Complete genome sequence of "Thiodictyon syntrophicum" sp. nov. strain Cad16T, a photolithoautotrophic purple sulfur bacterium isolated from the alpine meromictic Lake Cadagno
Luedin, Samuel M.; Pothier, Joël F.; Danza, Francesco; Storelli, Nicola; Frigaard, Niels-Ulrik; Wittwer, Matthias; Tonolla, Mauro

Published in:
Standards in Genomic Sciences

DOI:
10.1186/s40793-018-0317-z

Publication date:
2018

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY

Citation for published version (APA):
Luedin, S. M., Pothier, J. F., Danza, F., Storelli, N., Frigaard, N-U., Wittwer, M., & Tonolla, M. (2018). Complete genome sequence of "Thiodictyon syntrophicum" sp. nov. strain Cad16T, a photolithoautotrophic purple sulfur bacterium isolated from the alpine meromictic Lake Cadagno. Standards in Genomic Sciences, 13, [14]. https://doi.org/10.1186/s40793-018-0317-z
Complete genome sequence of “Thiodictyon syntrophicum” sp. nov. strain Cad16T, a photolithoautotrophic purple sulfur bacterium isolated from the alpine meromictic Lake Cadagno

Samuel M. Luedin1,2,3, Joël F. Pothier4, Francesco Danza1,2, Nicola Storelli1,2, Niels-Ulrik Frigaard5, Matthias Wittwer3 and Mauro Tonolla1,2*

Abstract

“Thiodictyon syntrophicum” sp. nov. strain Cad16T is a photoautotrophic purple sulfur bacterium belonging to the family of Chromatiaceae in the class of Gammaproteobacteria. The type strain Cad16 was isolated from the chemocline of the alpine meromictic Lake Cadagno in Switzerland. Strain Cad16T represents a key species within this sulfur-driven bacterial ecosystem with respect to carbon fixation. The 7.74-Mbp genome of strain Cad16T has been sequenced and annotated. It encodes 6237 predicted protein sequences and 59 RNA sequences. Phylogenetic comparison based on 16S rRNA revealed that Thiodictyon elegans strain DSM 232T the most closely related species. Genes involved in sulfur oxidation, central carbon metabolism and transmembrane transport were found. Noteworthy, clusters of genes encoding the photosynthetic machinery and pigment biosynthesis are found on the 0.48 Mb plasmid pTs485. We provide a detailed insight into the Cad16T genome and analyze it in the context of the microbial ecosystem of Lake Cadagno.

Keywords: Phototrophic sulfur bacteria, Chromatiaceae, Sulfur cycling, Meromictic lake, CRISPR, Okenone

Introduction

PSB belonging to the family of Chromatiaceae are generally found at the interface of aerobic and sulfidic-anaerobic zones that are exposed to sunlight such as stagnant, hypertrophic water bodies, littoral zones and bacterial mats [1]. The genus Thiodictyon was first described by Winogradsky in 1888 [2] and comprises two type strains, Thiodictyon elegans strain DSM 232T and Thiodictyon bacillosum strain DSM 234T. “Thiodictyon syntrophicum” sp. nov. strain Cad16T is the proposed type strain of the species Thiodictyon syntrophicum [3] within the family of Chromatiaceae of the genus Thiodictyon [4]. Cultures of strain Cad16T were isolated from the chemocline of the alpine meromictic Lake Cadagno (Ticino, Switzerland). This lake is characterized by high influx of sulfate, magnesium and calcium in the euxinic monimolimnion which favors the formation of a steep chemocline at 10 to 14 m depth [5, 6]. Within this zone a dense population (up to 107 cells per ml in summer) of mainly anaerobic phototrophic sulfur bacteria belonging to the PSB genera Chromatium, Lamprocystis, Thiodictyon, Thiocystis, and the GSB Chlorobium [7] is responsible for up to 40% of the total CO2 fixation measured in Lake Cadagno [8]. Strain Cad16T has been shown to be highly active in CO2 fixation both in situ and in vitro [9]. Furthermore, aggregation of strain Cad16T with SRBof the genus Desulfofapsa has been described [3]. In this publication we describe the first complete genome of strain Cad16T providing details especially on CO2 fixation, sulfur metabolism and on CRISPRs. The sequencing of strain Cad16T is part of a larger sequencing project that includes the key species of...
the microbial community from the anoxic layers of Lake Cadagno.

Organism information
Classification and features
Strain Cad16ᵀ is Gram-negative, the cells are oval-sphere shaped and 1.4–2.4 μm in diameter, non-motile, vacuolated and contain BChl a. Isolate Cad16ᵀ can grow as single cells, as well as in cell aggregates with up to 100 cells contained in EPS layer (Fig. 1). It was isolated from the chemocline of Lake Cadagno in a depth of 10–14 m where it grows in a non-obligate mutualistic association with sulfur-reducing bacteria of the genus Desulfocapsa [10]. Based upon morphology and partial 16S rRNA sequence analysis, the strain Cad16ᵀ was classified as a member of the genus Thiodictyon within the family Chromatiaceae before [10]. Figure 2 shows the phylogenetic placement of strain Cad16ᵀ (complete 16S rRNA sequence) in a 16S rRNA based maximum likelihood phylogenetic tree. The closest relatives of isolate Cad16ᵀ are T. bacillosum DSM 234ᵀ and T. elegans DSM 232ᵀ with 99% sequence identity (partial 16S rRNA sequences). A comparison of the strain Cad16ᵀ core genome with other whole genome sequenced PSB confirmed the phylogenetic placement (Additional file 1: Figure S1).

Strain Cad16ᵀ was anaerobically grown in Pfennigs medium [11], containing per liter: 0.25 g KH₂PO₄, 0.34 g NH₄Cl, 0.5 g MgSO₄·7H₂O, 0.25 g CaCl₂·2H₂O, 0.34 g KCl, 1.5 g NaHCO₃, 0.5 ml trace element solution SL₁₀, and 0.02 mg vitamin B₁₂ with 2 mM acetate in 100 mL serum bottles with rubber stoppers. The medium was prepared in a 2 L bottle with a N₂/CO₂ (80%/20%) gas phase. The medium was then reduced with 0.3 g l⁻¹ Na₂S·9H₂O (1.10 mM final concentration) and adjusted to a pH of 7.2. Cultures were incubated at 20–23 °C under photoheterotrophic conditions with 6 h light/dark photoperiods with a 40-W tungsten bulb placed at a distance of 60 cm from the cultures (ca. 10 μE m⁻² s⁻¹).

Different electron donors and carbon substrates were tested under phototrophic conditions by Peduzzi et al. [3, 10]. Photolithoautotrophic growth was observed under anoxic conditions with hydrogen sulfide, thiosulfate and elemental sulfur as electron donors. Thereby, elemental sulfur is stored within the periplasma as intermediate oxidation product (Fig. 1). The carbon sources acetate, butyrate, ethanol, formate, fructose, fumarate, glucose, glycerol, lactate, malate, propanol, propionate, pyruvate and succinate were added at 5 mM concentration, respectively. Strain Cad16ᵀ was observed to assimilate only acetate, pyruvate and fructose in the presence of sulfate and bicarbonate. Strain Cad16ᵀ was additionally tested for chemolithoautotrophic growth with bicarbonate under a headspace atmosphere containing 5% O₂, 10% CO₂ and 85% N₂, in the dark. Growth was observed with 0.02% hydrogen sulfid and 0.07% thiosulfate, or with 0.07% sulfide only, respectively. The pigments responsible for the purple-red color of strain Cad16ᵀ were analysed spectrometrically in vivo by Peduzzi et al. [3]. Local absorption maxima at 833 nm, 582 nm and 374 nm gave evidence for the presence of BChl a, and a at 528 nm for the carotenoid okenone, respectively [10].

A further characterization of strain Cad16ᵀ can be found in Table 1.

A circular representation of the genome sequence and annotation according to the COG criteria is shown in Fig. 3.

Genome sequencing information
Genome project history
Sampling was done in August 2001 using a Friedinger-type bottle on Lake Cadagno. Subsequent isolation and cultivation of strain Cad16ᵀ was done in Pfennig’s medium I [11]. gDNA was isolated in November 2014 and sequencing was performed in January 2015. Raw data was assembled in with the SMRTview assembly platform and annotated using the NCBI Prokaryotic Genome Annotation Pipeline. Completeness of the isolate Cad16ᵀ sequence was verified using the 31 single copy genes of the Amphora Net analysis platform [12].

The genome sequence was deposited in GenBank under BioProject PRJNA354524, with the accession numbers CP020370–CP020372. The key elements of the genome studied are listed in Table 2.

Growth conditions and genomic DNA preparation
Strain Cad16ᵀ was anaerobically grown in Pfennigs medium [11] Cells were collected by centrifugation for 15 min at 10,600 g. DNA was extracted using phenol/
chloroform/isoamylalcohol solution (25:24:1, v/v, Sigma, Buchs, Switzerland) following the protocol provided by Pacific Biosciences [13] in combination with phase lock gels (VWR International). gDNA was purified using AMPure beads (Agencourt, Beckman Coulter Life Sciences, Indianapolis, USA) following the E2612 protocol form New England Biolabs [14]. Purity of the DNA was tested using the Qbit UV/VIS absorption reader (Thermo Fisher Scientific, Rheinach, Switzerland).

**Genome sequencing and assembly**

The library construction and genome sequencing was done on the Pacific Biosciences RS II platform at the Functional Genomic Center Zurich, Zurich, Switzerland. A 10 kb SMRTbell library was constructed using the DNA Template Prep Kit 1.0 (Pacific Biosciences, Menlo Park, USA). SMRTbell template fragments over 10 kb length were used for creating a SMRT bell-Polymerase Complex with P6-C4 chemistry (Pacific Biosciences) according to the manufacturer instructions.

Four SMRT cells v3.0 (Pacific Biosciences) for PacBio RS II chemistry were used for sequencing. Separate sequencing quality reports for all four cells were created through the SMRT portal software.

The SMRT web portal was used for genome assembly with the RS_HGAP_Assembly.2 pipeline from the SMRT Analysis 2.3 server. The polished assembly consists of 153 scaffolds with a mean coverage of 175× and a N50 value of 6,849,178. Thereof, three scaffolds were distinctly longer (6.85, 0.50 and 0.43 Mb, respectively) and showed a coverage greater than 200×, whereas mean coverage dropped below a value of 50× for the remaining 150 scaffolds.

---

**Fig. 2** Phylogenetic tree indicating the position of “T. syntrophicum” sp. nov. strain Cad167 relative to other genomes sequenced within the family Chromatiaceae. The tree was inferred from 16S rRNA sequences (≥ 1300 bp) using the maximum likelihood algorithm of the IQ-TREE software [62]. 1000 bootstrap replicates were performed. Numbers at the nodes indicate the SH-aLRT support (%) and ultrafast bootstrap support (%), respectively. Asterisk mark strains with complete genome sequences available. Open circles as node shapes indicate bootstrap support values below 50%.
These three scaffolds showed self-similar ends in dot-plot graphs and could be circularized manually.

The genome was manually corrected for SNPs using MiSeq Illumina 300-bp paired-end reads from previous sequencing (unpublished data, N. Storelli, J.F. Pothier, M. Tonolla).

### Genome annotation

NCBI Prokaryotic Genome Annotation Pipeline (Annotation Software revision 4.1) NCBI Prokaryotic Genome Annotation Pipeline (Annotation Software revision 4.1) was used for gene calling and gene annotation. To identify CRISPR-Cas sequences the CRISPRFinder server was used [15]. The Pfam-A v29 database was used to predict Pfam domains [16]. Transmembrane domains were predicted with the webserver based TMHMM2 program [17] and signal peptides were predicted with SignalP 4.1 server [18].

### Genome properties

The complete genome of strain Cad16T comprises one circular chromosome (6,837,296 bp) and two circular plasmids pTs485 (484,824 bp) and pTs417 (416,864 bp) (Table 3). The average GC content for the chromosome, and plasmids pTs485 and pTs417, is 66.28%, 65.59 and 65.97%, respectively. A total of 6601 coding sequences were predicted. Thereof, 6237 were predicted to encode proteins whereas six rRNA, 49 tRNA and four ncRNA sequences were predicted. A putative function is assigned for 3471 (46.57%) protein encoding genes (Table 4). The classification of genes into COGs functional categories is given in Table 5. The replicons pTs485 and pTs417 could be made circular, have their own origin of replication each, but do not contain any RNA or house-keeping genes. Therefore, to our understanding, both pTs485 and pTs417 fulfill the plasmid definition.
Fig. 3 Schematic representation of the complete genome sequence of "Thiodictyon syntrophicum" sp. nov. strain Cad16T and gene classification according to the COG criteria. The genome comprises one circular chromosome (a) and two circular plasmids, pTs485 (b) and pTs417 (c). Beginning with the outermost circle, tracks are shown in the order of: (1) predicted PGC on pTs485 (green), (2) CRISPR arrays on the chromosome and pTs485 (alternating in blue and red) (3) protein coding genes on forward strand colored according to COG categories, (4) CDS (blue), tRNA (orange) and rRNA (violet) on forward strand, (5) CDS (blue), tRNA (orange) and rRNA (violet) on reverse strand, (6) protein coding genes on reverse strand colored according to COG categories, (7) GC content (black), (8) positive and negative GC skew (green and purple, respectively) and (9) genome region by kbp. GCView [63] was used to create this genome map.
Phototrophy

PSB typically transform light energy into chemical energy with the membrane bound type 2 photochemical reaction center. The chromosome of strain Cad16T encodes the core antenna proteins LH1, subunits PufA and PufB (THSYN_31145 and THSYN_31140), and the regulatory protein PufQ (THSYN_31110) upstream to the reaction center genes composed of reaction RC subunits PufL, PufM, and PufC (THSYN_31125–31,135). Additional two copies of subunits LH2 alpha and beta (THSYN_31115 and THSYN_31120), respectively, are encoded further downstream, as well as pairwise in two other clusters (THSYN_30995/31005/31030/31040 and THSYN_31000/3100531010/31035/31045), similar as described for the PSB *Allochromatium vinosum* DSM 180T [19]. The photosynthetic reaction center H subunit PuhA (THSYN_31405) and PucC (THSYN_31410) are clustered upstream with genes encoding RC-LH1 auxiliary proteins (THSYN_31390–31,400). Furthermore, a homologousHi-PIP (THSYN_25970) is found in strain Cad16T. It may function as the main electron donor to the photosynthetic reaction center similar as in *A. vinosum* [20].

The absorption spectrum of strain Cad16T shows strong absorption peaks at 374 nm, 582 nm and 833 nm which are characteristic for BChl a [10]. The genes for the complete enzymatic pathway from protoporphyrin to chlorophyllide, and further to BChl a (THSYN_31090–31,105, THSYN_31375, THSYN_31385, THSYN_31415–31,445, THSYN_31555, THSYN_32265–32,270), are clustered on pTs485. BChl a formation is thereby catalyzed by an anaerobic type of the Mg-protoporphyrin IX monomethyl ester oxidative cyclase (ChIE) (THSYN_31385) and a light independent protochlorophyllide reductase complex (ChLNb) (THSYN_31420–31,430) in strain Cad16T.

Extended insights from the genome sequence

### Table 2 Project information for “T. syntrophicum” sp. nov. strain Cad16T

| MIGS ID | Property                              | Term                                      |
|---------|---------------------------------------|-------------------------------------------|
| MIGS 31 | Finishing quality                     | Complete                                  |
| MIGS-28 | Libraries used                        | SMRT 10 kb (BluePippin size selection)    |
| MIGS 29 | Sequencing platforms                  | PacBio RS II                              |
| MIGS 31.2 | Fold coverage                        | 200x                                      |
| MIGS 30 | Assemblers                            | HGAP2                                     |
| MIGS 32 | Gene calling method                   | GeneMarkS+, software revision: 4.1       |
| Locus Tag |                                      | THSYN                                     |
| GeneBank ID |                                   | CP020370.1, CP020371.1, CP020372.1         |
| GeneBank Date of Release |                           | 07/12/2017                                |
| GOLD ID   |                                      | Gp0131589                                 |
| BIOPROJECT |                                  | PRJNA354524                               |
| MIGS 13  | Source Material Identifier           | NA                                        |
| Project relevance |                             | Environmental                             |

### Table 3 Summary of genome of “T. syntrophicum” sp. nov. strain Cad16T: one circular chromosome and two circular plasmids

| Label      | Size (Mb) | Topology | INSDC identifier | RefSeq ID |
|------------|-----------|----------|------------------|-----------|
| Chromosome | 6.84      | Circular | CP020370         | NA        |
| pTs417     | 0.42      | Circular | CP020371         | NA        |
| pTs485     | 0.49      | Circular | CP020372         | NA        |

### Table 4 Genome statistics for the “T. syntrophicum” sp. nov. strain Cad16T genome

| Attribute                  | Value | % of Total |
|----------------------------|-------|------------|
| Genome size (bp)           | 7,738,984 | 100.00     |
| DNA coding (bp)            | 6,663,511 | 86.10      |
| DNA G+C (bp)               | 5,124,386 | 66.22      |
| DNA scaffolds              | 3      | 100.00     |
| Total genes                | 6601   | 100.00     |
| Protein coding genes       | 6237   | 94.49      |
| RNA genes                  | 59     | 0.89       |
| tRNA genes                 | 6      | 0.09       |
| trRNA genes                | 49     | 0.74       |
| ncRNA genes                | 4      | 0.06       |
| Pseudo genes               | 305    | 4.62       |
| Genes in internal clusters | NA     | NA         |
| Genes with function prediction | 2737   | 41.46      |
| Genes assigned to COGs     | 3157   | 47.83      |
| Genes with Pfam domains    | 4675   | 70.82      |
| Genes with signal peptides | 436    | 6.61       |
| Genes with transmembrane helices | 1185 | 17.95     |
| CRISPR repeats             | 5      | –          |

NA = not applicable
Strain Cad16\textsuperscript{T} produces okenone as its sole carotenoid [10] and Crt proteins involved in carotenoid biosynthesis are found on pTs485. The complete synthesis of this keto-carotenoid is mediated through two novel types of carotenoid ketolases, the C-4/4$'$ ketolase CruO (THSYN_31065) and the oxygen dependent CruS bifunctional desaturase (THSYN_31070) [21]. The characteristic $\chi$-ring of okenone is introduced through the key enzymes CrtY and CrtU (THSYN_31055 and THSYN_31050) [21, 22].

Remarkably, most of the proteins involved in photosynthesis are encoded on plasmid pTs485, forming a PGC (Fig. 3) [23]. The highly modular character of the $pufLM$ and $pufC$ genes of $\alpha$, $\beta$ and $\gamma$-proteobacteria has been demonstrated previously [24, 25]. To our knowledge, this is the first description of a PGC being localized on a plasmid in a PSB species. Interestingly, the gene cluster is similarly organized as in the $\gamma$-proteobacterium Congregibacter litoralis strain KT71\textsuperscript{T} and as in members from the $\alpha$-proteobacteria families Rhodobacteraceae and Rhodospirillaceae, respectively.

Sulfur metabolism

For the photoautotrophic process of CO\textsubscript{2} assimilation in PSB, electrons derived from the oxidation of reduced sulfur compounds, are transferred to electron carriers NAD(P)$^+$ and ferredoxin through light energy. During photolithoautotrophic growth under anaerobic conditions, strain Cad16\textsuperscript{T} uses electrons from the oxidation of sulfide, thiosulfate and elemental sulfur as reducing equivalents [3]. Strain Cad16\textsuperscript{T} can use thiosulfate as an electron source during phototrophic growth [3]. No homologous genes for the thiosulfate oxidizing multi-enzyme complex SoxAX, could be found in the strain Cad16\textsuperscript{T} genome. However, $soxB$ (THSYN_26690) and clustered genes

### Table 5

| Code | Value | % age | Description |
|------|-------|-------|-------------|
| J    | 210   | 3.37  | Translation, ribosomal structure and biogenesis |
| A    | 1     | 0.02  | RNA processing and modification |
| K    | 144   | 2.31  | Transcription |
| L    | 276   | 4.43  | Replication, recombination and repair |
| B    | 0     | 0.00  | Chromatin structure and dynamics |
| D    | 42    | 0.67  | Cell cycle control, Cell division, chromosome partitioning |
| Y    | 0     | 0.00  | Nuclear structure |
| V    | 186   | 2.98  | Defense mechanisms |
| T    | 297   | 4.76  | Signal transduction mechanisms |
| M    | 274   | 4.39  | Cell wall/membrane biogenesis |
| N    | 10    | 0.16  | Cell motility |
| Z    | 0     | 0.00  | Cytoskeleton |
| W    | 0     | 0.00  | Extracellular Structures |
| U    | 62    | 0.99  | Intracellular trafficking and secretion |
| O    | 226   | 3.62  | Posttranslational modification, protein turnover, chaperones |
| X    | 97    | 1.56  | Energy production and conversion |
| C    | 245   | 3.93  | Energy production and conversion |
| G    | 122   | 1.96  | Carbohydrate transport and metabolism |
| E    | 167   | 2.68  | Amino acid transport and metabolism |
| F    | 49    | 0.79  | Nucleotide transport and metabolism |
| H    | 135   | 2.16  | Coenzyme transport and metabolism |
| I    | 93    | 1.49  | Lipid transport and metabolism |
| P    | 184   | 2.95  | Inorganic ion transport and metabolism |
| Q    | 29    | 0.46  | Secondary metabolites biosynthesis, transport and catabolism |
| R    | 308   | 4.94  | General function prediction only |
| S    | 1522  | 24.40 | Not in COGs |
| No COG | 1543 | 24.74 | |
| Multi COG | 320 | 5.13 | Multiple COG assignments |
encoding SoxYZ (THSYN_09005–09010) that binds thiosulfate were identified in the genome. Remarkably, this gene combination is found in several genome sequenced *Ectothiorhodospiraceae*. In contrast to the PSB *A. vinosum DSM 180<sup>T</sup> [26], no homologous sequence for the tetrahionate-forming thiosulfate dehydrogenase TsdA was found. However, a c4 cytochrome type TsdB homolog (THSYN_17090) was identified. Due to this unusual combination of genes involved in thiosulfate oxidation, further studies are needed to elucidate the thiosulfate oxidation pathways in strain Cad16<sup>T</sup>.

Initial sulfide and thiosulfate oxidation is immediately followed SGB formation in strain Cad16<sup>T</sup> (Fig. 1). In strain Cad16<sup>T</sup> the SGB structure is mediated through envelope SGP homologues to SgpA and SgpB (THSYN_20250 and THSYN_05960) from *Thi flaviovoccus mobilis* and *Thiocystis violascens*, respectively. The sequence of SgpC (THSYN_11025) shows homology to *Marichromatium* species SgpC/CV3. Predicted signal peptides suggest export of all three SGP proteins into the periplasm in Cad16<sup>T</sup>, as proposed for *A. vinosum DSM 180<sup>T</sup> [27].

Moreover, the genome of strain Cad16<sup>T</sup> encodes the membrane-bound sulfide: quinone oxidoreductases SqrD (THSYN_04215) and SqrF (THSYN_09305). These are possibly involved in the oxidation of sulfide in the periplasm.

The mode of sulfur transport across the inner membrane is not known for PSBs [28]. Organic persulfides such as glutathione or glutathione amide persulfide are proposed as possible candidates. In a next step, the rhodanese-like protein Rhd transfers the sulfur from the persulfide-carrier to the TusA protein in the cytoplasm. The further oxidation steps from sulfur to sulfite are typically mediated through the reverse acting dsr genes in PSB [29]. The strain Cad16<sup>T</sup> genes in the dsr cluster (THSYN_22480, THSYN_22490–22,545) are arranged in a highly conserved organization similar to *A. vinosum DSM 180<sup>T</sup>*; only missing dsrS that is non-essential for sulfur oxidation [30]. The DsrEFH complex mediates persulfate transfer from TusA onto DsrC. The persulfurated form of DsrC is then substrate for the cytoplasmic reverse-acting dissipatory sulfite reductase DsRAB that catalyzes the formation of sulfite. Finally, DsrMKJOP complex reduces DsrC [30].

The genome harbors three additional sulfur relay proteins similar to DsrC (THSYN_09485, THSYN_18820 and THSYN_22565) that could function as TusA homologues. In *A. vinosum DSM 180<sup>T</sup>* DsrC is able to bind DNA upstream the dsr cluster [31].

In strain Cad16<sup>T</sup>, soeABC (THSYN_16370–16,380) encode the sulfur-iron molybdoprotein complex that further oxidizes sulfite to sulfonate on the cytoplasmic site of the membrane [32]. Alternatively, strain Cad16<sup>T</sup> oxidizes sulfite via APS by APS-reductase AprBA (THSYN_16395 and THSYN_16400) and ATP sulfurylase Sat (THSYN_16390), as in other PSB [33, 34]. Thereby, the membrane-bound QmoABHdrCB-complex [35] (THSYN_16425–6440) possibly functions as an electron acceptor for the AprAB reductase complex since no aprM homolog was found in the strain Cad16<sup>T</sup> sequence. For the extra-cytoplasmic export of the final oxidation product sulfate, a SulP sulfate permease (THSYN_14085) homolog to *A. vinosum DSM 180<sup>T</sup>* is encoded in the strain Cad16<sup>T</sup> sequence.

Hydrogen uptake and consumption has been shown to be linked to sulfur metabolism in *Thiocapsa roseopersicina* BBS [36, 37]. Thereby, electrons from hydrogen oxidation in the periplasm by the hyn-type hydrogenase HydSL could be transferred via the Isp membrane complex to the disulfide bound to DsrC. In *A. vinosum DSM 180<sup>T</sup>*, transcription of isp1 and isp2 encoding the Isp hydrogenase subunits is upregulated during growth on sulfide [38]. The Isp complex is composed of two subunits, Isp1 and Isp2, that contain similar catalytic domains as DsrM and DsrK, respectively. Similarly, homologous Isp1 and Isp2 proteins (THSYN_28105 and THSYN_28100) may link sulfur to hydrogen metabolisms in strain Cad16<sup>T</sup>. In accordance, an increase in the sulfide concentration was observed while SGB were consumed by strain Cad16<sup>T</sup> during incubation in the dark (unpublished results, F. Danza).

Additionally, other [NiFe]-hydrogenases of the Hox and Hup type (THSYN_22655, THSYN_22660 and THSYN_28115) are found in the sequence that could mediate light-dependent H<sub>2</sub> evolution as proposed for *T. roseopersicina* [39, 40].

The Cad16<sup>T</sup> genome also harbors cys genes (THSYN_05020–05035) that are probably involved in sulfate assimilation under sulfur-limiting conditions. Furthermore, the genome also encompasses genes encoding the CydDC (THSYN_18930 and THSYN_18935) ATP-driven cysteine transport proteins [41].

**Autotrophic growth**

In PSB, CO<sub>2</sub> fixation is essentially achieved through the reductive pentose phosphate also known as the CBB cycle. In accordance, the strain Cad16<sup>T</sup> genome harbors the complete CBB enzymatic pathway. On the chromosome, the dimeric RuBis-CO form II (THSYN_13250) clusters with RuBis-CO activation protein subunits CbbR, CbbQ and CbbO, (THSYN_13245, THSYN_13255 and THSYN_13285). Interestingly, small and large RuBis-CO subunits form I (THSYN_29475 and THSYN_29480) cluster together with carboxysome shell and auxiliary proteins on plasmid pTs417 (THSYN_29485–29,520 and THSYN_29530–29,535). The carboxysome may allow efficient photoassimilation across varying CO<sub>2</sub> concentrations as proposed for *A. vinosum DSM 180<sup>T</sup>* [42]. Previous studies showed...
different expression regulation for RuBis-CO type I and type II genes from Cad16\textsuperscript{T} suggesting that only the type II is involved in the process of CO\textsubscript{2} fixation [8]. Interestingly, the plasmid pTs485 also harbors a RuBis-CO-like protein form III gene (THSYN_31160) upstream the PGC.

The missing sedoheptulose-1,7-bisphosphatase SBP is possibly bypassed by via the fructose-1,6-bisphosphatase (THSYN_25630). The genes gltA citrate synthase (THSYN_12620), fumA fumarate hydratase (THSYN_24360) and succCD succinyl-CoA ligase (THSYN_00880 and THSYN_00885) that are essential for the TCA cycle, and isocitrate lyase (THSYN_16275) and malate synthase (THSYN_00885) are found in the strain Cad16\textsuperscript{T} sequence. Recently a proteomic study about the late cycle, respectively, are identified in the strain Cad16\textsuperscript{T} sequence. Currently a proteomic study about the capacity of Cad16\textsuperscript{T} to fix CO\textsubscript{2} in the dark suggested the presence of a particular archaean DC/HB cycle [42]. However, no further genes coding for this DC/HB cycle were found. Also a complete set of genes coding for polyhydroxyalkanoic acid synthase PhaC (THSYN_06905) are found in the strain Cad16\textsuperscript{T} genome. Strain Cad16\textsuperscript{T} additionally encodes genes necessary for glycogen polymerisation. The glucose 1-phosphate adenyltransferase GlgC (THSYN_00810), the glycogen synthase GlgA (THSYN_11615) and the 1,4-alpha-glucan branching enzyme GlgB (THSYN_00805) allow the synthesis of glycogen.

Interestingly, strain Cad16\textsuperscript{T} also has the potential to produce the storage compound cyanophycin normally found in cyanobacteria [43], since the two subunits of the enzyme cyanophycin synthetase (THSYN_26990 and THSYN_26995) are found. Toggether, these finding provide genetic evidence for the high carbon fixation potential of strain Cad16\textsuperscript{T} in the dark [8, 44].

Anaerobic Fe(II)-oxidation was described for other \textit{Thiodictyon} strains [45, 46] and evidence of cryptic in situ iron cycling has been demonstrated recently [47]. In accordance with these findings, we found \textit{cbb3} type terminal cytochrome C oxidases (THSYN_06760–08775) possibly involved in Fe(II) driven carbon fixation in strain Cad16\textsuperscript{T} genome.

Strain Cad16\textsuperscript{T} grows chemoautotrophically under microaerobic conditions (5\% O\textsubscript{2}) with sulfide, thiosulfate, or sulfide only [3], as also observed in other PSB in vitro studies with \textit{Lamprocytis purpurea} [10, 48], \textit{Thiocystis violacea} and \textit{A. vinosum} [49]. In situ, strain Cad16\textsuperscript{T} is possibly exposed to low concentration of oxygen produced by oxygenic microbota at the mixolimnion-chemocline interface [8]. Accordingly, we observe genes encoding sod-type superoxide dismutases (THSYN_20405 and THSYN_22720), as well as fur and fur-type transcriptional regulators involved in peroxide stress response. In situ, strain Cad16\textsuperscript{T} is possibly exposed to oxygen produced by oxygenic microbota at the mixolimnion-chemocline interface [8].

**Nitrogen metabolism**

Furthermore, with the genes encoding NifB (THSYN_03975), NifD (THSYN_08880), NifH (THSYN_08885), NifK (THSYN_08875), NifT (THSYN_08870) NifW, NifZ and NifM (THSYN_10720, THSYN_10725 and THSYN_10730), NifX (THSYN_21435) and NifL (THSYN_24590) strain Cad16\textsuperscript{T} could possibly fix nitrogen. Genes encoding the multisubunit urease UreDEFG (THSYN_03745, 03750, 03760 and 03765) and the urea transporter UrtABCDE (THSYN_07940–07955, 03760, 07975) indicate the possible utilisation of urea.

**Transmembrane transport proteins**

Several membrane transport genes were found in the strain Cad16\textsuperscript{T} genome, including protein secretion system Type II, genes encoding the TAT pathway and several TRAP transporter genes, as well as genes encoding Ton-Tol type and ABC-type transporter complexes. Additionally, a complete TSS4 pilus machinery is encoded in six clusters dispersed on the strain Cad16\textsuperscript{T} chromosome. Notably, also structural components of TSS6 secretion system are found in two clusters on the chromosome (THSYN_11395–11,410) and on pTs485 (THSYN_32540-THSYN_32580). Two effector proteins of the VrgG family were identified. THSYN_15360 belongs to the vgr GE type Rhs family proteins similar sequences found in \beta-proteobacterial family of the \textit{Burkholderiaceae} whereas THSYN_32425 is conserved in \gamma-proteobacteria and contains a type IV Rhs element. Toggether, the secretion machinery allows strain Cad16\textsuperscript{T} to interact within the highly populated chemocline with up to 10\textsuperscript{7} bacterial cells per milliliter. The secretion and uptake mechanism may also play a key role in the cell-to-cell contact with \textit{Desulfovacta thiozymogenes}.

**Buoyancy regulation and chemotaxis**

Strain Cad16\textsuperscript{T} can possibly regulate buoyancy by gas vesicles that are formed with the encoded structural gas vesicle proteins. Whereas GvpA proteins forms the vesicle core (THSYN_11790, THSYN_11825, THSYN_15290, THSYN_18705 and THSYN_31215), GvpF (THSYN_111800 and THSYN_18685), GvpK (THSYN_11785) and GvpN (THSYN_11815 and THSYN_18695) further stabilize the structure. Proteins homologues to the transcriptional regulatory factors GvrA (THSYN_11850) and GvrC (THSYN_11830) from the enterobacterium \textit{Serratia} sp. ATCC 39006 are also found in Cad16\textsuperscript{T}.

The diurnal and sesonal behavior of vacuolated \textit{Chromatiaceae} has been described for different lakes [50, 51].
In strain Cad16<sup>T</sup> a diguanylate cyclase (THSYN_19835) is found upstream the circadian clock genes <i>kaiCBB</i> (THSYN_19820–19,830). These genes act together [52] and may synchronize optimal flotation within the chemocline.

**CRISPR-Cas systems**

Bacterial CRISPR-Cas systems provide a mechanism against bacteriophage infection and plasmid transformation [53]. A CRISPR locus is composed out of a 300–500 bp leader sequence, spacer sequences (21–72 bp), complementary to foreign DNA, and direct repeats (DRs, 24–40 bp) flanking them [53–55]. Adjacent <i>cas</i> genes encode protein that are co-transcribed with the CRISPR locus and interfere with invading DNA guided by the specific spacers [56, 57].

Five CRISPR repeat regions (CRR1–CRR5) were identified in the genome of strain Cad16<sup>T</sup>, four being located on the chromosome and one on the plasmid pTs485 (Fig. 3). The number of DRs ranges from 19 (CRR4) to 146 (CRR2) as seen in Table 6.

BLASTn analysis of the CRISPR DRs using the CRISPRfinder platform revealed similarities in CRR1, CRR2 and CRR4 to sequences of "<i>Thioalkalivibrio sulfidophilus</i>" (31 hits, 4 mismatches), whereas the DRs in CRR3 are similar to the ones in <i>Halothiobacillus neapolitanus</i> c2 (31 hits, 4 mismatches), whereas the DRs in CRR5 are similar to the ones found in <i>Vibrio alginolyticus</i> NBRC 15630 (1 hit, 5 mismatches).

Furthermore, three CRISPR-Cas loci were identified in the strain Cad16<sup>T</sup> sequence, containing <i>cas3</i> genes that are characteristic for type I CRISPR-Cas systems [58]. A complete CRISPR-Cas loci (THSYN_08045–08070) is located 201 bp upstream of CRR2 and is assigned to subtype I-U, containing the signature protein (THSYN_08055) of the GSU0054 family (TIGR02165 and a <i>cas3</i>, THSYN_08070) with C-terminal HD domain (TIGR01596) [58]. Another CRISPR array (THSYN_19240–19,290) is located 182 bp upstream of CRR3 and is classified as subtype I-C due to the <i>cas8c</i> gene and the lack of a <i>cas6</i> sequence. Additionally, an incomplete CRISPR-Cas locus (CRR5) is identified on plasmid pTs485, encoding for Cas2, Cas1, (THSYN_19240–19,245, THSYN_19265, THSYN_19275, THSYN_19285 and THSYN_19,290).

**Conclusions**

We report on the first complete genome sequence of "<i>Thiodictyon syntrophicum</i>" sp. nov. strain Cad16<sup>T</sup> and the metabolic versatility of this environmentally relevant organism. The observed carbon fixation potential can be explained by the highly developed photosynthesis machinery that is coupled to the sulfur and carbon metabolism. Within the changing conditions in the chemocline, strain Cad16<sup>T</sup> is able to optimally use light, different organic and inorganic carbon compounds, reduced sulfur, nitrogen and oxygen. The two 0.4 Mb plasmids found in Cad16<sup>T</sup> are unique for known PSB species and we report structural similarity to sequences from α- and γ-proteobacterial phototrophs. The availability of the complete genome sequence of strain Cad16<sup>T</sup> will facilitate further studies that elucidate its role as key species of the chemocline and the tight association with the <i>Desulfovarpa</i> sp. and the interaction with different PSB and GSB species present in the anoxic part of Lake Cadagno. Due to the limited molecular data on other Thiodictyon strains and no reference strains available, no (digital) DNA-DNA hybridization experiments could be performed. However, the result from phylogenetic analyses on 16S rRNA sequence level, comparative genomic analyses as well as a morphological and physiological differences (see above) indicate a novel species within the genus Thiodictyon.

The described isolate is therefore proposed as "<i>Thiodictyon syntrophicum</i>" sp. nov. strain Cad16<sup>T</sup>, a novel species within the genus Thiodictyon.

A formal description of the proposed novel species follows below:

**Description of “<i>Thiodictyon syntrophicum</i>" sp. nov. “<i>Thiodictyon syntrophicum</i>" (syn.trophi.cum. Gr. pref. Syn, together with; Gr. adj. <i>Trophikos</i>, nursing, tending or feeding; N.L. neut. Adj. <i>syntrophicum</i>, syntrophic).**

Gram-negative, cells are oval-round shaped and 1.4–2.4 µm in diameter, non-motile, vacuolated and contain BChl <i>a</i> and okenone. Growth as single cells, as well as in aggregates with up to 100 cells in a EPS layer.

| Localization | Name | CRISPR start | CRISPR end | CRISPR length (bp) | DR consensus | DR length | No. of spacers | CRISPR-Cas loci<sup>a</sup> |
|--------------|------|--------------|------------|---------------------|--------------|-----------|----------------|-----------------------------|
| chromosome   | CRR1CR1 | 1,879,131 | 1,881,639 | 2508 | GCTCAATGAGGCGCCGGAGAATTGCCGGCGGAAC | 36 | 34 | type I-U |
| CRR2CR2 | 1,883,646 | 1,894,325 | 10,679 | GCTCAATGAGGCGCCGGAGAATTGCCGGCGGAAC | 36 | 146 |  |
| CRR3CR3 | 4,626,522 | 4,629,249 | 2727 | GCATCGCCCGGCCAATTGGCCGGGCGCGGATTGAAAC | 37 | 37 | type I-C |
| CRR4CR4 | 5,078,034 | 5,085,199 | 7165 | GTTCCGCCGGCAGATTGCCGGCGCAATTGAAAC | 36 | 98 |  |
| pTs485 | CRSCRRS | 391,741 | 393,104 | 1363 | GTAGGCCCTACTCGAGCAGCCAAGGCTATTGAAC | 35 | 19 |  |

<sup>a</sup> CRISPR-Cas classification according to Makarova et al. [58]
Assimilation of elemental sulfur in intracellular sulfur globules. Grow photoautotrophically in Pfennig’s minimal medium with a doubling time of 121 h at 20–23 °C, a pH of 6.8–7.2, at 1 mM sulfide and a photoperiod of 12 h dark/12 h light. Dense cultures show a milky purple-red and milky color. Carbon assimilation via Calvin cycle. Following carbon substrates were utilized at a concentration of 5 mM: acetate, fructose and pyruvate. No growth was observed with 5 mM butyrate, ethanol, formate, fumarate, glucose, glyceral, lactate, malate, propanol, propionate and succinate, respectively. Chemolitotrophic growth was observed with 5% Oxygen and 0.02% hydrogen sulfide and 0.07% thiosulfate, or with 0.07% sulfide only, respectively.

The type strain Cad16T (=JCM 15483T =KCTC5955T) was isolated from a sulfidic chemocline in the alpine Lake Cadagno in Switzerland. The genome size of the type strain is 6.84 Mb (chromosome), contains two plasmids, pTs485 (0.49 Mb) and pTs417 (0.42 Mb) and the G + C content of the genome is 66.22%. The 16S RNA gene sequence of strain Cad16T is deposited under the GenBank/EMBL/DDJB accession number AJ511274. The complete genome sequence of the type strain Cad16T is deposited under the GenBank ID CP020370, CP020371 and CP020372. The type strain has been deposited both at the Japan Collection of Microorganisms (JCM 15483T) and at the Korean Collection for Type Cultures (KCTC 5955T).

Additional file

Additional file 1: Figure S1. Phylogenetic placement of T. syngenticum strain Cad16T within the other 12 Chromatiaceae species with a publicly available whole genome sequences. Additionally, the closely related phylogenetic lineages Nitrosococcus, Rheinhemera and Arsukabacterium are also included. Strain Cad16T is most closely related to L. purpurea DSM 4197. The maximum likelihood tree was inferred from 100 concatenated single-copy orthologues sequences [61] and a total of 1000 bootstrap replicates were performed. Numbers at the nodes indicate the SH-aLRT support (%) and ultrafast bootstrap support (%). OrthoMCL [64], was used to define at set orthologues proteins between these 23 species. Hundred single-copy orthologues were randomly chosen and aligned with MUSCLE [66]. The best-fit phylogenetic model and subsequent consensus tree computation, based on maximum-likelihood and 1000 bootstrap iterations, was performed with the IQ-TREE software [62]. Nodes with both, 100% SH-aLRT and ultrafast bootstrap support, are indicated with filled black circle symbols for convenience. (TIF 57220 kb)

Abbreviations

ABC: ATP-binding cassette; APS: adenosine-phosphosulfate; BCHl: bacteriochlorophyll; CAS: CRISPR associated; CBB: Calvin-Benson-Bassham cycle; CRISPRs: clustered regularly interspaced short palindromic repeats; DC/ HB: dicarboxylate/4-hydroxybutyrate cycle; DR: direct repeat; dsr: dissimilatory sulfite reductase; EPS: extracellular polymeric substances; GSB: green sulfur bacteria; HiPIP: high-potential iron-sulfur protein; HPr: high-potential iron-sulfur protein; PCC: photosynthesis gene cluster; PSB: purple sulfur bacteria; RC: reaction center; RubisCO: ribulose-1,5-bisphosphate carboxylase/oxygenase; SBR: sulfur reducing bacteria; SrP: type IV pilus; T6SS: type VI secretion system; TAT: twin-arginine translocation; TCA: tricarboxylic acid cycle; TRAP: tripartite ATP-independent periplasmic

Acknowledgements

We kindly thank Nicole Liechti, Corinne Oechslin, Pierre H.H. Schneeberger and Christian Bueret. We also thank the assistance of Andrea Patrignani from the Functional Genomics Center Zurich for performing the PacBio RS II sequencing.

Authors' contributions

SML and FD drafted the manuscript, carried out cultivation, DNA extraction and purification and microscopy. SML, JFP and MW participated in the genome assembly, correction and annotation. SML, JFP, FD, NS, MW, NUF and MT discussed and analyzed the data and revised the manuscript. MT, JFP and MW conceived of and supervised the study. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Author details

1. University of Geneva, Sciences III, Department of Botany and Plant Biology, Microbiology Unit, 1211 Geneva, Switzerland. 2. University of Applied Sciences of Southern Switzerland (SUPSI), Department of Environment, Constructions and Design (DACCD), Laboratory of Applied Microbiology (LMA), Via Mirasole 22A, 6500 Bellinzona, Switzerland. 3. Federal Office for Civil Protection, Speiez Laboratory, Biology Division, Speiez, Switzerland. 4. Zurich University of Applied Sciences (ZHAW), Institute of Natural Resource Sciences, Environmental Genomics and System Biology Research Group, Wädenswil, Switzerland. 5. University of Copenhagen, Department of Biology, Helsingør, Denmark.

Received: 24 July 2017 Accepted: 24 April 2018 Published online: 09 May 2018

References

1. Imhoff JF. The Family Chromatiaceae. In: Rosenberg E, Delong EF, Lory S, Stackebrandt E, Thompson F, editors. The Prokaryotes [Internet]. Springer Berlin Heidelberg, 2014 [cited 2017 Apr 24]. p. 151–78. Available from: http://link.springer.com/reviewworkentry/10.1007/3-642-38922-1_295.
2. Winogradsky S. Beiträge zur Morphologie und Physiologie der Bakterien. Heft 1. Zur Morphologie und Physiologie der Schwefelbakterien. Leipzig: Felix; 1888.
3. Peduzzi S, Storelli N, Welsh A, Peduzzi R, Hahn D, Perret X, et al. Candidate Thiocyst cyan syntrophiscum, sp. nov., a new purple sulfur bacteria isolated from the chemocline of Lake Cadagno forming aggregates and specific associations with Desulfofastacps sp. Syst Appl Microbiol. 2012:35:139–44.
4. Imhoff JF. Chromatiales Ord. Nov. In: Brenner DJ, Krieg NR, Staley JT, Garrity GM, Boone DR, De Vos P, et al., editors. Bergey’s manual® Syst. Bacteriol. Vol. two Proteobacteria part B Gammaproteobacteria. Boston: Springer US, 2005. p. 1–59. Available from: http://dx.doi.org/10.1007/0-387-28022-7_1.
5. Del Don C, Hanselmann KW, Peduzzi R, Bachofen R. The meromictic alpine Lake Cadagno: orographical and biogeochemical description. Aquat Sci. 2001;63:70–90.
6. Tonolla M, Storelli N, Danza F, Ravasi D, Peduzzi S, Posth NR, et al. Lake Cadagno: Microbial Life in Crenogenic Meromixis. In: Gulati RD, Zadereev ES, Cadagno: Microbial Life in Crenogenic Meromixis. In: Gulati RD, Zadereev ES, editors. Bergey's Manual® Syst. Bacteriol. Vol. 5. Berlin Heidelberg; 2014 [cited 2017 Apr 24]. p. 151–78.
7. Tonolla M, Peduzzi R, Hahn D. Long-term population dynamics of phototrophic sulfur bacteria in the chemocline of Lake Cadagno, Switzerland. Appl Environ Microbiol. 2005;71:3544–50.
8. Camacho A, Erez J, Chicote A, Fborin M, Squires MM, Lehnmann C, et al. Microbial microstratification, inorganic carbon photoassimilation and dark carbon fixation at the chemocline of the meromictic Lake Cadagno (Switzerland) and its relevance to the food web. Aquat Sci. 2001;63:91–106.
9. Storelli N, Peduzzi S, Saad MM, Frigaard NU, Perret X, Tonolla M. CO2 fixation in the chemocline of Lake Cadagno dominated by a few types of phototrophic purple sulfur bacteria. FEMS Microbiol Ecol. 2013;84:421–32.
10. Peduzzi S, Tonolla M, Hahn D. Isolation and characterization of aggregate-forming sulfate-reducing and purple sulfur bacteria from the chemocline of meromictic Lake Cadagno, Switzerland. FEMS Microbiol Ecol. 2003;45:29–39.
11. Eichler B, Pfennig N. A new purple sulfur bacterial from stratified freshwater lakes, Amoebobacter purpureus sp. nov. Arch Microbiol. 1988;149:395–400.
12. AmphoraNet: The webserver implementation of the AMPHORA2 metagenomic workflow suite [Internet]. [cited 2016 Jun 5]. Available from: http://www.sciencedirect.com/science/article/pii/S0033935X16301567

13. SharedProtocol-Extracting-DNA-using-PhenoEl-Chloroform.pdf [Internet]. [cited 2017 Jan 5]. Available from: http://www.pabcx.cn/wp-content/uploads/2015/06/SharedProtocol-Extracting-DNA-using-PhenoEl-Chloroform.pdf

14. Agencourt AMPure XP Bead Clean-up - NEBioNext Microbiome DNA Enrichment Kit (E2612) | NEB [Internet]. [cited 2017 Jan 5]. Available from: https://www.neb.com/protocols/2013/04/18/agencourt-ampure-xp-bead-clean-up-e2612

15. Grissa I, Vergnaud G, Pourcel C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res. 2007;35:3622–7.

16. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: new developments. Nucleic Acids Res. 2016;44:D279–85.

17. TMHHW Server, v. 2.0 [Internet]. [cited 2017 Apr 6]. Available from: http://www.cbdi.tsu-services/TMHHWM

18. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods. 2011;8:785–6.

19. Conron GE, Nagashima KVP, Matsuura K, Sakuragi Y, Wettasinghe R, Qin H, et al. Genes encoding light-harvesting and reaction center proteins from Chromatium vinosum. Photosynth Res. 1999;59:39–52.

20. Vegméglo A, Li J, Schoepf-Cothenet B, Pratt N, Knaff DB. The role of high-potential iron protein and cytochrome c₅₇ as alternative Electron donors to the reaction center of Chromatium vinosum. Biochemistry (Moscow). 2002;81:9868–75.

21. Vogl K, Bryant DA. Elucidation of the biosynthetic pathway for Okenone in Thiothrix sp. CAD16 leads to the discovery of two novel carotene ketolases. J Biol Chem. 2011;286:38521–32.

22. Vogl K, Bryant DA. Biosynthesis of the biomarker okenone: x-ring formation. Geobiology. 2012;10:205–15.

23. Bauer CE, Bugg JJ, Yang ZM, Mann LS. The superoperonic organization of genes for pigment biosynthesis and reaction center proteins is a conserved feature in Rhodobacter capsulatus: analysis of overlapping bchB and puhA transcripts. Mol Gen Genet. 1991;228:434–43.

24. Nagashima KVP, Vegméglo A, Fusuda N, Nagashima S, Shimada K, Inoue K. Exchange and complementation of genes coding for photosynthetic reaction center Core subunits among purple Bacteria. J Mol Evol. 2014;79:52–62.

25. Igarashi N, Harada J, Nagashima S, Matsuura K, Shimada K, Nagashima KVP. Horizontal transfer of the photosynthesis gene cluster and operon rearrangement in purple Bacteria. J Mol Evol. 2014;79:52–62.

26. Denkmann K, Grein F, Zigan R, Siemen A, Bergmann J, van Helmont S, et al. Thiosulfate dehydrogenase: a widespread unusual acidophilic c-type cytochrome. Environ Microbiol. 2004;70:722–8.

27. Pittman MS, Coker H, Wu G, Binet MB, Moir AJG, Poole RK. Cysteine is exported from the Escherichia coli cytoplasm by CcdC, an ATP-binding cassette-type transporter required for cytochrome assembly. J Biol Chem. 2002;277:49841–9.

28. Weissgerber T, Zignani R, Holdridge SR, Dettet JD, Han C, et al. Complete genome sequence of Chromatium vinosum DSM 180(T). Stand Genomic Sci. 2011;5:311–30.

29. Simon RD. Cyanophycin granules from the blue-green alga Anabaena cylindrica: a reserve material consisting of copolymers of aspartic acid and arginine. Proc Natl Acad Sci U S A. 1971;68:265–7.

30. Storelli N, Saad MA, Frigaard N-U, Perret X, Tonolla M. Proteomic analysis of the purple sulfur bacterium Candidatus “Thiothrix cyanocyanophytum” strain Cad16T isolated from Lake Cadagno. EUFPa Open Proteomics [Internet]. [cited 2013 Nov 28]; Available from: http://www.sciencedirect.com/science/article/pii/S2212968513001792

31. Erenreich A, Widdel F. Anaerobic oxidation of ferrous iron by purple bacteria, a new type of phototrophic metabolism. Appl Environ Microbiol. 1994;60:4517–26.

32. Croal LR, Johnson CM, Beard BL, Newman DK. Iron isotope fractionation by Fe(II)-oxidizing photoautotrophic bacteria. Geochim Cosmochim Acta. 2004;68:1227–42.

33. Berg J. Intensive cryptic microbial iron cycling in the low iron water column of the meromictic Lake Cadagno - Berg - 2016 - Environmental Microbiology - Wiley Online Library [Internet]. [cited 2016 Oct 23]; Available from: http://onlinelibrary.wiley.com/doi/10.1111/1462-2920.13587/full

34. Overmann J, Pfennig N. Continuous chemotrophic growth and respiration of Chromatiales species at low oxygen concentrations. Arch Microbiol. 1992;158:59–67.

35. Kampf C, Pfennig N. Capacity of chromatiales for chemotrophic growth. Specific respiration rates of Thiothrix violacea and Chromatium vinosum. Arch Microbiol. 1980;127:125–35.

36. Overmann J, Pfennig N. Boxyancy regulation and aggregate formation in Anoeobacter pumpeus from Mahoney Lake. FEMS Microbiol Lett. 1992;101:167–79.

37. Egli K, Wiglfl M, Fritz M, Klug J, Giess J, Bachofen R. Spatial and temporal dynamics of a plume of phototrophic microorganisms in a meromictic alpine lake using turbidity as a measure of cell density. Aquat Microb Ecol. 2004;35:105–13.

38. York A. Structural biology: the tick-tock of circadian clocks. Nat Rev Microbiol. 2017;15:256–7.

39. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, et al. CRISPR provides acquired resistance against viruses in prokaryotes. Science. 2007;315:1790–9.

40. Madica FAJ, Diaz-Villaseñor C, Soria E, Juez G. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. Mol Microbiol. 2000;36:244–6.

41. Bolotin A, Quinquis B, Sorokin A, Ehrlich SD. Clustered regularly interspaced short palindromic repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology. 2005;151:2551–61.
56. Al-Attar S, Westra ER, van der Oost J, SJJ B. Clustered regularly interspaced short palindromic repeats (CRISPRs): the hallmark of an ingenious antiviral defense mechanism in prokaryotes. Biol Chem. 2011;392:277–89.
57. Marraffini LA. CRISPR-Cas immunity in prokaryotes. Nature. 2015;526:55–61.
58. Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, et al. An updated evolutionary classification of CRISPR-Cas systems. Nat Rev Microbiol. 2015;13:722–36.
59. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci U S A. 1990;87:4576–9.
60. Skerman VBD, McGowan V, Sneath PHA. Approved Lists of Bacterial Names. Int J Syst Evol Microbiol. 1980;30:225–420.
61. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. Nat Genet. 2000;25:25–9.
62. 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;32:668–74.
63. Grin I, Linke D. GCView: the genomic context viewer for protein homology searches. Nucleic Acids Res. 2011;39 (suppl):W353–W356.
64. Chen F. OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. Nucleic Acids Res. 2006;34(90001):D363–D368.
65. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, Tatusova T, Thomson N, Allen MJ, Angiuoli SV, Ashburner M, Axelrod N, Baldau S, Ballard S, Boore J, Cochrane G, Cole J, Dawyndt P, De Vos P, dePamphilis C, Edwards R, Faruque N, Feldman R, Gilbert J, Gilina P, Göckner FO, Goldstein P, Guralnick R, Haft D, Hancock D, Hernjakob H, Hertz-Fowler C, Hugenholtz P, Joint I, Kagan L, Kane M, Kennedy J, Kowalchuk G, Kottmann R, Kolker E, Kravitz S, Kyrpides N, Leebens-Mack J, Lewis JE, Li K, Lister AL, Lord P, Maltsev N, Markowitz V, Martin J, Methe B, Mizrachi I, Moxon R, Nelson K, Parkhill J, Proctor L, White O, Sansone S-A, Spiers A, Stevens R, Swift P, Taylor C, Tateso Y, Tet T, Turner S, Usery D, Vaughan B, Ward N, Whetzel T, Gil IS, Wilson G, Wipat A. The minimum information about a genome sequence (MIGS) specification. Nat Biotechnol. 2008;26(5):541–547.
66. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32(5):1792–1797.