Süling J, Gorlenko V. *Thiocapsa marina* sp. nov., a novel, okenone-containing, purple sulfur bacterium isolated from brackish coastal and marine environments. *Int J Syst Evol Microbiol* 2004;54:1031–1036.
26. Guyoneaud R, Süling J, Petri R, Matheron R, Caumette P, et al. Taxonomic rearrangements of the genera *Thiocapsa* and *Amoebobacter* on the basis of 16S rDNA sequence analyses, and description of *Thiolamproyum* gen. nov. *Int J Syst Bacteriol* 1998;48:957–964.
27. Schott J, Griffin BM, Schink B. Anaerobic phototrophic nitrite oxidation by *Thiocapsa* sp. strain KS1 and *Rhodospseudomonas* sp. strain LQ17. *Microbiology* 2010;156:2428–2437.
28. Lunina ON, Bryantseva IA, Akimov VN, Rusanov II, Rogozin DYU, et al. Seasonal changes in the structure of the anoxygenic photosynthetic bacterial community in Lake Shunet, Khakassia. *Microbiology* 2007;76:368–379.
29. Thrash JC. Towards culturing the microbe of your choice. *Environ Microbiol Rep* 2021;13:36–41.
30. Marcy Y, Ouverney C, Bik EM, Lösekann T, Ivanova N, et al. Dissecting biological "dark matter" with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth. *Proc Natl Acad Sci USA* 2007;104:11889–11894.
31. Hemp J, Lückner S, Schott J, Pace LA, Johnson JE, et al. Genomics of a phototrophic nitrite oxidizer: insights into the evolution of photosynthesis and nitrification. *ISME J* 2016;10:2669–2678.
32. Kashtan N, Roggensack SE, Rodrigue S, Thompson JW, Biller SJ, et al. Single-cell genomics reveals hundreds of coexisting subpopulations in wild prochlorococcus. *Science* 2014;344:416–420.
33. Mason OU, Han J, Woyke T, Jansson JK. Single-cell genomics reveals features of a *Colwellia* species that was dominant during the deepwater horizon oil spill. *Front Microbiol* 2014;5:332.
34. Stepanauskas R. Single cell genomics: an individual look at microbes. *Curr Opin Microbiol* 2012;15:613–620.
35. Casamayor EO, Ferrera I, Cristina X, Borrego CM, Gasol JM. Flow cytometric identification and enumeration of photosynthetic sulfur bacteria and potential for ecophysiological studies at the single-cell level. *Environ Microbiol* 2007;9:1969–1985.
36. Rogozin DY, Degermendzhi AG. Hydraulically-operated thin-layer sampler for sampling heterogeneous water columns. *J Sib Fed Univ, Biol* 2008;1:111–117.
37. Siegl A, Kamke J, Hochmuth T, Piel J, Richter M, et al. Single-cell genomics reveals the lifestyle of Poribacteria, a candidate

- phylum symbiotically associated with marine sponges. *ISME J* 2011;5:61–70.
38. Hamp TJ, Jones WJ, Fodor AA. Effects of experimental choices and analysis noise on surveys of the “rare biosphere”. *Appl Environ Microbiol* 2009;75:3263–3270.
 39. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012;19:455–477.
 40. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, et al. DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* 2007;57:81–91.
 41. Lindgreen S. Adapter removal: easy cleaning of next-generation sequencing reads. *BMC Res Notes* 2012;5:337.
 42. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 2015;31:1674–1676.
 43. Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, et al. Binning metagenomic contigs by coverage and composition. *Nat Methods* 2014;11:1144–1146.
 44. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 2015;25:1043–1055.
 45. Seemann T. Barrnap 0.7: Rapid Ribosomal RNA Prediction. 2013. <https://github.com/tseemann/barrnap>
 46. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* 2012;40:D109–D114.
 47. Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* 2013;29:2933–2935.
 48. Nawrocki EP. Annotating functional RNAs in genomes using Infernal. *Methods Mol Biol* 2014;1097:163–197.
 49. Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* 2007;56:564–577.
 50. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-Tree: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 2015;32:268–274.
 51. Letunic I, Bork P. Interactive tree of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* 2019;47:W256–W259.
 52. Wu Y-W. ezTree: an automated pipeline for identifying phylogenetic marker genes and inferring evolutionary relationships among uncultivated prokaryotic draft genomes. *BMC Genomics* 2018;19:921.
 53. Bogorov LV. The properties of *Thiocapsa roseopersicina*, strain BBS, isolated from an estuary of the White Sea. *Mikrobiologija* 1974;43:326–332.
 54. Asao M, Takaichi S, Madigan MT. *Thiocapsa imhoffii*, sp. nov., an alkaliphilic purple sulfur bacterium of the family Chromatiaceae from Soap Lake, Washington (USA). *Arch Microbiol* 2007;188:665–675.
 55. Lunina ON, Bryantseva IA, Akimov VN, Rusanov II, Barinova ES, et al. Anoxygenic phototrophic bacterial community of Lake Shira (Khakassia). *Microbiology* 2007;76:469–479.
 56. Parks DH, Rinke C, Chuvochina M, Chaumeil P-A, Woodcroft BJ, et al. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat Microbiol* 2017;2:1533–1542.
 57. Pfennig N. Ecology of phototrophic purple and green sulfur bacteria. In: Schlegel HG and Bowien B (eds). *Autotrophic Bacteria*. New York: Springer-Verlag; 1989. pp. 97–116.
 58. Pfennig N. Genus *Chromatium*. In: Staley JT, Bryant MP, Pfennig N and Holt JC (eds). *Bergey's Manual of Systematic Bacteriology*, 1st edn, vol. 3. Baltimore, MD: Williams and Wilkins; 1989. pp. 1639–1643.
 59. Zehr JP, Jenkins BD, Short SM, Steward GF. Nitrogenase gene diversity and microbial community structure: a cross-system comparison. *Environ Microbiol* 2003;5:539–554.
 60. Schmidt K. Biosynthesis of carotenoids. In: Clayton RK and Sistrom WR (eds). *The Photosynthetic Bacteria*. New York: Plenum Press; 1978. pp. 729–750.
 61. Kovács ÁT, Rákhely G, Kovács KL. Genes involved in the biosynthesis of photosynthetic pigments in the purple sulfur photosynthetic bacterium *Thiocapsa roseopersicina*. *Appl Environ Microbiol* 2003;69:3093–3102.
 62. Olson JM, Stanton EK. Absorption and fluorescence spectra of bacterial chlorophylls in situ. In: Vernon LP and Seely GR (eds). *The Chlorophylls*. New York: Academic Press; 1966. pp. 381–398.
 63. Baumgarth N, Bigos M. Optimization of emission optics for multi-color flow cytometry. *Methods Cell Biol* 2004;75:3–22.
 64. Kondratieva EN. Interrelation between modes of carbon assimilation and energy production in phototrophic purple and green bacteria. In: Quale JR (ed). *Microbial Biochemistry: International Review of Biochemistry*, vol. 21. Baltimore, MD: University Park Press; 1979. pp. 117–175.
 65. Yeates TO, Kerfeld CA, Heinhorst S, Cannon GC, Shively JM. Protein-based organelles in bacteria: carboxysomes and related microcompartments. *Nat Rev Microbiol* 2008;6:681–691.
 66. Kimber MS. Carboxysomal carbonic anhydrases. In: Frost SC and McKenna R (eds). *Carbonic Anhydrase: Mechanism, Regulation, Links to Disease, and Industrial Applications*. Dordrecht: Springer; 2014. pp. 89–103.
 67. Trüper HG. CO₂-Fixierung und Intermediärstoffwechsel bei *Chromatium okenii* Perty. *Archiv Mikrobiol* 1964;49:23–50.
 68. Bonin P. Anaerobic nitrate reduction to ammonium in two strains isolated from coastal marine sediment: a dissimilatory pathway. *FEMS Microbiol Ecol* 1996;19:27–38.
 69. Trüper HG, Fischer U. Anaerobic oxidation of sulphur compounds as electron donors for bacterial photosynthesis. *Phil Trans R Soc Lond B* 1997;298:529–542.
 70. Brown CM, Herbert RA. Ammonia assimilation in purple and green sulphur bacteria. *FEMS Microbiol Lett* 1977;1:39–42.

Five reasons to publish your next article with a Microbiology Society journal

1. The Microbiology Society is a not-for-profit organization.
2. We offer fast and rigorous peer review – average time to first decision is 4–6 weeks.
3. Our journals have a global readership with subscriptions held in research institutions around the world.
4. 80% of our authors rate our submission process as ‘excellent’ or ‘very good’.
5. Your article will be published on an interactive journal platform with advanced metrics.

Find out more and submit your article at microbiologyresearch.org.