UNIVERSITÉ DE GENÈVE

Département de botanique et biologie végétale

Unité de microbiologie

FACULTÉ DES SCIENCES

Prof. Dr. Michael Hothorn

Prof. UAS Dr. Mauro Tonolla

Functional Genomics of Purple Sulfur Bacteria from Lake Cadagno

THÈSE

présentée à la Faculté des sciences de l'Université de Genève

pour obtenir le grade de Docteur ès sciences, mention biologie

par

Samuel M. Lüdin

de

Ramlinsburg (BL)

Thèse N° 5252

GENÈVE, Septembre 2018

UNIVERSITÉ DE GENÈVE

FACULTÉ DES SCIENCES

DOCTORAT ÈS SCIENCES, MENTION BIOLOGIE

Thèse de Monsieur Samuel LUDIN

intitulée :

«Functional Genomics of Purple Sulfur Bacteria from Lake Cadagno»

La Faculté des sciences, sur le préavis de Monsieur M. HOTHORN, professeur associé et directeur de thèse (Département de botanique et biologie végétale), Monsieur M. TONOLLA, docteur et codirecteur de thèse (Département de botanique et biologie végétale et Laboratory of applied microbiology University of applied sciences of Southern Switzerland, Bellinzona, Swizerland), Monsieur N. STORELLI, docteur (Laboratory of applied microbiologiy, University of applied sciences of Southern Switzerland, Bellinzona, Swizerland), Monsieur J. F. POTHIER, docteur (Institute of Natural Resource Sciences, Zurich University of Applied Sciences, School of Life Sciences and Facility Management, Wädenswil, Switzerland), Monsieur M. WITTWER, docteur (Biology Division, Spiez Laboratory, Spiez, Switzerland) et Madame P. JUNIER, docteuer (Laboratory of microbiology, Institute of biology, University of Neuchâtel, Switzerland), autorise l'impression de la présente thèse, sans exprimer d'opinion sur les propositions qui y sont énoncées.

Genève, le 3 septembre 2018

Thèse - 5252 -

Le Doyen

N.B. - La thèse doit porter la déclaration précédente et remplir les conditions énumérées dans les "Informations relatives aux thèses de doctorat à l'Université de Genève".

TABLE OF CONTENTS

SUMMARY

Microbial sulfur redox cycling is one of Earth's central biogeochemical processes and sulfur can be both, an essential structural part of proteins and catalytic Co-factors, and a source of electrons. The members of the family *Chromatiaceae* (purple sulfur bacteria; PSB) are prominent phototrophic sulfur oxidizing bacteria (PSOB) involved in sulfur oxidation coupled to anoxygenic photosynthesis. Additional chemotrophic growth in PSB has been observed under aerobic and anaerobic conditions. On today's Earth photic euxinia, characterized by strong opposing concentration gradients in oxygen and sulfide that are accessible to light, is relatively rare and restricted to stagnant coastal waters and intertidal flats, bacterial mats in wetlands and stratified lakes. The meromictic Lake Cadagno (Swiss Alps) is an ideal study site for euxinic microbial communities due to the accessibility and the detailed scientific record spanning from biogeochemistry to biophysics and microbial ecology. The PSOB community of Lake Cadgno has been studied for years *in situ* and strains have been isolated and characterized, however information on genomic level was missing. For the first time, with this study we provide novel ecological insights on the biological functions encoded in two PSOB species preforming whole genome sequencing. We then elucidated in-depth the day/night differences in carbon and energy metabolic pathways of PSB '*Thiodictyon syntrophicum*' nov. sp. str. Cad16^T *in situ* with a combination of quantitative proteomics and radiotracer uptake experiments. Finally we additionally demonstrated the use of genomic data as predictive biomarkers for rapid mass-spectrometry based typing of environmental isolates.

Whole genome sequencing was performed on two Lake Cadagno key species of the PSB group, 'Thiodictyon syntrophicum' nov. sp. str. Cad16^T and *Chromatium okenii* str. LaCa.
The 7.74-Mbp genome of strain Cad16^T was sequenced completely and 6'237 predicted protein sequences and 59 RNA sequences were annotated. Thereby complete pathways of sulfur oxidation, central carbon metabolism, photosynthesis, respiration and transmembrane transport were detected. Interestingly clusters of genes encoding the photosynthetic machinery and pigment biosynthesis are found on the 0.48 Mb plasmid pTs485. That may indicate horizontal gene transfer is common in the PSOB community studied.

Chromatium okenii was first described in 1838 and we produced the first complete genome of a large-celled *Chromatium* sp. from cell enrichments. Our findings on a genomic level confirm previous experiments on the obligate phototrophic metabolism of *C. okenii*. The

genome was also compared to complete sequences of other PSOB isolated from Lake Cadagno using core genome analysis. Genes involved in flagellar movement, chemotaxis and encoding S-layer-proteins were relatively overrepresented in strain LaCa. As *C. okenii* is involved in bioconvection, the local vertical mixing of the chemocline, the relative importance of genes involved in mobility and taxis were highlighted.

As it was previously found that the str. Cad16[T] (1.5 % off all PSOB cells) assimilates up to 26 % of all carbon fixed daily (day and night), the yet-vaguely defined dark CO_2 fixation mechanism was elucidated in more detail with a combination of long-term monitoring of the physicochemical parameters, ^{14}C-bicarbonate uptake-experiments and quantitative proteomics of *in situ* incubation in dialysis bags. During summer 2017 the chemocline was sinking from 12 to 14 m and a co-occurring bloom of cyanobacteria was observed. Sampling was performed at day and night in September. We found that CO_2 assimilation rates were higher during the light period but the relative change in the proteome (663 quantified proteins) was only 1 % of all porteins encoded in str. Cad16[T]. Oxidative respiration was thereby upregulated at light, whereas stress-related mechanisms prevailed during the night. We therefore concluded that the low average light availability (0.4 μmol quanta m^{-2} s^{-1}) was due to high cell concentrations and the oxygenated chemocline induced a combination of parallel phototrophic and chemotrophic growth in str. Cad16[T].

Matrix assisted laser desorption ionization - time of flight mass spectrometry (MALDI-TOF MS) is a qualitative diagnostic tool for microbial typing. The classical MALDI-TOF MS 'fingerprinting' method is limited by the m/z-database of previously characterized isolates. We now used the genomic information obtained from PSOB sequencing to predict putative biomarker proteins—in most cases light harvesting antenna proteins—for a more complete m/z database. For the fast and reliable MALDI-MS typing of novel PSB and purple non-sulfur bacteria (PNSB) isolates we performed explorative sample preparations that have been be further validated by isolating biomarkers and subsequent MS/MS protein or peptide sequencing.

For future research on the Lake Cadagno microbial ecology we propose to further undermine integrative in situ studies that combine the monitoring of environmental variables, such as hydrodynamics and nutrient availability, and functional meta-omics on protein or/and metabolic level, with predictive modelling.

RÉSUMÉ

Ce travail est une contribution à l'étude d l'écologie microbienne du Lac de Cadagno, un lac méromictique de Cadagno (Alpes suisses), aussi riche en sulfures, est un site d'étude idéal pour les communautés microbiennes euxiniques en raison de l'accessibilité et des données scientifiques détaillées allant de la biogéochimie à la biophysique et à l'écologie microbienne. Sur la Terre d'aujourd'hui, l,euxinia photique, les environnements qui présentent des gradients significatifs d'oxygène et de sulfures et qui sont accessibles à la lumière, se trouve rarement. Le cycle redox microbien du soufre est l'un des processus biogéochimiques primordiaux, car le soufre est à la fois une partie structurelle des protéins et de divers cofacteurs et une source d'électrons. Les membres de da famille des *Chromatiaceae* (bactéries pourpres sulfureuses; BPS) sont des bactéries phototrophes sulfo-oxydantes (BPSO) et la croissance chimiotrophique des BPS a été aussi observée dans des conditions aérobies et anaérobies.

Le séquençage du génome entier a été effectué sur deux espèces clés du lac Cadagno du groupe PSB, '*Thiodictyon syntrophicum*' gen. nov., sp. nov., souche Cad16^T et *Chromatium okenii* souche LaCa et mes résultats ont fait l'objet de deux publications .

Le génome de 7,74 Mbp de la souche Cad16^T a été séquencé et 6237 séquences de protéines et 59 séquences d'ARN ont été annotées. Nous avons détecté des voies complètes d'oxydation par le soufre, le métabolisme central du carbone, la photosynthèse et le transport transmembranaire. Il est intéressant de noter que des groupes de gènes codant pour la machinerie photosynthétique et la biosynthèse des pigments se trouvent sur le plasmide pTs485 de 0,48 Mo. Cela pourrait indiquer que le transfert horizontal de gènes est courant dans la communauté PSOB étudiée.

Chromatium okenii a été décrit pour la première fois en 1838 et nous avons séquencé et décodé le premier génome complet d'un *Chromatium* sp. à partir d'un enrichissement cellulaire. Nos résultats au niveau génomique confirment les expériences précédentes sur le métabolisme phototrophique et chimioautotrophique de *C. okenii*. Le génome a été comparé à des séquences complètes d'autres PSOB isolées du lac Cadagno à l'aide d'une analyse du génome de base. Les gènes codant pour la chimiotaxie et les protéines des flagelles étaient en nombre plus grand que prévu. Comme *C. okenii* est impliqué dans la bioconvection, le mélange vertical local de la chimiocline, l'importance relative des gènes impliqués dans la mobilité et le chemotaxis ont été mis en évidence.

Comme il a été précédemment constaté la souche Cad16T assimile jusqu'à 26 % de tout le carbone fixé par jour (jour et nuit) et le mécanisme de fixation du CO_2 à l'obscurité a été élucidé plus en détail avec une combinaison de surveillance à long terme des paramètres physico-chimiques, d'expériences d'incorporation du ^{14}C-bicarbonate et de protéomique quantitative de cultures incubées en sacs de dialyse in situ.

Au cours de l'été 2017, la chimiocline a baisseé de 12 à 14 m et une prolifération simultanée de cyanobactéries a été observée. Nous avons constaté que les taux d'assimilation du CO_2 étaient plus élevés pendant le jour et que le changement relatif du protéome (663 protéines quantifiées) n'était que de 1 % de toutes les proteins codeés par la souche Cad16^T. La respiration oxydative était ainsi surexprimée à la lumière, alors que les mécanismes liés au stress prévalaient pendant la nuit. Une explication possible était que la faible disponibilité de la lumière due à des concentrations cellulaires élevées et la chimiocline oxygénée induisaient une combinaison de croissance phototrophique et chimiotrophique parallèle chez Cad16T.

La spectrométrie de masse MALDI-TOF est un outil de diagnostic qualitatif pour le typage microbien. Malheureusement, la méthode classique dite 'fingerprinting' est limitée par la base de données des isolats préalablement caractérisés. Pour pallier à cet inconvénient, nous avons utilisé l'information génomique obtenue par séquençage PSOB pour prédire les protéines de biomarqueurs putatifs pour une base de données m/z plus complète. Dans la plupart des cas, il s'agissait des protéines antenne d'absorption de la lumière - Pour le typage MALDI-MS rapide et fiable des nouveaux isolats de bactéries PSB et de bactéries pourpres non sulfureuses (PNSB), nous avons effectué des méthodes de préparation qui ont été validées par l'isolement de biomarqueurs et le séquençage ultérieur des protéines par MS/MS. Pour les recherches futures sur l'écologie microbienne du lac Cadagno, nous proposons de poursuivre les études in situ intégratives qui combinent la surveillance des variables environnementales, telles que l'hydrodynamique et la disponibilité des nutriments, et les méta-omiques fonctionnelles sur le niveau protéique et/ou métabolique, avec la modélisation prédictive.

LIST OF ABBREVIATIONS

2D-DIGE	Two dimensional fluorescence difference gel electrophoresis
ACN	Acetonitrile
AFM	Atomic force microscopy
aLRT	Approximate likelihood-ratio test
AMP	adenosine monophosphate
BChl	Bacteriochlorophyll
BLAST	Basic local alignment search tool
bp	Base pairs
CARD-FISH	Catalysed-reporter deposition fluorescence in situ hybridization
CBB	Calvin-Benson-Bassham
CHCA	α-Cyano-4-hydroxycinnamic acid
Chl	Chlorophyll
CID	collision induced dissociation
CTD	Conductivity, temperature and depth sensor
DAPI	4′,6′-diamino-2-phenylindole
DC/HB	cycle Dicarboxylate/4-hydroxybutyrate cycle
DIC	Dissolved inorganic carbon
DSM(Z)	German Collection of Microorganisms and Cell Cultures
EDTA	Ethylenediaminetetraacetic acid
EM	Electron microscopy
EMBL	European Molecular Biology Laboratory
ESI	Electrospray ionization
FA	Formic acid
FACS	Fluorescence activated cell sorting
FCC	Flavocytochrome c
FCM	Flow cytometry
FISH	Fluorescent in situ hybridization
FGCZ	Functional Genomics Center Zurich
FSC	Forward scatter
FTU	Formazine turbidity unit
FWHM	Full
Ga	Billions years ago
GSB	Green sulfur bacteria
HGT	Horizontal gene transfer
HIFU	High intensity focused ultrasound
HP/HB cycle	3-hydroxypropionate/4-hydroxybutyrate cycle
HPLC	High-performance liquid chromatography
INDELs	Insertions and Deletions
IPA	Isopropanol
IRMS	Isotope ratio mass spectrometry

kyr BP	Kilo years before the present
LC-MS^2	Liquid chromatography-tandem mass spectrometry
LFQ	Label free quantification
LH(C)	Light harvesting complex
m.a.s.l	meters above sea level
MALDI-TOF MS	Matrix assisted laser desorption ionization – time-of-flight mass spectrometry
NAD(P)H	Nicotinamide adenine dinucleotide (phosphate-oxidase) reduced
nanoSIMS	Nano-scale secondary-ion mass spectrometry
NPP	Net primary production
OD	Optical density
ODZ	Oxygen deficient zones
ORP	Oxidation reduction potential
OTUs	Operational taxonomic units
pH	Power of hydrogen
PHA	Poly(3-hydroxyalkanoates)
PHB	Polyhydroxybutyrate
PMF	Peptide mass fingerprinting
PNSB	Purple nonsulfur bacteria
PSB	Purple sulfur bacteria
PSOB	Phototrophic sulfur oxidizing bacteria
QS	Quorum sensing
RC	Reaction center
rTCA	Reverse tricarboxylic acid cycle
RuBisCO	Ribulose-1,5-bisphosphate carboxylase oxygenase
SDC	Sodium deoxy cholate
SDS	Sodium dodecyl sulfate
SGB	Sulfur globule
SGP	Sulfur globule proteins
SMRT	Single-molecule real-time
SQR	Sulfide-quinone reductase
SRB	Sulfate-reducing bacteria
SSC	Side scatter
TCA	Tricarboxylic acid cycle
TFA	Trifluoroacetic acid
TDS	Top down sequencing
TOF	Time-of-flight
UV	Ultraviolet
WGS	Whole genome sequencing

LIST OF TABLES AND FIGURES

Tables

Supplementary Tables

Figures

Supplementary Figures

1 INTRODUCTION

1.1 Biogeochemistry of sulfur – then and now

Prokaryotes belong to the most ancient group of organisms and have shaped the Earth's continents, atmosphere and oceans since ~3.5 billion years from now (Ga) [1]. Whereas Eukaryotes have suffered from regular mass extinctions due to asteroid collisions and climate change, a continuous prokaryotic gene pool has sustained. Due to their vast abundance ($4–6 \times 10^{30}$ cells) and high doubling rates (hours to days) bacteria dominate many habitats such as the open ocean, the soil and the oceanic and terrestrial subsurfaces [2]. For aquatic habitats, estimated numbers are 1.2×10^{29} cells ($10^4–10^7$ cells ml^{-1}) for the ocean and a total of 2.3×10^{26} cells in freshwater (10^6 cells ml^{-1}) [3, 4]. With their metabolic flexibility, they are major transformers of macroelements in all ecosystems making inorganic molecules available to all organisms.

Life as we know it is based on macromolecules containing H, C, N, O, S, and P as major elements and their abundance is depending on tectonics [5]. The enzymatically driven redox-couples interplay with geochemical processes and thereby have shaped the redox conditions at planetary scale and sparked evolutionary transitions [5, 6]. The biotic production of organic carbon by autotrophy is linked to non-equilibrium redox reactions or the transformation of electromagnetic radiation (i.e. sunlight, in most cases) [1].

The biotic oxidation of organic carbon is exergonic and provides the energy for heterotrophic organisms. Whereas most of the redox-paired reactions required for redox cycling are each performed by different organism separately, some prokaryotes can also switch their metabolism to reversal (mixotrophy).

1.2 Photosynthesis coupled to sulfur oxidation

Of the solar energy that reaches the Earth's surface (160 W$\cdot$m^{-2}) about 0.13 % are effectively used by phototrophic organism [7]. Phototrophy is the transformation of light energy into a chemical potential across biological membranes ΔP (i.e. proton motive force: PMF) to generate ATP (Eq.1) and reduce redox carriers such as NAD(P)$^+$ and ferredoxin (Eq. 2). Two subsequent steps can be distinguished. The oxidation of a reductant A driven by light (Eq. 2)

produces reduced redox carriers and ΔP that is coupled to the reduction of an oxidized redox carrier R_{ox} and CO_2 –reduction (Eq. 3, dark reaction).

$$ADP + P_i \xrightarrow{\Delta P} ATP + H_2O \qquad (1)$$

$$2H_2A + 2R_{ox} \xrightarrow{light} 2A^0 + 2R_{red} \cdot H + \Delta P \qquad (2)$$

$$CO_2 + ATP + 2R_{red} \cdot H + H_2O \rightarrow \{CH_2O\} + ADP + P_i + 2R_{ox} \qquad (3)$$
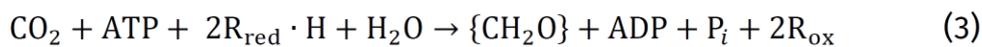
Thereby, most of the electron donors (e.g. H_2S) have not enough reducing power to reduce the redox couples directly. As a consequence, despite that both, chemolithoautotrophs and phototrophs, utilize common electron donors, the former need to spend ΔP on reverse electron flow to overcome this thermodynamic barrier, whereas the later profit from the additional transformed light energy and completely use ΔP for ATP and substrate transport, making them more efficient [8].

Light (Eq. 2) and dark reaction (Eq. 3) sum up to the so-called 'van Niel equation' (Eq. 4).

$$2H_2A + CO_2 \xrightarrow{light} \{CH_2O\} + H_2O + 2A^0 \qquad (4)$$
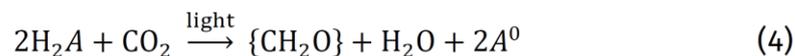
The term $\{CH_2O\}$ is thereby used for intracellular organic matter such as carbohydrates. In phototrophic sulfur oxidizing bacteria (PSOB) hydrogen sulfide (H_2S) is mainly used as an electron source (Eq. 5).

$$2H_2S + CO_2 \xrightarrow{light} \{CH_2O\} + H_2O + 2S^0 \qquad (5)$$
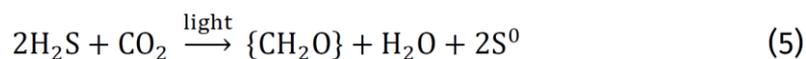
Sulfur cycling is thereby one of Earth's central biogeochemical processes. Sulfur has eight different oxidation states (allotropes from -2 to +6, Table 1.2-1) and can function as electron source or sink in prokaryotes in dissimilatory pathways within a large redox range [9]. Alternatively, assimilated sulfur is important for the

structure and function of proteins as in disulfide bridges and Fe-S-enzyme complexes and vitamins and sulfolipids in both prokaryotes and eukaryotes, as examples.

Table 1.2-1 | Biologically available inorganic sulfur compounds used by prokaryotes and their oxidation state

Compound	Chemical formula	Sulfur oxidation state
Sulfate	SO_4^{2-}	+6
Thiosulfate	$S_2O_3^{2-}$	+5 (sulfone S)/ -1 (sulfane S)
(Poly-)thionates	$^-O_3S-(S)_n-SO_3^-$	+5 (sulfone S)/ ±0 (inner S)
Sulfite	SO_3^{2-}	+4
Elemental Sulfur	S_n (S-rings) S_μ or S_∞ (S-polymers)	0
(Poly-)sulfides	S_n^{2-}	-1 (terminal S) / ±0 (inner S)
Sulfide	HS^- / S^{2-}	-2

1.2.1 Sulfur reduction in bacteria

Sulfur reducing bacteria (SRBs) are largely restricted to anaerobic environments such as freshwater, brackish water, marine, and haloalkaline habitats and sulfur in the scale of 7.2×10^9 t is reduced annually in the ocean sediment [10]. Sulfate and sulfite reduction are widespread in bacteria [11] and the order *Desulfobacterales* of the *δ-proteobacteria* have been studied in detail [12]. In *Desulfobacteraceae*, the dissimilatory sulfite reduction (DSR) pathway genes *dsrAB* and *dsrMKJOP* are conserved and *aprBA* encoding the dissimilatory adenosine-5'-phosphosulfate (APS) reductase and the membrane anchor *qmoABC*, are present as single gene copies each [12]. For anaerobic respiration, *Desulfovibrio* spp. utilize an incomplete TCA cycle and produce acetate and H_2S. The complete anaerobic respiration of short organic compounds such as acetate, butyrate and pyruvate and H_2 as electron acceptor was found in *Desulfocbacter* spp. (Eq. 6).

$$2H^+ + CH_3COO^- + SO_4^{2-} \rightarrow 2CO_2 + 2H_2O + HS^- \qquad (6)$$

Alternatively, SRB such as *Desulfovibrio* spp. and *Desulfobulbus* spp. also reduce $S_2O_3^{2-}$, SO_3^{2-} and elemental sulfur by disproportionation —where sulfur is reduced to products with different oxidation states—, a chemolithoautotrophic process often termed as 'inorganic fermentation' (Eq. 7)[13–15].

$$S_2O_3^{2-} + H_2O \rightarrow SO_4^{2-} + HS^- + H^+ \qquad (7)$$
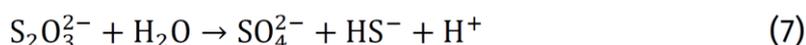
A model of today's sulfur cycle is given in Figure 1.2-1.
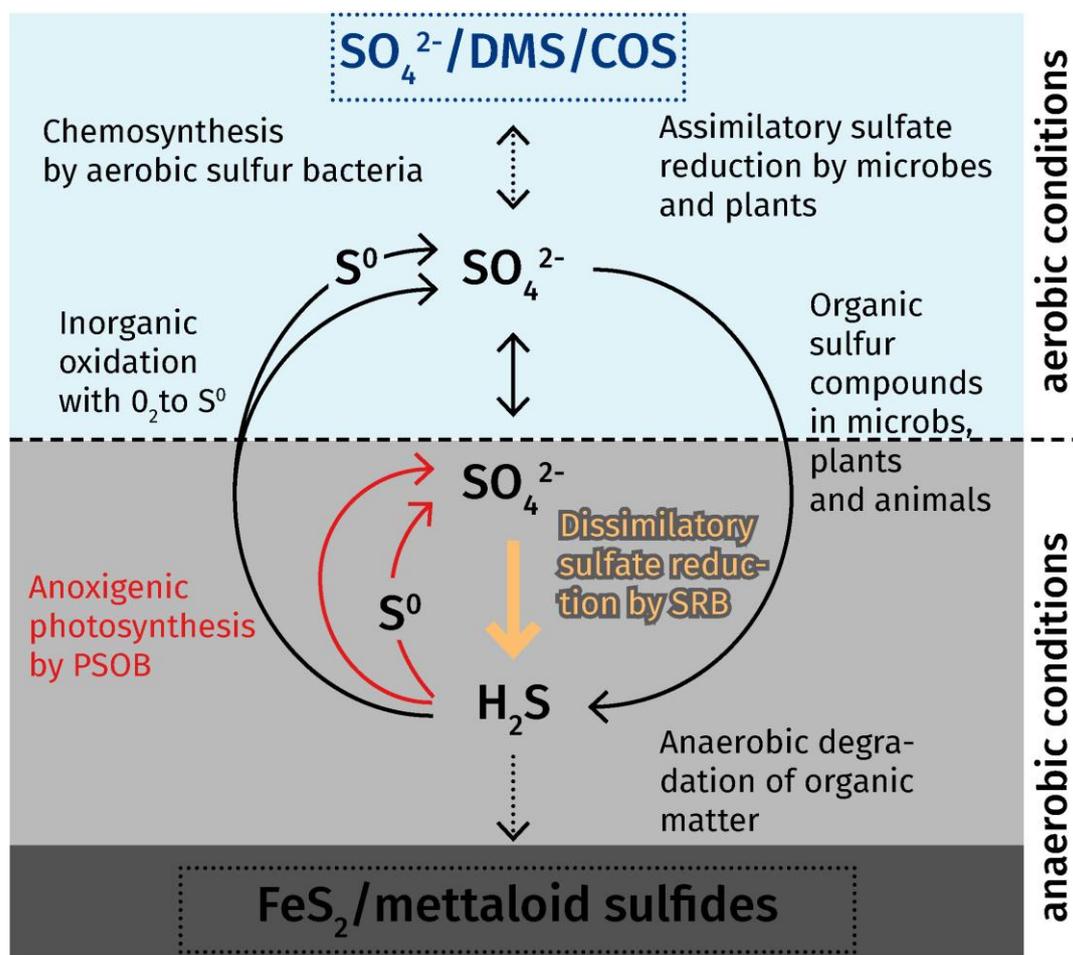
Figure 1.2-1 | Graphic summary of the global sulfur cycle. Pools of oxidized atmospheric and sediment-buried reduced sulfur pools are represented with dotted boxes.
COS: carbonyl sulfide, DMS: dimethyl sulfide, PSOB: phototrophic sulfur oxidizing bacteria, SRB: sulfate reducing bacteria. (Adapted from Pfennig [16] and Middelburg [17]).

Billion years of existence of chemotrophic sulfur reduction and phototrophic sulfur oxidation has thereby shaped the modern Earth, as I will elaborate in more detail in the next paragraph.

1.2.2 The microbial sulfur cycle on ancient Earth and today

On the early anoxic Earth, pyrite (FeS_2) was the initial source of all sulfur and volcanic outgassing and dissolution released large amount into the anaerobic surface completely covered by the ocean [17]. As metabolic activity discriminates isotopes depending on the enzymes involved, stable isotope fractation can be used to trace metabolic activity over space and time [18]. Thereby, depletion of ^{13}C in 3.7 Ga-old Archean sedimentary rocks marks the first indirect evidence of biogenic carbon cycling [19]. The changes in the $^{33}S/^{34}S$

isotope-ratio in sulfide and sulfate records in ocean sediments are used to elucidate the development of sulfur-driven microbial ecosystems. Sulfur disproportionation has been found be to as old as 3.5 billion years [20] and additional evidence for sulfate reduction was found in sulfidic-barites depleted in ^{34}S [21]. Although traces for the evolution of oxygenic phototrophic bacteria date back to 3.8 Ga [22], only the more recent rapid oxygenation of the atmosphere 2.3–2.2 Ga by oxygenic photosynthesis of cyanobacteria [23] let to accelerated weathering of pyrite and limited formation of evaporites [17] (Figure 1.2-2). However, during the Proterozoic the ocean remained anoxic [24] and evidence for the existence of PSOB are found in carotenoid-carbonates from 1.6 Ga [25, 26]. The endosymbiosis of cyanobacteria led to the evolution of phototrophic algae around 2.0 Ga [27], followed by a marked increase in atmospheric O_2 level. The newly available electron acceptors may then have fueled the evolution of heterotrophic prokaryotes [28] and eukaryotes that led to the radiation of multicellular organisms between 0.6 and 0.5 Ga [29, 30]. As a consequence, aquatic animals stirred the sedimentary sulfides, which resulted in an increase in ocean sulfate and the rate of deposition of sulfate evaporate minerals such as gypsum ($CaSO_4 \cdot H_2S$) in the Phanerozoic [31]. Today, sulfate (SO_4^{2-}) is abundant (28 mM) in the ocean and builds the major S reservoir, together with sediment-buried pyrite and in evaporites ($CaSO_4$) [17]
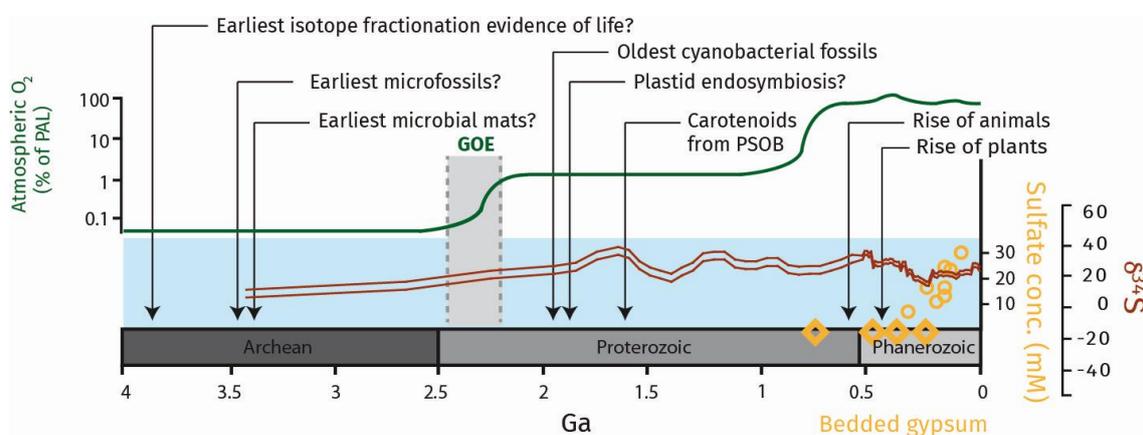
Figure 1.2-2 | A timeline of the last 4 Ga on Earth with major changes in ocean sulfur biogeochemistry, oxygenation of the atmosphere and key events in evolution indicated.
Range of Isotopic composition of sedimentary sulfate over time is indicated by red lines. The Cambrian–Precambrian speciation events at 0.542 Ga led to increased bioturbation in sediments that may have increased the formation of gypsum depositions (yellow diamonds). Abbreviations: Ga: Billion years ago, GOE: Great oxidation event, PAL: present atmospheric levels, PSOB: phototrophic sulfur-oxidizing bacteria. Adapted from Shih 2015 [32] and Canfield&Farquhar 2009 [31].

In the modern ecosystem, sulfur is further present as sulfide in the ocean and soil and as atmospheric sulfur dioxide [17]. However, habitats with illuminated steep redox gradients that allow anoxygenic photosynthesis are comparably rare on today's Earth. Anoxygenic phototrophs have been described in stagnant water bodies with incomplete mixing such as

ponds, ditches, lakes, solar lakes, ice-covered Antarctic lakes, lagoons, fjords and also large marine basins such as the Black Sea and the Baltic Sea [7]. Coastal flats, mangrove forests and bacterial mats in swamps and marshes allow also microbial sulfur-cycling. Taken together, these ecosystems are responsible for <1 % of the global CO_2 assimilation [7].

1.3 Biochemistry of metabolic pathways phototrophic sulfur oxidizing bacteria

In the following section, the metabolic capabilities of photosynthetic bacteria are summarized with the focus on species from sulfidic-anoxic habitats.

1.3.1 Anoxygenic phototrophic sulfur oxidizing bacteria

Phototrophy is widespread among prokaryotes (Figure 1.3-1a) and there are seven major phyla of bacteria that include strains encoding for photochemical reaction centers (RCs) and biosynthesize pigments. Whereas Proteobacteria, Chloroflexi and Gemmatimonadete contain type II RCs, Chlorobi, Firmicutes and Acidobacteria contain type I RCs, and Cyanobacteria that exert oxygenic photosynthesis possess both types of RC, respectively [33]. The evolutionary history of phototrophic prokaryotes is still under debate, as horizontal transfer of gene clusters encoding the photosynthesis machinery (photosynthesis gene cluster: PGC) and loss-of-function events have been described within the proteobacteria between distantly related groups [34, 35], especially within the polyphyletic *Rhodospirillaceae* group of *α-* and *β-proteobacteria* [36]. Six different pathways for inorganic carbon fixation that are not evolutionary linked to the RC type, have been described in phototrophic prokaryotes so far [37].

Anoxygenic phototrophic sulfur oxidizing bacteria (PSOB) belong to different phyla including the order *Chromatiales* and *Chlorobi* group. Importantly, sulfur oxidation by PSOB is considered as 'secondary primary production' [38] since PSOB are largely depending on the sulfide derived from the sulfate-reduzing SRB (e.g. *δ-proteobacteria*) that is coupled to oxidation of organic carbon. Sulfide and/or thiosulfate (SO_3^{2-}) oxidation was described in *Chloroflexi*, heliobacteria, for purple non-sulfur bacteria and some *cyanobacteria* as summarized in Brune 1989 and Dahl 2017 [36, 39]. In the further introduction, I will focus on

the order *Chromatiales* and phylum *Chlorobi* since they are central for the understanding of the ecosystem studied.

1.3.2   Photosynthesis in purple sulfur bacteria

The families of *Chromatiaceae* and *Ectothiorhodospiraceae* (purple sulfur bacteria: PSB) belonging to the γ-proteobacteria, stain Gram-negative and are prominent anoxygenic photosynthetic organisms. The *Chromatiaceae* found in freshwater can be broadly divided in two subgroups based on their morphology and physiology. The genus *Chromatium* contains large-celled mobile forms that are unable to oxidize thiosulfate and are strict phototrophs. The second group comprises small-celled, mostly immobile and often aggregate forming morphotypes of the genera *Lamprocystis, Thiocapsa* and *Thiodictyon* that show chemoheteroorganotrophic growth. Due to their intense purple-to-red coloration and a size up to some dozen μm these were among the first bacteria enriched directly from the environment 180 years ago [40].

In PSB light energy is efficiently captured by bacteriochlorophyll (BChl) *a* or *b* and accessory carotenoid pigments of the okenone or rhodopinal family [7]. Carotenoids have additional structural and photoprotective functions by either BChl-radical quenching and scavenging oxygen radicals [41, 42]. Upon excitation by light the pigments change into an excited electronic state. As excited delocalized π-electrons return to the ground state, that is accompanied by a small decrease in free energy, which is channeled as radiation-free inductive dipole resonance along the highly coordinated pigment arrays [7]. The pigments are bound noncovalently to membrane spanning heterodimeric antenna α/β-polypeptides in circular arrays that form light harvesting complexes (LH) [43]. There are two types of LH quaternary complexes described, called LH1 with peptides PufAB and LH2 with peptides PucAB. The LH1 peptides surround the reaction center (RC) to form the antenna core complex, with the LH2 forming an outer ring around the inner structure [44] in a variable stoichiometry depending on the organism and light availability [45]. In *Allochromatium vinosum*, the LH2 consists of 13 α/β-dimers and the LH1 out of 16 α/β-mers [46]. The pigment-LH cascade funnels electrons to the central type II quinone-type RC (RCII) where charges are separated (Figure 1.3-1 b). The RCII contains up to four subunits, Puf(C)LMH and transfers electrons from a periplasmic high potential electron carrier protein to the quinone pool via a closely associated tetraheme cytochrome [47], thereby entering a cyclic electron flow. Quinone is then reduced to quinol by two protons at the cytoplasmic side of the

inner membrane. The reduced quinol subsequently transfers electrons to a cytochrome bc complex (NADH-dehydrogenase). The membrane-bound cytochrome bc complex subsequently reduces the redox carrier proteins in the periplasm, and the redox cycle is re-initiated [48]. In *A. vinosum* the cytochrome bc complex preferentially reduces the high potential iron-sulfur protein (HiPIP) under photoautotrophic growth whereas the cytochrome c_8 (c_{551}) is used under photoheterotrophic growth conditions [49]. Co-occurring redox reactions in the cytochrome bc releases two H^+ into the periplasm and thereby establish the PMF. The PMF enables ATP generation through ATPases and the reduction of $NAD(P)^+$ by oxidation of quinol catalyzed by the NADH:quinone oxidoreductase.

1.3.3 Sulfur metabolism in purple sulfur bacteria

In PSB, the electrons needed to reduce $NAD(P)^+$ and ferredoxin during photosynthesis derive from the dissimilatory oxidation of the reduced inorganic sulfur compounds HS^-/S^{2-}, S^0 and $S_2O_3^{2-}$. Alternatively, H_2 and ferrous iron (Fe(II)) are also used as electron donors by some PSB [50]. The oxidation of reduced sulfur has been studied in the great detail for *A. vinosum* str. DSM 180^T (former *Chromatium* strain D). Whereas the uptake mechanism of the highly hydrophobic S^0 is unknown [48], H_2S is readily diffusing through the outer membrane and is oxidized to more hydrophilic polysulfide through the membrane-bound sulfide/quinone oxidoreductases (SQR) [51, 52] and/or a soluble dimeric flavocytochrome c sulfide dehydrogenase (FccAB) in the periplasm of PSB [53]. Thereafter, electrons gained are possibly directly fed to soluble redox carriers such as quinone or c-type cytochromes [48]. Noteworthy, the H^+ generated by sulfur oxidation in the periplasm directly contribute to the PMF [54].

The multi-enzyme Sox pathway is used to oxidize cysteine-bound thiosulfate ($S_2O_3^{2-}$) to sulfate [55]. Sulfur intermediates are constantly bound to SoxYZ via a C-ter cysteine [56] and SoxXA(K) mediates the formation of an S–S-bridge between the S^{1-} of thiosulfate and the cysteine of SoxY [57]. SO_4^{2-} is then released by hydrolysis involving SoxB and the remaining sulfane sulfur bound to SoxY is possibly transferred to SGB sulfur chains by SoxL [58].

Alternatively, $S_2O_3^{2-}$ is oxidized to tetrathionate ($S_4O_6^{2-}$) with the cytochrome c thiosulfate dehydrogenase TsdA, as shown for *A. vinosum* [59, 60]. The proteins involved in electron transfer from thiosulfate oxidation to the membrane redox chain are not known, but the c_4 cytochrome TsdB and HiPIP are possible electron acceptors in question [39, 61].
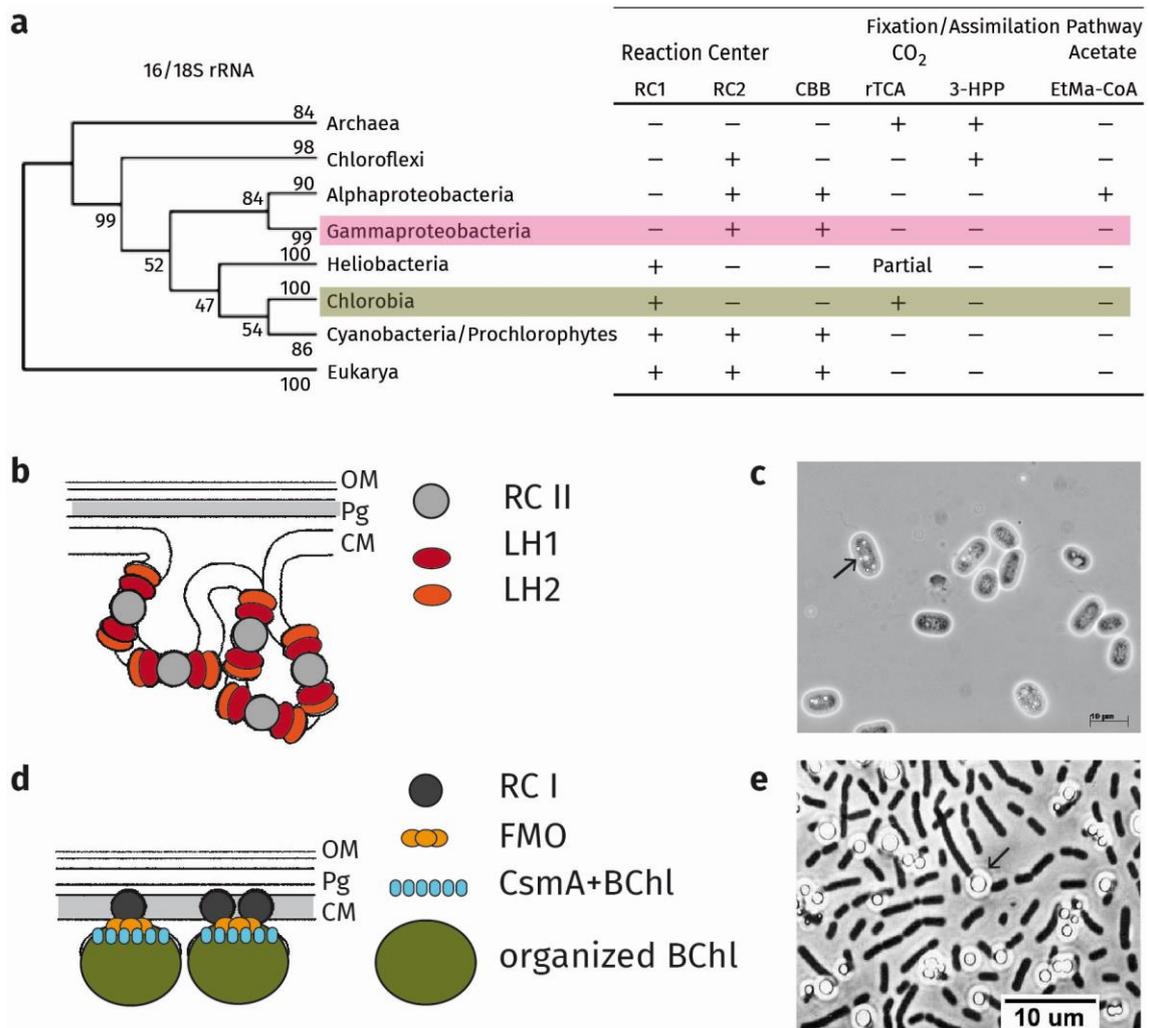
Figure 1.3-1 | Overview on the phylogeny of phototrophs with focus on PSB and GSB spp., the CO₂ fixation pathways used, the structure of their reaction centers and sulfur storage mechanism.

a) A total of 73 16S or 18S rRNA gene sequences from phototrophs were used to construct a Minimum Evolution phylogenetic tree shown as a cladogram. Archaea are used as a non-phototrophic outgroup. Numbers at nodes indicate bootstrap support in % from 1'000 replicates (left). The resulting tree was compared with the type of photosynthetic reaction center, type of CO₂ fixation pathway utilized by each lineage, and the presence or absence of the ethyl malonyl-CoA pathway for acetate assimilation (EtMa-CoA) that contains a novel heterotrophic CO₂ fixation reaction. (Modified from Hanson et al 2012 [8]). b) Graphical representation of the vesicular membrane that harbors the phototrophic apparatus in PSB sp. Light is taken up by BChl and pigments and electrons are shuttled through the peripheral LHs to the reaction center. c) Micrograph of *Chromatium okenii*, a typical representative of the large celled PSB. Arrow indicates sulfur globules within the cell. d) Schematic representation of the photosynthetic system of Chlorobium spp. Self-organized BChl aggregates collect photons and the energy is sequentially transferred through two BChl binding proteins: CsmA forming the baseplate and the trimeric Fenna–Matthew–Olson (FMO) protein, ending in the special pair of the photosynthetic reaction centers anchored within the cytoplasmic membrane. (b) and c) Adapted from Garcia-Pichel and Overmann, 2006. [7]) e) Micrograph of GSB *Chlorobium tepidum*, arrow indicates sulfur globules attached to the cell surface. (Image courtesy of N-U. Frigaard). Abbreviations: Bchl: Bacteriochlorophyll, CBB: Calvin-Benson-Bassham-cycle, rTCA: Reverse tricarboxylic acid cycle, 3-HPP: 3-hydroxypropionate/malyl-CoA cycle, CM: Cytoplasm membrane, CsmA:, EtMa-CoA: ethyl malonyl-CoA pathway OM: Outer membrane, Pg: peptidoglycan, RC: Reaction center, FMO: Fenna–Matthews–Olson a protein, LHC: Light harvesting complex.

Sulfur globules (SGB) with a diameter of 100–200 nm (Figure 1.3-1 c) are obligatory formed during sulfide oxidation. They encompass a 2–5 nm thick protein shell and the therein encapsulated zero-valent sulfur is stored as hydrophilic polysulfides terminated by organic residues ($R-S_n^{2-}-R$) [62] in the periplasm in *Chromatiaceae*. Up to three variants of the structural SGB proteins (SgpABCD) containing coiled-coil motives are present in *A. vinosum*, whereas only SgpA was found to be essential for the formation of SGB [54]. However the mechanism of polymerization and de-polymerization of the SGBs is unknown. The later process likely involves the formation of organic polysulfide intermediates such as glutathione amide persulfide that are further transported into the cytoplasm [48].

Thereupon, sulfane sulfur in the form of protein-bound persulfides via cysteine (Cys–SS⁻), is further transferred along a cascade of subsequent enzymatic multi-subunit complexes while bound to a rhodanese (Rhd) [63], or putatively to DsrE2A [64]. In *A. vinosum*, the first step is a transfer of sulfane sulfur from Rhd to TusA containing an N-term Cys–Pro–X–Pro sequence motif [64, 65]. TusA subsequently transfers sulfur to the hexameric DsrEFH protein a reversible mode [64, 66]. In a third step, sulfane sulfur is shuttled to DsrC [63] containing two conserved C-ter cysteines, Cys_A and Cys_B [67]. Finally, dissimilatory sulfite reductase (DsrAB) interacts with DsrC and catalyzes the formation of the bound persulfide to sulfite [68]. Two possible mechanisms of persulfide oxidation have been discussed. The substrate $DsrC-Cys_A-S^-$ could either be oxidized by DsrAB $DsrC-Cys_A-SO_3^-$ followed by sulfite release by S–S-bond formation in DsrC [67]. Alternatively the membrane bound DsrMKJOP complex, containing the heterodisulfide reductase-like subunit DsrK, could catalyze the formation of an internal DsrC tripeptide ($Cys_A-SSS-Cys_B$) that is subsequently reduced by DsrAB [69]. Electrons thereby released could be transferred to the iron-sulfur flavoprotein DsrL [68] that possibly directly reduces NAD^+ [39].

As a last step, SO_3^{2-} is further oxidized to SO_4^{2-} in two different cytoplasmic pathways in PSB and sulfate is then excreted by an unknown mechanism. The polysulfide reductase-like iron–sulfur molybdoprotein SoeABC in conjunction with SoxYZ as substrate directly oxidizes SO_3^{2-} at the inner side of the cytoplasm membrane [70]. Alternatively, two-step oxidation of SO_3^{2-} to SO_4^{2-} in the cytoplasm is catalyzed via adenosine-5′-phosphosulfate (APS) by the membrane bound trimeric APS reductase AprABM and the Sat ATP sulfurylase [48]. In both cases, electrons are transferred to the redox chain via menaquinone. Sulfur plays additionally an important role in non-photoautotrophic growth of PSB. SGBs are de-

polymerized under chemotrophic growth in the dark and the sulfide is excreted by PSB [71] whereas under photoorganoheterotrophic conditions with acetate present, PSB assimilate sulfate and thiosulfate via the Cys multi-enzyme pathway [48].

A graphical overview on sulfur oxidation in PSB is given in Figure 1.3-2.
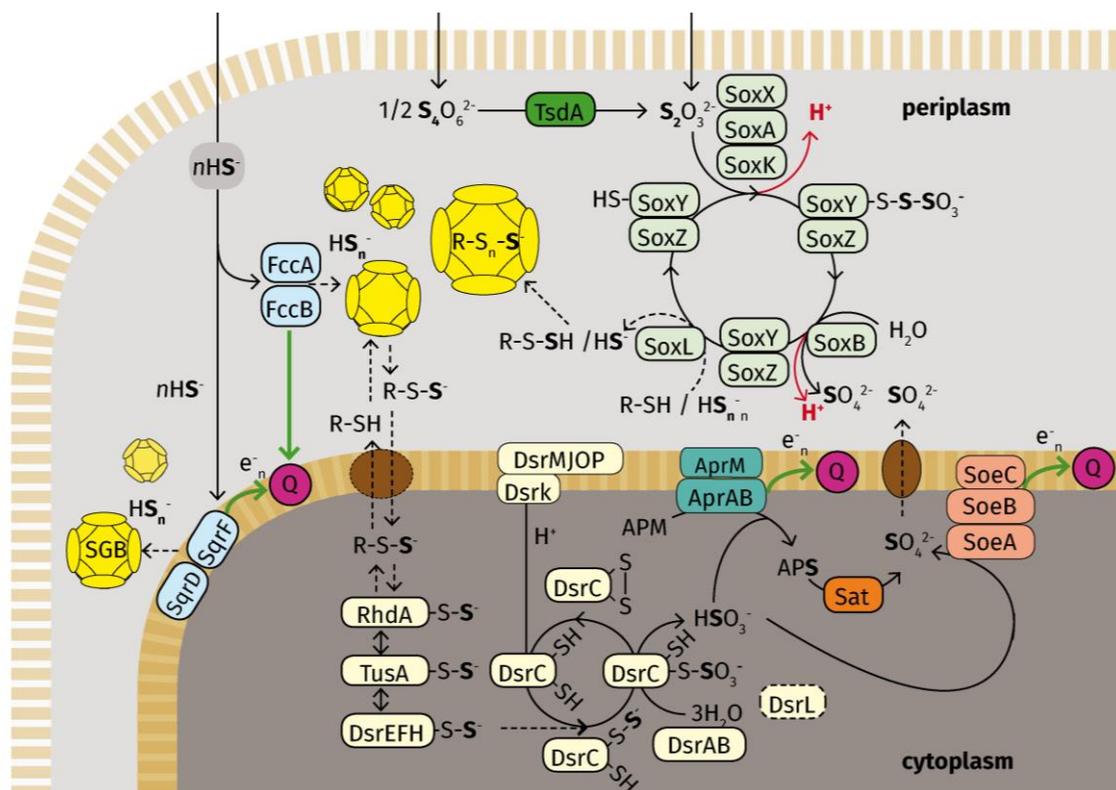
Figure 1.3-2 | Scheme of the complete dissimilatory sulfur oxidation pathway as described for PSB *Allochromatium vinosum*.
Reactions and enzymes with predicted involvement are depicted with dashed lines/outlines. Sulfide is diffusing freely into the periplasm where it is oxidized to elemental sulfur by the action of FccAB or SqrDF (light blue). Electrons are thereby transferred to the quinol-pool (pink) in the inner membrane. Sulfur globules are formed that contain polysulfides encapsulated by a protein cover (yellow). Polymerization and de-polymerization as well as import of polysulfone sulfur (R: organic residue) are not yet understood (brown). Sulfide oxidation to sulfite (light yellow) in the cytoplasm is mediated through a series of transfer reactions along RhdA, TusA and DsrEFH. The DsrAB finally oxidizes sulfane sulfur to sulfite in a cyclic process involving the carrier DsrC. Sulfite is either oxidized directly to sulfate via the membrane-bound iron-sulfur molybdoprotein SoeABC (light orange). Alternatively, sulfite is oxidized to APS by oxidative APS reductase AprABM (turquois) and further oxidized to sulfate by ATP sulfurylase Sat (orange). The mechanism of sulfate excretion is not known. Thiosulfate oxidation in the periplasm is mediated via the Sox –multi-enzyme complex (light green) or through the TsdA enzyme (green). Electron transfers are depicted in green. Produced protons are highlighted in red, as they are produced in the periplasm they contribute to the potential across the inner membrane (proton motive force). Stoichiometry was omitted due to graphical clarity. Abbreviations; APM: adenosine monophosphate, APS: adenosine-5'-phosphosulfate, Q: quinol pool, SGB: sulfur globules, (adapted from Weissgerber et al 2014 [72]).

1.3.4   Carbon metabolism in purple sulfur bacteria

The main CO_2 assimilation pathway in PSB is the reductive pentose or Calvin-Bassham-Benson cycle (CBB) [73, 74]. Ribulose 1,5-bisphosphate carboxylase/oxygenase (RuBisCo) catalyzes either the carboxylation or oxygenation of ribulose-1,5-bisphosphate (RuBP) with CO_2 (or O_2) and substrate competition between photorespiration and photosynthesis arises. In accordance to these finding, components of the CBB are also found in chemolithoautotrophs such as *Methylohalomonas* spp. of the *Ectothiorhodospiraceae* family [75] and also heterotrophs. The second key enzyme is phosphoribulokinase (PBK) that replenishes the RuBP pool by condensation of ribulose-5-phosphate and ATP. Products of CO_2 reduction are 3-phosphoglycerate (3PG) and sugar phosphates. The production of onebiomass unit (i.e. 3PG) thereby consumes 6 oxidized redox carriers and 9 ATPs (Eq. 8).

$$3CO_2 + 6\,(NADH + H^+) + 9ATP \;\rightarrow\; 3\{CH_2O\} + 3H_2O + 6NAD^+ + 9\,(ADP + P_i) \quad (8)$$
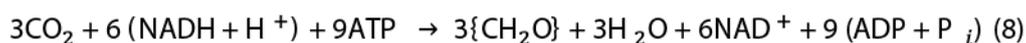
While CO_2 assimilation is the central function of the CBB cycle in photoautotrophic and mixotrophic prokaryotes, purple non-sulfur bacteria such as *Rhodobacter* sp. still express CBB genes and CO_2 is fixed under heterotrophic growth with light. As reducing equivalents are thereby produced in excess, the CBB thereby maintaining the redox balance [76]

Complementary, oxaloacetate and malate are replenished through anaplerotic CO_2-fixation pathways and are used to synthesize biomass in the tricarboxylic acid cycle (TCA) in PSB [37]. In the PSB *A. vinosum* citrate synthase, malate synthase G and isocitrate lyase of the glyoxylate cycle producing 4C-units from acetyl-CoA condensation are expressed under photoautotrophic conditions [77].

Quantum yield is the ratio of carbon fixed to light energy (i.e. quantum) absorbed. For PSB the theoretical ratio of 8.5 mol CO_2 (mol quanta)$^{-1}$ (0.12) was found to be close to measured value of 12 ±1.5 mol CO_2 (mol quanta)$^{-1}$ (0.083) [36]. This slightly lower value is due to the production of ATP from APS during oxidation via AprAB and Sat [36].

The growth of PSB on organic carbon sources was proposed by van Niel [78] and has been successfully demonstrated since for several PSBs. Whereas *A. vinosum* photosynthetically grows with malate as electron and carbon source under anaerobic conditions —i.e.

photoorganoheterotrophic— [77], aerobic mixotrophic growth with CO_2 and acetate in the dark was described for small-celled *Allochromatium*, *Thiocapsa* and *Thiocystis* spp. [79] and different PSB were successfully cultivated under aerobic chemotrophic conditions with reduced sulfur compounds and acetate in the dark over a time of 20 days [80].

1.3.5 Key metabolic pathways in green sulfur bacteria

The *Chlorobiaceae* (green sulfur bacteria; GSB) consists of 15 described species and is a sister taxon of the *Bacteroides*/*Flavobacterium* phylum [81]. The characteristic ovoid antenna complexes called chlorosomes (Figure 1.3-1 d) are 70–180 nm long and 30–60 nm wide and consists of ~10'000 BChl *c*, *d* or *e* molecules organized in crystalline rods and isoprenoid quinones. A combination of monogalactosyl diglyceride molecules and variants of the Csm peptides (e.g. CsmA) builds up chlorosome envelope [82, 83]. A variety of BChl homologues and carotenoids of the isorenieratene or chlorobactene group [84] ensure an efficient excitation energy transfer even at low incident light intensities as low as 0.0022–0.00075 μmol quanta m^{-2} s^{-1} [85]. The average GSB quantum yield is twice as high (0.25) as compared to PSB or other anoxygenic phototrophs due to the small redox potential within the photosynthesis apparatus [81]. Subsequently, the Fenna–Matthews–Olson a protein (FMO) funnels energy from the antenna structure onto the FeS type-I RC. There soluble ferredoxin is reduced that feeds electrons directly, or by reducing NAD^+, into the reverse tricarboxylic acid cycle (rTCA) for CO_2 fixation, omitting cyclic electron flow [81]. As in the rTCA reduced ferredoxin ($E^{o'}$~ -500 mV) is utilized as electron donor by two enzymes, α-oxoglutarate synthase and pyruvate synthase, strongly reducing conditions are required for GSB [7]. The rTCA requires 5 ATP to synthesize one 3PG and does not allow reverse reactions [36]. However, GSB can growth mixotrophically with electron donors, CO_2 and small organic acids such as acetate and pyruvate [81]. In GSB glucose accumulation is induced during N and P-limited photosynthetic growth and allows for the subsequent fermentative oxidation in the dark [86] that was confirmed by *in situ* incubation ^{13}C-uptake experiments with *C. clathratiforme* [87]. Reduced sulfur (H_2S, S^0 and eventually $S_2O_3^{2-}$) and H_2 are used as electron sources and produced S^0 is stored in SGBs attached to the cell (Figure 1.3-1e). An overview is given by Frigaard [88]. Initial sulfide oxidation is mediated through the FccAB and/or Sqr enzymes in the periplasm and the multi-enzyme Sox complex oxidizes $S_2O_3^{2-}$. The elemental sulfur thereby produced it exported out to the attached SGBs, and re-imported into the cell in the form of organic-polysulfanes by a yet unknown process. Intracellular sulfide is

fed into the Dsr pathway and adenylylsulfate (also called APS) is subsequently produced by adenylylsulfate reductase AprAB. The Final oxidation to sulfate step is mediated by ADP sulfurylase or ATP sulfurylase, respectively [89]. Electrons thereby obtained are fed to menaquinone or soluble cytochrome c_{555} and a membrane-bound monoheme cytochrome c_{551} at the periplasmic side of the inner membrane to re-oxidize ferredoxin [88]. A scheme of the sulfur oxidizing pathways in GSB is given in Figure 1.3-3.
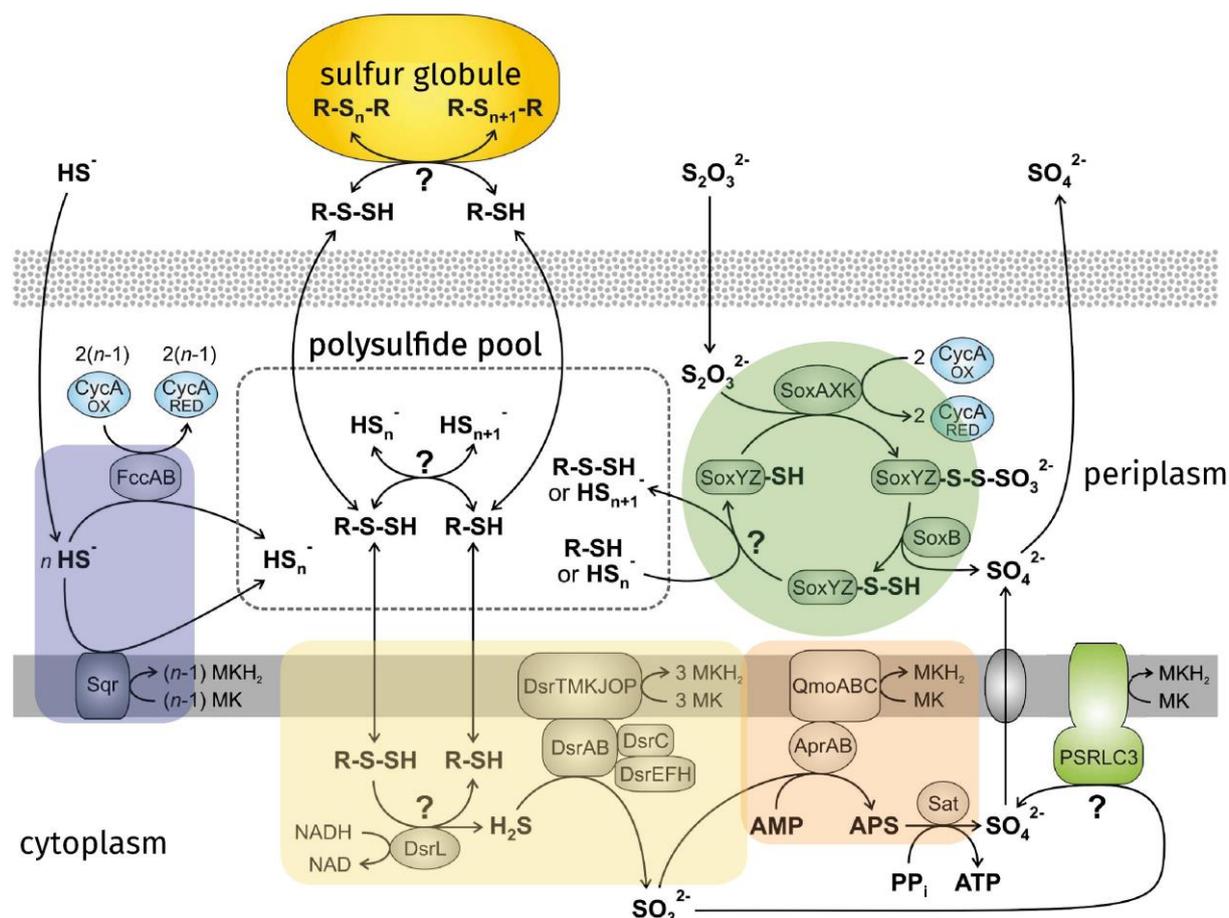
Figure 1.3-3 | Graphical summary of the predicted pathways of sulfur oxidation in GSB.
Initial sulfide oxidation is mediated through the flavoprotein FCC or sulfide:quinone oxidoreductase SQR in the periplasm (blue box). As a difference to PSB, intra-membranous menaquinol functions as a redox shuttle in GSB. The SOX multi-enzyme pathway is oxidizing thiosulfate to sulfate and elemental sulfur in the periplasm (green box). The extracellular sulfur globules are connected to a putative pool of polysulfides (HSn– and possibly organic R-Sn-H). The mechanism of polysulfide transport is thereby largely unknown. Complete oxidation of the polysulfanes to sulfate is channeled through the DSR system (yellow box). Sulfite is oxidized by the APR and SAT (orange box), or the Polysulfide reductase-like complex 3 (PSRLC3), APM: adenosine monophosphate, APS: adenosine-5′-phospho-sulfate, MK: menaquinone. (Adapted from Gregersen et al, 2011 [90])

1.3.6 PSB and GSB niche separation

Due to low their light adaptation, the utilization of reduced sulfur and the danger of oxygen-singlet toxicity PSB and GSB preferably flourish in stagnant deep waters and densely layered microbial mats. Thereby the water molecules, Chl-containing oxygenic phototrophs and the concentration of humic substances above the chemocline define both intensity and spectrum at large [91, 92] (Figure 1.3-4). Furthermore, physical factors such as internal seiches and currents then define the lateral extent and thickness of the chemocline. There are examples of ecosystems where PSB and GSB each flourish exclusively. In Lake Mahoney (British Columbia, Ca) a dominant population of PSB *Lamprocystis purpurea* thrives under euxinia with a sulfide concentration of 30–35 mM [93] and *Thiocapsa* sp. grow in tidal flats at only mm-thick oxic-sulfidic gradients in Scapa Flow (Orkney Islands) [94]. Whereas only GSB are found in a sub-glacial lake in Antarctica [95], in the Black sea at a depth of 70 m [96] and close to hydrothermal vent absorbing geothermal radiation at 2'391m depth [97]. However, in most habitats GSB and PSB population overlap, with GSB preferably growing below PSB plates due to their absorption characteristics (Figure 1.3-4) and the relative higher efficiency in light and carbon utilization [98–100]. Competition within GSB was found to be based on variable carotenoid content of different GSB sp. [101], whereas in PSB sulfide affinity and oxidation rates, as well as the possibility of aerobic chemotrophic growth and mobility pattern have been discussed as possible competitive strategies [102, 103]. Co-cultivation of PSB and GSB revealed syntrophy based on exchange of substrates of incomplete sulfur oxidation processes [104] whereas GSB cells form numerically defined aggregates with a central mobile chemotrophic bacteria that show phototaxis and chemotaxis towards carbon and reduced sulfur sources [105]. Aggregate forming PSB 'Thiodictyon syntrophicum' sp. nov. strain Cad16^T has been found to include SRB that help to maintain sulfur cycling in sulfide depleted environment [106]. Interestingly, also several symbiotic chemotrophic isolates closely related to PSB have been obtained from bivalves (shipworms) and gastropods (marine snails) and Annelida (tube worms) [107–109].
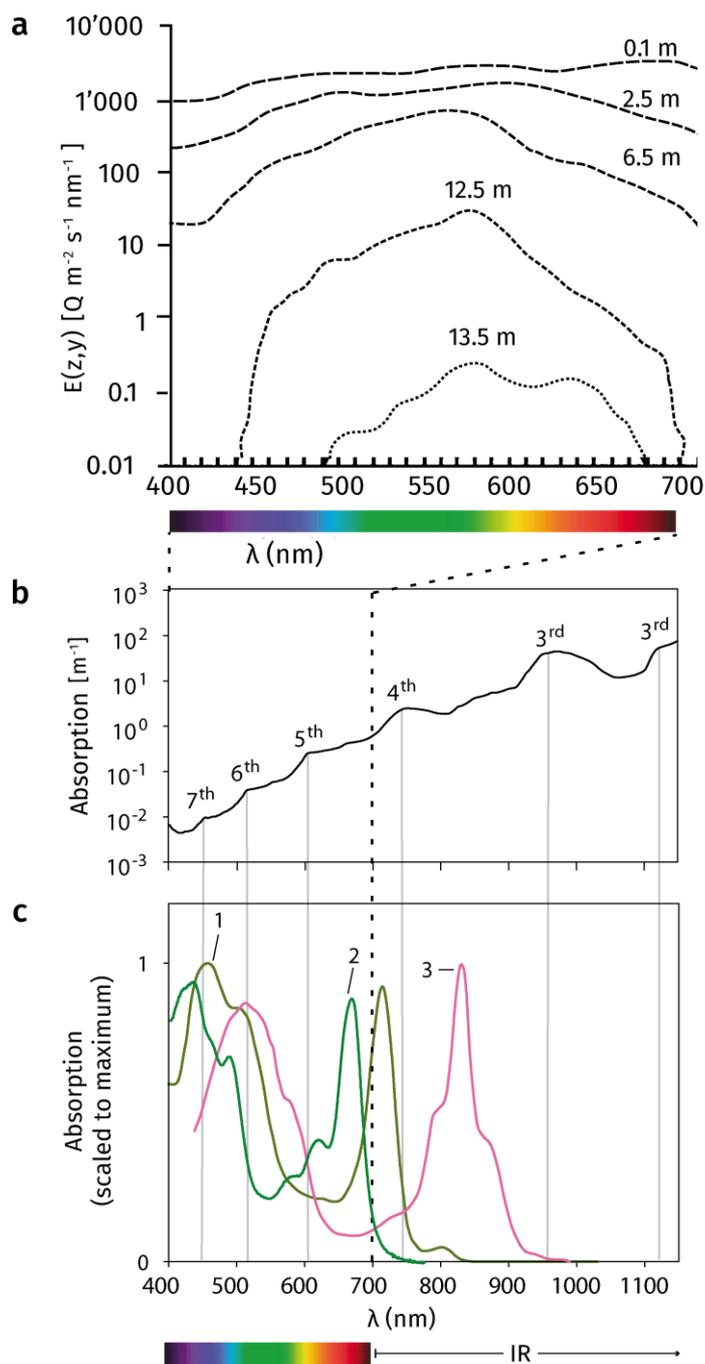
Figure 1.3-4| Sunlight is absorbed in water through a combination of the agitation of vibrational modes of the water molecule and the sum of pigments from phototrophic organisms.

a) Underwater light spectrum at different depths of a permanently stratified alpine lake (Lake Cadagno) measured in August 1985. At the surface as substantial fraction of solar radiation is present in the wavelength range between 400 and 750 nm. A 100–1'000× reduction in light intensity is measured at 12.55 m depth compared to the surface (0.1 m). Green to red light prevails at 13.55 m. b) Absorption spectrum (log-scale) of pure water from 400–1'100 nm. The H_2O has three vibrational modes and the wavebands of the matching energy requirements are absorbed. Grey lines indicate harmonics of the stretching and bending vibrations. c) Absorption spectra of phototrophic microorganism found in lakes 1) GSB *Chlorobium phaeoclathratiforme* [BChl *e* and isorenieratene and *β*-isorenieratene], 2) green algae *Chlamydomonas* sp. [Chl a], b) and 3) PSB *Chromatium okenii* [BChl *a* and okenone] that show the adaptation to the radiation-niche defined by the water column. The visible spectrum is indicated by a spectral color bar and the range is denoted by a vertical, dotted line in b) and c). IR: infrared. Adapted from del Don et al 2001 [122] (a) and Stomp et al 2007 [92] (b,c).

1.4 Microbial ecosystems in permanently stratified lakes – Lake Cadagno as an example

The physical conditions largely define biological and chemical processes in the environment studied. As the biogeochemical processes are studied over a wide time range (billion years to hours) both, dynamic and recurring (eco-) systems are needed. Permanently stratified (meromictic) lakes are therefore ideal since anoxic bottom water preserves chemical and biological remnants, in- and outflow of organisms and nutrients are more readily assessed, the catchment area is defined and they are easy to access compared to the open ocean [110]. The vertical stability of layers in lakes largely depends on seasonal temperature changes, salinity, bathymetry and the grade of exposure to external factors such as storms, avalanches and landslides. Additionally, the anaerobic microbial degradation of organic matter leads to increased concentrations of solutes (e.g. HCO_3^-, HS^- and NH_4^+) that stabilizes the pycnocline and eventually lead to a permanent stratification [111]. Thereby, the water column is characterized by a partly oxygenated upper mixolimnion, the chemocline, a transition zone with steep gradients of dissolved oxygen, redox potential and salinity, and the anoxic and strongly reduced monimolimnion at the bottom [111]. As a consequence, these stable physical and chemical conditions contribute to a zonation of microbial communities that allows the measurement of biogeochemical fluxes. Prime examples of the 177 meromictic lakes studied worldwide are Mahoney Lake in Canada [112], Lake Rogoznica at the Croatian coast [113], Lake Faro on Sicily [114], Lake Kaiike in Japan [115], Spanish karst lakes [116], the saline-alkaline lakes in Central Asia [117] and Lake Cadagno in the Swiss alps [118]. In the following paragraph I will concentrate on Lake Cadagno to elaborate on the microbial processes typically associated with meromixis.

1.4.1 Geology and formation of Lake Cadagno – a short environmental history

The Piora valley (Ticino, Switzerland) at around 2'000 m.a.s.l. that spans the southern Alps 8 km in an east west direction encompasses Lake Cadagno (Figure 1.4-1a) and 10 other lakes. The geology of the Gotthard region (Switzerland) has been described by Garwood in 1906 [119] and Krige 1917 [120] and has recently been updated and revised in the Swiss Geological Survey's geological maps [121]. It is lined by two [122] metamorphic crystalline rocks in the north and in the south (Lukmanier layer) whereas the porous bedrock along the
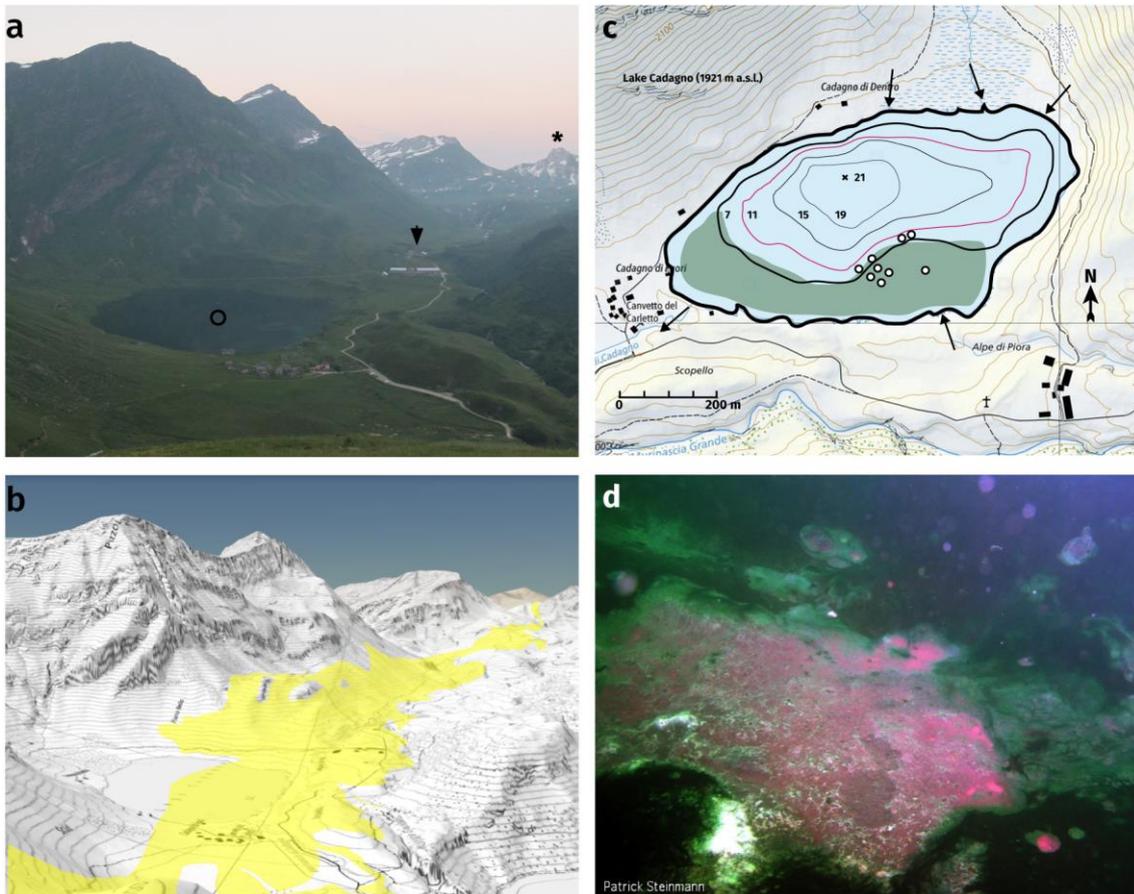
Figure 1.4-1 | Characteristics of Lake Cadagno

a).View along the upper Piora valley from the West to East with Lake Cadagno (circle), Alpe di Piora (arrow) and the dolomitic Pizzo Colombe (2'545 m, asterisk). Lake Cadagno (46°33'N, 8°43'E) at 1'921 m a.s.l. has a maximal depth of 21 m (mean depth 9.27 m), a surface area of 261'043 m² (0.26 km2) and a volume of 2'419'850 m³ (2'419'850'000 l) [122]. Early signs of deforestation have been dated back to 2'800 kyr BP [123] and the valley has been in use as alp since the middle age [124]. The lake is surrounded by meadows that are grazed by the herd of 200 milk cows of the nearby Alpe di Piora in summer season (June-August). Dense green alder (Alnus viridis) shrubs cover the southern slopes. b) Map with a 3D-representation of the geological map of the Piora valley imposed .The same viewpoint as in a). In yellow: Röti-formation, dolomite with gypsum. Source: swisstopo.ch,

Geo25 atlas. c) Bathymetric map of Lake Cadagno and other geographical features (summer situation). Arrows indicate tributaries and the effluent of surface water. The pink contour-line denotes the contour line (1'480 m) at the chemocline measured at an average depth of 11 m. The deepest point of the lake is marked with a cross. Open circles indicate the locations of the subaquatic spring rich in sulfate, bicarbonate, Ca²⁺ and Mg²⁺ (~5 g s⁻¹). The littoral zone shaded green is covered by stoneworts (Chara spp.) and pondweeds (Potamogeton spp.). Adapted from Ref. [122] and including map material form swisstopo.ch (1:25'000 map). d) Cyanobacteria, Beggiatoa sp. and purple sulfur bacteria form dense bacterial mats around the subaquatic springs (Image courtesy of Patrick Steinmann, 1998).

valley bottom consisting of Rauwacke and Dolomite containing gypsum (Piora Zone) (Figure 1.4-1b). The retraction of glaciers around 12'000 years ago formed the Lake Cadagno basin and left a moraine that formed a natural dam [125]. The lake is fed by surface water from the crystalline northern slope. A ~2'300 year-long transition time between oxic and anoxic lake conditions followed. The thawing of the permafrost allowed the percolation of solute-rich water through the karstic system into the lake (some l s⁻¹), thereby allowing the formation of a

crenogenic meromixis. Importantly, a short period between thermic stratification in fall and the ice cover during winter from December to May prevents complete thermic mixing [122]. Regular floods between 11'950 to 10'350 kyr BP however disturbed the anoxic conditions, transporting allochthonous material and O_2 into the lower lake layers [125]. Permanent euxinic conditions since 9'800 years have been concluded based on a stable Mo-concentration in the sediments and remnants of anaerobic phototrophic sulfur oxidizing bacteria (SOB) of the families *Chromatiaceae* (PSB) and *Chlorobiaceae* (GSB) [126, 125]. Anthropogenic influence is mainly due to the deposition of heavy metals between 1950–1990 [127] and the summerly nutrient input from manure of the alpine dairy.

1.4.2   The mixolimnion of Lake Cadagno – primary production and metazoa

The Lake Cadagno mixolimnion (2'318'000 m^3) spans down to around 11 m and is fed by surface tributaries with low-salt water throughout the summer and fall (Figure 1.4-1c). Thermal mixing occurs in the first 5 m due to diel temperature changes (0–20 °C) and the thermocline is found at 7–8 m [122]. Low concentrations of inorganic nitrogen (NO_2^-, NO_3^- and NH_4^+, respectively) at ~1 μM and < 0.03 μM for soluble phosphorous are characteristic [128]. A graphical summary of the different microbial metabolisms studied in Lake Cadagno is given in Figure 1.4-2. The oxygenic algal and microbial community of Lake Cadagno has been studied by Dueggeli and Brutschy in the early 20th century [129, 130]. The primary production rates by phytoplankton and *cyanobacteria* in summer were found to be in the range of 4.9 and 7.8 mg C m^{-3} h^{-1} in the first 10 m, and up to 45.9 mg C m^{-3} h^{-1} at 10.5 m depth, respectively [128]. Whereas UV radiation has been found to inhibit oxygenic photoassimilation down to one meter in Lake Cadagno [131–133], Cyanobacteria occur throughout the mixolimnion and even occur at a depth of 15 m [128, 134]. The chlorophyte *Echinocoleum* spp. dominate the epilimion down to 8 m where diatoms (*Fragilaria* and *Cyclotella spp.) and Cryptomonas* spp. were found from 9 m to the oxic-anoxic boundary at 11 m depth [118, 128]. The zooplankton community consists of cladocera (*Daphnia* and *Bosmino* spp.), rotifers (*Conochilus* and *Asplanchna* spp.), copeopods (*Acanthodiaptomu* sp.) and cyclopoid species (*Cyclops* sp.). Interestingly, ~50 % of the zooplankton's carbon input derives from grazing on the SOB in the chemocline [128]. Additionally, fish (*Salmo trutta* ssp.) have been introduced seasonally.
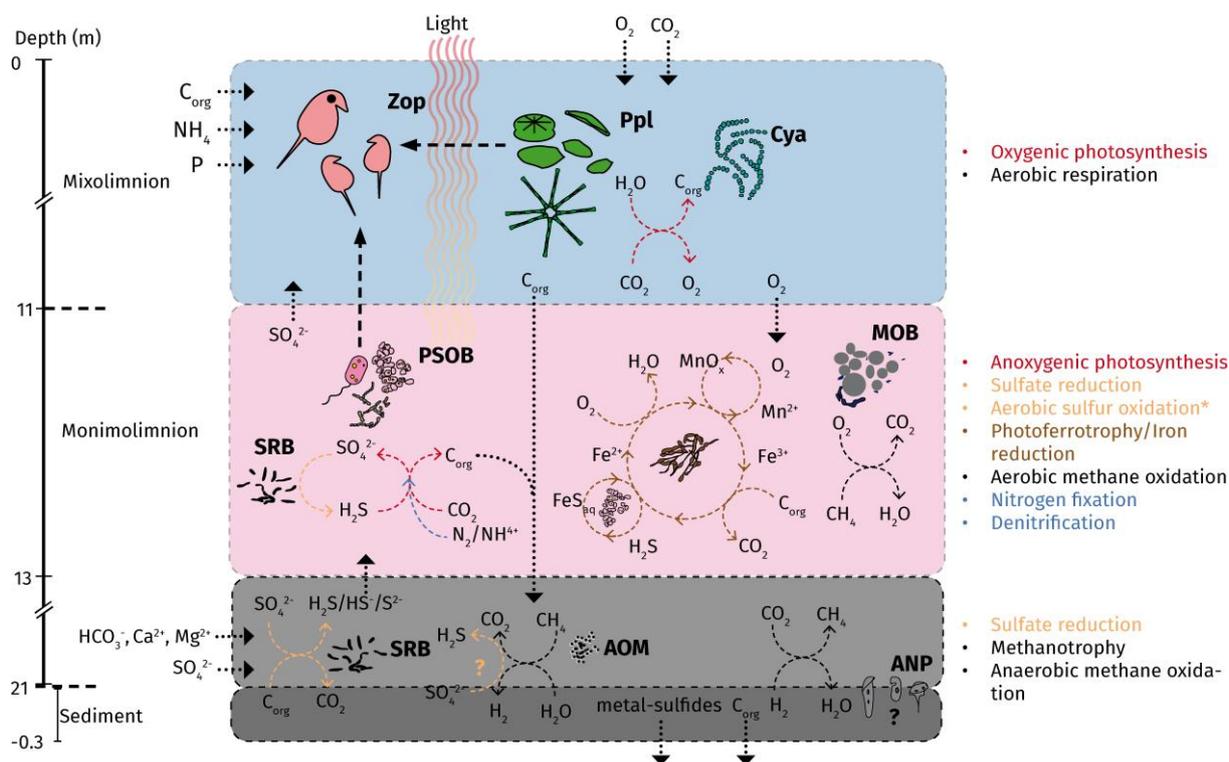
Figure 1.4-2 | Schematic overview of the microbial processes studied in Lake Cadagno.

Typical microbial metabolisms that have been described for other meromictic lakes can be found on the right. The lake can be divided in four different vertical zones based on water chemistry and light availability (not drawn to scale).The upper mixolimnion (blue) is dominated by oxygenic phototrophs (chlorophytes, diatoms and *Cryptomonas* and *cyanobacteria*) that profit from the light and nutrients [135]. Zooplankton is both feeding on primary producers in the oxic part, and occasionally dips down to the oxic anoxic transition zone at around 12 m [128]. Aerobic sulfur oxidation has been predicted around the benthic sulfur springs [122] Detritus is sinking downwards in the anoxic part whereas sulfate is diffusing upwards. The Monimolimnion consists of a photic (rose, 11–13 m) and an aphotic part (grey, below 13 m). Within the chemocline (rose) different coupled redox-transformations between and within members of the same species have been described: Sulfur oxidation and reduction by SOB [87, 106], simultaneous nitrogen uptake and denitrification by GSB [136] and methane oxidation dependent on in situ oxygenic photosynthesis (AOM) from α- γ-proteobacteria (*Methylorosula*, *Sphingomonadaceaean* and *Methylococcaceae* sp.), respectively [137] and cryptic iron redox-cycling by several bacteria (PSB sp., Rhodomicrobium sp., Rhodoferax sp.) [138]. A wide range of heterotrophs along the water column have been found in several studies based on 16S rRNA gene sequences [118]. However their role and impact on the Lake Cadagno ecosystem is unknown. Within the lower sulfidic monimolimnion and on the water-sediment transition zone (dark grey) sulfate reduction by SRB [139] and methanotrophy prevail [140]. The role of anaerobic protists in the top sediment has still to be elucidated [141]. Different arrows indicate following processes: triangle wide, line dotted: diffusion; triangle, line dashed: grazing; simple, line dashed: metabolic transformations Abbreviations: ANP: anaerobic Protists, AOM: Anaerobic Oxidation of Methane, Cya: Cyanobacteria, MOB: Methane Oxidizing Bacteria, Ppl: Phytoplankton, SOB: Sulfur Oxidizing Bacteria, SRB: Sulfate Reducing Bacteria, Zop: Zooplankton. Not drawn to scale. Due to better visual clarity stoichiometry was omitted.

The subaquatic sources are found in the south-western littoral zone (Figure 1.4-1c) that is overgrown by stoneworts (*Chara globularis*.) and that is locally covered with white deposits of chemically oxidized sulfate and purple and green bacterial mats (*Chromatium, Chloroflexi* and Cyanobacteria.) [142, 143] (Figure 1.4-1d). The permanent sulfate supply within the benthic sediment allows for high microbial sulfide production rates [139]. High values of sulfide (4 mM) and high ionic strength (9–10 mM) were measured *in situ* and sulfidic water

flows lateral and downwards along the sediment surface and accumulates in the lower lake basin [142].

1.4.3 The oxic-anoxic interface – multiple microbial metabolisms co-exist in the chemocline

Early microbial studies on the Lake Cadagno and the neighboring Lake Ritom focused on the visible water coloration and turbidity that was found at around a depth of 10 to 12 m and correspond to rapid changes in the water chemistry [129, 144]. On Lake Ritom a hydroelectric dam was built in 1918 and as a consequence the PSOB population vanished within one year [145]. The Lake Cadagno chemocline is defined by steep gradients in oxygen and sulfide concentrations, as well as in the redox potential and conductivity. With a volume of ~100'000 m^3 it equals 10% of the lake volume. Both passive vertical chemocline displacement and extension (centimeters to meter) over diel and seasonal scales, as well as internal mixing processes have been described [146–148]. Whereas oxygen is depleted below detection (0.1 μM), an increase in sulfide from 0 to 0.2 μM corresponds to maximum of total phosphorous (~6 μM) and to a high concentration of BChl a (350 mg m^{-3}) and biomass of 38–120 mg m^{-3} [149]. The relative light availability drops from ~4.7% to ~0.4 % surface PAR that correlates with increased turbidity within some cm vertical distance [149] (see also Figure 1.3-4 b). A SOB bacterial plate has been observed qualitatively to seasonally occur in summer and fall over the last 100 years [130, 144, 145, 150, 151]. This dense community (>10^7cells ml^{-1}) has been further described with molecular methods since the 1990ies using temporal temperature gradient gel electrophoresis (TTGE), 16S RNA-gene sequencing, fluorescent in situ hybridization (FISH) and flow cytometry (FCM) [152–154, 134]. The community of PSB with *Chromatium okenii*, 'Thiodictyon syntrophicum' str. Cad16^T, *Lamprocystis* spp. and *Thiocystis* spp. regularly grows up to 10^5 cells ml^{-1} from June to October [154] in the years 1985–2000. In winter the population diminishes to about 25 % [155]. Whereas *C. okenii* dominate in numbers from June to August, *Lamprocystis* spp. and Cad16^T dominate in fall and winter during some seasons [153, 155]. Over the years examined, the SOB community made up around 30% between 1998 and 1999 [152, 155], to 50-75% between 2000 and 2007 [156] and 24% in 2016 [134] of the total bacterial cells, respectively.

In situ functional studies on PSB physiology at different depths of the chemocline encompassed BChl synthesis rate measurements of *C. okenii* [157], *Kautsky*-fluorescence

kinetics of phototrophic reaction centers [158], sulfur-cycling rates [159] and carbon storage dynamics [160]. To sum up, these studies suggest a low-light adapted population of, PSB *C. okenii* especially, that readily adapts the BChl *a*-synthesis to the rapidly-changing light availability. Together with the ability to store S^0, polyglucose and PHB, *C. okenii* might be able to still actively move below the chemocline in winter with insufficient light [161]. SOB photoassimilation has been further studied with *in situ* $H^{14}CO_3^-$ incubations [128, 162] and ^{13}C-stable isotope analysis of single cells by nano-scale secondary-ion mass spectrometry (nanoSIMS) [163]. Together, these findings have shown that PSB "*T. syntrophicum*" str. Cad16^T and *Lamprocystis purpurea* and *Thiocystis chemoclinalis* str. CadH11^T were responsible for 25.9, 5.8 and 0.2 % (measured in ng C ml^{-1}) [163], whereas *C. okenii* was responsible for up 70% of the total carbon photoassimilation (measured in nM h^{-1}) [162] , respectively. Additionally, PSB were also accounted for c. 41 % of the total NH_4^+ assimilation rate measured (133 nmol N h^{-1}) [163]. Interestingly, str. Cad16^T forms aggregates with SRB (*Desulfocapsa* sp.) allowing for sulfur cycling within the diffusion barrier [164] A change in the SRB-to-PSB cell-ratio from 1:2 to 2:1 was found for the aggregates between 1998 and 2004 possibly depending on trends in sulfide and light availability due to SOB population dynamics [165].

An increase in numbers of the GSB *Chlorobium clathratiforme* was noticed first in 1999 and it dominated the following decade with up 1×10^7 cells ml^{-1} equal to 95 % of all PSOB cells in 2007 [155]. The population has been described to be largely clonal comparing the 16S RNA-gene [156]. Despite high cell concentrations, *C. clathratiforme* contributed only to 2–15 % of the photoassimilation with large cell-to-cell differences [162, 163], whereas a metaproteomic study showed the possibility of glucose fermentation in *C. clathratiforme* under sub-optimal light conditions [87]. For the other low abundant GSB (1×10^3 cells ml^{-1}) *C. phaeobacteroides,* simultaneous nitrification and de-nitrification was measured *in situ* [136]. Another study combined flow cell sorting with nanoSIMS analysis found N-coupled carbon fixation in the absence of NH_4^+ in *C. phaeobacteroides* [166]. To conclude, the GSB community shows physiological population heterogeneity both in carbon uptake and nitrogen fixation and adaption to low-light conditions and heterotrophic growth. Noteworthy, during the last ten years the GSB population was again reduced to lower cell concentrations as observed before the year 2000 [134]. Taken together, the SOB account for 41% of the total photoassimilation of the entire lake [128]. Despite these studies, still 25–66 % carbon assimilations is unexplained during daytime [162, 163] and oxygenic photoassimilation and

chemoautotrophic "dark" fixation can make up to 25% and 50% of the total carbon fixation, respectively [128].

Interestingly, cryptophyta [128] and Cyanobacteria [134] where found to thrive at the chemocline down to 15 m depth, indicating the possibility of *in situ* oxygenic photoassimilation and nitrogen fixation [136]. In accordance, micro-oxic conditions between 1 and 0.02 μM below 10 m have been measured with micro-optodes during daytime [137]. These conditions might still allow for light-driven aerobic methane oxidation (MOB) [137] and ferrous iron (Fe^{2+}) oxidation [138] (Figure 1.4-2). Additionally, PSB might be able to grow chemoautotrophically by respiration [128, 138]. An increase in NH_4^+ (10–70 μM), soluble phosphorus (0.7–4.1 μM) [128], CH_4 (>20 μM) and up 300 μM total sulfide (H_2S + HS^- + S^{2-}) [167] is measured down to 13–21 m in the lower euxinic and dark monimolimnion.

1.4.4 Dark, anoxic and cold – biomineralisation and sulfur deposition in the Lake Cadagno sediment

About 114'800 m^2 of the lake bottom is covered by meter-thick black to grey fine-grained sediments. The composition of both, the plankton and the chemocline SOB community is mirrored to some extend in the euxinic sediment record of silica-shells, DNA fragments and carotenoid-remnants [126, 168]. With an average sedimentation rate of 5 mm year^{-1} and a lack of bioturbation, layers of different lithologies are preserved that allow paleo-climatic reconstructions [125]. The high organic carbon input form the waters above fuels the biomineralisation involving SRB within the top 0.15 m of the sediment and one third of sulfide diffuses upwards into the monimolimnion [127, 139]. Thereby, extensive sulfate reduction leads also to abiotic Zn, Cd and Pb precipitation in the upper 0.05–0.2 m [127] that correlates with an increase of endospore-forming *Firmicutes* [169]. Within the lower 0.5 m Fe(II) and Mn(II) are produced and again consumed at 0.3 m where FeS_2 and MnS are formed [139]. Moreover, predicted microbial and archaeal CH_4 production [141, 170] is coupled to anaerobic methane oxidation (AOM) by of AOM-associated archaea at a depth of c. 0.2 m [140]. The role of Ciliphora (ciliates) present harboring endosymbiotic methanogens has still to be elucidated [141].

1.5   Meta-omics technologies to study microbial ecology

As this thesis is focused on understanding the complex microbial community in the chemocline of Lake Cadagno I will outline the theory and technology applied. To provide a larger context for these discussions, a brief historical overview of the concepts of microbial ecology and techniques to study microbial diversity will be given in this paragraph.

Earl microbiologist in the 19th century focused on human pathogens and Koch developed the concept of microbial pathogenesis by isolating, cultivating and re-infecting hosts with 'pure' cultures [171]. If successfully isolated, organisms were classified microscopically by morphology, and pathogenicity or/and physiology [172]. Bacteria in the environment were described in the early 19th century [40] and microbiologist became soon aware of the inherent complexity of bacterial communities. As a consequence, the subtle nutritional needs and community interactions have often complicated their cultivation, isolation and detailed phenotypic description of the organisms found. In contrast to Koch's concept of purity, Sergei N. Winogradsky developed a 'dirty' conception where ecosystems were imported and reconstructed *in vitro*, most notably he developed the Winogradsky column [173]. Thereby one could follow bacterial zonation by the bare eye and metabolic activities could be followed by quantifying metabolic rates. As the broad strokes of metabolic networks could be elucidated thereby, the numbers and type of involved species often remained elusive.

The discoveries of heredity at a molecular level in the years from 1940–1970 led to an early hypothesis on protein sequence similarity as a measure of relatedness between organisms [174]. With the possibility of DNA clone-library sequencing in the mid 1980ies [175] and the concept of the 16S rRNA and 18S rRNA genes as universal phylogenetic markers [176], microbial phylogeny was revolutionized [177] and was also soon used to profile microbial communities quantitatively [178]. Subsequently, the advances in both, soft- and hardware for genome based high-throughput computational methods in the last 20 years have enabled insights into diverse microbial ecosystems beyond mere description.

1.5.1   Bacterial Diversity and Genome Sequencing

The study of complete genomes of species (genomics) and those of communities (metagenomics) was and is largely driven by the advances in high-throughput sequencing technology that enable fast, reliable and low-cost whole genome sequencing (WGS) of microbes [179]. Whereas *Sanger* sequencing is based on sequence chain-termination, the second and third generation sequencing technologies (illumina, 454 and ion torrent) rely on surface-bound PCR amplification of short DNA fragments (25–500 bp) and the detection of polymerization reactions of fluorescent nucleotides or the release of hydrogen thereby, respectively. In contrast, fourth and fifth generation sequencing technologies such as PacBio and nanopore sequencing use the complete input DNA without amplification and produce longer segments [180]. As sequencing technology has constantly improved and the price per base sequenced is still dropping, an exponential increase of novel genomes sequenced is observed, with currently around 22'500 new genome projects per year [181]. However, the quality of sequence assembly and annotation has not kept pace with the increase in sequence data. As prokaryotic genome assembly and annotation are largely automated [182, 183], a plethora of different algorithms have been developed. As a consequence, large variations in assembly [184] and annotation quality have emerged leading to erroneous datasets, a problematic that is further increasing [185]. Contaminating sequences are an additional problem arising from erroneous assemblies, especially in large polyploid genomes [186].

WGS allows further improving phylogeny of prokaryotes compared to the 16S rRNA gene based methodology [187, 188] as organisms are compared on the level of complete sets of ribosomal proteins (ribosomal multi-locus sequence typing; MSLT) [189], entire core genomes [190] or the average nucleotide identity (ANI) [191] or a combination thereof [192]. Additionally taking into account large intra-genomic rearrangements [193], extrachromosomal DNA such as plasmids [194] and bacteriophages [195], and horizontal gene transfer (HGT) [196] has led to an increasingly complex tree of life mapping out bacterial diversity [197, 198]. A number of statistical methods have been developed to estimate species richness, diversity and evenness based on metagenomic data as summarized in Locey and Lennon [199] and Widder *et al* [200].

Different microbial habitats have been studied on metagenomic level such as acid mining drainage biofilm [201], a meromictic lake [202], frozen lakes [203] the human gut

microbiome [204], a whale carcass [205] and subway stations in New York [206], as examples.

However, sequencing gives only information on the 'potentiality' not directly on function and includes also non-living bacteria [207]. Therefore findings rather draw from correlation than causality. Therefore, microbial ecosystem functioning is ideally studied on the level of translation (transcriptomics) and transcription (proteomics).

1.5.2   Quantitative transcriptome and proteome analysis

Transcription analysis is used to detect changes in physiology upon stimuli at the level of gene transcription and RNA based regulatory networks. Technology has evolved from the targeted microchip to mRNAseq quantitation based on short-read 3rd generation sequencing, massively expanding both resolution and coverage. However, problematic for the analysis are the inherent instability of the RNA molecule and the lack of correlation between the transcription levels and actual protein abundance [208].

Therefore proteins are the preferable analyte to mirror pathways and structures actually expressed. A proteomics experiment can be divided into following steps: sample preparation, peptide or protein isolation and separation, peptide/protein mass spectrometry (MS)-analysis and identification and quantitation. A detailed overview on methods and application in microbial ecology is given by Wöhlbrand *et al* [208].

As proteins are composed of 21 amino acids that allow for enormous structural and physiochemical versatility this renders them difficult to analyze. Parameters such as mass, relative abundance, hydrophobicity, polarity and posttranslational modifications are highly variable within an organism. As a consequence, an unbiased extraction protocol has to be combined with loss-free separation steps. Crucial measures are cold temperatures and protease inhibitors that reduce proteolysis. Cells can be homogenized by biochemical, chemical, physical methods, or by a combination of the former, respectively and subcellular compartments are fractioned through subsequent chemical lysis (e.g. a change from aqueous buffers to organic solvents) and ultra-centrifugation. Bile-salt (e.g. sodium deoxy cholate) based methods have thereby been proven to be superior detergents with low bias on hydrophobicity [209, 210].

As smaller protein fragments can be measured with better resolution in mass spectrometry (MS) and are better dissolved, the protein lysates are subsequently hydrolyzed by the protease trypsin or /and other proteases, a process also called bottom-up proteomics. As trypsin specifically cuts peptide bonds at the carboxyl side of arginine or lysine, fragments can be reliably predicted from genomic data. Label free quantitation of peptides is based on a high number of technical replicates (i.e. LC-runs) and is preferentially used for differential expression study as in this thesis.

The following, de-complexation and separation steps of entire proteins was done in SDS-PAGE gel-based 1D and 2D-iso-electric focusing (IEF) systems that have been largely replaced today by high performance liquid chromatography (HPLC) of peptides. Due to the on-line high through-put capability and high sensitivity, nano HPLC systems with a flow of 300 nl min^{-1}, that include a one dimensional reverse phase (RP) or a RP combined with ion-exchange chromatography (IEC) which are coupled directly to theMS. In electro spray ionization (ESI), acidified analytes leaving the HPLC are sprayed into an electrical potential between outlet and mass analyzer that leads to charged drops [211]. As drops constantly evaporate while moving toward the orifice, increasing charge repulsion leads to further droplets, that finally results in the desorption of multiple protonated analytes into gas phase at a high vacuum. The analytes are then entering a series of multiple quadrupole (Q) mass filters. Normally three units termed Q_1–Q_3 are installed sequentially (triple-quad; TQMS) where the first and last function as mass selectors, the middle figures as a collision cell to induce collision induced dissociation (CID) for tandem MS (MS^2). The analyte-ion is modulated by pulsed radio-frequency (*rf*) fields and thereby enters a mass dependent- flight path and is then selectively transferred to the Q_2. There, the ion is fragmented and the resulting fractions are accelerated into the Q_3 where *rf*-based selection for the mass analyzer, that normally is a time-of-flight (TOF) detector [212]. Another common mass analyzer is the orbitrap cell [213] with 140'000 FWHM resolution, <1 ppm mass accuracy, a mass range of about 50–6000 m/z and high frequency [214]. The MS^2 peptides detected are then searched in customized database that contain translated CDS sequences derived from genomic data (i.e. genomes, metagenomes and mRNAseq) and/or protein data from previous experiments. An overview is given in e.g. [215, 216]. Common search algorithms are MASCOT [217], MaxQuant [218] and MS-GF+ [219].

Metaproteomic studies have extended form simple acid-mine drainage communities [220], a termite microbiome [221], to complex coral reefs [222], algal-bacterial communities [223, 224] and cyanobacterial exoproteomes [225, 226]. Different PSOB isolates and communities have been studied with proteomics so far. In order to understand the photoautotrophic and mixotrophic metabolic pathways in PSB *A. vinosum,* a combination of genomics, transcriptomics, proteomics and metabolomics experiments was used *in vitro* [72, 77, 227, 228]. Similarly, growth GSB *Chlorobaculum tepidum* on different sulfur compounds revealed different expression pattern of the Dsr and Sox pathways *in vitro* [229] and a metaproteomic analysis of a functional clonal GSB population showed differential expression dependent on light availability [87].

1.6 Proteogenomics for mass spectrometry based bacterial typing

Matrix-assisted laser-desorption and ionization time-of-flight mass-spectrometry (MALDI-TOF MS; from here on MALDI-MS) developed in 1985 [230] is a versatile and fast MS method that detects a large variety of chemical compounds from protons to macro-molecular complexes, with only minimal sample preparation needed. A matrix is a substance that is excitable through UV-laser pulses. The matrix is mixed with the sample forming crystal-like structures. Under vacuum, a UV laser beam is directed towards the sample that leads to the sample desorption and the disruption of the matrix. Thereby charges are being transferred by a complex process and molecules are ionized. The grade of sample ionization is non-linear and depends amongst other things on the amount, the intrinsic charge and the size of the analyte, that in consequence, hinders quantitation and concurrent analysis of mixed samples by MALDI-MS [231]. As voltage is applied along a flight tube, ions accelerated and separated and their m/z signals are detected as a function of time.

Biomolecules have been already studied by MALDI-MS as early as 1990 [232]. Subsequently, numerous studies utilized 'protein fingerprinting', that compares spectra from entire cells with an m/z database from complementary characterized isolates (e.g. [232–235]), a concept where, however, identity and quantity of proteins measured is unknown. This MALDI typing method is therefore limited by extend of the m/z-database that consist mostly of clinical isolates of bacteria and fungi [236, 237]. In some projects, MALDI-MS typing has been used to study phytoplankton [238] and environmental isolate of *E. coli* [239].

Ribosomal subunits have been suggested as potential MALDI-MS biomarkers, since these about 52 proteins are found conserved in all species, however with some mass differences reflecting phylogeny [189]. Furthermore, HGT of ribosomal genes is limited, the subunits are highly abundant, intrinsically positively charged and in the mass range of 2'000–30'000 Da, thereby allowing reliable MALDI-MS detection [240–242]. These putative masses must however be further verified by separately isolating proteins and subsequent MS^2 protein/peptide sequencing [243]. The recent development of MALDI-MS that allow for on-plate top-down protein sequencing (TDS), might allow a more straight forward initial characterization of 1D-seperated protein mixtures from cells lysates [244, 245]. The *de novo* sequencing of proteins furthermore allows the correction of incorrect CDS annotations, a process called proteogenomics [246], and the detection of post-translational modifications [247].

Biomarker characterization in combination with the increasing number genomic sequences from WGS, complete genomes from singe cells (MGA) and metagenomes from different environments, allows curating an ever expanding m/z-database. An increasing number of publically available and commercial m/z biomarker databases based on genomics already exist [248–250]. As traits such as antibiotic resistance or pathogenicity are protein encoded and typical metabolites can be simultaneously detected, the biomarker dataset can be expanded to strain or species specific targeted research [251, 252]. Another recent development is single-cell MALDI-MS, where levels of low molecular metabolites such as fatty acids are quantified semi-quantitatively over a large number of individual cells [253].

1.7 Aim and Objectives

The microbial ecosystem of Lake Cadagno has been described in some detail, while the biological functions within and between the microbes have only been partly understood.

The importance of PSOB for the primary production — both in the light and in the dark — was previously confirmed with the isolation of small-celled PSB strain Cad16^T being highly active in C-assimilation. However the detailed dark carbon fixation mechanism remained unclear. The observed high carbon and nitrogen uptake rates *in situ*, as well as synchronized swimming behavior of large celled PSB *C. okenii* pointed out to the importance of this organism for the lake ecology. In order to understand and compare the pathways and

structures at disposal in depth, we have sequenced the complete genomes of both organism using 4th generation long-read sequencing. The encoded functions might then allow drawing conclusions on differentiation and the potential ecological niches that can be filled by the different PSB.

The genomic information is the corner stone for further quantitative proteomics studies to compare different physiological states at the level of pathways. For str. $Cad16^T$ we have designed an *in situ* incubation experiment that enabled us to study physiology at day and night while monitoring important environmental factors.

Beside the ecological studies using 'omics', the need for fast and reliable tools for the identification of novel isolates is given as there are many bacterial species not yet isolated in the chemocline. In order to reliably and rapidly characterize novel PSOB environmental isolates, we have developed a MALDI-MS pipeline including sample preparation, HPLC chromatography, MS^2 analysis and a genomic database that allows the verification of genome-derived m/z markers.

My thesis is therefore structured in the following four chapters:

I.   Characterization of strain "*T. syntrophicum*" str. $Cad16^T$ with WGS (Chapter 2.1)

II. WGS of enriched *C. okenii* and comparative genomics with other PSOB from Lake Cadagno (Chapter 2.2)

III. Investigation of the CO_2-assimilation and central carbon metabolism of 'T. syntrophicum' str. $Cad16^T$ under light and dark conditions *in situ*, using a combination of quantitative proteomics and carbon radiotracers (Chapter 2.3).

IV. Development of a top-down MALDI-MS based m/z biomarker database for the rapid characterization of novel PSOB isolates.

2 RESULTS

2.1   Research Paper I

-47-

Complete genome sequence of '*Thiodictyon syntrophicum*' sp. nov. strain Cad16^T, a photolithoautotrophic purple sulfur bacterium isolated from the alpine meromictic Lake Cadagno

Samuel M. Luedin, Joël F. Pothier, Francesco Danza, Nicola Storelli, Niels-Ulrik Frigaard, Matthias Wittwer and Mauro Tonolla

Statement of contribution

I performed DNA extraction, sequence assembly and annotation and contributed to data interpretation and manuscript preparation.

Research objective

In order to understand the full metabolic potential we sequenced the complete genome of str. Cad16^T using long-read technology.

Standards in
Genomic Sciences

REPORT

Open Access

CrossMark

Complete genome sequence of *"Thiodictyon syntrophicum"* sp. nov. strain Cad16[T], a photolithoautotrophic purple sulfur bacterium isolated from the alpine meromictic Lake Cadagno

Samuel M. Luedin[1,2,3], Joël F. Pothier[4], Francesco Danza[1,2], Nicola Storelli[1,2], Niels-Ulrik Frigaard[5],
Matthias Wittwer[3] and Mauro Tonolla[1,2*] iD

Abstract

"Thiodictyon syntrophicum" sp. nov. strain Cad16[T] is a photoautotrophic purple sulfur bacterium belonging to the family of *Chromatiaceae* in the class of *Gammaproteobacteria*. The type strain Cad16[T] was isolated from the chemocline of the alpine meromictic Lake Cadagno in Switzerland. Strain Cad16[T] represents a key species within this sulfur-driven bacterial ecosystem with respect to carbon fixation. The 7.74-Mbp genome of strain Cad16[T] 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 232[T] 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 Cad16[T] 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 232[T] and *Thiodictyon bacillosum* strain DSM 234[T]. *"Thiodictyon syntrophicum"* sp. nov. strain Cad16[T] is the proposed type strain of the species *"Thiodictyon syntrophicum"* [3] within the family of *Chromatiaceae* of the genus *Thiodictyon* [4]. Cultures of

strain Cad16[T] 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 10^7 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 CO_2 fixation measured in Lake Cadagno [8]. Strain Cad16[T] has been shown to be highly active in CO_2 fixation both in situ and in vitro [9]. Furthermore, aggregation of strain Cad16[T] with SRBof the genus *Desulfocapsa* has been described [3]. In this publication we describe the first complete genome of strain Cad16[T] providing details especially on CO_2 fixation, sulfur metabolism and on CRISPRs. The sequencing of strain Cad16[T] is part of a larger sequencing project that includes the key species of

---

* Correspondence: mauro.tonolla@supsi.ch
[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 (DACD), Laboratory of Applied Microbiology (LMA), Via Mirasole 22A, 6500 Bellinzona, Switzerland
Full list of author information is available at the end of the article

the microbial community from the anoxic layers of Lake Cadagno.

Organism information
Classification and features

Strain Cad16[T] 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 Cad16[T] 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 Cad16[T] was classified as a member of the genus *Thiodictyon* within the family *Chromatiaceae* before [10]. Figure 2 shows the phylogenetic placement of strain Cad16[T] (complete 16S rRNA sequence) in a 16S rRNA based maximum likelihood phylogenetic tree. The closest relatives of isolate Cad16[T] are *T. bacillosum* DSM 234[T] and *T. elegans* DSM 232[T] with 99% sequence identity (partial 16S rRNA sequences). A comparison of the strain Cad16[T] core genome with other whole genome sequenced PSB confirmed the phylogenetic placemant (Additional file 1: Figure S1).

Strain Cad16[T] was anaerobically grown in Pfennigs medium [11], containing per liter: 0.25 g KH_2PO_4, 0.34 g NH_4Cl, 0.5 g $MgSO_4·7H_2O$, 0.25 g $CaCl_2·2H_2O$, 0.34 g KCl, 1.5 g $NaHCO_3$, 0.5 ml trace element solution SL_{10}, and 0.02 mg vitamin B_{12} with 2 mM acetate in 100 mL serum bottles with rubber stoppers. The medium was prepared in a 2 l bottle with a N_2/CO_2 (80%/20%) gas

phase. The medium was then reduced with 0.3 g l^{-1} $Na_2S·9H_2O$ (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 phototautotrophic 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 Cad16[T] was observed to assimilate only acetate, pyruvate and fructose in the presence of sulfide and bicarbonate. Strain Cad16[T] was additionally tested for chemolithoautrophic growth with bicarbonate under a headspace atmosphere containing 5% O_2, 10% CO_2 and 85% N_2, 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 Cad16[T] 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 at 528 nm for the carotenoid okenone, respectively [10].

A further characterization of strain Cad16[T] 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 Cad16[T] 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 Cad16[T] 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 Cad16[T] was anaerobically grown in Pfennigs medium [11] Cells were collected by centrifugation for 15 min at 10,600 g. DNA was extracted using phenol/

Fig. 1 Phase-contrast photomicrograph of "T. syntrophicum" sp. nov. strain Cad16[T]. The elementary sulfur globule inclusions are visible as black dots within the cell

Fig. 2 Phylogenetic tree indicating the position of "T. syntrophicum" sp. nov. strain Cad16[T] 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%

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 U*V*/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.

Results

Luedin *et al. Standards in Genomic Sciences* (2018) 13:14   Page 4 of 13

Table 1 Classification and general features of *"T. syntrophicum"* sp. nov. strain Cad16[T] according to the MIGS recommendations [65]

MIGS ID	Property	Term	Evidence code[a]
	Classification	Domain *Bacteria*	TAS [3, 59]
		Phylum *Proteobacteria*	TAS [3]
		Class *Gammaproteobacteria*	TAS [3]
		Order *Chromatiales*	TAS [3]
		Family *Chromatiaceae*	TAS [3]
		Genus *Thiodictyon*	TAS [2, 60]
		Species ""	TAS [3]
		Strain: Cad16[T]	TAS [3]
	Gram stain	Negative	TAS [3]
	Cell shape	Coccus	TAS [3]
	Motility	Non-motile	TAS [3]
	Sporulation	No	NAS
	Temperature range	5–25 °C	TAS [3]
	Optimum temperature	20–23	TAS [3]
	pH range; Optimum	6.8–7.5	TAS [3]
	Carbon source	CO_2, acetate, pyruvate, fructose	TAS [3]
MIGS-6	Habitat	Fresh water, alpine meromictic lake	TAS [3]
MIGS-6.3	Salinity	Not determined	NAS
MIGS-22	Oxygen requirement	Aerotolerant	TAS [3]
MIGS-15	Biotic relationship	Free-living	TAS [3]
MIGS-14	Pathogenicity	Non-pathogen	NAS
MIGS-4	Geographic location	Switzerland, Ticino	TAS [3]
MIGS-5	Sample collection	08.28.2001	TAS [3]
MIGS-4.1	Latitude	46°33′ N	TAS [3]
MIGS-4.2	Longitude	8°43′ E	TAS [3]
MIGS-4.4	Altitude	1923 m	TAS [3]

[a]Evidence codes – *IDA* Inferred from Direct Assay, *TAS* Traceable Author Statement (i.e., a direct report exists in the literature), *NAS* Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project [10]

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 Cad16[T] 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.

Luedin *et al. Standards in Genomic Sciences* (2018) 13:14

Page 5 of 13

Fig. 3 Schematic representation of the complete genome sequence of "Thiodictyon syntrophicum" sp. nov. strain Cad16[T] 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

Table 2 Project information for "T. syntrophicum" sp. nov. strain Cad16^T

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	200×
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

Extended insights from the genome sequence
Phototrophy

PSB typically transform light energy into chemical energy with the membrane bound type 2 photochemical reaction center. The chromosome of strain Cad16^T 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 180^T [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 homologousHiPIP (THSYN_25970) is found in strain Cad16^T. It may function as the main electron donor to the photosynthetic reaction center similar as in *A. vinosum* [20].

The absorption spectrum of strain Cad16^T 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 (ChlE) (THSYN_31385) and a light independent proto-chlorophyllide reductase complex (ChlLNB) (THSYN_31420–31,430) in strain Cad16^T.

Table 4 Genome statistics for the "T. syntrophicum" sp. nov. strain Cad16^T 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
rRNA genes	6	0.09
tRNA 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

Table 3 Summary of genome of "T. syntrophicum" sp. nov. strain Cad16^T: 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

Results

Luedin *et al. Standards in Genomic Sciences* (2018) 13:14   Page 7 of 13

Table 5 Number of genes associated with general COG functional categories of the genome of "T. syntrophicum" sp. nov. strain Cad16^T. The percentage is set relative to the total number of protein coding genes in the genome

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	Function unknown
No COG	1543	24.74	Not in COGs
Multi COG	320	5.13	Multiple COG assignments

Strain Cad16^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 χ-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 α, β and γ-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 γ-proteobacterium *Congregibacter litoralis* strain KT71^T and as in members from the α-proteobacteria families *Rhodobacteraceae* and *Rhodospirillaceae*, respectively.

Sulfur metabolism

For the photoautotrophic process of CO_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^T uses electrons from the oxidation of sulfide, thiosulfate and elemental sulfur as reducing equivalents [3]. Strain Cad16^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^T genome. However, *soxB* (THSYN_26690) and clustered genes

Results

Luedin *et al. Standards in Genomic Sciences* (2018) 13:14   Page 8 of 13

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[T] [26], no homologous sequence for the tetrathionate-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[T].

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

Moreover, the genome of strain Cad16[T] 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[T] genes in the *dsr* cluster (THSYN_22480, THSYN_22490–22,545) are arranged in a highly conserved organization similar to *A. vinosum* DSM 180[T], 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 dissimilatory 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[T] DsrC is able to bind DNA upstream the *dsr* cluster [31].

In strain Cad16[T], *soeABC* (THSYN_16370–16,380) encode the sulfur-iron molybdoprotein complex that further oxidizes sulfite to sulfate on the cytoplasmic site of the membrane [32]. Alternatively, strain Cad16[T] 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[T] sequence. For the extra-cytoplasmic export of the final oxidation product sulfate, a SulP sulfate permease (THSYN_14085) homolog to *A. vinosum* DSM 180[T] is encoded in the strain Cad16[T] 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[T], 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[T]. In accordance, an increase in the sulfide concentration was observed while SGB were consumed by strain Cad16[T] 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_2 evolution as proposed for *T. roseopersicina* [39, 40].

The Cad16[T] 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_2 fixation is essentially achieved through the reductive pentose phosphate also known as the CBB cycle. In accordance, the strain Cad16[T] 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_2 concentrations as proposed for *A. vinosum* DSM 180[T] [42]. Previous studies showed

Results

Luedin *et al. Standards in Genomic Sciences* (2018) 13:14   Page 9 of 13

different expression regulation for RuBis-CO type I and type II genes from Cad16^T suggesting that only the type II is involved in the process of CO_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 *sucCD* succinyl-CoA ligase (THSYN_00880 and THSYN_00885) that are essential for the TCA cycle, and isocitrate lyase (THSYN_16275) and malate synthase (THSYN_15655) that are essential for the glyoxylate cycle, respectively, are identified in the strain Cad16^T sequence. Recently a proteomic study about the capacity of Cad16^T to fix CO_2 in the dark suggested the presence of a particular archael DC/HB cycle [42]. However, nofurther genes coding for this DC/HB cycle were found. Also a complete set of genes coding for polyhydroxyalkanoic acid synthase PhaC (THSYN_06910) and poly-(3-hydroxybutyrate) depolymerase PhaE (THSYN_06905) are found in the strain Cad16^T genome.

Strain Cad16^T additionally encodes genes necessary for glycogen polymerisation. The glucose 1-phosphate adenylyltransferase 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.

Interestingely, strain Cad16^T also has the potential to produce the storage compound cyanophycin normally found in *caynobacteria* [43], since the two subunits of the enzyme cyanophycin synthetase (THSYN_26990 and THSYN_26995) are found.

Togheter, these finding provide genetic evidence for the high carbon fixation potential of strain Cad16^T in the dark [8, 44].

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

Strain Cad16^T grows chemoautotrophically under microaerobic conditions (5% O_2) with sulfide, thiosulfate, or sulfide only [3], as also observed in other PSB in vitro studies with *Lamprocystis purpurea* [10, 48], *Thiocystis violacea* and *A. vinosum* [49]. In situ, strain Cad16^T is possibly exposed to low concentration of oxygen produced by oxygenic microbiota at the mixolimnion-chemocline interface [8]. Accordingly, we observe genes encoding *sod*-type superoxide dismutases (THSYN_20405 and THSYN_22720), as well as *fnr* and *fur*-type

transcriptional regulators involved in peroxide stress response. In situ, strain Cad16^T is possibly exposed to oxygen produced by oxygenic microbiota 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^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^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^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 β-proteobacterial family of the *Burkholderiaceae* whereas THSYN_32425 is conserved in γ-proteobacteria and contains a type IV Rhs element. Togheter, the secretion machinery allows strain Cad16^T to interact within the highly populated chemocline with up to 10^7 bacterial cells per milliliter. The secretion and uptake mechanism may also play a key role in the cell-to-cell contact with *Desulfocapsa thiozymogenes.*

Buoyancy regulation and chemotaxis
Strain Cad16^T can possibly regulate buyoncy 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), GvpFL (THSYN_11800 and THSYN_18685), GvpK (THSYN_11785) and GvpN (THSYN_11815 and THSYN_18695) further stabilize the structure. Proteins homologoues to the transcriptional regulatory factors GvrA (THSYN_11850) and GvrC (THSYN_11830) from the enterobacterium *Serratia* sp. ATCC 39006 are also found in Cad16^T.

The diurnal and sesonal behavior of vacuolated *Chromatiaceae* has been described for different lakes [50, 51].

Results

Luedin *et al. Standards in Genomic Sciences* (2018) 13:14   Page 10 of 13

In strain Cad16^T a diguanylate cyclase (THSYN_19835) is found upstream the circadian clock genes *kaiCBB* (THSYN_19820–19,830). These genes act togheter [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 *cas* 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^T, 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 DRsusing the CRISPRfinder platform revealed similarities in CRR1, CRR2 and CRR4 to sequences of "*T. mobilis*" 8321 (57 hits, 2 mismatches) and "*Thioalkalivibrio sulfidophilus*" HL-EbGr7 (63 hits, 3 mismatches). The DRs found in CRR3 are similar to the ones in *Halothiobacillus neapolitanus* c2 (31 hits, 4 mismatches), whereas the DRs in CRR5 are similar to the ones found in *Vibrio alginolyticus* NBRC 15630 (1 hit, 5 mismatches).

Furthermore, three CRISPR-Cas loci were identified in the strain Cad16^T sequence, containing *cas3* genes that are characteristic for type I CRIPSR-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 *cas3*, 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 *cas8c* gene and the lack of a *cas6* 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 "*Thiodictyon syntrophicum*" sp. nov. strain Cad16^T 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^T 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^T 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^T will facilitate further studies that elucidate its role as key species of the chemocline and the tight association with the *Desulfocapsa* 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 a morphological and physiological differences (see above) indicate a novel species within the genus *Thiodictyon*.

The decribed isolate is therefore proposed as "*Thiodictyon syntrophicum*" sp. nov. strain Cad16^T, a novel species within the genus *Thiodictyon*.

A formal description of the proposed novel species follow below:

Description of "*Thiodictyon syntrophicum*" sp. nov.

"*Thiodictyon syntrophicum*" (syn.tro'phi.cum. Gr. pref. Syn, together with; Gr. adj. *Trophikos*, nursing, tending or feeding; N.L. neut. Adj. *syntrophicum*, syntrophic).

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

Table 6 CRISPR-Cas loci detected in "T. syntrophicum" sp. nov. strain Cad16^T genome

Localization	Name	CRISPR start	CRISPR end	CRISPR length (bp)	DR consensus	DR length	No. of spacers	CRISPR-Cas loci[a]
chromosome	CRR1CRR1	1,879,131	1,881,639	2508	GCTTCAATGAGGCCGCGGCGAATTCGCCGCGGAAAC	36	34	type I-U
	CRR2CRR2	1,883,646	1,894,325	10,679	GCTTCAATGAGGCCGCGGCGAATTCGCCGCGGAAAC	36	146	
	CRR3CRR3	4,626,522	4,629,249	2727	GCATCGCCCGGCCAATTGGCCGGGCGCGGATTGAAAC	37	37	type I-C
	CRR4CRR4	5,078,034	5,085,199	7165	GTTTCCGCGGCGAATTCGCCGCGGCCTCATTGAAGC	36	98	–
pTs485	CRR5CRR5	391,741	393,104	1363	GTAGCGCTACTCCGAGCCGCAAAGGCTATTGAAAC	35	19	–

[a] CRISPR-Cas classification according to Makarova et al. [58]

Luedin *et al. Standards in Genomic Sciences* (2018) 13:14

Page 11 of 13

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 photoperiode 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, glycerol, lactate, malate, propanol, propionate and succinate, respectively. Chemolitoautotrophic growth was observed with 5% Oxygen and 0.02% hydrogen sulfid and 0.07% thiosulfate, or with 0.07% sulfide only, respectively.

The type strain Cad16^T (=JCM 15483^T =KCTC5955^T) 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 Cad16^T is deposited under the GenBank/EMBL/DDBJ accession number AJ511274. The complete genome sequence of the type strain Cad16^T is deposited under the GenBank ID CP020370, CP020371 and CP020372. The type strain has been deposited both at the Japan Collection of Microorganisms (JCM 15483^T) and at the Korean Collection for Type Cultures (KCTC 5955^T).

Additional file

> **Additional file 1: Figure S1.** Phylogenetic placement of "T. syntrophicum" strain Cad16^T within the other 12 Chromatiaceae species with a publicly available whole genome sequences. Additionally, the closely related phylogenetic lineages *Nitrosococcus*, Rheinheimera and Arsukibacterium are also included. Strain Cad16^T 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-5′-phosphosulfate; BChl *a*: bacteriochlorophyll *a*; Cas: CRISPR assosiated; 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; PGC: photosynthesis gene cluster; PSB: purple sulfur bacteria; RC: reaction center; Rubis-CO: ribulose-1,5-bisphosphate carboxylase/oxygenase; SGB: sulfur globule; SMRT: single molecule real-time; SRB: sulfur reducing bacteria; T4P: type IV pilus; T6SS: type VI protein secretion system; TAT: twin-arginine translocation; TCA: tricarboxylic acid cycle; TRAP: tripartite ATP-independent periplasmic

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 (DACD), Laboratory of Applied Microbiology (LMA), Via Mirasole 22A, 6500 Bellinzona, Switzerland. [3]Federal Office for Civil Protection, Spiez Laboratory, Biology Division, Spiez, 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.

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/referenceworkentry/10.1007/978-3-642-38922-1_295.
2. Winogradsky S. Beiträge zur Morphologie und Physiologie der Bacterien. Heft 1. Zur Morphologie und Physiologie der Schwefelbacterien. Leipzig: Felix; 1888.
3. Peduzzi S, Storelli N, Welsh A, Peduzzi R, Hahn D, Perret X, et al. *Candidatus "Thiodictyon syntrophicum"*, sp. nov., a new purple sulfur bacterium isolated from the chemocline of Lake Cadagno forming aggregates and specific associations with *Desulfocapsa* 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, Degermendzhi AG, editors. Ecol Meromictic Lakes. Cham: Springer International Publishing; 2017 [cited 2017 Jun 16]. p. 155–86. Available from: http://link.springer.com/10.1007/978-3-319-49143-1_7
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, Florín M, Squires MM, Lehmann 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 N-U, Perret X, Tonolla M. CO_2 assimilation in the chemocline of Lake Cadagno is 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;45:29–37.
11. Eichler B, Pfennig N. A new purple sulfur bacterium 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/S0378111913014091

13. SharedProtocol-Extracting-DNA-usinig-Phenol-Chloroform.pdf [Internet]. [cited 2017 Jan 5]. Available from: http://www.pacb.com/wp-content/uploads/2015/09/SharedProtocol-Extracting-DNA-usinig-Phenol-Chloroform.pdf

14. Agencourt AMPure XP Bead Clean-up - NEBNext 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:W52–7.

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

17. TMHMM Server, v. 2.0 [Internet]. [cited 2017 Apr 6]. Available from: http://www.cbs.dtu.dk/services/TMHMM/

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. Corson 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. Verméglio A, Li J, Schoepp-Cothenet B, Pratt N, Knaff DB. The role of high-potential Iron protein and cytochrome c_8 as alternative Electron donors to the reaction Center of *Chromatium vinosum* [†]. Biochemistry (Mosc). 2002;41:8868–75.

21. Vogl K, Bryant DA. Elucidation of the biosynthetic pathway for Okenone in *Thiodictyon* 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: χ-ring formation. Geobiology. 2012;10:205–15.

23. Bauer CE, Buggy JJ, Yang ZM, Marrs BL. The superoperonal 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;228:433–44.

24. Nagashima KVP, Verméglio A, Fusada 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. 52:333–41.

26. Denkmann K, Grein F, Zigann R, Siemen A, Bergmann J, van Helmont S, et al. Thiosulfate dehydrogenase: a widespread unusual acidophilic c-type cytochrome. Environ Microbiol. 2012;14:2673–88.

27. Pattaragulwanit K, Brune DC, Trüper HG, Dahl C. Molecular genetic evidence for extracytoplasmic localization of sulfur globules in *Chromatium vinosum*. Arch Microbiol. 1998;169:434–44.

28. Frigaard N-U, Dahl C. Sulfur Metabolism in Phototrophic Sulfur Bacteria. In: Poole RK, editor. Adv. Microb. Physiol. [Internet]. Academic Press; 2008 [cited 2016 Mar 16]. p. 103–200. Available from: http://www.sciencedirect.com/science/article/pii/S0065291108000027

29. Pott AS, Dahl C. Sirohaem sulfite reductase and other proteins encoded by genes at the dsr locus of *Chromatium vinosum* are involved in the oxidation of intracellular sulfur. Microbiology. 1998;144:1881–94.

30. Grein F, Pereira IAC, Dahl C. Biochemical characterization of individual components of the *Allochromatium vinosum* DsrMKJOP transmembrane complex aids understanding of complex function in vivo. J Bacteriol. 2010; 192:6369–77.

31. Grimm F, Dobler N, Dahl C. Regulation of dsr genes encoding proteins responsible for the oxidation of stored sulfur in *Allochromatium vinosum*. Microbiology. 2010;156:764–73.

32. Dahl C, Franz B, Hensen D, Kesselheim A, Zigann R. Sulfite oxidation in the purple sulfur bacterium *Allochromatium vinosum*: identification of SoeABC as a major player and relevance of SoxYZ in the process. Microbiology. 2013;159:2626–38.

33. Hipp WM, Pott AS, Thum-Schmitz N, Faath I, Dahl C, Trüper HG. Towards the phylogeny of APS reductases and sirohaem sulfite reductases in sulfate-reducing and sulfur-oxidizing prokaryotes. Microbiology. 1997;143:2891–902.

34. Parey K, Demmer U, Warkentin E, Wynen A, Ermler U, Dahl C. Structural, biochemical and genetic characterization of dissimilatory ATP Sulfurylase from *Allochromatium vinosum*. PLoS One. 2013;8:e74707.

35. Meyer B, Kuever J. Homology modeling of dissimilatory APS reductases (AprBA) of sulfur-oxidizing and sulfate-reducing prokaryotes. PLoS One. 2008;3:e1514.

36. Tengölics R, Mészáros L, Győri E, Doffkay Z, Kovács KL, Rákhely G. Connection between the membrane electron transport system and Hyn hydrogenase in the purple sulfur bacterium, *Thiocapsa roseopersicina* BBS. Biochim Biophys Acta BBA - Bioenerg. 1837;2014:1691–8.

37. Laurinavichene T. The effect of sulfur compounds on H_2 evolution/consumption reactions, mediated by various hydrogenases, in the purple sulfur bacterium, *Thiocapsa roseopersicina* - Springer [Internet]. [cited 2016 Jun 20]. Available from: http://link.springer.com/article/10.1007%2Fs00203-007-0260-7

38. Weissgerber T, Sylvester M, Kröninger L, Dahl C. A comparative quantitative proteomic study identifies new proteins relevant for sulfur oxidation in the purple sulfur bacterium *Allochromatium vinosum*. Appl Environ Microbiol. 2014;80:2279–92.

39. Rákhely G, Laurinavichene TV, Tsygankov AA, Kovács KL. The role of Hox hydrogenase in the H_2 metabolism of *Thiocapsa roseopersicina*. Biochim Biophys Acta BBA - Bioenerg. 2007;1767:671–6.

40. Rákhely G, Kovács ÁT, Maróti G, Fodor BD, Csanádi G, Latinovics D, et al. Cyanobacterial-type, Heteropentameric, NAD^+-reducing NiFe hydrogenase in the purple sulfur photosynthetic bacterium *Thiocapsa roseopersicina*. Appl Environ Microbiol. 2004;70:722–8.

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

42. Weissgerber T, Zigann R, Bruce D, Chang Y-J, Detter JC, Han C, et al. Complete genome sequence of *Allochromatium vinosum* DSM 180(T). Stand Genomic Sci. 2011;5:311–30.

43. 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.

44. Storelli N, Saad MM, Frigaard N-U, Perret X, Tonolla M. Proteomic analysis of the purple sulfur bacterium Candidatus "Thiodictyon syntrophicum" strain Cad16T isolated from Lake Cadagno. EuPA Open Proteomics [Internet]. [cited 2013 Nov 28]; Available from: http://www.sciencedirect.com/science/article/pii/S2212968513000172

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

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

47. 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

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

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

50. Overmann J, Pfennig N. Bouyancy regulation and aggregate formation in *Amoebobacter purpureus* from Mahoney Lake. FEMS Microbiol Lett. 1992;101:67–79.

51. Egli K, Wiggli M, Fritz M, Klug J, Gerss 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.

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

53. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, et al. CRISPR provides acquired resistance against viruses in prokaryotes. Science. 2007;315:1709–12.

54. Mojica FJM, Díez-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.

55. Bolotin A, Quinquis B, Sorokin A, Ehrlich SD. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology. 2005;151:2551–61.

Luedin *et al. Standards in Genomic Sciences* (2018) 13:14

Page 13 of 13

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;32:268–74.
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, Baldauf S, Ballard S, Boore J, Cochrane G, Cole J, Dawyndt P, De Vos P, dePamphilis C, Edwards R, Faruque N, Feldman R, Gilbert J, Gilna P, Glöckner FO, Goldstein P, Guralnick R, Haft D, Hancock D, Hermjakob 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, Martiny 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, Tateno Y, Tett A, Turner S, Ussery 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.

2.1.1 Additional files

Figure S 2.1-1 | Phylogenetic placement of 'T. syntrophicum' strain Cad16ᵀ within the other 12 *Chromatiaceae* species with a publicly available whole genome sequences.

Additionally, the closely related phylogenetic lineages *Nitrosococcus, Rheinheimera* and *Arsukibacterium* 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.

2.2   Research Paper II

The complete genome sequence of *Chromatium okenii* strain LaCa, a purple sulfur bacterium with a turbulent life

Samuel M Luedin, Nicole Liechti, Raymond P Cox, Francesco Danza, Niels-Ulrik Frigaard, Nicole R. Posth, Joël F Pothier, Samuele Roman, Nicola Storelli, Matthias Wittwer and Mauro Tonolla

Statement of contribution

I performed DNA extraction, sequence assembly and annotation and contributed to sampling, culture enrichment, data interpretation and manuscript preparation.

Research objective

Chromatium okenii was described in the Lake Cadagno chemocline as early as 1918 and in situ studies on light harvesting kinetics [157], on microbial inorganic nitrogen and carbon fixation [163] and bioconvection [148] have highlighted the fundamental role of *C. okenii* for the chemocline ecosystem. We therefore sequenced the complete genome from *C. okenii* str. LaCa enrichment and compared it to other sequenced PSOB isolates from Lake Cadagno in order to better understand the biological functions encoded in the context of the chemocline community.

2.2.1 Abstract

Purple sulfur bacteria (PSB) are phototrophic γ-proteobacteria oxidizing reduced sulfur in intertidal flats and stagnant water bodies. *Chromatium okenii* was first described in 1838 and has since been found in many stratified sulfidic environments worldwide. In this study, the first complete genome of a large celled PSB of the genus *Chromatium* sp. is described in detail. Single-Molecule Real-Time (SMRT) sequencing was used to sequence the *C. okenii* strain LaCa genome from cells enriched from water samples of the permanently stratified alpine Lake Cadagno in the Swiss Alps. The 3.78 Mb genome contains 3'016 protein coding genes and 67 RNA genes. Our findings on a genomic level confirm previous studies on the phototrophic and chemoautotrophic metabolism of *C. okenii* and the putative biological functions are discussed from an ecosystem perspective. The genome was further compared to complete sequences of other phototrophic sulfur bacteria isolated from the same habitat using core genome analysis. Genes involved in flagellar movement, chemotaxis and encoding S-layer-proteins were relatively overrepresented in strain LaCa. The phenotypically described rapid response to changing external factors and the S-layer acting as a putative defense against bacterial predation and phage infection are reflected in the assembly. The complete genome of *C. okenii* thereby provides one basis for further functional studies on bioconvection and host-predator interaction.

2.2.2 Introduction

Chromatium okenii belonging to the purple sulfur bacteria (PSB, family *Chromatiaceae*), was already described in the environment as blooms in 1836 by the pioneering microbiologists Ehrenberg and Weisse [40] as: "*Monas corpore cylindrico, aequabili, parumpcr curvato, ter quaterve longiore quam lato, utrinque rotundato, 1/192 lineae attingens, volutando procedens, vacillans, rubra; socialis*", and was later examined in more detail by Maximilian Perty in 1852 [254] and Sergei Winogradsky in 1888 [255]. Early microscopic studies were focused on bacterial behavior under variable light and sulfur availability [256, 257]. *C. okenii* showed scotophobotaxis – i.e. the immediate change of direction of an organism when experiencing a decrease in light intensity over time – negative aerotaxis and positive chemotaxis towards H_2S *in vitro* [258–260]. Different populations of *C. okenii* have since been described in freshwater ecosystems worldwide such as lakes, ponds and bacterial mats

[261–265]. A strain of *C. okenii* was isolated in 1960 by Schlegel and Pfennig from a pond in Germany [266] however the culture is not available anymore. The cells are rod-shaped, 5.0 to 6.5 *μ*m thick and 8.0 to 10.0 *μ*m long, stain Gram-negative, are motile through flagella and contain okenone and bacteriochlorophyll *a* (BChl *a*) as the main photosynthesis pigments (Figure 2.2-1a) [267]. Different intracellular storage compounds such as polyhydroxybutyrate (PHB), glucose, polyphosphate and elemental sulfur that function as carbon and or electron donors in PSB, respectively, have been described for *C. okenii* [268].
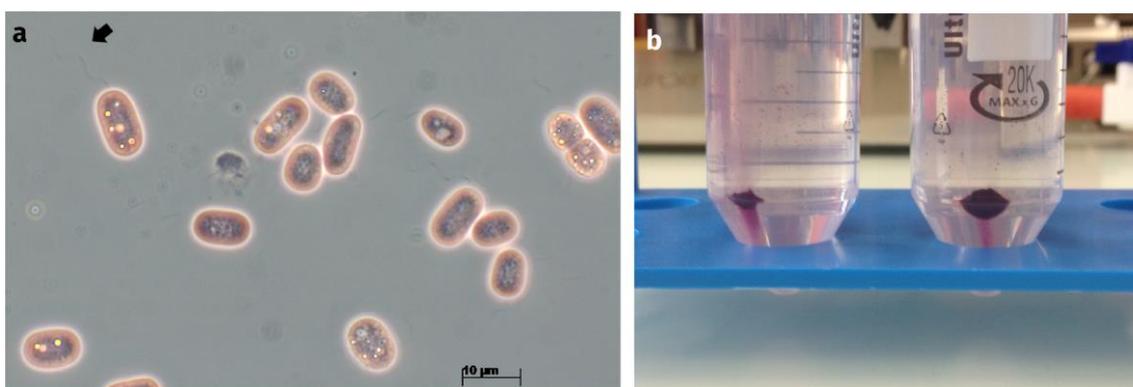
Figure 2.2-1 | Morphology and Macroscopic appearance of *Chromatium okenii* str. LaCa.
a) Microscopic image of *C. okenii* str. LaCa cells. Intracellular sulfur globules are visible as yellow, highly-refractive spheres. The polar flagellar tuft is visible (black arrow). b) *C. okenii* cell pellets enriched from a water sample from Lake Cadagno.

Lake Cadagno is a crenogenic, meromictic alpine lake that harbors a dense population of anoxygenic phototrophic sulfur bacteria of the families *Chromatiaceae* such as *Chromatium*, *Thiodictyon*, *Lamprocystis* and *Thiocystis* spp. and *Chlorobiaceae* (green sulfur bacteria; GSB) of the *Chlorobium* genus. This heterogeneous community of up to 10^7 cells ml^{-1} grows by anoxygenic photosynthesis with around 0.2 mM sulfide present at the chemocline at 12 m depth [118, 122]. The presence of an ancestral *Chromatium* sp. has also been shown in a 16S rRNA-gene study from sediments of Lake Cadagno for the past 9'450 years [168] and for the last 20 years, a single *Chromatium okenii* strain has been described for the chemocline with fluorescent in situ hybridization (FISH) and 16S rRNA gene analysis [153, 155, 134]. Whereas in the chemocline of Lake Cadagno, *C. okenii* represents only 1-10% of all bacterial cells and 22–83% of the phototrophic community [157, 161, 155, 136, 134], it comprises to 72% of the total biovolume [161]. *C. okenii* was found to assimilate up to 70% of total carbon and 40% of ammonium during the day, with a doubling time of 5 to 7 days [162, 163]. These findings were supported by a carbon isotope fractionation assay that attributed 36–52% of the total bulk δ^{13}C-signal to *C. okenii* in October and June, respectively [269]. *C. okenii* may also represents a major food source for zooplankton in Lake Cadagno, as grazing on *C. okenii* by the ciliate *Trimyema compressum* was demonstrated *in vitro* [270]. The population of *C.*

okenii varies with season. Regular cell concentration monitoring with FISH and flow cytometry revealed a seasonal pattern with dominance of *C. okenii* over small celled PSB from July to September and a rapid population decline within two weeks in October in Lake Cadagno [155, 161]. However, due to a mixing event in 2000 and the following massive bloom of GSB *C. clathratiforme*, less predictable *C. okenii* population dynamics have been described between the years 2000–2005 [154], suggesting that environmental influences may have a long lasting impact on microbial community composition [154].

The importance of *C. okenii* in bioconvection in Lake Cadagno has been discussed theoretically [147]. Interestingly, a spatial and temporal association of convection in zones with high concentrations of *C. okenii* (10^5– 10^6 cells ml^{-1}) was later inferred *in situ* [148]. Additionally, a functional study has recently shown the short time dynamics in sulfide uptake of *C. okenii* and putative interactions between *C. okenii* and the GSB *Chlorobium phaeobacteroides* [134].

This study provides the first complete and annotated genome for a member of the large celled PSB genus *Chromatium*, *C. okenii* str. LaCa, enriched from Lake Cadagno. Moreover, the availability of sequenced PSB and GSB isolates of the Lake Cadagno allowed us a core genome comparison and to elucidate on strain specific biological functions. The enrichment and the complete genome of *C. okenii* str. LaCa are essential components towards the better understanding of the nutrient fluxes and interactions within this highly balanced microbial ecosystem of the Lake Cadagno chemocline.

2.2.3 Material and Methods

Chemicals

All chemicals were purchased from Sigma-Aldrich AG (Buchs, Switzerland), if not further specified.

Enrichment

Physiochemical measurements on Lake Cadagno were made with an YSI 6000 profiler (Yellow Springs Inc., Yellow Springs OH, USA) on the 14th of July 2016. In order to

understand carbon isotope fractionation of PSB and GSB strains, *C. okenii* was previously enriched to high purity by sedimentation and dilution, however cultivation was not established [269]. We used a comparable approach for this study. Samples were collected with a 1 l *Ruttner*-type sampling bottle (Hydrobios Apparatebau GmbH, Germany) taken at depths with maximum turbidity and rapid changes in redox-potential between 11.6–12.0 m depth, indicating a dense bacterial population at the chemocline, as previously described [147]. The bottles were brought immediately to the lab and were placed at natural illumination (2'000 lux PAR) at 16 °C for 6 hours. The purple precipitates thus obtained (Figure 2.2-1b) were identified by cell morphology with light microscopy as *Chromatium* sp. based on previous descriptions in the literature (e.g. [271]). Other bacteria were also present, however only in low numbers (<10%). Cells were collected with a 10 ml pipette and transferred to 50 ml tubes. The cells were then centrifuged 10 min at 15 *g* at room temperature (RT). The supernatant was carefully discarded and the residual 10 ml were collected and combined in 100 ml serum bottles. The bottles were then filled up with filtered (0.45 μm) chemocline water. A rubber plug was applied and a vacuum was generated with a suction pump to remove O_2. Using a syringe, 100 μl of a 35mM Na_2S $9 \cdot H_2O$ solution was added to each sample, resulting in a final Na_2S $9 \cdot H_2O$ concentration of 0.03 mM Sub-samples were frozen at -20 °C for DNA extraction.

DNA Extraction, Sequencing and Genomic Analysis

Frozen samples were thawed on ice and cells were collected by centrifugation for 15 min at 10'600 *g*. Genomic DNA (gDNA) was extracted with phenol/chloroform/isoamylalcohol solution (25:24:1, *v/v*) adhering to the protocol provided by *Pacific Biosciences* in combination with phase lock gels for phase separation (VWR International, Radnor, USA). gDNA was concentrated and washed using *AMPure* beads (Agencourt, Beckman Coulter Life Sciences, Indianapolis, USA) following the E2612 protocol from New England Biolabs [272]. Concentration of the DNA was assessed using the *Qubit™* UV/VIS absorption reader (Thermo Fisher Scientific, Rheinach, Switzerland).

The library construction and Single-Molecule Real-Time sequencing (SMRT) 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 *SMRTbell*-Polymerase Complex with P6-C4 chemistry (Pacific Biosciences) according to the manufacturer instructions. Two *SMRT* cells v3.0 (Pacific Biosciences) for *PacBio RS II* chemistry were used for sequencing. Sequencing quality reports were created through the *SMRT* portal software.

PacBio RSII high quality reads were assembled using the *canu* assembler v1.4, [RRID:SCR_015880] [273]. Genes were annotated using the NCBI Prokaryotic Genome Annotation Pipeline *GeneMarkS+*, v4.3 [RRID:SCR_011930]. *PhiSpy* v2.3 [274] and *VIRSorter* v1.0.3 [275] were used to detect phage and prophage related sequences. *EggNOG* [RRID:SCR_002456] [276] was used to classify the predicted genes into COG-categories and *OrthoVenn* [277] was used to classify gene families and visualize clustering. Genes were classified with the *blastKOALA* tool to *KEGG* categories [RRID:SCR_012773] [278]. *AmphoraNet* [RRID:SCR_005009] [279] and *BUSCO* [RRID:SCR_015008] [280] were used to asses genome completeness and contamination.

Phylogenetic Analysis

Roary [281] was used to compare the core genomes of sequenced *Chromatiaceae*. Out of this dataset, 100 single-copy orthologues were selected randomly and their sequences aligned with MUSCLE [282]. The best-fit phylogenetic model and subsequent consensus tree estimation, based on maximum-likelihood and 1'000 bootstrap iterations, was performed with the *IQ-TREE* platform [283].

The *C. okenii* str. LaCa full length 16S rRNA gene sequence (CXB77_RS15475) was used to search 16S gene NCBI database for related sequences with BLASTn [RRID:SCR_001598]. The online tool IQtree [284] was used to create the phylogenetic trees based on the alignment with MAFFT v7. 215 [285] [RRID:SCR_011811]. A combination of 1'000 bootstrap iterations and 1'000 aLRT replications were performed. FigTree v1.4.3 [286] was used to render phylogenetic trees.

Data availability

The complete, corrected and annotated genomic data is available at NCBI under the GenBank assembly accession: GCA_002958735.1 and RefSeq assembly accession: GCF_002958735.1.

2.2.4   Results and Discussion

Genome Structure and Phylogeny

The *de novo* sequencing of an enrichment culture of *C. okenii* str. LaCa was successfully done with a PacBio RSII system using two SMRTcells. A total of 45 contigs was obtained with a total length of 3'784,749 bp, a N50 of 448'938 bp and a L50 of 3. The GC content was found to be 49.8% (Table 2.2-1). Details on the sequencing output can be found under supplementary Figure S 2.2-1 a. Three long contigs (PPGH01000034.1, PPGH01000035.1 and PPGH01000037.1) with a coverage between 24× and 27× possibly form the chromosome (suppl. Figure S 2.2-1b). Further 3 contigs (PPGH01000013.1, PPGH01000018.1 and PPGH01000038.1) were associated and showed an average coverage of 25× (22–27×). Additional 4 putative plasmids (PPGH01000033.1, PPGH01000043.1, PPGH01000024.1 and PPGH01000027.1) were identified (suppl. b). The other 35 shorter contigs with coverage < 22× showed partial or complete overlap with other contigs. The genome was considered complete due to the high number of single copy core genes (*BUSCO*; 129 complete /single copy genes, *amphoraNet*; 40 genes homologous to *A. vinosum*) and acomplete set of tRNA genes and no contaminating sequences. COG classification of the 3'016 protein coding genes resulted in 2'022 assigned proteins (

Table 2.2-2).

Table 2.2-1| Genome statistics for *Chromatium okenii* str. LaCa

Attribute	Value	% age of Total
Genome size (bp)	3'784'749	100.0
DNA coding (bp)	2'686'967	71.0
DNA G+C (bp)	1'884'805	49.8
DNA scaffolds	45	100.0
Total genes	3'792	100.0
Protein coding genes	3'016	79.5
RNA genes	67	1.8
rRNA genes	3, 3, 3 (5S, 16S, 23S)	0.2
tRNA genes	53	1.4
ncRNA genes	5	0.1
Pseudo genes	708	18.7
Genes in internal clusters	NA	NA
Genes with function prediction	1'726	45.5
Genes assigned to COGs	2'291	60.4
Genes KEGG predictions	1'335	35.2
Genes with Pfam domains	2'274	60.0

Genes with signal peptides	163	4.3
Genes with transmembrane helices	556	14.7
CRISPR repeats	2	0.1

Table 2.2-2| Clusters of Orthologous Genes (COG) functional categories of *C. okenii* str. LaCa.

Code	Value	% age	Description
J	132	4.38	Translation, ribosomal structure and biogenesis
A	2	0.07	RNA processing and modification
K	67	2.22	Transcription
L	182	6.03	Replication, recombination and repair
B	1	0.03	Chromatin structure and dynamics
D	39	1.29	Cell cycle control, Cell division, chromosome partitioning
Y	0	0.00	Nuclear structure
V	61	2.02	Defense mechanisms
T	232	7.69	Signal transduction mechanisms
M	117	3.88	Cell wall/membrane biogenesis
N	51	1.69	Cell motility
Z	0	0.00	Cytoskeleton
W	0	0.00	Extracellular Structures
U	60	1.99	Intracellular trafficking and secretion
O	82	2.72	Posttranslational modification, protein turnover, chaperones
X	0	0.00	Energy production and conversion
C	120	3.98	Energy production and conversion
G	55	1.82	Carbohydrate transport and metabolism
E	75	2.49	Amino acid transport and metabolism
F	39	1.29	Nucleotide transport and metabolism
H	74	2.45	Coenzyme transport and metabolism
I	31	1.03	Lipid transport and metabolism
P	63	2.09	Inorganic ion transport and metabolism
Q	30	0.99	Secondary metabolites biosynthesis, transport and catabolism
R	0	0.00	General function prediction only
S	569	18.87	Function unknown
Multi COG	59	1.96	Multiple COG assignments
Single COG	1'963	65.09	single COG assignments
No COG	994	32.96	Not in COGs

Phylogenetic 16S rRNA gene analysis revealed a 99 % sequence identity with *C. okenii* DSM 169[T] and *C. okenii* str. LaCa groups with *Allochromatium* and *Thiocystis* spp. (Figure 2.2-2). When alternatively comparing a subset of 100 core genes, *C. okenii* is most similar to *T. violascens* and *A. vinosum* (supplementary Figure S2.2-2). The *C. okenii* DSM 169[T] type strain or other *Chromatium* spp. are not available anymore in repositories [287]. Therefore a more detailed genomic comparison within the genus *Chromatium* was not possible.
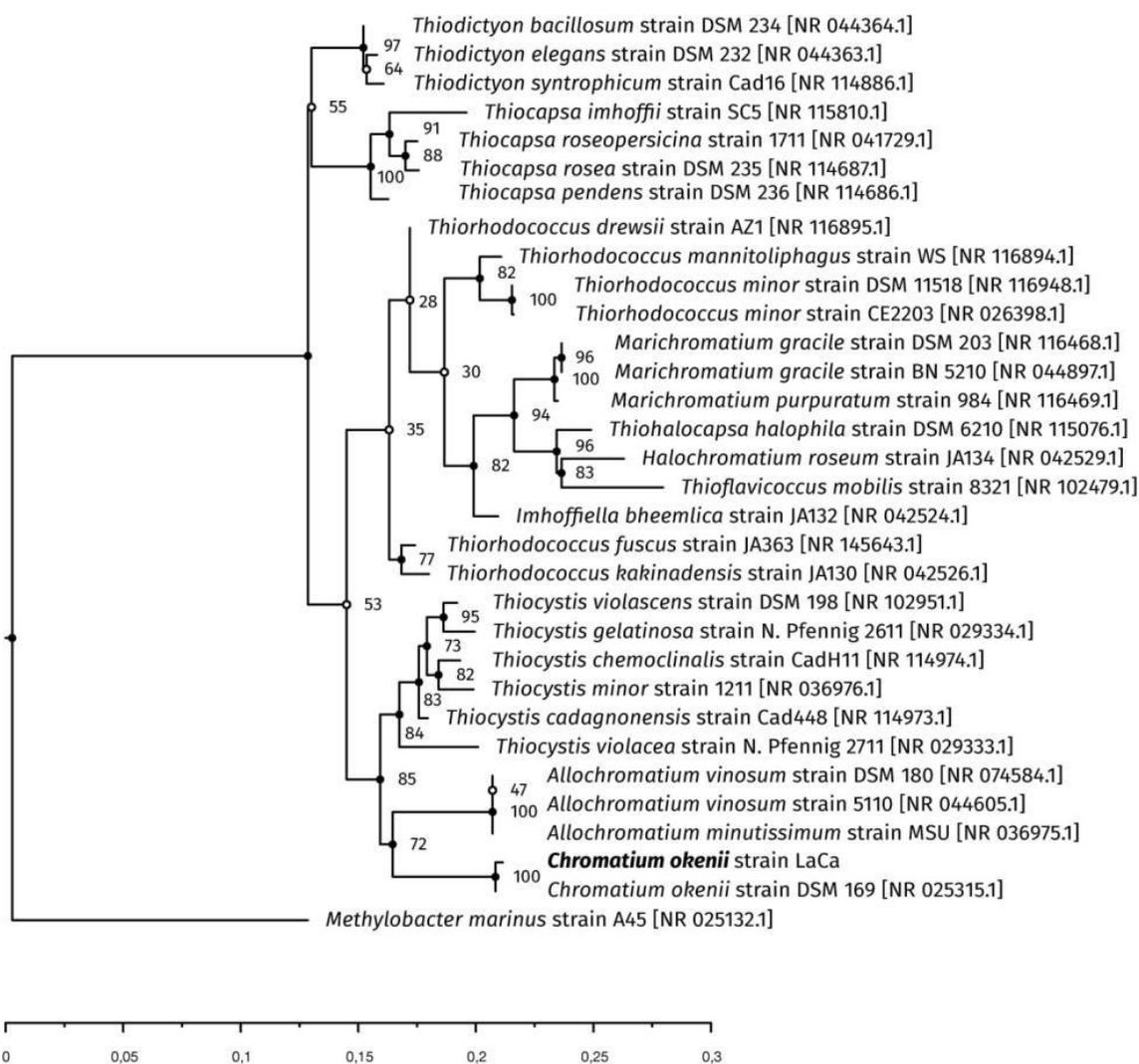
Figure 2.2-2 | Phylogenetic relationship of *C. okenii* str. LaCa based on 16S rRNA gene sequences.
IQTree **[283]** was used to calculate a consensus tree combining 1'000 bootstrap iterations and 1'000 aLRT replications. Branch lengths were optimized by maximum likelihood on original alignment. Full node circles indicate bootstrap support above 60%. Scale bar denotes genomic distance in base-pair (bp) substitutions per 100 bp.

Genome Features

The encoded metabolic pathways and structures will be discussed in detail below.

Photosynthesis and Chemotrophy

An extended system of photosynthetic membranes has been described for *C. okenii* [288]. Within these membranes, PSB employ type II reaction centres (RCs) to transform light energy into chemical energy. In the strain LaCa genome a canonical photosystem II-type RC is encoded in two clusters, containing two *pufAB* (CXB77_RS07530, CXB77_RS07535,

CXB77_RS07560 and CXB77_RS07565), *pufL* (CXB77_RS07555) and *pufM* (CXB77_RS07550) genes, a RC complex subunit H *puhA* (CXB77_RS09135) and a putative photosynthetic complex assembly protein (CXB77_RS09130). Further light harvesting complex LHC I and LHC II genes (CXB77_RS02170 and CXB77_RS02175) and two additional pairs of *pufAB* genes were predicted (CXB77_RS10755–CXB77_RS10765). Multiple copies of LHC polypeptides are thought to be an adaptation of phototrophic organisms to changes in light availability [7] and were also found in 'Thiodictyon syntrophicum' str. Cad16[T] and *A. vinosum* DSM180[T][289]. Following from this, low light adaptation of *C. okenii* has been described by measuring fluorescence kinetics *in situ* [158] and quantum yields below the optimum have been reported [149].

Light is taken up efficiently by photosynthetic pigments and the energy gained is then further transferred to the RC in femtosecond transfers in PSB. The carotenoid okenone [290] and BChl *a* [266] are synthesized in *C. okenii*. Complete sets of genes encoding for BChl *a* synthesis (CXB77_RS09140–CXB77_RS09180) and of the carotenoid okenone (*crt* and *cru*) were detected. Notably, a carotenoid 3,4-desaturase *crtD*-homologue of the C-4/4' ketolase *cruO*-type (CXB77_RS02160) [291] was identified next to the *crtC* hydroxyneurosporene synthase (CXB77_RS02155) homologous to *MariChromatium purpuratum* DSM 1591[T]. *C. okenii* showed a stable BChl *a* to protei ratio over 3 months sampling *in situ* [157]. The BChl *a* dark synthesis rate was thereby independent of sampling depth, however small changes in light intensity (0.06 mol quanta m^{-2} h^{-1} in average) had an impact on BChl *a* synthesis rates [157]. Taken together, may adapt to changes in light availability and quality by modulating energy uptake efficiency by different combinations of LHC antenna proteins and pigments concentrations, respectively.

During photosynthesis, soluble electron-carrier cytochromes enable cyclic electron flow by transporting electrons from the cytochrome *bc*1 complex back to the RCs. In strain LaCa the high potential iron sulfur protein (HiPIP; CXB77_RS02565) might function as the main high potential cytochrome as in *A. vinosum* DSM180[T] [292]. The cytochrome *c*551/*c*552 (CXB77_RS07885) may function as an additional RC reductant under strictly autotrophic conditions [49]. Other soluble electron carriers present were a cytochrome *c'* (CXB77_RS15975), a cytochrome *c*4 (CXB77_RS04500) homologous to *Thiocapsa roseopersicina* and two soluble *c*-type cytochromes (CXB77_RS01960–CXB77_RS01970 and CXB77_04510) similar to *A. vinosum* DSM180[T].

The greatly diminishes in numbers in the chemocline of Lake Cadagno during winter at low light availability of >0.4 μmol quanta·m^{-2} s^{-1} [155]. In other stratified lakes, motile *Chromatium* sp. were detected in viable non-dividing states below the chemocline [293] and survival over 1.5 year in the dark was observed [294]. Interestingly, upward swimming of *C. okenii* in dark conditions has been inferred by observation of bioturbation at night in summer [147] and attachment of cells at the underside of sediment traps in spring in Lake Cadagno [161]. These findings indicate the importance of dark metabolism for *C. okenii* persistence. Over the season, a minimum oxygen concentration in spring [155] and a lower oxycline due to thermal mixing of the mixolimnion have been observed from October to December [122, 269]. Furthermore, locally produced oxygen (< 20 nmol l^{-1}) by oxygenic photosynthesis has been reported in summer [137]. Together, these observations define low concentrations of oxygen with the monimolimnion throughout the year. Whereas aerobic sulfur oxidation yields about 25–30% of the energy of anaerobic photosynthesis [295], it still might be important for long time survival and provide additional energy at the micro-oxic upper layer in summer. In accordance, a complete respiratory chain was found in *C. okenii* str. LaCa. Encoded are a NADH-quinone oxidoreductase (CXB77_RS02240–CXB77_RS02300 and CXB77_RS02310), a succinate-dehydrogenase (CXB77_RS02325–CXB77_RS02335 and CXB77_RS09655), as well as a multi-subunit terminal cytochrome *bd* oxidase (CXB77_RS12155 and putatively CXB77_RS12145 and CBX77_RS12150) both involved, in the photosynthetic and respiratory electron chain. Taken together, the encoded metabolism in *C. okenii* str. LaCa may enable for both, low level chemotrophic substrate respiration and anoxygenic photosynthesis in Lake Cadagno, as generally suggested for PSB by Kämpf and Pfennig [79].

In contrast, chemotrophic incubations at room temperature with two strains of *C. okenii* under a 5% O$_2$ atmosphere showed no significant growth after 5 days in darkness [79]. Additionally, no correlation between light availability and specific dark fixation rates has been observed *in situ* in a chemocline population dominated by *C. okenii* [149]. Consequently, the experimentally described metabolism and the biological functions encoded in the genome of *C. okenii* do not readily explain the high dark total fixation rates found in Lake Cadagno [128, 162].

Sulfur Metabolism

Reduced sulfur compounds such as H_2S, S^0 and SO_3^{2-} are used as reductants by *C. okenii* during photolitho-autotrophic growth [296]. Light energy is then used to transfer the electrons to $NAD(P)^+$ and ferredoxin for subsequent CO_2 fixation. *C. okenii* str. LaCa encodes a flavocytochrome *c* (FccAB; CXB77_RS06380) and the sulfide:quinone oxidoreductases (SqrD and SqrF; CXB77_RS06755 and CXB77_RS12425) for initial H_2S oxidation in the periplasm to form sulfur globules (SGBs) containing S^0. SGBs are surrounded by sulfur globule proteins (SGPs) that are folded into structures resembling collagen [297]. Two putative SgpA copies with N-terminal signal peptides were found (CXB77_RS07855 and CXB77_RS14820). This is important, as a homologue SgpA is essential to build intact sulfur globules in *A. vinosum* [298]. Furthermore, the canonical dissimilatory sulfite oxidation pathway (Dsr) that enables sulfite production in the cytoplasm [39] is also found completely conserved in one cluster (CXB77_RS03215–CXB77_RS03270) in *C. okenii* and shows gene synteny to other *Chromatiaceae*. Interestingly, two *arsR* family transcriptional regulator genes (CXB77_RS06260 and CXB77_RS07240) possibly involved in H_2S-dependent gene regulation [299] were detected. Furthermore, the trimeric adenylylsulfate reductase alpha and beta-subunits AprAB (CXB77_RS17245 and CXB77_RS17240) that is anchored by the CoB--CoM heterodisulfide reductase multi subunit complex (CXB77_RS04305–CXB77_RS04320) were found in strain LaCa. To complete sulfur oxidation, a Sat sulfate adenylyltransferase (CXB77_RS09675) and the dissimilatory-type SoeABC type enzyme (CXB77_RS11845–CXB77_RS11855) are encoded. An additional cluster of sulfur carrier proteins TusA (XB77_RS15940) and DsrE2 (CXB77_RS15945) [65] putatively involved in sulfur oxidation was detected. Furthermore, a cytochrome *b*561 (CXB77_RS01235) and an octaheme cytochrome *c* (CXB77_RS01240) homologues to *A. vinosum* DSM180[T] was found. Both proteins are conserved among PSB and have been upregulated in *A. vinosum* DSM180[T] with sulfide as sole electron donor [77]. Notably, no genes encoding Sox proteins necessary for thiosulfate ($S_2O_3^{2-}$) oxidation were detected [48], which is in accordance with previous experimental results [300]. Moreover, no genes of the adenylyl-sulfate kinase Cys-pathway for assimilatory sulfate reduction were found, confirming previous experimental findings that no sulfate uptake is observed for *C. okenii* [300]. Finally, hydrogenases were not predicted in the genome of *C. okenii* that excludes H_2 as a source of electrons [296, 300].

Nitrogen and Phosphate Assimilation

In the *C. okenii* str. LaCa genome *nif* genes, putatively involved in nitrogen fixation, are found spread throughout the genome as in *A. vinosum* DSM180[T] [227]. The genes for the dimeric nitrogenase molybdenum-iron protein *nifDK* (CXB77_RS12525 and CXB77_RS12530) and the nitrogenase iron protein *nifH* (CXB77_RS12535) indicate a diazotrophic metabolism. Transcriptional regulation of N_2-uptake is under the control of the two-component sensor histidine kinases NtrX and NtrY (CXB77_RS03520/ CXB77_RS03525), the nitrogen regulatory protein P-II (CXB77_RS11185) and of *nifA* (CXB77_RS10450) and the oxygen sensor *nifL* (CXB77_RS10445). Both polyphosphate kinase and exopolyphosphatase, were also present in the genome of this strain, however polyphosphate accumulation was not detected *in situ* [160]. Moreover, the genome of *C. okenii* str. LaCa revealed also the presence of genes for ammonium assimilation, glutamate synthase and glutamine synthetase. Whereas NH_4^+-consumption of *C. okenii* was measured *in situ* [163] and modelled for the chemocline [136], other N-uptake mechanisms have still to be elucidated.

Carbon Metabolism

Carbon fixation kinetics driven by photosynthesis was also studied in *C. okenii*. The Calvin-Benson-Bassham-cycle (CBB) is the central carbon fixation pathway for PSB [300–302] and for *C. okenii*, a complete Calvin-Benson-Bassam cycle (CBB) with the one *cbbM* ribulose 1,5-biphosphate carboxylase/oxygenase (RuBisCO) form II (CXB77_RS09535), two regulatory genes *cbbQ* (CXB77_RS09540) and *cbbO* (CXB77_RS09550), and phosphoribulokinase PrkB (CXB77_RS15420) is encoded in the genome. Interestingly, also a putative RbcL RuBisCO-like protein (CXB77_RS07520) is present. The RbcL is possibly involved in methionine salvage as it is known for the purple bacteria *Rhodopseudomonas palustris* [303, 304], or in stress response or sulfur metabolism as found in both the GSB *Chlorobium tepidum* [305] and the purple bacteria *Rhodospirillum rubrum* [306]. In contrast to other small celled PSB isolated from Lake Cadagno such as "T. syntrophicum" str. Cad16[T] and *Lamprocystis* spp., no hypothetical carboxysome-like components and RuBisCO form I (*cbbL* and *cbbS*) were detected in the *C. okenii* str. LaCa genome.

Different carbon storage mechanism have been described for PSB [307, 308] that may function as energy and reductant reserves. Glycogen storage in *C. okenii* str. LaCa is mediated through glucose-1-phosphate adenylyltransferase and 1,4-alpha-glucan (glycogen) branching enzyme (CXB77_RS16905). Additionally, a complete tricarboxylic acid (TCA) cycle and enzymes for glycolysis were encoded in the genome. Furthermore, storage of carbon as PHB under nitrogen limitation was described *in vitro* [160] and a high average C:N ratio of 14.8 for *C. okenii*,that potentially could induce carbon storage mechanisms has been previously reported [163]. Consequently, genes PhaC (CXB77_RS16475) and PhaE (CXB77_RS16480) involved in PHB synthesis and de-polymerization are present. It is known that *C. okenii* oxidizes glycogen to PHB and that stored sulfur is used as an electron sink by reduction into H_2S [301]. Additionally, the putatively aerobic oxidation of sulfur was found to be favoured over glucose oxidation to acetate and CO_2 for *C. okenii* under *in situ* dark conditions [160]. However, storage compounds were depleted within hours and PHB inclusions were not observed under *in situ* conditions in Lake Cadagno [160], obscuring the role of PHB for long time survival of *C.okenii*.

Membrane Transport and Bacterial S-layer

Similar to other PSB species such as *A vinosum* DSM 180[T] or "*T. syntrophicum*" str. Cad16[T], the genome of *C. okenii* str. LaCa contains both a Type IV pilus (CXB77_RS13225, CXB77_RS13640–CXB77_RS13660 and CXB77_RS13745–CXB77_RS13760) and a Type VI secretion system (CXB77_RS0616–CXB77_RS06190 and CXB77_RS12705–CXB77_RS12755). Additionally, a general secretion (Sec) and twin-arginine translocation (Tat) systems is encoded. Furthermore, ABC-type transporters for di-, oligopeptide, lipoprotein, phosphate and molybdenum uptake, as well as Tol and TRAP and Co^{2+}, $Mg^{2+}$and Ni^{2+}-uptake systems are present.

The primary function of surface layer (S-layer) is to stabilize bacterial cells against mechanical, thermal and osmotic stress [309]. Moreover, the S-layer has also been speculated to protect against bacterial predation and bacteriophage infection [309, 310]. In *C. okenii*, the S-layer consists of conical shaped hexagonal lattice components with a diameter of 13 nm that are regularly spaced by 19 nm which extend 25 nm from the surface [288]. In accordance, 2 putative exported S-layer proteins (CXB77_RS09990 and CXB77_RS09995) and a FhaB-like protein (CXB77_RS10005) similar to alkaline phosphatases in *Microcystis* ssp were found. The S-layer proteins might be secreted through a Type I secretion homologues SapDEF

system (CXB77_RS09940–CXB77_RS09950) [309]. We further find a putative SapC protein (CXB77_RS08915) missing a signal peptide homologues to *Halorhodospira halochloris* (HH1059_1773).

The *C. okenii* population was observed to diminish rapidly in Lake Cadagno within a period of days in October [155]. The increase in *C. okenii* cells over the summer months possibly leads to metabolic stress and an increased sedimentation rate could lead to conditions of high predator pressure. Epibionts were reported for *C. okenii* in Lake Cadagno [163] and for other large-celled *Chromatium* species [311]. These epibionts are described as scavengers that feed on inactive *Chromatium* cells [312, 313] and may lead to this population collapse [314]. However, no sequences related to *Bdellovibrio, Daptobacter* or *Vampirococcus*-type bacterial scavengers were detected in Lake Cadagno chemocline enrichment samples. While this data is currently unavailable, we expect to find more epibionts in non-viable, sedimented cells in samples from the monimolimnion. The importance of bacteriophages for aquatic microbial communities has been recognized [315, 316] however only few studies have focused on stratified systems [317–319]. In this study, several putative prophage and phage sequences are present in the *C. okenii* str. LaCa sequence (suppl. Table S 2.2-1). Additionally, eleven Rha-type phage regulatory proteins were detected. The putative phage sequences and phage-related genes indicate the occurrence of phages targeting the chemocline community. Therefore, the function of the S-layer against phage attachment and the antagonistic Type 6 secretion system in defense against predation has to be further studied.

Flagellum and Chemotaxis

C. okenii is motile through a tuft of around 40 lophotrichous flagella with a length of 20 to 30 μm [258, 288]. The direction of movement is defined by the action of either pulling or pushing flagella that rotate clockwise or counter clockwise, respectively [320]. A complete set of genes encoding the basal body, hook and filament were found in one cluster. Additional genes *flg, flh*, *che* and *fli* were grouped together with *motA* and *motB* genes. A histidine–aspartate phosphorelay (HAP) based system [321] that includes chemotaxis genes *cheABRWYZ* and in total 31 putative chemoreceptor genes (MCP: methyl-accepting chemotaxis protein) of the TAP or TLPA family were found. Notably, a putative *aer* aerotaxis sensor receptor protein (CXB77_RS12890), a bacteriophytochrome (CXB77_RS05740) and two putative blue-light-activated histidine kinases (CXB77_RS09475 and CXB77_RS08785) were identified.

Furthermore, a putative circadian input kinase *A* (CXB77_RS08775) was detected, however no complete *kaiABC* regulatory relay was found. Also only an incomplete set of genes involved in acyl homoserine lactone (AHL) mediated quorum sensing was detected with components of the SagS-HptB-HsbR (swarming activity and biofilm formation) two-component regulatory system/cAMP/Vfr signaling (CXB77_RS11060 and CXB77_RS08790) and a putative transcriptional activator protein LasR (CXB77_RS11545).

Overmann and Pichel-Garcia speculate that motile PSB have an advantage over PSB with gas vacuole in conditions with light intensities above 0.2 μmol quanta m^{-2} s^{-1} [7]. In Lake Cadagno light intensities between 5.8 μmol quanta·m^{-2} s^{-1} [155] and between 35 μmol quanta m^{-2} s^{-1} [157] were measured at the upper border of the bacterial layer during summer, whereas a ten-fold decrease in light intensity within the bacterial layer at the chemocline is described [122]. Moreover, an inverse correlation between available light and thickness of the bacterial plume has been observed in Lake Cadagno [157]. *In vitro* the swimming velocity of *Chromatium minus* seems to be determined by external sulfide concentration and light intensity[322]. Interestingly, for *C. minus* both swimming speed and run time is relatively higher and longer under low light intensity, than under high light-intensity, a phenomena that last over hours [322]. With an observed swimming speed of $(2.7 \pm 1.4) \times 10^{-5}$ m s^{-1} [148, 323] (0.97 m h^{-1}) the chemocline can be crossed within minutes to hours. Importantly, these accumulations of motile, dense cells may even induce bioturbation [148]. Together, these findings indicate that *C. okenii* would benefit from actively moving to the upper boarder of the chemocline under non light-limiting conditions via control of scotophobotaxis, and negative O_2 and positive H_2S chemotaxis, respectively. However, temporal vertical mobility patterns of *C. okenii* were described as diel [262, 324, 325] or stochastic [147, 148] and the intervened signalling pathways that controls and coordinates swimming in *C. okenii* have to be further studied. A graphical summary of the features discussed above is given by Figure 2.2-3.
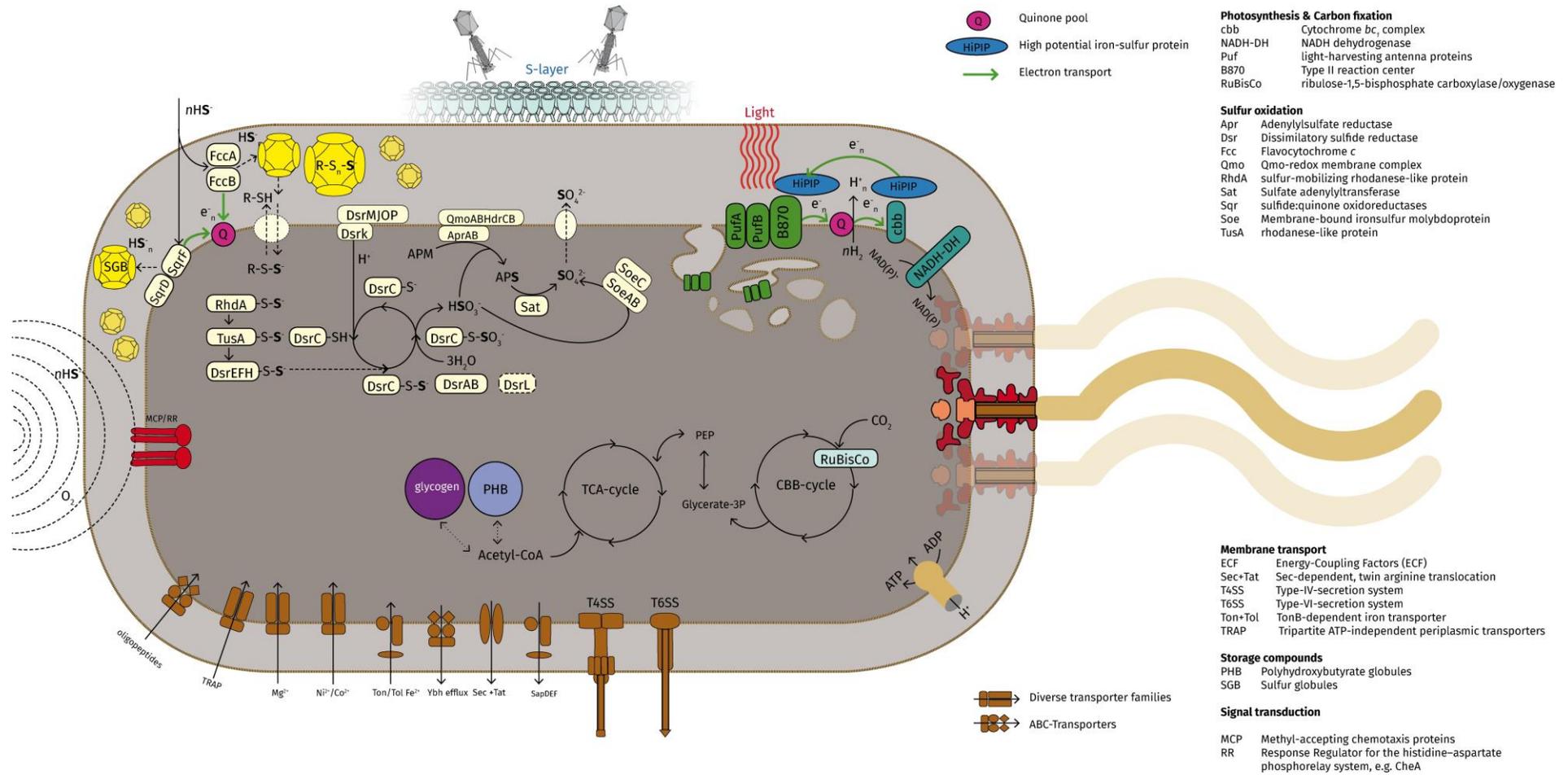
Figure 2.2-3 | Schematic *Chromatium okenii* cell with the discussed metabolic and structural features indicated.

(Phage scheme adopted from https://en.wikipedia.org/wiki/File:PhageExterior.svg#filelinks)

Comparative genomics

Large celled *Chromatium* species have been described as metabolically less flexible in comparison with the non-motile small-celled PSB [7, 296] and might therefore be forced to adapt to changing conditions by moving along the optimal gradients within short timespans. Orthologue gene families can be used to compare the metabolic, structural and behavioral potential between organisms [326]. *OrthoVenn* was used to create a dataset of annotated gene clusters to compare phototrophic sulfur bacteria population of Lake Cadagno. Thereby, the genomes of previously isolated PSB ("T. syntrophicum" str. Cad16^T and *L. purpurea* str. CadA31) and GSB (*C. phaeoclathratiforme* Bu-1) were compared to PSB *C. okenii* str. LaCa. With this approach, we sought to find genes potentially explaining the co-occurrence of different anoxygenic phototrophic sulfur bacteria in the chemocline. In total 10'632 genes were included, and the 4 species encompassed 4536 gene clusters, 3902 orthologous gene clusters (at least containing two species) and 386 single-copy gene clusters (Figure 2.2-4).
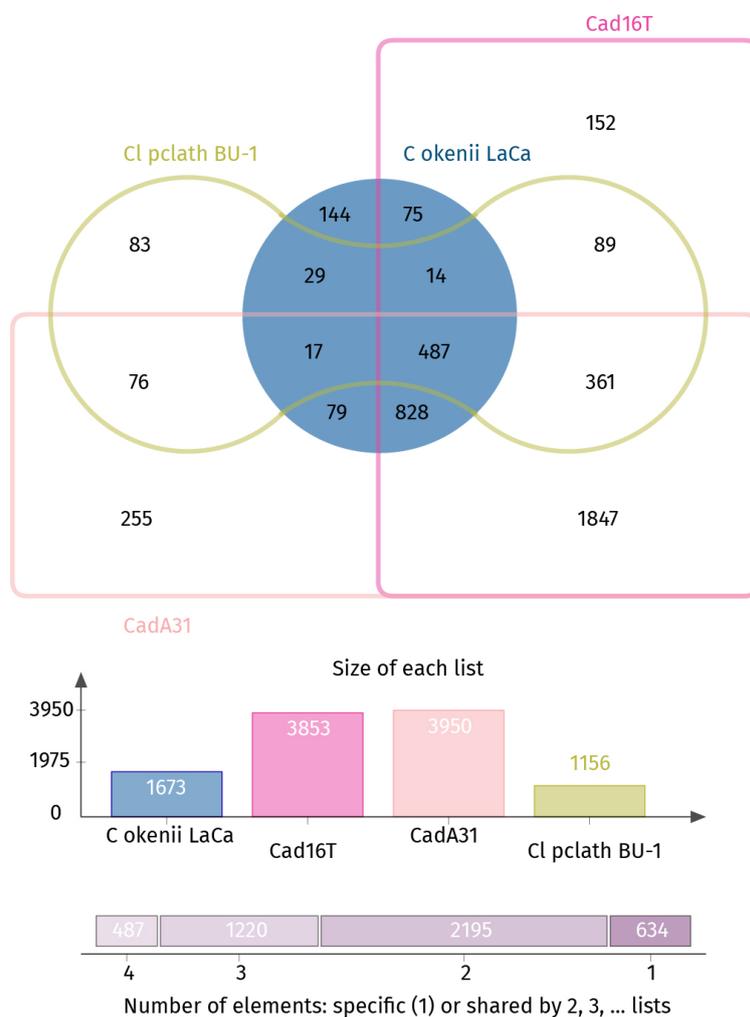
Figure 2.2-4 | Venn diagram showing the shared orthologous gene clusters among 4 sulfur oxidizing bacteria found in Lake Cadagno. Denoted the number of clusters (orthologues or paralogs) shared between phototrophic sulfur bacteria sp. Each cluster contains at least 2 genes.

Orthologous gene clusters shared by PSB ($n = 828$) were enriched for the GO-terms involved in protein export and membrane insertion as well as light harvesting complex components and cyclic electron flow indicating the primary phototrophic lifestyle and the membrane bound enzymes involved (light harvesting proteins and reaction centers as examples). Whereas GSB *C. phaeoclathratiforme* Bu-1 was enriched for chlorosome components among others, Cad 16^T was enriched for chitinase activity and extracellular and outer membrane components and CadA31 for phage related sequences and processes, respectively.

RuBisCO type II (CbbM) and RuBisCO-like (RPL) type IV proteins were found conserved in all PSB examined here. Interestingly, the heterodimeric RuBisCO type I (CbbLS) was found only in the small celled PSB and is missing in *C. okenii*. Additionally, all PSB studied contained cytochrome *d* ubiquinol oxidases (CydAB), whereas only small-celled PSB encoded a ccb3 type cytochrome c oxidase (Table 2.2-3).

Table 2.2-3| Genome features and growth characteristics of *Chromatium okenii* str. LaCa and of selected purple Sulfur bacteria. CBB: Calvin-Benson-Bassham cycle

Genome features	*C. okenii* str. LaCa	' *T. syntrophicum*' str. Cad16^T	*L. purpurea* str. CadA31*	*T. violascens* DSM 198^T	*A. vinosum* DSM 180^T
Genome size [Mb]	3.78	7.74	7.19	5.02	3.67
Number of scaffolds	1	1	302	1	1
Number of contigs	45	3	302	1	3
Average G+C content [%]	49.8	66.2	64.3	62.6	64.2
Number of genomic objects (CDS, fragment CDS, r/tRNA)	3'792	6'601	7'314	4'555	3'366
Number of coding sequences (CDS)	3'016	6'237	7'255	4'317	3'302
Motility	+	-	-	+	+
Carbon fixation	CBB	CBB	CBB	CBB	CBB
Thiosulfate oxidation	-	+	+	+	+
Chemotrophic growth	-	+	+	+	+
Hydrogenases	-	+	+	+	+
Catalases	-	+	+	+	+
cbb3 type terminal cytochrome *c* oxidase	-	+	+	+	+
Pigments	BChl *a*, okenone	BChl *a*, okenone	BChl *a*, okenone	BChl *a*, rhodopinal	BChl *a*, spirilloxanthin
Vitamin B12	+	+	+	+	+
Generation time [h]	120–168	121	33–35	NA	13–20

* A draft genome of *L. purpurea* str. CadA31 is available at our lab.

Out of the 3'016 protein coding genes identified, 144 exclusive gene-clusters were present in the genome of *C. okenii*. GO-enrichment analysis within this group resulted in over-representation of GO-terms connected to chemotaxis, flagellar movement, components of the S-layer and arginine uptake (Figure 2.2-4). The microbial utilization of free amino acids in lakes has been described as an important driver for bacterial community function [327] and arginine ammonification has been used as a proxy for respiration of microbial communities [328, 329]. Sporadic cell lysis of cyanobacteria and sulfur oxidizing bacteria blooms as described for bacterioplankton could be the source of high local concentrations of dissolved organic matter (DOM) within the Lake Cadagno chemocline. For *C okenii* str. LaCa, arginine could function as an additional carbon source not available to other PSB and GSB, and also provide N due to the high C:N ratio of 3, as proposed for other freshwater bacteria [329].

C. okenii possesses an approximately $7\times$ larger cell volume [163] and a $30\times$ reduced surface-to-volume ratio compared to small celled PSB. As bacterial cell size influences metabolic activity and internal organization [330, 331], an adaptation in transcriptional regulation, functional compartmentalization and genome organization (i.e. polyploidy) that may be fundamentally different between *C. okenii* and small-celled PSB and GSB can be expected. However, when considering the uniform coverage and the lack of allele variants of the assembly, respectively, no evidence of multiple chromosomes in *C. okenii* str. LaCa, was found.

2.2.5   Conclusions

In the study presented, we could confirm previous experimental findings on metabolic activity [296, 301, 79, 160] on the genomic level. The *C. okenii* genome encodes the CBB-cycle and a type II RC as typically found in PSB, however *sox*-proteins, hydrogenases and the Cys sulfate assimilation pathway are missing. Carbon and nitrogen utilization genes were similar between *C. okenii* str. LaCa and other PSB and show redundancy with *A. vinosum* DSM 180^{T}. Interestingly, also cytochrome *d* ubiquinol oxidases were found in all known PSB genomes of Lake Cadagno, indicating the possibility of a respiratory metabolism on oxidized organic carbon compounds such as glucose. In contrast, the arrangement of RuBisCO type II together with *cbbQ* and *cbbO* genes has been associated with obligate autotrophs [332]. Type II RuBisCO was found to be constitutively expressed in PSB "*T. syntrophicum*" str. Cad16^{T} [162] and it was suggested to be important for cofactor re-oxidation as described in purple non sulfur bacteria [303]. However, the absence of both, a CO_2 concentration mechanism and

type I RuBisCO (CbbLS) in *C. okenii*, as well as the low CO_2 affinity of RuBisCO form II [333], may limit the cell functioning to high CO_2 concentrations in the environment.

The *C. okenii* population in the Lake Cadagno is exposed to environmental factors that vary on the short-term(minutes to hours), such as availability of light, reduced electron donors and oxygen, disturbances of the water column (e.g. internal waves and seiches) and grazing pressure [128, 334]. Seasonal factors such as an increase in total cell numbers within the chemocline from spring to autumn, changes in the day to night length ratio and the 3–5 months of ice-cover from November to March (i.e. light availability and quality) add up additional complexity.

In spring, under low light availability, the reported relative higher sulfide affinity and the benefit of the larger dark-to-light hours ratio may, in turn, give *C. okenii* an advantage over small celled PSB as observed *in vitro* [102]. Additionally the low phototrophic population density of ~ 25% of the summer community [155] may minimize self-shading [335] . With the immediate onset of aerobic photosynthesis after the ice-melt additional energy may be obtained for chemotrophic microaerophilic growth. Predation rates and phage numbers may also be low.

In conclusion, the multiple factors that influence *C. okenii* str. LaCa behavior have to be further entangled. The sensing of short- and long term changes in the environment were found to have left an imprint in the *C. okenii* genome by the relative over-representation of genes involved in motility and sensing. Chemo- and scotophobotaxis, quorum sensing and diel and seasonal behavioral patterns have to be integrated in further studies when studying bioconvection. Future studies on genomic heterogeneity or, and diversity in gene regulation on single cell level could give further evidence of the ecological relevance of *C. okenii* [134, 163].

Conflict of Interest

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

Author Contributions

Acknowledgments

2.2.6 Supplementary Data

Tables

Table S 2.2-1| Phage and prophage sequences detected in the *C. okenii* str. LaCa genome.

	Contig (accession)	Start	Stop	Length (bp)	No. Of genes	coverage (canu assembly)
detected prophages	PPGH01000021.1	5'655	16'320	10'666	7	7.7
	PPGH01000023.1	266	10'497	10'232	7	2.3
	PPGH01000037.1	919'636	952,286	32,651	34	-
	PPGH01000038.1	70'225	136'412	66'188	66	-
detected phages	PPGH01000025.1	1	13'879	13'879	15	7.2
	PPGH01000028.1	1	14'664	14'664	18	3.9
	PPGH01000033.1	1	15'163	15'163	20	4.6
	PPGH01000045.1	1	16'950	16'950	19	55
	PPGH01000027.1	1	38'611	38'611	54	6.4

Figures

Figure S 2.2-1| Graphical summery of the *Chromatium okenii* LaCa genome sequencing and assembly.
a) Graph of the PacBio RSII output with the distribution of the read lengths after quality filtering. Sequencing of 2 SMRT cells resulted in total 611'754'382 bp, 95'063 reads with a N50 of 8'942 and a mean length of 6'435 bp after filtering. b) canu assembly scheme of the contigs of the *C. okenii* enrichment using the bandage [336]. The 3.78 Mb chromosome encompasses 6 contigs that could not be circularized.

Figure S2.2-2 | Phylogenetic tree including *C. okenii* str. LaCa and 12 *Chromatiaceae* with WGS data available.
Furthermore, the related phylogenetic lineages Nitrosococcus, Rheinheimera and Arsukibacterium are represented. Strain LaCa is most closely related to *Thiocystis violascens* DSM 198T. The consensus tree is based on best-model maximum-likelihood estimation and 1'000 bootstrap iterations. The node numbers indicate bootstrap support below 75%.

2.3   Research paper III

Anoxygenic Photosynthesis and Dark Carbon Metabolism in the Purple Sulfur Bacterium 'Thiodictyon syntrophicum' nov. strain Cad16[T]

Samuel M Luedin, Nicola Storelli, Francesco Danza, Samuele Roman, Andreas Bruder, Matthias Wittwer, Joël F Pothier and Mauro Tonolla

Statement of contribution

I performed sample preparation, protein extraction and field measurements and contributed to scintillation measurement, data interpretation and manuscript preparation. Please note that the methods part contains standard text on LC-MS settings and data analysis provided by the proteomics service center.

Research objective

Strain Cad16[T] was previously shown to fix a major fraction of the bulk CO_2 under both light and dark conditions [162] and was suggested to be involved in Fe(II) driven aerobic metabolism [138] in the lake Cadagno chemocline. However evidence on genomic and proteomic level was only fragmentary [337] . In order to understand the biological functions expressed under phototrophic and chemotrophic growth conditions of str. Cad16[T] an *in situ*

experiment conceived. It included cultures of str. Cad16^T incubated in dialysis bags for three months while physiochemical conditions were monitored longitudinally. Upon sampling, physiological differences of str. Cad16^T were compared utilizing quantitative proteomics in combination with CO_2-uptake measurements.

2.3.1 Abstract

The microbial ecosystem of the meromictic Lake Cadagno (Ticino, Swiss Alps) has been studied intensively to understand metabolic functions driven by the highly abundant anoxygenic phototrophic sulfur bacteria of the families Chromatiaceae and Chlorobiaceae. It was found that the sequenced isolate "*Thiodictyon syntrophicum*" nov. sp. str. Cad16^T, belonging to the Chromatiaceae, may fix 26% of all bulk inorganic carbon in the chemocline at day and night. With this study, we elucidated the mode of dark carbon fixation of str. Cad16^T with a combination of long-term monitoring of key physicochemical parameters, ^{14}C-incorporation experiments and quantitative proteomics of *in situ* dialysis bag incubations of pure cultures. During the study period in summer 2017, the chemocline sank from 12 to 14 m accompanied by a bloom of cyanobacteria. Sampling was performed both day and night for several sampling campaigns during this period. While CO_2 assimilation rates were higher during the light period, the relative change in the proteome (663 quantified proteins) was only 1% of all CDS encoded in str. Cad16^T. Oxidative respiration was thereby upregulated at light, whereas stress-related mechanisms prevailed during the night. These results indicate that the low light availability due to high cell concentrations and the oxygenation of the chemocline induced a combination of parallel phototrophic and chemotrophic growth in str. Cad16^T.
The complete proteome data have been deposited to the ProteomeXchange with identifier PXD010641.

2.3.2 Introduction

Inorganic carbon, nitrogen and sulfur are cycled in diverse microbial metabolism networks in pared redox reactions and organic compounds are thereby produced in the scale of 10^{15} g year^{-1} worldwide [1]. Light has been used as a source of energy in phototrophic anoxygenic bacteria since at least 3.85 Ga (billion years before the present) [338]. Thereby, the transfer of electrons along a gradient of carrier molecules with sequentially lower potential allows the generation of energy bound to phosphate (ATP) and reductants [e.g., NAD(P)H and reduced ferredoxin] for carbon fixation. The electrons needed to replenish the oxidized electron acceptor pool derive from the oxidation of reduced sulfur species such as sulfide, sulfite and thiosulfate, or H_2 and Fe (II) [339]. The anoxygenic photosynthetic purple sulfur bacteria (PSB) of the family Chromatiaceae are found widespread in aquatic, sulfidic oxygen minimum-zones where light is still available [267]. As an adaption to low light availability

around 0.1–20 μmol m^2 s^{-1} and only limited wavelength (450–600 nm, and infrared above 750 nm), PSB contain pigments of the carotenoid and bacteriochlorophyll *a* and *b* (BChl) classes [340], as well as multiple copies of antenna peptides (LHC), to subtly modulate charge separation within the membrane bound type II reaction center (RC) [227, 341]. Carbon is typically fixed through the Calvin-Benson-Bassham (CBB) cycle [342]. In order to store both, reduction-equivalents and oxidized carbon, PSB intracellularly concentrate elemental sulfur-chains (S-S$_n^0$) in protein covered globules (SGBs) [343] and glycogen and polyhydroxybutyrate (PHB) [307], respectively. These storage assimilates may subsequently allow for chemotrophic growth in the dark [71, 160].

The possibility of chemolithoautotrophic growth of PSB under microaerophilic conditions in the dark has been proposed by van Niel [78] and has been described first for *Thiocapsa roseopersicina* BBS [344]. The PSB *T. roseopersicina* inhabiting shallow tidal flats are especially adapted to the daily changes of oxygen concentrations and have to compete over substrates with chemotrophic non purple sulfur bacteria (*Thiobacillus* spp.) and *Beggiatoaceae* spp. [295]. Chemoheterotrophic, chemoautotrophic and mixotrophic growth has since been shown for different PSB spp. [345–347, 80, 348, 349]. Several strains of *Allochromatium vinosum* and *T. roseopersicina* have shown mixotrophic growth under a 5% oxygen atmosphere and acetate, and different reduced sulfur compounds [347]. The ecological significance and the impact on biogeochemistry of PSB were extensively studied in permanently stratified lakes [114, 262, 149, 152, 202, 113]. In Lake Cadagno (Piora valley, Swiss Alps), sub-aquatic springs in gypsum rich dolomite provide a steady inflow of solute-rich water. In combination with solute-poor surface water, a stable and steep gradient in redox potential, salinity, sulfide and oxygen concentrations at around 12 m depth is formed [122]. Within this chemocline a dense population of phototrophic sulfur oxidizing bacteria (PSOB) of the family *Chromatiaceae* and *Chlorobiaceae* (GSB; green sulfur bacteria) thrive to dense populations with up to 10^7 cells ml^{-1} between June and October [118]. *In situ* chemocline incubation experiments with ^{14}C-uptake in Lake Cadagno [128, 162] and other lakes [350], as well with nanoSIMS ^{13}C-stable-isotope labelling [163] revealed both light driven and dark carbon fixation of PSB. Thereby it was found, that the population of PSB isolate 'Thiodictyon syntrophicum' sp. nov. str. Cad16^T (str. Cad16^T, thereafter) [164] assimilated 26% of the total carbon the chemocline during dark and light incubations [162]. Alternatively, the relative contribution of different PSB and GSB spp. to total carbon assimilation normalized to biomass during daytime was estimated with stable isotope analysis for Lake Cadagno. Thereby, str. Cad16^T only photosynthetically fixed 1.3 to 2% of the carbon, as estimated from

the daily bulk δ^{13}C-mass balance [269]. Additional insight came from an *in vitro* quantitative proteomics study with str. Cad16[T] growing anaerobically under light and dark with 1 mM H$_2$S [337]. Photosynthesis driven growth of str. Cad16[T] resulted in the relative >1.5× expression of 22 proteins. Most notably, the poly(R)-hydroxyalkanoicacid synthase subunit PhaE and the phasin (PhaP) involved in the synthesis of PHB were found. In contrast, among the 17 proteins overexpressed under dark conditions, three enzymes of the dicarboxylate/4-hydroxybutyrate (DC/HB) cycle were detected, indicating dark carbon fixation through this typically *archeal* pathway. The complete genome of str. Cad16[T] gave evidence of the biological functions encoded [351]. Similar to *Allochromatium* spp. or *Lamprocystis spp.*, str. Cad16[T] expresses a type II (quinone type) RC, the membrane-bound protein cascade of cyclic electron transport to generate ATP and reverse electron transport to produce NAD-(P)H, and also contains a *cbb*3 type cytochrome *c*. The later may enable microaerobic growth and Fe(III) oxidation of str. Cad16[T] [138]. However, no genetic evidence for the possible syntrophic relationship within aggregates of *Desulfocapsa* sp., and also only incomplete Sox and no thiosulfate dehydrogenase Tsd proteins, responsible for SO$_3^{2-}$ oxidation, as previously described for str. Cad16[T] by Peduzzi and colleagues [164], were found. By studying the metabolic pathways of PSB str. Cad16[T] in more detail, we believe that important information about key metabolic mechanisms can be obtained.

With this study, we wanted to elucidate differences between day and night in the metabolism of PSB str. Cad16[T] *in situ* at the chemocline of Lake Cadagno with a combination of CO$_2$ assimilation analysis using ^{14}C-scintillation and proteomics using label-free quantitation tandem mass spectrometry (LFQ-MS^2). The carbon assimilation rates and protein profiles obtained, thereby revealed a mixotrophic metabolism of str. Cad16[T] influenced by the unique microaerobic *in sit*u in summer 2017.

2.3.3 Materials and Methods

Study Site and Field Measurements

The *in situ* incubations were placed in Lake Cadagno between July 13 2017 to September 23 2017 attached to a mooring (46°,33''05,1" N/8°,42',43'0" E, max. depth 18 m) (suppl. Figure S 2.3-1b and c). A CTD (Conductivity, Temperature, Depth) 115 probe (Sea & Sun

Technology GmbH, Germany) equipped with temperature, salinity, oxygen, sulfide, redox potential, chlorophyll *a* (Chl *a*) and turbidity sensors was used to measure physicochemical profiles. Profiles from the July 13 2017 were taken as an estimate on chemocline depth (suppl. Figure S2.3-2). Sulfide and Chl *a* sensors were not functioning in September, therefore this data is missing. HOBO UA-002-64 Pendant® data loggers (Onset Computer Corporation, MA, USA) were programmed for 60 min intervals of relative light (Lux; 180–1200 nm) and temperature measurements. Two sensors were placed at the subsurface (0.05 m) and other pairs 0.4 m apart at the upper and lower part of the rig, respectively (suppl. Figure S 2.3-1c). An empirical conversion factor of Lux=0.018 PAR and 0.016 PAR (μmol quanta m^{-2} s^{-1}) was used for the surface and below the water, respectively [352]. Data was analyzed with HOBOware v3.7.14 (Onset Computer Corporation, MA, USA).

Flow Cytometry for Cell Enumeration

Flow cytometry (FC) based cell enumeration was performed as in Danza and colleagues [134]. In short, phototrophic bacteria were identified using 50 μl sub-samples in triplicates with a BD Accuri C6 cytometer (Becton Dickinson, San José, CA, USA). A forward scatter threshold of FSC-H 10'000 was applied to exclude abiotic particles. A second red fluorescent (FL3-A) threshold above 1'100 was applied to select for cells emitting autofluorescence due to Chl and BChl. The flow rate was set to 66 μL min^{-1}. SSC values were used as an indication for internal cell complexity as in [134].

Biovolume and Biomass Calculation

The biovolume was calculated for strain Cad16^{T} assuming a median diameter of 2 μm (range; 1.4-2.4 μm) for a spherical cell. Biomass was estimated using the conversion factor determined for Lake Cadagno PSB strains 4.5 fmol C/μm^{-3} [163].

Uptake rate calculation

We took the estimate from Camacho *et al* [128] that half of the carbon is fixed through oxygenic and anoxygenic photosynthesis, respectively. Musat and colleagues estimated

anoxygenic carbon assimilation for PSB *C.okenii* and GSB *C. clathratiforme* to be 70 % and 15 % of the total photosynthetic CO_2 assimilation, respectively [163]. We therefore calculated the uptake rates for the three populations as follow (Eq. 9) where A_{day} is the total uptake rate per cell of the phototrophic community and F the fraction as estimated in [163]:

$$A_{Pop} = A_{day} * 0.5 * F \tag{9}$$

Bacterial Cultures and Media

Strain Cad16[T] was isolated in 2003 [106] and is grown in pure culture in our lab (Figure S 2.3-1a). Autotrophic Pfennig's medium II was used to grow str. Cad16[T] [353]. The medium contained; 0.25 g of KH_2PO_4 L^{-1}, 0.34 g of NH_4Cl L^{-1}, 0.5 g of $MgSO_4 \cdot 7H_2O$ L^{-1}, 0.25 g of $CaCl \cdot 2H_2O$ L^{-1}, 0.34 g of KCl L^{-1}, 1.5 g of $NaHCO_3$ L^{-1}, 0.02 mg of vitamin B12 L^{-1} and 0.5 mL of trace element solution SL12 L^{-1}. The medium was autoclaved under a 80 % N_2 / 20 % CO_2 atmosphere [354] and 1.1 mM $Na_2S \cdot 9H_2O$ was added aseptically. The pH was adjusted to 7.0. Cultures were grown in 500 ml glass bottles at RT and under a 12/12 h light/dark-regime with a 60 W incandescent lamp (6 μ mol quanta m^{-1} $s^{-}1$). Cells were grown up to a concentration of around 3×10^6 cells ml^{-1}. Cell concentrations were measured by flow cytometry.

Cellulose dialysis bags with a 14 kDa cutoff (D9777-100FT, Sigma-Aldrich, Buchs, CH) were rinsed for 1.5 h in Na_2CO_3 (40g l^{-1}) and 0.01 M EDTA at 60 °C wile stirring. The bags were cleaned with ddH_2O, cut into 0.6 m long pieces, closed by a knot on one end and autoclaved for 20 min at 121 °C. On site, about 80 ml of bacteria culture were filled randomly in each bag, which was closed, attached to a rig and installed in the chemocline within 30 min. In total, 18 dialysis bags where placed at 12 m depth from 13 July to 23 August 2017 and then lowered to 14 m.

¹⁴C-Incubations

The scintillation experiment was performed as in Storelli *et al* [162]. In short, three dialysis bag samples were pooled together randomly, 7 ml $NaH^{14}CO_3$ ($NaH^{14}CO_3$; 1.0 mCi; 8.40 mCi $mmol^{-1}$, 20 μCi mL^{-1};Cat.No NEC-086S Perkin-Elmer, Zurich, Switzerland) were added and

incubated in cleaned and autoclaved 50 ml translucent Duran® glass bottles (SCHOTT AG, Mainz, Germany). Six replicates of str. Cad16[T] cultures were incubated for 4 hours during day (1:00–4:00 pm) and night (9:00 pm –12:00 am), respectively. Chemocline background fixation rates were determined in 50 ml chemocline samples. Filtered chemocline lake water (0.45 μm) was used as negative control to estimate the ^{14}C background. Upon retrieval, the amount of β-activity (^{14}C) assimilated by microbes during the incubation time was measured in the laboratory following standard method that included acidification and bubbling of the samples [355].

The inorganic dissolved carbon concentration was determined with the CaCO$_3$ *Merck Spectroquant* kit No. 1.01758.0001 and the *Merck spectroquant Pharo 100* photospectrometer (Merck & Cie, Schaffhausen, Switzerland). Samples were taken at 14 m depth, filtered with 0.45 μm filters, pH was tested with indicator paper to lie within 6.8–7.0 and triplicate samples were measured.

Scintillation was done on a *Guardian 1414* liquid scintillation counter (Perkin Elmer Wallac, MA, USA) running with the WinSpectral v1.40 software. Raw data was statistically analyzed using *t*-test based statistics were performed using Microsoft excel.

Protein Extraction and Digest

For the proteome samples, cells were transferred to 50 ml tubes upon retrieval and stored at 4 °C in the dark. Samples were brought to the lab within 30 min and centrifuged 10'000 *g* at 4 °C for 10 min. The supernatant was discarded and the cell pellets were re-suspended in 1× PBS pH 7.0 and 1% EDTA-free Protease Inhibitor Cocktail (*v/v*; Thermo Fisher Scientific, Rheinach, Switzerland) and frozen at -20 °C until further processing.

The cells were thawed lysed in 5% SDC (w/w) in 100 mM ammonium-bicarbonate buffer containing 1% EDTA-free Protease Inhibitor Cocktail (*v/v*; Thermo Fisher Scientific, Rheinach, Switzerland) and sonicated for 15 min at 200 W at 10 °C with a *Bioruptor* ultrasonicator (Diagenode SA, Belgium). Samples were then shipped to the Functional Protein Genomic Center Zurich (FGCZ) on dry ice for further processing. Protein concentration was estimated using the *Qubit*® Protein Assay Kit (Thermo Fisher Scientific, Rheinach, Switzerland). The samples were subsequently processed using a commercial *iST*

Kit (PreOmics, Germany) with an updated version of the protocol. Briefly, 50 μg of protein were solubilized in 'Lysis' buffer, boiled at 95 °C for 10 min and processed with High Intensity Focused Ultrasound (HIFU) for 30 s setting the ultrasonic amplitude to 85%. The samples were then transferred to the cartridge and digested by adding 50 μl of the 'Digest' solution. After 60 min of incubation at 37 °C the digestion was stopped with 100 μl of 'Stop' solution. The solutions in the cartridge were removed by centrifugation at 3800 *g*, while the peptides were retained in the *iST*-filter. Finally the peptides were washed, eluted, dried and re-solubilized in 'LC-Load' buffer for Tandem Mass spectrometry (MS^2)-Analysis.

Liquid Chromatography and MS2-Analysis

MS^2 analysis was performed on a *QExactive* mass spectrometer coupled to a nano *EasyLC 1000* HPLC (Thermo Fisher Scientific, Rheinach, Switzerland). Initial solvent composition was 0.1% formic acid for channel A and 0.1% formic acid, 99.9% acetonitrile for channel B, respectively. For each sample 4 μL of peptides were loaded on a commercial Acclaim *PepMap^{TM}* Trap Column (75 μm x 20 mm; Thermo Fisher Scientific, Rheinach, Switzerland) followed by a *PepMap^{TM}* RSLC C18 Snail Column (75 μm $\times$ 500 mm; Thermo Fisher Scientific, Rheinach, Switzerland). The peptides were eluted at a flow rate of 300 nL min^{-1} by a gradient from 5 to 22% B in 79 min, 32% B in 11 min and 95% B in 10 min. Samples were acquired in a randomized order. The mass spectrometer was operated in data-dependent mode (DDA), acquiring a full-scan MS spectra (300−1,700 *m/z*) at a resolution of 70'000 at 200 *m/z* after accumulation to a target value of 3'000'000, followed by HCD (higher-energy collision dissociation) fragmentation on the twelve most intense signals per cycle. HCD spectra were acquired at a resolution of 35'000, using a normalized collision energy of 25 and a maximum injection time of 120 ms. The automatic gain control (AGC) was set to 50'000 ions. Charge state screening was enabled and singly and unassigned charge states were rejected. Only precursors with intensity above 8'300 were selected for MS^2 (2% underfill ratio). Precursor masses previously selected for MS^2 measurement were excluded from further selection for 30 s, and the exclusion window was set at 10 ppm. The samples were acquired using internal lock mass calibration on *m/z* 371.1010 and 445.1200.

Protein Identification and Label Free Protein Quantification

The acquired raw MS^2 data were processed by *MaxQuant* v1.4.1.2, followed by protein identification using the integrated Andromeda search engine. Each file is kept separate in the experimental design to obtain individual quantitative values. Spectra were searched against a forward str. Cad16^T database (6'237 coding genes), concatenated to a reversed decoyed *fasta* database and common protein contaminants (NCBI Assembly No. ASM281377v1; release date: 2017/12/07). Carbamidomethylation of cysteine was set as fixed modification, while methionine oxidation and N-terminal protein acetylation were set as variable. Enzyme specificity was set to trypsin/P allowing a minimal peptide length of 7 amino acids and a maximum of two missed-cleavages. Precursor and fragment tolerance was set to 10 ppm and 0.05 Da for the initial search, respectively. The maximum false discovery rate (FDR) was set to 0.01 for peptides and 0.05 for proteins. Label free quantification was enabled and a 2 min window for match between runs was applied. The re-quantify option was selected. For protein abundance the intensity (Intensity) as expressed in the protein groups file was used, corresponding to the sum of the precursor intensities of all identified peptides for the respective protein group. Only quantifiable proteins (defined as protein groups showing two or more razor peptides) were considered for subsequent analyses. Protein expression data were transformed (hyperbolic arcsine transformation) and missing values (zeros) were imputed using the *missForest R*-package v1.4 [356]. The protein intensities were normalized by scaling the median protein intensity in each sample to the same values.

Scaffold v4.8.4 (Proteome Software Inc., Portland, OR) was used to validate MS^2 based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 42,0% probability to achieve an FDR less than 0,1% by the Peptide Prophet algorithm with *Scaffold* [357] delta-mass correction. Protein identifications were accepted if they could be established at greater than 54.0% probability to achieve an FDR less than 1'0% and contained at least 2 identified peptides. Protein probabilities were assigned by the *Prophet algorithm* [358]. Proteins that contained similar peptides and could not be differentiated based on MS^2 analysis alone were grouped to satisfy the principles of parsimony. Proteins sharing significant peptide evidence were grouped into clusters. For the two-group analysis the statistical testing was performed using (paired) *t*-test on transformed protein intensities (hyperbolic arcsine transformation). Proteins were called significantly

differentially expressed if linear fold-change exceeded 2-fold and the *q*-value from the *t*-test was below 0.01.

As an alternative method to find differentially expressed proteins, we used the correlation adjusted t-Score algorithm provided by the R-package sda v1.3.7 [359] to further analyze the dataset of 1'333 identified proteins with MaxQuant.

Protein Functional Annotation

BlastKOALA [278] and eggNOG v4.5.1 [276] were used to classify the proteins into functional categories. The complete KEGG-dataset for str. Cad16^T can be found under Ref. [360].

Genomic Data Availability

The complete genome of strain Cad16^T [351] is available under the GenBank assembly GCA_002813775.1

Proteomic Data Availability

The complete proteomic data of strain Cad16^T have been deposited to the ProteomeXchange Consortium [361] via the PRIDE partner repository [362] under the accession PXD010641 and project DOI 10.6019/PXD010641.

2.3.4   Results

Physicochemical parameters from July to August 2017

During July 13 to September 13 2017 different physical and chemical parameters were measured alongside the incubations in order to monitor the *in situ* chemocline conditions and adjust the incubation depth, if necessary. The passive HOBO logger values were analyzed after retrieval at the end of the experiment. Fluctuations in light intensity were positively correlated with the predicted surface radiation and negatively associated with cloud cover,

respectively (suppl. Figure S2.3-3 a). Relative average surface light intensity at 12:00 pm was 1'448.9 μmol quanta m^{-2} s^{-1} (range: 4.3–3'769.6) during the time period measured with the surface loggers (suppl. Figure S2.3-3b). The surface temperature was stable at an average of 15 °C. However, high temperatures up to 40 °C were measured, due the solar heating of the logger casing (suppl. Figure S2.3-3 b).The number accumulated sunlight were above average for the weeks observed (suppl. Figure S2.3-4). From July 13 to August 23 2017 at 12 m depth (top of the rig) an average of 3.7 μmol quanta m^{-2} s^{-1} (0.2–26.2) was measured. At the lower rig, at 12.4 m depth, a mean of 1.3 μmol quanta m^{-2} s^{-1} (0.04–7.8) was registered (suppl. Figure S2.3-3 c and d). We had additionally access to CTD data from a parallel project of Sepúlveda Steiner and colleagues where two CTD profiles were taken daily. Between July and August 2017, changes of the turbidity, oxygen and CHL *a* profiles in the daily CTD profiles indicated that the chemocline had been sinking from 12 to 13–14.5 m depth (suppl. Figure S2.3-5 Sepúlveda Steiner and Bouffad, pers. comm). To ensure chemocline conditions for the incubations, we adjusted the depth of the rig from 12 to 14 m depth. This resulted in an unexpected reduced relative light intensity at novel depth at 14 m for the days August 23 to September 12 2017, as registered with the light loggers. Only an average of 0.4 μmol quanta m^{-2} s^{-1} (0.2–4.0) was measured at 14 m depth from August 24 to September 13 2017, and no light was measured at 14.4 m depth after August 23 2017. Temperature was stable around 5 °C (4.6–6.1) at both incubation depths with a positive trend over the months (suppl. Figure S2.3-3 c and d).

Chemical and Physical Analysis of Lake Cadagno at Sampling Date

Additional physicochemical measurements and endpoint-sampling was done at September 12 and 13 2017. The weather was cloudless and with only weak wind. Two CTD profiles at 1:30 pm and 9:00 pm showed a comparable situation for temperature, dissolved oxygen and conductivity (Figure 2.3-1). Water temperature was stable at 12 °C at the surface down to the thermocline at 9.5 m, where after it dropped to 5 °C at 16 m depth at both time points. Dissolved oxygen was measured at 8.5 mg l^{-1} (265.6 μM) throughout the mixolimnion, by 7 m down to 9.5 m the concentration steadily declined to 6 mg l^{-1}, and then more rapidly to 2 mg l^{-1} at 10.5 m depth, at both time points measured. At 14 m depth, 0.4 mg l^{-1} (12.5 μM) and 0.3 mg l^{-1} (9.4 μM) were measured during day and night, respectively. Conductivity increased along the profile from 0.13 in the mixolimnion to 0.22 mS cm^{-1} hypolimnion, both at day and night. In contrast, a pronounced turbidity peak (30 FTU) was observed at a depth of 13 m at

1:30 pm whereas a broader distribution of the FTU values (6–16 FTU) from 13 to 14 m depth was observed at 9:00 pm. Water samples taken at 1:30 pm from 14 m depth showed a milky and pink coloration, characteristic for a concentrated PSB community. The total inorganic dissolved carbon concentrations measured at 14 m depth were 1.26 mM at 2:00 pm and 1.46 mM 9:00 pm.

Flow Cytometry based Enumeration

Initial str. Cad16^T average cell concentrations were on average 3.1×10^6 cells ml^{-1} in July, as measured by FC. The rigged cultures were checked on 23 August 2017 and all dialysis bags were found intact and cells were judged healthy due to the turbid-pinkish appearance. When retrieved for sampling on 12 September 2017, all dialysis bags were still intact, the population was uniformly distributed within the dialysis bags, and the cells grew to a mean concentration of 9.3×10^6 cells ml^{-1}. No significant difference in cell concentration and in the SSC/FSC signature was measured between the two randomly assigned sampling groups (P=0.74). In total, the str. Cad16^T cultures grew 3-fold from July to September.

The average cell concentration in the lake sample taken at 14 m was 4.2×10^6 cells ml^{-1} at 1:30 pm, whereas it was 1.69×10^5 cells ml^{-1} at 9:00 pm. However, the later value has to be questioned as the FC count is below the 0.45 μm filtered negative control in contrast to the similar turbidity values (4-5 FTU) measured at the two time points at 14 m depth (Figure 2.3-1). FC revealed that the phototrophic microbial community at 1:30 pm (4.23×10^6 cells ml^{-1}) consists mainly of *C. okenii*, *Chl. clathratiforme* and *cyanobacteria* with 1.48×10^6 cells ml^{-1}, 7.45×10^5 cells ml^{-1} and 1.49×10^6 cells ml^{-1}, representing 35, 35 and 17 % of the total phototrophic population respectively.

In situ **Carbon Fixation Rates**

Absolute carbon fixation rates at the chemocline were comparable between day and night when tested, with medians of 757 nM h^{-1} and 587 nM h^{-1}, respectively (Figure 2.3-2). Strain Cad16^T fixed carbon in both conditions tested (Figure 2.3-3). For str. Cad16^T ^{14}C-bicarbonate median uptake rates normalized per cell were significantly different with 1'074 amol C cell^{-1} h^{-1} (range: 937–1'585 amol C cell^{-1} h^{-1}) during the day, and 834 amol C cell^{-1} h^{-1} (range; 650–969 amol C cell^{-1} h^{-1}) during the night (Figure 2.3-3). When the uptake rates were normalized
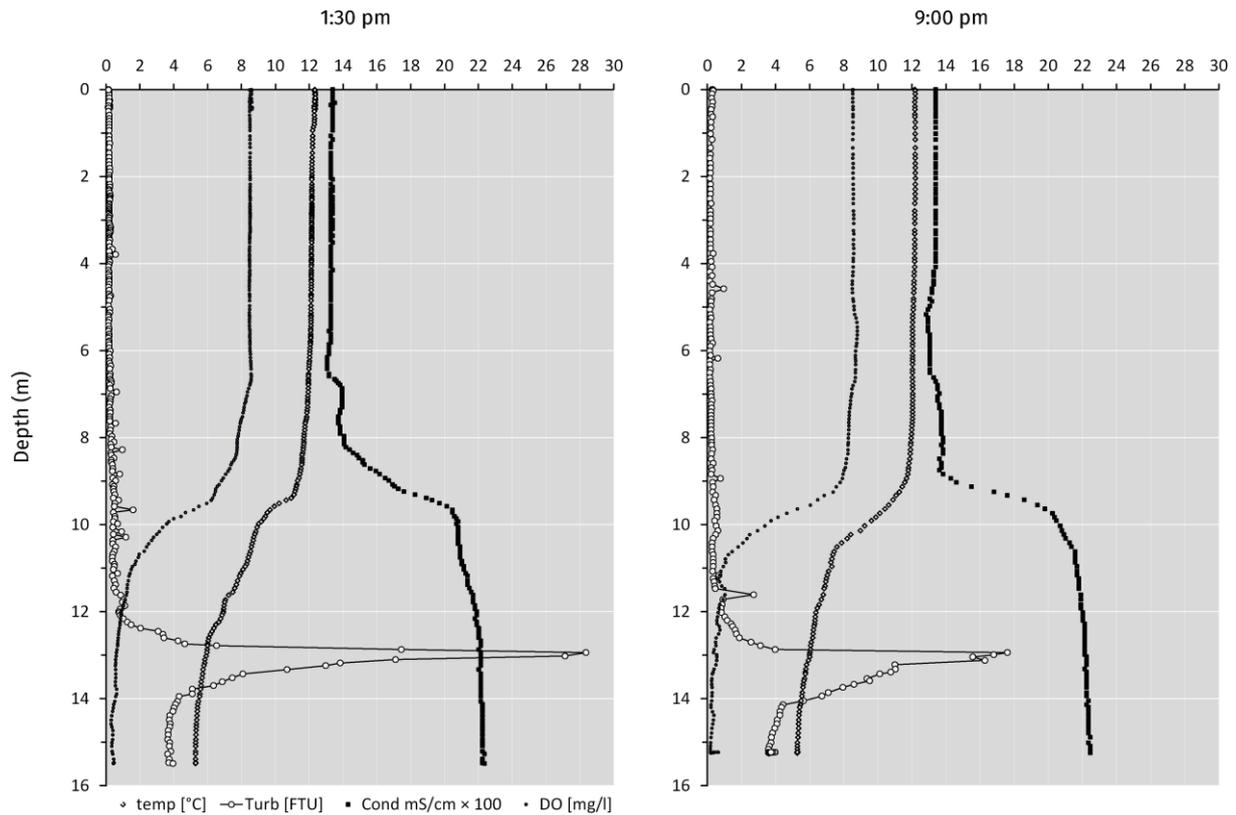
Figure 2.3-1 | CTD depth profiles from Lake Cadagno on September 12 (1:30 pm) and (9:00 pm) 2017.
Measurements were taken from the platform. Probe was equilibrated for 5 min at 0.5 m depth before measuring.
Temperature (◊), FTU: Formazin Turbidity Unit (○), Cond: Conductivity (■), DO: Dissolved Oxygen (●),
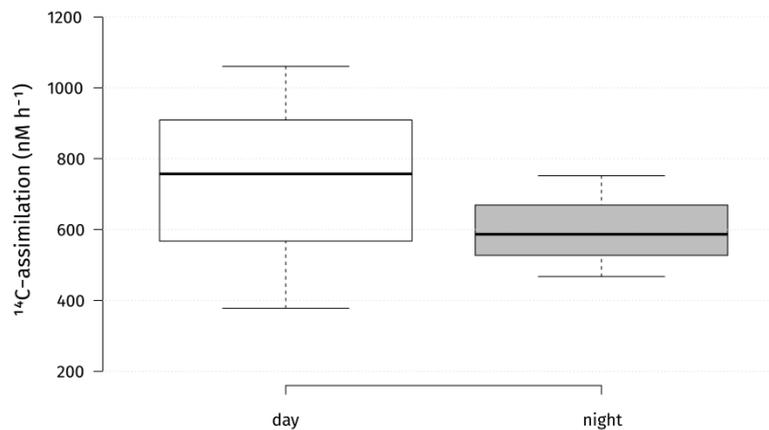
Figure 2.3-2| Absolute microbial 14C-uptake rates in the chemocline of Lake Cadagno during day and night at a depth of 14 m.
Data derived from 3 biological replicates per condition. Center lines show the medians; box limits indicate the 25th and 75th percentiles as determined by *R* software; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles, outliers are represented by dots. n = 3 sample points.
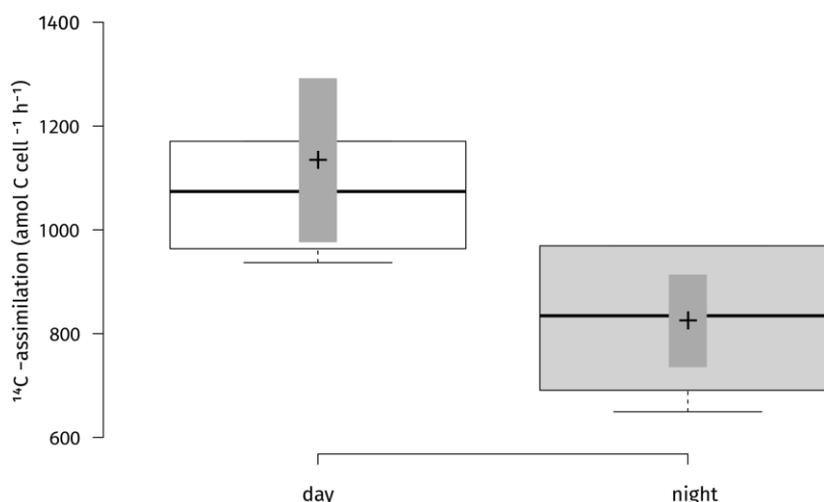
Figure 2.3-3 | Carbon uptake rates per cell for strain Cad16T cultures for two conditions (day /night) during 4 h incubation in situ.

[14]C-scintillation experiments were performed on 6 biological replicates. Two side *t*-test statistics was applied. The difference in uptake rates between the two conditions was statistically significant at $p < 0.05$ (p=0.02). Center lines show the medians; box limits indicate the 25th and 75th percentiles as determined by R software;
.

whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles, outliers are represented by dots; crosses represent sample means; bars indicate 83% confidence intervals of the means; data points are plotted as open circles. n = 6 sample points

to average biovolume (3.6 μm^3 cell^{-1}), 316 amol C μm^{-3} h^{-1} with light and 230 amol C μm^{-3} h^{-1} in the dark were obtained, and a carbon-based doubling time range of 5 to 23 h was calculated for str. Cad16^T. For the dominant phototrophic populations of the chemocline we calculated uptake rates in amol C cell^{-1} h^{-1}only at day because of uncertain FC values at night. Thereby *C. okenii*, GSB *Chlorobium* spp. and *cyanobacteria* assimilated in average 173.35, 73.70 and 245.68 amol cell^{-1} h^{-1}, respectively.

Total Proteins identified with LC-MS^2-LFQ

A total of 11 samples, 5 samples for category 'light' and 6 samples for category 'dark', were processed and the protein quantification was performed using the MaxQuant package. We used corrected t-test based statistics in order to identify and quantify proteins. The samples Cad16T_dia_7 ('light') and Cad16T_dia_12 ('dark') were identified as outliers in cluster analysis and were excluded from further data analysis. Therefore, the data analysis was made with 4 samples of the category 'light' and 5 samples of the category 'dark'. Overall a total of 1'333 proteins (21 % of the total coding CDS) with at least 2 peptides were identified.

Proteins Quantified with LC-LFQ-MS²-

Peptide identifications were accepted if they could be established at >42.0 % probability to achieve an FDR less than 0.1% with Scaffold delta-mass correction, resulting in 12,576 spectra included. Protein identifications were accepted if they could be established at > 54.0% probability to achieve an FDR less than 1.0 % and contained at least 2 identified peptides. The number of quantified proteins per condition was constant, with an average of 374 for 'day' and 354 for 'night', respectively. Between 102 and 663 proteins could be quantified in each biological replicate (7.2–37 % of all IDs). Consequently, 684 proteins were quantified over all samples. Thereof, 21 contaminants were excluded. The remaining 663 CDS were classified with blastKOALA and EggNOG, with 627 annotated CDS for (99%) COG and 460 annotated CDS (69.4%) for blastKOALA, respectively.

As expected, many of the proteins with unchanged abundance belonged to the functional categories energy conversion, genetic information processing, carbohydrate and amino metabolism and protein modification (Table 2.3-1). Among the most abundant proteins detected were the F_0F_1 ATPase subunits (AUB84561.1, AUB81565.1, AUB81565.1 and AUB84563.1) and the chaperons (GroEL; AUB81575.1, AUB80010.1, AUB84066.1, DnaK; AUB84026.1). Since cell growth is dependent on protein synthesis, we found 36 ribosomal subunits as well as elongation factor Tu (AUB80476.1) to be equally abundant.

The cells always contained the established components of the dissimilatory sulfate reduction pathway such as ATP-sulfurylase Sat (AUB82369.1), the adenylylsulfate reductase AprAB (AUB82371.1, AUB82370.1) and the sulfite reductase Dsr complex (AUB83448.1–AUB83455.1). A Sqr sulfide:quinone reductase and a glutathione amide reductase GarA homolog to *A vinosum* putatively involved in intracellular sulfur shuttling [48] were additionally present. In PSB sulfur oxidation provides the electrons for cyclic electron transport driven by light. Consequently, in str. Cad16^T PufMCL (AUB85378.1–AUB85380.1) and PuhA (AUB85431.1) subunits forming the RC II and six different PufAB antenna proteins (AUB85355.1–AUB85357.1, AUB85363.1, AUB85361.1, AUB85710.1) were expressed. Additionally we found also enzymes for BChl synthesis, terpenoid backbone biosynthesis and carotenoid synthesis. Noteworthy, elements of the reduction pathways driven by photosynthesis are shared with oxidative phosphorylation in PSB. We found in total 23 protein subunits involved in substrate respiration including the NADH dehydrogenase

subunits NuoCDEFG an HoxFU, the cytochrome reductase CytB and Cyt1, seven F-type ATPase subunits and two *cbb3* cytochrome *c* oxidase subunits. PSB use the ATP and NAD(P)H derived from photosynthesis to fix CO_2 through the CBB cycle. For str. $Cad16^T$ a complete CBB cycle with the key enzymes CbbM/RbcL RuBisCO form II (AUB81831.1) and phosphoribulokinase PrkB (AUB79979.1) were present. The fixed carbon enters the central carbon metabolism as 3-Phospho-D-glycerate. In both growth conditions, str. $Cad16^T$ contains enzymes for glycolysis and gluconeogenesis, as well as pyruvate oxidation, the glyoxylate cycle and the citrate cycle (TCA cycle) in unvaried abundance. Additionally, the presence of malic enzyme (MaeB; AUB82893.1) may allow the entry of malate into the central carbon pathway via pyruvate, is shown for *A. vinosum* [363]. In both conditions examined, the PHB synthase subunits PhaC and PhaE are expressed (AUB80707.1 and AUB84676.1). Also enzymes necessary for amino acid biosynthesis and Co-factor and vitamin synthesis were expressed under both groups analyzed.

In a second step, we analyzed the proteome data for significant changes between growth conditions. However, we found a large within-group variation and no proteins were significantly differently expressed between the two growth conditions using (paired) corrected *t*-test on transformed protein intensities (q.mod <0.01).

Table 2.3-1 | Functional categories for the proteins quantified for str. Cad16T. EggNOG [276] and blastKOALA [278]was used to classify proteins according to COG and KEGG classification, respectively.

COG Code	Description	No. Entries
J	Translation, ribosomal structure and biogenesis	64
K	Transcription	15
L	Replication, recombination and repair	7
D	Cell cycle control, Cell division, chromosome partitioning	5
V	Defense mechanisms	3
T	Signal transduction mechanisms	19
M	Cell wall/membrane biogenesis	28
N	Cell motility	1
U	Intracellular trafficking and secretion	10
O	Posttranslational modification, protein turnover, chaperones	50
C	Energy production and conversion	93
G	Carbohydrate transport and metabolism	35
E	Amino acid transport and metabolism	52
F	Nucleotide transport and metabolism	12
H	Coenzyme transport and metabolism	27
I	Lipid transport and metabolism	18
P	Inorganic ion transport and metabolism	37
Q	Secondary metabolites biosynthesis, transport and catabolism	6

KEGG Functional Category	
Genetic information processing	78
Carbohydrate metabolism	78
Energy metabolism	41
Metabolism of cofactors and vitamins	29
Amino acid metabolism	28
Environmental information processing	27
Nucleotide metabolism	17
Metabolism of terpenoids and polyketides	11
Lipid metabolism	9
Metabolism of other amino acids	5

Proteins Differentially Expressed

The expression dataset was alternatively analyzed using correlation-adjusted t-scores (CAT scores) in order to additionally address the correlative structure of the dataset as only 4.5% of the proteins are differentially expressed. Thereby, 60 proteins were found differentially expressed (1 % of all coding CDS) at a false discovery rate of lfdr<0.05.

During the 'day' period 21 CDS were found relatively more abundant for str. Cad16^T. Thereof, all CDS were annotated with eggNOG (Table 2.3-2). Growing in the light, str. Cad16^T over-expressed four proteins involved in oxidative phosphorylation that were subunits of NADH:quinone oxidoreductase, cytochrome bc1 complex respiratory unit and AtpC subunit of F-type ATPase. Two enzyme involved in the central carbon pathway were abundant, the glycogen synthase GlgA and the glucose-6-phosphate isomerase GPI involved in glycolysis.

The acetolactate synthase I/III small subunit associated with thiamine synthesis and the 7,8-dihydroneopterin aldolase FolB involved in tetrahydrofolate biosynthesis was additionally found an ABC-2 type transport system and Twin-arginine translocation (Tat) system were detected involved membrane transport. Additionally, chaperone-type proteins DnaJ and proteolytic ClpX were also over-expressed.

In the dark 39 proteins were shown to be more abundant and thereof 35 entries (89 %) were COG annotated (Table 2.3-3). Analysis of this data showed that the NAD(P)H isocitrate dehydrogenase Idh1 responsible first carbon oxidation from oxaloacetate to 2-oxoglutarate in the TCA cycle as well as a FabH 3-oxoacyl-[acyl-carrier-protein] synthase III involved in

Table 2.3-2 | List of proteins more abundant in the light period. Lfdr: local false discovery rate, cat: t-score for each group and feature the cat score of the centroid versus the pooled mean [364].

Accession	Name	COG cat.	cat score	lfdr
AUB81546.1	cytochrome c1	C	5.44	9.53E-11
AUB83791.1	NADH-quinone oxidoreductase subunit I	C	4.35	8.32E-05
AUB84560.1	F0F1 ATP synthase subunit epsilon	C	3.51	0.00022589
AUB83997.1	acetolactate synthase small subunit	E	2.88	0.01230503
AUB81260.1	glucose-6-phosphate isomerase	G	3.63	0.00022589
AUB83261.1	starch synthase	G	4.48	2.53E-06
AUB79823.1	thiazole synthase	H	2.74	0.02928445
AUB81699.1	dihydroneopterin aldolase	H	3.41	0.00128609
AUB83078.1	DNA topoisomerase I	L	2.98	0.00380615
AUB84659.1	outer membrane lipoprotein carrier protein LolA	M	5.43	2.03E-10
AUB82245.1	HflK protein	O	6.34	1.24E-13
AUB82609.1	ATP-dependent protease ATP-binding subunit ClpX	O	3.38	0.00128609
AUB84025.1	molecular chaperone DnaJ	O	6.17	1.24E-13
AUB79510.1	hypothetical protein	S	2.84	0.01230503
AUB81403.1	hypothetical protein	S	3.53	0.00022589
AUB82197.1	hypothetical protein	S	3.54	0.00022589
AUB82687.1	hypothetical protein	S	3.96	0.00012143
AUB83592.1	histidine kinase	S	3.51	0.00128609
AUB84052.1	Host attachment protein	S	3.51	0.00022589
AUB84690.1	hypothetical protein	U	3.02	0.00380615
AUB80231.1	multidrug ABC transporter ATP-binding protein	V	3.83	0.00012143

fatty acid biosynthesis initiation and elongation were abundant. The adenylylsulfate reductase subunit alpha was found, and the APS reductase AprA is responsible for sulfite oxidation to 5'-adenylyl sulfate [72]. Strain Cad16^T further expressed proteins associated to cell division (FtsZ), cell wall formation CpxP, Lysine and branched amino acid synthesis and nucleotide metabolism (ppx-gppA; exopolyphosphatase and rutE; 3-hydroxypropanoate dehydrogenase). Elements of two secretion pathways were identified, a putative polar amino acid transport system and Type II general secretion system. Additional three proteins detected are involved in stress response. The BolA-family transcriptional regulators, is a general stress-responsive regulator, whereas rubrerythrin may provide oxidative stress protection via catalytic reduction of intracellular hydrogen peroxide. Furthermore, an ATP-dependent serine protease mediates the degradation of proteins and transitory regulatory proteins, and thereby ensures cell homeostasis.

Table 2.3-3 | List of proteins more abundant in the dark period. Lfdr: local false discovery rate, cat: t-score for each group and feature the cat score of the centroid versus the pooled mean [364].

Accession	Name	COG cat	cat. score	lfdr
AUB80187.1	isocitrate dehydrogenase (NADP)	C	2.94	0.0038062
AUB81294.1	oxidoreductase	C	3.54	0.0002259
AUB82371.1	adenylylsulfate reductase alpha subunit	C	3.03	0.0038062
AUB83156.1	NADH dehydrogenase NAD(P)H nitroreductase	C	3.03	0.0038062
AUB84003.1	rubrerythrin	C	2.54	0.0444337
AUB80125.1	cell division protein FtsZ	D	2.67	0.0292845
AUB79575.1	extracellular solute-binding protein, family 3	E	2.62	0.0444337
AUB81750.1	decarboxylase	E	2.63	0.0292845
AUB83785.1	4-hydroxy-tetrahydrodipicolinate synthase	E	2.94	0.0041797
AUB84954.1	acetolactate synthase	E	2.76	0.0195093
AUB84325.1	proline iminopeptidase	E	2.91	0.0041797
AUB82337.1	adenylosuccinate synthase	F	2.54	0.0444337
AUB79762.1	ubiquinone biosynthesis regulatory protein kinase UbiB	H	2.86	0.012305
AUB80927.1	uBA THIF-type NAD FAD binding	H	2.64	0.0292845
AUB83525.1	synthase	I	5.36	6.662E-10
AUB81024.1	lysyl-tRNA synthetase	J	3.31	0.0012861
AUB79657.1	OmpA MotB	M	5.26	9.242E-09
AUB80121.1	cell wall formation (By similarity)	M	3.04	0.0038062
AUB83127.1	choloylglycine hydrolase	M	2.98	0.0038062
AUB83783.1	conserved repeat domain protein	M	2.58	0.0444337
AUB83066.1	twitching motility protein	N, U	3.14	0.0038062
AUB79946.1	heat-shock protein	O	2.65	0.0292845
AUB82608.1	endopeptidase La	O	5.02	1.012E-07
AUB85620.1	DnaK-related protein	O	2.78	0.012305
AUB81090.1	protein of unknown function (DUF971)	S	3.55	0.0002259
AUB82842.1	dinitrogenase iron-molybdenum cofactor biosynthesis protein	S	3.30	0.0038062
AUB83253.1	nitrogen fixation protein	S	2.96	0.0038062
AUB83860.1	tellurite resistance protein TehB	S	2.76	0.012305
AUB83943.1	short-chain dehydrogenase reductase SDR	S	2.91	0.0041797
AUB84318.1	transposase	S	2.99	0.0038062
AUB84331.1	dienelactone hydrolase	S	2.91	0.0041797
AUB85354.1	protein of unknown function (DUF2868)	S	3.78	0.0002259
AUB82296.1	Ppx GppA phosphatase	T	2.79	0.012305
AUB82910.1	YciI-like protein	T	2.63	0.0292845
AUB79943.1	general secretion pathway protein G	U	2.96	0.0038062
AUB80110.1	hypothetical protein		3.21	0.0038062
AUB81753.1	ATPase		4.76	1.203E-06
AUB82624.1	hypothetical protein		2.81	0.012305
AUB82926.1	hypothetical protein		2.99	0.0038062

2.3.5 Discussion

In the study presented, we compared light to dark carbon fixation metabolism of PSB str. Cad16^T through a combination longitudinal monitoring of physicochemical condition of the Lake Cadagno chemocline, scintillation and quantitative proteomics of cultures incubated *in situ*.

Monitoring Data

During the incubation experiment from July to September 2017 the chemocline was monitored with CTD and passive light and temperature. This detailed record allowed us to understand the prevailing physicochemical conditions experienced by the str. Cad16^T cultures. Thereby, the average light availability was 10× lower than previously measured at the chemocline, whereas the available DIC concentration was comparable to 2013 [162]. CTD measurements (this study) and additional FC enumeration [365] of the chemocline, revealed dense PSB and GSB populations (10^6 cells ml^{-1}) comparable in numbers to previous years, however at around a depth 2 m below the average 11-12 m [118]. Together with the additional cyanobacterial bloom down to the monimolimnion [134] the phototrophic microbial community inflicted self-shading below a depth of 13 to 14 m. The resulting low-light conditions at 14 m depth possibly reduced net photosynthesis and subsequent carbon storage capacity and growth for str. Cad16^T as observed for *Chromatium* spp. in the Lakes Cisó and Vilar [366, 98].

Whereas the monimolimnion has been typically described as anoxic [122], recurrent blooms of oxygenic plankton [128, 365] and optode-based measurements of dissolved oxygen revealed micro-oxic conditions below 20 nM at the chemocline [137]. We are just now learning to understand the consequences thereof for the PSOB.

Carbon Uptake Rates

Measured chemocline CO_2-fixation rates at both conditions were within the ranges previously obtained for Lake Cadagno, that were between 85 to 8'000 nM h^{-1} with light, and between 27 to 7'608 nM h^{-1} in dark conditions, respectively (suppl. Table S 2.3-1) [128, 163, 136, 162]. Comparable CO_2-assimilation rates have been also measured in other stratified sulfureta as in Spanish karstic lakes with up to 3'000 nM h^{-1} both with light or in the dark; [367]. It was estimated that only half of the bulk carbon is fixed by anoxygenic photosynthesis in the Lake Cadagno chemocline [128]. Furthermore, PSB *C. okenii* was found to be responsible for 70% of the total anoxygenic phototrophic CO_2 assimilation [163] and the concentration was 10^6 cells ml^{-1}—35 % of the total phototrophic microbes— at 14 m depth in this study. Taken together, *C okenii* may account for an average 173.35 amol C cell^{-1} h^{-1} at the light, that is 60-

times lower that previously observed [163]. The average GSB assimilation rate with 73.7 amol C cell^{-1} h^{-1} was higher than observed (1–30), but within the same log [163]. Oxygenic photosynthesis contributes 245.68 amol C cell^{-1} h^{-1} to the C-fixation that is 10–100 less than observed in the Baltic sea [368]. As oxygenic photosynthesis is inhibited by elevated sulfide concentrations common in the chemocline [369], that may have reduced the photosynthetic efficiency of *cyanobacteria.*

For str. Cad16^T, the median uptake rate of 1'073.9 amol C cell^{-1} h^{-1} during the day was within the range measured for other PSB (100–30'000 amol C cell^{-1} h^{-1}) [163, 162], however 10× lower at when compared a previous *in situ* ^{14}C-assimilation study with strain Cad16^T (around 12'000 amol C cell^{-1} h^{-1}) [162]. In the former experiment, autotrophic CO_2 assimilation rates of str. Cad16^T were additionally measured in vitro, with values between 8'541.7–18'541.7 amol C cell^{-1} h^{-1} with light, and around 2'916.7 amol C cell^{-1} h^{-1} in the dark, respectively. For C-fixation at night, highly variable C-assimilation rates in chemocline bulk samples were measured, ranging from 7–45 % of the rates measured with light [149]. Schanz and colleagues thereby found a positive correlation between the photosynthetically driven increase in overall biomass and the dark fixation rate. In our study, the median dark fixation rate of 834.4 amol C cell^{-1} h^{-1} was significantly lower than during the day and again around 10× lower than measured by Storelli and colleagues [162]. This overall discrepancy may be explained by the 10-fold lower light availability that possibly reduced photosynthetic carbon assimilation and carbon storage (see above). Furthermore the sampling times in our study were adapted to the natural light-dark hours. This in order to account for possible circadian effects, whereas in the former study incubations were performed in parallel at daytime, using clear and opaque sample bottles.

The differences between the study results may further be explained with the varied cell counting methodologies. Fluorescent *in situ*-hybridization (FISH) was used to count cells in the study of Storelli *et al* [162]. In the study presented, we additionally did not control for alive or dead cells and FC was used which might have led to a relative overestimation of the str. Cad16^T cell concentration as a consequence. In contrast, str. Cad16^T forms aggregates, and some of them stuck to the dialysis bags, reducing the number of planktonic cells for counting.

When normalized to biomass, the carbon uptake values at light were higher as previously measured for PSB *C. okenii* and *Lamprocystis* sp. [163]. Uptake rates normalized to

biovolume depend on the cell volume estimation (i.e. volume depends on r^3) and the conversion factors chosen. We used values previously used in literature (see methods). For str. Cad16[T], the thereby calculated carbon-based doubling time was between 4 and 23 h, which is in contrast to previous estimates of 121 h *in vitro* [164], and a median of 333.6 h (25th percentile; 196.8 h, 75th percentile; 652.8 h) for the bulk biomass in the chemocline [149]. In contrast, the average doubling time obtained by FC enumeration is 948.0 h (39.5 d) which is 40× longer than our ^{14}C-uptake based calculations, and 8× longer than *in vitro* [164]. We did not control for the presence and metabolic activity of the syntrophic sulfate reducing *Desulfocapsa* sp. nov. str. Cad626 [69], as it was observed that str. Cad16[T] lab cultures lost the initially co-occurring str. Cad626 after several passages.

The significantly higher inorganic C-uptake rate during the day when compared to the night suggests active photosynthesis of str. Cad16[T] at low light intensities, as it was described for other PSB [149, 370]. Noteworthy, the presence of up to 3×10^5 ml^{-1} oxygenic phototrophic microbes down to 16 m depth [134] resulted in a partly oxygenated chemocline with around 0.6 mg O_2 l^{-1} (19 μM), as measured with CTD. As a consequence, the chemocline waters retained some of the produced oxygen during the night (Figure 2.3-1). Taken together, str. Cad16[T] may have used the O_2 present as electron acceptor during mixotrophic growth under both conditions. As a consequence, some of the CO_2 assimilated might have been constantly respired with thiosulfate as electron donor as found for *T. roseopersicina* str. M1 [295]. In accordance, str. Cad16[T] grew in the dark with thiosulfate and 5% O_2 *in vitro* [106]. Interestingly however, we found only SoxXY and SoxB and no TsdA thiosulfate-oxidizing enzyme homologues encoded in the str. Cad16[T] genome [351]. As the complete Sox-complex is essential for the complete thiosulfate oxidation to sulfate in *A vinosum* DSM 180[T] [59], str. Cad16[T] possibly uses an alternative mechanism. One option may be thiosulfate uptake through CysTWA (AUB80378.1, AUB80379.1 and AUB80380.1) and CysP (AUB80377.1) and oxidation to sulfite via the intermediate S-sulfocysteine by cysteine synthase B CysM (AUB82938.1) and possibly monothiol glutaredoxin of the Grx4 family (AUB83488.1) as suggested by Dahl [39]. However, with the methods applied above we could not determine the relative contributions of photosynthetic or chemotrophic activity to the total increase in biomass. To conclude, ^{14}C-fixation rates and FC enumeration indicated an actively growing str. Cad16[T] population during the months incubated. The findings on carbon assimilation rates are further reflected in the proteome, as discussed below.

Stable Fraction of the Strain Cad16^T Proteome

In order to understand differences in metabolism the str. Cad16^T light/dark proteome has been studied extensively. Previously, Storelli and colleagues performed a comparable proteomics study on light/dark metabolism, str. Cad16^T *in vitro* with 2D-difference gel electrophoresis (2D-DIGE), where they identified around 1,400 proteins, quantified 56 and, thereof, found 37 differentially expressed proteins [337]. With six biological replicates instead of three, and the increased resolution and high-throughput capacity of LC-MS^2 compared to 2D-DIGE [208], we expected to substantially increase protein identification and quantitation in comparison. Overall, the number of identified proteins was comparable in both studies, with 1'400 in Ref [337] to 1'333 in the study presented. Noteworthy, LC-MS^2 analysis returns sequence information on all proteins identified, whereas 2D-DIGE is limited to a chosen subset, based on the relative intensity and size of the protein spots on the gel. Therefore, we could quantify as much as 663 proteins that are 12× more than in the former study. Nevertheless, the number of unique proteins with different abundance between the two growth conditions was comparable, with 62 in this study versus 37 in the former one, respectively. Interestingly, in this study we retrieved only 28 of these proteins in the constantly expressed fraction of the proteome.

In addition to technical differences, the former study was performed *in vitro* under an anaerobic atmosphere, with higher light intensities (6 μmol quanta m^{-2} s^{-1}) and at 20 °C, which may have an influence on metabolic rates and relative activity. In conclusion, the results of the present study can only be compared to some extend with previous findings.

The photosynthesis apparatus was abundant in both growth conditions, as several LHC proteins were detected. In accordance, also the enzymatic pathway for BChl *a* and carotenoid synthesis was expressed. In *T. roseopersicina* BChl *a* was absent at prolonged exposure to 60 μM O$_2$ in darkness [371]. In contrast, for str. Cad16^T we did not observe a loss of pigmentation. This might suggest that BChl *a* abundance was not regulated on expression level and/or the micro-oxic conditions did not influence BChl *a* biosynthesis.

The central role of dissimilatory sulfur oxidation during photosynthesis is well established for PSB [301] and proteins involved were found expressed in str. Cad16^T. The proteins Sqr, Dsr and Sat were present and SGBs were also observed microscopically in both conditions. The CBB cycle is central in purple bacteria, not only for autotrophic carbon fixation, but also to regenerate the pool of reduced co-factors NAD[P]H$_2$ [333]. Strain Cad16^T contains two forms

of RuBisCo, RbcAB form I and RbcSL form II. In a previous study Storelli and colleagues [162] detected constitutive transcription of the *rbcL* gene under autotrophic condition *in vitro* under a 12/12 h dark/light regime, whereas the form I *rbcA* was induced by light. In contrast both *rbc* genes were transcribed equally under heterotrophic conditions with acetate, with and without light, respectively [162]. In our study we detected RbcL to be equally abundant in both conditions and no other RuBisCo subunits were found. Therefore, the sole presence of the dimeric form II RuBisCo may underline the importance of the CBB cycle mediated CO_2 fixation in maintaining the redox-balance under chemo or mixotrophic growth at low light and dark conditions, as described for purple bacteria [372].

In str. Cad16^T, the CsrA (AUB84364.1) seems to be the main carbon storage regulator where it was detected under both conditions. Glycolysis under mixotrophic conditions might thereby be regulated through mRNA transcription and stability as in *A. vinosum* [72]. We additionally found enzymes involved in the central carbon pathways TCA, EMP and Glyoxylate cycle in unvaried abundance. Noteworthy, isocitrate lyase was found expressed, that is involved in the glyoxylate cycle, that prevents loss of CO_2 and ensures production of $NAD[P]H_2$ otherwise occurring through the isocitrate dehydrogenase and 2-oxoglutarate dehydrogenase in the TCA [373]. Further, the malic enzyme was abundant, that generates oxaloacetate via malate through anaplerotic reactions without ATP [37]. The ccb3 cytochrome *c* oxidase found in both conditions is used in aerobic respiration and additionally used for FeS oxidation and it was speculated that str. Cad16^T is also involved in both aerobic [138] and anaerobic cryptic iron cycling, as found for *Thiodictyon* str. F4 [374].

Differentially Expressed Proteome

In the light condition proteins of the oxidative respiration pathway were upregulated, indicating an active substrate respiration with light as in *T. roseopersicina* [348]. However, the cytochrome *bc* and the NADH-dehydrogenase complex and redox carrier molecules are also used in cyclic electron transport during photosynthesis in PSB. In the light period studied, both chemotrophic and phototrophic metabolism compete for electrons in str. Cad16^T. We also found evidence for glycolysis / gluconeogenesis since GPI was overexpressed. Interestingly, the glycogen synthase GlgA was additionally found abundant in the light. Glycogen synthesis and sulfur oxidation was found to be ineffectively regulated in *A. vinosum* [375] and we might speculate that in str. Cad16^T both processes are highly active

even at slow growth, allowing for both, intracellular sulfur and glycogen accumulation. Altogether, these results indicate an active phototrophic and chemotrophic metabolism competing for electrons in the light.

In the dark, the micro-oxic condition may have led to the production of reactive oxygen species in str. Cad16^T that would explain the upregulated proteins involved in stress induced damage-control.

Interestingly, AprA was more abundant in the night period, indicating a relative higher activity of sulfite oxidation possibly coupled to chemotrophy. SGBs are consumed under dark autotrophic conditions in PSB *C. okenii* and *C. minus* observed *in vitro* [160]*,* however the side scatter values measured did not vary between the two sample groups in this study when monitored with FC. This may be due to glycogen and PHB storage inclusions, structures that add up additional structural complexity and de-polymerization kinetics may be different to those of SGB. Supporting evidence thereof is given by the unchanged presence of enzymes involved in of glycogen and PSB synthesis.

In summary, the 60 protein found differentially expressed represent only about 1 % off all protein coding CDS and about 5 % of the identified proteins. Therefore, their impact on metabolic pathways has to be further examined.

2.3.6 Conclusions

PSB str. Cad 16^T is metabolically flexible and growth photo- as well as chemotrophically in the light as shown in this study. In dark conditions, low levels of oxygen may enable respiration of different small organic molecules. In order to understand the dark carbon metabolism, uptake experiments with labelled acetate and or pyruvate should be therefore included in the future. The long-term observation of the chemocline revealed the relative importance of oxygenic phototrophs on the oxygenation of the chemocline. However, it is not clear if O_2 or reduced sulfur is used a terminal electron acceptor in str. Cad16^T. The oxidation of Fe(II) to fix CO_2 should be tested for str. Cad16^T *in vitro*, both under microaerobic and anaerobic conditions as found in Lake Cadagno in 2017. To better estimate the relative metabolic contributions of the different energy metabolism at microaerobic conditions with light a control using oxygenic photosynthesis-inhibitors [376] should be included in future experiments. The contribution of phytoplankton to dark carbon uptake [377] has also not yet

been elucidated in Lake Cadagno and a sequential size dependent fractionation of sub-samples would be interesting. In order to further understand transcriptional control over the light to dark metabolism also mRNA sequencing experiments would be of need. To complete the understanding, a metabolomics study would give insight into the relative amount of metabolic intermediates produced under different regimes. We further observed a large variability in the C-uptake rates between different studies that cannot be readily explained by differences in the chemocline community composition and/or technical variances. Therefore, NanoSIMS experiments [378] on Cad16[T] may help to determine the within-population heterogeneity in inorganic C-assimilation under variable growth regimes.

Author Contributions

SML, NS, FD, JFP, AB and MT conceived the study. SML, NS, FD and SR installed the mooring and performed field measurement and sampling. NS and FD prepared scintillation samples. FD did flow cytometry cell enumeration. SML extracted total protein and performed scintillation measurements. SML, AB, MW and JFP and MT analyzed physicochemical, proteomic and scintillation data. SML, NS; FD and MT prepared the manuscript. All authors contributed to writing and agreed on the manuscript before review.

Acknowledgements

2.3.7   Supplementary data

Supplementary Tables

Table S 2.3-1| Absolute carbon uptake rates of the microbial community in the Lake Cadagno chemocline from different studies.

|	Conditions				
	light		dark		
Photosynthesis	oxygenic	anoxygenic		depth [m]	Reference
Carbon Uptake Rates [nM h⁻¹]	3'825	4'175	7'608	11.5	Camacho *et al* 2001 [128]
	1'200		-	11.5	Musat *et al* 2008 [163]
	85		25	12.5	Halm *et al* 2009 [136]
	6'187		4'812	12	Storelli *et al* 2013 [162]
	987		-	11.4	Berg *et al* 2016 [138]

Supplementary Figures
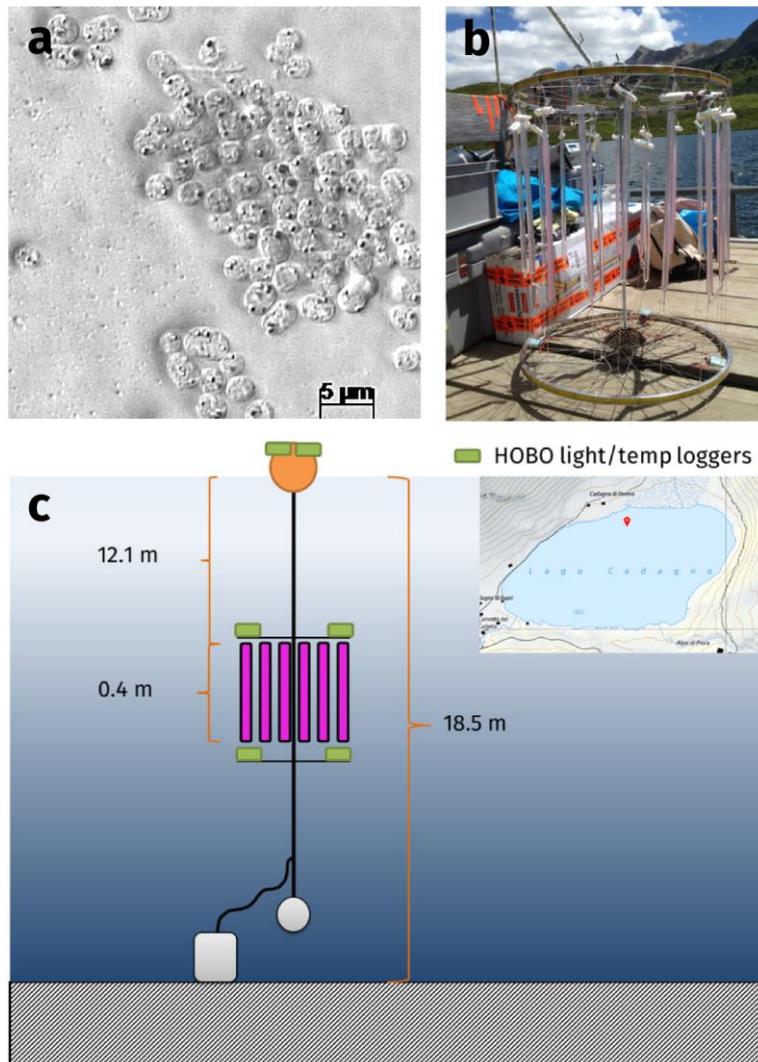
Figure S 2.3-1 | Depiction of strain 'Thiodictyon syntrophicum' sp. nov. strain Cad16^T cells, sampling site and the experimental setup.

a) Phase-contrast microscopic image of pure cultures of str. Cad16^T with sulfur inclusions visible as highly refractive particles. b) Set-up of the strain Cad16^T cultures in dialysis tubes attached to a support grid. c) Mooring scheme and location of the incubation experiment (initial depth, July to August) on Lake Cadagno. In pink indicated the dialysis tubes, in green the HOBO temperature and light sensors.
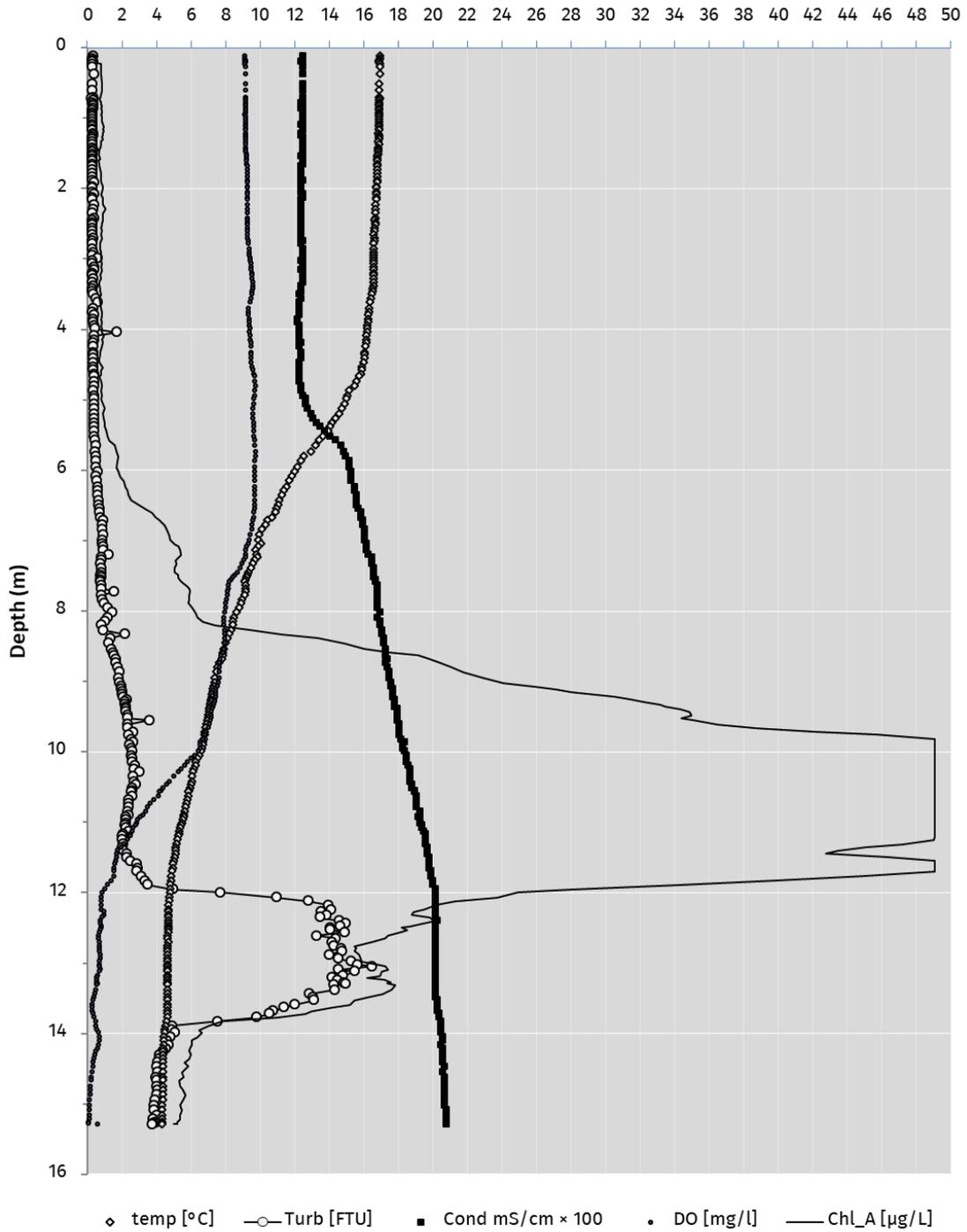
Figure S2.3-2 | CTD profile on July 13 2017 at 2:44 pm above the deepest point of Lake Cadagno. Chlorophyll a (––), Temperature (◊), FTU: Formazin Turbidity Unit (○), Cond: Conductivity (■), DO: Dissolved Oxygen (●),
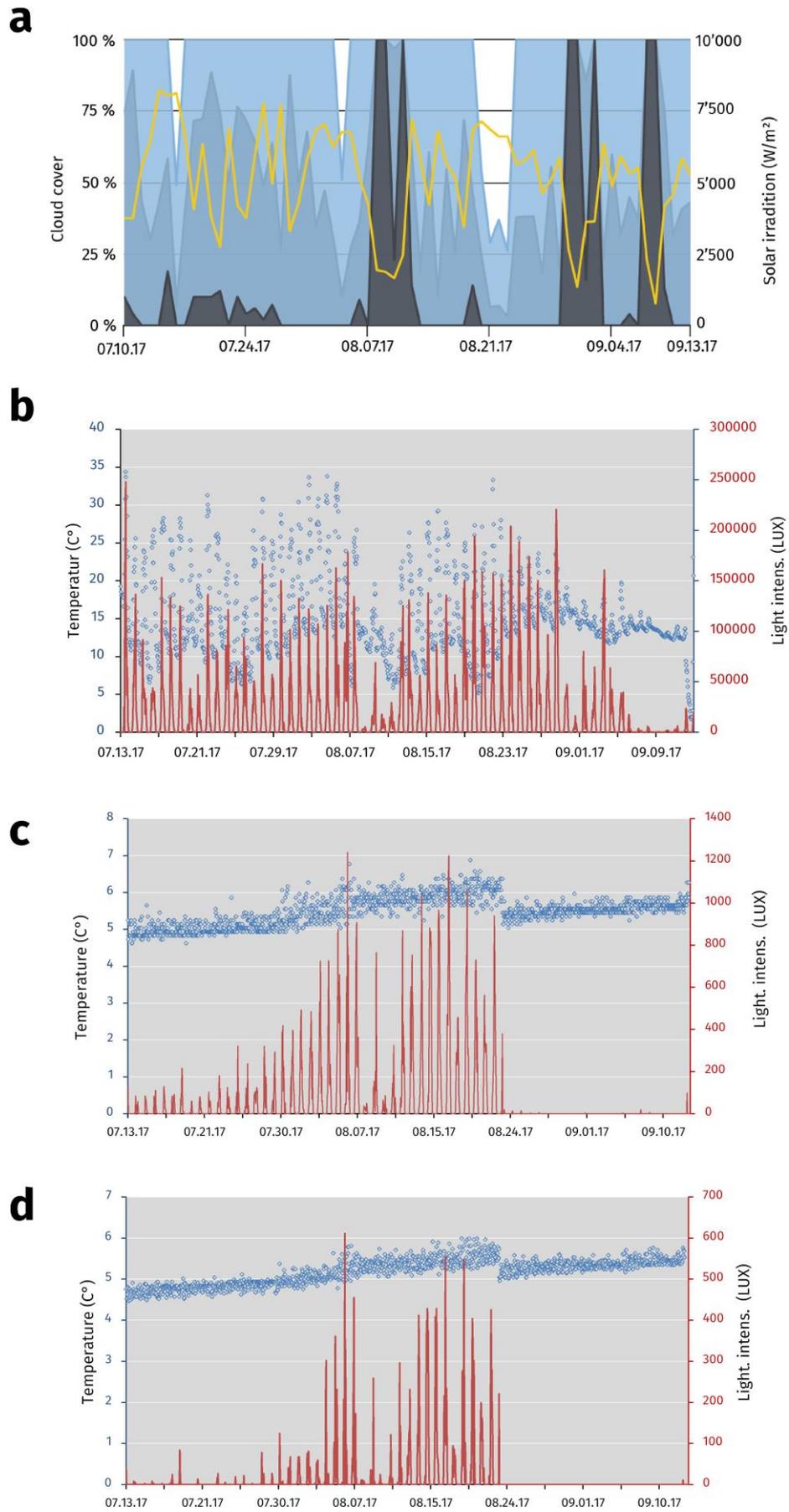
Figure S2.3-3 | Meteorological data for the Piora valley and temperature and relative light availability at different depths of the mooring in Lake Cadagno from July 13 to September 13 2017.

a) Sun light and cloud cover graph for the Piora valley from July 13 to September 13 2017. Data from meteoblue.com (yellow; sum of the daily shortwave radiation in W m-2, blue; maximal daily cloud cover in %, light blue; mean daily cloud cover in %, grey; minimal daily cloud cover in %) b) Temperature (in °C, blue circles) and average light available (in Lux, red line) at the surface buoy of Lake Cadagno. The data logger was partly immersed in water, dampening the light and temperature readings (see values in August) c) A steady increase in average temperature and light availability from July to August is visible. The increase in the available light can be explained through the downward movement of photosynthetic bacteria over the season as in the suppl. Fig 8. Re-positioning of the rig on the August 23 2017 results in both, a drop in temperature, and light availability. Available light was reduced to an average of 0.4 µmol m-2 s-1. d) Low light availability and temperature are characteristic for the depth of around 12.4 m. As in c), temperature and available light values are steadily rising from July to August. No light was detected during daytime at 14.4 m depth after repositioning in August. Data was logged in hourly intervals for b)-d).
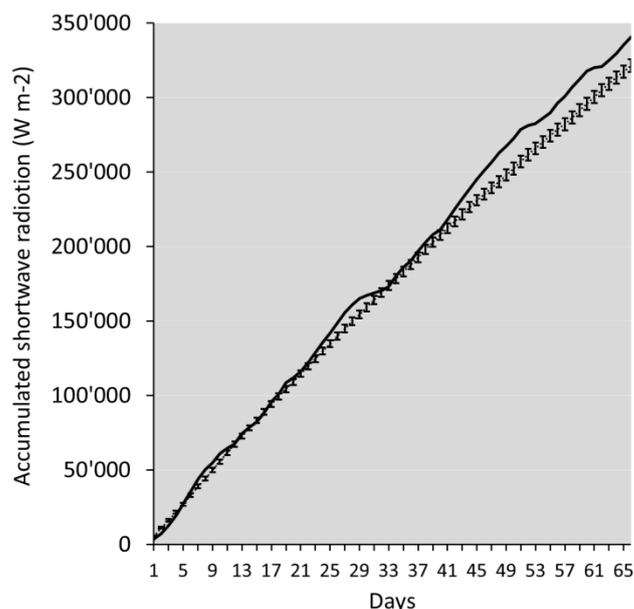
Figure S2.3-4 | Accumulated surface shortwave radiation for 66 days from July 10 to September 12 2017 at Piora valley. The values for 2017 show above-average values (solid line). The mean values from 1985 to 2017 for the same time period are represented by the dotted line. Error bars indicate standard-error. Global radiation (diffuse and direct) on a horizontal plane given in W m-2. Values derived from simulation data from meteoblue.com.
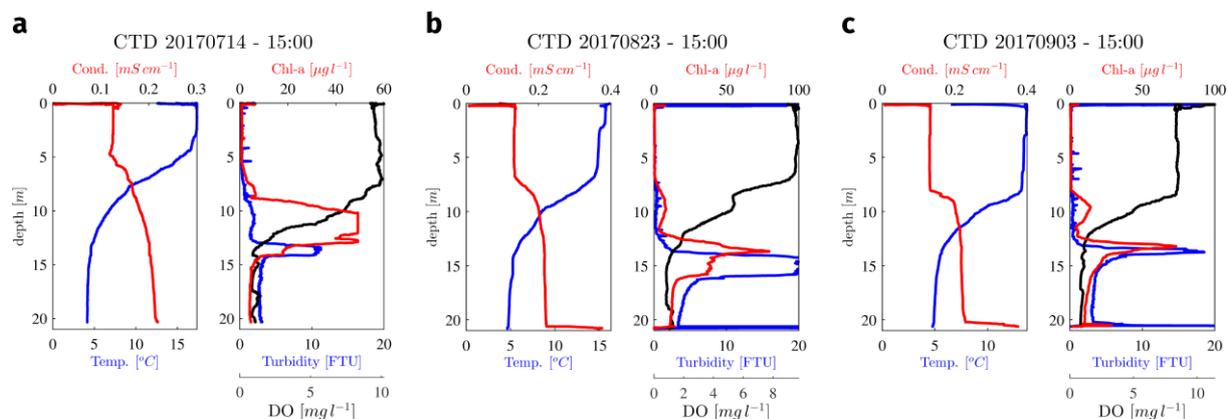
Figure S2.3-5 | Vertical CTD profiles from Lake Cadagno from July to September 2017 at the deepest point of the lake. Temperature, conductivity, dissolved oxygen, chlorophyll a and turbidity are displayed for: a) July 14 2017 b) August 23 2017 c) September 03 2017 at 3:00 pm. Formazin Turbidity Unit (FTU). The data is courtesy of Oscar Sepúlveda Steiner, Damien Bouffard and Johny Wüest. Plots courtesy of O. Sepúlveda Steiner, APHYS Laboratory, EPFL, Switzerland.
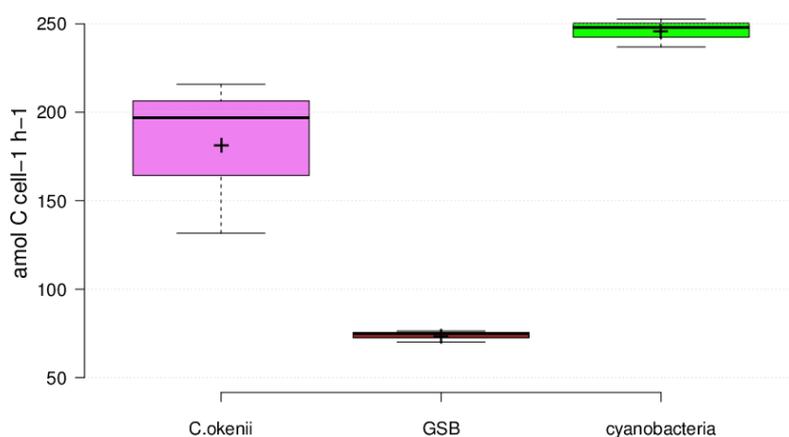
Figure S2.3-6 | Average inorganic uptake rates at the chemocline of the three populations counted with FC at 1:30 pm. *Chromatium okenii* (pink), GSB *Clathratiforme* (brown) and *cyanobacteria* (green). Center lines show the medians; box limits indicate the 25th and 75th percentiles as determined by R software; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles, outliers are represented by dots; crosses represent sample means. n = 3 sample points.

2.4 Research paper IV

MALDI-TOF MS/MS m/z biomarker for the identification of novel isolates of anoxygenic phototrophs

Samuel M. Luedin, René A. Brunisholz, Joël F. Pothier, Matthias Wittwer and Mauro Tonolla

Unpublished technical note

Statement of contribution

I performed sampling, isolation, sample preparation, protein extraction, nanoHPLC and MS experiments, and contributed to data analysis, interpretation and manuscript preparation.

2.4.1 Introduction

MALDI-TOF MS is a qualitative diagnostic tool for microbial "fingerprinting" however sequence and quantity of protein detected remains unknown. The MALDI-TOF MS typing method is therefore limited by the m/z-database of previously characterized isolates and mostly restricted to clinical microbiology [240]. However the constantly increasing number genomic information from different environments allows the prediction of putative biomarker proteins *in silico* for a more complete m/z database [233]. These putative masses —in most cases ribosomal subunits—can be further validated by isolating biomarkers and subsequent MS/MS protein sequencing [379]. In microbial ecology, MALDI was used to type rhizobial plant-symbionts [242], aquatic filamentous fungi [380] and cyanobacteria [238], as examples. The parallel organism identification and quantitation of a functional trait with MALDI-MS has been demonstrated for antibiotic resistance in bacteria [381] and indirect functional metabolic quantitation assays have been developed for the detection of cyanobacterial toxins [382], antibiotic resistance [252] and low molecular metabolites [253]. PSOB pigments and proteins were studied with MALDI-MS [383] and a first fingerprint typing was used to compare 16S mRNA taxonomy of PSB isolates [164]. For the fast and reliable MALDI-MS typing of novel PSB and PNSB isolates we developed a pipeline including explorative sample preparations and a validation step of the biomarker is proposed.

2.4.2 Methods

Bacterial cultures and enrichments

PSB 'Thiodictyon syntrophicum' sp. nov. str. Cad16[T], *Lamprocystis roseopersicina* and *Thiocystis violascen*s were cultivated in Pfennig's minimal medium [353] to an OD_{600} of 0.6 as in Ref. [164]. *Chromatium okenii* was enriched through sedimentation of the cells out of water samples from the Lake Cadagno chemocline that were left in the dark over night at 4°C. *Escherichia coli* str. K12 DH5α and *Yersinia pseudotuberculosis* were cultivated on TSA agar as positive controls. A sample collection of frozen and lyophilized (-20 °C) purple non-sulfur bacteria (PNSB) lysates and isolated proteins was obtained from R. Brunisholz (R. bruniholz, pers. commun.). Bacteria were originally cultivated as in Ref [384]. Protein samples were obtained by the method as described in [384].

Protein extraction

All chemicals were purchased from Sigma-Aldrich/Merck (Sigma-Aldrich, Buchs, Switzerland) if not further specified.

Three different extraction methods for bacterial cells were tested.

- Extraction with 25 % formic acid (FAEX from now) [245].
- Extraction with 5% SDC and 50 mM ammonium-bicarbonate (SDCEX) [385].
- Water-free extraction with methylene chloride/MeOH(*v/v*) and 0.1 M ammonium- acetate (WFEX) [384].

PSB cells were collected in 1.5 ml tubes and centrifuged for 5 min at 20'000 g at 4 °C. Two subsequent wash steps with 1×PBS pH7 and dd H$_2$O followed. In all three extraction protocols, cells were sonicated for 15 min at 200 W at 10 °C with a *Bioruptor* ultrasonicator (Diagenode SA, Belgium) in 300 μl of the protocol-specific lysis buffer. FAEX extracts in 25 % FA were diluted 1:3 with 20/80/0.1 ACN/H$_2$O/TFA and centrifuged for 30 min at 20'000 g before chromatography. SDCEX samples were sonicated and later precipitated with mixture of chloroform/MeOH/H$_2$O (1:4:3 *v/v*) by centrifugation for 10 min at 20'000 g. Proteins accumulated at the organic/aqueous interphase and the top layer was carefully removed, and the residue was then washed twice with 100 % MeOH and dried for 5 min at RT. The pellet was then re-suspended in 5 μl -20 °C cold 80 % FA, incubated at -20 °C for 10 min and subsequently diluted 1:10 to a volume of 50 μl with 20/80/0.1 ACN/ H$_2$O/TFA. Cells pellets were completely dried in a SPD1010P1-115 Savant speed-vac (ThermoScientific AG, Reinach, Switzerland) and then re-suspended in methylene chloride/MeOH (*v/v*) and 0.1 M ammonium-acetate before sonication in the WFEX protocol. All samples were stored at -20 °C.

Chromatography

A nanoHPLC Dionex3000 (ThermoFisher AG, Reinach, Switzerland) under the control of the chromeleon software v7.1 (ThermoFisher AG, Reinach, Switzerland) was used to separate proteins. A RP-C18 Acclaim Pepmap 0.075×150 mm column (ThermoFisher AG, Reinach, Switzerland) was used

with a flow of 300 nl min^{-1} and a linear gradient from 80 % A (H$_2$O 0.1 % TFA) to 90 % B (80/15/5/0.08 IPA/ACN/ H$_2$O/TFA) in 60 min was applied.

Fractions were spotted in 20 s intervals (100 nl on spot) with a SunCollect MALDI Spotter (SunChrom Wissenschaftliche Geräte GmbH, Friedrichsdorf Germany) on a 384 spot μFocus plate (Hudson Surface Technology Inc., Old Tappan, US).

Tryptic digest

The stock trypsin GOLD (Promega Corporation, Madison, US) was activated and diluted 1:100 in 50 mM ammonium-acetate pH 7.8. Protein sample spots were overlaid with 1.5 μl trypsin solution and briefly re-suspended by pipetting. On-plate trypsinisation was performed for 30 min in a SunDigest humid chamber (SunChrom Wissenschaftliche Geräte GmbH, Friedrichsdorf Germany) at 37 °C and 95 % relative humidity.

Sample deposition and matrix preparation

Cells, purified cell lysates or undigested HPLC fractions were analyzed with MALDI-MS. Cells and complete lysates were directly applied (1.5 μl) on the plate and dried. Matrix was overlaid subsequently (1.5 μl), and the matrix/analyte mixture was re-suspended by pipetting and was allowed to dry at 20 °C under a flow hood. α-Cyano-4-hydroxycinnamic acid (CHCA) was used at a concentration of 10 mg ml^{-1} for proteins and 3 mg ml^{-1} for peptides, dissolved in 50/50/0.1 ACN/H$_2$O/TFA (v/v). Sinapinic acid was used at a concentration of 10 mg ml^{-1} dissolved in 50/50/0.1 ACN/H$_2$O/TFA (v/v). 'Super DHB' (90/10 DHB and 5-methoxysalicylic Acid) was 10 mg ml^{-1} in 70/30/0.1 ACN/H$_2$O/TFA (v/v). Matrices were sonicated for 30 min after preparation and stored in the dark at 4°C.

MALDI-MS settings

MALDI-MS and collision induced dissociation (CID) MS^2 experiments were done both with AXIMA™ performance mass spectrometer (Shimadzu-Biotech Corp., Kyoto, Japan) and data was preprocessed with the Launchpad™ software v2.9 (Shimadzu-Biotech Corp., Kyoto, Japan). The acceleration voltage was set to 20 kV and the laser frequency was 50 Hz. Spectra were obtained in positive linear mode 2'000 – 30'000 m/z. The ion gate was set to 1'900 Da

and the spectra were pulsed extraction optimized at 10'000 Da. Advanced peak cleanup settings were the following: peak width of 80, smoothing filter 50 and baseline 500 channels, respectively. Peak detection was set to a dynamic threshold apex with a baseline of 0.125 and a response of 1.25. ASCII files of the m/z peak-list were exported and used for further analysis.

External calibration in linear mode was done with *Escherichia coli* str. DH5α at the beginning of each MS experiment by calibrating the spectrum against a list of ribosomal subunits from *Escherichia coli* str. K12 in the range 3'000–20'000 m/z. Sample spots were scanned in circular raster with 50 *μ*m spacing and 100 shots with 20 consecutive shots—adding up to 2'000 single spectra—were taken.

MALDI-MS^2 CID settings

Spectra were obtained in positive reflectron mode 700 – 4'500 m/z at a frequency of 50 Hz. The ion gate was set to 690 Da and the spectra were pulsed extraction optimized at 2'300 Da. Isotopically resolved peak cleanup settings were the following for precursor fragments: peak width of 5 channels. Peak detection was set to threshold 25% centroid with a dynamic baseline of 1 and a response of 1. Monoisotopic peak picking parameters were set to a minimal mass of 700 Da, a max. mass of 4'500 Da, a minimum number of isotopes of 2 with a max. intensity variation of 80 and allowing for a minimal peak distribution overlap of 80 %.

In a first step, peptide match fingerprinting (PMF) was done with MASCOT ncbiRefseq + WGS information of Lake Cadagno PBS, allowing for two miss-cleavages and a monoisotopic tolerance 0.3 Da was set. Statistically significant PMF precursors were selected using a channel width of 8 collision induced dissociation with helium gas (3 bar inlet pressure) was performed. Monoisotopic peak picking parameters were off. Laser was set to +20 % a.u. power than in PMF. Ion gate was set 'off'.

In the second peptide sequencing step, MASCOT MS/MS settings were the following: 2 trypsin misscleavages, 0.5 Da average peptide tolerance, carbamidomethyl as fixed and Met-oxidation as variable modifications and and MS^2 tolerance of 0.8 Da and a peptide charge of +1 for MALDI-TOF-TOF. Fourty peaks were used for identification and the precursor was set as monoisotopic. ASCII files of the m/z peak-list were exported and used for further analysis.

External calibration in reflectron mode was done with the MASCAL2 ProteoMass™ (Sigma-Aldrich/Merck KGaA, Darmstadt, Germany) peptide mixture (10 pM on-plate concentration) at the beginning of each MS^2 experiment,and then each 10th spot during measurements, by calibrating the spectrum against a list of five peptides and the matrix peaks in the range 200–4'500 m/z. Sample spots were scanned in circular raster with 50 μm spacing and 50 spots with 50 consecutive shots—adding up to 2'500 single spectra—were taken.

MALDI-MS^2 TDS settings

All Bruker rapifleX™ MALDI-TOF MS(Bruker GmbH, Bremen, Germany) top-down sequencing (TDS) experiments were done by Dr. Volker Sauerland with the standard settings as in Ref. [386].

2.4.3   Results

Linear MALDI-MS

Linear MALDI-MS measurement of directly smeared PSB isolates and *C. okenii* enrichments revealed different m/z spectra. Intensive m/z signals between 4'000-10'000 Da thereby match with the predicted masses of LH peptide masses (Figure 2.4-1). Different cell lysates from purple bacteria were also anlysed (Figure 2.4-2)
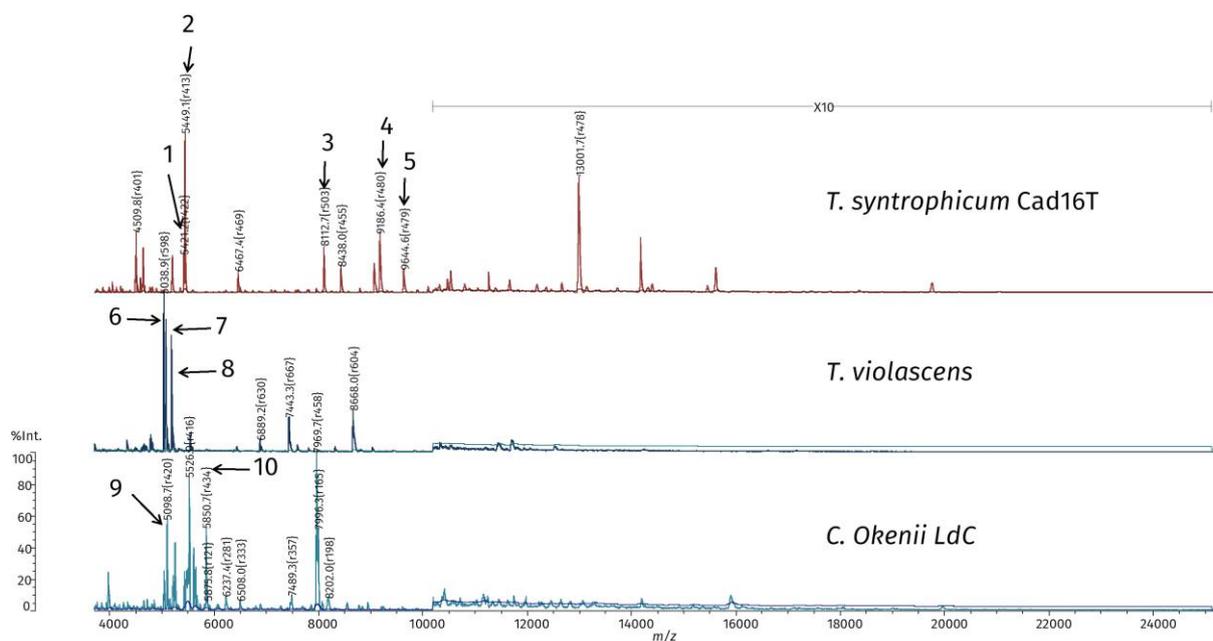
Figure 2.4-1 | Linear, positive MALDI-TOF MS spectra of PSB isolates 'Thiodictyon syntrophicum' str. cad16ᵀ, Thiocystis violascens DSM 198 and Chromatium okenii str. LaCa. The following m/z signals match LH antenna peptides, for str. Cad16T; 1: AUB85356.1, 2: AUB85381.1 3: AUB85382.1, 4: AUB85355.1 and 5: AUB85363.1, for. *T. violascens*; 6: WP_014778400.1 7: WP_014778398.1 and 8: WP_014778412.1, and *C. okenii*; 9: WP_105073895.1 and 10: WP_105073882.1.
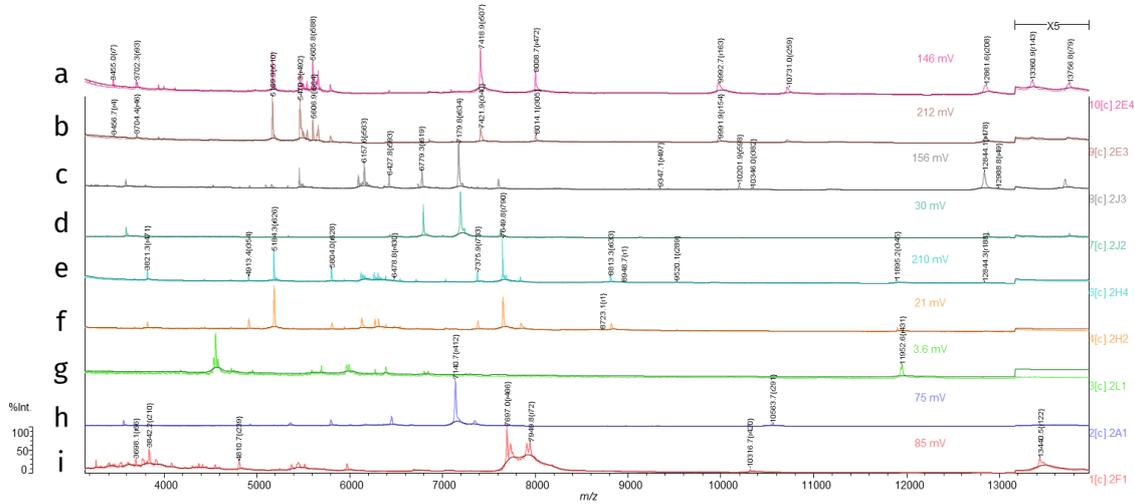
Figure 2.4-2 | MALDI-MS spectra of different Purple non-sulfur bacteria show dominant signals between 5'000 to 9'000 Da in positive linear mode. a) *Rp. palustris*; sDHB 20mg ml-1, b) *Rp. palustris*; sDHB 20 mg ml-1, c) *Rs. rubrum* s1; sDHB 20 mg ml-1, d) *Rs. rubrum* s; SA 10 mg ml-1, e) *Rp. marina*; sDHB 20 mg ml-1, f) *Rp. marina*; SA 10 mg ml-1, g) *Rp. acidophila* 7050; SA 10 mg ml-1, h) *Rc. tenuis*; SA 10 mg ml-1, i) *Rc. gelatinosus*; SA 10 mg ml-1 Abbreviations: SA: Sinapinic acid, sDHB: super DHB

We further analyzed previously isolated single LH *α/β*-antenna proteins from *Afifella marina* DSM 2698 (before *Rhodopseudomonas marina*) [387] and retrieved matching m/z signals, however with slight modifications possibly derived from oxidation (+ 16 Da) (Figure 2.4-3).
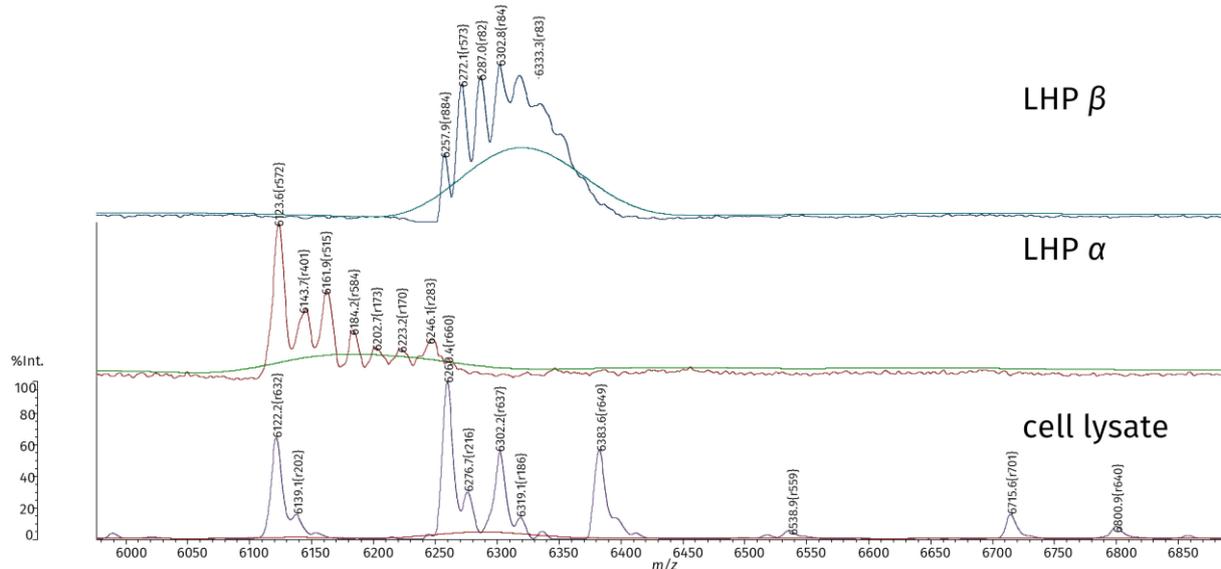
Figure 2.4-3 | *Afifella marina* DSM 2698 purified LH β, LH α and whole cell WFEX lysate Da in positive linear mode (top to bottom). Storage of the LH fractions at –80 °C for 26 years has led to oxidation of the proteins whereas the intact cells have been freshly lysed and show less modification.

Protein extraction protocols

FAEX protocol was not further considered since the MALDI-MS peaks were indicating degradation (formylation) of the proteins and nanoHPLC chromatograms showed only few signals. Differential extraction of proteins was tested on str. Cad16^T were WFEX showed different m/z profile than SDCEX. Masses between 6'000–8'000 Da were better recovered with the WFEX protocol. SDCEX resulted in more intense signals in the range 8'300–10'500 m/z (Figure 2.4-4). SDCEX produced more MALDI peaks and was therefore further used to produces lysates for subsequent nanoHPLC separation.
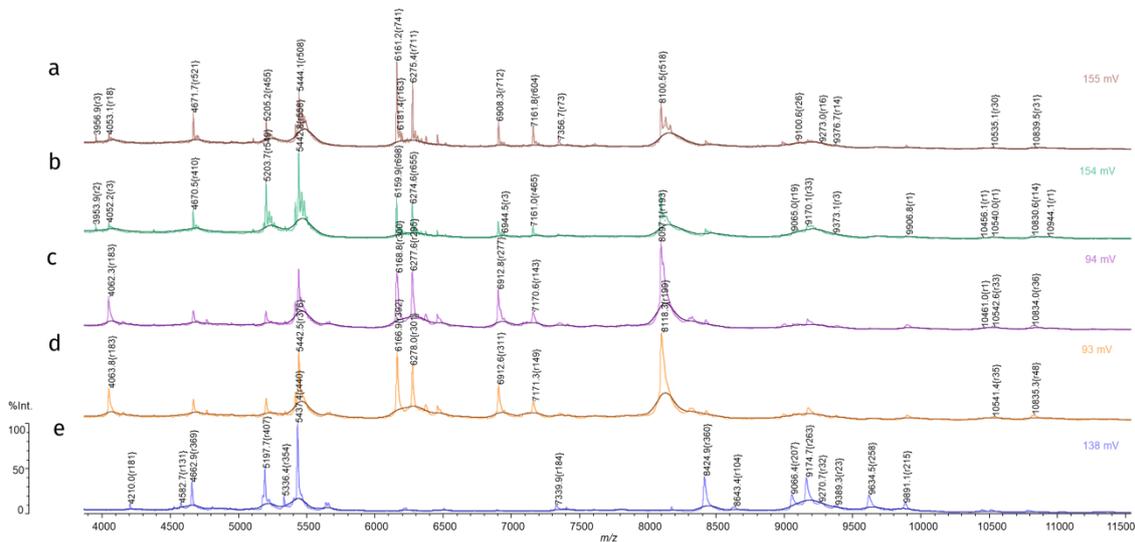
Figure 2.4-4 | MALDI linear positive spectra of '*Thiodictyon syntrophicum*' str. Cad16^T a-d) technical replicates of WFEX lysates. e) SDCEX lysate

The WFEX and SDCEXprotein extraction methods were also compared in *Afifella marina* DSM 2698 (Figure 2.4-5).
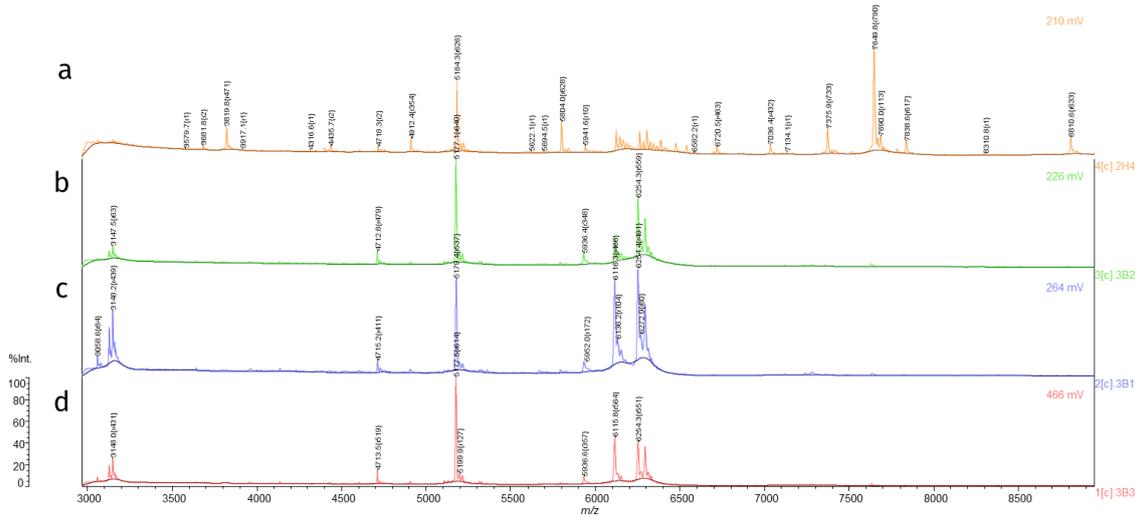
**Figure 2.4-5 | Linear positive MALDI-MS spectra from *Afifella marina* DSM 2698 a) MALDI SDCEX lysate b-d) Spectra of WFEX lysate replicates of *Afifella marina* DSM 2698. The LH antenna proteins are enriched with this method compared to a).**

NanoHPLC coupled to MALDI-MS^2 CID

Complete SDCEX lysates were separated with C18-RP DI plate (Figure 2.4-6) and collected by spotting 100 nl fractions directly on a 384 MALDI plate (min 20–60; about 100–120 fractions). When analyzed in linear MALDI-MS, the protein fractions were imperfectly separated as peaks at around 5'400 were found in >10 consecutive fractions (data not shown).
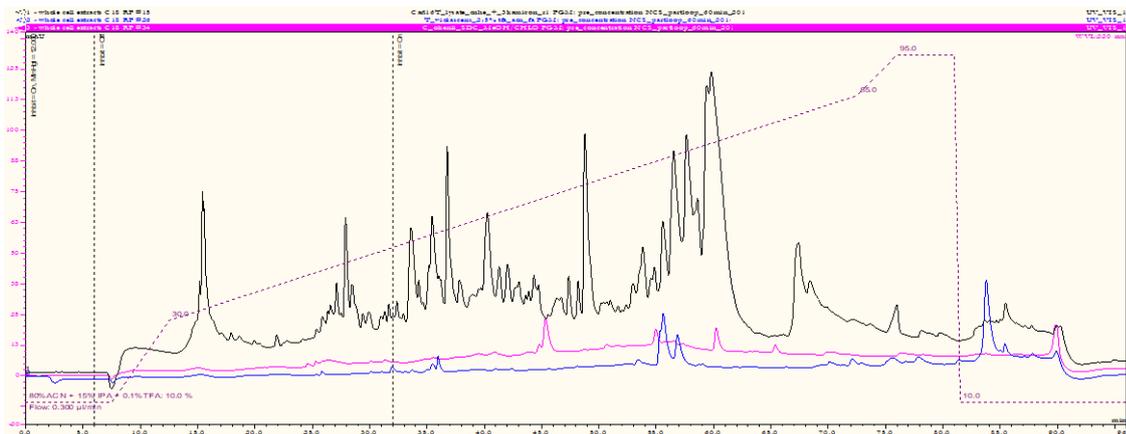
Figure 2.4-6 | Typical chromatograms of SDCEX lysates from different PSB '*Thiodictyon syntrophicum*' str. Cad16^T (black), *Chromatium okenii* (pink) and *Thiocystis violascens* (blue). Fractions of 100 nl were taken between 20–60 min elution time. Gradient is depicted as dashed red line.

On-plate digestion was tested with BSA before (results not shown) and Cad16^T fractions successfully digested in 30 min at 37 °C at 97 % humidity. Protein identification by MALDI-MS PMF and subsequent MS^2 peptide sequencing was performed in the positive reflectron

mode. Thereby we identified 21 proteins with PMF and of 12 the identity was confirmed by de novo peptide sequencing (Table 2.4-1).

Table 2.4-1 | List of proteins identified for '*Thiodictyon syntrophicum*' str. Cad16ᵀ in the SDCEX protocol measured with the AXIMA™ performance MS² set-up. A custom str. Cad16 ᵀ database was imported into MASCOT. MALDI-MS was performed in reflectron linear mode with and monoisotopic MH+ ions were identified at 0.3 Da with M-oxidation as variable modification with 1 missed trypsin cleavage. MS² was performed at 0.5 Da average mass tolerance, with a fragment mass tolerance of ±1 Da and 1+ charged peptides with 1 missed trypsin cleavage.

GenBank accession	Avg. MW [Da]; mature chain	Name
AUB79572.1	4'560.21	hypothetical protein
AUB79612.1	13'122.38	transcriptional regulator
AUB79908.1	20'460.25	hypothetical protein
AUB80110.1	19'641.60	pilus assembly protein
AUB80540.1	12'984.84	sulfur globule protein CV1
AUB80653.1	93'73.30	hypothetical protein
AUB80827.1	15'471.00	H-NS histone family protein
AUB81326.1	19'760.00	OmpA family protein
AUB82176.1	16'245.12	hypothetical protein
AUB82414.1	10'477.00	DUF4258 domain-containing protein
AUB82781.1	18'912.88	EF-hand domain-containing protein
AUB83362.1	13'662.00	diacylglycerol kinase
AUB83455.1	12'648.00	TusE/DsrC/DsvC family sulfur relay protein
AUB83885.1	37'394.00	hypothetical protein
AUB83947.1	8'964.00	DUF465 domain-containing protein
AUB84191.1	10'176.00	RnfH family protein
AUB85046.1	12'052.00	putative BrnA_antitoxin
AUB85355.1	9'039.00	light-harvesting protein
AUB85357.1	8'934.00	light-harvesting protein
AUB85363.1	9'626.46	light-harvesting protein
AUB85710.1	11'392.00	light-harvesting protein

To verify the findings with the AXIMA performance MALDI we used the rapifleX system to analyze on-plate digested fractions of Cad16ᵀ. MASCOT search of the str. Cad16ᵀ database resulted in 10 unique, 5 multiple and 10 uncertain identification from totally 50 fractions. Fractions were not well separated as indicated in the m/z versus fraction graph (Figure 2.4-7a). Proteins identified were YeaQ/YmgE of the transglycosylase-associated protein family(AUB84949.1), a type II toxin-antitoxin system VapC family toxin (AUB83009.1), the DNA polymerase subunit beta (AUB80723.1) and a helicase SNF2 (AUB81621.1). LH1 PufAB peptides have been also identified multile times, however some sequence ambiguity occurred (Figure 2.4-7 b). Alignment of all LH antenna proteins of str. Cad16ᵀ shows the hydrophobic core sequence that is conserved and leads to sequence ambiguity (supplementary Figure S 2.4-1).
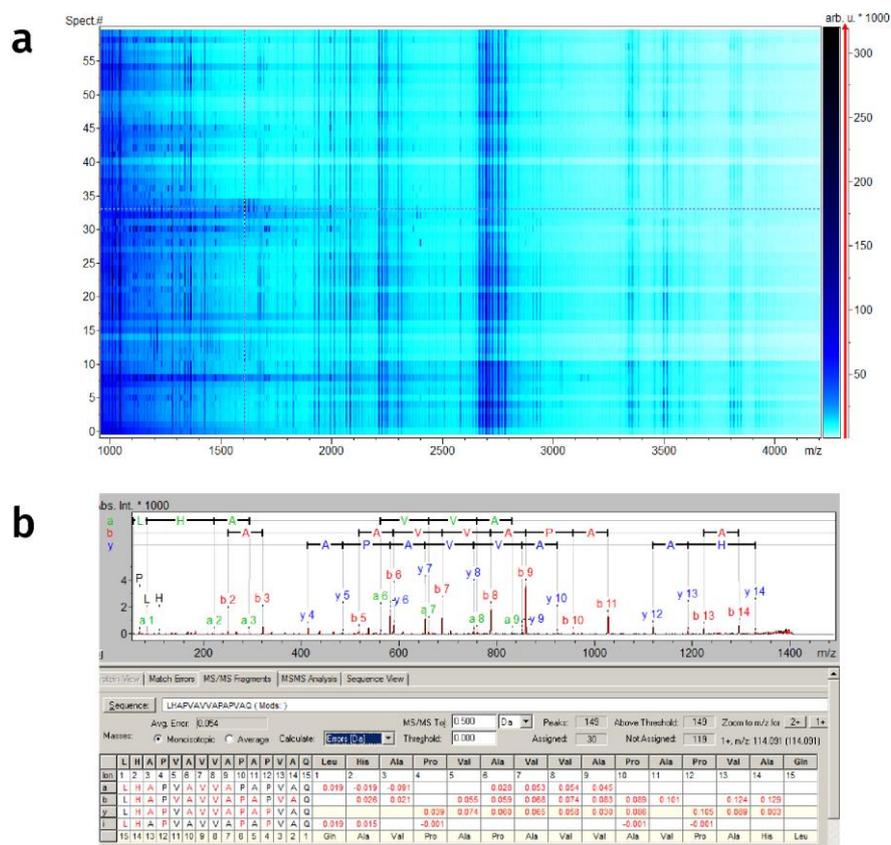
Figure 2.4-7 | CID experiments plots a) Chromatographic fractions versus CID fragments plot shows incomplete separation since CDI fragments are similar between fractions analyzed. b) Detailed report on a fragment of an LH antenna peptides AUB85355.1 and AUB85357.1

NanoHPLC coupled to MALDI-MS^2 TDS

The potential for protein TDS was successfully tested on the Axima MALDI with 1 mg/ml BSA (data not shown), however we were unable to demonstrate TDS for LC-fractions of bacteria.

For PSB or PNSB species we were unable to obtain positive TDS identification. However we could identify the OmpA membrane protein of *Y. pseudotuberculosis* with a sequence coverage of 17 % (Figure 2.4-8).
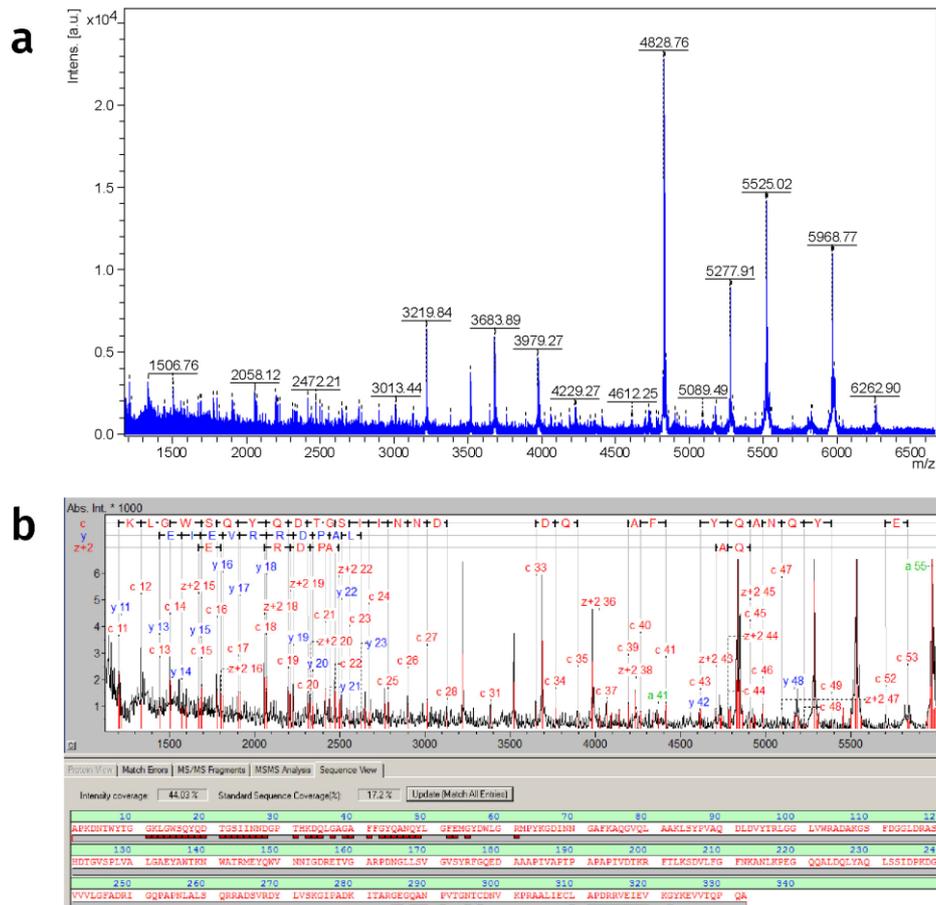
Figure 2.4-8 | TDS experimental plots a) TDS spectra of *Yersinia pseudotuberculosis* nanoHPLC fraction with the *de novo* peptide sequence analysis depicted in b)

2.4.4 Discussion

With the technical note above we laid the groundwork the application of MALDI-MS biotyping of environmental isolates of PSB and PNSB. Direct MALDI-MS analysis revealed rich spectra in the m/z range 2'000–30'000 for both PSB and PSNB. Long term storage at -80 °C had no impact on the quality of the spectra, as showed with PSNB frozen cell stocks.

Membrane bound and hydrophobic proteins were successfully detected for PSB str. Cad16^T when using the bottom-up approach. The masses of the proteins detected were within the MALDI-MS biotyping window from 2'000–30'000 m/z. The LH antenna pigments were highly abundant in all extraction methods tested and were successfully *de novo* sequenced for PSB str. Cad16T.

As nanoHPLC-MALDI-MS TDS was successful with *Y. pseudotuberculosis* we will further adapt the chromatographic set up and use C4 and Solid core columns to further optimize

separation of hydrophobic proteins. With the future availability of a rapifleX system, we will test further settings for biomarker validation of PSB and PNSB. The lower detection limit for pure proteins is thereby ~3 pg μl^{-1} (R. Brunisholz, personal communication) and lies within the loading limits of the nanoHPLC. In a next step we will integrate putative LH-CDS into the PAPMID™ biomarker protein database [248].

Acknowledgements

We want to specially thank Valentin Pflüger from mabritec AG and Nadia Schürch from the Spiez Laboratory for their constant support of this project.

2.4.5 Supplementary data
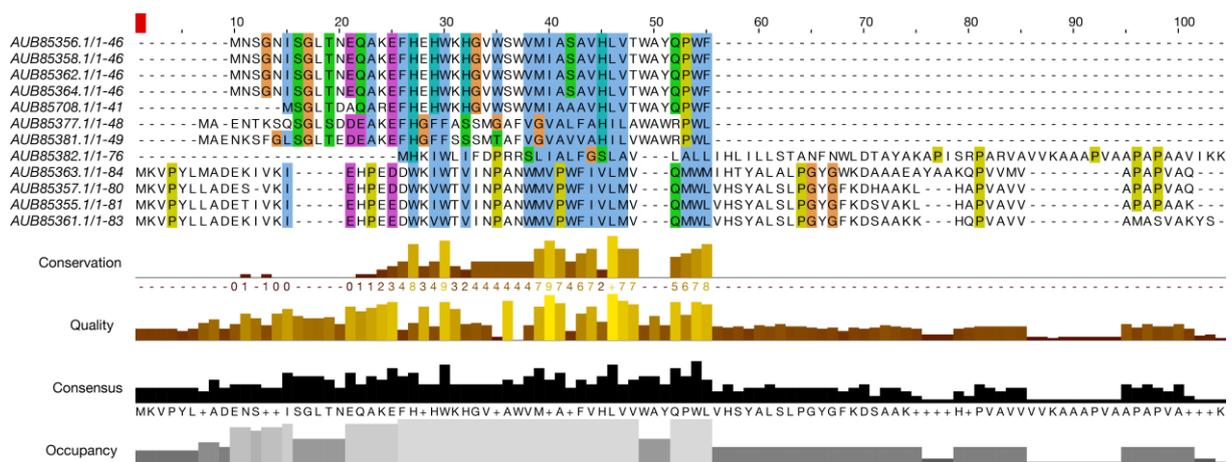
Figure S 2.4-1 | Sequence alignment of the LH proteins from '*Thiodictyon syntrophicum*' str. Cad16^T shows conserved membrane spanning core motif.
PufAB alignment was done with CLUSTAL omega v1.2.4 **[388]**. A conserved hydrophobic core is common among all LH-peptides (blue). CLUSTAL color scheme is found at: http://www.jalview.org/help/html/colourSchemes/clustal.html. Conservation: conservation of the overall alignment with less then 25% gaps, Quality: alignment quality based on Blossum62 score, Consensus: percentage identity; PID, Occupancy: Number of aligned positions

3   DISCUSSION AND OUTLOOK

In perspective of Global Change, the anthropogenic influence on aquatic environments has gained renewed interest in microbial ecology. Ocean warming in combination with eutrophication may lead to more frequent occurrences of algal blooms [389]. As a consequence, the increase in anaerobic biomass degradation stimulates euxinia [390] and results in expanding oxygen deficient zones (OMZs) [391]. Due to the accessibility and detailed scientific record spanning from biogeochemistry to biophysics and microbial ecology, the meromictic Lake Cadagno therefore remains an excellent study site for euxinic microbial communities. We therefore have to bring the 'Omics' to the field.

In this thesis I describe the *de novo* whole genome sequencing of '*T. syntrophicum*' str. Cad16^T and *C. okenii* str. LaCa, two key microbial species for the Lake Cadagno euxinic chemocline. This led to insight into the encoded metabolic pathways and structures and allowed for the comparison thereof between different PSOB isolates from Lake Cadagno. I then used the genomic information to study the *in situ* carbon and energy metabolism of str. Cad16^T by functional proteomics. I additionally demonstrated the use of genomic information to generate putative biomarkers for MALDI-MS based microbial typing.

3.1   What did we learn from PSB genomics?

In a first publication, we presented the complete genome of str. Cad16^T in detail and elaborate on the encoded functions and on the genome architecture, which is described in the section 2.1. We thereby found conserved sulfur, carbon and energy pathways similar to other PSB studied such as *A. vinosum*. We showed the capacity of str. Cad16^T for oxidative respiration by the encoded cbb3 type cytochrome *c* oxidase, carboxysome like structures and a form I RuBisCO. Interestingly however, we did not find enzymes for thiosulfate oxidation that would allow for mixotrophic growth. Notably, for the time in PSB, we found the complete photosynthesis, and parts of the pigment biosynthesis pathways, encoded on a 0.5 Mb plasmid. This may facilitate horizontal gene transfer within and between PSB populations. Indirect evidence for phage infection was given by the defensive CRISPR-Cas elements found. Recent re-sequencing of str. Cad16^T utilizing the MinION nanopore sequence technology confirmed the genome structure of str. Cad16^T comprising a 6.8 Mb chromosome and two plasmids with 0.5 and 0.4 Mb, respectively (Michela Ruinelli, pers. commun.).

We further studied the genomic potential of PSB isolates in context of the microbial community. Therefore, the PSB large-celled *C. okenii* str. LaCa was sequenced *de novo* using long-read technology. The problematic of inserts and deletions (INDELs) in long-read sequencing data [392] may have led to the many pseudo genes in the *C. okenii* annotation and large repetitive sequences may be present, hindering genome closure. In section 2.2 we gave a detailed record on the biological functions encoded and connected them to previous experimental findings on metabolism and mobility. As the pathways found are confirming the strict photoautotrophic metabolism describe for *C. okenii*, we did a core genome analysis with more metabolically versatile small-celled PSB and photoautotrophic GSB species found in Lake Cadagno, respectively. Based on these findings, we pointed out the relative importance of mobility, taxis and the bacterial S-layer—providing both stability and possibly protection from phage infection and bacterial *predation*—for *C. okenii* str. LaCa, as these functions are overrepresented in the genome. The PSB genomes analyzed mainly differ in the chemotrophic energy pathways and the structures allowing for passive or active movement. Our genomic findings reflected experimental data previously obtained, where in general, large celled PSB were shown to be metabolically less flexible than small-celled representatives [79]

Interestingly, the *Chromatiaceae* spp. sequenced so far show genome sizes in the range of 3–7 Mb, that puts the 6.8 Mb genome of str. Cad16^T at the upper, and the *C. okenii* LaCa with 3.8 Mb at the lower egde, respectively. Interestingly, despite the c. two-fold difference in the genes encoded, we did not find major variability in the metabolic pathways encoded between str. Cad16^T and *C. okenii* LaCa, leaving a large part of CDS without predicted function in the former species. That might be the consequence of a lack of reliable annotation due to missing related PSB sequences in the databases used. With the increasing number *Chromatiaceae* species that are being sequenced and experimented with however, functional annotation of PSB will further improve.

The evolutionary implication of genome architecture and size in prokaryotes has been discussed before [393, 394]. Factors such as nutrient limitation typically lead to genome streamlining—often accompanied with a decrease in cell size—, a phenomena prominently observed in marine SAR-type bacteria [395]. For Lake Cadagno, nutrient limitation might not be a strong selection factor as nutrient input is rather high. As data on the effective population size of PSOB is missing—assuming a large part of the PSOB being in a non-diving state—we

cannot further elucidate on the differential effects of selection and genetic drift on the PSOB studied.

Genetic accessibility is necessary to make knock-out or gain-of-function studies as demonstrated in *A. vinosum* [396]. As we found extrachromosomal elements in both, str. Cad16^T and *C. okenii* str. LaCa, we expect them to be naturally competent species. The transformation of Lake Cadagno isolates has not yet been attempted, but should be a priority in future functional studies. For *C. okenii* it would be especially useful to create knock-out strains deficient in motility and/or taxis. This would then allow studying the onset of bioconvection by external stimuli in great detail *in vitro*. Furthermore, GFP-fusion components of the TSS6 would be ideal to study their function in PSB isolates, as it has been done for *Vibrio* spp. [397]. Furthermore, stochastic processes in gene regulation within bacterial population was studied successfully in with single-cell microfluidics in combination with fluorescently labelled functional markers [398].

3.2   Beyond omics — The possible ecological consequences of cellular spatial organization and mobility pattern of large-celled *C. okenii*

We have described functions and structures related to active guided swimming and cell surface/ S-layer are overrepresented in the genome of the large-celled *C. okenii*. For *C. okenii* LaCa these traits might be of competitive character in the interplay with the chemocline community. Importantly, neither active movement nor cell size are traits that are *a priori* governed by gene regulation.

An increase in cell size imposes modifications onto cell organization, as the consequent increase in diffusion time of internal factors between source and target negatively influences metabolic rates and signaling speed [331]. As an example, multiple chromosomes in *Achromatium* spp. may allow a more efficient transcription and regulation of cell homeostasis in sectors [399]. As diffusion coefficients of molecules are in the range of ca 10^{-6} cm^2 s^{-1}, that results in 274 s traffic time—the average time during 'molecule X' meets any other molecule inside the cell—in a 6.5 μm wide *C. okenii* cell, which in turn may effectively slow down metabolic rates. Furthermore, the relative decrease of the surface-area-to-volume ratio diminishes the effective uptake and export rates of nutrients and effectors —i.e. leads to

diffusion limitation—, and the relative increase in internal capacity for redox or carbon storage is daunted by a decrease in buoyancy. For *C okenii* we calculate a surface-area-to-volume ranging from 0.7–0.9 μm^{-1} for rod -shaped cells that however might be forced to larges values in elongated 'cigar-shaped' cells as observed under sulfur starvation (own observations). Furthermore, the average cell density of *C. okenii* was seen to be stable and *in situ* not depending on active sulfur oxidation [148]. Collectively, these hypotheses have to be verified for *C. okenii* with experimental data.

The spatial organization and internal transport kinetics of (bacterial) cells has gained renewed interest as high-resolution microscopy technologies have been developed the last 30 years. Optical fluorescence microscopy is limited to a resolution of some 100 nm by the diffraction limit of light, but super-resolution microscopy now allows the detection of structures of ~10 nm diameter (reviewed by [400]) and surface structures can be studied at 1 nm resolution with atomic force microscopy (AFM) *in vivo* [401]. Finally, the high resolution cryo-tomography EM have led to the discovery of numerous novel (internal) prokaryotic structures at near atomic-resolution [402] and it would be favorable to use this techniques in order to understand *C okenii* cell architecture. Especially the study of SGB formation and structure would be interesting to study in *C. okenii*. The content of SGB could be further studied by energy dispersive X-ray (EDX) [403] and raman spectroscopy [404].

An accumulation of highly concentrated cells induces bioconvection [148] and *C. okenii* is thereby actively re-shaping the chemocline in Lake Cadagno. For PSB *C. minus* and *C. okenii* positive chemotaxis towards H$_2$S and negative chemotaxis towards O$_2$, and additionally, scotophotaxis was demonstrated [260, 322]. If bioconvection is used as a proxy to infer active swimming of *C. okenii*, the mechanisms responsible for the stochastic onset of this phenomena at day and night, as described by Sommer *et al* [148], as well as conditions that led to its ceasing have yet to be discovered. As both, diurnal [325] and circadian behavior, as well as swarm like swimming was found for *C okenii* [260], additional evidence thereof was given in this study by fragments of both a circadian and a quorum sensing signaling detected in the genome of *C. okenii* str. La Ca. Therefore, the multi-leveled molecular mechanisms which govern direction and coordinates collective movements in *C. okenii* have to further be studied, and the now available genome provides information of the potential thereof.

Swimming speed was shown to be regulated by different factors in *C. minus*. Importantly, the energy demand increases with the square of swimming speed [331]. Under low light intensity (10 μmol quanta$\cdot$m^{-2} s^{-1}) the addition of sulfide let to an increase in swimming speed whereas under light saturation (100 μmol quanta$\cdot$m^{-2} s^{-1}) the relative swimming speed was decreasing In accordance, periplasmic sulfur oxidation process producing H$^+$ that are coupled to the photosynthesis rate may directly fuel the flagellum motor. Importantly, the average *C. okenii* cell is around 10 μm long which allows to break out of the 'diffusion halo'[331] — a surrounding substrate deplete zone due to rapid diffusion at rate of 10^{-6} cm^2 s^{-1}— and thereby gains more nutrients through active swimming.

3.3 Carbon metabolism in 'T. syntrophicum' str. Cad16^T and the role of oxygen

In order to understand the carbon fixation and energy metabolism of str. Cad16^T in the dark beyond the 'potentiality', we used the genomic information as a base for functional proteomics *in situ*, as outlined in chapter 2.3. We combined longitudinal monitoring of the physicochemical conditions by CTD and FC during three months adaption with carbon-radioisotope assimilation and quantitative proteomics for the two endpoint conditions 'day' and 'night'. Monitoring revealed microaerobic conditions at the chemocline possibly due to the oxygenic photosynthesis of *cyanobacteria*. As mirrored in the proteome, the combination of oxygen and low-light intensity led to a physiological adaptation of str. Cad16^T to mixotrophic metabolism — photosynthesis competing on electrons with respiration of simple carbon compound such as acetate with sulfate as electron acceptor — during the day and chemotrophic growth at night, respectively.

The 10-fold lower C-assimilation rates found for str. Cad16^T compared to previous results [162], may indicate that inorganic CO_2 is mainly used to sustain the redox balance during aerobic growth. Within the literature on total carbon assimilation rates of Lake Cadagno chemocline we find large variability between the values obtained, with ~100× and ~300× fold differences for day and night, respectively. When we compared the radiocarbon techniques used, classical ^{14}C-scintillation gives relatively higher values than the MS-based ^{13}C-approach, partly because of lacking information on dark fixation during the day in the ^{13}C-method [405]. In accordance with our findings, variation between and within the radiocarbon

methods have been previously reported [405]. As both methods are well suited to detect trends in carbon fixation, comparisons between the techniques have to be judged critically.

With the methods applied, we could not further infer respiration and we did not measure dissolved organic carbon in the environment. Therefore, an approach as used by Kleiner and colleagues, that combines quantitative bottom-up proteomics and isotope ratio mass spectrometry (IRMS) of stable carbon isotope ratios (δ^{13}C) in peptides [406], would be more appropriate in order to disentangle the possible mixotrophism in str. Cad16^T. I therefore propose to further test metabolic activity of str. Cad16^T under aerobic conditions with and without light at prolonged time using stable isotope uptake analysis and mRNAseq and/or proteomics. The inclusion of sulfur-uptake kinetic studies, as done for *in vitro* incubations of GSB *Prosthecochloris* spp. enrichments of the Chesapeake Bay estuary [407] and PSB str. Cad16^T and *C. okenii* LaCa [408], will give further evidence of the relative importance of sulfur oxidation under aerobic conditions. Furthermore the potential for aerobic Fe oxidation of str. Cad 16^T has to be tested in vitro with the upper mentioned methods and the additional spectrophotometrical monitoring of Fe concentrations.

3.4 How can we improve our understanding of microbial communities in euxinia over space and time?

For future research on Lake Cadagno microbial ecology I propose to lay the focus the following topics; 1) Metagenomics of the chemocline and the adjacent bacterial matts in order to gain an in-depth understanding of the community in high taxonomic resolution, 2) functional omics that combine environmental monitoring, quantitation of mRNA and/or proteins and the metabolic activity on community or single-cell level, and 3) the application of advanced statistics and modeling of microbial interactions and feedback loops with the environment based on the data obtained (Figure 3.4-1).

Whereas the keystone species concept [409] provides evidence of universal taxa that provide basic biogeochemical processes and ecosystem functions, the complementary "rare biosphere" concept [410], based on genomic evidence on the enormous bacterial diversity, is reinforced by an increasing complex marine microbial census [411].
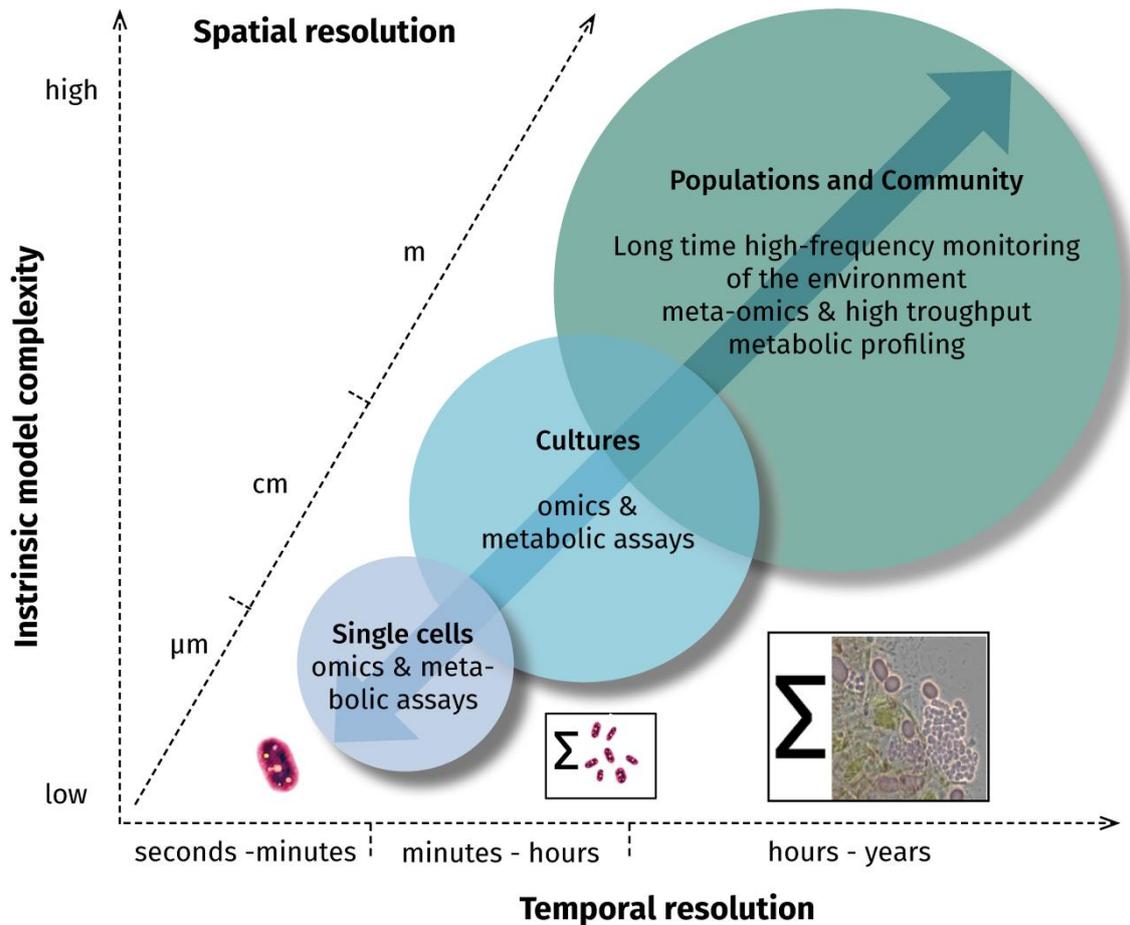
Figure 3.4-1| Schematic representation of the spatial and temporal dimensions faced by microbial ecology research on Lake Cadagno. As the size (e.g. volume and the microbes therein) and timeframes of the projects expand, the complexity of the research increases simultaneously.

As the study of bacterial taxonomy and community composition is driven beyond the canonical 16S rRNA standard towards WGS metagenomics, it is time to apply these developments in the research on Lake Cadgano. Several metagenomic surveys of stratified aquatic systems have been performed As examples, Hamilton and colleagues successfully reconstructed the microbial sulfur cycling of Mahoney Lake by metagenomic binning [202] and the thriving GSB population in Ace Lake was studied in great detail with meta-omics [95, 412]. The recent discovery of the more widespread distribution of *dsrAB* genes in heterotrophic marine bacteria such as *Verrucomicrobia* spp. [11], the fundamental role of the SAR202 cluster (*Chloroflexi*) in large- scale marine sulfite-oxidation [413], as well as the crucial role in sulfur cycling of the rare *Arcobacter* spp. in coastal waters [414] and the identification of freshwater *β*-proteobacteria *Sulfuricella denitrificans* skB26 and *Sulfuritalea hydrogenivorans* sk43H involved in sulfur cycling [415]. Collectively, these examples have shed a light on the increasing complexity of the microbial sulfur cycling. We therefore might also expect a more complex structure of the sulfur cycling in Lake Cadagno. In consequence, the interaction of PSOB with the abundant heterotrophic bacteria in the chemocline [143]

should be studied in more detail, as either competition for 'niche-spaces', or a more mutualistic interaction based on specialization and complementarity as in marine microbial networks [416], can be expected. As we found evidence on the possibility of horizontal gene transfer (HGT) between PSB species in Lake Cadagno, the concept of a 'pangenome'[417] and the ever-difficult definition of the microbial species concept [418, 419] demand for metagenomic approaches in order to understand distribution and redundancy biological functions encoded. Thereby, the rapid advances of 4th generation sequencing MinION will allow rapid, increasingly cheap and highly resolved metagenomic studies, and an already ongoing project on Lake Cadagno has been initiated in the path of this thesis.

Trophic-level interactions in the Lake Cadagno chemocline are possibly governed by bottom-up factors such as phage lysis and predation. As an example, the population of *C. okenii* of $\sim 10^5$ cells ml^{-1} suffers a sudden yearly collapse in October [141,156], a process possibly triggered by bottom-up effects as observed in cyanobacteria bloom collapse [420]. Bacterial predation was described for large-celled *Chromatium* spp. in Spanish Karst lakes [314] and epibionts were discovered on *C. okenii* of Lake Cadagno [163]. As we sequenced a community enriched for *C. okenii* —and thereby, also the epibionts attached —we expected to find genomic sequences from predatory bacteria. However, we did not find any matching sequences to known bacterial predators or scavengers. However, we find a series of putative (pro) phages in the *C. okenii* LaCa enrichment in September 2016. Collectively, these findings stress the importance for an in-depth metagenomic survey that includes bacteriophages.

As interaction in microbial communities is difficult to understand on the sole promise of the genomic 'potentiality' [421], a more integrating approach is needed in the follow up, including physicochemical measurements such as hydrological dynamics and nutrient concentrations and flux, as well as quantitation of metabolites and proteins expressed on community level.

In this thesis, we demonstrated the importance of regular active and passive monitoring during the *in situ* incubation experiments with str. Cad16[T]. The future use of automated sensory system such as VERTEX [422] could further facilitate high-frequency monitoring and will give additional information on the horizontal distribution of physicochemical conditions met by the microbes. Recent studies on bioconvection and community dynamics that involved

regular field measurements in a weekly to hourly frequency gave valuable insight to the physical and chemical bottom-up factors that influence the chemocline community growth characteristics over time [134, 148]. Ongoing studies on biophysical phenomena will provide even better time-resolved physiochemical vertical profiles of Lake Cadagno (O. Sepùlveda, pers. commun.). However, microbial physiology and behavior themselves exert changes in the physical and chemical environment and one might end up in a 'hen-and-egg'-like conundrum.

Therefore, the effect of external factors such as light and nutrient availability or hydrological dynamics on the microbial population will have to be disentangled from biotic factors carefully, since correlation of physicochemical factors with changes in cell numbers and complexity alone falls short of explaining functional interactions within the community and the effect of top-down factors, such as viral lysis and predation, affecting total microbial biomass distribution on the taxa involved. To further produce evidence for cause–effect linkages between abiotic factors and microbial community composition in Lake Cadagno, we need to design more conclusive functional assays

As we described for str. Cad16[T], biological functions are best studied in a combination of omics-technologies in combination with testing for metabolic activity *in vitro*. This concept can be differently scaled depending on the research question, as recent omics-technology and the development of automated analysis tools allow for both, the study of numerous individual cells within communities and the description of entire communities with unseen resolution.

The progress in single-cell sequencing [423] might soon enable to mirror some of the phenotypic heterogeneity observed in bacterial populations [424] on genomic level and the development of FACS coupled to single-cell RNA-sequencing [425], nanoSIMS [166] and, possibly, single-cell MALDI-MS [426] will allow to trace short-time adaptions in gene regulation, translation and metabolic activity, respectively. As ongoing large scale metaproteomic projects on OMZs [427] demonstrate the recent advance in metaproteomics, it would be interesting to use this approach also in Lake Cadagno. A first metaproteomic study on the GSB dominated chemocline in 2008 —a then relatively simple community—revealed photosynthetic carbon fixation and dark fermentative activity. It would therefore be interesting to compare these findings with the current, more 'rich' community composition which resembles the one describe pre-2000 era [365].

The littoral bacterial-matt community in Lake Cadagno has not been studied in detail so far. A 16S rRNA based consensus of twelve samples from matts around different sublacustrin sulfate-rich sources (see Figure 1.4-1d) revealed a complex community dominated by oxygenic phototrophic Phormidiaceae (*cyanobacteria*), and (potentially) sulfur oxidizing anoxygenic bacteria of the families Chlorobiaceae, Helicobacteraceae (*ε-proteobacteria*), Chlorothrixaceae (Chloroflexi) and the Comamonadaceae (*β-proteobacteria*) [143]. Chromatiaceae spp. were also detected in most of these samples and it would be interesting to isolate and sequence these strains, and /or do a metagenomic study, in order to compare them to chemocline community. We could expect genomic and physiological adaption to a life at sharp gradients of light availability, O_2 and H_2S. Meta-transcriptomic analysis in combination with microsensors at light and dark conditions could additionally reveal sulfur and carbon metabolisms driven by photosynthesis and chemotrophic growth as suggested by Ravasi [143].

Long-time survival of Lake Cadagno PSBs during the all-winter freezing periods has been studied only patchily and a PSB have been thought to re-emerge from 'seed-banks' in the sediment, or secondly from the overwintering pelagic chemocline community [161, 155]. This hypothesis is confirmed by the account on the successful reviving some years old PSB cultures that were stored at 4°C in the dark in our lab (Samuele Roman, per. commun). As a third option, bacterial matts could possibly function as save harbors for the chemocline population during winter, since light availability is higher than in the lower sulfidic water column. With *cyanobacteria* present at the epilimnic zone, we might further expect local oxygenation that could fuel chemotrophic metabolism in PSB when light availability is limited. In spring, PSB would then laterally invade the water column.

Despite the relative stability of the creogenic meromixis in Lake Cadagno, regime changes in the PSOB community have been observed regularly. On one hand, the Lake Cadagno sediment bares a historic record dating back to 10'000 years from now, on how avalanches and landslides have affected the microbial community composition for the following decades [125]. On the other hand, more recent GSB *C. clathratiforme* blooms possibly induced by hydrological mixing [154], the seasonal PSOB succession and the autumnal *C. okenii* population collapse [141,156], as well as the unusually strong presence of oxygenic cyanobacteria in the chemocline in 2016 and 2017 [134] gave additional evidence of ongoing regime-changes. In the recent research on regime changes in the seasonally stratified Lake

Vechten, modelling which included physicochemical data and regular 16S rRNA gene sequencing throughout the season, was used to define tripping points of non-reversible changes from oxic-to-anoxic states [100, 428]. With the increasing amount of highly resolved data on both, physicochemical conditions and microbial community composition over space and time, an adapted modelling and prediction of feedbacks between abiotic processes and microbial community dynamics will be possible for the Lake Cadagno system.

Coming back to the initial claim on the importance of research on Lake Cadagno for Global change driven increase in OMZs, this ecosystem itself will be affected by changes. As Global Change may lead to a further increase in average temperatures during winter in the Alps [429], we expect to see shorter freezing periods for Lake Cadagno. Additionally, more intense rainfall events are projected that in turn, may increase landslide-frequency [430]. In consequence, both phenomena may lead to a destabilization of the meromixis. Taken together, the fascinating microbial community in Lake Cadagno will face new challenges ahead that will leave characteristic traces in their ongoing record of existence, which will be as valuable for future researchers, as it has already been proven to be for the past research of the last 100 years.

4 REFERENCES

1. Falkowski PG, Fenchel T, Delong EF. The Microbial Engines That Drive Earth's Biogeochemical Cycles. Science. 2008;320:1034–9.

2. Whitman WB, Coleman DC, Wiebe WJ. Prokaryotes: The unseen majority. Proc Natl Acad Sci. 1998;95:6578–83.

3. Fry JC. In: Austin B, editor. Methods in aquatic bacteriology. Chichester; New York: Wiley; 1988. p. 27–72.

4. Ducklow HW, Carlson CA. Oceanic Bacterial Production. In: Advances in Microbial Ecology. Springer, Boston, MA; 1992. p. 113–81. doi:10.1007/978-1-4684-7609-5_3.

5. Williams RJP. The natural selection of the chemical elements. Cell Mol Life Sci CMLS. 1997;53:816–29.

6. Smith JM, Szathmáry E. The major transitions in evolution. Oxford [etc: Oxford Univ. Press; 2010.

7. Overmann J, Garcia-Pichel F. The Phototrophic Way of Life. In: Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E, editors. The Prokaryotes. New York, NY: Springer New York; 2006. p. 32–85. doi:10.1007/0-387-30742-7_3.

8. Hanson TE, Alber BE, Tabita FR. Phototrophic CO2 Fixation: Recent Insights into Ancient Metabolisms. In: Burnap R, Vermaas W, editors. Functional Genomics and Evolution of Photosynthetic Systems. Dordrecht: Springer Netherlands; 2012. p. 225–51. http://link.springer.com/10.1007/978-94-007-1533-2_9. Accessed 22 Mar 2016.

9. Kaplan IR, Rittenberg SC. MICROBIOLOGICAL FRACTIONATION OF SULPHUR ISOTOPES. J Gen Microbiol. 1964;34:195–212.

10. Jørgensen BB, Kasten S. Sulfur Cycling and Methane Oxidation. In: Marine Geochemistry. Springer, Berlin, Heidelberg; 2006. p. 271–309. doi:10.1007/3-540-32144-6_8.

11. Anantharaman K, Hausmann B, Jungbluth SP, Kantor RS, Lavy A, Warren LA, et al. Expanded diversity of microbial groups that shape the dissimilatory sulfur cycle. ISME J. 2018;:1.

12. Rosenberg E. The prokaryotes: *deltaproteobacteria* and *epsilonproteobacteria*. 4th edition. New York: Springer; 2014.

13. Bak F, Cypionka H. A novel type of energy metabolism involving fermentation of inorganic sulphur compounds. Nature. 1987;326:891–2.

14. Krämer M, Cypionka H. Sulfate formation via ATP sulfurylase in thiosulfate- and sulfite-disproportionating bacteria. Arch Microbiol. 1989;151:232–7.

15. Thamdrup B, Finster K, Hansen JW, Bak F. Bacterial Disproportionation of Elemental Sulfur Coupled to Chemical Reduction of Iron or Manganese. Appl Environ Microbiol. 1993;59:101–8.

16. Pfennig N. The phototrophic bacteria and their role in the sulfur cycle. Plant Soil. 1975;43:1–16.

17. Middelburg J. The geochemical sulfur cycle. In: Lens P, Hulshoff PW, editors. Environmental technologies to treat sulfur pollution. London: IWA Publishing; 2000. p. 33–46.

18. Holland HD. Systematics of the isotopic composition of sulfur in the oceans during the Phanerozoic and its implications for atmospheric oxygen. Geochim Cosmochim Acta. 1973;37:2605–16.

19. Ohtomo Y, Kakegawa T, Ishida A, Nagase T, Rosing MT. Evidence for biogenic graphite in early Archaean Isua metasedimentary rocks. Nat Geosci Nat Geosci. 2013;7:25–8.

20. Philippot P, Zuilen MV, Lepot K, Thomazo C, Farquhar J, Kranendonk MJV. Early Archaean Microorganisms Preferred Elemental Sulfur, Not Sulfate. Science. 2007;317:1534–7.

21. Shen Y, Buick R, Canfield DE. Isotopic evidence for microbial sulphate reduction in the early Archaean era. Nature. 2001;410:77–81.

22. Rosing MT, Frei R. U-rich Archaean sea-floor sediments from Greenland – indications of >3700 Ma oxygenic photosynthesis. Earth Planet Sci Lett. 2004;217:237–44.

23. Blank Carrine E. Evolutionary timing of the origins of mesophilic sulphate reduction and oxygenic photosynthesis: a phylogenomic dating approach. Geobiology. 2004;2:1–20.

24. Canfield D. A new model for proterozoic ocean chemistry. Nature. 1998;396.

25. Brocks JJ, Love GD, Summons RE, Knoll AH, Logan GA, Bowden SA. Biomarker evidence for green and purple sulphur bacteria in a stratified Palaeoproterozoic sea. Nature. 2005;437:866–70.

26. Brocks JJ, Schaeffer P. Okenane, a biomarker for purple sulfur bacteria (*Chromatiaceae*), and other new carotenoid derivatives from the 1640 Ma Barney Creek Formation. Geochim Cosmochim Acta. 2008;72:1396–414.

27. Douglas AE, Raven JA. Genomes at the interface between bacteria and organelles. Philos Trans R Soc B Biol Sci. 2003;358:5–518.

28. Raymond J, Segrè D. The Effect of Oxygen on Biochemical Networks and the Evolution of Complex Life. Science. 2006;311:1764.

29. Towe KM. Oxygen-Collagen Priority and the Early Metazoan Fossil Record. Proc Natl Acad Sci. 1970;65:781–8.

30. Bowring SA, Grotzinger JP, Isachsen CE, Knoll AH. Calibrating Rates of Early Cambrian Evolution. Sci -N Y THEN Wash-. 1993;261:1293.

31. Canfield DE, Farquhar J. Animal evolution, bioturbation, and the sulfate concentration of the oceans. Proc Natl Acad Sci. 2009;106:8123–7.

32. Shih PM. Photosynthesis and early Earth. Curr Biol. 2015;25:R855–9.

33. Cardona T. A fresh look at the evolution and diversification of photochemical reaction centers. Photosynth Res. 2014;126:111–34.

34. 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. 52:333–41.

35. Nagashima KVP, Hiraishi A, Shimada K, Matsuura K. Horizontal transfer of genes coding for the photosynthetic reaction centers of purple bacteria. J Mol Evol. 45:131–6.

36. Brune DC. Sulfur oxidation by phototrophic bacteria. Biochim Biophys Acta. 1989;:189–221.

37. Tang K-H, Tang YJ, Blankenship RE. Carbon Metabolic Pathways in Phototrophic Bacteria and Their Broader Evolutionary Implications. Front Microbiol. 2011;2. doi:10.3389/fmicb.2011.00165.

38. Pfennig N. General physiology and ecology of phototrophic bacteria. Photosynth Bact. 1978;:3–18.

39. Dahl C. Sulfur Metabolism in Phototrophic Bacteria. In: Hallenbeck PC, editor. Modern Topics in the Phototrophic Prokaryotes. Springer International Publishing; 2017. p. 27–66. doi:10.1007/978-3-319-51365-2_2.

40. Ehrenberg CG. Die Infusionsthierchen als vollkommene Organismen Atlas. Atlas. Leipzig: Voss; 1838.

41. Lang HP, Hunter CN. The relationship between carotenoid biosynthesis and the assembly of the light-harvesting LH2 complex in *Rhodobacter sphaeroides*. Biochem J. 1994;298 (Pt 1):197–205.

42. Polívka T, Frank HA. Molecular Factors Controlling Photosynthetic Light Harvesting by Carotenoids. Acc Chem Res. 2010;43:1125–34.

43. Magdaong NCM, Niedzwiedzki DM, Goodson C, Blankenship RE. Carotenoid-to-Bacteriochlorophyll Energy Transfer in the LH1–RC Core Complex of a Bacteriochlorophyll *b* Containing Purple Photosynthetic Bacterium *Blastochloris viridis*. J Phys Chem B. 2016;120:5159–71.

44. Robert B, Cogdell RJ, Grondelle R van. The Light-Harvesting System of Purple Bacteria. In: Light-Harvesting Antennas in Photosynthesis. Springer, Dordrecht; 2003. p. 169–94. doi:10.1007/978-94-017-2087-8_5.

45. Young CS, Beatty JT. Multi-level Regulation of Purple Bacterial Light-harvesting Complexes. In: Light-Harvesting Antennas in Photosynthesis. Springer, Dordrecht; 2003. p. 449–70. doi:10.1007/978-94-017-2087-8_16.

46. Kereïche S, Bourinet L, Keegstra W, Arteni AA, Verbavatz J-M, Boekema EJ, et al. The peripheral light-harvesting complexes from purple sulfur bacteria have different 'ring' sizes. FEBS Lett. 2008;582:3650–6.

47. Nitschke W, Jubault-Bregler M, Rutherford AW. The reaction center associated tetraheme cytochrome subunit from *Chromatium vinosum* revisited: A reexamination of its EPR properties. Biochemistry (Mosc). 1993;32:8871–9.

48. Frigaard N-U, Dahl C. Sulfur Metabolism in Phototrophic Sulfur Bacteria. In: Poole RK, editor. Advances in Microbial Physiology. Academic Press; 2008. p. 103–200. http://www.sciencedirect.com/science/article/pii/S0065291108000027. Accessed 16 Mar 2016.

49. Verméglio A, Li J, Schoepp-Cothenet B, Pratt N, Knaff DB. The Role of High-Potential Iron Protein and Cytochrome c_8 as Alternative Electron Donors to the

Reaction Center of *Chromatium vinosum* [†]. Biochemistry (Mosc). 2002;41:8868–75.

50. Widdel F, Schnell S, Heising S, Ehrenreich A, Assmus B, Schink B. Ferrous iron oxidation by anoxygenic phototrophic bacteria. Nature. 1993;362:834–6.

51. Schütz M, Maldener I, Griesbeck C, Hauska G. Sulfide-Quinone Reductase from *Rhodobacter capsulatus*: Requirement for Growth, Periplasmic Localization, and Extension of Gene Sequence Analysis. J Bacteriol. 1999;181:6516–23.

52. Griesbeck C, Schutz M, Schodl T, Bathe S, Nausch L, Mederer N. Mechanism of sulfide-quinone reductase investigated using site-directed mutagenesis and sulfur analysis. Biochemistry (Mosc). 2002;41:11552–65.

53. Meyer TE, Cusanovich MA. Discovery and characterization of electron transfer proteins in the photosynthetic bacteria. In: Govindjee, Beatty JT, Gest H, Allen JF, editors. Discoveries in Photosynthesis. Berlin/Heidelberg: Springer-Verlag; 2005. p. 455–70. doi:10.1007/1-4020-3324-9_44.

54. Pattaragulwanit K, Brune DC, Trüper HG, Dahl C. Molecular genetic evidence for extracytoplasmic localization of sulfur globules in *Chromatium vinosum*. Arch Microbiol. 1998;169:434–44.

55. Friedrich CG, Rother D, Bardischewsky F, Quentmeier A, Fischer J. Oxidation of Reduced Inorganic Sulfur Compounds by Bacteria: Emergence of a Common Mechanism? Appl Environ Microbiol. 2001;67:2873–82.

56. Sauvé V, Bruno S, Berks BC, Hemmings AM. The SoxYZ Complex Carries Sulfur Cycle Intermediates on a Peptide Swinging Arm. J Biol Chem. 2007;282:23194–204.

57. Ogawa T, Furusawa T, Nomura R, Seo D, Hosoya-Matsuda N, Sakurai H, et al. SoxAX Binding Protein, a Novel Component of the Thiosulfate-Oxidizing Multienzyme System in the Green Sulfur Bacterium *Chlorobium tepidum*. J Bacteriol. 2008;190:6097–110.

58. Welte C, Hafner S, Krätzer C, Quentmeier A, Friedrich CG, Dahl C. Interaction between Sox proteins of two physiologically distinct bacteria and a new protein involved in thiosulfate oxidation. FEBS Lett. 583:1281–6.

59. Hensen D, Sperling D, Trüper HG, Brune DC, Dahl C. Thiosulphate oxidation in the phototrophic sulphur bacterium *Allochromatium vinosum*. Mol Microbiol. 2006;62:794–810.

60. Brito JA, Denkmann K, Pereira IAC, Archer M, Dahl C. Thiosulfate Dehydrogenase (TsdA) from *Allochromatium vinosum* STRUCTURAL AND FUNCTIONAL INSIGHTS INTO THIOSULFATE OXIDATION. J Biol Chem. 2015;290:9222–38.

61. Denkmann K, Grein F, Zigann R, Siemen A, Bergmann J, van Helmont S, et al. Thiosulfate dehydrogenase: a widespread unusual acidophilic *c*-type cytochrome. Environ Microbiol. 2012;14:2673–88.

62. Franz B, Lichtenberg H, Dahl C, Hormes J, Prange A. Utilization of "elemental" sulfur by different phototrophic sulfur bacteria (*Chromatiaceae, Ectothiorhodospiraceae*): A sulfur K-edge XANES spectroscopy study. J Phys Conf Ser. 2009;190:012200.

63. Stockdreher Y, Venceslau SS, Josten M, Sahl H-G, Pereira IAC, Dahl C. Cytoplasmic Sulfurtransferases in the Purple Sulfur Bacterium *Allochromatium vinosum*: Evidence for Sulfur Transfer from DsrEFH to DsrC. PLoS ONE. 2012;7:e40785.

64. Stockdreher Y, Sturm M, Josten M, Sahl H-G, Dobler N, Zigann R, et al. New Proteins Involved in Sulfur Trafficking in the Cytoplasm of *Allochromatium vinosum*. J Biol Chem. 2014.

65. Liu L-J, Stockdreher Y, Koch T, Sun S-T, Fan Z, Josten M, et al. Thiosulfate Transfer Mediated by DsrE/TusA Homologs from Acidothermophilic Sulfur-oxidizing Archaeon *Metallosphaera cuprina*. J Biol Chem. 2014;289:26949–59.

66. Dahl C, Schulte A, Stockdreher Y, Hong C, Grimm F, Sander J, et al. Structural and Molecular Genetic Insight into a Widespread Sulfur Oxidation Pathway. J Mol Biol. 2008;384:1287–300.

67. Venceslau SS, Stockdreher Y, Dahl C, Pereira IAC. The "bacterial heterodisulfide" DsrC is a key protein in dissimilatory sulfur metabolism. Biochim Biophys Acta BBA - Bioenerg. 2014;1837:1148–64.

68. Dahl C, Engels S, Pott-Sperling AS, Schulte A, Sander J, Lübbe Y, et al. Novel Genes of the dsr Gene Cluster and Evidence for Close Interaction of Dsr Proteins during Sulfur Oxidation in the Phototrophic Sulfur Bacterium *Allochromatium vinosum*. J Bacteriol. 2005;187:1392–404.

69. Santos AA, Venceslau SS, Grein F, Leavitt WD, Dahl C, Johnston DT, et al. A protein trisulfide couples dissimilatory sulfate reduction to energy conservation. Science. 2015;350:1541–5.

70. Dahl C, Franz B, Hensen D, Kesselheim A, Zigann R. Sulfite oxidation in the purple sulfur bacterium *Allochromatium vinosum*: identification of SoeABC as a major player and relevance of SoxYZ in the process. Microbiology. 2013;159:2626–38.

71. Gemerden H van. On the ATP generation by *Chromatium* in darkness. Arch Microbiol. 1968;64:118–124.

72. Weissgerber T, Sylvester M, Kröninger L, Dahl C. A Comparative Quantitative Proteomic Study Identifies New Proteins Relevant for Sulfur Oxidation in the Purple Sulfur Bacterium *Allochromatium vinosum*. Appl Environ Microbiol. 2014;80:2279–92.

73. Bassham JA, Benson AA, Calvin M. The Path of Carbon in Photosynthesis Viii. the Role of Malic Acid. J Biol Chem. 1950;185:781–7.

74. Tabita FR. The Biochemistry and Metabolic Regulation of Carbon Metabolism and CO_2 Fixation in Purple Bacteria. In: Anoxygenic Photosynthetic Bacteria. Springer, Dordrecht; 1995. p. 885–914. doi:10.1007/0-306-47954-0_41.

75. Sorokin DY, Trotsenko YA, Doronina NV, Tourova TP, Galinski EA, Kolganova TV, et al. *Methylohalomonas lacus* gen. nov., sp. nov. and *Methylonatrum kenyense* gen. nov., sp. nov., methylotrophic *gammaproteobacteria* from hypersaline lakes. Int J Syst Evol Microbiol. 2007;57:2762–9.

76. Wang X, Falcone DL, Tabita FR. Reductive pentose phosphate-independent CO2 fixation in *Rhodobacter sphaeroides* and evidence that ribulose bisphosphate carboxylase/oxygenase activity serves to maintain the redox balance of the cell. J Bacteriol. 1993;175:3372–9.

77. Weissgerber T, Dobler N, Polen T, Latus J, Stockdreher Y, Dahl C. Genome-Wide Transcriptional Profiling of the Purple Sulfur Bacterium *Allochromatium vinosum* DSM 180^T during Growth on Different Reduced Sulfur Compounds. J Bacteriol. 2013;195:4231–45.

78. Niel CBV. The Bacterial Photosyntheses and Their Importance for the General Problem of Photosynthesis. In: Advances in Enzymology and Related Areas of Molecular Biology. Wiley-Blackwell; 1941. p. 263–328. doi:10.1002/9780470122464.ch8.

79. Kämpf C, Pfennig N. Capacity of *Chromatiaceae* for chemotrophic growth. Specific respiration rates of *Thiocystis violacea* and *Chromatium vinosum*. Arch Microbiol. 1980;127:125–35.

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

81. Overmann J. The Family *Chlorobiaceae*. In: Dr MDP, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E, editors. The Prokaryotes. Springer New York; 2006. p. 359–78. doi:10.1007/0-387-30747-8_13.

82. Blankenship RE, Matsuura K. Antenna Complexes from Green Photosynthetic Bacteria. In: Light-Harvesting Antennas in Photosynthesis. Springer, Dordrecht; 2003. p. 195–217. doi:10.1007/978-94-017-2087-8_6.

83. Chung S, Frank G, Zuber H, Bryant DA. Genes encoding two chlorosome components from the green sulfur bacteria *Chlorobium vibrioforme* strain 8327D and *Chlorobium tepidum*. Photosynth Res. 1994;41:261–75.

84. Borrego CM, Garcia-Gil LJ. Rearrangement of light harvesting bacteriochlorophyll homologs as a response of green sulfur bacteria to low light intensities. Photosynth Res. 1994;45:21–30.

85. Manske AK, Glaeser J, Kuypers MMM, Overmann J. Physiology and phylogeny of green sulfur bacteria forming a monospecific phototrophic assemblage at a depth of 100 meters in the Black Sea. Appl Env Microbiol. 2005;71:8049–60.

86. Sirevåg R, Ormerod JG. Synthesis, storage and degradation of polyglucose in *Chlorobium thiosulfatophilum*. Arch Microbiol. 1977;111:239–44.

87. Habicht KS, Miller M, Cox RP, Frigaard N-U, Tonolla M, Peduzzi S, et al. Comparative proteomics and activity of a green sulfur bacterium through the water column of Lake Cadagno, Switzerland. Environ Microbiol. 2011;13:203–15.

88. Meyer TE, Donohue TJ. Cytochromes, Iron-Sulfur, and Copper Proteins Mediating Electron Transfer from the Cyt bc_1 Complex to Photosynthetic Reaction Center Complexes. In: Anoxygenic Photosynthetic Bacteria. Springer, Dordrecht; 1995. p. 725–45. doi:10.1007/0-306-47954-0_34.

89. Frigaard NU, Bryant DA. Genomic insights into the sulfur metabolism of phototrophic green sulfur bacteria. Sulfur Metab Phototrophic Org. 2008;:60–76.

90. Gregersen LH, Bryant DA, Frigaard N-U. Mechanisms and Evolution of Oxidative Sulfur Metabolism in Green Sulfur Bacteria. Front Microbiol. 2011;2. doi:10.3389/fmicb.2011.00116.

91. Parkin TB, Brock TD. The effects of light quality on the growth of phototrophic bacteria in lakes. Arch Microbiol. 1980;125:19–27.

92. Stomp M, Huisman J, Stal LJ, Matthijs HCP. Colorful niches of phototrophic microorganisms shaped by vibrations of the water molecule. ISME J. 2007;1:271–82.

93. 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;62:3251–8.

94. van Gemerden H, Tughan CS, de Wit R, Herbert RA. Laminated microbial ecosystems on sheltered beaches in Scapa Flow, Orkney Islands. FEMS Microbiol Lett. 1989;62:87–101.

95. Ng C, DeMaere MZ, Williams TJ, Lauro FM, Raftery M, Gibson JA, et al. Metaproteogenomic analysis of a dominant green sulfur bacterium from Ace Lake, Antarctica. ISME J. 2010;4:1002–19.

96. Overmann J, Cypionka H, Pfennig N. An extremely low-light-adapted phototrophic sulfur bacterium from the Black Sea. Limnol Oceanogr. 1992;:150–5.

97. Beatty JT, Overmann J, Lince MT, Manske AK, Lang AS, Blankenship RE. An obligately photosynthetic bacterial anaerobe from a deep-sea hydrothermal vent. Proc Natl Acad Sci USA. 2005;102:9306–10.

98. Guerrero R, Montesinos E, Pedrós-Alió C, Esteve I, Mas J, Van Gemerden H, et al. Phototrophic sulfur bacteria in two Spanish lakes: vertical distribution and limiting factors. Limnol Oceanogr. 1985;30:919–931.

99. Caldwell DE, Tiedje JM. The structure of anaerobic bacterial communities in the hypolimnia of several Michigan lakes. Can J Microbiol. 1975;21:377–85.

100. Diao M, Huisman J, Muyzer G. Spatio-temporal dynamics of sulfur bacteria during oxic--anoxic regime shifts in a seasonally stratified lake. FEMS Microbiol Ecol. 2018;94. doi:10.1093/femsec/fiy040.

101. Montesinos E, Guerrero R, Abella C, Esteve I. Ecology and Physiology of the Competition for Light Between Chlorobium limicola and Chlorobium phaeobacteroides in Natural Habitats. APPL Env MICROBIOL. 1983;46:10.

102. Van Gemerden H. Coexistence of organisms competing for the same substrate: an example among the purple sulfur bacteria. Microb Ecol. 1974;1:104–119.

103. Hans van Gemerden, Emilio Montesinos, Jordi Mas, Ricardo Guerrero. Diel Cycle of Metabolism of Phototrophic Purple Sulfur Bacteria in Lake Ciso (Spain). Limnol Oceanogr. 1985;30:932–43.

104. Gemerden HV, Mas J. Ecology of Phototrophic Sulfur Bacteria. In: Anoxygenic Photosynthetic Bacteria. Springer, Dordrecht; 1995. p. 49–85. doi:10.1007/0-306-47954-0_4.

105. Fröstl JM, Overmann J. Physiology and tactic response of the phototrophic consortium "*Chlorochromatium aggregatum*." Arch Microbiol. 1998;169:129–35.

106. 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;45:29–37.

107. Distel DL, Altamia MA, Lin Z, Shipway JR, Han A, Forteza I, et al. Discovery of chemoautotrophic symbiosis in the giant shipworm *Kuphus polythalamia* (*Bivalvia*: *Teredinidae*) extends wooden-steps theory. Proc Natl Acad Sci. 2017;114:E3652–8.

108. Nakagawa S, Shimamura S, Takaki Y, Suzuki Y, Murakami S, Watanabe T, et al. Allying with armored snails: the complete genome of *gammaproteobacterial* endosymbiont. ISME J. 2014;8:40–51.

109. Bright M, Giere O. Review article. Microbial symbiosis in *Annelida*. Symbiosis. 2005;38:1–46.

110. Gulati RD, Zadereev ES, Degermendzhi AG, editors. Ecology of Meromictic Lakes. Cham: Springer International Publishing; 2017. doi:10.1007/978-3-319-49143-1.

111. Boehrer B, Schultze M. Stratification of lakes. Rev Geophys. 2008;46. doi:10.1029/2006RG000210.

112. Overmann J, Tilzer MM. Control of primary productivity and the significance of photosynthetic bacteria in a meromictic kettle lake. Mittlerer Buchensee, West-Germany. Aquat Sci. 1989;51:261–78.

113. Pjevac P, Korlević M, Berg JS, Bura-Nakić E, Ciglenečki I, Amann R, et al. Community Shift from Phototrophic to Chemotrophic Sulfide Oxidation following Anoxic Holomixis in a Stratified Seawater Lake. Appl Environ Microbiol. 2015;81:298–308.

114. Genovese S, Trüper HG. Characterization of Photosynthetic Sulfur Bacteria Causing Red Water in Lake Faro (messina Sicily). Limnol Oceanogr. 1968;13:225 &.

115. Matsuyama M. Disturbance of *Chromatium* Population at Mid-depth of Lake Kaiike. Microbes Environ. 2002;17:134–8.

116. Casamayor EO, Mas J, Pedrós-Alió C. In Situ Assessment on the Physiological State of Purple and Green Sulfur Bacteria through the Analyses of Pigment and 5S rRNA Content. Microb Ecol. 2001;42:427–37.

117. Sorokin DY. The Microbial Sulfur Cycle at Extremely Haloalkaline Conditions of Soda Lakes. Front Microbiol. 2011;2. doi:10.3389/fmicb.2011.00044.

118. 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, Degermendzhi AG, editors. Ecology of Meromictic Lakes. Cham: Springer International Publishing; 2017. p. 155–86. doi:10.1007/978-3-319-49143-1_7.

119. Garwood EJ. The tarns of the canton Ticino. Q J Geol Soc Lond. 1906;62:165–93.

120. Krige LJ. Petrographische Untersuchungen im Val Piora u. Umgebung. 1918.

121. Bianconi F, Schweiz, Bundesamt für Landestopografie, Schweiz, Geologische Landesaufnahme. Ambrì-Piotta. 2015.

122. Del Don C, Hanselmann KW, Peduzzi R, Bachofen R. The meromictic alpine Lake Cadagno: orographical and biogeochemical description. Aquat Sci. 2001;63:70–90.

123. Stapfer A. Pollenanalytische Untersuchungen im Val Piora (Tessin) : ein Beitrag zur Klima- und Vegetationsgeschichte der Nacheiszeit. Geogr Helvetica. 1991;46:156–64.

124. Quinto (TI). Il Comune di Quinto: storia di un comune alpino sulla via delle genti. Quinto: Comune di Quinto; 2005.

125. Wirth SB, Gilli A, Niemann H, Dahl TW, Ravasi D, Sax N, et al. Combining sedimentological, trace metal (Mn, Mo) and molecular evidence for reconstructing past water-column redox conditions: The example of meromictic Lake Cadagno (Swiss Alps). Geochim Cosmochim Acta. 2013;120:220–38.

126. Züllig H. Pigmente phototropher Bakterien und ihre Bedeutung für die Seenforschung (mit Ergebnissen aus dem Lago Cadagno, Rotsee und Lobsigensee). Schweiz Z Für Hydrol. 1985;:87–126.

127. Birch L, Hanselmann KW, Bachofen R. Heavy metal conservation in Lake Cadagno sediments: Historical records of anthropogenic emissions in a meromictic alpine lake. Water Res. 1996;30:679–87.

128. Camacho A, Erez J, Chicote A, Florín M, Squires MM, Lehmann 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.

129. Düggeli M. Die Schwefelbakterien. Neujahrsbl Naturforschenden Ges Zür. 1919;121:43.

130. Brutschy A. Die Algenflora des Val Piora. Z Für Hydrol. 1931;5:1–120.

131. Neale PJ, Litchman E, Sobrino C, Callieri C, Morabito G, Montecino V, et al. Quantifying the response of phytoplankton photosynthesis to ultraviolet radiation: Biological weighting functions versus in situ measurements in two Swiss lakes. Aquat Sci. 2001;63:265–285.

132. Callieri C, Morabito G, Huot Y, Neale PJ, Litchman E. Photosynthetic response of pico- and nanoplanktonic algae to UVB, UVA and PAR in a high mountain lake. Aquat Sci. 2001;63:286–93.

133. Pasini P, Bachofen R. Influence of UV-radiation on the primary production of two high mountain lakes in the Piora Region. Lake Cadagno Meromictic Alp Lake. 1998;63:65–70.

134. Danza F, Storelli N, Roman S, Lüdin S, Tonolla M. Dynamic cellular complexity of anoxygenic phototrophic sulfur bacteria in the chemocline of meromictic Lake Cadagno. PLOS ONE. 2017;12:e0189510.

135. Schanz F, Stalder S. Phytoplankton summer dynamics and sedimentation in the thermally stratified Lake Cadagno. Doc DellIstituto Ital Idrobiol. 1998;63.

136. Halm H, Musat N, Lam P, Langlois R, Musat F, Peduzzi S, et al. Co-occurrence of denitrification and nitrogen fixation in a meromictic lake, Lake Cadagno (Switzerland). Environ Microbiol. 2009;11:1945–58.

137. Milucka J, Kirf M, Lu L, Krupke A, Lam P, Littmann S, et al. Methane oxidation coupled to oxygenic photosynthesis in anoxic waters. ISME J. 2015;9:1991–2002.

138. Berg JS, Michellod D, Pjevac P, Martinez-Perez C, Buckner CRT, Hach PF, et al. Intensive cryptic microbial iron cycling in the low iron water column of the meromictic Lake Cadagno: A cryptic microbial iron cycle. Environ Microbiol. 2016;18:5288–302.

139. HANSELMANN K, HUTTER R. Geomicrobiological coupling of sulfur and iron cycling in anoxic sediments of a meromictic lake: sulfate reduction and sulfide sources and sinks in Lake Cadagno. :14.

140. Schubert CJ, Vazquez F, Lösekann-Behrens T, Knittel K, Tonolla M, Boetius A. Evidence for anaerobic oxidation of methane in sediments of a freshwater system (Lago di Cadagno). FEMS Microbiol Ecol. 2011;76:26–38.

141. Wagener S, Schulz S, Hanselmann K. Abundance and distribution of anaerobic protozoa and their contribution to methane production in Lake Cadagno (Switzerland). FEMS Microbiol Lett. 1990;74:39–48.

142. Lehmann C, Bachofen R. Images of concentrations of dissolved sulphide in the sediment of a lake and implications for internal sulphur cycling. Sedimentology. 1999;46:537–44.

143. Ravasi D. Indagine sulla diversità microbica del Lago di Cadagno tramite spettrometria di massa MALDI-TOF e pirosequenziamento. Bellinzona: LMA; 2011.

144. Bachmann H. Das Phytoplankton der Pioraseen nebst einigen Beitr igen zur Kenntnis ties Phytoplanktons schweizerischer Alpenseen. :58.

145. Düggeli M. Hydrobiologische Untersuchungen im Pioragebiet. Z Für Hydrol. 1924;2:65–205.

146. Wuest A. Interactions in Lakes: Biology as Source of dominant physical Forces. Limnologica /. 1994;24:93.

147. Egli K, Wiggli M, Fritz M, Klug J, Gerss 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;35:105–113.

148. Sommer T, Danza F, Berg J, Sengupta A, Constantinescu G, Tokyay T, et al. Bacteria-induced mixing in natural waters. Geophys Res Lett. 2017;44. doi:10.1002/2017GL074868.

149. Schanz F, Fischer-Romero C, Bachofen R. Photosynthetic Production and Photoadaptation of Phototrophic Sulfur Bacteria in Lake Cadagno (Switzerland). Limnol Oceanogr. 1998;43:1262–9.

150. Jaag O, Märki E. Untersuchungen über die Trübungsverhältnisse, die Durchsichtigkeit und die Wasserfarbe in schweizerischen Gewässern. 1964. https://link.springer.com/content/pdf/10.1007%2FBF02505731.pdf. Accessed 13 Mar 2018.

151. Del Don C, Hanselmann K, Peduzzi R, Zullig H. Phototrophic Bacteria in the Redox Transition Zone of Lago Cadagno, a Meromictic, Alpine Lake. Experientia. 1985;41:554–554.

152. Tonolla M, Demarta A, Peduzzi R, Hahn D. In situ analysis of phototrophic sulfur bacteria in the chemocline of meromictic Lake Cadagno (Switzerland). Appl Environ Microbiol. 1999;65:1325–30.

153. Bosshard PP, Santini Y, Grüter D, Stettler R, Bachofen R. Bacterial diversity and community composition in the chemocline of the meromictic alpine Lake Cadagno as revealed by 16S rDNA analysis. FEMS Microbiol Ecol. 2000;31:173–182.

154. 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.

155. Tonolla M, Peduzzi S, Hahn D, Peduzzi R. Spatio-temporal distribution of phototrophic sulfur bacteria in the chemocline of meromictic Lake Cadagno (Switzerland). FEMS Microbiol Ecol. 2003;43:89–98.

156. Gregersen LH, Habicht KS, Peduzzi S, Tonolla M, Canfield DE, Miller M, et al. Dominance of a clonal green sulfur bacterial population in a stratified lake. FEMS Microbiol Ecol. 2009;70:30–41.

157. Fischer C, Wiggli M, Schanz F, Hanselmann KW, Bachofen R. Light environment and synthesis of bacteriochlorophyll by populations of *Chromatium okenii* under natural environmental conditions. FEMS Microbiol Ecol. 1996;21:1–9.

158. Joss A, Mez K, Känel B, Hanselmann KW, Bachofen R. Measurement of Fluorescence Kinetics of Phototrophic Bacteria in the Natural Environment. J Plant Physiol. 1994;144:333–8.

159. Luthy L, Fritz M, Bachofen R. In Situ Determination of Sulfide Turnover Rates in a Meromictic Alpine Lake. Appl Environ Microbiol. 2000;66:712–7.

160. Del Don CD, Hanselmann KW, Peduzzi R, Bachofen R. Biomass composition and methods for the determination of metabolic reserve polymers in phototrophic sulfur bacteria. Aquat Sci. 1994;56:1–15.

161. Bosshard PP, Stettler R, Bachofen R. Seasonal and spatial community dynamics in the meromictic Lake Cadagno. Arch Microbiol. 2000;174:168–74.

162. Storelli N, Peduzzi S, Saad MM, Frigaard N-U, Perret X, Tonolla M. CO2 assimilation in the chemocline of Lake Cadagno is dominated by a few types of phototrophic purple sulfur bacteria. FEMS Microbiol Ecol. 2013;84:421–432.

163. Musat N, Halm H, Winterholler B, Hoppe P, Peduzzi S, Hillion F, et al. A single-cell view on the ecophysiology of anaerobic phototrophic bacteria. Proc Natl Acad Sci U S A. 2008;105:17861–6.

164. Peduzzi S, Storelli N, Welsh A, Peduzzi R, Hahn D, Perret X, et al. Candidatus "*Thiodictyon syntrophicum*", sp. nov., a new purple sulfur bacterium isolated from the chemocline of Lake Cadagno forming aggregates and specific associations with Desulfocapsa sp. Syst Appl Microbiol. 2012;35:139–44.

165. Decristophoris PMA, Peduzzi S, Ruggeri-Bernardi N, Hahn D, Tonolla M. Fine scale analysis of shifts in bacterial community structure in the chemocline of meromictic Lake Cadagno, Switzerland. J Limnol. 2009;68:16–24.

166. Zimmermann M, Escrig S, Hübschmann T, Kirf MK, Brand A, Inglis RF, 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. Microb Physiol Metab. http://journal.frontiersin.org/article/10.3389/fmicb.2015.00243/abstract. Accessed 10 Nov 2015.

167. Dahl TW, Anbar AD, Gordon GW, Rosing MT, Frei R, Canfield DE. The behavior of molybdenum and its isotopes across the chemocline and in the

sediments of sulfidic Lake Cadagno, Switzerland. Geochim Cosmochim Acta. 2010;74:144–63.

168. Ravasi DF, Peduzzi S, Guidi V, Peduzzi R, Wirth SB, Gilli A, et al. Development of a real-time PCR method for the detection of fossil 16S rDNA fragments of phototrophic sulfur bacteria in the sediments of Lake Cadagno: Real-time PCR detection of fossil 16S rDNA fragments. Geobiology. 2012;10:196–204.

169. Bueche M, Wunderlin T, Roussel-Delif L, Junier T, Sauvain L, Jeanneret N, et al. Quantification of Endospore-Forming Firmicutes by Quantitative PCR with the Functional Gene spo0A. Appl Environ Microbiol. 2013;79:5302–12.

170. Bottinelli M. Approche moléculaire à l'étude des bactéries sulfato-réductrices et des Archaea méthanogènes dans les sédiments des lacs Cadagno et Rotsee. [s.n.]; 2008.

171. Koch R. Zur Untersuchung von pathogenen Organismen. 2010. doi:10.25646/5071.

172. Stanier RY, Niel The Main Outlines of Bacterial Classification. Bacteriol. 1941, 42:4 437–466

173. Grote M. Petri dish versus Winogradsky column: a longue durée perspective on purity and diversity in microbiology, 1880s–1980s. Hist Philos Life Sci. 2018;40:11.

174. Zuckerkandl E, Pauling L. Evolutionary Divergence and Convergence in Proteins. In: Evolving Genes and Proteins. Elsevier; 1965. p. 97–166. doi:10.1016/B978-1-4832-2734-4.50017-6.

175. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci. 1977;74:5463–7.

176. Woese CR, Fox GE. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. Proc Natl Acad Sci. 1977;74:5088–90.

177. Woese CR. Bacterial evolution. Microbiol Rev. 1987;51:221–71.

178. Giovannoni SJ, DeLong EF, Schmidt TM, Pace NR. Tangential flow filtration and preliminary phylogenetic analysis of marine picoplankton. Appl Environ Microbiol. 1990;56:2572–5.

179. Land M, Hauser L, Jun S-R, Nookaew I, Leuze MR, Ahn T-H, et al. Insights from 20 years of bacterial genome sequencing. Funct Integr Genomics. 2015;15:141–61.

180. Welcome to the $1000 genome. Opiniomics. 2014. http://www.opiniomics.org/welcome-to-the-1000-genome/. Accessed 25 Mar 2018.

181. JGI GOLD | Statistics. https://gold.jgi.doe.gov/statistics. Accessed 22 Jun 2018.

182. Richardson EJ, Watson M. The automatic annotation of bacterial genomes. Brief Bioinform. 2013;14:1–12.

183. Koren S, Treangen TJ, Hill CM, Pop M, Phillippy AM. Automated ensemble assembly and validation of microbial genomes. BMC Bioinformatics. 2014;15:126.

184. Mardis E, McPherson J, Martienssen R, Wilson RK, McCombie WR. What is Finished, and Why Does it Matter. Genome Res. 2002;12:669–71.

185. Denton JF, Lugo-Martinez J, Tucker AE, Schrider DR, Warren WC, Hahn MW. Extensive error in the number of genes inferred from draft genome assemblies. PLoS Comput Biol. 2014;10:e1003998.

186. Fierst JL, Murdock DA. Decontaminating eukaryotic genome assemblies with machine learning. BMC Bioinformatics. 2017;18:533.

187. McInerney J, Cummins C, Haggerty L. Goods-thinking vs. tree-thinking. Mob Genet Elem. 2011;1:304–43.

188. Chun J, Rainey F. Integrating genomics into the taxonomy and systematics of the Bacteria and Archaea. Int J Syst Evol Microbiol. 2014;64:316–24.

189. Jolley KA, Bliss CM, Bennett JS, Bratcher HB, Brehony C, Colles FM, et al. Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. Microbiology. 2012;158 Pt 4:1005–15.

190. Huang CH, Hsiang T, Trevors JT. Comparative bacterial genomics: defining the minimal core genome. Antonie Van Leeuwenhoek. 2013;103:385–98.

191. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int J Syst Evol Microbiol. 2007;57:81–91.

192. Lee I, Ouk Kim Y, Park S-C, Chun J. OrthoANI: An improved algorithm and software for calculating average nucleotide identity. Int J Syst Evol Microbiol. 2016;66:1100–3.

193. Periwal V, Scaria V. Insights into structural variations and genome rearrangements in prokaryotic genomes. Bioinforma Oxf Engl. 2015;31:1–9.

194. Smalla K, Jechalke S, Top EM. Plasmid Detection, Characterization, and Ecology. Microbiol Spectr. 2015;3. doi:10.1128/microbiolspec.PLAS-0038-2014.

195. Touchon M, Moura de Sousa JA, Rocha EP. Embracing the enemy: the diversification of microbial gene repertoires by phage-mediated horizontal gene transfer. Curr Opin Microbiol. 2017;38:66–73.

196. García-Aljaro C, Ballesté E, Muniesa M. Beyond the canonical strategies of horizontal gene transfer in prokaryotes. Curr Opin Microbiol. 2017;38:95–105.

197. Parks DH, Rinke C, Chuvochina M, Chaumeil P-A, Woodcroft BJ, Evans PN, et al. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. Nat Microbiol. 2017;2:1533–42.

198. Spang A, Caceres EF, Ettema TJG. Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life. Science. 2017;357:eaaf3883.

199. Locey KJ, Lennon JT. Scaling laws predict global microbial diversity. Proc Natl Acad Sci. 2016;113:5970–5.

200. Widder S, Allen RJ, Pfeiffer T, Curtis TP, Wiuf C, Sloan WT, et al. Challenges in microbial ecology: building predictive understanding of community function and dynamics. Nature. 2016. http://dspace.mit.edu/handle/1721.1/110256. Accessed 23 Jun 2018.

201. Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM, et al. Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature. 2004;428:37–43.

202. Hamilton TL, Bovee RJ, Thiel V, Sattin SR, Mohr W, Schaperdoth I, et al. Coupled reductive and oxidative sulfur cycling in the phototrophic plate of a meromictic lake. Geobiology. 2014;12:451–68.

203. Bertilsson S, Burgin A, Carey CC, Fey SB, Grossart H-P, Grubisic LM, et al. The under-ice microbiome of seasonally frozen lakes. Limnol Oceanogr. 2013;58:1998–2012.

204. Li J, Jia H, Cai X, Zhong H, Feng Q, Sunagawa S, et al. An integrated catalog of reference genes in the human gut microbiome. Nat Biotechnol. 2014;32:834–41.

205. Cavalett A, Castro da Silva MA, Toyofuku T, Mendes R, Taketani RG, Pedrini J, et al. Dominance of *Epsilonproteobacteria* associated with a whale fall at a 4204 m depth - South Atlantic Ocean. Deep-Sea Res Part Ii-Top Stud Oceanogr. 2017;146:53–8.

206. Bahcall OG. Metagenomics: Urban microbiome. Nat Rev Genet. 2015;16:194–5.

207. Emerson JB, Adams RI, Román CMB, Brooks B, Coil DA, Dahlhausen K, et al. Schrödinger's microbes: Tools for distinguishing the living from the dead in microbial ecosystems. Microbiome. 2017;5:86.

208. Wöhlbrand L, Trautwein K, Rabus R. Proteomic tools for environmental microbiology—A roadmap from sample preparation to protein identification and quantification. PROTEOMICS. 2013;13:2700–2730.

209. Zhou J, Zhou T, Cao R, Liu Z, Shen J, Chen P, et al. Evaluation of the Application of Sodium Deoxycholate to Proteomic Analysis of Rat Hippocampal Plasma Membrane. J Proteome Res. 2006;5:2547–53.

210. Glatter T, Ahrné E, Schmidt A. Comparison of Different Sample Preparation Protocols Reveals Lysis Buffer-Specific Extraction Biases in Gram-Negative Bacteria and Human Cells. J Proteome Res. 2015;14:4472–85.

211. Emmett MR, Caprioli RM. Micro-electrospray mass spectrometry: Ultra-high-sensitivity analysis of peptides and proteins. J Am Soc Mass Spectrom. 1994;5:605–13.

212. Guilhaus M, Selby D, Mlynski V. Orthogonal acceleration time-of-flight mass spectrometry. Mass Spectrom Rev. 19:65–107.

213. Hu Q, Noll RJ, Li H, Makarov A, Hardman M, Cooks RG. The Orbitrap: a new mass spectrometer. J Mass Spectrom. 2005;40:430–43.

214. Q Exactive™ Hybrid Quadrupole-Orbitrap™ Mass Spectrometer. https://www.thermofisher.com/order/catalog/product/IQLAAEGAAPFALGMAZR. Accessed 26 Jun 2018.

215. Sadygov RG, Cociorva D, Iii JRY. Large-scale database searching using tandem mass spectra: Looking up the answer in the back of the book. Nat Methods. 2004;1:195–202.

216. Edwards NJ. Protein Identification from Tandem Mass Spectra by Database Searching. In: Protein Bioinformatics. Humana Press, New York, NY; 2017. p. 357–80. doi:10.1007/978-1-4939-6783-4_17.

217. Koenig T, Menze BH, Kirchner M, Monigatti F, Parker KC, Patterson T, et al. Robust Prediction of the MASCOT Score for an Improved Quality Assessment in Mass Spectrometric Proteomics. J Proteome Res. 2008;7:3708–17.

218. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol. 2008;26:1367–72.

219. Kim S, Pevzner PA. MS-GF+ makes progress towards a universal database search tool for proteomics. Nat Commun. 2014;5:5277.

220. Ram RJ, Verberkmoes NC, Thelen MP, Tyson GW, Baker BJ, Blake RC 2nd, et al. Community proteomics of a natural microbial biofilm. Science. 2005;308:1915–20.

221. Warnecke F, Warnecke F, Luginbuhl P, Ivanova N, Ghassemian M, Richardson TH, et al. Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. Nature. 2007;450. doi:10.1038/nature06269.

222. Wegley Kelly L, Haas AF, Nelson CE. Ecosystem Microbiology of Coral Reefs: Linking Genomic, Metabolomic, and Biogeochemical Dynamics from Animal Symbioses to Reefscape Processes. mSystems. 2018;3. doi:10.1128/mSystems.00162-17.

223. Krohn-Molt I, Alawi M, Förstner KU, Wiegandt A, Burkhardt L, Indenbirken D, et al. Insights into Microalga and *Bacteria* Interactions of Selected Phycosphere Biofilms Using Metagenomic, Transcriptomic, and Proteomic Approaches. Front Microbiol. 2017;8. doi:10.3389/fmicb.2017.01941.

224. Singh RP, Reddy CRK. Unraveling the Functions of the Macroalgal Microbiome. Front Microbiol. 2016;6. doi:10.3389/fmicb.2015.01488.

225. Christie-Oleza JA, Armengaud J, Guerin P, Scanlan DJ. Functional distinctness in the exoproteomes of marine *Synechococcus*. Environ Microbiol. 2015;17:3781–94.

226. Rubiano-Labrador C, Bland C, Miotello G, Armengaud J, Baena S. Salt Stress Induced Changes in the Exoproteome of the Halotolerant Bacterium Tistlia consotensis Deciphered by Proteogenomics. PLoS ONE. 2015;10. doi:10.1371/journal.pone.0135065.

227. Weissgerber T, Zigann R, Bruce D, Chang Y-J, Detter JC, Han C, et al. Complete genome sequence of *Allochromatium vinosum* DSM 180(T). Stand Genomic Sci. 2011;5:311–30.

228. Weissgerber T, Watanabe M, Hoefgen R, Dahl C. Metabolomic profiling of the purple sulfur bacterium *Allochromatium vinosum* during growth on different reduced sulfur compounds and malate. Metabolomics. 2014;10:1094–112.

229. Falkenby LG, Szymanska M, Holkenbrink C, Habicht KS, Andersen JS, Miller M, et al. Quantitative proteomics of *Chlorobaculum tepidum*: insights into the sulfur metabolism of a phototrophic green sulfur bacterium. FEMS Microbiol Lett. 2011;323:142–50.

230. Karas M, Bachmann D, Hillenkamp F. Influence of the wavelength in high-irradiance ultraviolet laser desorption mass spectrometry of organic molecules. Anal Chem. 1985;57:2935–9.

231. Zenobi R, Knochenmuss R. Ion formation in MALDI mass spectrometry. Mass Spectrom Rev. 1998;17:337–66.

232. Karas M, Bahr U. Laser desorption ionization mass spectrometry of large biomolecules. TrAC Trends Anal Chem. 1990;9:321–5.

233. Pineda FJ, Lin JS, Fenselau C, Demirev PA. Testing the Significance of Microorganism Identification by Mass Spectrometry and Proteome Database Search. Anal Chem. 2000;72:3739–44.

234. Fenselau C, Demirev PA. Characterization of intact microorganisms by MALDI mass spectrometry. Mass Spectrom Rev. 20:157–71.

235. Christner M, Trusch M, Rohde H, Kwiatkowski M, Schlüter H, Wolters M, et al. Rapid MALDI-TOF Mass Spectrometry Strain Typing during a Large Outbreak of Shiga-Toxigenic *Escherichia coli*. PLOS ONE. 2014;9:e101924.

236. MALDI Biotyper Systems - Mass Spectrometry and Separations | Bruker. https://www.bruker.com/products/mass-spectrometry-and-separations/maldi-biotyper-systems.html. Accessed 26 Jun 2018.

237. VITEK® MS. bioMérieux Clinical Diagnostics. Mass spectrometry microbial identification system. Accessed 26 Jun 2018.

238. Imanishi SY, Nakayama T, Asukabe H, Harada K. Application of MALDI Biotyper to cyanobacterial profiling. Rapid Commun Mass Spectrom. 31:325–32.

239. Tagg KA, Ginn AN, Partridge SR, Iredell JR. MALDI-TOF Mass Spectrometry for Multilocus Sequence Typing of *Escherichia coli* Reveals Diversity among Isolates Carrying bla CMY-2 -Like Genes. PLOS ONE. 2015;10:e0143446.

240. Ryzhov V, Fenselau C. Characterization of the Protein Subset Desorbed by MALDI from Whole Bacterial Cells. Anal Chem. 2001;73:746–50.

241. Sun L, Teramoto K, Sato H, Torimura M, Tao H, Shintani T. Characterization of ribosomal proteins as biomarkers for matrix-assisted laser desorption/ionization mass spectral identification of *Lactobacillus plantarum*. Rapid Commun Mass Spectrom RCM. 2006;20:3789–98.

242. Ziegler D, Pothier JF, Ardley J, Fossou RK, Pflüger V, Meyer S de, et al. Ribosomal protein biomarkers provide root nodule bacterial identification by MALDI-TOF MS. Appl Microbiol Biotechnol. 2015;99:5547–62.

243. Armengaud J. Defining Diagnostic Biomarkers Using Shotgun Proteomics and MALDI-TOF Mass Spectrometry. In: Diagnostic Bacteriology. Humana Press, New York, NY; 2017. p. 107–20. doi:10.1007/978-1-4939-7037-7_6.

244. Fagerquist CK. Top-down proteomic identification of bacterial protein biomarkers and toxins using MALDI-TOF-TOF-MS/MS and post-source decay. Rev Anal Chem. 2013;32:127–133.

245. Vogel G, Cuénod A, Mouchet R, Strauss A, Daubenberger C, Pflüger V, et al. Functional characterization and phenotypic monitoring of human hematopoietic stem cell expansion and differentiation of monocytes and macrophages by whole-cell mass spectrometry. Stem Cell Res. 2018;26:47–54.

246. Kucharova V, Wiker HG. Proteogenomics in microbiology: Taking the right turn at the junction of genomics and proteomics. PROTEOMICS. 2014;14:2360–675.

247. Chen B, Lietz CB, Li L. Coupling matrix-assisted ionization with high resolution mass spectrometry and electron transfer dissociation to characterize intact proteins and post-translational modifications. Anal Bioanal Chem. 2017;:1–11.

248. PAPMID DB - mabritec. http://mabritec.com/papmid.html. Accessed 26 Jun 2018.

249. Tomachewski D, Galvão CW. Ribopeaks: a web tool for bacterial classifica- tion through m/z data from ribosomal proteins. 2018;:3.

250. Cheng D, Qiao L, Horvatovich P. Toward Spectral Library-Free Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry Bacterial Identification. J Proteome Res. 2018. doi:10.1021/acs.jproteome.8b00065.

251. Russell SC, Edwards N, Fenselau C. Detection of Plasmid Insertion in Escherichia coli by MALDI-TOF Mass Spectrometry. Anal Chem. 2007;79:5399–406.

252. Kostrzewa M, Sparbier K, Maier T, Schubert S. MALDI-TOF MS: an upcoming tool for rapid detection of antibiotic resistance in microorganisms. Proteomics Clin Appl. 2013.

253. Krismer J. Single-Cell Mass Spectrometry for High-Throughput Lipid Phenotyping of *Chlamydomonas reinhardtii*. Doctoral Thesis. ETH Zurich; 2016. doi:10.3929/ethz-a-010797620.

254. Perty M. Zur Kenntniss kleinster Lebensformen: nach Bau, Funktionen, Systematik, mit Specialverzeichniss der in der Schweiz beobachteten. Bern: Jent & Reinert; 1852.

255. Winogradsky S. Beiträge zur Morphologie und Physiologie der Bacterien. Heft 1. Zur Morphologie und Physiologie der Schwefelbacterien. Leipzig: Felix; 1888.

256. Buder J. Zur Kenntnis der phototaktischen Richtungsbewegungen. In: Jahrbücher für wissenschaftliche Botanik. 1917. p. 105–220.

257. Engelmann TW. *Bacterium photometricum*. Arch Für Gesamte Physiol Menschen Tiere. 1883;30:95–124.

258. Buder J. Zur Kenntnis des Thiospirillum jenense und seiner Reaktionen auf Lichtreize. Jahrbücher für wissenschaftliche Botanik. 1915;56. http://www.worldcat.org/title/kenntnis-des-thiospirillum-jenense-und-seiner-reaktionen-auf-lichtreize/oclc/918412503&referer=brief_results. Accessed 17 Jan 2018.

259. Schlegel HG. Vergleichende Untersuchungen über die Lichtempfindlichkeit einiger Purpurbakterien. Univerität Halle; 1950.

260. Pfennig N. [Observations on swarming of *Chromatium okenii*]. Arch Für Mikrobiol. 1962;42:90–5.

261. Manten A. The isolation *Chromatium okenii* and its behaviour in different media. Antonie Van Leeuwenhoek. 1942;8:164–8.

262. Sorokin YI. Interrelations between sulphur and carbon turnover in meromictic lakes. Arch Hydrobiol. 1970;66:391–446.

263. Matsuyama M, Shirouzu E. Importance of Photosynthetic Sulfur Bacteria, *Chromatium* sp. as an Organic Matter Producer in Lake Kaiike. Jpn J Limnol Rikusuigaku Zasshi. 1978;39:103–11.

264. Martínez-Alonso M, Méndez-Álvarez S, Ramírez-Moreno S, González-Toril E, Amils R, Gaju N. Spatial Heterogeneity of Bacterial Populations in Monomictic Lake Estanya (Huesca, Spain). Microb Ecol. 2008;55:737–50.

265. Note sur la structure d'une Bactériacée, le *Chromatium okenii* - Bulletin de la Société Botanique de France - Volume 56, Issue 4. http://www.tandfonline.com/doi/abs/10.1080/00378941.1909.10831411. Accessed 26 Feb 2016.

266. Schlegel HG, Pfennig N. Die Anreicherungskultur einiger Schwefelpurpurbakterien. Arch Für Mikrobiol. 1960;38:1–39.

267. Imhoff JF. The Family *Chromatiaceae*. In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F, editors. The Prokaryotes. Springer Berlin Heidelberg; 2014. p. 151–78. doi:10.1007/978-3-642-38922-1_295.

268. Schlegel HG. [Uptake substances of *Chromatium okenii*]. Arch Für Mikrobiol. 1962;42:110–6.

269. Posth NR, Bristow LA, Cox RP, Habicht KS, Danza F, Tonolla M, et al. Carbon isotope fractionation by anoxygenic phototrophic bacteria in euxinic Lake Cadagno. Geobiology. 2017;15:798–816.

270. Schulz S, Wagener S, Pfennig N. Utilization of various chemotrophic and phototrophic bacteria as food by the anaerobic ciliate *Trimyema compressum*. Eur J Protistol. 1990;26:122–31.

271. Tonolla M, Demarta A, Hahn D, Peduzzi R. Microscopic and molecular in situ characterization of bacterial populations in the meromictic Lake Cadagno. Lake Cadagno Meromictic Alp Lake. 1998;:31–44.

272. Agencourt AMPure XP Bead Clean-up - NEBNext Microbiome DNA Enrichment Kit (E2612) | NEB. https://www.neb.com/protocols/2013/04/18/agencourt-ampure-xp-bead-clean-up-e2612. Accessed 5 Jan 2017.

273. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 2017;27:722–36.

274. Akhter S, Aziz RK, Edwards RA. PhiSpy: a novel algorithm for finding prophages in bacterial genomes that combines similarity- and composition-based strategies. Nucleic Acids Res. 2012;40:e126–e126.

275. Roux S, Enault F, Hurwitz BL, Sullivan MB. VirSorter: mining viral signal from microbial genomic data. PeerJ. 2015;3:e985.

276. Huerta-Cepas J, Forslund K, Szklarczyk D, Jensen LJ, von Mering C, Bork P. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. bioRxiv. 2016;:076331.

277. Wang Y, Coleman-Derr D, Chen G, Gu YQ. OrthoVenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. Nucleic Acids Res. 2015;43:W78–84.

278. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. J Mol Biol. 2016;428:726–31.

279. Wu M, Scott AJ. Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2. Bioinformatics. 2012;28:1033–4.

280. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31:3210–2.

281. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, et al. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics. 2015;31:3691–3693.

282. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32:1792–7.

283. 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;32:268–74.

284. Trifinopoulos J, Nguyen L-T, von Haeseler A, Minh BQ. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Res. 2016;44:W232–5.

285. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform. 2017. doi:10.1093/bib/bbx108.

286. Rambaut A. FigTree. Institute of Evolutionary Biology, University of Edinburgh.; 2016. http://tree.bio.ed.ac.uk/software/figtree/. Accessed 13 Apr 2018.

287. DSMZ: Catalogue Microorganisms. https://www.dsmz.de/catalogues/catalogue-microorganisms.html. Accessed 12 Apr 2018.

288. Hageage GJ, Gherna RL. Surface Structure of *Chromatium okenii* and *Chromatium weissei*. J Bacteriol. 1971;106:687–90.

289. Corson 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.

290. Schmidt K, Jensen SL, Schlegel HG. Die Carotinoide der *Thiorhodaceae*. Arch Für Mikrobiol. 1963;46:117–26.

291. Vogl K, Bryant DA. Elucidation of the Biosynthetic Pathway for Okenone in *Thiodictyon* sp. CAD16 Leads to the Discovery of Two Novel Carotene Ketolases. J Biol Chem. 2011;286:38521–32.

292. Schoepp B, Parot P, Menin L, Gaillard J, Richaud P, Vermeglio A. In vivo Participation of a High Potential Iron-Sulfur Protein as Electron Donor to the Photochemical Reaction Center of *Rubrivivax gelatinosus*. Biochemistry (Mosc). 1995;34:11736–42.

293. Matsuyama M. Upward Movement of *Chromatium* sp. in the H_2S-Layer of Lake Kaiike Causing a Bloom at its Upper Boundary. Jpn J Limnol Rikusuigaku Zasshi. 1995;56:205–9.

294. Matsuyama M, Moon S-W. A Bloom of Low Light-adapted *Chromatium* sp. In Lake Kaiike. Jpn J Limnol Rikusuigaku Zasshi. 1998;59:79–85.

295. Wit R de, Gemerden H van. Chemolithotrophic growth of the phototrophic sulfur bacterium *Thiocapsa roseopersicina*. FEMS Microbiol Ecol. 1987;3:117–26.

296. Thiele HH. Die Verwertung einfacher organischer Substrate durch *Thiorhodaceae*. Arch Für Mikrobiol. 1968;60:124–38.

297. Brune DC. Isolation and characterization of sulfur globule proteins from *Chromatium vinosum* and *Thiocapsa roseopersicina*. Arch Microbiol. 1995;163:391–9.

298. Prange A, Engelhardt H, Trüper HG, Dahl C. The role of the sulfur globule proteins of *Allochromatium vinosum*: mutagenesis of the sulfur globule protein genes and expression studies by real-time RT-PCR. Arch Microbiol. 2004;182:165–74.

299. Shimizu T, Shen J, Fang M, Zhang Y, Hori K, Trinidad JC, et al. Sulfide-responsive transcriptional repressor SqrR functions as a master regulator of sulfide-dependent photosynthesis. Proc Natl Acad Sci. 2017;114:2355–60.

300. Trüper HG, Schlegel HG. Sulphur metabolism in *Thiorhodaceae* I. Quantitative measurements on growing cells of *Chromatium okenii*. Antonie Van Leeuwenhoek. 1964;30:225–38.

301. Trüper HG. CO2-Fixierung und Intermediärstoffwechsel bei *Chromatium okenii* Perty. Arch Für Mikrobiol. 1964;49:23–50.

302. Trüper HG. Sulphur metabolism in *Thiorhodaceae*. II. stoichiometric relationship of CO2 fixation to oxidation of hydrogen sulphide and intracellular sulphur in *Chromatium okenii*. Antonie Van Leeuwenhoek. 1964;30:385–94.

303. Badger MR, Bek EJ. Multiple Rubisco forms in proteobacteria: their functional significance in relation to CO_2 acquisition by the CBB cycle. J Exp Bot. 2008;59:1525–41.

304. Miller AR, North JA, Wildenthal JA, Tabita FR. Two Distinct Aerobic Methionine Salvage Pathways Generate Volatile Methanethiol in *Rhodopseudomonas palustris*. mBio. 2018;9:e00407-18.

305. Hanson TE, Tabita FR. Insights into the stress response and sulfur metabolism revealed by proteome analysis of a *Chlorobium tepidum* mutant lacking the Rubisco-like protein. Photosynth Res. 2003;78:231–48.

306. Imker HJ, Singh J, Warlick BP, Tabita FR, Gerlt JA. Mechanistic Diversity in the RuBisCO Superfamily: A Novel Isomerization Reaction Catalyzed by the RuBisCO-Like Protein from *Rhodospirillum rubrum*. Biochemistry (Mosc). 2008;47:11171–3.

307. Schlegel HG. Die Speicherstoffe von *Chromatium okenii*. Arch Für Mikrobiol. 1962;42:110–6.

308. Mas J, Van Gemerden H. Storage products in purple and green sulfur bacteria. Anoxygenic Photosynth Bact. 1995;:973–90.

309. Fagan RP, Fairweather NF. Biogenesis and functions of bacterial S-layers. Nat Rev Microbiol. 2014;12:211–22.

310. Koval SF, Hynes SH. Effect of paracrystalline protein surface layers on predation by *Bdellovibrio bacteriovorus*. J Bacteriol. 1991;173:2244–9.

311. Bavendamm W. Die farblosen und roten Schwefelbakterien des süss-und salzwassers: Grundlinien zu einer Monographie. G. Fischer; 1924.

312. Esteve I, Guerrero R, Montesinos E, Abellà C. Electron microscope study of the interaction of epibiontic bacteria with *Chromatium minus* in natural habitats. Microb Ecol. 1983;9:57–64.

313. Guerrero R, Pedros-Alio C, Esteve I, Mas J, Chase D, Margulis L. Predatory prokaryotes: Predation and primary consumption evolved in bacteria. Proc Natl Acad Sci U S A. 1986;83:2138–42.

314. Gaju N, Esteve I, Guerrero R. Distribution of Predatory Bacteria That Attack *Chromatiaceae* in a Sulfurous Lake. Microb Ecol. 1992;24:171–9.

315. Proctor LM, Fuhrman JA. Viral mortality of marine *bacteria* and *cyanobacteria*. Nature. 1990;343:60–2.

316. Wommack KE, Colwell RR. Virioplankton: *Viruses* in Aquatic Ecosystems. Microbiol Mol Biol Rev. 2000;64:69–114.

317. Witzel K-P, Demuth J, Schütt C. *Viruses*. In: Microbial Ecology of Lake Plußsee. Springer, New York, NY; 1994. p. 270–86. doi:10.1007/978-1-4612-2606-2_13.

318. Heldal M, Bratbak G. Production and decay of viruses in aquatic environments. Mar Ecol Prog Ser. 1991;:205–212.

319. Tuomi P, Torsvik T, Heldal M, Bratbak G. Bacterial population dynamics in a meromictic lake. Appl Environ Microbiol. 1997;63:2181–8.

320. Pfennig N. *Chromatium okenii* (*Thiorhodaceae*): bio-convection aero and photoactic behaviour. Konstanz: Institut für den wissenschaftlichen Film; 1965.

321. Wadhams GH, Armitage JP. Making sense of it all: bacterial chemotaxis. Nat Rev Mol Cell Biol. 2004;5:1024–37.

322. Mitchell JG, Martinez-Alonso M, Lalucat J, Esteve I, Brown S. Velocity changes, long runs, and reversals in the *Chromatium minus* swimming response. J Bacteriol. 1991;173:997–1003.

323. Vaituzis Z, Doetsch RN. Motility Tracks: Technique for Quantitative Study of Bacterial Movement. Appl Microbiol. 1969;17:584–8.

324. Pfennig N. Eine vollsynthetische Nährlösung zur selektiven Anreicherung einiger Schwefelpurpurbakterien. Naturwissenschaften. 1961;48:136–136.

325. Gervais F. Diel vertical migration of *Cryptomonas* and *Chromatium* in the deep chlorophyll maximum of a eutrophic lake. J Plankton Res. 1997;19:533–50.

326. Koonin EV. Orthologs, Paralogs, and Evolutionary Genomics. Annu Rev Genet. 2005;39:309–38.

327.   Tranvik LJ, Jørgensen NOG. Colloidal and dissolved organic matter in lake water: Carbohydrate and amino acid composition, and ability to support bacterial growth. Biogeochemistry. 1995;30:77–97.

328.   Bonde TA, Nielsen T, Miller M, Sørensen J. Arginine ammonification assay as a rapid index of gross N mineralization in agricultural soils. Biol Fertil Soils. 2001;34:179–84.

329.   Bertilsson S, Eiler A, Nordqvist A, Jørgensen NOG. Links between bacterial production, amino-acid utilization and community composition in productive lakes. ISME J. 2007;1:532–44.

330.   Levin PA, Angert ER. Small but Mighty: Cell Size and Bacteria. Cold Spring Harb Perspect Biol. 2015;7:a019216.

331.   Schulz HN, Jørgensen and BB. Big Bacteria. Annu Rev Microbiol. 2001;55:105–37.

332.   Kusian Bernhard, Bowien Botho. Organization and regulation of cbb CO2 assimilation genes in autotrophic bacteria. FEMS Microbiol Rev. 2006;21:135–55.

333.   McKinlay JB, Harwood CS. Carbon dioxide fixation as a central redox cofactor recycling mechanism in bacteria. Proc Natl Acad Sci. 2010;107:11669–75.

334.   Takahashi M, Ichimura S. Vertical Distribution and Organic Matter Production of Photosynthetic Sulfur Bacteria in Japanese Lakes. Limnol Oceanogr. 1968;13:644–55.

335.   Camacho A, Vicente E. Carbon photoassimilation by sharply stratified phototrophic communities at the chemocline of Lake Arcas (Spain). FEMS Microbiol Ecol. 1998;25:11–12.

336.   Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualisation of de novo genome assemblies. bioRxiv. 2015;:018333.

337.   Storelli N, Saad MM, Frigaard N-U, Perret X, Tonolla M. Proteomic analysis of the purple sulfur bacterium Candidatus "*Thiodictyon syntrophicum*" strain Cad16[T] isolated from Lake Cadagno. EuPA Open Proteomics. 2014. doi:10.1016/j.euprot.2013.11.010.

338.   Hohmann-Marriott M, Blankenship RE. Variable fluorescence in green sulfur bacteria. Biochim Biophys Acta. 2007;1769:106–13.

339.   Blankenship RE, Madigan MT, Bauer CE, editors. Anoxygenic Photosynthetic Bacteria. Dordrecht: Springer Netherlands; 1995. http://link.springer.com/10.1007/0-306-47954-0. Accessed 19 Apr 2016.

340.   Pfennig N, Markham MC, Liaaen-Jensen S. Carotenoids of *Thiorhodeceae*. 8. Isolation and characterization of a *Thiothece*, *Lamprocystis* and *Thiodictyon* strain and their carotenoid pigments. Arch Für Mikrobiol. 1968;62:178–91.

341. Wagner-Huber R, Brunisholz RA, Bissig I, Frank G, Suter F, Zuber H. The primary structure of the antenna polypeptides of *Ectothiorhodospira halochloris* and *Ectothiorhodospira halophila*. Four core-type antenna polypeptides in *E. halochloris* and *E. halophila*. Eur J Biochem FEBS. 1992;205:917–25.

342. Tabita FR. Molecular and Cellular Regulation of Autotrophic Carbon Dioxide Fixation in Microorganismst. 1988;:36.

343. Prange A, Dahl C, Truper HG, Behnke M, Hahn J, Modrow H, et al. Investigation of S-H bonds in biologically important compounds by sulfur K-edge X-ray absorption spectroscopy. Eur Phys J D. 2002;20:589–96.

344. Bogorov LV. O svoĭstvakh *Thiocapsa roseopersicina*, shtamm BBS, vydelennogo iz éstuariia Belogo moria. Mikrobiologiia. 1974;43:326–32.

345. Hurlbert RE. Effect of Oxygen on Viability and Substrate Utilization in *Chromatium*. J Bacteriol. 1967;93:1346–52.

346. Kondratieva EN, Zhukov VG, Ivanovsky RN, Petushkova YP, Monosov EZ. The capacity of phototrophic sulfur bacterium *Thiocapsa roseopersicina* for chemosynthesis. Arch Microbiol. 1976;108:287–92.

347. Kämpf C, Pfennig N. Chemoautotrophic growth of *Thiocystis* violacea, *Chromatium gracile* and *C. vinosum* in the dark at various O_2-concentrations. J Basic Microbiol. 1986;26:517–31.

348. Schaub BEM, Gemerden H van. Simultaneous phototrophic and chemotrophic growth in the purple sulfur bacterium *Thiocapsa roseopersicina* M1. FEMS Microbiol Ecol. 1994;13:185–95.

349. Rákhely G, Laurinavichene TV, Tsygankov AA, Kovács KL. The role of Hox hydrogenase in the H_2 metabolism of *Thiocapsa roseopersicina*. Biochim Biophys Acta BBA - Bioenerg. 2007;1767:671–6.

350. 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.

351. Luedin SM, Pothier JF, Danza F, Storelli N, Frigaard N-U, Wittwer M, et al. Complete genome sequence of "*Thiodictyon syntrophicum*" sp. nov. strain Cad16T, a photolithoautotrophic purple sulfur bacterium isolated from the alpine meromictic Lake Cadagno. Stand Genomic Sci. 2018.

352. Thimijan RW. Photometric, Radiometric, and Quantum Light Units of Measure A Review of Procedures for Interconversion. 1983;18:5.

353. Eichler B, Pfennig N. A new purple sulfur bacterium from stratified freshwater lakes, *Amoebobacter purpureus* sp. nov. Arch Microbiol. 1988;149:395–400.

354. Widdel F, Bak F. Gram-Negative Mesophilic Sulfate-Reducing Bacteria. In: The Prokaryotes. Springer, New York, NY; 1992. p. 3352–78. doi:10.1007/978-1-4757-2191-1_21.

355. Gächter R, Mares A, Tilzer MM. Determination of phytoplankton production by the radiocarbon method: a comparison between the acidification and bubbling method (ABM) and the filtration techn1que. J Plankton Res. 1984;6:359–64.

356. Stekhoven DJ, Bühlmann P. MissForest—non-parametric missing value imputation for mixed-type data. Bioinformatics. 2012;28:112–8.

357. Keller A, Nesvizhskii AI, Kolker E, Aebersold R. Empirical Statistical Model To Estimate the Accuracy of Peptide Identifications Made by MS/MS and Database Search. Anal Chem. 2002;74:5383–92.

358. Nesvizhskii AI, Keller A, Kolker E, Aebersold R. A statistical model for identifying proteins by tandem mass spectrometry. Anal Chem. 2003;75:4646–58.

359. Ahdesmäki M, Strimmer K. Feature selection in omics prediction problems using cat scores and false nondiscovery rate control. Ann Appl Stat. 2010;4:503–19.

360. KEGG PATHWAY: Metabolic pathways - Candidatus Thiodictyon syntrophicum. https://www.kegg.jp/kegg-bin/show_pathway?tsy01100. Accessed 1 Jun 2018.

361. ProteomeXchange Datasets. http://proteomecentral.proteomexchange.org/cgi/GetDataset. Accessed 4 Jun 2018.

362. Vizcaíno JA, Côté RG, Csordas A, Dianes JA, Fabregat A, Foster JM, et al. The Proteomics Identifications (PRIDE) database and associated tools: status in 2013. Nucleic Acids Res. 2012;41:D1063–9.

363. Sahl HG, Trtiper HG. Malie Enzyme of *Chromatium vinosum*. Arch Microbiol. 1980;127:8.

364. Zuber V, Strimmer K. Gene ranking and biomarker discovery under correlation. Bioinformatics. 2009;25:2700–7.

365. Danza F, Ravasi D, Storelli N, Roman S, Luedin SM, Bueche M, et al. Microbial diversity in the water column of meromictic Lake Cadagno and evidence for seasonal dynamics. 2018.

366. Montesinos E, Esteve I. Effect of algal shading on the net growth and production of phototrophic sulfur bacteria in lakes of the Banyoles karstic area. SIL Proc 1922-2010. 1984;22:1102–5.

367. Casamayor EO, Llirós M, Picazo A, Barberán A, Borrego CM, Camacho A. Contribution of deep dark fixation processes to overall CO2 incorporation and large vertical changes of microbial populations in stratified karstic lakes. Aquat Sci. 2012;74:61–75.

368. Klawonn I, Nahar N, Walve J, Andersson B, Olofsson M, Svedén JB, et al. Cell-specific nitrogen- and carbon-fixation of *cyanobacteria* in a temperate marine system (Baltic Sea). Environ Microbiol. 18:4596–609.

369.   Miller SR, Bebout BM. Variation in sulfide tolerance of photosystem II in phylogenetically diverse cyanobacteria from sulfidic habitats. Appl Environ Microbiol. 2004;70:736–44.

370.   Gemerden H van. Survival of *Chromatium vinosum* at low light intensities. Arch Microbiol. 1980;125:115–21.

371.   Witt R de, Gemerden H van. Growth of the phototrophic purple sulfur bacterium *Thiocapsa roseopersicina* under oxic/anoxic regimens in the light. FEMS Microbiol Ecol. 1990;6:69–76.

372.   Laguna R, Joshi GS, Dangel AW, Luther AK, Tabita FR. Integrative control of carbon, nitrogen, hydrogen, and sulfur metabolism: the central role of the Calvin-Benson-Bassham cycle. Adv Exp Med Biol. 2010;675:265–71.

373.   Kornberg HL. Aspects of Terminal Respiration in Microorganisms. Annu Rev Microbiol. 1959;13:49–78.

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

375.   Beeftink HH, van Gemerden H. Actual and potential rates of substrate oxidation and product formation in continuous cultures of *Chromatium vinosum*. Arch Microbiol. 1979;121:161–7.

376.   Draber Wilfried, Tietjen Klaus, Kluth Joachim F., Trebst Achim. Herbicides in Photosynthesis Research. Angew Chem Int Ed Engl. 1991;30:1621–33.

377.   Znachor P, Nedoma J. Importance of dissolved organic carbon for phytoplankton nutrition in a eutrophic reservoir. J Plankton Res. 2010;32:367–76.

378.   Behrens S, Kappler A, Obst M. Linking environmental processes to the *in situ* functioning of microorganisms by high-resolution secondary ion mass spectrometry (NanoSIMS) and scanning transmission X-ray microscopy (STXM): NanoSIMS and STXM in environmental microbiology. Environ Microbiol. 2012;14:2851–69.

379.   Wynne C, Fenselau C, Demirev PA, Edwards N. Top-Down Identification of Protein Biomarkers in Bacteria with Unsequenced Genomes. Anal Chem. 2009;81:9633–42.

380.   Cornut J, De Respinis S, Tonolla M, Petrini O, Bärlocher F, Chauvet E, et al. Rapid characterisation of aquatic hyphomycetes by matrix assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS). Mycologia. 2017. http://repository.supsi.ch/8227/. Accessed 18 Feb 2018.

381.   Hart PJ, Wey E, McHugh TD, Balakrishnan I, Belgacem O. A method for the detection of antibiotic resistance markers in clinical strains of Escherichia coli using MALDI mass spectrometry. J Microbiol Methods. 2015;111:1–8.

382.   Lauceri R, Austoni M, Caviglia F, Kamburska L, Lami A, Morabito G, et al. Coupling a bio-accumulator organism and MALDI-TOF MS: an early warning

detection system for microcystins in water bodies. J Appl Phycol. 2017;29:2979–88.

383. Persson S, Sönksen CP, Frigaard N-U, Cox RP, Roepstorff P, Miller M. Pigments and proteins in green bacterial chlorosomes studied by matrix-assisted laser desorption ionization mass spectrometry. Eur J Biochem. 2000;267:450–456.

384. Brunisholz RA, Suter F, Zuber H. Structural and spectral characterisation of the antenna complexes of *Rhodocyclus gelatinosus*. FEBS J. 1994;222:667–675.

385. Nomura E, Katsuta K, Ueda T, Toriyama M, Mori T, Inagaki N. Acid-labile surfactant improves in-sodium dodecyl sulfate polyacrylamide gel protein digestion for matrix-assisted laser desorption/ionization mass spectrometric peptide mapping. J Mass Spectrom. 2004;39:202–7.

386. Suckau D, Resemann A. MALDI Top-Down Sequencing: Calling N- and C-Terminal Protein Sequences with High Confidence and Speed. J Biomol Tech JBT. 2009;20:258–62.

387. Meckenstock RU, Brunisholz RA, Zuber H. The light-harvesting core-complex and the B820-subunit from *Rhodopseudomonas marina*. Part I. Purification and characterisation. FEBS Lett. 1992;311:128–34.

388. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011;7:539.

389. Huisman J, Codd GA, Paerl HW, Ibelings BW, Verspagen JMH, Visser PM. Cyanobacterial blooms. Nat Rev Microbiol. 2018;:1.

390. Breitburg D, Levin LA, Oschlies A, Grégoire M, Chavez FP, Conley DJ, et al. Declining oxygen in the global ocean and coastal waters. Science. 2018;359:eaam7240.

391. Keeling RF, Koertzinger A, Gruber N. Ocean Deoxygenation in a Warming World. Annu Rev Mar Sci. 2010;2:199–229.

392. Watson M. Mind the gaps - ignoring errors in long read assemblies critically affects protein prediction. bioRxiv. 2018;:285049.

393. Kuo C-H, Moran NA, Ochman H. The consequences of genetic drift for bacterial genome complexity. Genome Res. 2009;19:1450–4.

394. Bobay L-M, Ochman H. The Evolution of Bacterial Genome Architecture. Front Genet. 2017;8. doi:10.3389/fgene.2017.00072.

395. Giovannoni SJ, Cameron Thrash J, Temperton B. Implications of streamlining theory for microbial ecology. ISME J. 2014;8:1553–65.

396. Pattaragulwanit K, Dahl C. Development of a genetic system for a purple sulfur bacterium: Conjugative plasmid transfer in *Chromatium vinosum*. Arch Microbiol. 1995;164:217–22.

397. Vettiger A, Basler M. Type VI Secretion System Substrates Are Transferred and Reused among Sister Cells. Cell. 2016;167:99-110.e12.

398. Potvin-Trottier L, Luro S, Paulsson J. Microfluidics and single-cell microscopy to study stochastic processes in bacteria. Curr Opin Microbiol. 2018;43:186–92.

399. Ionescu D, Bizic-Ionescu M, Maio N, Cypionka H, Grossart H-P. Community-like genome in single cells of the sulfur bacterium A *Chromatium oxaliferum*. Nat Commun. 2017;8:455.

400. Gahlmann A, Moerner WE. Exploring bacterial cell biology with single-molecule tracking and super-resolution imaging. Nat Rev Microbiol. 2014;12:9–22.

401. Bitler A, Dover RS, Shai Y. Fractal properties of cell surface structures: A view from AFM. Semin Cell Dev Biol. 2018;73:64–70.

402. Dobro MJ, Oikonomou CM, Piper A, Cohen J, Guo K, Jensen T, et al. Uncharacterized bacterial structures revealed by electron cryotomography. bioRxiv. 2017;:108191.

403. Sun C, Jiang F, Gao W, Li X, Yu Y, Yin X, et al. Scanning electron microscopy coupled with energy-dispersive X-ray spectrometry for quick detection of sulfur-oxidizing bacteria in environmental water samples. Chin J Oceanol Limnol. 2017;35:185–91.

404. Berg JS, Schwedt A, Kreutzmann A-C, Kuypers MMM, Milucka J. Polysulfides as Intermediates in the Oxidation of Sulfide to Sulfate by *Beggiatoa* spp. Appl Environ Microbiol. 2014;80:629–36.

405. Sakamoto M, Tilzer MM, Gächter R, Rai H, Collos Y, Tschumi P, et al. Joint field experiments for comparisons of measuring methods of photosynthetic production. J Plankton Res. 1984;6:365–83.

406. Kleiner M, Dong X, Hinzke T, Wippler J, Thorson E, Mayer B, et al. Metaproteomics method to determine carbon sources and assimilation pathways of species in microbial communities. Proc Natl Acad Sci. 2018;115:E5576–84.

407. Findlay AJ, Bennett AJ, Hanson TE, Luther GW. Light-dependent sulfide oxidation in the anoxic zone of the Chesapeake Bay can be explained by small populations of phototrophic bacteria. Appl Environ Microbiol. 2015;:AEM.02062-15.

408. Danza F. F. Danza PhD Thesis.pdf. Université de Genève; 2018.

409. Banerjee S, Schlaeppi K, van der Heijden MGA. Keystone taxa as drivers of microbiome structure and functioning. Nat Rev Microbiol. 2018. doi:10.1038/s41579-018-0024-1.

410. Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal PR, et al. Microbial diversity in the deep sea and the underexplored "rare biosphere." Proc Natl Acad Sci U S A. 2006;103:12115–20.

411. Kleindienst S, Grim S, Sogin M, Bracco A, Crespo-Medina M, Joye SB. Diverse, rare microbial taxa responded to the *Deepwater Horizon* deep-sea hydrocarbon plume. ISME J. 2016;10:400–15.

412. Lauro FM, DeMaere MZ, Yau S, Brown MV, Ng C, Wilkins D, et al. An integrative study of a meromictic lake ecosystem in Antarctica. ISME J. 2011;5:879–95.

413. Mehrshad M, Rodriguez-Valera F, Amoozegar MA, López-García P, Ghai R. The enigmatic SAR202 cluster up close: shedding light on a globally distributed dark ocean lineage involved in sulfur cycling. ISME J. 2018;12:655–68.

414. Kalenitchenko D, Le Bris N, Peru E, Galand PE. Ultra-rare marine microbes contribute to key sulfur related ecosystem functions. Mol Ecol. 2018. doi:10.1111/mec.14513.

415. Watanabe T, Kojima H, Fukui M. Complete genomes of freshwater sulfur oxidizers *Sulfuricella denitrificans* skB26 and *Sulfuritalea hydrogenivorans* sk43H: genetic insights into the sulfur oxidation pathway of *betaproteobacteria*. Syst Appl Microbiol. 2014;37:387–95.

416. Christie-Oleza JA, Sousoni D, Lloyd M, Armengaud J, Scanlan DJ. Nutrient recycling facilitates long-term stability of marine microbial phototroph–heterotroph interactions. Nat Microbiol. 2017;2:17100.

417. McInerney JO, McNally A, O'Connell MJ. Why prokaryotes have pangenomes. Nat Microbiol. 2017;2:17040.

418. Cohan FM. Bacterial Species and Speciation. YSTEMATIC Biol. 2001;50:12.

419. Bobay L-M, Ochman H. Biological Species Are Universal across Life's Domains. Genome Biol Evol. 2017;9:491–501.

420. Hannen EJ van, Zwart G, Agterveld MP van, Gons HJ, Ebert J, Laanbroek HJ. Changes in Bacterial and Eukaryotic Community Structure after Mass Lysis of Filamentous Cyanobacteria Associated with Viruses. Appl Environ Microbiol. 1999;65:795–801.

421. Abreu NA, Taga ME. Decoding molecular interactions in microbial communities. FEMS Microbiol Rev. 2016;40:648–63.

422. Schill F, Bahr A, Martinoli A. Vertex: A New Distributed Underwater Robotic Platform for Environmental Monitoring. In: Distributed Autonomous Robotic Systems. Springer, Cham; 2018. p. 679–93. doi:10.1007/978-3-319-73008-0_47.

423. Woyke T, Doud DFR, Schulz F. The trajectory of microbial single-cell sequencing. Nat Methods. 2017;14:1045–54.

424. Raj A, van Oudenaarden A. Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. Cell. 2008;135:216–26.

425. Chen Z, Chen L, Zhang W. Tools for Genomic and Transcriptomic Analysis of Microbes at Single-Cell Level. Front Microbiol. 2017;8. doi:10.3389/fmicb.2017.01831.

426. Krismer J, Steinhoff RF, Zenobi R. Single-cell MALDI Tandem Mass Spectrometry: Unambiguous Assignment of Small Biomolecules from Single Chlamydomonas reinhardtii Cells. Chim Int J Chem. 2016;70:236–9.

427. Project: The ProteOMZ Expedition: Investigating Life Without Oxygen in the Pacific Ocean | BCO-DMO. https://www.bco-dmo.org/project/685696. Accessed 3 Jul 2018.

428. Bush T, Diao M, Allen RJ, Sinnige R, Muyzer G, Huisman J. Oxic-anoxic regime shifts mediated by feedbacks between biogeochemical processes and microbial community dynamics. Nat Commun. 2017;8. doi:10.1038/s41467-017-00912-x.

429. Casty C, Wanner H, Luterbacher J, Esper J, Bohm R. Temperature and precipitation variability in the european Alps since 1500. Int J Climatol. 2005;25:1855–80.

430. Huggel C, Clague JJ, Korup O. Is climate change responsible for changing landslide activity in high mountains?: CLIMATE CHANGE AND LANDSLIDES IN HIGH MOUNTAINS. Earth Surf Process Landf. 2012;37:77–91.

431. Edwards A, Debbonaire AR, Sattler B, Mur LA, Hodson AJ. Extreme metagenomics using nanopore DNA sequencing: a field report from Svalbard, 78 N. 2016. http://biorxiv.org/lookup/doi/10.1101/073965. Accessed 16 Sep 2016.

432. Loman NJ, Quinlan AR. Poretools: a toolkit for analyzing nanopore sequence data. Bioinformatics. 2014;:btu555.

433. Kim D, Song L, Breitwieser FP, Salzberg SL. Centrifuge: rapid and sensitive classification of metagenomic sequences. Genome Res. 2016;26:1721–9.

434. Breitwieser F. pavian: Interactive analysis of metagenomics data. R. 2017. https://github.com/fbreitwieser/pavian. Accessed 11 Sep 2017.

435. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–10.

436. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinforma Oxf Engl. 2010;26:589–95.

5 APPENDIX

5.1 Unpublished experimental data and protocols

5.1.1 Proteomic data from in vitro str. Cad16T cultivated at light and dark

In 2016 we did a first *in vitro* experiment on dark carbon fixation of strain Cad16^T using LFG-LC-MS^2 based protoemics. However, due to technical difficulties at the experimental setting (high temperatures and missfunctions of the light timer) and the ambiguos results in the protein profiles were we did not sea a distinct grouping according to day and night, we abandoned the project. Therefore, we attached below the report on protein quantitation provided by the FGCZ.

FGCZ Two-Group Analysis: p1669 Statistics for a Quantitative Protein Matrix

Functional Genomics Center Zurich

24 February, 2017

Contents

1 Input Matrix

Experiment is called: p1669

1.1 The Numbers

The protein matrix is filtered like this:

- Minimum number of peptides / protein: 2
- Maximum of missing values per protein : 6

The total number of proteins in this experiment is: 1552

2 The Groupings

Condition	# samples
dark	5
light	5

Here the files in each group:

Condition	Raw.file
dark	20170117_04_dark_Cad16T_9
dark	20170117_05_dark_Cad16T_7
dark	20170117_08_dark_Cad16T_8
dark	20170117_09_dark_Cad16T_10
dark	20170117_13_dark_Cad16T_6
light	20170117_02_light_Cad16T_3_rep
light	20170117_03_light_Cad16T_1
light	20170117_07_light_Cad16T_2
light	20170117_10_light_Cad16T_5_
light	20170117_12_light_Cad16T_4

2

3   Proteins Used for Quantitation

The input matrix has the following structure (Figure 1).
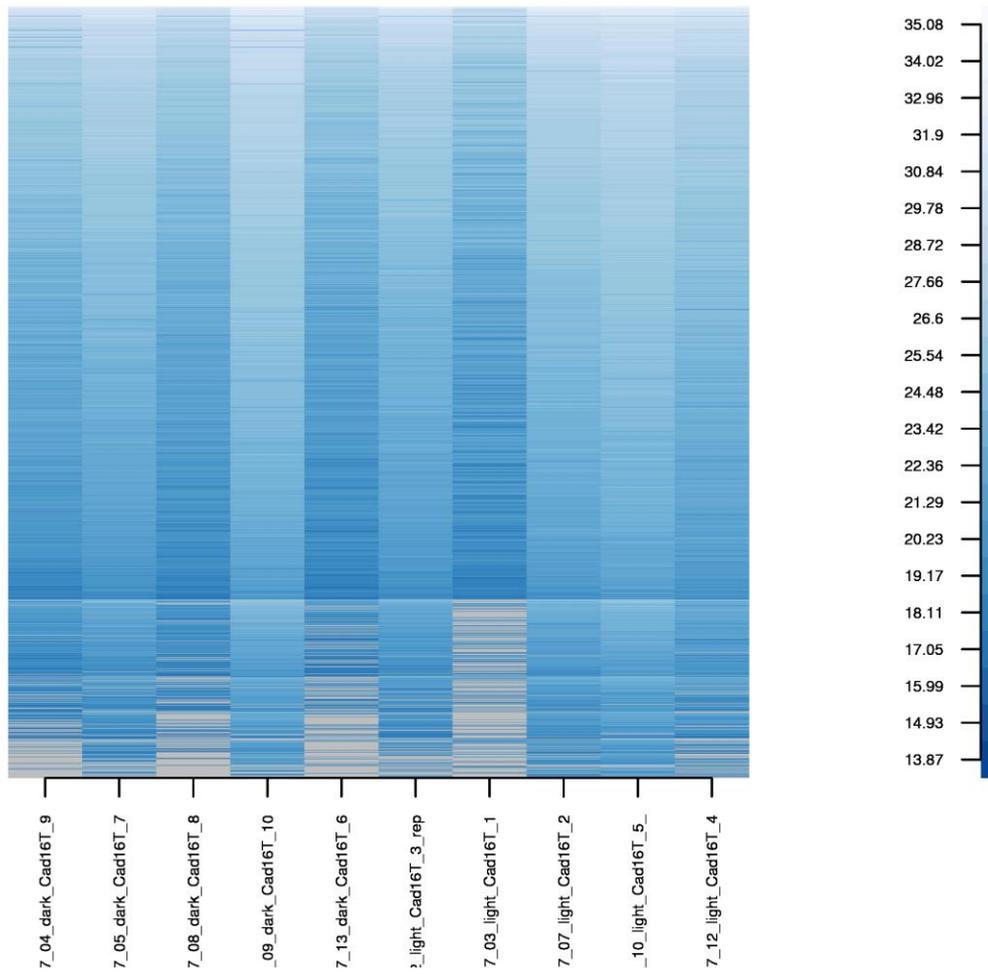
Figure 1: Heatmap for quantifyable proteins sorted by missigness and intensity (log2). Gray - missing values

3
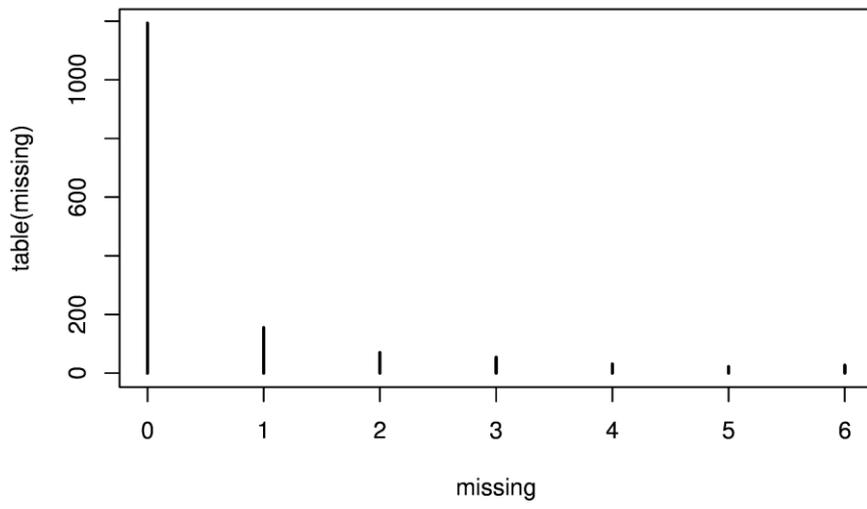
Figure 2: # of missing values per protein

4

4 Distribution of Intensities

Shown in Figure 3 are the un-normalized values while in Figure 4 are the z transformed values (subtracted median and divided by variance).
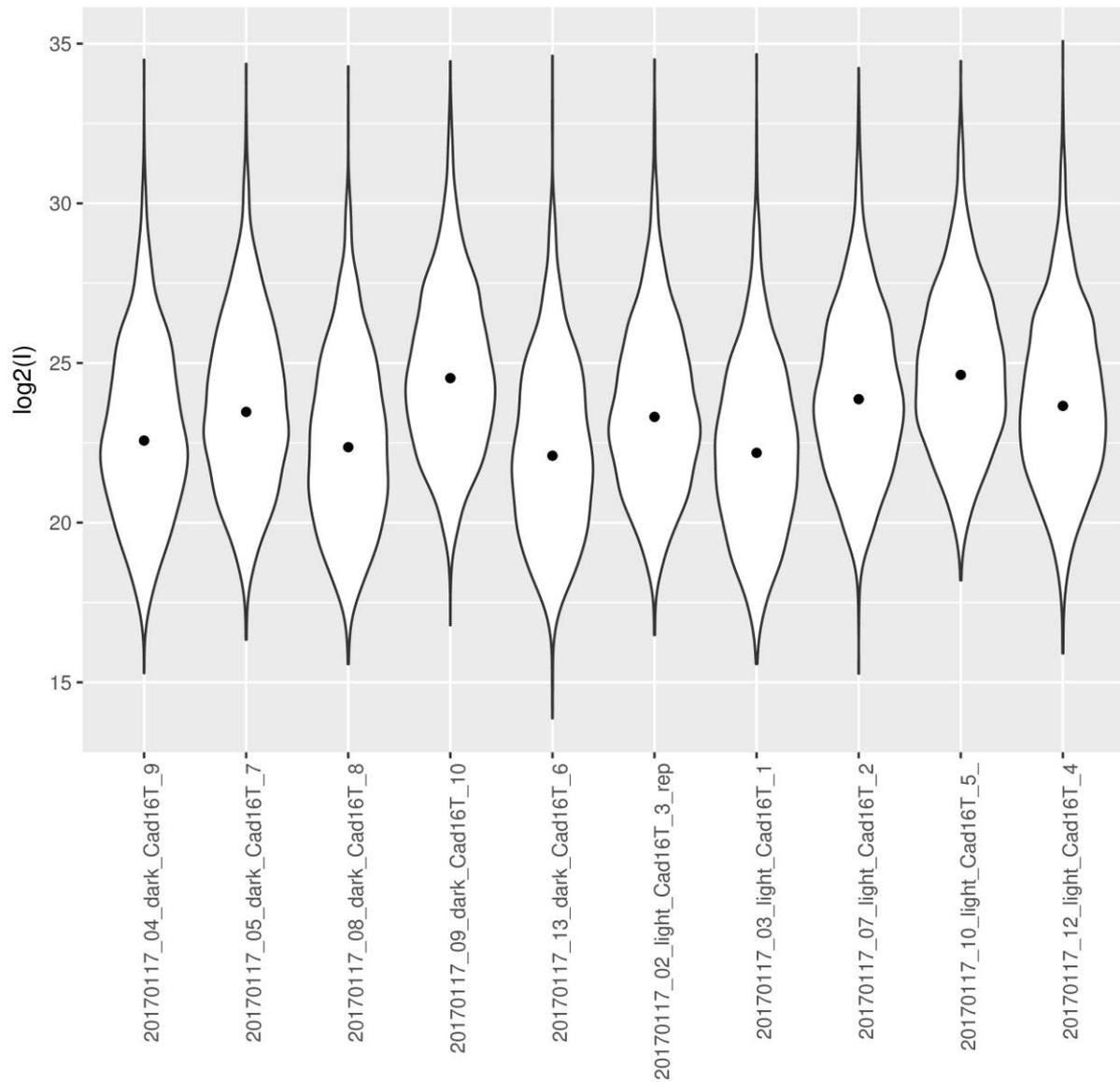
Figure 3: Density plot for quantifyable proteins (log2 transformed)

The effect of the normalization (z transformation) is visualized in Figure 5. It shows the log2-median value which is subtracted from each individually sample. Large differences in these values are critical.

5
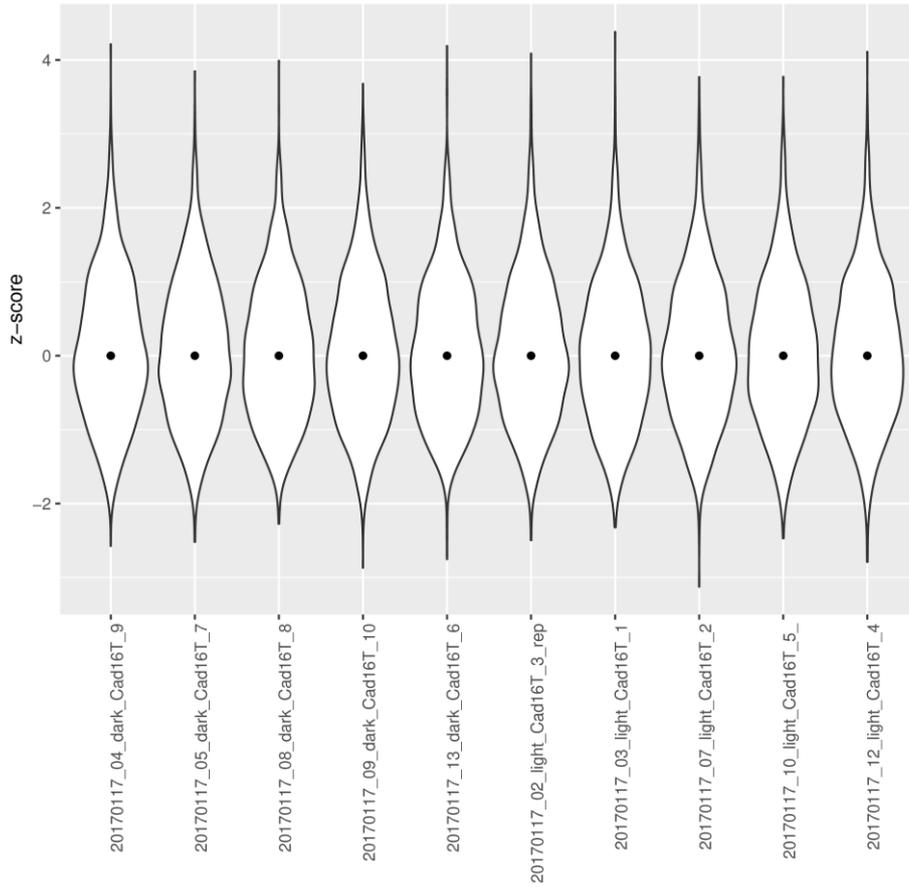
Figure 4: Density plot for normalized values (z-score)

6
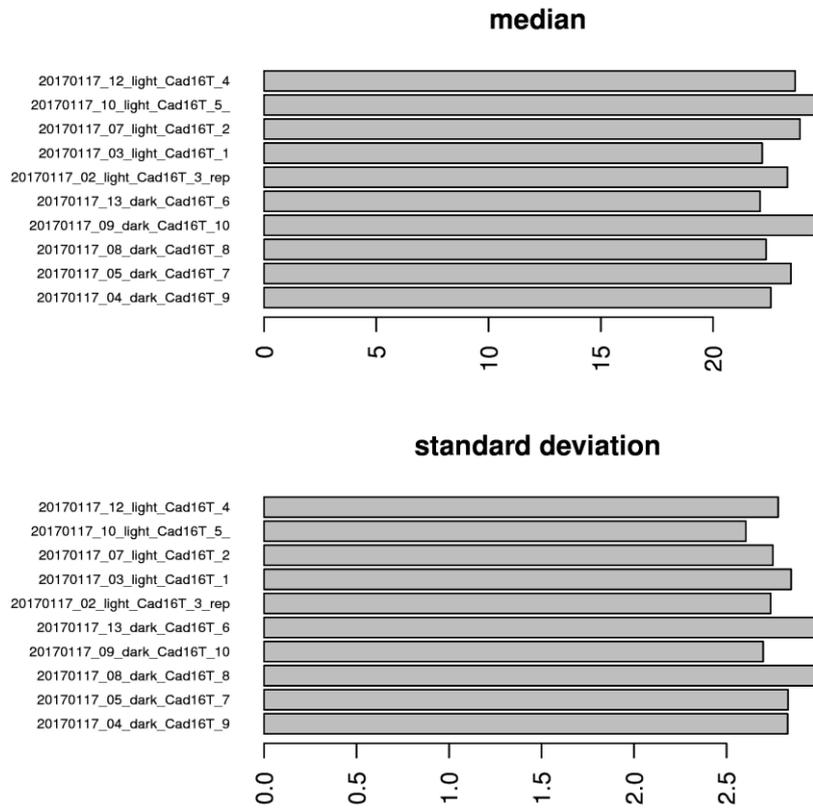
Figure 5: Median values subtracted for normalization.

7

Table 1: median of cv

condition	cv
dark	84.97600
light	60.33416
all	71.78575

5 Coefficients of Variations
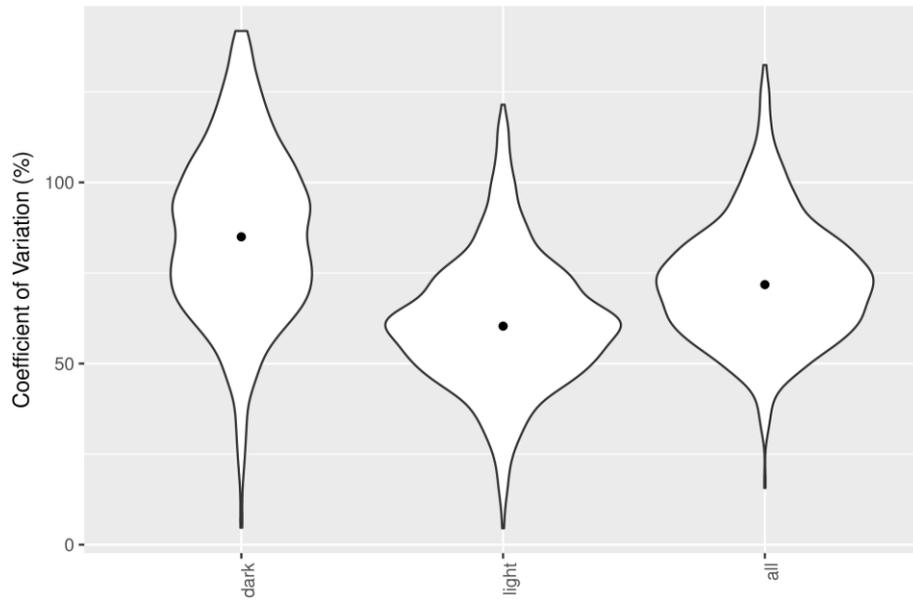
Figure 6: distribution of protein CV within condition and overall

Table 2: median of sd

condition	sd
dark	0.1509679
light	0.1895577
all	0.1995221
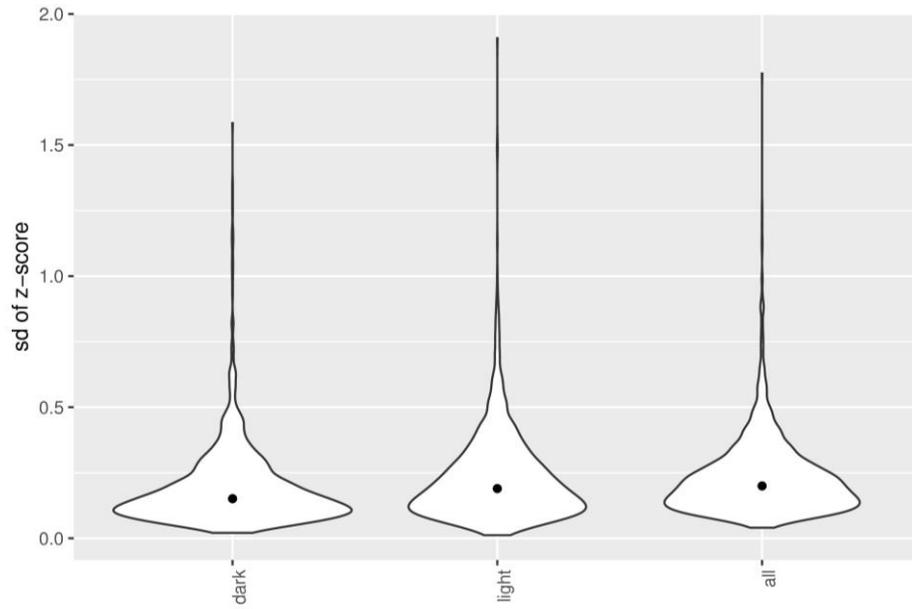
Figure 7: Distribution of protein standard deviation (after sample normalization and scaling) within conditions and overall

6 Heatmaps and Clustering for Samples and Proteins
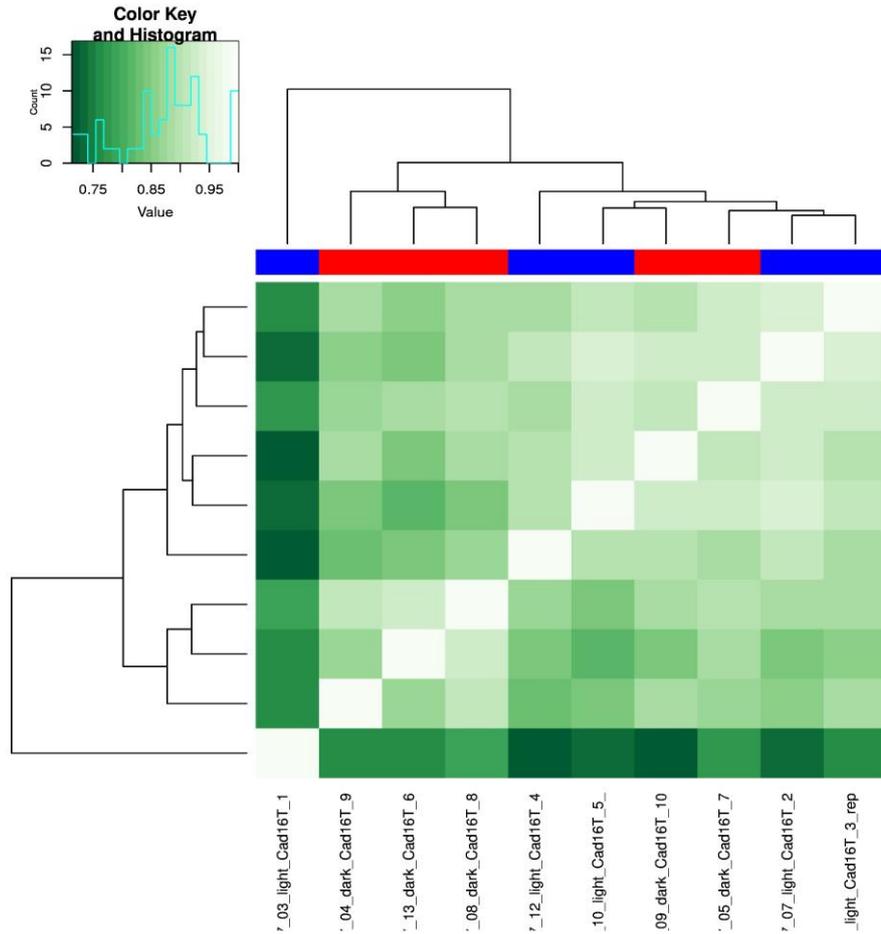
Figure 8: Heatmap of correlations (spearman) between samples.

In Figure 9 and Figure 8 we show how samples are clustering depending on their correlation and on the protein expression profiles.

7 FGCZ - Two Group Analysis

Here we show the result of a two group analysis for the normalized matrix using the bioconductor limma package.

10
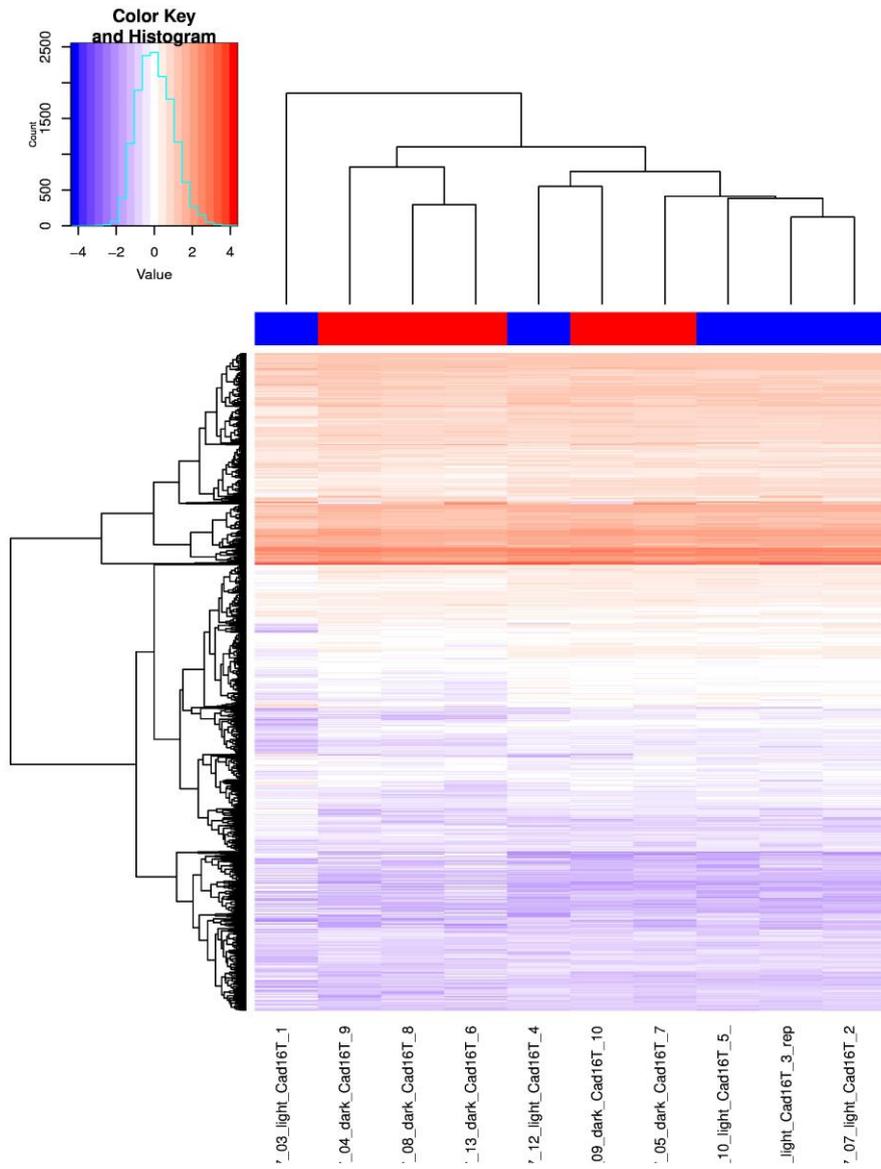
Figure 9: Heatmap for normalized and scaled values (z-scores).

11

Table 3: Most significant regulated proteins.

regulation	Accession	log2(light/dark)	moderated qvalue	(light/dark)
up	fig_572262_15_peg_1923	-4.281819	0.0007426	0.0514096
up	fig_572262_15_peg_1065	-2.779778	0.0233355	0.1456141
up	fig_572262_15_peg_5789	-2.047856	0.0309668	0.2418433
up	fig_572262_15_peg_2590	-3.181695	0.0341163	0.1102083
up	fig_572262_15_peg_2026	-3.064102	0.0428704	0.1195676

7.2 Adjusted p-values (q-values) (normal t-test)

Significant calls are made with a p Value smaller 0.01.

Table 4: Most significant regulated proteins.

regulation	Accession	log2(light/dark)	p-value	(light/dark)
up	fig_572262_15_peg_1923	-4.281819	0.0035351	0.0514096

7.3 Output Data Description

columns
ProteinName
TopProteinName
nrPeptides
effectSize
p.ord
p.mod
q.ord
q.mod
log2FC
nrNAs
20170117_04_dark_Cad16T_9.raw

7.4 References

The q-values and p-values were computed using the bioconductor package limma.

7.5 Disclaimer and Acknowledgements

This document was generated using Rmarkdown and processes text files which are generated with a label-free quantitation software such as MaxQuant or Progenesis. The obtained results should be validated orthogonally as well (e.g. with Western blots). The Functional Genomics Center Zurich does not provide any kind of guarantee of the validity of these results.

5.1.2 Collaborations

- Danza F, Storelli N, Roman S, Lüdin S, Tonolla M. Dynamic cellular complexity of anoxygenic phototrophic sulfur bacteria in the chemocline of meromictic Lake Cadagno. PLOS ONE. 2017;12:e0189510.

- Danza F, Ravasi D, Storelli N, Roman S, Luedin SM, Bueche M, et al. Microbial diversity in the water column of meromictic Lake Cadagno and evidence for seasonal dynamics. 2018. In submission

5.2 Reports

MinION@Cadagno – Exploring the functional potential of the Lake Cadagno microbiome through nanopore sequencing, a pilot study

Nicole Liechti, Corinne P. Oechslin, Matthieu Bueche, Samuel M. Luedin, Francesco Danza, Nicola Storelli, Joël F. Pothier, Christian Beuret, Matthias Wittwer and Mauro A. Tonolla

Unpublished pilot study

Statement of contribution

I contributed to sampling, sample preparation, DNA extraction and data interpretation and manuscript preparation.

Research objective

The aim of the study was to test for the technical feasibility of future nanopore based metagenomic sequencing of the complex microbial community of Lake Cadagno under field conditions.

5.2.1 Introduction

The microbial community of Lake Cadagno (Ticino, Switzerland) has been studies for decades and represents an ancient ocean homologue due to its permanent stratification. It is situated in the Swiss Alps at about 2'000 meters above sea level [122]. Bottom water rich in sulfide provides the living condition for a dense population of mostly anoxygenic photosynthetic bacteria, belonging to the families of *Chromatiaceae* and *Chlorobiaceae* [154]. However, the microbial population has only been classified by 16S rRNA gene sequences [143]. This allows only a limited insight into the complex community with a taxonomic resolution to the genus level. As with the sequencing decreasing cost and ease of DNA sample preparation more bacterial ecosystems are studied using meta-omics. Thereby,

the functional potential can be studied, as well as insight into phylogeny and evolution of microbial communities can be gained in unseen resolution.

The recent development of nanopore-based sequencing techniques, fast library preparation protocols, as well as hard- and software enables on site genomic studies with minimal lab equipment [431]. The Research Center in the Piora valley provides an ideal training ground to do nanopore-based metagenomics in a remote, minimal lab setting. As we expect the autumnal community shift in the interphase (chemocline at 12–14 m depth) from the dominating large celled *Chromatium okenii* (up to 10^5 cells ml^{-1}) to small celled aggregated *Thiodictyon* and *Lamprocystis* type PSB in September, we plan to take samples then to obtain a profile from a population with a larger beta-diversity. This may help to further disentangle the different biochemical cycles present in Lake Cadagno. With this study we tested feasibility of on-site sequencing at Lake Cadagno.

5.2.2 Methods

Experiments were performed at the Alpine Biology Center in Piora located at 2'000 m a.s.l. The field lab was set up in a simple field lab equipped with a centrifuge for 50 ml tubes and 1.5ml Eppendorf tubes as well as -20 °C and 4 °C fridges, water bath and heat block (Figure 5.2-1). In Mountain areas of this elevation the average day temperature is around 10 °C and drops to 0 °C during night in September. Therefore, low temperatures may affect experimental conditions and slow down the processes of sequencing library preparation. The conditions for the sequencing process are then controlled by the MinIon device.

Figure 5.2-1 | Overview on the minimal lab equipment used for sample preparation, DNA extraction, library preparation and on-site sequencing in this study

DNA extraction and purification

For cell lysis 200 μl AVL buffer (Qiagen, Switzerland), 20 μl Proteinase K, 4 μl RNaseA (Qiagen) as well as approximate 25mg Lysozyme was added. After incubation for 5 min at room temperature (warmed by hand) cells were lysed by incubating at 56°C for 30 min. DNA was extracted using Qiagen DNeasy Blood and Tissue kit (Qiagen) according to manufactures protocol performing two AW1 washing steps and elution of the DNA in to steps using 100 μl prewarmed nuclease free water. The amount of DNA was measured using the dsDNA HS assay 2.0 on a Qubit fluorometer (Thermo Fisher). To concentrate the DNA, sample was purified using Agencount AMPure Beads (Beckmann, Coulter) according to manufactures protocol and eluted in 50 μl water.

DNA library preparation and MinION sequencing

Nanopore sequencing library was prepared using Rapid Sequencing Kit RAD003 (Nanopore, Oxford, UK) with a total DNA input of 200 μg in 7.5 μl. For the reaction 1.5 ml DNA low bind Tubes (Eppendorf, Switzerland) was used. Tagmentation step was performed for 2 min at 30 °C in water bath and inactivation step at 80 °C for 2 min using a heat block. Prior to

sequencing, the Flow Cell QC was conducted, and the number of active pores was examined. Finally, the library was loaded to the SpotON Flow Cell R9.4 and sequenced for 48h. As temperature in the mountain areas at time point of sampling drops to around 0°C during the night, the MinION device was enclosed in a Styrofoam box to prevent gadget from cooling during sequencing run.

Binning & Assembly

'Live' base calling was performed during sequencing using MinKNOW(version). Poretools v0.6.0 [432] (Loman and Quinlan, 2014) was used to extract reads in fastq format from base called fas5 files and to perform basic statistical analysis. For the taxonomic classification of the reads, *de novo* assembled genomes of *C. okenii* str. LaCa str. Cad16^T, and str. CadA31 were added to the NCBI nr database (accessed 2018). Reads with a minimal length of 500bp were then classified using Centrifuge [433]. Output files were visualized using Pavian [434], an R based application for visualizing metagenomic datasets. In a second step, reads were aligned to the same database (NCBI nr + 3 additional genomes) using blastn (default parameters) [435]. Further, reads were mapped to the species of interest (*C. okenii*, str. Cad16^T, str. CadA31) using bwa mem specifying *-ont* option for nanopore data [436].

5.2.3 Results

Sequencing on MinIon during 48 h resulted in a total of 40'017 reads with a mean read length of 3'336 bp. For further analysis reads shorter than 500 bp were removed, resulting in 23'531 reads for downstream analysis. By using centrifuge in combination with the NCBI nt database with 3 additional de novo assemblies originated from Lake Cadagno bacterial strains, a total of 19'003 reads could be taxonomically classified. Out of them, 11'979 are considered as Eukaryotic sequences while 6'697 belong to Bacteria, 36 to Archaea and 88 to viral species (Figure 5.2-2). The PSB *Chromatium okenii* dominates the community whereas members of the class *Actinobacteria*, *α-* and *β- proteobacteria* are also abundant. The dominance of *C. okenii* findings is in accordance with the flow cytometry based counting in August 2017 [134].
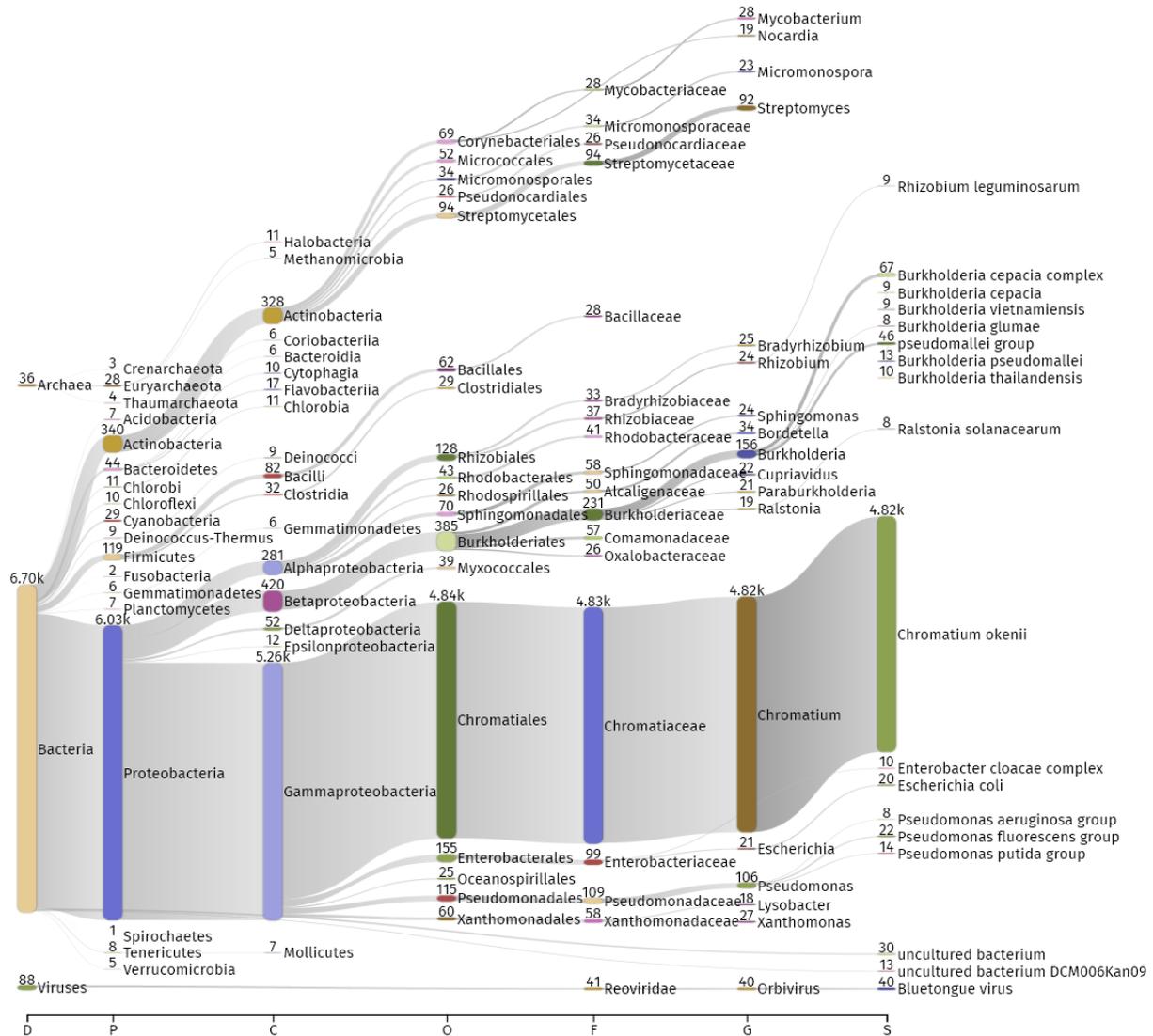
Figure 5.2-2 | Taxonomic tree of the Lake Cadagno metagenomic chemocline sequences after centrifuge classification using the NCBI nr database and the complete genomes from PSB str. Cad16T, CadA31 and *C. okenii* LaCa. Numbers indicate the reads allocated for each taxonomic level. Scale indicates the hierarchical taxonomic level from left to right; D: Domain, P: phylum C: class, O: order, F: family, G: genus and S: species.

5.2.4   Discussion

We showed the feasibility of metagenomic sequencing under field conditions at a simple laboratory setting at 2'000 m.a.s.l. However total DNA concentration was low and possible contaminations with eukaryotic occurred. The sample composition might has been biased by the centrifugation step, thereby enriching for large celled PSB *C. okenii*. Therefore we plan to repeat the experiment in September 2018 with an improved sampling protocol involving tangential flow filtration (TFF), CTD measurements and FC, DAPI and FISH enumeration of the samples. Salt free extraction buffers possibly improve protease and RNAase performance during DNA purification. As taxonomic classification is dependent on the database, we will

include also a dataset with illumina based metagenomic bins from 2013/2014 (Petra Pjevac, University of Vienna, personal communications)

6 ACKNOWLEDGEMENTS