



Department of Plant and Microbial Biology

Master's Thesis

**Ecological and Genomic Characterization of
Streamlined, Rhodopsin-bearing Freshwater
Microbes**

Melissa Kaspar

Supervised by Dr. Michaela M. Salcher

Group of Prof. Dr. Jakob Pernthaler

Limnological Station, Kilchberg

University of Zurich

July 2017

Summary

'*Candidatus Methylospumilus* sp.' is a proposed bacterial taxon isolated from the plankton of lakes. This genus includes two suggested species, which are tentatively named '*Candidatus Methylospumilus planktonicus*' and '*Candidatus Methylospumilus turicensis*'. They are characterized by a small, streamlined genome (1.3 Mb and 1.75 Mb, respectively) and methylotrophic lifestyle, which allows them to utilize substrates without carbon-carbon bounds for assimilation, i.e. methanol or methylamine. By utilizing these alternative carbon resources instead of the more abundant dissolved organic matter (DOM), methylotrophs play an important role in the overall carbon cycle, since they recycle carbon that is not easily accessible by other organisms.

In the scope of the presented thesis, more strains of '*Ca. Methylospumilus planktonicus*' were sequenced and put in phylogenomic relation to each other, and to the closely related, marine OM43 clade. Thereby, evidence was supplied that '*Ca. Methylospumilus planktonicus*' is apparently not only consisting of an unitary species, but is divided into three groups. Moreover, an analysis of the percentage of common proteins suggests that the marine OM43 clade also belongs to the same genus of '*Ca. Methylospumilus* sp.'. It is assumed that two habitat transitions of the ancestral lineage have occurred, firstly from the sediment of lakes to the freshwater plankton, and secondly from limnic systems to marine environments.

Another feature of '*Ca. Methylospumilus* sp.' is the presence of rhodopsins (light-driven proton pumps), meaning that the candidate genus can further utilize light to produce energy in the form of ATP. To examine the influence of rhodopsins on growth of '*Ca. Methylospumilus planktonicus*', a growth assay was conducted with conditions varying in light-exposure and carbon concentration. The response of the bacteria to the various conditions was quantified by measuring their cell densities by flow cytometry, and additionally by assessing the cell volumes and morphotype-frequencies by epifluorescence microscopy of DAPI-stained cells.

There were tendencies of a fitness-advantage in light-grown cultures, as the cell densities were higher, more biomass could be assimilated, and the carbon uptake was higher when exposed to light. Furthermore, there was a shift from Vibrio-shaped forms to rounder morphotypes over time in both light- and dark-grown cultures. Unfortunately, it was not possible to examine differences under starvation conditions, in which the dependence on rhodopsins might have been more apparent.

All in all, this thesis presents an extensive characterization of '*Ca. Methylospumilus planktonicus*' in terms of genomic properties and growth behavior in varying conditions and can thus contribute to the official approval of the novel species description.

Contents

1 Introduction	1
1.1 Methylotrophy in limnic systems	1
1.2 Rhodopsins: light-driven proton pumps	2
1.3 Aims of the thesis	5
2 Material and methods	6
2.1 Genomics	6
2.1.1 DNA isolation and whole genome sequencing	6
2.1.2 Phylogenomics	7
2.2 Growth experiment	8
2.2.1 Biomass and morphotype assessment	9
2.3 Preparations for transcriptomics	10
3 Results	12
3.1 Genomics of ' <i>Ca. Methylospumilus</i> sp.'	12
3.1.1 Sequencing and characterization of novel strains	12
3.1.2 Phylogenetic distances	14
3.1.3 Metabolic features	18
3.2 Growth experiment	22
3.2.1 Cell densities	24
3.2.2 Differences in biomass	24
3.2.3 Carbon uptake	26
3.2.4 Differences in morphotypes	27
4 Discussion	28
4.1 Evolution of ' <i>Ca. Methylospumilus</i> sp.'	28
4.1.1 Hypothesized habitat transitions	28
4.1.2 Lateral gene transfer of rhodopsin genes	30
4.2 Growth experiment	32
4.2.1 Insights and limits	32
5 Final remarks	35
References	i
Supplementary	vii
Statement of authorship	xvii

List of Figures and Tables

1	Categorization of rhodopsins	3
2	Absorption spectrum of rhodopsins	4
3	Previous work on rhodopsin-bearing bacteria	4
4	Experimental setup	8
5	Overview of the genome-sequenced strains	13
6	Overview of genomic features	13
7	16S rRNA tree	15
8	COG-based tree	15
9	ANI-matrix	16
10	POCP-matrix	17
11	Simplified illustration of methane metabolism	20
12	Schematic overview of methane metabolism	20
13	Methanol dehydrogenases tree	21
14	Rhodopsin tree	21
15	pH values over time	22
16	Microscope image of MMS-VI-25	23
17	Microscope image of MMS-M-34	23
18	Growth curves	25
19	Biomass of cells	25
20	Carbon uptake - graph	26
21	Shift of morphotypes over time	27
22	Habitat transitions	29
23	Phylogenomic tree of rhodopsins in diverse groups	31
24	Retinal synthesis cassette	31
25	pH differences by dark/light or by strain	33
S9	Carbon uptake - values	xiii

Acknowledgment

First of all, I thank my supervisor **Dr. Michaela M. Salcher** for providing a highly interesting project which allowed me to explore various methods both in the lab and on the computer. I was lucky to have such an excellent supervision.

I also thank **Prof. Dr. Jakob Pernthaler** for the opportunity to do my Master's Thesis at the institute, and **Dr. Stefan Neuenschwander** for all the help at the GDC.

Furthermore I thank all members of the **Limnological Station** in Kilchberg for an unforgettable time and all the laughs.

It would not have been possible without my **family**. Thank you for all the support during my studies, and for believing in me.

Most importantly, I thank **Dominik Schöni** for all the help, technical and emotional support, and steady coffee supply to my desk. Thank you for everything.

1 Introduction

1.1 Methylotrophy in limnic systems

Freshwater ecosystems are shaped by a variety of abiotic and biotic factors, influencing the temporal and spatial distribution of microbes throughout the water column. The influence of provided nutrients on the bacterial abundances is well observable in Lake Zurich, a large, prealpine, oligomesotrophic lake, where the annual natural complete water turnover has been deferred as a consequence of global warming [Posch et al., 2012]. Without the holomixis, the oxygen-rich surface water is not brought to deeper layers, which leads to hypoxic conditions in the hypolimnion and also to stronger thermal stratification. Further, nutrients such as phosphorus and nitrogen accumulate in the deep hypolimnion without being mixed to the surface, thus, leading to an artificial oligotrophication. These conditions favor growth of highly competitive organisms, and thus increase the pressure of nutrient limitation even further. In light of this competition, the various bacterial groups have adapted differently to the limnic environment, for example by pursuing an evasive, defensive strategy or by specializing on alternative energy sources.

Methylotrophy is an example for such a specialized metabolism, which allows microbes to utilize substrates without carbon-carbon bonds (C1 compounds), such as methane, methanol or methylamine [Chistoserdova, 2011; Anthony, 1982]. Methanotrophs, a subclassification of methylotrophs, are able to metabolize sedimentary methane [Jacob et al., 2005], or to oxidate methane released by methanogenesis of other microbes [Kuivila et al., 1988; Crowe et al., 2011]. Oxidation of methane in turn partly releases methanol, making it available to non-methanotrophic methylotrophs. Other methanol sources in lakes are associated with decay of algae [Galbally and Kirshtein, 2002; Jacob et al., 2005], or allochthonous input from the catchment. By utilizing these alternative carbon resources instead of the more abundant dissolved organic matter (DOM), methylotrophs play an important role in the overall carbon cycle since they recycle carbon that is not easily accessible by other organisms. Aquatic methylotrophic members can be found in Alpha-, Beta- and Gammaproteobacteria, Verrucomicrobia, and Actinobacteria [Chistoserdova, 2015]. The most common aquatic methylotroph is the marine OM43 clade within Betaproteobacteria [Giovannoni et al., 2008]. This clade contains several cultured strains, which are all characterized by a small, streamlined genome and a planktonic, free-living lifestyle. Several freshwater sister clades of the OM43 have been described based on their 16S rRNA genes [Zwart et al., 1998], including the "sequence type LD28", which was further investigated by Salcher et al. in 2015 [Salcher et al., 2015]. They described a novel genus within *Methylophilaceae* isolated from Lake Zurich, which they proposed to name '*Candidatus Methylopumilus* sp.'. This genus includes the sequence type LD28, tentatively named '*Candidatus Methylopumilus planktonicus*', and a newly isolated sister taxon with the suggested name '*Candidatus Methylopumilus turicensis*'. As they are closely related to members of the marine OM43 clade, the candidate genus is also methylotrophic. Genomic analy-

sis revealed that its members lack a methane dehydrogenase and thus are dependent on methanol or methylated amines for carbon uptake. Furthermore, they have a reduced, streamlined genome which consists of 1.3 megabases (Mb) in '*Candidatus Methylospumilus planktonicus*' and 1.8 Mb in '*Candidatus Methylospumilus turicensis*', respectively.

More strains have been isolated since, some members of the species of '*Ca. Methylospumilus planktonicus*' have also been detected in the Řimov Reservoir in Czech Republic. As mentioned above, the "sequence type LD28" has been described before, and 16S rRNA sequences highly similar to LD28 and thus highly similar to '*Ca. Methylospumilus planktonicus*' have been frequently detected in many other aquatic systems [Newton et al., 2011]. This includes Lake Loosdrecht (NL), from where it originally has been described [Zwart et al., 1998], Toolik Lake ("Conc33", AK, USA) [Bahr et al., 1996], Adirondack Mountain lakes ("ACK-C30", NY, USA) [Hiorns et al., 1997], several Scandinavian lakes [Lindström et al., 2005], or Piburger See ("PIB-38" et seq., AT) [Salcher et al., 2008], to name a few.

It is thus hypothesized that '*Ca. Methylospumilus planktonicus*' is very widely distributed in limnic systems all over the world, suggesting that its specialized metabolism and adaption by genome-streamlining and other mechanisms makes it highly competitive in the oligotrophic pelagic.

1.2 Rhodopsins: light-driven proton pumps

Another notable feature of both the marine OM43 clade and the candidate genus is the possession of rhodopsins. These are light-driven ion pumps, generating energy in the form of ATP and thus enabling a photoheterotrophic lifestyle. While chlorophyll systems have a highly complex structure, consisting of several substructures, rhodopsins on the other hand consist of only two parts: one opsin-protein and a retinal bound to it [Pinhassi et al., 2016]. Despite their simple composition, as of 2014 there were seven described types of rhodopsins occurring in *Bacteria*, *Archaea*, *Eukarya* and reportedly even in viruses (Figure 1a) [Béjà and Lanyi, 2014; Pinhassi et al., 2016]. However, several studies imply a much higher number of rhodopsin variants and show that they are more widespread throughout all three domains of life than previously thought, especially in aquatic environments. A recent study by Govorunova et al. suggests a categorization of the microbial rhodopsin superfamily into 13 subgroups (see Figure 1b) [Govorunova et al., 2017].

While '*Ca. Methylospumilus planktonicus*' encodes for two kinds of rhodopsins (proteo- and xanthorhodopsins), '*Ca. Methylospumilus turicensis*' and the marine OM43 clade only carry one type (proteorhodopsins, or xanthorhodopsins respectively). Proteorhodopsins are the most common rho-dopsins on earth and are mainly found in aquatic bacteria such as the Proteobacteria and Bacteroidetes groups, but also in members of the *Archaea* domain. Xanthorhodopsins, however, have another molecule bound to the opsin in addition to the retinal, that is a carotenoid named

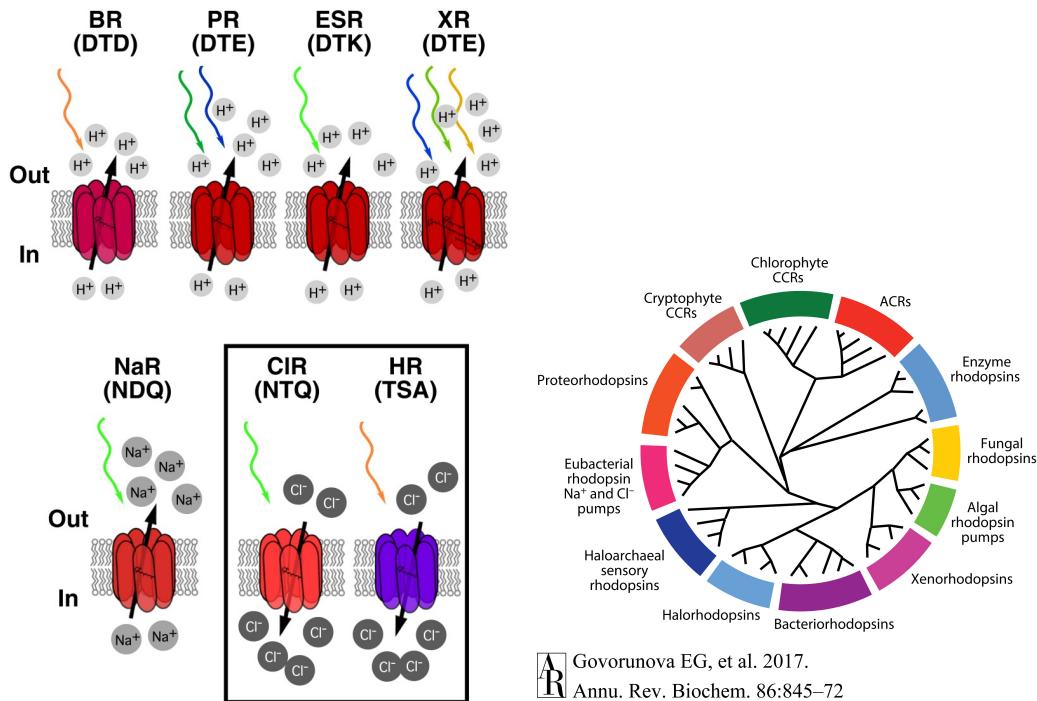


Figure 1: Proposed categorizations of microbial rhodopsins.

salinixanthin [Balashov et al., 2005]. Even though proteorhodopsins and xanthorhodopsins may be similar in their setup (aside from salinixanthin), they are clearly distinct by their sequence identity [Luecke et al., 2008]. Another difference between these two rhodopsin types is their absorption spectrum; while proteorhodopsins are tuned to blue or green light (absorption peaks around 490 nm or 540 nm [Béjà et al., 2001]), the light absorption of xanthorhodopsins shows peaks at 457 nm, 487 nm and 521 nm [Balashov et al., 2005], enabling them to utilize a slightly broader spectrum of light in the blue to green range (see Figure 2) [Fuhrman et al., 2008].

It is unclear if rhodopsins contribute to cellular activity as a crucial energy provider, or if they are utilized as a subordinate backup solution e.g. in starvation periods. To examine their relevance, and also their effective contribution to microbial growth, experiments with various rhodopsin-carrying bacteria have been conducted, such as with the extremely halophilic bacterium *Salinibacter ruber*, in which xanthorhodopsins were described for the first time [Balashov et al., 2005; Boichenko et al., 2006]. Other aquatic, rhodopsin-carrying bacteria such as *Polaribacter* sp. [González et al., 2008], 'Candidatus Pelagibacter' [Steindler et al., 2011], *Dokdonia* sp. [Palovaara

et al., 2014] and a marine *Vibrio* strain [Gómez-Consarnau et al., 2010] have also been studied in regard of quantitative measures linked to growth and/or proton pump activity (see Table 3).

These studies have shown that changes observed during bacterial growth can be partly attributed to rhodopsins either due to differences in light conditions or by comparison with non-photoheterotrophic bacteria. Furthermore, not only biomass-accumulation is influenced, but also the uptake of various molecules, or the morphological appearance.

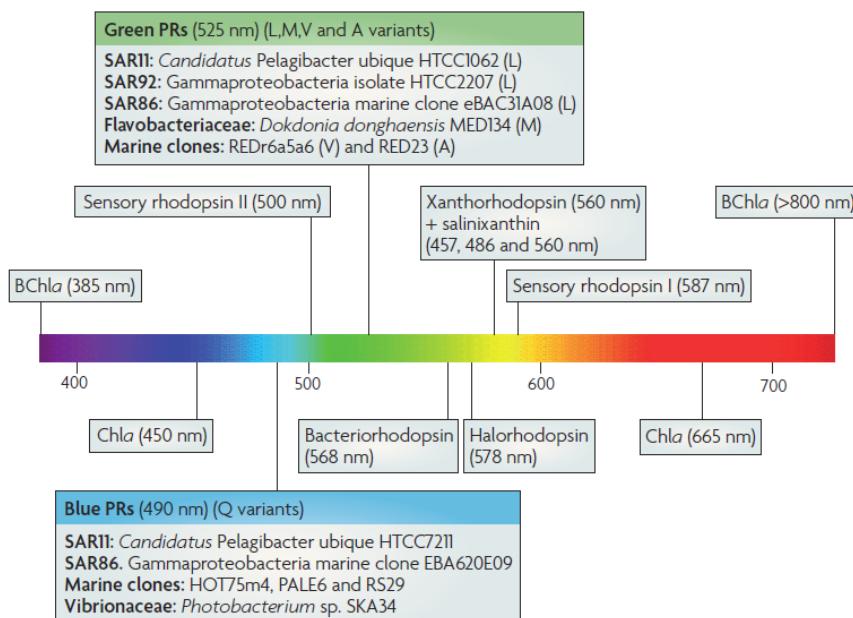


Figure 2: Absorbance maxima for various rhodopsin- and chlorophyll-types.

Name, Affiliation	Author	Source	Rhodopsin	Measurement
1 <i>Salinibacter ruber</i> M31, Bacteroidetes	Boichenko2006	Crystallizer ponds (Spain)	XR & HR	Oxygen uptake
2 <i>Polaribacter</i> sp. MED152, Bacteroidetes	Gonzalez2008	Mediterranean Sea (Spain, 0.5m)	PR	Bicarbon uptake
3 <i>Ca. Pelagibacter</i> HTCC1062, Alphaproteobacteria (SAR11 clade)	Steindler2011	Pacific Ocean (USA, OR, 10m)	PR	Oxygen concentration, ATP content, Taurine uptake, morphology
4 <i>Dokdonia</i> sp. MED134, Bacteroidetes	Palovaara2014	Mediterranean Sea (Spain, 0.5m)	PR	Bicarbon uptake
5 <i>Vibrio</i> sp. AND4, Gammaproteobacteria	Gomez-Consarnau2010	Andaman Sea (Thailand, 2m)	PR	OD

Table 3: Earlier growth experiments with rhodopsin-bearing bacteria.

1.3 Aims of the thesis

The thesis aims to present a comprehensive, extensive study on the candidate genus '*Candidatus Methylopumilus* sp.', a limnic, planktonic member of Betaproteobacteria, characterized by a metabolism specialized on methylotrophy and the possession of rhodopsins.

Two main questions were addressed in this work:

1. Research question 1: What is the phylogenomic context of '*Ca. Methylopumilus* sp.' and what particular genomic features are present?

In the previous work of Salcher et al. [Salcher et al., 2015], one strain of each '*Ca. Methylopumilus* planktonicus' and '*Ca. Methylopumilus turicensis*' have been sequenced. One aim of the thesis is to sequence more strains of the genus, from both Lake Zurich and Czech samples. This serves for a more comprehensive characterization of '*Ca. Methylopumilus* sp.' in regard of genomic features e.g. genome size and coding density, identifying genes crucial for methylotrophy or other metabolic pathways, and phylogenomic relation to previously studied, closely related marine strains (OM43 clade).

2. Research question 2: How does '*Ca. Methylopumilus* planktonicus' react to light in varying carbon concentrations?

The bacterium possesses proteins for autotrophic metabolism in the form of light-dependent proton pumps (rhodopsins). It is of interest, how rhodopsins contribute to the cellular energy source under the influence of light and varying carbon concentrations. Previous work has shown changes in morphology, carbon uptake and pH in comparable organisms. These effects remain to be tested in '*Ca. Methylopumilus* planktonicus'.

2 Material and methods

2.1 Genomics

2.1.1 DNA isolation and whole genome sequencing

Prior to this work, a total of 157 strains were isolated from three different lakes by dilution to extinction [Salcher et al., 2015; Salcher and Simek, 2016]: Lake Zurich (CH, 5m and 100m depth), Řimov Reservoir (CZ, 0.5m and 30m depth) and Lake Medard (CZ, 0.5m depth). Bacteria were cultivated in Artificial Lake Water (ALW) [Zotina et al., 2003] containing only methanol and methyamine as carbon sources. 29 cultures were processed to pellets by high-speed centrifugation.

DNA was extracted from bacterial pellets by using the Qiagen MagAttract Kit (Qiagen, Hilden, DE) according to the manufacturer's protocol. PCR was performed using the general bacterial 16S rRNA primers GM3f and GM4r [Muyzer et al., 1995] and 1 μ l DNA extract. PCR products were purified using GenElute PCR Clean-Up (Sigma-Aldrich, St Louis, MO, USA) and sent to the Department of Systematic and Evolutionary Botany, University of Zurich for Sanger Sequencing. The obtained sequences were analyzed in Geneious (version 10.0.7, Biomatters, NZ ⁽¹⁾) and BLAST [Altschul et al., 1990], to confirm affiliation with *Methylophilaceae*.

Suitable candidates were then processed for whole genome sequencing using the KAPA Hyper Prep Kit (Kapa Biosystems, Wilmington, MA, USA) and the Illumina MiSeq platform (Illumina, San Diego, CA, USA), with the help of the Genomic Diversity Center (ETH Zurich, CH). The subsequent computational pipeline included trimming the reads by using `trimmmatic` [Lohse et al., 2012], contig assembly by `SPAdes` [Bankevich et al., 2012], mapping by `bowtie2` [Langmead et al., 2009] and checking for a circular genome by `check-circular` (Rohit Ghai, unpublished). Circular sequences were uploaded to RAST [Aziz et al., 2008] for genome annotation. Annotated sequences were imported to Geneious to perform orientation of the circular genome, to browse annotations and to check for low coverage regions by mapping the raw reads to the assembled contigs.

For non-circular sequences, multiple contigs that could not be joined, low coverage regions (<10-fold coverage), or regions with many N-spacers, the reads were mapped to the contigs for a second time with the Geneious reads mapper to eventually extend the contigs. A further solution was to design forward and reverse primers for the regions in question to perform a PCR. Resulting PCR-products were processed for Sanger Sequencing as previously described. This procedure was also applied to Single Nucleotide Polymorphisms (SNPs), or other peculiar regions.

⁽¹⁾Geneious, <http://www.geneious.com>

2.1.2 Phylogenomics

The resulting genomes were put in relation in phylogenetic distance trees on the basis of:

- (a) 16S rRNA sequence
- (b) conserved proteins
- (c) rhodopsin proteins
- (d) methanol-dehydrogenase proteins

To extract the genes for (a), (c) and (d), the built-in workflow creator in Geneious was used. These sequences were translated (with the exception of the 16S rRNA sequences) and aligned, and finally arranged in a maximum-likelihood tree by FastTree 2 [Price et al., 2010] with a bootstrap number of 100.

For the comparison of conserved proteins (b), the protein sequences of whole genomes were classified according to the database of Clusters of Orthologous Groups of proteins (COG) [Tatusov et al., 2000]. For each genome, the protein sequences that had the best accordance to a COG and were common to all analyzed genomes were then concatenated and aligned by using Kalign [Lassmann and Sonnhammer, 2005]. Again, a maximum-likelihood tree was built using FastTree 2 with a bootstrap number of 100.

Additionally, the Average Nucleotide Identity (ANI) [Goris et al., 2007] and the Percentage Of Conserved Proteins (POCP) [Qin et al., 2014] were examined. ANI is a method comparable to the conventional DNA-DNA-hybridization to determine relatedness between bacteria, but it is more robust, not prone to horizontal gene transfer and allows for a higher resolution on subspecies-level [Konstantinidis and Tiedje, 2005]. The threshold to distinguish two different species is an ANI score of <95%. POCP is based on the assumption that "two species belonging to the same genus would share at least half of their proteins" [Qin et al., 2014], and uses BLAST to determine conserved proteins between two genomes. If the POCP value between two genomes is above 50%, it is proposed that they belong to the same genus.

For a more comprehensive comparison, the following genomes of other methylotrophic bacteria were included in the phylogenomic analysis:

1. Marine OM43 clade:
 - a) KB13 [Huggett et al., 2012]
 - b) HTCC2181 [Giovannoni et al., 2008]
 - c) MBRS-H7 [Jimenez-Infante et al., 2016]
2. *Methylotenera versatilis* strain 79 (Chistoserdova, L.⁽²⁾), isolated from lake sediment.
3. *Methylphilus* sp. strain TWE2 [Xia et al., 2015], isolated from tap water.

⁽²⁾"Genomes of fifty methylotrophs isolated from Lake Washington", http://genome.jgi.doe.gov/GenoWashington_2/GenoWashington_2.info.html

2.2 Growth experiment

A growth assay was designed to test the response of the newly isolated strains to varying carbon concentrations and to light-exposed or dark conditions. For this purpose, the following two strains were used:

- (a) **Strain 1: MMS-VI-25** '*Ca. Methylophilus planktonicus*', encodes two types of rhodopsins (xantho- and proteorhodopsin).
- (b) **Strain 2: MMS-M-34** *Methylophilus* sp., similar to '*Ca. Methylophilus* sp.' concerning growth rate and density with no rhodopsins genes.

Cultures of each strain have been previously grown in artificial lake water (ALW), containing 1 mM methanol and 0.1 mM methylamine. Cell densities were determined using a CytoFLEX flow cytometer (Beckman Coulter, Brea, CA, USA) with 2% formaldehyde fixed cells stained with 2% SYBR green. Based on these concentrations, new cultures were prepared for the growth experiment with a starting cell density of approximately 25'000 bacteria/ml. The two strains were cultivated in triplicates in different set-ups varying in light conditions (light/dark) and concentration of C1 substrates in the culture medium (low/medium/high). A schematic illustration is shown in Figure 4. Light intensity for the light-exposed conditions was 40 μ Einstein (≈ 0.002 lux), equivalent to the light conditions of Lake Zurich in 5m depth. A white light lamp emitting light in the green spectrum was used as artificial light source. The samples for incubation in darkness

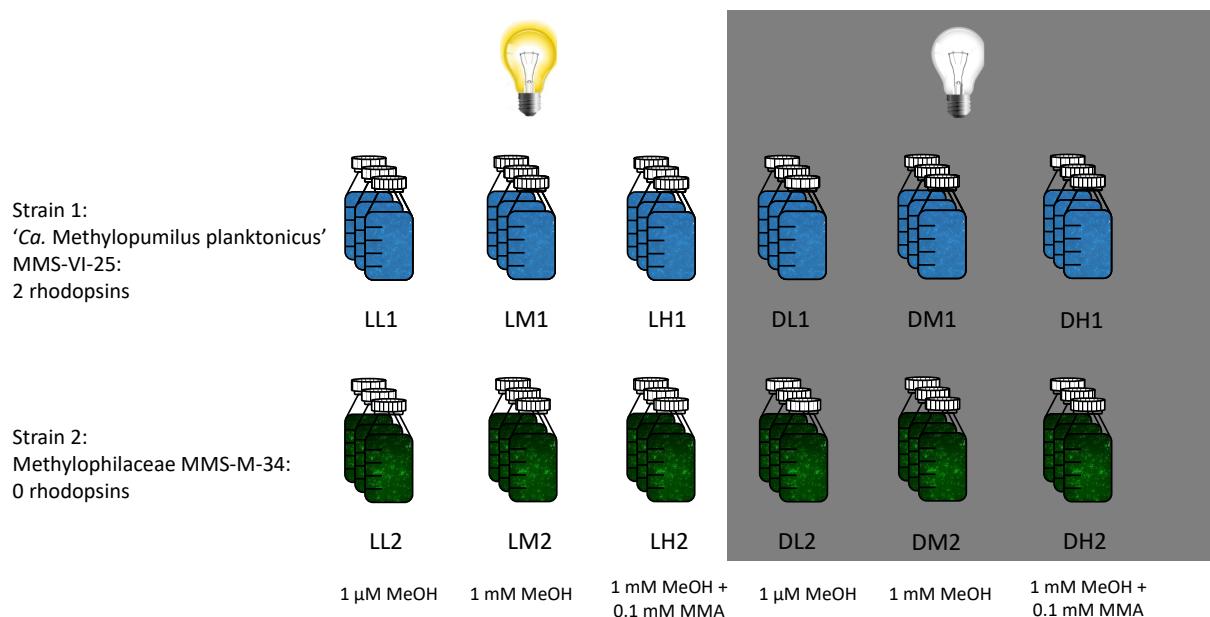


Figure 4: Experimental setup of the growth experiment. Nomenclature: First position [D/L] depicts the light condition (dark/light), second position [L/M/H] the carbon concentration (low/medium/high), and the third position [1/2] the strain identifier (1 = MMS-VI-25, 2 = MMS-M-34). MeOH = methanol, MMA = methylamine

were thoroughly wrapped in aluminum foil. The base medium used for all samples was ALW (350 ml each) without supplements; different C1 concentrations were achieved by adding 1 μM methanol (low), or 1 mM methanol (medium), or 1 mM methanol and 0.1 mM methylamine (high) respectively. Bacteria were cultivated in TPP tissue culture flasks at 20 °C without shaking. Sub-samples of 5 ml were taken twice a week and fixed with formaldehyde (2% final concentration) to determine cell density as previously described. After 47 days, the sampling interval was extended to once a week. After 61 days, sampling volume was reduced to 2 ml.

A table with the sampling days and their corresponding total day can be found in the Supplementary (Table S7).

50 ml were collected from all experimental conditions from the filtrates of the filtering step for the planned transcriptomics project (see Section 2.3 below), and measured with a pH-meter (inoLab, WTW series 720. Xylem, NY, USA).

Additionally, the genomes of '*Ca. Methylospumilus planktonicus*' MMS-VI-25 and *Methylophilaceae* MMS-M-34 were uploaded in the KEGG mapper⁽³⁾ [Kanehisa, 2000] to reconstruct metabolic pathways, such as the specialized methane metabolism.

2.2.1 Biomass and morphotype assessment

On days 12, 40 and 82 of the growth experiment (sampling days 3, 11 and 18), aliquots of formaldehyde-fixed samples (200 μl of strain1, 500 μl of strain2) were filtered onto polycarbonate filters (0.2 μm pore size, Millipore) and stained with 70 μl DAPI for 7 minutes. Filters were transferred onto diagnostic glass slides and inspected with a Zeiss AxioImager M.1 microscope (Carl Zeiss, Jena, DE). Image-based assessment of cell sizes for at least 100 cells per sample was done with LUCIA G (Laboratory Imaging Prague, Czech Republic) by using a macro developed by Posch et al. [Posch et al., 2009]. The resulting files containing the values regarding cell width, length, maximum and minimum feret were processed with a self-written Python-script⁽⁴⁾, which calculated cell volume (Eq. 1) and carbon content (Eq. 2) [Loferer-Krössbacher et al., 1998] as follows:

$$\text{Volume}[\mu\text{m}^3] = \text{width}^2 * \pi/4 * (\text{length} - (\text{width}/3)) \quad (1)$$

$$\text{Carbon}[fg/\text{cell}] = \text{Volume}^{0.86} * 218 \quad (2)$$

The calculated carbon content per cell was multiplied by the cell density (cells/ml) to derive the total carbon content per milliliter for each sample, a number referred to as "biomass".

⁽³⁾KEGG, <http://www.kegg.jp/kegg/mapper.html>

⁽⁴⁾Python Software Foundation, <https://www.python.org/>

Furthermore, it was assessed how much carbon the bacteria assimilated from the added sources at the beginning of the growth experiment. The initial carbon concentration was calculated by the combined molecular weight of all carbon-containing molecules (methanol and/or methylated amines and vitamins) in the ALW medium. Then, the mean biomass for each setup at days 12, 40 and 82 (sampling days 3, 11 and 18) were calculated and divided by the initial carbon concentration to assess the relative amount of carbon that was taken up from the medium by the bacteria at the three different time points.

The processed filters and resulting microscope images were also used to determine the various occurring morphotypes, again using the protocol provided by Posch et al. 2009, which categorizes the cells into five different morphotypes: Large rods, Vibrio-shaped, large cocci, small rods or small cocci.

The majority of the statistical and graphical analyses was done with R⁽⁵⁾ and RStudio respectively [RStudio Team, 2015].

2.3 Preparations for transcriptomics

To further assess the underlying metabolic mechanisms of '*Ca. Methylopumilus* sp.', it was planned to examine the transcriptional profile during the different growth stages. For this purpose, 20-50 ml of culture were filtered onto white polycarbonate filters (0.2 µm pore size, Millipore), which were immediately frozen in liquid nitrogen and then stored at -80°C. This was done at sampling day 2 (after 8 days in total) and sampling day 7 (after 26 days in total). Another filtering step will take place after the thesis' deadline.

RNA extraction was tested with the following kits:

1. MoBio PowerBiofilm RNA Isolation Kit (now part of the Qiagen group, Qiagen, Venlo, NL)
2. Lexogen SPLIT RNA Extraction Kit (Lexogen, Vienna, AT)
3. Qiagen RNeasy Plant Mini Kit
4. Qiagen RNeasy Micro Kit

Unfortunately, all test runs with test-cultures of MMS-VI-25 and MMS-M-34 were not successful. Inconveniently, it did not work either when the cell concentration of the test sample was higher than in the samples of the growth experiment. However, the Qiagen RNeasy Micro Kit worked well for filters of lake samples, for which 2 l of lake water were filtered onto 0.22 µm pore size polycarbonate filters, which equals approximately 100 billion cells in total per filter.

To extract enough RNA from the remaining cultures of the growth experiment, other approaches would be to centrifuge larger volumes (at least 100 ml) of sample mixed with a RNA-stabilizing

⁽⁵⁾R Foundation for Statistical Computing, <http://www.r-project.org>

reagent such as *RNAlater* (ThermoFisher Scientific, MA, USA), as described by Gifford et al. [Gifford et al., 2016]. For the previously processed filters, a kit that focuses on single cell quantities of RNA must be tested.

If a successful procedure to extract RNA can be established, a cDNA transcription and subsequent sequencing on an Illumina MiSeq platform would follow. The sequences would be analyzed in regard of type of transcript (e.g. by metabolic category) and number of transcripts per cell. An RNA depletion or subtractive hybridization to get rid of small, residual RNA strands or of the overrepresented 16S and 23S rRNA transcripts respectively, is unfavorable since the total RNA amount would assumably be very low, meaning that there is a risk of losing transcripts in these selection processes. A computational approach must be considered to filter out the relevant transcripts.

3 Results

3.1 Genomics of '*Ca. Methylospumilus* sp.'

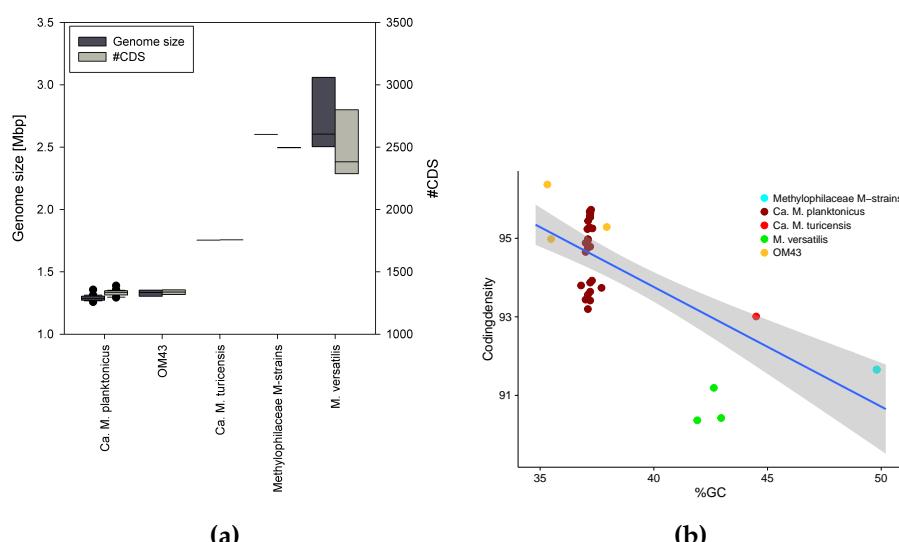
3.1.1 Sequencing and characterization of novel strains

In total, 29 previously unknown, limnic, methylotrophic bacteria were sequenced, which according to the 16S rRNA Sanger Sequencing belong to the family of Methylophilaceae (see Table 5). Analysis of the whole genome sequencing revealed that 26 of the 29 isolates can be classified as '*Ca. Methylospumilus* planktonicus'. They are characterized by a genome size of around 1.3 Mb, a low GC-content of approximately 37%, around 1300 coding DNA sequences (CDS) and a relatively high coding density (~95%). The three remaining sequenced isolates originate from Lake Medard in Czech Republic (labeled with the prefix "M", such as M-34) and can be classified as a novel limnic, methylotrophic bacterial species, as the differences are also apparent in several genomic characteristics; their genome is bigger (2.6 Mb), has a higher GC-content (49.8%), a higher number of CDS and a lower coding density (91.66%).

Table 5 also contains the first two described strains from Salcher et al. (2015), which are '*Ca. Methylospumilus* planktonicus', strain MMS-2-53 (original lineage identifier "LD28"), and '*Ca. Methylospumilus* turicensis', strain MMS-10A-171 (original lineage identifier "PRD01a001B"). Unfortunately, no additional strains of the latter species have been sequenced.

The genome of the closely related OM43 clade exhibits similar characteristics as the candidate genus, i.e. a small size of around 1.3 Mb, high coding density and low GC content (Figure 6). The M-strains share more similarities with *Methylotenera versatilis* in terms of genome size and number of CDS, but have a higher GC content. Of the compared genomes, the size aligns well with the number of CDS, in fact there is a high correlation of 0.998 (Figure S1). Furthermore, the GC content shows a negative correlation with the coding density, meaning the higher the coding density, the lower the GC content, a common pattern in small, streamlined genomes.

	Affiliation	Strain	Origin	Genome size (bp)	GC(%)	# CDS	Coding density(%)
1	<i>Ca. Methylopumilus</i> planktonicus	MMS-2-53	Lake Zurich	1'356'428	37.00	1389	94.88
2	<i>Ca. Methylopumilus</i> planktonicus	MMS-VA-83	Lake Zurich	1'275'265	37.11	1302	95.44
3	<i>Ca. Methylopumilus</i> planktonicus	MMS-VB103	Lake Zurich	1'304'290	37.00	1346	94.65
4	<i>Ca. Methylopumilus</i> planktonicus	MMS-VB-110	Lake Zurich	1'313'706	37.16	1349	95.27
5	<i>Ca. Methylopumilus</i> planktonicus	MMS-VB-116	Lake Zurich	1'311'334	37.08	1342	95.23
6	<i>Ca. Methylopumilus</i> planktonicus	MMS-VB-14	Lake Zurich	1'305'510	37.10	1359	94.98
7	<i>Ca. Methylopumilus</i> planktonicus	MMS-VB-60	Lake Zurich	1'272'502	37.17	1318	95.28
8	<i>Ca. Methylopumilus</i> planktonicus	MMS-VB-8	Lake Zurich	1'299'320	37.10	1341	94.77
9	<i>Ca. Methylopumilus</i> planktonicus	MMS-VB-91	Lake Zurich	1'302'879	37.20	1346	94.79
10	<i>Ca. Methylopumilus</i> planktonicus	MMS-VI-126	Lake Zurich	1'284'758	37.30	1326	95.25
11	<i>Ca. Methylopumilus</i> planktonicus	MMS-VI-155	Lake Zurich	1'302'967	37.00	1351	93.44
12	<i>Ca. Methylopumilus</i> planktonicus	MMS-VI-180	Lake Zurich	1'312'289	37.09	1348	94.93
13	<i>Ca. Methylopumilus</i> planktonicus	MMS-VI-25	Lake Zurich	1'295'732	37.70	1338	93.74
14	<i>Ca. Methylopumilus</i> planktonicus	MMS-VI-255	Lake Zurich	1'303'963	37.20	1331	93.42
15	<i>Ca. Methylopumilus</i> planktonicus	MMS-VI-257	Lake Zurich	1'312'437	37.10	1349	94.97
16	<i>Ca. Methylopumilus</i> planktonicus	MMS-VI-38	Lake Zurich	1'272'877	37.10	1315	93.56
17	<i>Ca. Methylopumilus</i> planktonicus	MMS-RVI-8	Řimov Res.	1'269'643	37.21	1299	95.65
18	<i>Ca. Methylopumilus</i> planktonicus	MMS-RIV-25	Řimov Res.	1'269'747	37.20	1336	95.54
19	<i>Ca. Methylopumilus</i> planktonicus	MMS-RI-1	Řimov Res.	1'287'884	36.80	1319	93.80
20	<i>Ca. Methylopumilus</i> planktonicus	MMS-RI-41	Řimov Res.	1'305'374	37.10	1341	93.20
21	<i>Ca. Methylopumilus</i> planktonicus	MMS-RII-56	Řimov Res.	1'280'948	37.20	1329	93.64
22	<i>Ca. Methylopumilus</i> planktonicus	MMS-RIV-17	Řimov Res.	1'258'664	37.18	1301	95.68
23	<i>Ca. Methylopumilus</i> planktonicus	MMS-RIV-28	Řimov Res.	1'262'052	37.24	1293	95.73
24	<i>Ca. Methylopumilus</i> planktonicus	MMS-RIV-30	Řimov Res.	1'268'464	37.20	1296	95.59
25	<i>Ca. Methylopumilus</i> planktonicus	MMS-RIV-70	Řimov Res.	1'276'045	37.28	1349	93.92
26	<i>Ca. Methylopumilus</i> planktonicus	MMS-RVI-13	Řimov Res.	1'277'839	37.20	1328	93.87
27	<i>Ca. Methylopumilus</i> planktonicus	MMS-RVI-9	Řimov Res.	1'269'873	37.20	1297	95.54
28	<i>Ca. Methylopumilus</i> turicensis	MMS-10A-171	Lake Zurich	1'754'988	44.50	1757	93.01
29	Methylophilaceae	MMS-M-34	Lake Medard	2'602'980	49.80	2495	91.66
30	Methylophilaceae	MMS-M-37	Lake Medard	2'602'980	49.80	2498	91.66
31	Methylophilaceae	MMS-M-51	Lake Medard	2'602'980	49.80	2497	91.66

Table 5: Overview of the sequenced strains and their genomic properties.**Figure 6:** Visualized genomic properties of the sequenced strains compared to reference genomes. (a) Genome sizes and numbers of CDS (b) Coding density values plotted against GC-content.

3.1.2 Phylogenetic distances

A maximum likelihood tree built from the 16S rRNA sequences confirms the genomic classification of the new strains to '*Ca. Methylospumilus planktonicus*', distinctly separated from the M-strains (see Figure 7). This is in concordance with the Average Nucleotide Identity (ANI) in Figure 9. As mentioned before, the cut-off value to distinguish between two species is <95%. According to this, the strains depicted as '*Ca. Methylospumilus planktonicus*' are technically three distinct species, even though the 16S rRNA comparison suggests otherwise. Notably, the biggest cluster seen in the ANI-matrix consists of strains that derive from both Lake Zurich and the Řimov reservoir, while other two have their origin in either Lake Zurich (prefix "V") or the Řimov reservoir (prefix "R").

The classification of the '*Ca. Methylospumilus planktonicus*' and '*Ca. Methylospumilus turicensis*' into the same genus is confirmed by the Percentage Of Conserved Proteins (POCP), as shown in Figure 10. Remarkably, the strains of the marine OM43 clade share more than 50% of conserved proteins with '*Ca. Methylospumilus planktonicus*' and '*Ca. Methylospumilus turicensis*', suggesting that they belong to the same genus.

The M-strains however are clustered with *Methylophilus sp.* TWE2 in all described analyses, proving an affiliation with the *Methylophilus* genus within the family of Methylophilaceae. However, the results concerning the three M-strains were all identical, and thus have the exact same genome.

In a maximum likelihood tree built on the base of 330 common proteins according to COG classifications (Figure 8), the same patterns are visible. The variance within '*Ca. Methylospumilus planktonicus*' is higher than in the maximum likelihood tree based on the 16S rRNA sequences, again suggesting that the candidate species consists of several groups instead of only one. Based on the common protein by COG analysis, the marine OM43 clade is related more closely to '*Ca. Methylospumilus planktonicus*' than to the latter's sister species '*Ca. Methylospumilus turicensis*'.

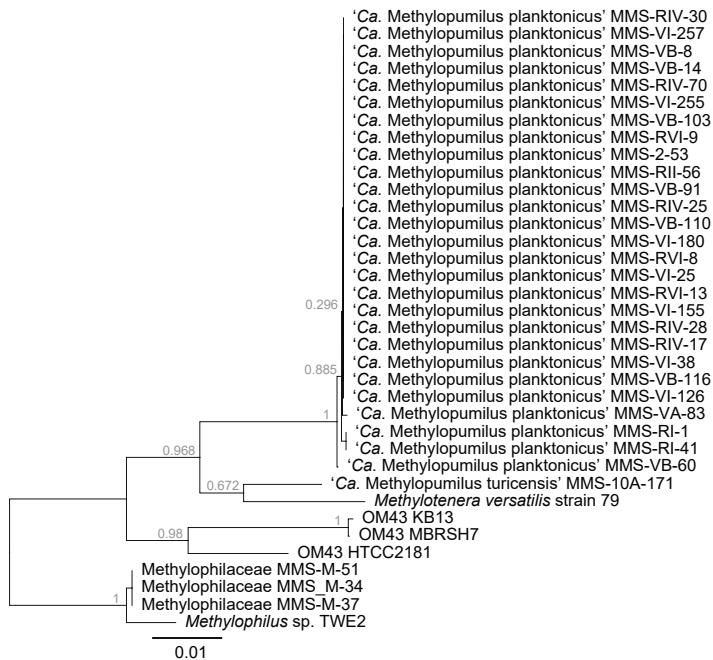


Figure 7: Bootstrapped maximum likelihood tree of the 16S rRNA sequences of the 28 'Ca. Methylophilus' sp.' strains, the three M-strains (Methylophilaceae sp.), three marine OM43 strains and two other methylotrophic, aquatic specimen (*Methylophilus* sp. TWE2, *Methylotenera versatilis* strain 79). The scale bar at the bottom represents 1% estimated sequence divergence. Bootstrap values are shown in gray.

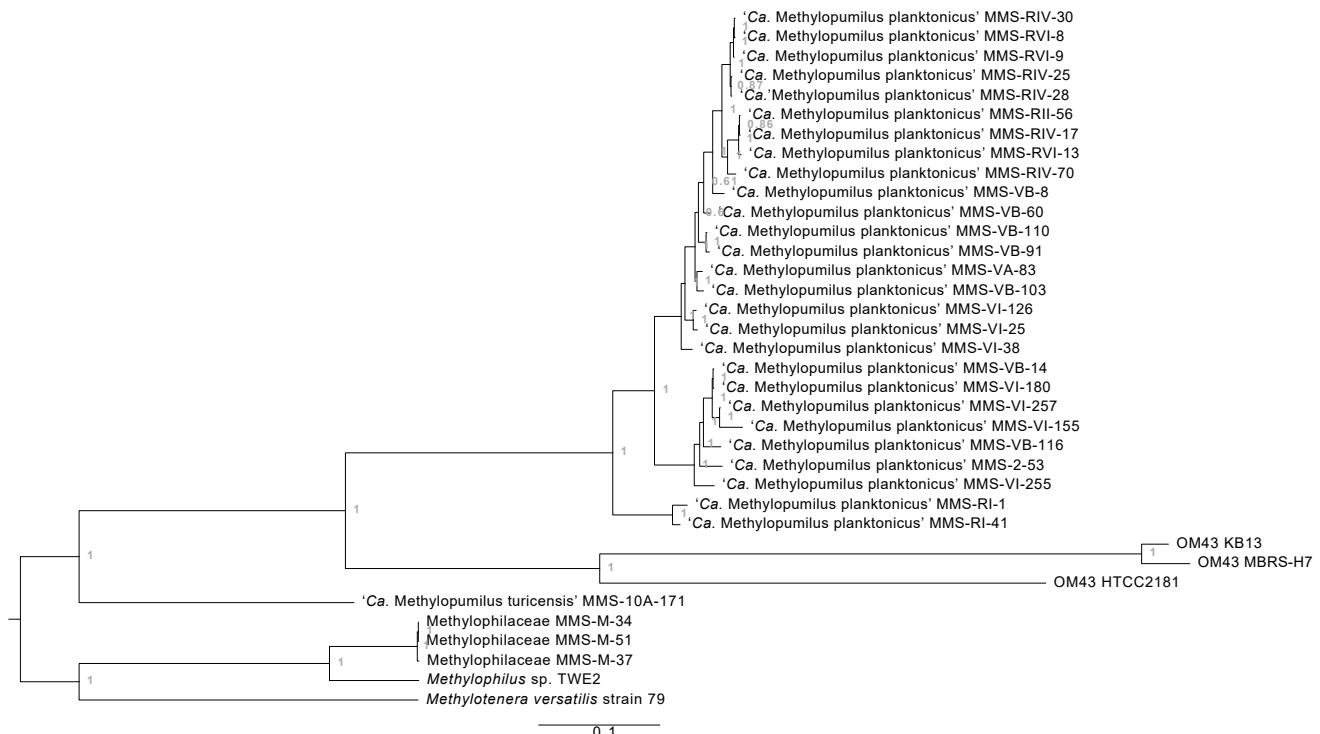


Figure 8: Maximum-likelihood tree of 330 common conserved proteins in the set of compared genomes, based on COG-classifications. Here the scale bar represents 10% estimated sequence divergence.

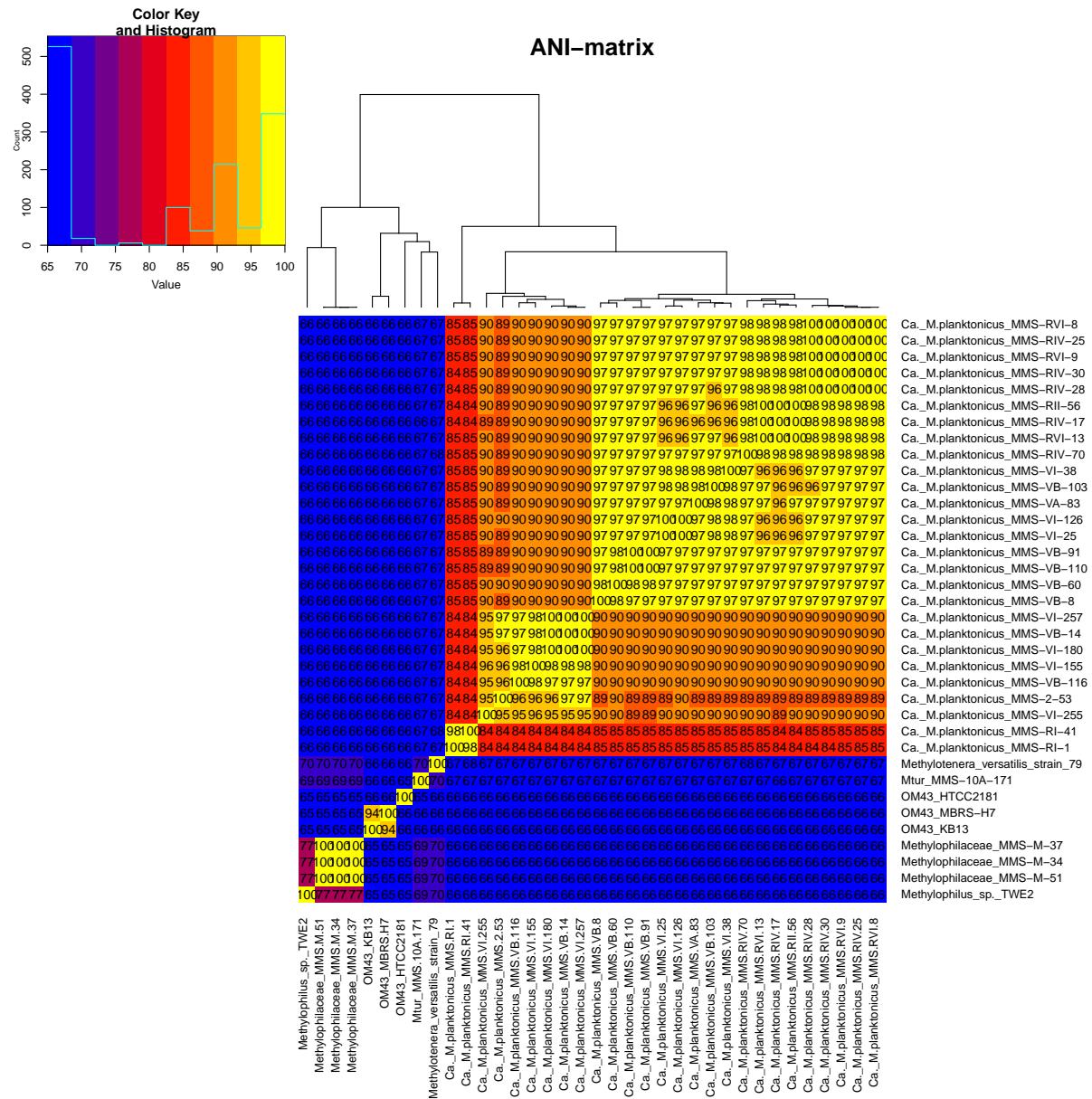


Figure 9: ANI-matrix (Average Nucleotide Identity).

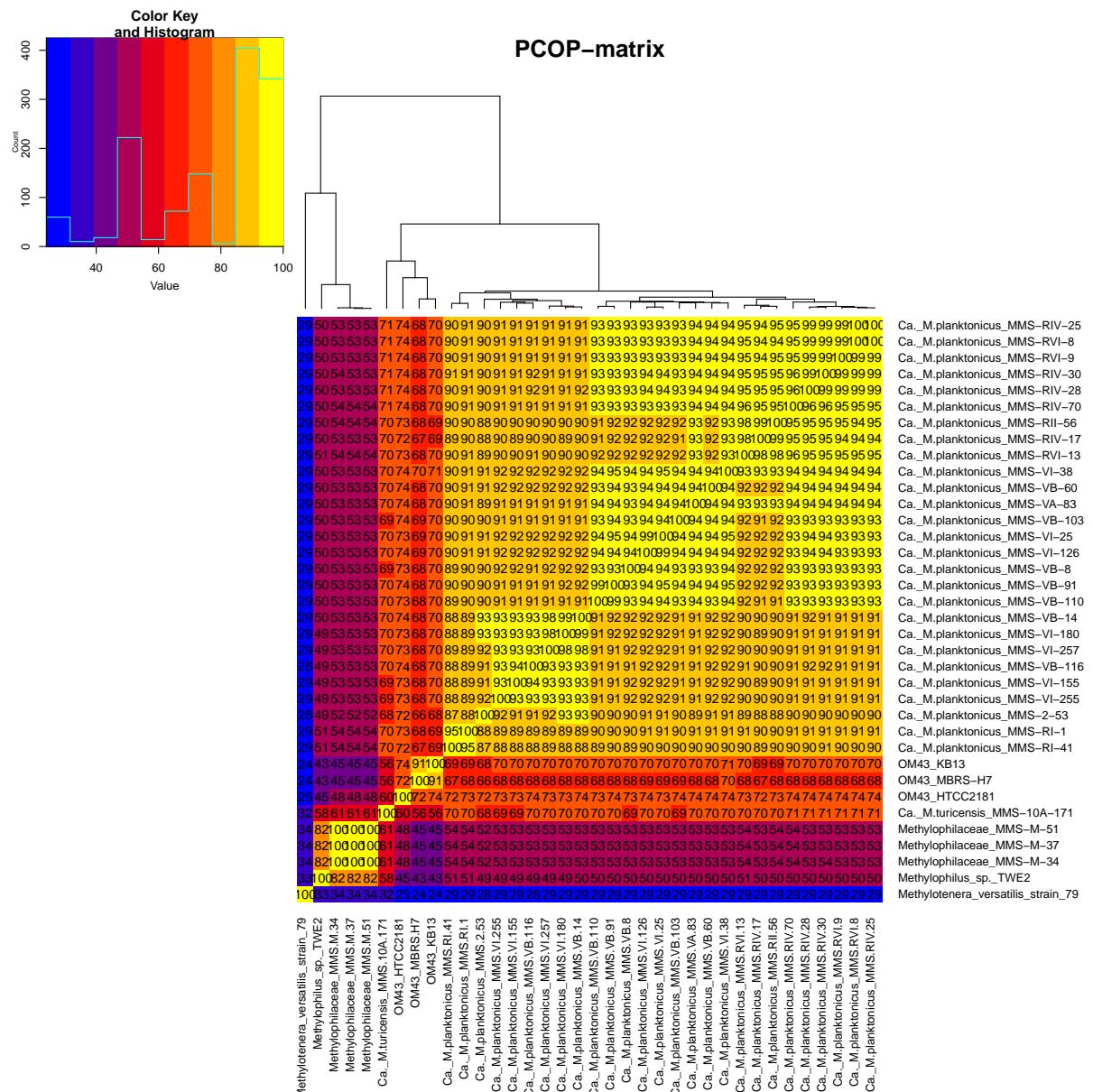


Figure 10: PCOP-matrix (Percentage Of Conserved Proteins).

3.1.3 Metabolic features

Genome analysis of '*Ca. Methylopumilus planktonicus*' MMS-VI-25 and *Methylophilaceae* MMS-M-34 revealed that both possess several specialized enzymes in order to metabolize C1-substrates. These enzymes and their role in methane metabolism are shown in Figure 12. A simplified illustration of the general methane metabolism is shown in Figure 11. In total, 21 enzymes were identified to be relevant for C1 utilization in the MMS-M-34 strain, of which 12 enzymes were also identified in MMS-VI-25. Apparently, the latter does not carry an enzyme for C1 metabolism which is not found in MMS-M-34 too.

The candidate genus and the M-strain are strict non-methanotrophs, since they do not carry any methane monooxygenases to convert methane to methanol. However, they do carry methanol dehydrogenases (MDHs) to convert methanol to formaldehyde, which then can be further metabolized to enter the RuMP-cycle (ribulose monophosphate-cycle) [Chistoserdova, 2011; Salcher et al., 2015]. Formaldehyde can also be converted to formate, which then is transformed to CO₂. CO₂ that is not respired and released, can be further used for C1 assimilation via methylene-tetrahydrofolate or the serine-cycle. Another option to utilize CO₂ in scope of C1 assimilation is the CBB cycle (Calvin-Benson-Bassham cycle). Mostly, non-methylotrophic autotrophs rely on the CBB pathway, but also methanotrophs such as *Verrucomicrobia* [Chistoserdova, 2011]. However, there are no genes present in '*Ca. Methylopumilus planktonicus*' MMS-VI-25 nor in *Methylophilaceae* MMS-M-34 that are related to the CBB cycle.

However, there are some apparent gaps in the reconstructed metabolism of '*Ca. Methylopumilus planktonicus*' MMS-VI-25 and *Methylophilaceae* MMS-M-34 (Figure 12). For example, there are no enzymes detected for the tetrahydromethanopterin-metabolism (THMPT) in MMS-VI-25, and also the genes needed for metabolizing methylamines were absent.

There are various kinds of MDHs found in methylotrophs, such as *mxaF*, *mdh2* or *xoxF*. As it is illustrated in Figure 13, '*Ca. Methylopumilus planktonicus*' and '*Ca. Methylopumilus turicensis*' both encode the *xoxF* variant, which is also present in the marine OM43 strains. The *Methylophilaceae* M-strains on the other hand, do not carry *xoxF*, but four other kinds of potential methanol dehydrogenases, each of them closely related to the four MDHs of *Methylophilius* sp. TWE2. A query with BLAST revealed that one is most likely *mxaF*, while the others did not have a distinct match to other previously identified MDHs.

A summary of the enzymes present in MMS-VI-25 and/or MMS-M-34 shown in Figure 12 is listed in the Supplementary (Table S2).

As it was elaborated in the introduction, there exist several variants of rhodopsins. Figure 14 illustrates that all of the newly isolated strains of '*Ca. Methylopumilus planktonicus*' carry both proteo- and xanthorhodopsins. In '*Ca. Methylopumilus turicensis*' only proteorhodopsins are present (upper cluster in the tree), and the marine OM43 strains only carry xanthorhodopsins (bottom cluster). The divergence between the proteorhodopsins in '*Ca. Methylopumilus planktonicus*' and '*Ca. Methylopumilus turicensis*' range between 72-73%. Moreover, the xanthorhodopsins of '*Ca. Methylopumilus planktonicus*' and the OM43 clade only share an identity of 57-63%.

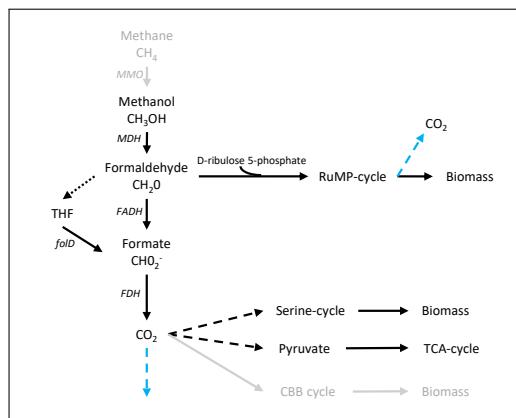


Figure 11: Simplified diagram of methane metabolism. Abbreviations: THF = tetrahydrofolate, MMO = methane monooxygenase, MDH = methanol dehydrogenase, FADH = formaldehyde dehydrogenase, FDH = formate dehydrogenase, fold = bifunctional enzyme (methylenetetrahydrofolate dehydrogenase; methenyltetrahydrofolate cyclohydrolase). Dotted line = spontaneous reaction, dashed line = abbreviated pathway, blue dashed line = respiration. Gray text = modules associated with methane metabolism, but not present in MMS-VI-25 nor in MMS-M-34.

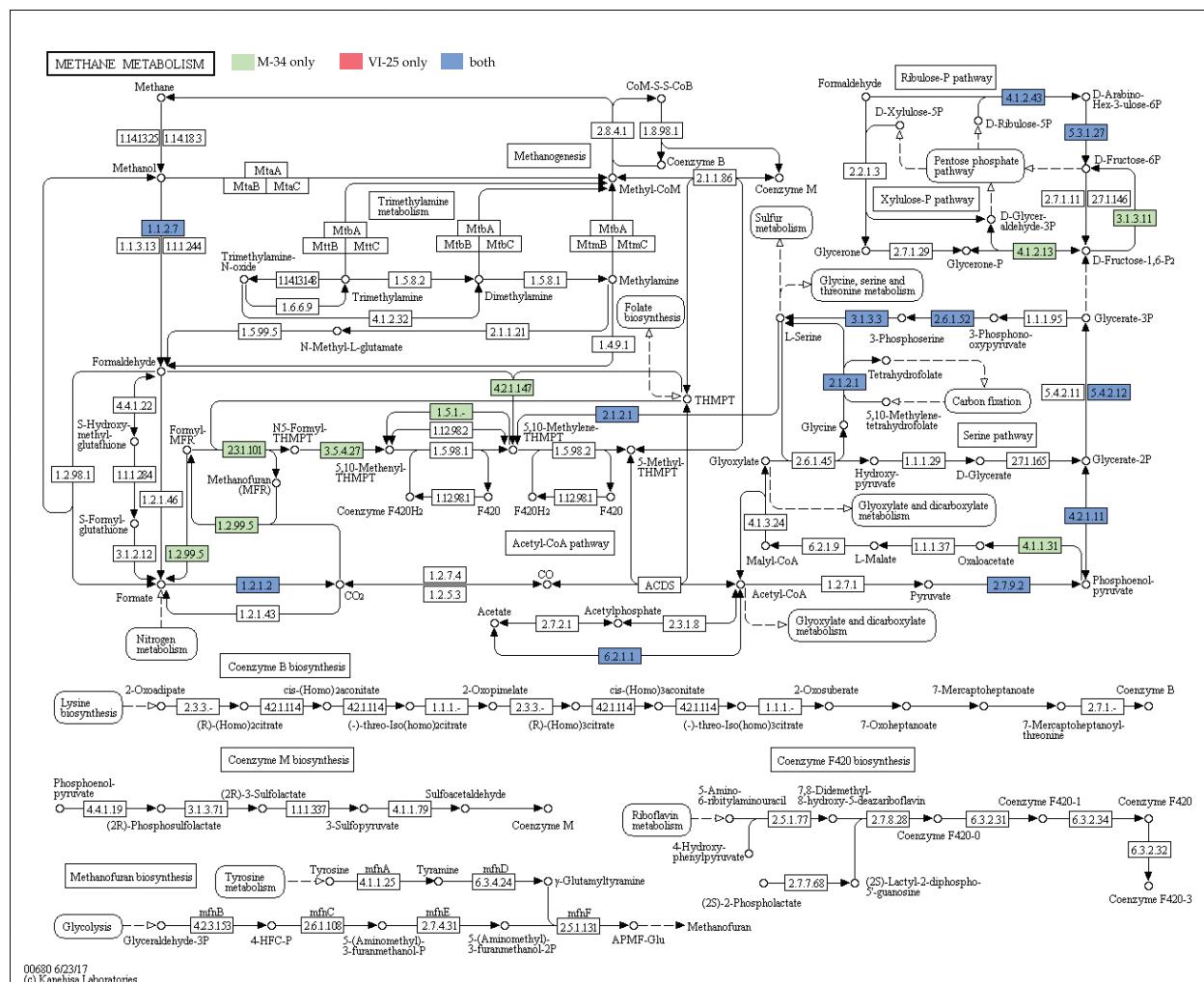


Figure 12: Reconstructed pathway of methane metabolism by KEGG. The boxes are labeled with the KEGG E.C. number.

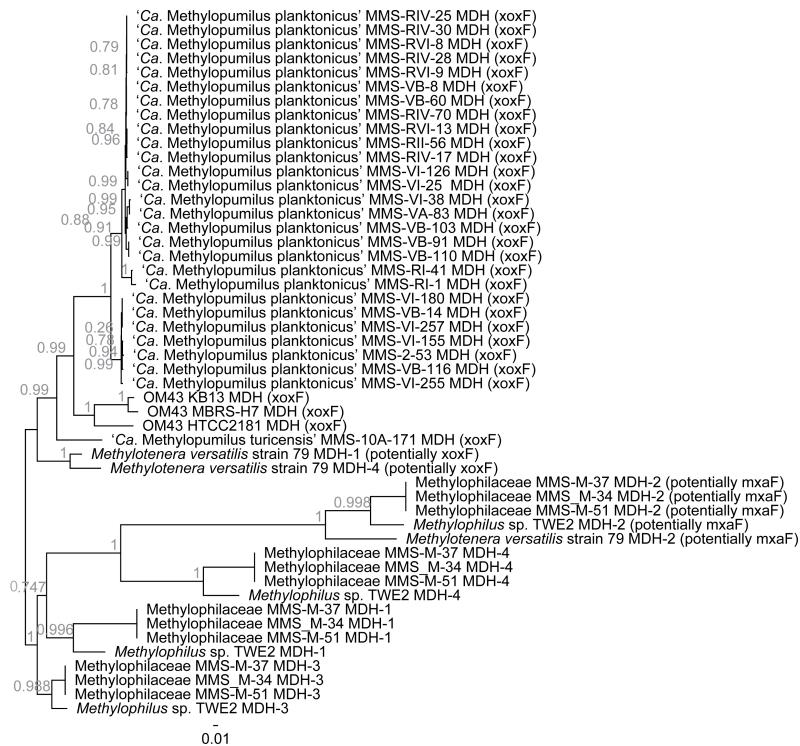


Figure 13: Maximum-likelihood tree of the methanol dehydrogenases present in the newly isolated strains and reference genomes. For the four groups below, an exact identification of the MDH type was not feasible, therefore the proteins are numbered by their order in the genome (MDH1-MDH4).

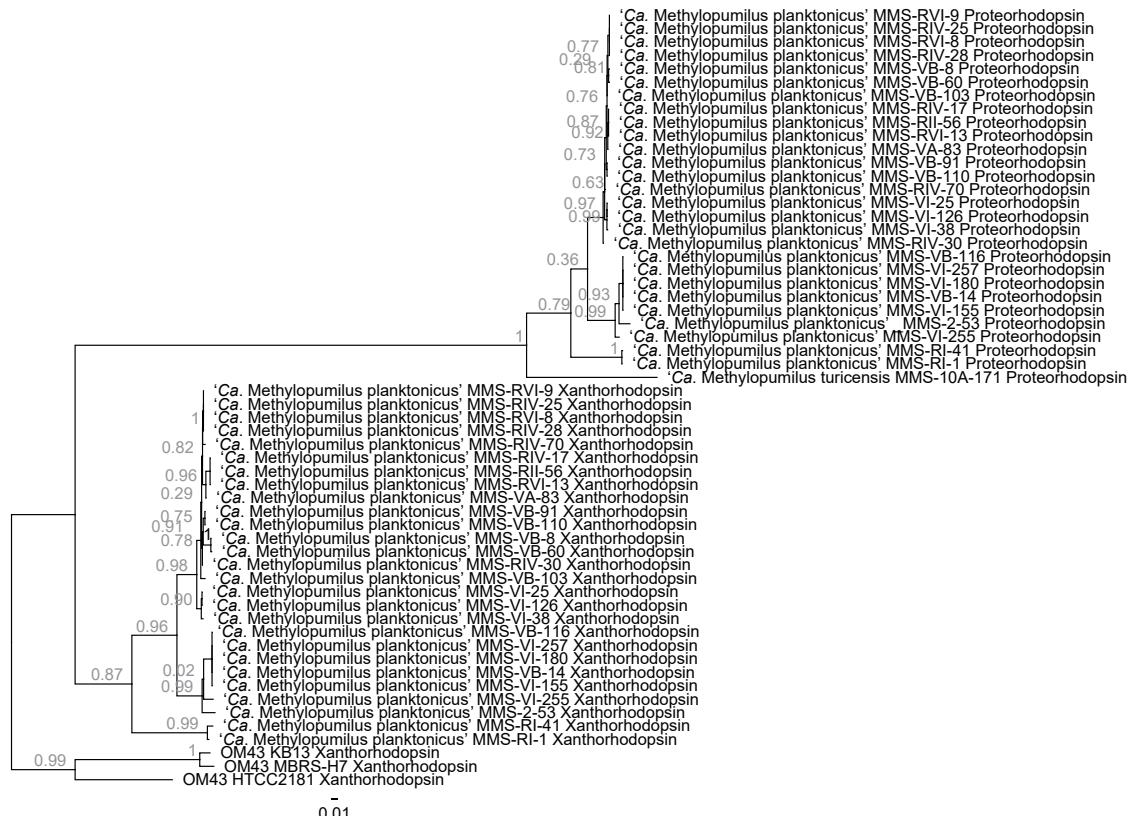


Figure 14: Maximum-likelihood tree of the rhodopsins present in '*Ca. Methylospumilus planktonicus*', '*Ca. Methylospumilus turicensis*' and in three members of the marine OM43 clade.

3.2 Growth experiment

At the time of thesis submission, the growth experiment had been running for 124 days in total and 23 samplings had been conducted. Visual assessment revealed that cultures grew without contamination, but lack visible turbidity suggested low density of bacteria. Microscopic analysis revealed that cells of '*Ca. Methylophilus planktonicus*' MMS-VI-25 had a curved rod shape (similar to *Vibrio*) and were small in size ($\sim 1 \mu\text{m}$), as is shown in Figure 16. *Methylophilaceae* MMS-M-34 on the other hand appeared bigger ($\sim 3 \mu\text{m}$) and in a straight rod shape (Figure 17).

During the growth experiment, the cultures also underwent changes in pH. Figure 15 compiles data on pH values measured at three time points. The pH decreased from a mean value of 9.055 on the first day to a mean value of 7.71 after 26 days in all samples. Notably, the setup in light-exposed, low and also medium carbon concentration with strain 1 (LL1 and LM1) showed the lowest pH levels after 26 days. There was no significant difference between the light-exposed cultures of strain 1 and the dark-grown (t-test p-value = 0.4853), nor between the cultures of strain 1 and strain 2.

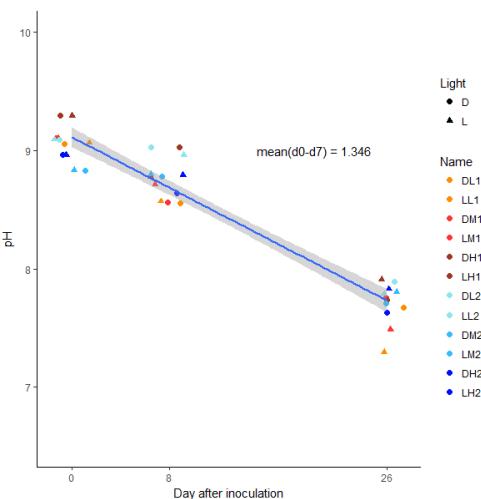


Figure 15: pH at the beginning of the experiment, after 8 days and after 26 days for all setups.

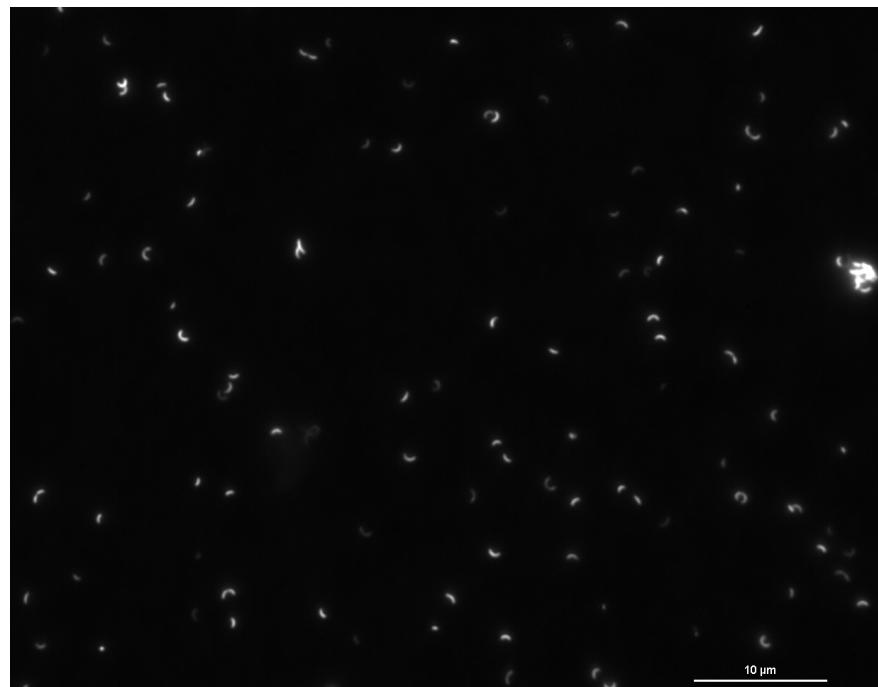


Figure 16: Epifluorescent microscope image of '*Ca. Methylophilus planktonicus*' MMS-VI-25 (DAPI-stained) in the early stationary phase.

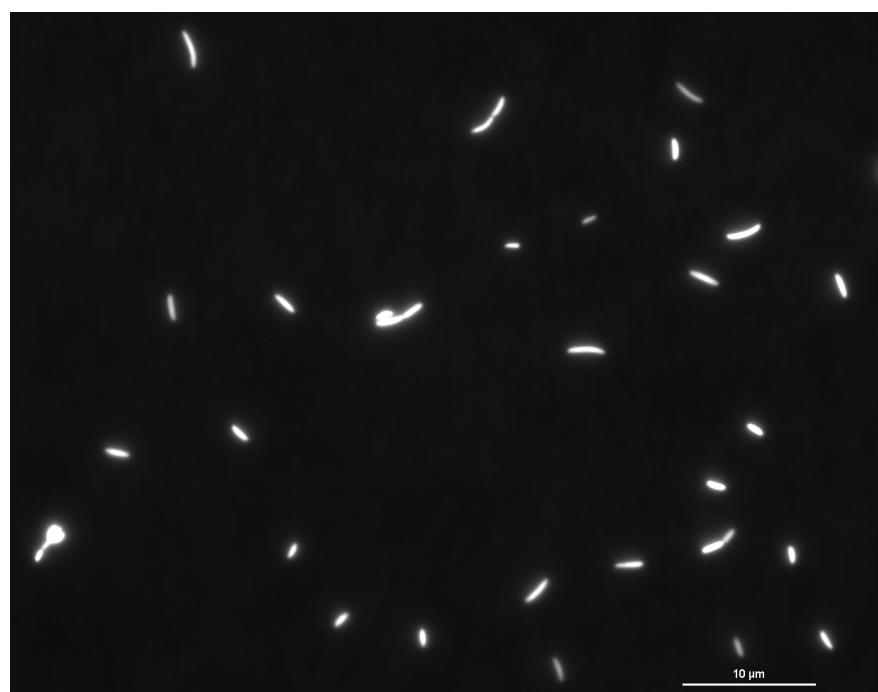


Figure 17: Epifluorescent microscope image of *Methylophilaceae* MMS-M-34 (DAPI-stained) in the early stationary phase.

3.2.1 Cell densities

As mentioned in Section 2.2, the initial bacterial concentration aimed at 25'000 cells per milliliter. Assessment by flow cytometry exhibited values around 27'466 (± 2119) cells/ml for strain 1 (MMS-VI-25), but only 5802 (± 1120) cells/ml for strain 2 (MMS-M-34) (Figure 18). Strain 1 started to grow exponentially at first and stabilized at around 32.8×10^6 cells/ml in low carbon concentrations and 19.1×10^6 cells/ml in medium carbon concentration after 19 days. In high carbon concentrations, the population stabilized after 43 days at around 47.4×10^6 cells/ml.

Strain 2 exhibited a short lag-phase of approximately five days at the beginning and reached the stationary phase after 19 days. The cell densities in strain 2 stabilized around 3.8×10^6 cells/ml in low carbon concentrations, 6.3×10^6 cells/ml in medium carbon concentrations and 4.6×10^6 cells/ml in high carbon concentrations. Differences between light and dark conditions were negligible at the early stationary phase for both strains.

With the exception of the cultures of strain 1 grown in low carbon concentrations (both in light and darkness), and of strain 2 in low carbon concentrations and in light, all cultures further increased their cell density. After 110 days (late stationary phase), the mean cell density of strain 1 samples was 38.24×10^6 cells/ml, and 10.58×10^6 cells/ml in strain 2.

Performing a Welch Two Sample t-test showed a significant difference of cell density between cultures grown in light or in the dark in the medium carbon concentrated conditions of strain 1 (sampling day 22, $p=0.075$), and in the culture of strain 2 in low carbon concentrations (sampling day 22, $p=0.0003$). However, in cultures of strain 1, the light-exposed samples generally exhibited a higher cell density than their dark-grown counterparts. After 110 days, the mean cell density of light-exposed strain 1 cultures was 1.28-fold higher than in the dark conditions (42.93×10^6 cells/ml vs. 33.56×10^6 cells/ml). On the contrary, strain 2 had overall lower cell densities in light-grown cultures; the mean cell density was 0.27-fold lower in cultures exposed to light than in dark conditions (8.64×10^6 cells/ml vs. 12.51×10^6 cells/ml).

A table containing all p-values resulting from Welch Two Sample t-tests on the difference between cell densities in light and dark conditions can be found in the Supplementary (Table S3). The cell densities and their calculated mean of sampling day 22 are listed in Table S4.

3.2.2 Differences in biomass

The comparison of the biomass of both strains for three sampling days ($d_3 = 12$ total days, $d_{11} = 40$ total days, $d_{18} = 82$ total days) is visualized in Figure 19.

In strain 1 (top row), there is an increase of biomass over time and with increased carbon concentration in the medium. Notably, the samples kept in light-exposed conditions were generally higher in all carbon concentrations at sampling day 11 and sampling day 18. There was a significant difference between light and dark conditions at sampling day 11 for the cultures in medium

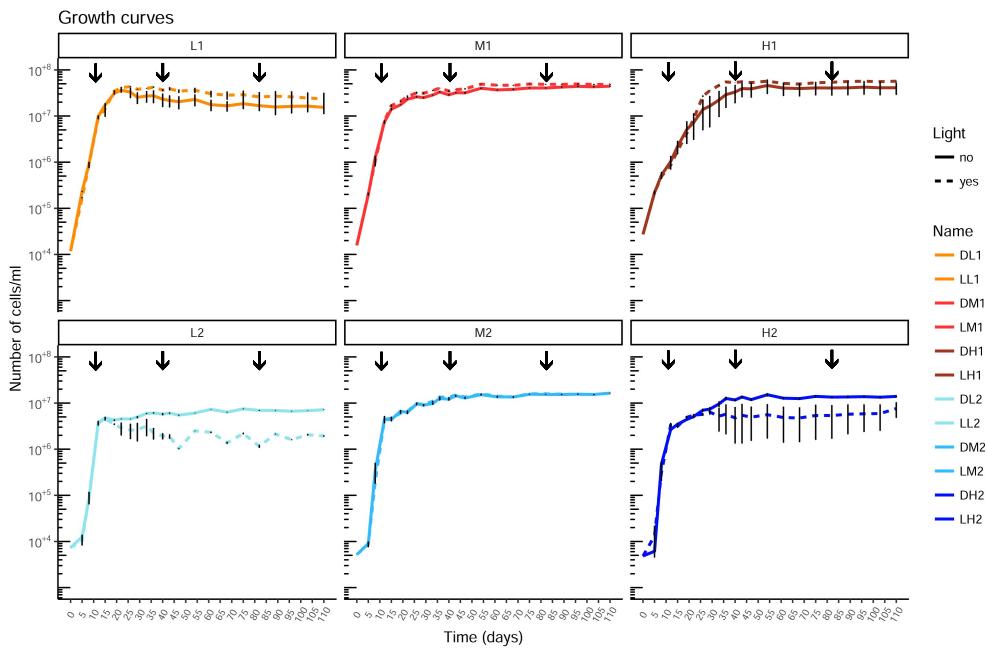


Figure 18: Cell concentrations of the experimental conditions varying in light-exposure (dark/light), carbon concentrations (low/medium/high) and bacterial strain (strain 1/strain 2). Strain 1 is shown in the top row in red colors, strain 2 in the bottom row in blue colors. The arrows mark the days from which a biomass and morphotype assessment was conducted.

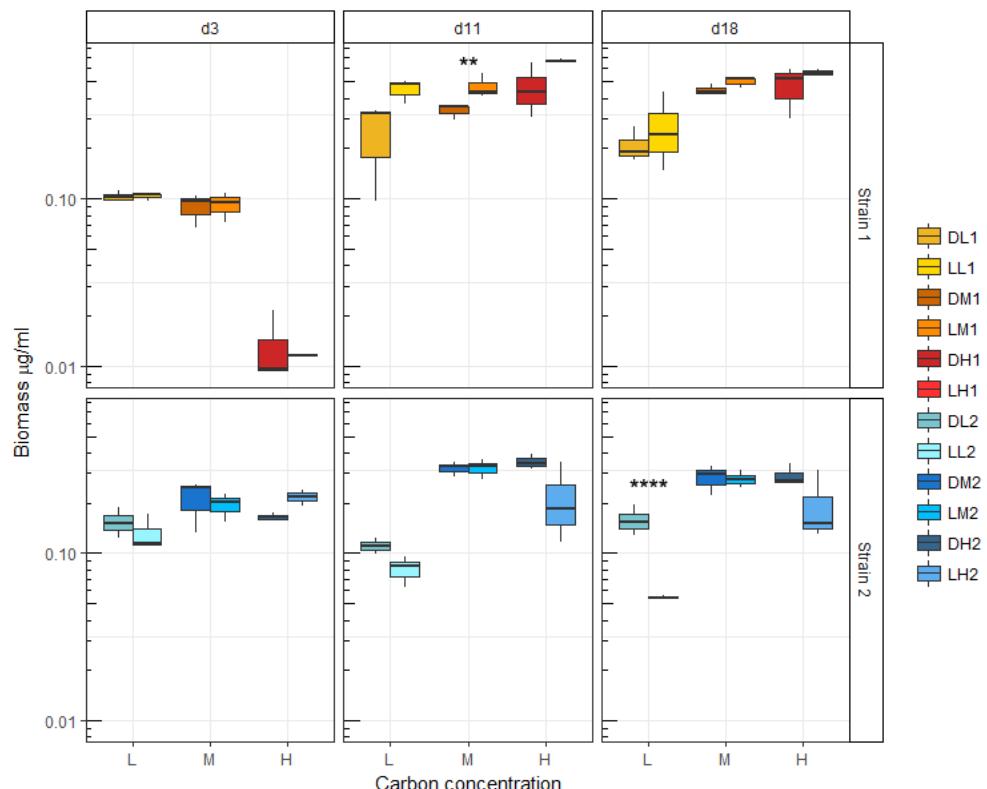


Figure 19: Biomass[µg C/ml] per growth condition sample. Significance codes: $<0.001 = \text{****}$, $<0.01 = \text{***}$, $<0.05 = \text{**}$, $<0.1 = *$

concentrated carbon (t -test, $p=0.035$).

Strain 2 shows no clear trend over time, nor are there apparent differences between the carbon concentrations except for the low concentrated carbon setups. However, it appears that there is slightly less biomass present in light-exposed conditions. The most striking difference is seen at sampling day 18 in the low carbon concentration, where the biomass is significantly higher for the cultures kept in darkness (t -test, $p=0.0074$).

An overview of the resulting p -values of the performed Welch Two Sample t -tests on the differences in biomass can be found in the Supplementary, Table S8.

3.2.3 Carbon uptake

After 40 and 82 days, the ratio between the initial carbon amount and the carbon present as biomass was between 2.77% and 5.0% for strain 1, and between 1.49% and 2.65% for strain 2 respectively in medium and high carbon concentrated cultures (Figure 20). Unfortunately, in supposedly low concentrated carbon conditions, the carbon incorporation values were above 100%. Possible explanations are elaborated in Section 4 (Discussion).

The cultures of strain 1 which were exposed to light exhibited a slightly higher carbon uptake than their dark-grown counterparts (medium concentrated carbon: 4.14% in light vs. 3.69% in the dark, high concentrated carbon: 4.26% in light vs. 3.55% in the dark). Cultures of strain 2 showed a noticeably lower carbon uptake in light-grown, high carbon concentrated conditions, and generally had a lower carbon incorporation than strain 1.

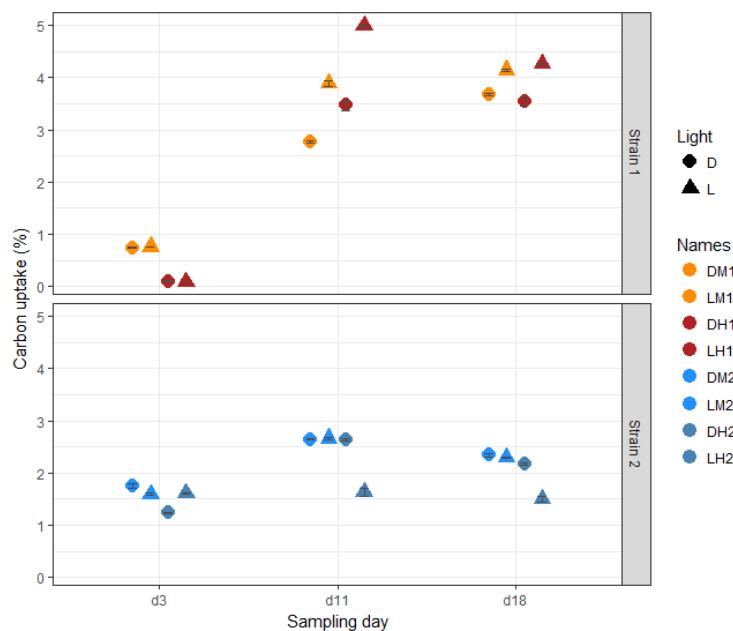


Figure 20: Relative carbon uptake for three sampling days. The values for the samples in the low concentrated carbon medium are omitted.

3.2.4 Differences in morphotypes

As mentioned before, '*Ca. Methylophilaceae MMS-M-34*' appear as small and curved ("Vibrio-shaped"), while *Methylophilaceae MMS-M-34* are bigger and rod-shaped. These morphotypes were dominant at sampling day 3, as seen in Figure 21. In strain 1, around 70% of the cells had a Vibrio-shaped form in the beginning, which decreased over time to around 23%. Meanwhile, the percentage of small cocci only made up approximately 9% in the exponential phase, but at the late stationary phase their frequency increased to 50% of all morphotypes. The differences between dark and light conditions are marginal; there is a tendency of a higher percentage of Vibrio-shaped morphotypes in the dark-grown cultures than the ones exposed to light (see also Supplementary Figure S11 for stacked bar plots of all carbon concentrations). In light-grown cultures, there was a slightly higher frequency of small cocci or small rods.

In strain 2, large rods were dominant at the beginning of the growth experiment (90%), but towards the end of the growth assay they only represented 24-30% of the morphotypes. In contrast, the smaller morphotypes like small cocci increased in frequency from 0-1% to around 37%.

Stacked bar plots of all carbon concentrations can be found in the Supplementary (Figure S11 and S12).

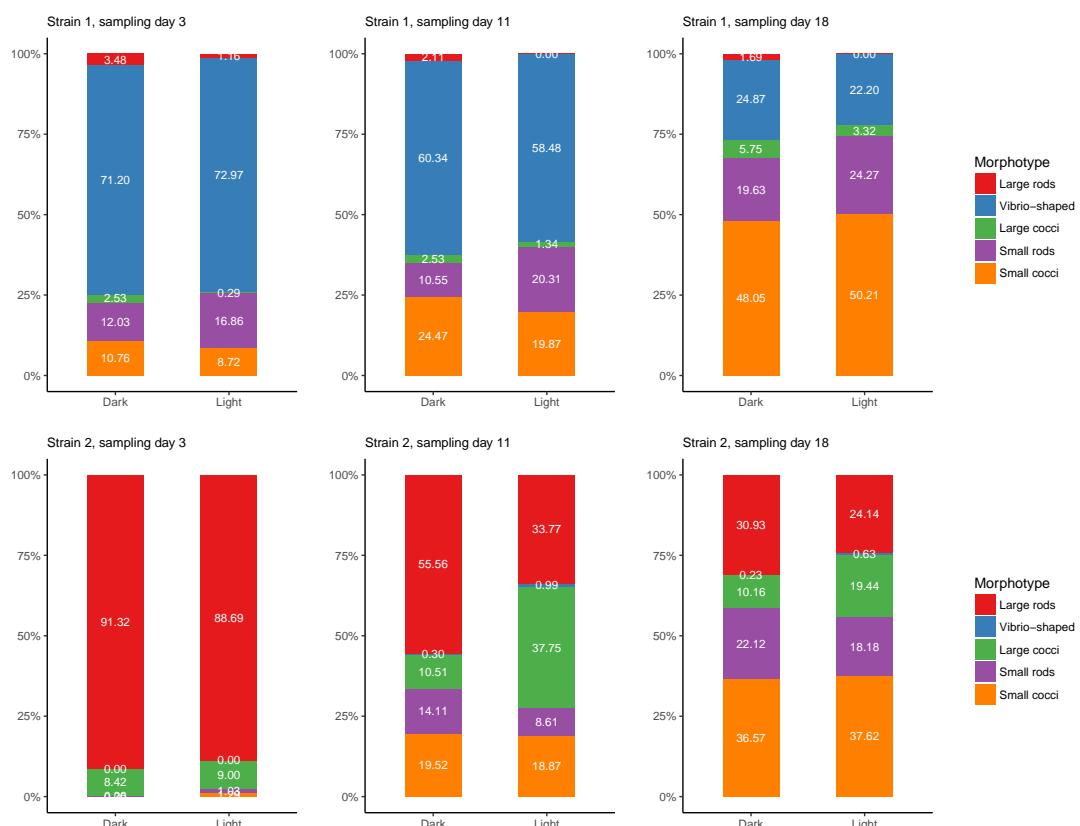


Figure 21: Distribution of morphotypes over time in light and dark conditions

4 Discussion

4.1 Evolution of '*Ca. Methylospumilus* sp.'

In the following sections, the findings of this thesis will be put in context in terms of evolutionary processes, in order to answer the first research question proposed in Section 1.3 (Aims of the thesis);

What is the phylogenomic context of '*Ca. Methylospumilus* sp.' and what particular genomic features are present?

4.1.1 Hypothesized habitat transitions

As '*Ca. Methylospumilus planktonicus*', '*Ca. Methylospumilus turicensis*', and the members of the marine OM43 clade are similar in regard of many genomic characteristics, and share a POCP score above 50%, it is assumed that they belong to the same genus of '*Ca. Methylospumilus* sp.'. A suggested name for the OM43 lineage is '*Ca. Methylospumilus marinus*'.

Given how closely these strains are related, it is hypothesized that there must have been a transition from either limnic systems to the marine environment, or vice versa. The first option seems more likely, since a) the OM43 clade occurs mostly on coastal regions, where the allochthonous input is higher than in the open water [Walsh et al., 2013], and b) the closest relatives to the candidate genus are methylotrophic bacteria living in freshwater sediments (e.g. *Methylotenera versatilis*). It is thus proposed that the ancestral '*Ca. Methylospumilus*' species was found in a sedimentary habitat and then transitioned to the freshwater plankton. From there, a distribution throughout all connected water bodies was possible, including propagation into the oceans (Figure 22).

However, the transition from freshwater to saltwater cannot simply happen as a matter of fact, since the high salinity in oceans was proven to be one of the factors that hedge bacterial lineages in distinctive environments. The salinity is often too extreme for bacteria that are not specially adapted to allow for casual transitions, which, thus, rarely happen [Lozupone and Knight, 2008]. A common strategy to adapt to the osmotic challenge is to actively combat the salt gradient by synthesizing or transporting intracellular solutes accordingly [Walsh et al., 2013]. These solutes can be organic compounds such as amino acids or sugars, or inorganic. However, synthesizing amino acids and the like, is costly. Some marine bacteria solve this by relying on NQR (Na^+ -translocating NADH:quinone oxidoreductase), a membrane-bound enzyme. By pumping out Na^+ ions, NQR creates an inwardly directed Sodium Motive Force (SMF). As a consequence, ubiquinol is produced, and Na^+ -dependent, active transporters of amino acids and sugars are facilitated, thus proving to be a crucial tool to survive the high salinity in oceans.

These genes coding for the NQR complex were acquired by the OM43 clade from the marine Alphaproteobacterium *Rhodobacteraceae* sp. [Jimenez-Infante et al., 2016]. It is believed that this was accomplished by lateral gene transfer (LGT), a crucial factor to facilitate the transition by adopting metabolic features from the surrounding bacterial community.

Furthermore, it may be advantageous to have a specialized adaption to a specific ecological niche for a successful habitat transition. If the niche exists in mostly similar conditions in the new environment, only minor adaptions are necessary, as it was the case in OM43. Bacteria with a specialized methylotrophic metabolism paired with the ability to utilize light as an energy provider with rhodopsins only needed a mechanism to cope with the salt concentration in marine systems, which was achieved by acquiring NQR from a marine lineage by lateral gene transfer.

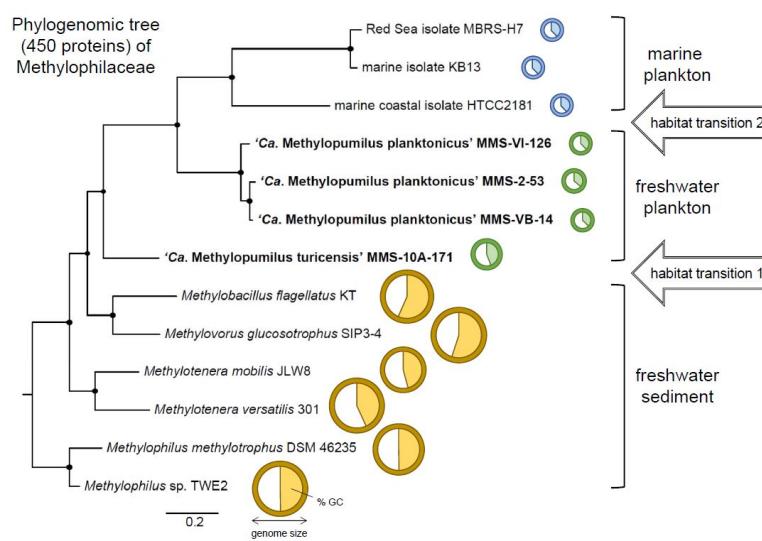


Figure 22: Phylogenetic tree based on COG annotations of several sedimentary, limnic and marine strains. Figure courtesy of Dr. M. M. Salcher.

4.1.2 Lateral gene transfer of rhodopsin genes

Similarly to NQR, the genes for rhodopsins were most likely obtained by means of lateral gene transfer. There is evidence that there were transfers occurring between different phyla such as Gammaproteobacteria and Actinobacteria, but also even between different domains (*Bacteria* to *Archaea*) [Pinhassi et al., 2016]. However, it is difficult to trace back the ancestry of rhodopsins. When analyzing the rhodopsin genes of various bacterial groups, Pinhassi et al. noticed a differentiation not only between, but also within genera (such as *Vibrio* or *Glaciecola*), where the rhodopsin phylogenomically clustered with either those of Alpha- or Gammaproteobacteria. Furthermore, proteorhodopsins found in Bacteroidetes commonly cluster together, clearly distant from the proteorhodopsins in Proteobacteria. This pattern can also be observed when comparing the rhodopsins of '*Ca. Methylopumilus planktonicus*', '*Ca. Methylopumilus turicensis*', an OM43 member, and the previously studied strains mentioned in the introduction (Section 1.2, Figure 3). A phylogenomic tree of these rhodopsin genes is shown in Figure 23.

Overall, lateral gene transfer of rhodopsin genes was not a rare event in the evolution of bacterial lineages, but rather happened quite frequently [Pinhassi et al., 2016]. A notable remark is that the synthesis of rhodopsins is coupled to the expression of carotenoid synthesis genes in order to correctly produce the retinal. In most rhodopsin-carrying bacteria, these consist of the four *crtEIBY* genes, flanked by a *blh* gene. For a functional rhodopsin, all six genes have to be present, and are therefore often found as a contiguous cassette in the genome (Figure 24). As the genes for the carotenoid synthesis are not present in most genomes by default, the whole cassette and not just the gene for rhodopsin alone had to be acquired by lateral gene transfer. This has evidently happened in '*Ca. Methylopumilus* sp.'. For '*Ca. Methylopumilus planktonicus*', the closest related rhodopsins according to BLAST are found in the Betaproteobacteria *Variovorax* (proteorhodopsin) and *Janthinobacterium* (xanthorhodopsin), but also in the Alphaproteobacterium *Sphingomonas* (xanthorhodopsin), which are all freshwater bacteria. Thus, the ancestral rhodopsins of '*Ca. Methylopumilus* sp.' were most likely transferred from a limnic Proteobacterium, and as the phylogenomic distances suggest, the lateral gene transfer occurred not just recently.

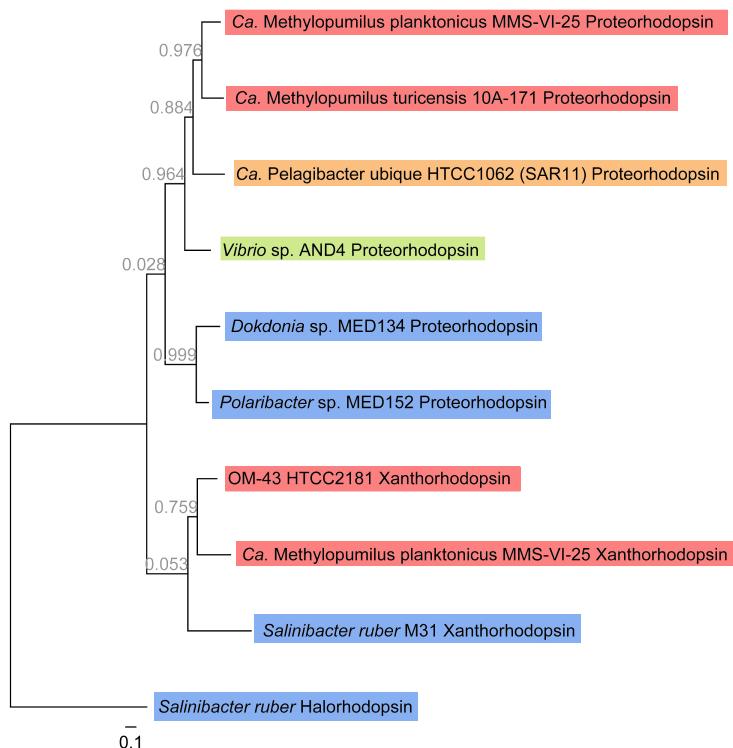


Figure 23: Phylogenomic tree of the rhodopsins of the candidate genus, and the rhodopsins from previous studies. Red = Betaproteobacteria, orange = Alphaproteobacteria, green = Gammaproteobacteria, blue = Bacteroidetes.

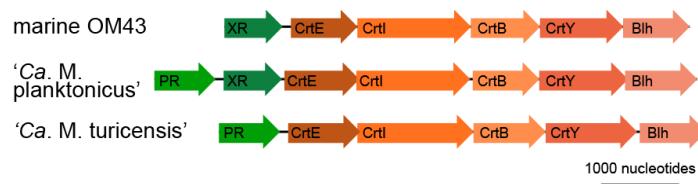


Figure 24: Schematic illustration of the Rhodopsin-CrtEIBY-Blh-cassette. PR = proteorhodopsin, XR = xanthorhodopsin. Figure courtesy of Dr. M. M. Salcher

4.2 Growth experiment

In this section, the results of the growth experiment are discussed in order to answer the second research question;

How does '*Ca. Methylospumilus planktonicus*' react to light in varying carbon concentrations?

4.2.1 Insights and limits

As previously described, there were striking differences between the conditions at the end of the growth experiment: The cultures of strain 1 (MMS-VI-25, two rhodopsins) were generally growing at a higher number in light than in darkness; the mean cell density was 1.28-fold higher in light-grown cultures than in dark conditions. On the contrary, strain 2 (MMS-M-34, no rhodopsins) had a clearly lower cell density in light-exposed setups than in dark conditions, exhibiting a fold change of 0.27.

Moreover, the biomass analysis showed that strain 1 had a clearly higher biomass when grown in light than in darkness, even more so in a high concentrated carbon medium, as expected. Strain 2 on the other hand showed generally lower biomass values, even though their cell size appeared to be bigger.

Cultures of strain 1 grown in light took up more carbon than their counterparts in the dark. As 4-5% carbon uptake seem very low, it must be considered that in the OM43 strain HTCC2181 approximately three quarters of the utilized carbon was oxidized to CO₂, while only a quarter was incorporated [Halsey et al., 2012]. Thus, it is hypothesized that the effective carbon uptake in '*Ca. Methylospumilus planktonicus*' was higher than the incorporated carbon implies. Assuming a CO₂ exhalation similar to the OM43 strain HTCC2181, the carbon utilization would range between 12-15%.

Another striking observation was that while the morphotypes shifted from Vibrio-shaped variants to small cocci or small rods, the biomass remained relatively stable. If the morphotypes would have gotten remarkably smaller, then the calculated carbon per cell would also be less, resulting in a lower biomass. This indicates that the Vibrio-shape is not necessarily smaller, and that the round, cocci-shaped form may be more efficient considering the surface-to-volume ratio. This also applies to the morphotypes found in strain 2; there, the large rods were replaced by small cocci, but the biomass did not decrease notably. There were negligible morphological differences between cultures grown in light or in the dark. This is contrary to the expectations derived from previous studies on *Ca. Pelagibacter ubique* [Steindler et al., 2011], where cells grown in the dark showed a distinct morphology compared to light-grown cells. However, this thesis' findings are based on light microscopy with DAPI-stained cells, whereas Steindler et al. analyzed their cells with a scanning electron microscopy.

In brief, the findings indicate the following:

1. **The rhodopsin-carrying strain was more successful in light-exposed, high-carbon conditions in terms of cell density, biomass accumulation and carbon uptake.**
2. **The strain without rhodopsins experienced a presumably light-inflicted damage, leading to a decrease in cell densities and biomass accumulation relative to cultures kept in the dark.**
3. **With a decreasing carbon availability in the culture medium, bacteria tended to shrink to smaller, rounder morphotypes.**

However, there are limits to these statements. If there was a substantial effect of rhodopsins, the pH of the surrounding fluid must have been noticeably lower, since rhodopsins are proton pumps that pump H⁺-ions from the intracellular to the extracellular space. As there was no significant difference in pH between dark/light conditions, nor between strain 1 and strain 2 (Figure 25), the potential activity of rhodopsin pumps did not have an influence on the acidity of the medium. Nevertheless, as described in Section 3.2 (Material and methods), the lowest pH values after 26 days were observed in the setups "LL1" and "LM1" (Light-Low-Strain1 and Light-Medium-Strain1). This might be an indication for an induced change of pH in light-exposed conditions, but it is not entirely certain if this can be attributed to the activity of rhodopsins.

Furthermore, as mentioned previously in the results (Section 3.2.3 Carbon uptake), the amount of accumulated carbon from the medium was >100% in the supposedly low carbon-setups. One possible explanation is that a mistake was made in the initial carbon amendment. However, in this case it would be expected of all other experiments that the results for "low carbon" setups to perform similarly to either the medium or high carbon concentrated setups, which was not the

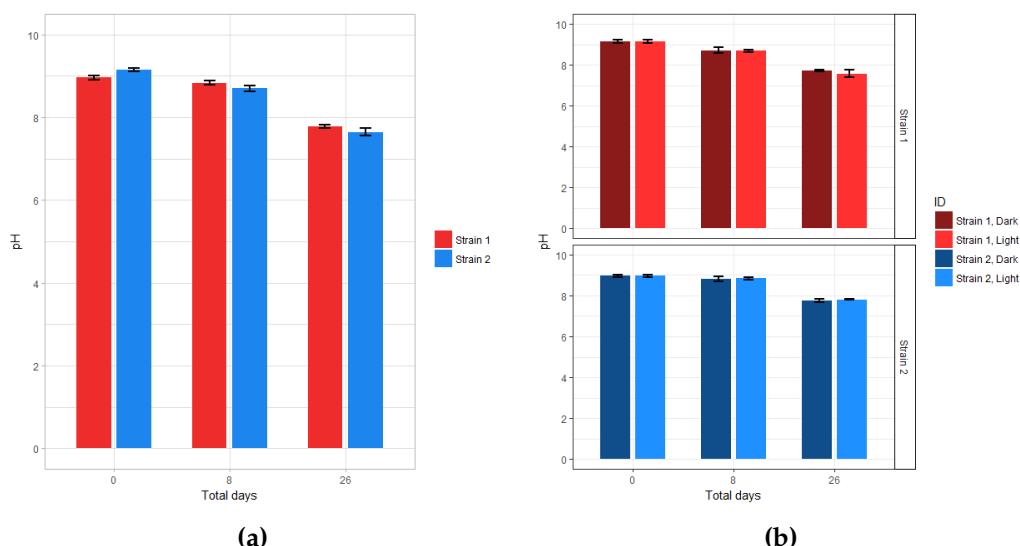


Figure 25: pH differences by strains (a), or by strains and dark/light (b)

case. Instead, there were significant differences in the low carbon conditions compared to medium or high conditions, for example in regard to cell densities (Figure 18 Growth curves), or when comparing the biomass (Figure 19 Biomass). This means, there must have been a deviation of the calculated, intended low carbon concentration when preparing the culture medium. This can be checked by quantifying the remaining methanol concentration in the cultures by photometric means [Mangos and Haas, 1996].

It would have been informative to observe the cell densities and biomass in even more severe starving conditions, which unfortunately could not be tested. Findings of previous works on other rhodopsins-bearing, aquatic bacteria, presented that rhodopsins were a crucial factor for surviving carbon-starvation ([Gómez-Consarnau et al., 2010; Steindler et al., 2011]). However, in the presented growth experiment cell densities were stable until the thesis' deadline in all conditions, indicating sufficient C-resources remaining in the medium. This can also be a consequence of the deviation of provided carbon in the supposedly low carbon concentrated medium.

5 Final remarks

The 26 newly sequenced genomes of '*Ca. Methylospumilus planktonicus*' helped to extend the knowledge about the phylogenomic context of the candidate genus. Especially the finding of the phylogenomic clusters in '*Ca. Methylospumilus planktonicus*' which suggest a categorization into three species instead, was not expected. Considering that the candidate species (formerly known as "LD28") has already been detected in numerous lakes around the world, it can be expected that many more sister species exist. As mentioned previously, there are implications of OM43 belonging to the same candidate genus of '*Ca. Methylospumilus sp.*', which thus includes both limnic and marine members. The three strains of the marine OM43 clade used here for comparison also showed a shared ANI-score below 95%, suggesting that they are distinct species, too. Overall, according to the presented findings, the genus of '*Ca. Methylospumilus sp.*' has a high species diversity.

Regarding the growth experiment there were some limitations, as previously described. In future growth experiments, lower methanol concentrations should be tested, also to induce starvation conditions after a more convenient running time. Furthermore, an analysis of the transcriptome would reveal the relevance of genes involved in C1-assimilation or rhodopsin synthesis. It could help to verify the dependence on rhodopsins in various conditions, or to close the gaps regarding the underlying metabolic processes of methylotrophy in the candidate genus. Also, there are still 192 "hypothetical proteins" in '*Ca. Methylospumilus planktonicus*' MMS-VI-25, which can be an artifact of the annotation process, however eventually one of these hypothetical proteins can be attributed to a function in scope of the transcriptional studies.

In conclusion, this thesis presents a comprehensive study of '*Ca. Methylospumilus planktonicus*' concerning phylogenomic context and growth-related responses in various growth conditions. The newly accomplished findings add to those of previous studies on aquatic methylotrophs, thus contributing to an extensive description of the novel genus '*Ca. Methylospumilus sp.*'.

References

- [Altschul et al., 1990] Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology*, 215:403–10.
- [Anthony, 1982] Anthony, C. (1982). *The biochemistry of methylotrophs*. Academic Press.
- [Aziz et al., 2008] Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kubal, M., Meyer, F., Olsen, G. J., Olson, R., Osterman, A. L., Overbeek, R. A., McNeil, L. K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G. D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., and Zagnitko, O. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC genomics*, 9:75.
- [Bahr et al., 1996] Bahr, M., Hobbie, J. E., and Sogin, M. L. (1996). Bacterial diversity in an arctic lake: a freshwater SAR11 cluster. *Aquat Microb Ecol*, 11(3):271–277.
- [Balashov et al., 2005] Balashov, S. P., Imasheva, E. S., Boichenko, V. A., Anton, J., Wang, J. M., and Lanyi, J. K. (2005). Xanthorhodopsin: a proton pump with a light-harvesting carotenoid antenna. *Science (New York, N.Y.)*, 309:2061–4.
- [Bankevich et al., 2012] Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotnik, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., and Pevzner, P. A. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology : a journal of computational molecular cell biology*, 19:455–77.
- [Béjà and Lanyi, 2014] Béjà, O. and Lanyi, J. K. (2014). Nature's toolkit for microbial rhodopsin ion pumps. *Proceedings of the National Academy of Sciences*, 111(18):6538–6539.
- [Béjà et al., 2001] Béjà, O., Spudich, E. N., Spudich, J. L., Leclerc, M., and DeLong, E. F. (2001). Proteorhodopsin phototrophy in the ocean. *Nature*, 411(6839):786–789.
- [Boichenko et al., 2006] Boichenko, V. A., Wang, J. M., Antón, J., Lanyi, J. K., and Balashov, S. P. (2006). Functions of carotenoids in xanthorhodopsin and archaerhodopsin, from action spectra of photoinhibition of cell respiration. *Biochimica et biophysica acta*, 1757(12):1649–1656.
- [Chistoserdova, 2011] Chistoserdova, L. (2011). Modularity of methylotrophy, revisited. *Environmental Microbiology*, 13(10):2603–2622.
- [Chistoserdova, 2015] Chistoserdova, L. (2015). Methylotrophs in natural habitats: current insights through metagenomics. *Applied Microbiology and Biotechnology*, 99(14):5763–5779.
- [Crowe et al., 2011] Crowe, S. A., Katsev, S., Leslie, K., Sturm, A., Magen, C., Nomosatryo, S., Pack, M. A., Kessler, J. D., Reeburgh, W. S., Roberts, J. A., González, L., Douglas Haffner, G., Mucci, A., Sundby, B., and Fowle, D. A. (2011). The methane cycle in ferruginous Lake Matano. *Geobiology*, 9(1):61–78.

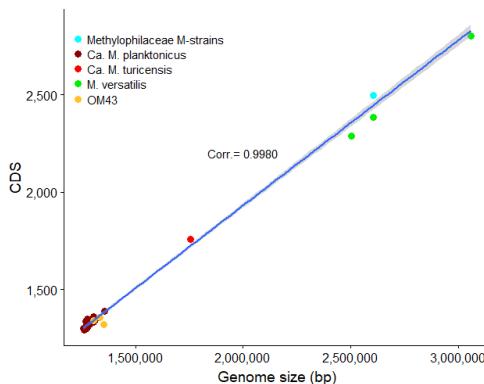
- [Fuhrman et al., 2008] Fuhrman, J. A., Schwalbach, M. S., and Stingl, U. (2008). Proteorhodopsins: an array of physiological roles? *Nat Rev Micro*, 6(6):488–494.
- [Galbally and Kirstine, 2002] Galbally, I. E. and Kirstine, W. (2002). The production of methanol by flowering plants and the global cycle of methanol. *Journal of Atmospheric Chemistry*, 43(3):195–229.
- [Gifford et al., 2016] Gifford, S. M., Becker, J. W., Sosa, O. A., Repeta, D. J., and DeLong, E. F. (2016). Quantitative transcriptomics reveals the growth- and nutrient-dependent response of a streamlined marine methylotroph to methanol and naturally occurring dissolved organic matter. *mBio*, 7(6).
- [Giovannoni et al., 2008] Giovannoni, S. J., Hayakawa, D. H., Tripp, H. J., Stingl, U., Givan, S. A., Cho, J.-C., Oh, H.-M., Kitner, J. B., Vergin, K. L., and Rappé, M. S. (2008). The small genome of an abundant coastal ocean methylotroph. *Environmental Microbiology*, 10(7):1771–1782.
- [Gómez-Consarnau et al., 2010] Gómez-Consarnau, L., Akram, N., Lindell, K., Pedersen, A., Neutze, R., Milton, D. L., González, J. M., and Pinhassi, J. (2010). Proteorhodopsin phototrophy promotes survival of marine bacteria during starvation. *PLOS Biology*, 8(4):1–10.
- [González et al., 2008] González, J. M., Fernández-Gómez, B., Fernández-Guerra, A., Gómez-Consarnau, L., Sánchez, O., Coll-Lladó, M., del Campo, J., Escudero, L., Rodríguez-Martínez, R., Alonso-Sáez, L., Latasa, M., Paulsen, I., Nedashkovskaya, O., Lekunberri, I., Pinhassi, J., and Pedrós-Alió, C. (2008). Genome analysis of the proteorhodopsin-containing marine bacterium Polaribacter sp. MED152 (Flavobacteria). *Proceedings of the National Academy of Sciences*, 105(25):8724–8729.
- [Goris et al., 2007] Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P., and Tiedje, J. M. (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *International Journal of Systematic and Evolutionary Microbiology*, 57(1):81–91.
- [Govorunova et al., 2017] Govorunova, E. G., Sineshchekov, O. A., Li, H., and Spudich, J. L. (2017). Microbial rhodopsins: Diversity, mechanisms, and optogenetic applications. *Annual review of biochemistry*, 86:845–872.
- [Halsey et al., 2012] Halsey, K. H., Carter, A. E., and Giovannoni, S. J. (2012). Synergistic metabolism of a broad range of C1 compounds in the marine methylotrophic bacterium HTCC2181. *Environmental Microbiology*, 14(3):630–640.
- [Hiorns et al., 1997] Hiorns, W. D., Methé, B. A., Nierzwicki-Bauer, S. A., and Zehr, J. P. (1997). Bacterial diversity in Adirondack mountain lakes as revealed by 16S rRNA gene sequences. *Applied and environmental microbiology*, 63:2957–60.
- [Huggett et al., 2012] Huggett, M. J., Hayakawa, D. H., and Rappé, M. S. (2012). Genome sequence

- of strain HIMB624, a cultured representative from the om43 clade of marine Betaproteobacteria. *Standards in Genomic Sciences*, 6(1):11–20.
- [Jacob et al., 2005] Jacob, D. J., Field, B. D., Li, Q., Blake, D. R., de Gouw, J., Warneke, C., Hansel, A., Wisthaler, A., Singh, H. B., and Guenther, A. (2005). Global budget of methanol: Constraints from atmospheric observations. *Journal of Geophysical Research: Atmospheres*, 110(D8):n/a–n/a. D08303.
- [Jimenez-Infante et al., 2016] Jimenez-Infante, F., Ngugi, D. K., Vinu, M., Alam, I., Kamau, A. A., Blom, J., Bajic, V. B., and Stingl, U. (2016). Comprehensive genomic analyses of the OM43 clade, including a novel species from the red sea, indicate ecotype differentiation among marine methylotrophs. *Applied and Environmental Microbiology*, 82(4):1215–1226.
- [Kanehisa, 2000] Kanehisa, M. (2000). *Post-genome Informatics*. Oxford University Press.
- [Konstantinidis and Tiedje, 2005] Konstantinidis, K. T. and Tiedje, J. M. (2005). Genomic insights that advance the species definition for prokaryotes. *Proceedings of the National Academy of Sciences of the United States of America*, 102(7):2567–2572.
- [Kuivila et al., 1988] Kuivila, K. M., Murray, J. W., Devol, A. H., Lidstrom, M. E., and Reimers, C. E. (1988). Methane cycling in the sediments of Lake Washington. *Limnology and Oceanography*, 33(4):571–581.
- [Langmead et al., 2009] Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultra-fast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, 10(3):R25.
- [Lassmann and Sonnhammer, 2005] Lassmann, T. and Sonnhammer, E. L. (2005). Kalign - an accurate and fast multiple sequence alignment algorithm. *BMC Bioinformatics*, 6(1):298.
- [Lindström et al., 2005] Lindström, E. S., Kamst-Van Agterveld, M. P., and Zwart, G. (2005). Distribution of typical freshwater bacterial groups is associated with pH, temperature, and lake water retention time. *Applied and environmental microbiology*, 71:8201–6.
- [Loferer-Krössbacher et al., 1998] Loferer-Krössbacher, M., Klima, J., and Psenner, R. (1998). Determination of bacterial cell dry mass by transmission electron microscopy and densitometric image analysis. *Applied and environmental microbiology*, 64:688–94.
- [Lohse et al., 2012] Lohse, M., Bolger, A. M., Nagel, A., Fernie, A. R., Lunn, J. E., Stitt, M., and Usadel, B. (2012). RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic Acids Research*, 40(W1):W622.
- [Lozupone and Knight, 2008] Lozupone, C. A. and Knight, R. (2008). Species divergence and the measurement of microbial diversity. *FEMS Microbiol Rev*, 32(4):557–578. 18435746[pmid].
- [Luecke et al., 2008] Luecke, H., Schobert, B., Stagno, J., Imasheva, E. S., Wang, J. M., Balashov, S. P., and Lanyi, J. K. (2008). Crystallographic structure of xanthorhodopsin, the light-

- driven proton pump with a dual chromophore. *Proceedings of the National Academy of Sciences*, 105(43):16561–16565.
- [Mangos and Haas, 1996] Mangos, T. J. and Haas, M. J. (1996). Enzymatic determination of methanol with alcohol oxidase, peroxidase, and the chromogen 2,2'-azinobis(3-ethylbenzthiazoline-6-sulfonic acid) and its application to the determination of the methyl ester content of pectins. *Journal of Agricultural and Food Chemistry*, 44(10):2977–2981.
- [Muyzer et al., 1995] Muyzer, G., Teske, A., Wirsén, C. O., and Jannasch, H. W. (1995). Phylogenetic relationships of Thiomicrospira species and their identification in deep-sea hydrothermal vent samples by denaturing gradient gel electrophoresis of 16S rDNA fragments. *Arch. Microbiol.*, 164(3):165–172.
- [Newton et al., 2011] Newton, R. J., Jones, S. E., Eiler, A., McMahon, K. D., and Bertilsson, S. (2011). A guide to the natural history of freshwater lake bacteria. *Microbiology and Molecular Biology Reviews*, 75(1):14–49.
- [Palovaara et al., 2014] Palovaara, J., Akram, N., Baltar, F., Bunse, C., Forsberg, J., Pedros-Alio, C., Gonzalez, J. M., and Pinhassi, J. (2014). Stimulation of growth by proteorhodopsin phototrophy involves regulation of central metabolic pathways in marine planktonic bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 111:E3650–8.
- [Pihassi et al., 2016] Pihassi, J., DeLong, E. F., Beja, O., Gonzalez, J. M., and Pedros-Alio, C. (2016). Marine bacterial and archaeal ion-pumping rhodopsins: Genetic diversity, physiology, and ecology. *Microbiology and molecular biology reviews : MMBR*, 80:929–54.
- [Posch et al., 2009] Posch, T., Franzoi, J., Prader, M., and Salcher, M. M. (2009). New image analysis tool to study biomass and morphotypes of three major bacterioplankton groups in an alpine lake. *Aquat Microb Ecol*, 54(2):113–126.
- [Posch et al., 2012] Posch, T., Koster, O., Salcher, M. M., and Pernthaler, J. (2012). Harmful filamentous cyanobacteria favoured by reduced water turnover with lake warming. *Nature Clim. Change*, 2(11):809–813.
- [Price et al., 2010] Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2 - approximately maximum-likelihood trees for large alignments. *PLOS ONE*, 5(3):1–10.
- [Qin et al., 2014] Qin, Q.-L., Xie, B.-B., Zhang, X.-Y., Chen, X.-L., Zhou, B.-C., Zhou, J., Oren, A., and Zhang, Y.-Z. (2014). A proposed genus boundary for the prokaryotes based on genomic insights. *Journal of bacteriology*, 196:2210–5.
- [RStudio Team, 2015] RStudio Team (2015). *RStudio: Integrated Development Environment for R*. RStudio, Inc., Boston, MA.
- [Salcher and Simek, 2016] Salcher, M. and Simek, K. (2016). Isolation and cultivation of planktonic freshwater microbes is essential for a comprehensive understanding of their ecology. *Aquatic Microbial Ecology*, 77(3):183–196.

- [Salcher et al., 2015] Salcher, M. M., Neuenschwander, S. M., Posch, T., and Pernthaler, J. (2015). The ecology of pelagic freshwater methylotrophs assessed by a high-resolution monitoring and isolation campaign. *The ISME Journal*.
- [Salcher et al., 2008] Salcher, M. M., Pernthaler, J., Zeder, M., Psenner, R., and Posch, T. (2008). Spatio-temporal niche separation of planktonic Betaproteobacteria in an oligo-mesotrophic lake. *Environmental Microbiology*, 10(8):2074–2086.
- [Steindler et al., 2011] Steindler, L., Schwalbach, M. S., Smith, D. P., Chan, F., and Giovannoni, S. J. (2011). Energy starved *Candidatus Pelagibacter Ubique* substitutes light-mediated ATP production for endogenous carbon respiration. *PLOS ONE*, 6(5):1–10.
- [Tatusov et al., 2000] Tatusov, R. L., Natale, D., Garkavtsev, I. V., Tatusova, T., Shankavaram, U. T., Rao, B., Kiryutin, B., Galperin, M., Fedorova, N. D., and Koonin, E. (2000). The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Research*, 29(1):22–28.
- [Walsh et al., 2013] Walsh, D. A., Lafontaine, J., and Grossart, H.-P. (2013). *On the Eco-Evolutionary Relationships of Fresh and Salt Water Bacteria and the Role of Gene Transfer in Their Adaptation*, pages 55–77. Springer New York, New York, NY.
- [Xia et al., 2015] Xia, F., Zou, B., Shen, C., Zhu, T., Gao, X.-H., and Quan, Z.-X. (2015). Complete genome sequence of *Methylophilus* sp. TWE2 isolated from methane oxidation enrichment culture of tap-water. *Journal of biotechnology*, 211:121–2.
- [Zotina et al., 2003] Zotina, T., Köster, O., and Jüttner, F. (2003). Photoheterotrophy and light-dependent uptake of organic and organic nitrogenous compounds by *Planktothrix rubescens* under low irradiance. *Freshwater Biology*, 48(10):1859–1872.
- [Zwart et al., 1998] Zwart, G., Hiorns, W. D., Methé, B. A., van Agterveld, M. P., Huismans, R., Nold, S. C., Zehr, J. P., and Laanbroek, H. J. (1998). Nearly identical 16S rRNA sequences recovered from lakes in north america and europe indicate the existence of clades of globally distributed freshwater bacteria. *Systematic and Applied Microbiology*, 21(4):546 – 556.

Supplementary

**Figure S1:** Genome sizes plotted against Coding DNA sequences (CDS). Correlation value: 0.9980345

E.C.	Substrate	Product	KEGG description	Class	Pathway
1.1.2.7	Methanol	Formaldehyde	methanol dehydrogenase	Oxidoreductases	Methane metabolism
1.2.99.5	Formyl-MFR	Formate	formylmethanofuran dehydrogenase	Oxidoreductases	Methane metabolism
1.2.1.2	Formate	CO2	formate dehydrogenase	Oxidoreductases	Methane metabolism
1.2.99.5	CO2	FormylMFR	formylmethanofuran dehydrogenase	Oxidoreductases	Methane metabolism
2.3.1.101	Formyl-MFR	N5-Formyl-THMPT	formylmethanofuran-tetrahydromethanopterin	Transferases	Methane metabolism
3.5.4.27	N5-Formyl-THMPT	5,10-Methenyl-THMPT	formyltransferase methenyltetrahydromethanopterin cy-	Hydrolases	Methane metabolism
5,10-Methenylene-THMPT	5,10-Methenyl-THMPT	5,10-Methenyl-THMPT	clohydrolase	Oxidoreductases	Methane metabolism
4.2.1.147	Formaldehyde	NA	formaldehyde-activating enzyme	Lyases	Methane metabolism
2.1.2.1	L-Serine	5,10-Methenylene-THMPT	serine aldolase, L-serine hydroxymethyltransferase	Transferases	Methane metabolism, Glycine Serine and Threonine metabolism
2.1.2.1	Glycine	L-Serine	glycine hydroxymethyltransferase	Transferases	Methane metabolism, Glycine Serine and Threonine metabolism
2.1.2.1	5,10-Methenylene-THF	THF (tetrahydrofolate)	NA	Transferases	Methane metabolism
6.2.1.1	Acetyl-CoA	Acetate	acetate-CoA ligase	Ligases	Methane metabolism, Glycolysis/Gluconeogenesis, Pyruvate metabolism
2.7.9.2	Pyruvate	Phosphoenolpyruvate	phosphoenolpyruvate synthase	Transferases	Methane metabolism, Pyruvate metabolism
4.1.1.31	Phosphoenolpyruvate	Oxaloacetate	phosphoenolpyruvate carboxylase	Lyases	Methane metabolism, Pyruvate metabolism
4.2.1.11	Phosphoenolpyruvate	Glycerate-2P	phosphopyruvate hydratase	Lyases	Methane metabolism, Glycolysis/Gluconeogenesis
5.4.2.12	Glycerate-2P	Glycerate-3P	phosphoglycerate mutase	Isomerases	Methane metabolism, Glycolysis/Gluconeogenesis, Glycine Serine and Threonine metabolism
2.6.1.52	3-Phosphoxypyruvate	3-Phosphoserine	phosphoserine transaminase	Transferases	Methane metabolism, Glycine Serine and Threonine metabolism, Vitamin B6 metabolism
3.1.3.3	3-Phosphoserine	L-Serine	phosphoserine phosphatase	Hydrolases	Methane metabolism, Glycine Serine and Threonine metabolism
4.1.2.13	D-Fructose-1,6-P2	D-Glyceraldehyde-3P	fructose-bisphosphate aldolase	Lyases	Methane metabolism, Glycolysis/Gluconeogenesis, Pentose phosphate pathway, Fructose and mannose metabolism
3.1.3.11	D-Fructose-1,6-P2	D-Fructose-6P	fructose-bisphosphatase	Hydrolases	Methane metabolism, Glycolysis/Gluconeogenesis, Pentose phosphate pathway, Fructose and mannose metabolism, RuMP-cycle
4.1.2.43	Fromaldehyde	D-Arabino-Hex-3-ulose-6P	3-hexulose-6-phosphate synthase	Lyases	Methane metabolism, Pentose phosphate pathway, RuMP-cycle
5.3.1.27	D-Arabino-Hex-3-ulose-6P	D-Fructose-6P	6-phospho-3-hexuloseisomerase	Isomerases	Methane metabolism, Pentose phosphate pathway, RuMP-cycle

Table S2: List of enzymes required for methane metabolism that are present in MMS-VI-25 and/or MMS-M-34.

	Carbon concentration		
	Low	Medium	High
d3 strain 1	0.2804	0.5411	0.8027
d3, strain 2	0.9018	0.9220	0.1815
d11, strain 1	0.2964	0.0169	0.3207
d11, strain 2	0.0122	0.2662	0.1929
d18, strain 1	0.3465	0.0037	0.4138
d18, strain 2	0.0011	0.1614	0.1738
d22, strain 1	0.5024	0.0749	0.3599
d22, strain 2	0.0003	0.9297	0.1638

Table S3: Resulting p-values of Welch Two Sample t-test of the cell densities.

Strain 1		Carbon concentration			Mean	SD
		Low	Medium	High		
Light		23'512'927.18	48'490'484.21	56'776'331.95	42'926'581.11	14'138'158.47
Dark		15'437'479.82	44'307'318.54	40'942'313.96	33'562'370.77	12'889'648.77
Difference		8'075'447.36	4'183'165.66	15'834'017.99	9'364'210.34	4'842'951.75
Factor L:D		1.523106586	1.094412521	1.386739694	1.279009203	
Strain 2		Carbon concentration			Mean	SD
		Low	Medium	High		
Light		1'935'815.32	16'402'869.73	7'591'179.18	8'643'288.07	5'952'820.94
Dark		7'178'658.90	16'429'361.15	13'916'490.23	12'508'170.09	3'905'670.86
Difference		5'242'843.58	26'491.42	6'325'311.05	3'864'882.02	2'749'892.82
Factor L:D		0.26966253	0.998387556	0.545480869	0.691011396	

Table S4: Cell density numbers at the last sampling day.

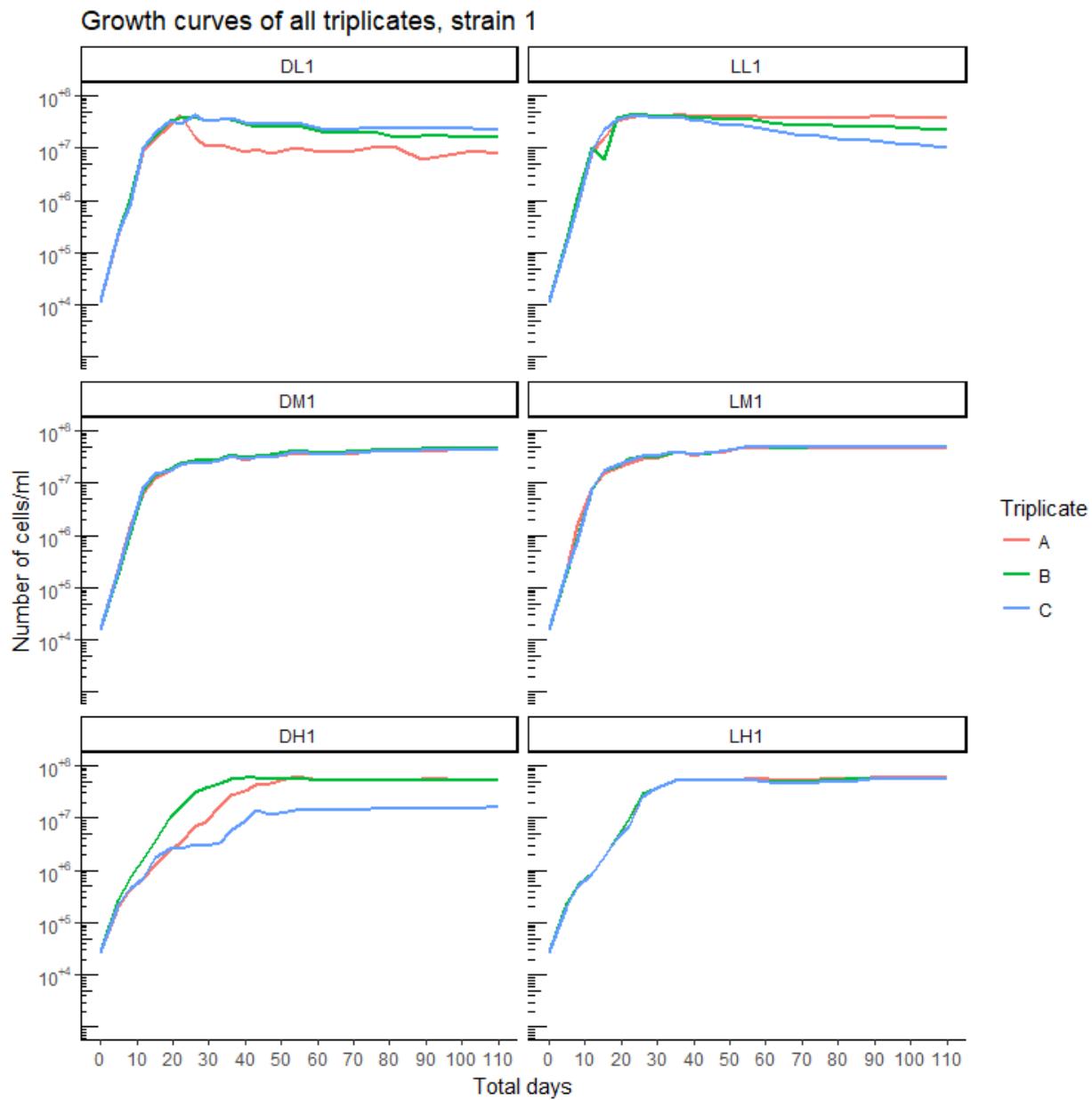


Figure S5: Cell density numbers of all triplicates of the growth experiment. Outliers (e.g. LL1B around day 15) were remeasured, but could not be resolved.

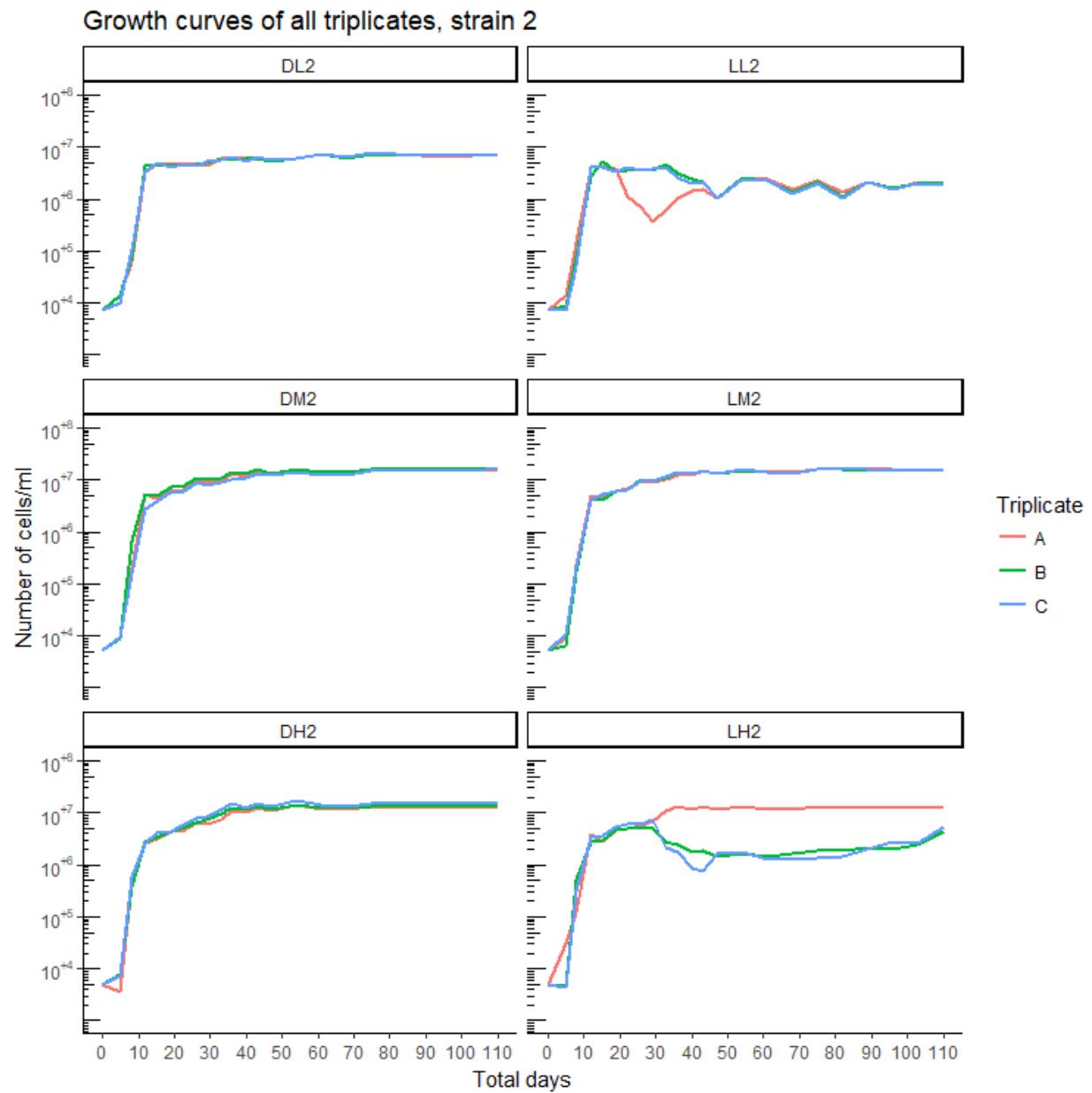


Figure S6: Cell density numbers of all triplicates of the growth experiment. As mentioned above, outliers were remeasured, but could not be resolved.

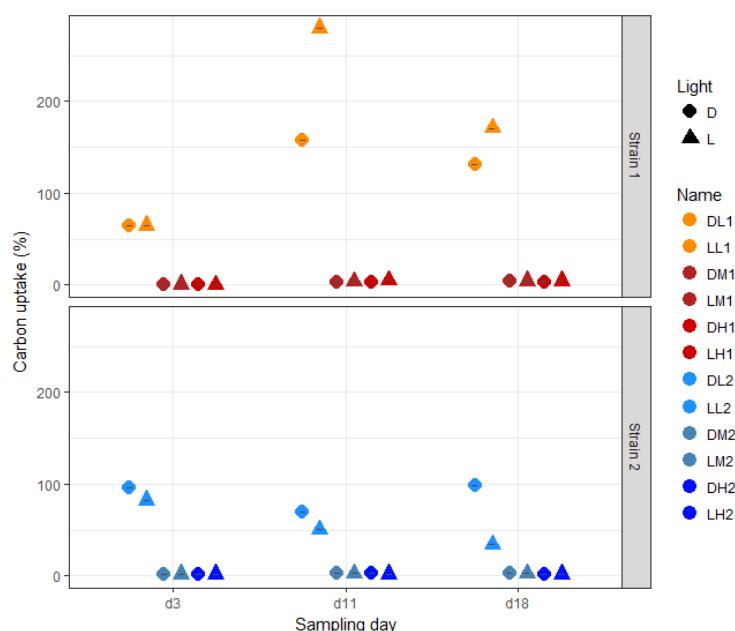
Sampling day	Total Days	Hours
d0	0	0
d1	5	120
d2	8	192
d3	12	288
d4	15	360
d5	19	456
d6	22	528
d7	26	624
d8	29	696
d9	33	792
d10	36	864
d11	40	960
d12	43	1032
-----		-----
d13	47	1128
d14	54	1296
d15	61	1464
d16	68	1632
d17	75	1800
d18	82	1968
d19	89	2136
d20	96	2304
d21	103	2472
d22	110	2640

Table S7: Sampling days conversion to total days and hours. The dashed line indicates the change of sampling interval from twice a week to once a week.

	Carbon concentration		
	Low	Medium	High
d3, Strain 1	0.920861	0.766940	0.843629
d3, Strain 2	0.416470	0.910933	0.168122
d11, Strain 1	0.202329	0.034698	0.207265
d11, Strain 2	0.226689	0.875642	0.204617
d18, Strain 1	0.625460	0.360004	0.437647
d18, Strain 2	0.007420	0.857959	0.210027

Table S8: Welch Two Sample t-test p-values of the biomass calculations, testing the difference between light and dark conditions.

	Sampling day		
	d3	d11	d18
	(%)	(%)	(%)
DL1	64.38	157.27	130.96
LL1	64.74	279.82	170.80
DM1	0.74	2.77	3.69
LM1	0.76	3.89	4.14
DH1	0.10	3.47	3.55
LH1	0.09	5.00	4.26
LL2	82.30	50.20	34.19
DL2	95.84	69.05	98.71
DM2	1.75	2.65	2.35
LM2	1.59	2.66	2.30
DH2	1.23	2.64	2.18
LH2	1.61	1.63	1.49

Table S9: Carbon uptake relative to the start concentration.**Figure S10:** Carbon uptake over time for all cultures, including the presumably faulty low carbon conditions.

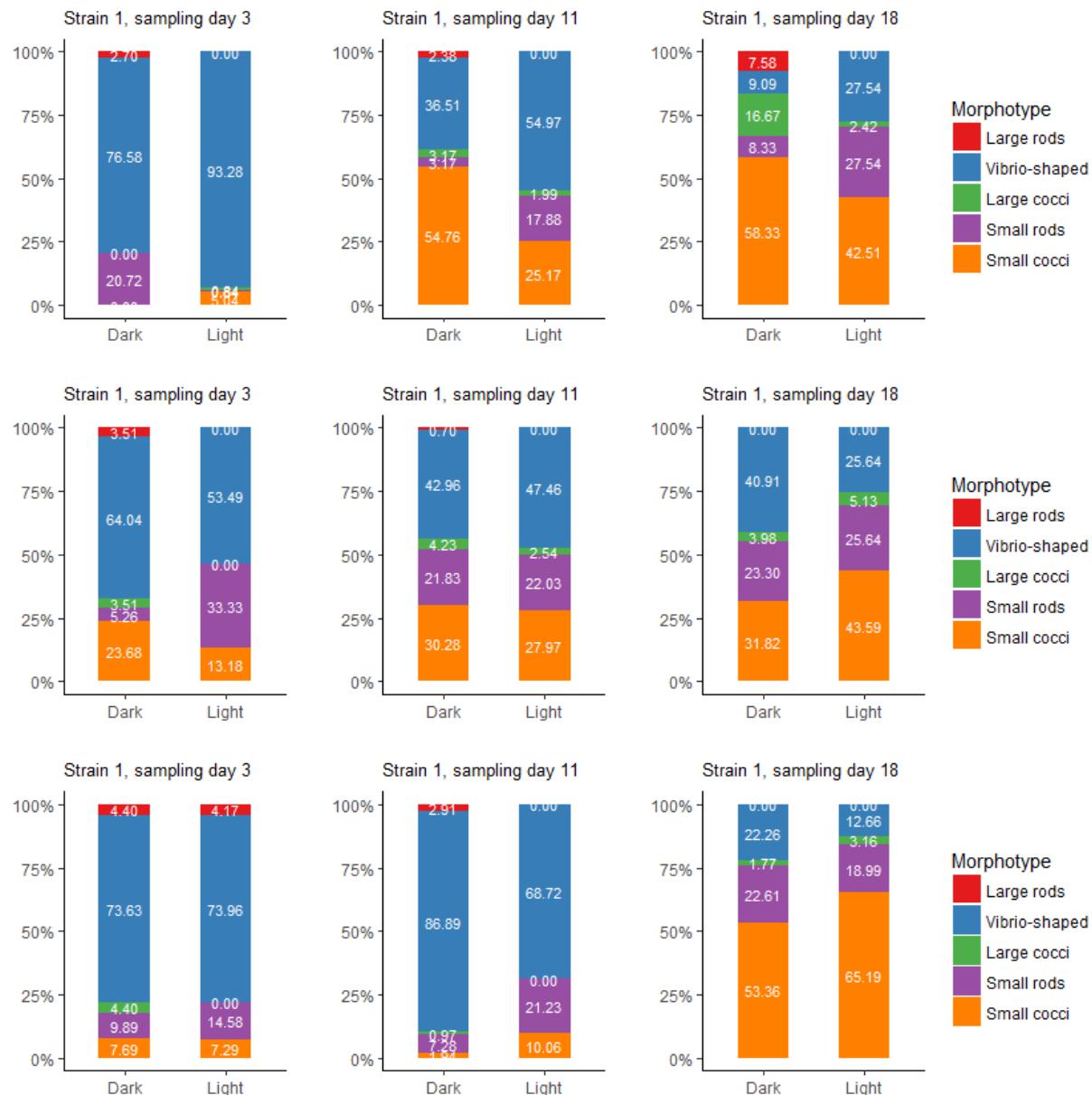


Figure S11: Distribution of morphotypes in strain 1 over time in light and dark conditions. Top row: low carbon concentrated medium, middle row: medium concentration, bottom row: high concentration.

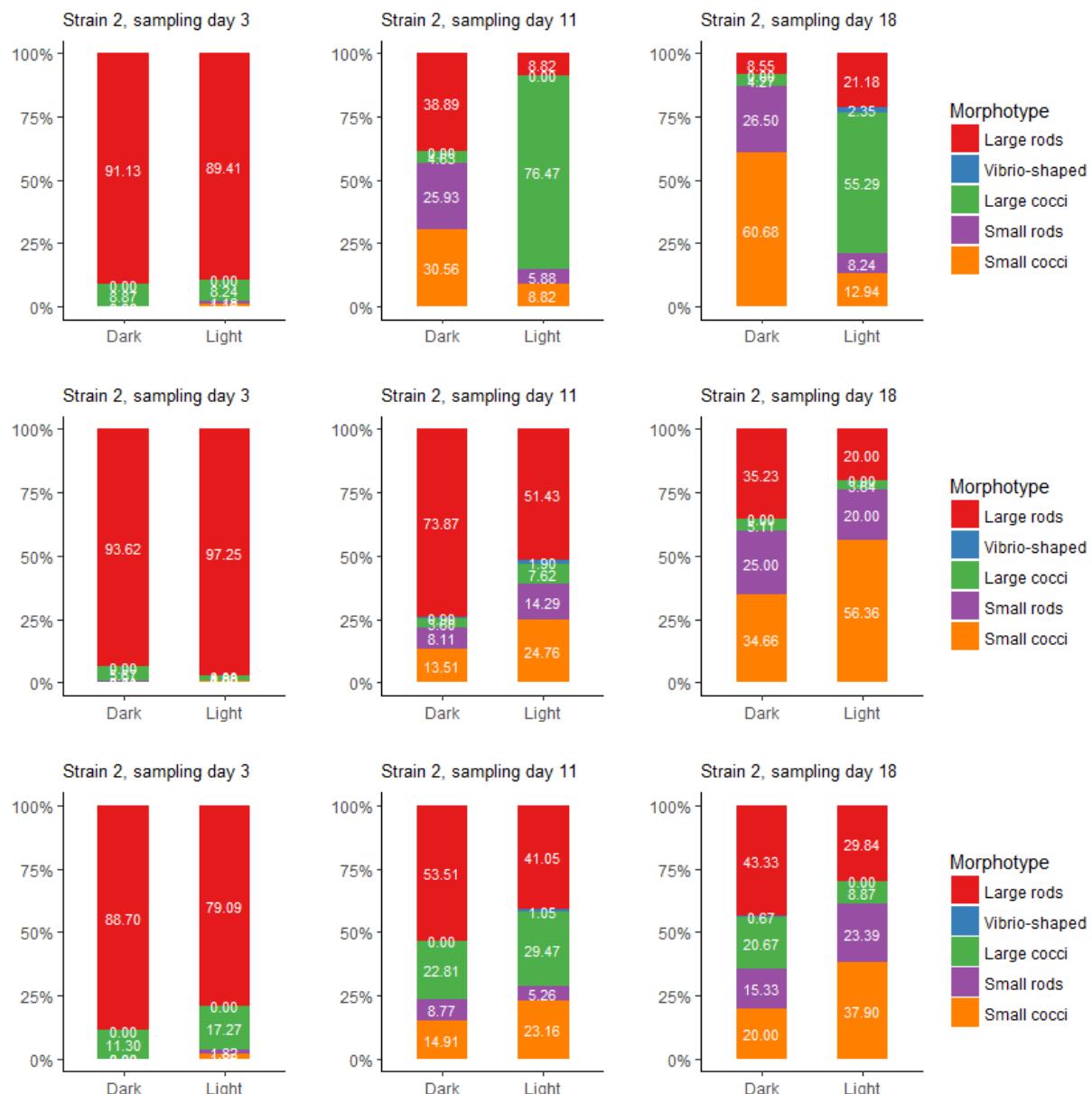


Figure S12: Distribution of morphotypes in strain 2 over time in light and dark conditions. Top row: low carbon concentrated medium, middle row: medium concentration, bottom row: high concentration.

Strain 1			Strain 2				
	d3	d11	d18	d3	d11	d18	
DL1A	102	132	157	DL2A	108	53	24
DL1B	137	257	206	DL2B	153	22	97
DL1C	133	102	164	DL2C	140	117	74
LL1A	79	242	271	LL2A	119	126	149
LL1B	138	209	187	LL2B	157	103	102
LL1C	136	223	182	LL2C	148	101	78
DM1A	76	162	213	DM2A	87	166	177
DM1B	117	215	188	DM2B	183	91	251
DM1C	134	170	274	DM2C	142	145	114
LM1A	87	215	185	LM2A	72	138	176
LM1B	133	213	239	LM2B	145	113	128
LM1C	141	236	239	LM2C	160	103	171
DH1A	70	284	378	DH2A	67	112	174
DH1B	132	202	270	DH2B	183	168	158
DH1C	109	133	152	DH2C	141	109	184
LH1A	55	250	221	LH2A	65	145	190
LH1B	122	198	221	LH2B	147	69	125
LH1C	128	240	274	LH2C	152	198	167

Table S13: Number of measured cells per sample for the biomass assessment.

	d3	d11	d18
DL1A	111	126	132
LL1A	119	151	207
DM1A	114	142	176
LM1A	129	118	117
DH1A	91	206	283
LH1A	96	179	158
DL2A	124	108	117
LL2A	170	102	85
DM2A	141	111	176
LM2A	109	105	110
DH2A	115	114	150
LH2A	110	95	124

Table S14: Number of measured cells for the morphotype assessment.

Statement of authorship

I declare that I have used no other sources and aids other than those indicated. All passages quoted from publications or paraphrased from these sources are indicated as such, i.e. cited and/or attributed. This thesis was not submitted in any form for another degree or diploma at any university or other institution of tertiary education.

Place, Date

Signature of author

