The University of South Bohemia in České Budějovice

Faculty of Science

Curated Database Construction for sMMO-Containing Methanotrophs

Bachelor thesis

Christina Zuser

Consultant: Mgr. Vojtěch Tláskal, Ph.D.

Supervisor: Dr. Anne Daebeler

České Budějovice 2023

Zuser, C., (2023). Curated Database Construction for sMMO-Containing Methanotrophs. Bc Thesis, in English. - p. 41, Faculty of Science, The University of South Bohemia, České Budějovice, Czech Republic.

Annotation

The significance of methane as an energy source is discussed, specifically focusing on understanding the ecology and metabolism of methanotrophs belonging to the families *Methylococcaceae*, *Methylocystaceae*, and *Beijerinckiaceae*. It includes applying bioinformatic tools, primer design, and a comparative analysis of metagenome-assembled genomes using the NCBI database, shedding light on phyla representation discrepancies. The study contributes to understanding methanotrophic diversity.

Declaration

I declare that I am the author of this qualification thesis and that in writing it, I have used the sources and literature displayed in the list of used sources only.

In České Budějovice on – 13.08.2023

Christina Zuser

Acknowledgments

I express my gratitude to the Institute of Soil Biology and Biogeochemistry, Biology Centre CAS, for permitting me to conduct my work at their facility.

I would also like to extend my appreciation to my advisor, Mgr. Vojtěch Tláskal, Ph.D., for providing valuable assistance with analyses, software, and insightful information throughout the whole process of working on my thesis.

Abstract

Methane (CH_4) plays a critical role in the global carbon cycle, being a significant greenhouse gas with negative environmental impacts. This study investigates the ecological and metabolic characteristics of methanotrophs, bacteria capable of utilizing CH_4 as their sole energy and carbon source. The families *Methylococcaceae*, *Methylocystaceae*, and *Beijerinckiaceae* are discussed, along with their distinct characteristics and habitats.

The study also delves into cytoplasmic soluble methane monooxygenase (sMMO), an enzyme responsible for CH_4 oxidation. Bioinformatic tools and approaches, including the NCBI database, sequence alignment, phylogenetic trees, primer design, and HMM models, are employed to analyse genetic relationships and design specific primers for *mmoX* genes in methanotrophs with sMMO. The study reveals the genetic diversity of *mmoX* genes within bacterial families, providing valuable insights for future ecological studies.

Comparative analysis of genomes retrieved from the Metagenome-Assembled Genomes (MAGs) catalogue and the NCBI database shows discrepancies in database size and phyla representation. The occurrence of sMMO in individual datasets is estimated. Due to its cultivation-independent character, the MAGs catalogue offers a more diverse range of methanotrophic taxa. At the same time, the NCBI database is comprehensive but may possess biased representation due to well-studied model organisms.

Overall, this research enhances our understanding of the diversity of methanotrophs, and the potential use of bioinformatic tools to analyse genetic data, contributing to a broader knowledge of methanotrophs' diversity and ecological importance.

Table of Contents

1. Introduction

1.1. Methane and its Relevance in the Ecosystem

Playing a fundamental role in the global carbon cycle, methane (CH_4) is an essential energy source for several microbial taxa and an important greenhouse gas. The production of CH_4 is achieved by methanogenesis, a microbial process that occurs in wetlands, rice paddies, the digestive tracts of termites and other animals, and other anoxic systems. In these habitats it is produced by specialized microbes called methanogens and represents the end-product of the anaerobic decomposition of organic matter. Methanotrophs oxidize CH_4, they are bacteria able to use CH_4 as an energy and carbon source. Despite its significance, accurately determining the global CH_4 budget remains challenging due to uncertainties surrounding its sources and sinks (Lee, 2019).

Increased CH_4 levels, for example, cannot fully be explained by increased emissions from wetlands. Warmer weather conditions leading to reduced hydroxyl radical (OH) concentrations in the atmosphere also contribute to a higher CH_4 concentration (Peng et al., 2022). Another study suggests that CH_4 emissions from beech trees are likely due to CH_4 production in stem tissues, not soil (Machacova et al., 2023). Including CH_4 emissions from beech trees in forest greenhouse gas inventories is crucial, as excluding them may overestimate the CH_4 sink capacity of forest systems (Machacova et al., 2023).

By acting as a greenhouse gas, CH_4 adds to climate change by trapping heat in the atmosphere via radiative forcing. Since CH_4 has an 84 times greater warming potential than CO_2 within 20 years, little changes in its concentration have dramatic effects (Zhang et al., 2023).

Wetlands, such as rice paddies, are known for their carbon-rich anoxic soils, making them the primary source of atmospheric CH_4. They emit a significant amount of CH_4, ranging from 142–284 Tg annually. Methanotrophs can consume up to 50% of the CH_4 produced in some soils, and their activity helps to prevent CH_4 from accumulating in the atmosphere (Kirschke et al., 2013).

Marine ecosystems cover almost 70% of the Earth's surface, but only contribute a small amount of CH_4 to the atmosphere, with an average of 12 Tg CH_4 per year (Lee, 2019). The release of CH_4 from aquatic systems involves various processes, such as diffusion, seepage, resuspension, or bioturbation of the sediments. The production of CH_4 typically occurs in

deposits or anoxic bottom water layers. Even oxic waters can be supersaturated with CH_4 (Lee, 2019).

In contrast, freshwater ecosystems, which comprise about 4% of the Earth's surface, can contribute up to 100 Tg CH_4 per year (Lee, 2019). Lakes specifically have a significant impact on the global CH_4 budget, contributing 6–16% to the total. As a result, they have been the subject of extensive research on CH_4 dynamics among freshwater systems (Tang et al., 2016).

Understanding the sources and sinks of CH_4 and developing strategies to reduce their emissions is crucial to mitigate the negative impacts of this vital gas on the environment (Lee, 2019).

1.2. Methanotroph Ecology and Metabolism

Methanotrophs are a unique group of bacteria that can use CH_4 as their only energy and carbon source (Op den Camp et al., 2009). Aerobic methanotrophs are found in habitats where CH_4 and oxygen are available. They belong to the bacterial phyla *Proteobacteria*, *Verrucomicrobia*, and *Methylomirabilota*, formerly candidate division NC10, and *Actinobacteria*. Notably, the first description of an aerobic methanotroph of the phylum *Actinobacteria* was published after the work for this thesis had been conducted (van Spanning et al., 2022). Therefore, the approach used in this thesis did not consider actinobacterial methanotrophs. Most cultivated methanotrophs belong to the proteobacterial classes *Gammaproteobacteria* and *Alphaproteobacteria* (Dedysh & Knief, 2018).

The oxidation of CH_4 in methanotrophs begins with the conversion of CH_4 into methanol by the methane monooxygenase (MMO) enzyme. Methanotrophs can express two types of MMO: the soluble methane monooxygenase (sMMO), found within the cytoplasm, and the particulate methane monooxygenase (pMMO), which is connected with the membrane. In the next step, CH_4 is oxidized to formaldehyde, which can be assimilated or oxidized to carbon dioxide by different pathways.

Methanotrophs can fix carbon using either the ribulose monophosphate (RuMP) cycle or the serine cycle, dividing them into type I and type II methanotrophs. Methanotrophs of type I (*Gammaproteobacteria*) use the RuMP cycle, while type II methanotrophs (*Alphaproteobacteria*) utilize the serine cycle. There is also an independent group called type X, which combines characteristics of types I and II (Gęsicka et al., 2021). Methanotrophs exhibit variations in the location of the intracytoplasmic membrane, the ability to fix nitrogen,

and significant phospholipid fatty acids, contributing to their classification and diversity (Knief, 2015).

1.2.1. Family *Methylococcaceae*

The family *Methylococcaceae* belongs to the *Gammaproteobacteria* class. *Methylococcaceae* methanotrophs, especially those of the genera *Methylococcus* and *Methylocaldum*, are commonly found in habitats characterized by high CH_4 concentrations and low oxygen levels. These two genera also contain type X methanotrophs that prefer higher temperatures for growth. Their habitats include peatlands, rice fields, and sediments of oxygen-limited marine or freshwater ecosystems (Dedysh & Knief, 2018).

The genus *Methylococcus* contains a species called *Methylococcus capsulatus*, a round-shaped bacterium lacking flagella and capable of forming cysts (Bowman, 2006). *M. capsulatus* possesses a Gram-negative cell envelope coated with a capsule made of exopolysaccharide, leading to the aggregation of cells. Certain *Methylococcus* strains can utilize methylamine, formate, or formamide as their sole carbon and energy sources. *M. capsulatus* has an oxygen-sensitive nitrogenase enzyme, which enables it to assimilate atmospheric nitrogen. The predominant fatty acids discovered in the *Methylococcus* species were identified as 16:0. The primary coenzyme involved in respiration is 18-methylene-ubiquinone-8, also known as MQ-8. They can tolerate higher temperatures, ranging from 40–50°C (Bowman, 2006).

Methylocaldum is a bacterial genus characterized by its ability to form motile, rod-like coccoidal cells (Bowman, 2006). Some species of *Methylocaldum* can develop spherical cells with Azotobacter-type cysts, and certain other species can form heat-resistant cysts (Whittenbury et al., 1970). They possess Gram-negative cell walls. Members of the genus *Methylocaldum* utilize nitrate or ammonia as nitrogen sources. Their fatty acid composition is primarily composed of 16:0, 16:1ω7c. Additionally, the primary respiratory quinone is MQ-8. *Methylocaldum* strains display a wide range of growth temperatures, from mesophilic to thermophilic, with optimal conditions varying between 20°C–55°C. (Bowman, 2006).

Methylomagnum ishizawai is the only cultured species in the genus *Methylomagnum*. Its cells are motile rods with a single polar flagellum and exhibit the typical intracytoplasmic membrane systems found in type I methanotrophs, forming bundles of membrane vesicles. These bacteria rely exclusively on CH_4 or methanol as their carbon and energy source and can utilize various nitrogen sources except nitrogen gas. The prominent fatty acids in the cells are

C16:1v7c, C16:0, and C14:0. The primary respiratory quinone detected is MQ-8. *M. ishizawai* is classified as neutrophilic, non-thermotolerant, and mesophilic. It thrives optimally at temperatures between 31–33°C and in a pH range of 6.8–7.4. (Khalifa et al., 2015).

Members of the genus *Methylomicrobium* are short rods that can exist singly or in pairs (Bowman, 2006). These organisms are capable of movement and possess one polar flagellum. They have a Gram-negative cell wall enclosed by a slimy capsule. Although *Methylomicrobium* strains cannot form cysts to overcome harsh conditions, they contain the energy storage compounds poly-β-hydroxybutyrate and polyphosphate granules, which aid in surviving during substrate limitation. They primarily utilize CH_4 or methanol as their carbon and energy sources. They can use nitrate and ammonia salts, casamino acids, yeast extract, and various amino acids as nitrogen sources. The most prevalent fatty acids of the *Methylomicrobium* species are 16:1ω5c, 16:1ω5t, 16:1ω8c, 16:1ω7c, and 16:0. The primary respiratory quinone identified in these species is MQ-8. *Methylomicrobium* species prefer mesophilic conditions, with optimal growth temperatures ranging from 10–30°C. Their growth thrives within a pH range of 6.0–9.0 (Bowman, 2006).

The genus *Methylovulum* contains aerobic members, which are nonmotile coccoids or short rods (Iguchi et al., 2016). Their colonies have a pale brown colour, a convex and smooth appearance, and lack cyst formation. They possess distinctive disk-shaped membrane vesicles and are classified as type I methanotrophs. These bacteria are obligate methanotrophs, relying solely on CH_4 and methanol as their carbon and energy sources, unable to utilize multicarbon compounds. Nitrate, ammonium salts, and glutamine are nitrogen sources for *Methylovulum* species (Iguchi et al., 2016). The fatty acid composition of *Methylovulum miyakonense* cells includes 14:0, 16:1ω8c, 17:0, and 17:1ω6c (Bussmann et al., 2021). Ubiquinone Q8 is the only quinone in the respiratory system (Deutzmann et al., 2014). They thrive in neutral pH conditions, with optimal growth at 25–30°C. They also tolerate lower temperatures, down to 5°C (Iguchi et al., 2016).

1.2.2. Family *Methylocystaceae*

Methylocystaceae bacteria belong to the phylum *Alphaproteobacteria*. They are often found in environments with moderate CH_4 concentrations and higher oxygen levels than other methanotrophs, such as forests, highlands, and agricultural systems (Dedysh & Knief, 2018).

The cells of *Methylocystis* strains are small, immotile, and have either a kidney bean or rod shape (Bowman, 2006). *Methylocystis* strains depend exclusively on CH_4 or methanol as their carbon and energy source. These strains possess catalase, cytochrome c oxidase, and an oxygen-sensitive nitrogenase enzyme for nitrogen fixation but thrive only under strict aerobic conditions. *Methylocystis* species exhibit a fatty acid composition mainly composed of 18:1ω8c, 18:1ω7c, and 18:0. The predominant quinone is Q-8. *Methylocystis* strains prefer mesophilic and neutrophilic environments, with optimal growth occurring at temperatures between 25–35°C and pH levels ranging from 6.5–7.5 (Bowman, 2006).

The genus *Methylosinus* comprises two primary species: *Methylosinus trichosporium* and *Methylosinus sporium* (Bowman, 2006). Both species share similar dimensions, ranging from 0.5–1.5 µm in width and 1.5–3 µm in length. However, they differ in shape, with *M. trichosporium* exhibiting a pear-shaped rod-like morphology, while *M. sporium* has a curved morphology. *Methylosinus* strains synthesize poly-β-hydroxybutyrate as a carbon reserve and possess standard Gram-negative cell walls. These strains are strictly aerobic methanotrophs, relying exclusively on CH_4 or methanol as carbon and energy sources. They can perform nitrogen fixation using an oxygen-sensitive nitrogenase enzyme. *Methylosinus* species have phospholipid fatty acids predominantly composed of 18:1ω8c, 18:1ω7c, and 16:1ω7c. The primary quinone observed is Q-8. *Methylosinus* strains are mesophilic and neutrophilic, with optimal growth at approximately 30°C and pH 6.5–7.0 (Bowman, 2006).

1.2.3. Family *Beijerinckiaceae*

The *Beijerinckiaceae* family of the phylum *Alphaproteobacteria* consists of Gram-negative aerobic bacteria with a tolerance to moderate acidity (Dedysh et al., 2016). These bacteria exhibit a wide range of metabolic capabilities. Methanotrophic members of this family are found in acidic soil environments such as Sphagnum bogs, tundra wetlands, and forest soils (Dedysh & Knief, 2018).

Methylocella bacteria are bipolar straight or curved rods that lack motility. Some species can form cysts. They utilize various carbon energy sources, including acetate, pyruvate, succinate, malate, propionate, and ethanol. Additionally, they possess the ability to fix nitrogen. The primary fatty acid composition of *Methylocella* species includes 18:1ω7c (Dedysh et al., 2016). *Methylocella palustris* exhibits optimal growth within a pH range of 5.0–5.5, with an optimal range of 15–20°C (Dedysh et al., 2000), similar to other species in this genus (Dedysh et al., 2004).

Methyloferula spp. are rod-shaped bacteria that can be straight or curved, forming rosettes (Dedysh et al., 2016). They lack mobility and do not produce cysts. These organisms rely exclusively on CH_4 or methanol as their carbon and energy sources. Additionally, they can fix nitrogen (Dedysh et al., 2016). The primary fatty acids in *Methyloferula* species, like *Methyloferula stellata,* consist of 18:1ω7c, and their major quinone is Q-10. These organisms are moderately acidophilic and mesophilic, capable of growth with an optimal pH of 4.8–5.2 and an optimum temperature of 20–23°C (Vorobev et al., 2011).

1.2.4. Family *Mycobacteriaceae*

Within the *Actinobacteria* genus, the *Mycobacteriaceae* family broadens the recognized scope of methanotrophic growth, surpassing the boundaries of *Proteobacteria* and *Verrucomicrobia*. *Mycobacterium methanotrophicum* was recognized as a CH_4 consumer within this family highlights its skill in utilizing CH_4 aerobically and suggests the occurrence of similar species in environments abundant in both CH_4 and acidity (van Spanning et al., 2022).

1.3. Methane Monooxygenase and its Variants

As mentioned, the primary function of MMO is the selective oxidation of CH_4 into methanol (Samanta et al., 2022).

1.3.1. pMMO

The enzyme pMMO is membrane-bound and found in almost all aerobic methanotrophic species except in strains from *Methyloferula stellata* (Haque et al., 2020) and the genus *Methylocella*. These specific strains only contain sMMO (Dedysh & Knief, 2018).

The functional pMMO enzyme comprises three subunits, in particular, PmoA, PmoB, and PmoC. These subunits are designated as α, β, and γ based on their decreasing molecular weights. PmoB has a molecular weight of 45 kDa, PmoA has a molecular weight of 27 kDa, and PmoC has a molecular weight of 23 kDa. While its primary substrate is CH_4, pMMO exhibits activity toward certain straight-chain hydrocarbons and linear alkenes (Lee, 2019).

The active sites within the pMMO enzyme are specific to the radius of the substrate, and there is evidence of multiple binding sites within the monomeric subunits. The functional protein pMMO requires binding three copper ions at the A, B, and C sites and one zinc ion at the C

site (Lee, 2019). The availability of copper ions plays a fundamental role in determining the expression and use of pMMO or sMMO in methanotrophs that contain both enzymes. These methanotrophs typically switch to sMMO under low copper availability. Additionally, the X-ray structure of pMMO reveals an empty hydrophilic cavity referred to as the D site. This cavity consists of specific amino acids from the PmoA and PmoC subunits and may have functional implications which are not fully understood yet (Lee, 2019).

1.3.2. sMMO

The sMMO is a metalloenzyme in the cytoplasm (Gęsicka et al., 2021) that utilizes a diiron cluster in its active site to transform CH_4 into methanol (Srinivas et al., 2020). The presence of the sMMO enzyme has been confirmed in several methanotroph species. *Methylococcus*, *Methylovulum*, *Methylocaldum*, *Methylomicrobium*, *Methylomagnum*, *Methylocystis*, and *Methylosinus* are among the species that possess the sMMO and pMMO enzymes. Notably, the genus *Methylocella* and the species *Methyloferula stellata* exclusively rely on the sMMO enzyme (Lee, 2019).

The sMMO enzyme relies on three vital components: the large hydroxylase (MMOH) with a molecular weight of 251 kDa, the small regulatory protein (MMOB), with 15.9 kDa, and the reductase (MMOR), with 37.8 kDa, which facilitates the catalytic activity. The relative amounts of MMOB and MMOR bound to MMOH directly impact the enzyme's catalytic efficiency and product distribution. MMOH and MMOB/MMOR interaction involves multiple binding sites characterized by different dissociation constants (Lee, 2019).

MMOH, the most substantial constituent of sMMO, forms a homodimer that comprises α, β, and γ subunits. MMOB is encoded by the *mmoB* gene and MMOR by the *mmoC* gene (Lee, 2019). The genes coding for the sMMO proteins are arranged in an operon, including *mmoX*, *mmoY*, *mmoB*, *mmoZ*, and *orfY*, also called MMOD and *mmoC* (Dalton, 2005). The transcription of sMMO genes is regulated by specific promoters and can be influenced by the availability of copper (Lee, 2019).

1.4. Bioinformatic Tools – NCBI, Phylogenetic Trees, Primer Design, HM Models, Genomes versus MAGs

Bioinformatic tools enable practical analysis and modelling of biological phenomena. Bioinformatics plays a crucial role in nucleic acid sequence analysis and the design of

oligonucleotide primers for PCR experiments (Singh & Kumar, 2001). Below is a summary of selected bioinformatic tools and their applications in microbial ecology research.

The National Center for Biotechnology Information (NCBI) is a part of the United States National Institutes of Health and hosts databases that are important resources for bioinformatic analyses. One of the major databases is GenBank for DNA sequences. It encompasses a wide range of information, including literature, health, genomes, genes, proteins, and chemicals. Through the Entrez system, users can perform search and retrieval functions across these databases (Agarwala et al., 2016).

MAFFT is an online tool for multiple sequence alignment (MSA) of amino acid or nucleotide sequences. It offers various options and additional functions for preprocessing and postprocessing MSA, including interactive sequence selection and phylogenetic inference (Katoh et al., 2018).

Phylogenetic trees are crucial in biology for understanding species relationships, evolutionary transitions, gene origins, adaptation, morphological evolution, and demographic changes. Building accurate trees involves addressing challenges, using reliable data analysis methods, and mitigating errors (Kapli et al., 2020). IQ-TREE is a fast and efficient maximum-likelihood tree inference method that explores tree space effectively, often finding high likelihoods (Nguyen et al., 2015a). The interactive Tree Of Life (iTOL) platform is an online service designed to showcase, manipulate, and interactively annotate phylogenetic trees (Letunic & Bork, 2021).

Primer-BLAST is a bioinformatic tool provided by NCBI for designing PCR primers specific to a given template sequence (Ye et al., 2012). Optimizing the design of synthetic oligonucleotide primers is vital for the success of PCR experiments, enabling accurate and specific amplification of desired sequences, by considering factors such as sequence, length, and melting temperature (Delghandi et al., 2022).

The accelerated HMMER3 software package, which utilizes profile Hidden Markov Models (HMMs) and probabilistic inference methods, dramatically enhances the speed and sensitivity of homology searches in sequence databases. It achieves a rate comparable to the Basic Local Alignment Search Tool (BLAST), a commonly used tool to find regions of similarity between sequences while maintaining a negligible loss in sensitivity (Eddy, 2011).

The term genomes refers to the entire genetic information of an organism (Mahner & Kary, 1997), whereas the term Metagenome-assembled genomes (MAGs) refers to the genetic

information of a single (micro) organism that has been reconstructed from metagenomic data (Setubal, 2021a). MAGs extend the knowledge of bacterial and archaeal diversity, enabling comparative analyses, metabolic modelling, and the exploration of genomic characteristics of uncultivated microorganisms with implications for ecosystem processes (Nayfach et al., 2021).

2. Aims

This study aims to achieve several objectives. The first objective involves gathering existing sequences of *mmoX* genes sourced from methanotrophs containing sMMO. These collected sequences will then be aligned, and an HMM will be constructed. Furthermore, the diversity of the *mmoX* gene will be clarified by creating a similarity tree. By utilizing the *mmoX* genes' information, the study designs primers that are specific to sMMO. Lastly, the research will extend to explore the diversity of methanotrophs across various databases.

3. Materials and Methods

In this study, the research was focused solely on utilizing bioinformatic tools.

3.1. Multiple Alignment Construction

In the context of this investigation, the method of multiple alignments was employed to analyse the genetic sequences of methanotrophs with genes for sMMO in their genomes. A list of methanotrophs available in Dedysh 2018 was used to retrieve the names of individual taxa with sMMO (Table 1).

Table 1: sMMO-containing methanotrophs. Taxa marked by an asterisk were used for BLAST and to construct a phylogenetic tree.

Species	**Taxonomy (Class: Order: Family)**
Methylomicrobium japanense	*Gammaproteobacteria: Methylococcales: Methylococcaceae*
Methylocaldum marinum	
Methylococcus capsulatus	
*Methylomagnum ishizawai**	
*Methylovulum miyakonense**	
Methylosinus trichosporium	*Alphaproteobacteria: Rhizobiales: Methylocystaceae*
*Methylosinus sporium**	
*Methylocystis heyeri**	
Methylocystis hirsuta	
Methylocystis bryophila	
*Methylocella palustris**	*Alphaproteobacteria: Rhizobiales: Beijerinckiaceae*
*Methylocella silvestris**	
*Methylocella tundrae**	
Methyloferula stellata	

The NCBI GenBank database (https://www.ncbi.nlm.nih.gov) was accessed to download either the partial or complete nucleotide and amino acid sequences of the *mmoX* genes for the specific species listed in Table 1. The downloaded sequences were merged into a single FASTA file to facilitate further analysis. This FASTA file was then uploaded to the MAFFT online service for a multiple sequence alignment with default settings (Katoh et al., 2018). MAFFT employs the F-INS-i technique, an iterative refinement method that integrates pairwise alignment data into the process of forming a multiple alignment. This approach

involves a customized version of the fasta34 program, with the Smith-Waterman optimization turned off, to gather local alignment information during the process. (Katoh et al., 2005)

The resulting aligned sequences from MAFFT were saved as a FASTA file, which was subsequently uploaded to the online version of IQ-TREE for phylogenetic tree construction using the maximum likelihood algorithm. The integrated Model Finder algorithm found the best-fit model to be a "LG4m" mixture model and the phylogenetic tree was calculated with 1,000 bootstraps iterations (Nguyen et al., 2015b). The corresponding TREEFILE was downloaded and uploaded to the interactive Tree of Life (iTOL) platform for visualization and analysis (Letunic & Bork, 2021). The resulting tree allowed for the identification of outliers and optimization of the multiple alignments.

3.2. Phylogenetic Tree Construction

To construct the phylogenetic tree, a tBLASTn search on NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi) was performed with the MmoX amino acid sequences of the species in Table 1 that are marked with an asterisk and which were selected as representative species for each order. TBLASTn is a specific mode of the BLAST tool hosted by NCBI that performs alignments between protein sequences and a nucleotide query that has been translated (Gerts et al., 2006). Amino acid sequences from bacterial organisms were used as the target database, and sequences from uncultured/environmental samples were excluded. The search parameters were set to a percentage identity above 95%, except for a single hit above 90%. Additionally, a query coverage above 90% of its length was required.

All aligned sequences of methanotrophs with a *mmoX* gene that met these parameters were downloaded. The nucleotide sequences were then converted into their corresponding amino acid sequences through translation using the online tool Sequence Manipulation Suite (https://www.bioinformatics.org/sms2/translate.html). Protein and DNA sequences were analysed and formatted using the Sequence Manipulation Suite, a collection of JavaScript programs (Stothard, 2000).

Next, the downloaded sequences from BLAST and the original amino acid sequences were merged into a single FASTA file. The approach from chapter 3.1. was repeated with this larger dataset (n=42) to obtain a phylogenetic tree of all found MmoX amino acid sequences (Figure 1).

3.3. Primer Design

In the primer design process, the following steps were undertaken. Firstly, the Primer-BLAST tool available on NCBI (https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?GROUP_TARGET=on) was utilized for primer design (Ye et al., 2012). The sequences from the phylogenetic tree were grouped according to tree clusters to find primers working for multiple strains and species with highly similar *mmoX* genes. In Primer-BLAST the default parameters were employed except that the primer melting temperature range was set to be between 45°C to 65°C, that the database used for Primer-BLAST was RefSeq RNA, and that the appropriate family, either *Methylococcaceae*, *Methylocystaceae* or *Beijerinckiaceae*, of organisms was selected. Following these steps, 21 primers were designed. Afterward, the In Silico PCR software (http://insilico.ehu.es/PCR/) was used for primer evaluation and problem prevention by ensuring specificity, identifying potential mismatches in binding sites, and avoiding amplification of unwanted amplicons (Yu et al., 2011). The parameters were maintained at the default settings, and the available genetic sequences of interest were utilized to evaluate the primer pair's specificity.

3.4. HMM Design

An HMM for amino acid sequences of all *mmoX* genes from section 3.2. was built using the software HMMER v3.3.1. (Eddy, 2011). A custom script was developed to match the sequences listed in section 3.5. against the created HMM profile, identify homologous sequences, and predict functional similarities using the “hmmsearch” function (Eddy, 2011). The script is deposited at Zuser, 2023.

3.5. High-Quality Genome Database

Several steps were undertaken to describe the distribution of *mmoX* genes in cultivated and uncultivated bacterial genomes. A local copy of the RefSeq database (https://www.ncbi.nlm.nih.gov/genome) obtained from NCBI by colleagues at ISBB was used to predict genes in 24,959 curated genomes, as available in February 2022, using Prodigal (Hyatt et al., 2010). The “hmmsearch” function of the software HMMER v3.3.1 (Eddy, 2011) was then used together with the *mmoX* HMM, from section 3.4., to find *mmoX* genes in the list of predicted genes. The resulting hits were filtered at E-value 1E-10 and compiled into a single text file for further analysis (Zuser, 2023).

In addition to the RefSeq database, a MAGs database was utilized in this study to represent uncultivated bacterial diversity. The MAGs dataset was obtained from the supplementary table of the publication "A genomic catalog of Earth's Microbiomes," which provided 9,143 high-quality MAGs with predicted genes (Nayfach et al., 2021, p. 13). As described above, the "hmmsearch" function was used together with the *mmoX* HMM to identify possible *mmoX* genes in the MAGs. A script was written in R (R Core Team, 2022) using RStudio (RStudio Team 2020) to process the output. The script enabled identifying the gene hits, filtering them at E-value 1E-15, and saving them in an Excel file (Zuser, 2023).

Furthermore, to gain insights from the gene hits obtained from both databases, various plots were generated using the R-language and packages "readxl" (Wickham et al., 2023) and "ggplot2" (Wickham, 2016). These analyses allowed for a comprehensive exploration and interpretation of the data, facilitating a deeper understanding of the characteristics and relationships of the identified *mmoX* genes within the high-quality genome and MAGs database.

4. Results

4.1. A reference MmoX Phylogenetic Tree

A phylogenetic tree was constructed with the MmoX sequences of the cultured strains and the sequences identified via BLAST (Figure 1). The strains without an accession number contain mmoX gene sequences from table 1, whereas those strains with an accession number are those with *mmoX* sequences obtained through BLAST searches.

Figure 1: Unrooted maximum likelihood tree of MmoX protein sequences. MmoX in Methylococcaceae strains (red), MmoX in Methylocystaceae strains (violet), and MmoX in Beijerinckiaceae strains (green). Strains without an accession number are from the mmoX *gene sequences of table 1 with an asterisk. Strains with the accession numbers are the sequences obtained through BLAST searches. The scale bar shows an estimation of 0.1 substitutions per amino acid site.*

The unrooted tree revealed distinct clusters corresponding to the phylogeny of the *mmoX*-hosting species, with those belonging to the families *Methylococcaceae*, *Methylocystaceae*,

and *Beijerinckiaceae* branching distinctly. The branches connecting these families appeared relatively long, indicating substantial genetic divergence between them.

The different strains of a species cluster near one another in the phylogenetic tree. Notably, the sequences of *Methylosinus sporium* strains showed an interesting pattern, with two forms of MmoX clustering relatively further from other *M. sporium* strains and with a high sequence similarity to other taxa.

4.2. Multiplex Primers

This section presents the investigation results for multiplex primers for seven groups of *mmoX* genes in methanotroph strains according to clusters in the phylogenetic tree. The strains were grouped into groups A to G based on their genetic relatedness (Figure 2).

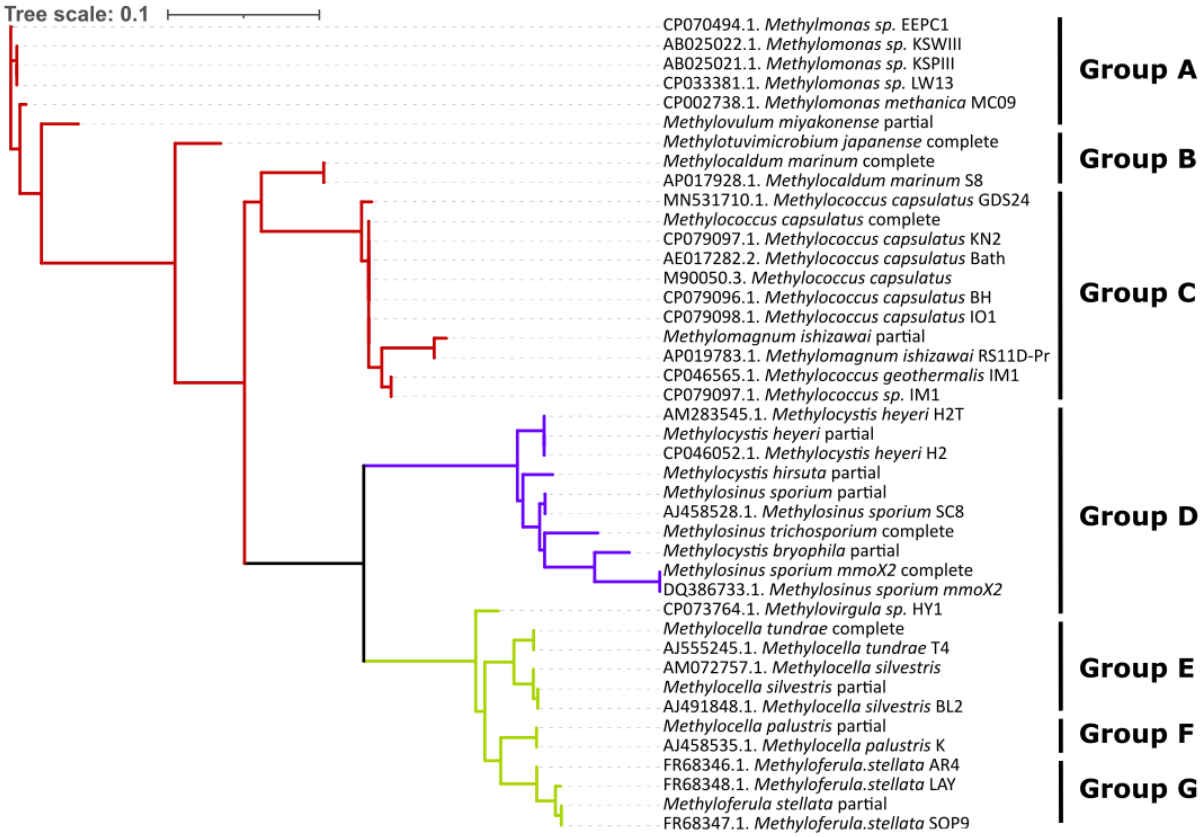

Figure 2: Unrooted maximum likelihood tree of MmoX protein sequences. MmoX in Methylococcaceae strains (red), MmoX in Methylocystaceae strains (violet), and MmoX in Beijerinckiaceae strains (green). Strains without an accession number are from the mmoX *gene sequences of table 1 with an asterisk. Strains with the accession numbers are the sequences obtained through BLAST searches. The scale bar shows an estimation of 0.1 substitutions per amino acid site. Groups on the right indicate the different target groups identified for primer generation. This figure represents the same phylogenetic tree as depicted in figure 1 but with a rectangular format.*

Tables 2 to 8 provide a comprehensive list of the primers identified for each group.

For *Methylococcaceae*, multiple primer pairs were found for the following three groups: Group A comprises strains from the genera *Methylomonas* and *Methylovulum*. The three primers found pairs exhibit remarkable similarity in both the forward and backward primer sites and the melting temperatures from 58.79–60.18°C (Table 2).

Table 2: Group A identified primers for mmoX in Methylomonas and Methylovulum.

Group A-Primers	Sequences (5'->3')	Length	Start	Stop	Tm/°C
1-Forward	TGGAAAGGCATGAAACGGGT	20	541	560	60.18
1-Reverse	GCCGTACTCGAACATCATGC	20	879	860	59.42
2-Forward	GTGGAAAGGCATGAAACGGG	20	540	559	59.76
2-Reverse	CGTACTCGAACATCATGCCC	20	877	858	58.79
3-Forward	TGTGGAAAGGCATGAAACGG	20	539	558	59.04
3-Reverse	CCGTACTCGAACATCATGCC	20	878	859	58.79

Group B comprises strains from the genera *Methylocaldum* and *Methylotuvimicrobium*, which exhibit notable genetic divergence, as evidenced by their positioning in the phylogenetic tree (Figure 2). Despite this genetic dissimilarity, primer pairs targeting these strains were successfully identified. Notably, the three primer pairs identified for group B shared significant similarities concerning both the forward and backward primer sequences and their respective melting temperatures, mirroring the characteristics observed for group A (Table 3).

Table 3: Group B identified primers for mmoX in Methylocaldum and Methylotuvimicrobium.

Group B-Primers	Sequences (5'->3')	Length	Start	Stop	Tm/°C
1-Forward	CCAACGGCGACGAAATCAC	19	677	695	59.87
1-Reverse	AAGTCGTGATGCGCCCAGTA	20	1037	1018	61.88
2-Forward	GCCAACGGCGACGAAATC	18	676	693	59.90
2-Reverse	GTCGTGATGCGCCCAGTA	18	1035	1018	59.82
3-Forward	GCCAACGGCGACGAAATCA	19	676	694	61.67
3-Reverse	AACAAGTCGTGATGCGCCC	19	1040	1022	61.33

Group C encompasses nine strains affiliated with the genus *Methylococcus* and with strains from *Methylomagnum ishizawai*. The multiplex primer analysis revealed notable differences in the melting temperatures among the three determined primer pairs, with the highest observed melting temperature of 62.7°C recorded for the third forward primer, as seen in table 4. Additionally, it is worth noting that the first and third primer pairs exhibited strikingly similar start- and endpoints. In contrast, the second primer pair featured a forward primer located approximately at position 431 and a corresponding backward primer around 783 (Table 4).

Table 4: Group C identified primers for mmoX genes in Methylococcus and Methylomagnum ishizawai.

Group C-Primers	Sequences (5'->3')	Length	Start	Stop	Tm/°C
1-Forward	GAATACAACGCCATCGCCG	19	340	358	59.72
1-Reverse	GGTGGGTGTGGCGGATTT	18	448	431	60.28
2-Forward	AAATCCGCCACACCCACC	18	431	448	60.28
2-Reverse	CGGATCGTTGGCGATGGA	18	783	766	59.89
3-Forward	GAATACAACGCCATCGCCGC	20	340	359	62.70
3-Reverse	GGTGGGTGTGGCGGATTTC	19	448	430	61.34

Group D encompasses all found strains belonging to the family *Methylocystaceae* from Figure 2. Notably, the three primer pairs identified for this group exhibit distinct melting temperatures compared to those identified for the other groups. The melting temperatures for these primer pairs range from 55.69–57.90°C. Furthermore, the second and third primer pairs share strikingly similar starting and endpoint positions. In contrast, the first primer pair consists of a forward primer around position 550 and a backward primer around position 690. Furthermore, the primer pairs were relatively short, consisting of a length of 16 or 17 nucleotides (Table 5).

Table 5: Group D identified primers for mmoX in Methylocystaceae.

Group D-Primers	Sequences (5'->3')	Length	Start	Stop	Tm/°C
1-Forward	TGGAAGGGCATGAAGCG	17	541	557	57.18
1-Reverse	GCGTGATCTCGTCGCC	16	697	682	57.81
2-Forward	CTCGAGGTCGGCGAATA	17	328	344	55.69
2-Reverse	CGCTTCATGCCCTTCCA	17	557	541	57.18
3-Forward	CTCGAGGTCGGCGAAT	16	328	343	55.75
3-Reverse	CGCGCTTCATGCCCTT	16	559	544	57.90

Group E consists of a single *Methylovirgula* strain and strains belonging to the genus *Methylocella*. The three determined primer pairs for this group exhibit comparable melting

temperatures. Additionally, these primer pairs demonstrate remarkable similarity in their starting and end points (Table 6).

Table 6: Group E identified primers for mmoX in Methylovirgula and Methylocella spp.

Group E-Primers	Sequences (5'->3')	Length	Start	Stop	Tm/°C
1-Forward	ATCGGACCGCTTTGGAAGG	19	529	547	60.08
1-Reverse	TGTCTTCTTCATCCGGCAGC	20	1105	1086	60.39
2-Forward	TCGGACCGCTTTGGAAGG	18	530	547	59.65
2-Reverse	TCCATGTCTTCTTCATCCGGC	21	1109	1089	60.13
3-Forward	ATCGGACCGCTTTGGAAGGG	20	529	548	62.54
3-Reverse	CTTCTTCATCCGGCAGCGA	19	1102	1084	60.15

Group F and group G represent closely related strains of the *Beijerinckiaceae* family. However, despite extensive efforts, no suitable primer was identified to amplify targets common to both groups. Only primers appropriate for amplifying *mmoX* genes of different strains within the same species were found. Regarding the *Methylocella palustris* species, which comprise group F, it was observed that all three primer pairs exhibited melting temperatures close to approximately 60°C. Notably, the third primer pair displayed markedly different start and stop positions than the first two (Table 7).

Table 7: Group F identified primers for mmoX in Methylocella palustris.

Group F-Primers	Sequences (5'->3')	Length	Start	Stop	Tm/°C
1-Forward	AATCCGCCACACCCATCAAT	20	249	268	60.03
1-Reverse	AAGGCGTTGTTGAGATCGGT	20	641	622	59.97
2-Forward	ATGAAATCCGCCACACCCAT	20	245	264	60.03
2-Reverse	CAGGGCTCGACCTTGAACTT	20	719	700	59.96
3-Forward	TCAAGGACGAGCGTCAGTTC	20	35	54	60.04
3-Reverse	CGCGAGATAGCCGTTCTTCT	20	234	215	59.97

In the context of group G, encompassing the *Methyloferula stellata* strains, it is noted that all three primer pairs exhibited closely similar melting temperatures, approximately around 60°C. However, it is crucial to highlight that despite the similarities in melting temperatures, these primer pairs are localized at distinct genomic locations (Table 8).

Table 8: Group G identified primers for mmoX in Methyloferula stellata.

Group G-Primers	Sequences (5'->3')	Length	Start	Stop	Tm/°C
1-Forward	TACGAGGATTGGGGTGGGAT	20	721	740	60.03
1-Reverse	CATCCTTCTTTGCATCGCGG	20	808	789	59.97
2-Forward	TCGTCGCTATCACCGAATGG	20	443	462	59.97
2-Reverse	ATCCCACCCCAATCCTCGTA	20	740	721	60.03
3-Forward	TCGCGTCGAACTTCCTTGAA	20	107	126	59.97
3-Reverse	CTCGTAGACCCAGCGATTCC	20	726	707	59.97

By in silico PCR analysis, the identified primers were assessed regarding their theoretical functionality. The results consistently indicated the prediction of a single amplification band for every primer pair, underscoring the specificity of the primers. An example of such in silico PCR is depicted in Figure 3. A predicted single band for group A, specifically the genus *Methylomonas*, derived from the use of primer pair 1, is displayed.

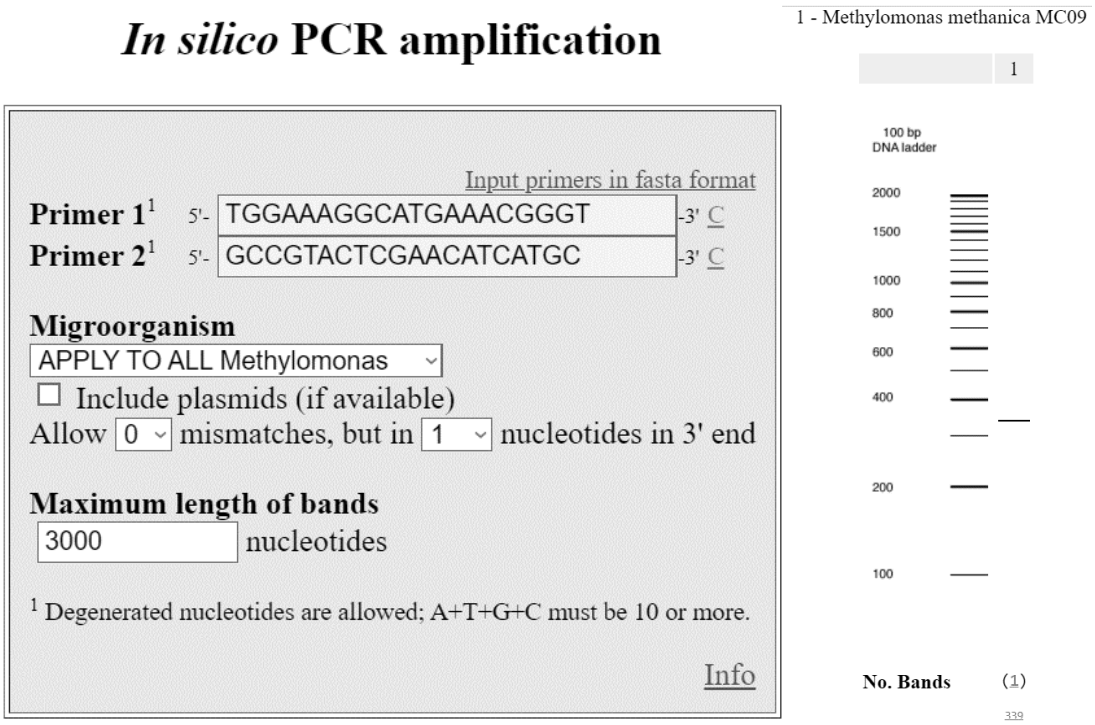

Figure 3: In silico PCR amplification of the first mmoX primer pair for group A.

4.3. Database Comparison

The outcomes yielded by the comparative analysis of genes retrieved from the MAGs database are laid out in this section. A total of 9,143 MAGs were included in the database, with three entries being unclassified and, therefore, not attributed to a specific phylum. The dataset encompassed a diverse range of microbial taxa, with 105 identified phyla.

Figure 4 illustrates the phylum distribution in the MAGs dataset using a pie chart. Notably, three dominant phyla emerged as the most prominent groups: *Proteobacteria*, *Bacteroidota*, and *Firmicutes A*, accounting for 20.9%, 19.8%, and 14.3 % of the total entries, respectively.

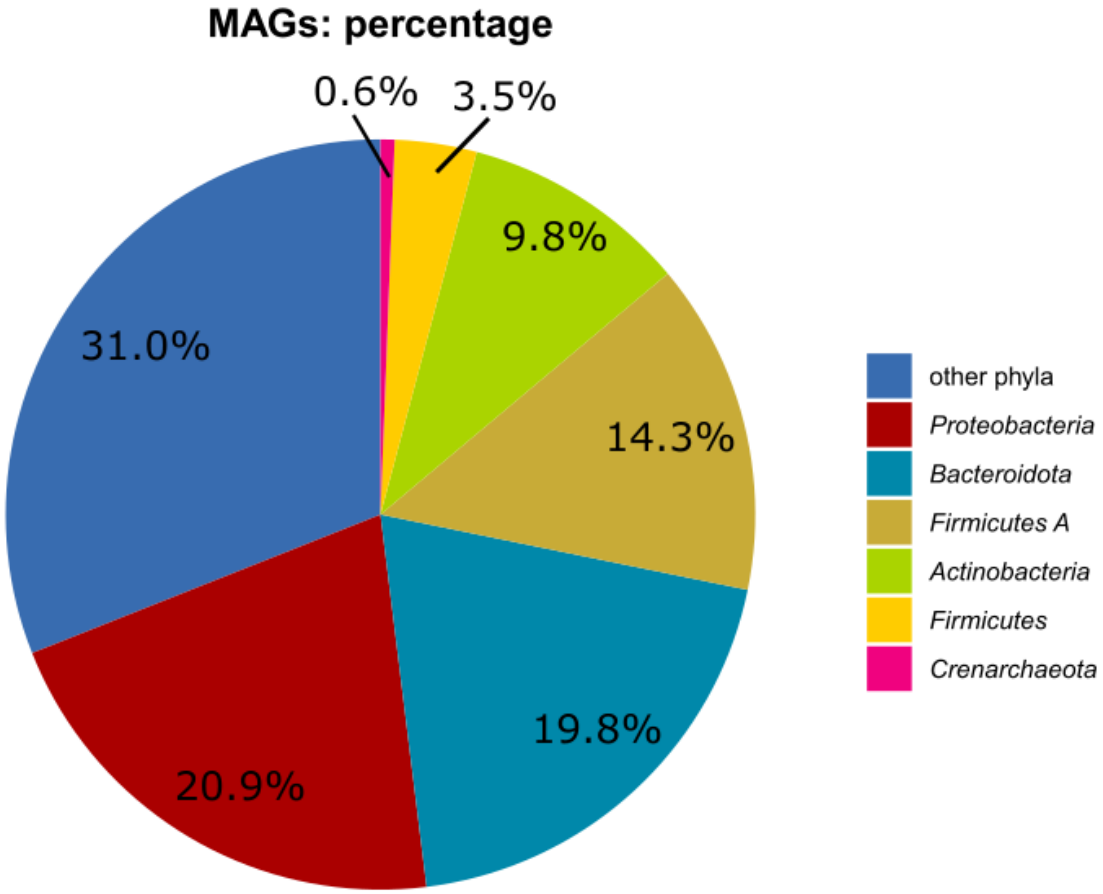

Figure 4: Pie chart representing phylum composition of MAGs database from Nayfach et al. (2021).

In the subsequent stage, a *mmoX* HMM that had been custom-built using selected *mmoX* sequences (see figure 2) was used to search among the MAGs for those containing a *mmoX* gene. The outcomes of this search are presented in figure 5 (B), while 5 (A) illustrates the total number of MAGs within each respective phylum. In total, 132 out of 9,143 genomes (1.4%) showed the presence of the *mmoX* gene. Out of all MAGs affiliated with the phylum *Proteobacteria* (1,915 MAGs), a mere 4.8% (93 MAGs) were found to contain *mmoX* genes. Of all MAGs belonging to other phyla, between 8.7% and 0.5% were found to contain *mmoX* genes.

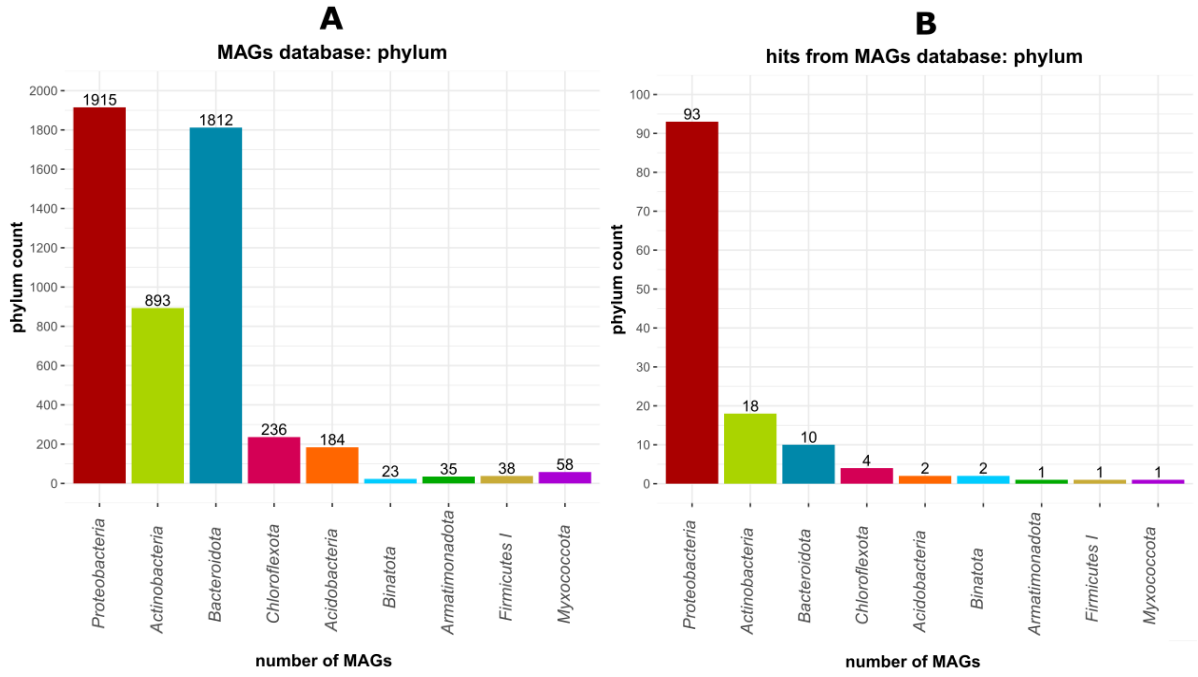

Figure 5: The overall number of MAGs in selected phyla (A) and MAGs with mmoX (B) of the MAGs database from Nayfach et al. (2021).

A total of 24,959 genomes were included in the database generated from NCBIs RefSeq entries, representing a comprehensive set of cultivated microbial taxa. In contrast to the MAGs database, the genome dataset encompassed a narrower phylogenetic diversity of microbial taxa, with 37 identified phyla (Figure 6). Notably, reflecting the reduced diversity of lower abundant phyla as compared to the MAGs database, the cumulative percentage of less prevalent phyla is considerably lower, with only 3.5%. Here, the three most abundant phyla were *Proteobacteria*, *Firmicutes*, and *Actinobacteria*, accounting for 58.0%, 21.7%, and 9.8% of the total genomes, respectively.

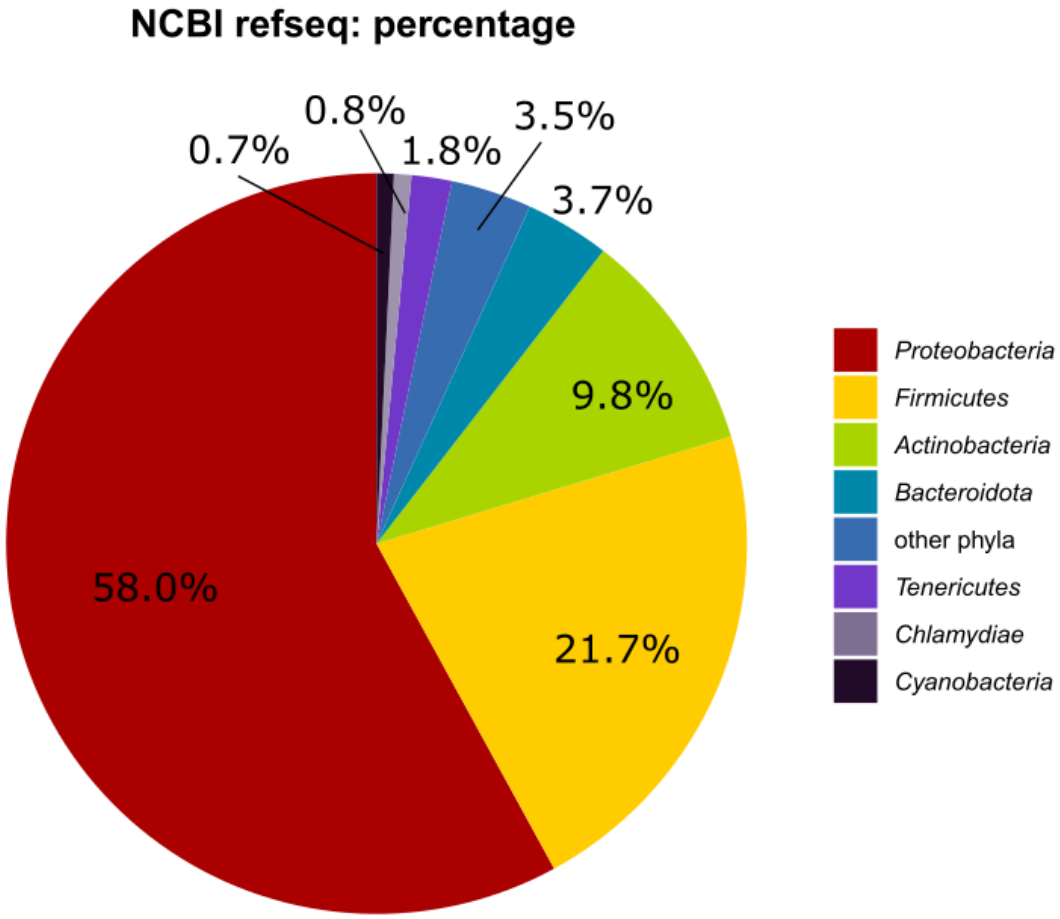

Figure 6: Pie chart showing bacterial phylum composition of NCBI RefSeq database retrieved in February 2022.

Next, a comparative search within the genomes from the NCBI RefSeq database, using the mmoX HMM, was conducted (Figure 7 (B)). In total, 757 out of 24,959 genomes (3.0%) showed the presence of the *mmoX* gene. The analysis revealed that out of the 14,529 proteobacterial genomes, 586 contained *mmoX* genes, which accounts for approximately 4% of the total genomes affiliated with this particular phylum. Remarkably, genomes with *mmoX* present were even rarer among other phyla, with none exceeding 7%.

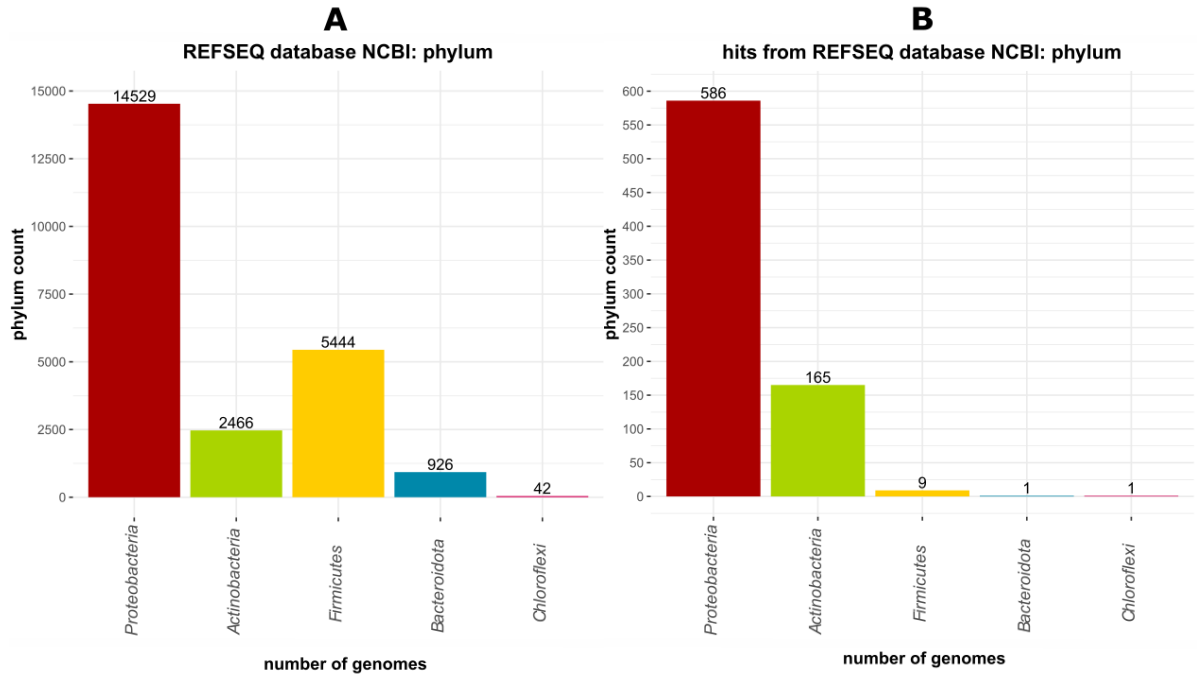

Figure 7: The overall number of genomes in selected phyla from the total NCBI RefSeq database (A) and from those containing the mmoX gene (B).

5. Discussion

The results obtained from the phylogenetic tree construction and the subsequent investigation into multiplex primers provided valuable insights into the genetic relationships and potential primer candidates for the *mmoX* genes in different bacterial species within the families *Methylococcaceae*, *Methylocystaceae*, and *Beijerinckiaceae*. The unrooted phylogenetic tree (Figure 1) clearly illustrated the distinct clustering of strains according to their respective phylogenetic families, indicating substantial genetic divergence among them.

Within the same species, the phylogenetic tree revealed variations in the genetic relatedness of different strains. The observation in *Methylosinus sporium* was particularly intriguing, where the *mmoX2* gene appeared more divergent from the *mmoX* gene in the same strain than from *mmoX* genes of other strains within the same species (Ali, 2006). This finding suggests the presence of distinct genotypic variations of target genes. Variations might be caused by evolutionary processes, like horizontal gene transfer (Dougherty et al., 2014), which may have led to a more significant divergence between these two gene variants within the same species. Such processes were identified to influence also other mono oxygenase, such as pMMO and ammonia monooxygenase (Khadka et al., 2018).

The investigation of possible *mmoX* primer pairs for specific groups of methanotrophs (Figure 2) successfully identified primer candidates for most groups. Even though it was impossible to design primer pairs to cover broader groups, primer pairs in groups A (*Methylomonas*) and B (*Methylotuvimicrobium* and *Methylocaldum*) exhibited significant similarity in their sequences and melting temperatures, indicating their potential applicability in the simultaneous amplification of multiple strains within these groups. Conversely, group F (*Methylocella palustris*) and group G (*Methyloferula stellata*), representing closely related species of the *Beijerinckiaceae* family, did not yield suitable primers to amplify targets common to both groups. However, the primers identified specific targets within the same species, but different strains were noteworthy.

The in-silico PCR analysis further validated the specificity of the identified primers, predicting a single amplification band for each primer pair. This characteristic is crucial for accurate and reliable amplification in subsequent laboratory experiments.

Overall, these results shed light on the genetic diversity and relatedness of *mmoX* genes in various bacterial taxa within the *Methylococcaceae*, *Methylocystaceae*, and *Beijerinckiaceae* families. The successful identification of multiplex primers for several groups opens possibilities for simultaneous amplification and targeted analysis of multiple species using the

FISH technique (Batani et al., 2019) or similar approaches. However, the inability to find primers amplifying common targets in closely related strains emphasizes the need for further exploration and optimization of primer design strategies.

The conclusions drawn from this study provide significant contributions to a better understanding of the genetic landscape of *mmoX* genes in methanotrophic bacteria and lay the groundwork for future investigations into their functional implications and ecological significance. Further research could focus on experimental validation of the identified primers and their application in environmental studies to unravel the diversity and dynamics of these essential genes in natural habitats.

The results obtained from comparing the MAGs and NCBI RefSeq databases provided valuable insights into their composition and scope differences. The MAGs database, despite containing a smaller number of total entries, namely 9,143, encompassed a more diverse range of microbial taxa with 105 identified phyla. In contrast, the NCBI RefSeq database included a much larger number of entries of 24,959 but exhibited a narrower range of microbial taxa with only 37 recognized phyla. This discrepancy in database size and phyla representation can be attributed to several factors, precisely the cultivation-independent technique of resolving MAGs and, in contrast to that, cultivation-biased genome collection at NCBI RefSeq with an over-representation of medically relevant microbial taxa.

One of the main reasons for the MAGs database's smaller size compared to NCBI RefSeq is the data generation technique. The MAGs database contains genomes assembled from metagenomic data. Metagenomic studies involve the sequencing of DNA directly from environmental samples, allowing for the exploration of diverse and uncultivable microbial communities without a primer-based PCR bias (Parks et al., 2017). However, assembling MAGs is complex and computationally intensive (Mattock & Watson, 2023), resulting in a limited number of MAGs available compared to the vast amount of raw metagenomic data in a single dataset. As a result, the MAGs database comprised of MAGs from numerous datasets captures a broader snapshot of microbial diversity, encompassing novel and less prevalent phyla that might need to be better represented in the NCBI database (Setubal, 2021b).

On the other hand, the NCBI database is a comprehensive repository that archives a wide array of biological data, including genomes from cultured isolates, whole-genome shotgun sequences, and metagenomic data. It is a centralized resource for researchers worldwide, and data submission to NCBI is relatively straightforward. As a result, it accumulates many

genomes from various studies, leading to a more significant total number of accessions. Genomes are often complete and contain fewer errors and less contamination than MAGs (Mattock & Watson, 2023). Additionally, the NCBI database includes a significant amount of data from well-studied model organisms and medically relevant species, such as human-associated bacteria, which might dominate the database and reduce the representation of less prevalent phyla (https://www.ncbi.nlm.nih.gov).

The identification of *mmoX* genes yielded candidate sMMO-containing genomes in 3% of genomes from the NCBI RefSeq database and in 1.4% of genomes from the MAGs database. Due to the character mentioned above in these datasets, the latter number most likely better represents the natural occurrence of sMMO in the microbial tree of life. Notably, the identification presented here was based on in silico sequence similarity comparison, and thus the actual function of targeted genes might differ. The comparison unveiled sMMO or sMMO-like genes in taxa in which CH_4 monooxygenases are identified very rarely, *Actinobacteria* (van Spanning et al., 2022) and *Binatota* (Murphy et al., 2021) or have not been identified yet, for example in *Bacteroidota*, *Chloroflexota*, and others. It will be an exciting avenue for further exploration to confirm or disprove the presence of a functional sMMO in microbes of these latter phyla.

6. Conclusion

In conclusion, this study encapsulated the significance of CH_4 in ecosystems, the ecology and metabolism of methanotrophs, and the diversity of vital families within this group based on various research papers. An overview of the involvement of MMO variants, pMMO and sMMO, in CH_4 oxidation was also provided.

The phylogenetic analysis revealed distinct clustering of MmoX in strains within families and variations in genetic relatedness within species. Multiplex *mmoX* primers were successfully identified for several methanotroph groups, offering the possibility of simultaneous amplification and other analysis possibilities. However, challenges were encountered with closely related strains. The study contributes to understanding the genetic landscape of *mmoX* genes.

Database search for MmoX genes in available microbial genomes from both culture-dependent and culture-independent genome repositories showed the presence of the gene in numerous bacterial phyla, with the majority of genes present in *Proteobacteria*, *Firmicutes*, *Actinobacteria* and *Bacteroidota*. While methanotrophs from *Proteobacteria* are widely recognized and *Actinobacteria* have been described only recently, other identified phyla contain yet unknown methanotrophs or genomes with MmoX-like genes that serve for other functions. Identification of the rest of the methanotrophic pathway is another step in the unambiguous determination of methanotrophic abilities in these taxa.

Overall, this research emphasizes the importance of bioinformatic tools in microbial ecology research and lays the groundwork for future investigations into CH_4-related processes. Further validation of identified primers and their application in environmental studies and continued investigations on the presence of sMMO in new microbial phyla will contribute to a better understanding of methanotrophic bacteria's diversity and dynamics in natural habitats.

7. List of References

Ali, H., Scanlan, J., Dumont, M. G., & Murrell, J. C. (2006). Duplication of the mmoX gene in *Methylosinus sporium*: cloning, sequencing and mutational analysis. *Microbiology*, *152*(10), 2931-2942. https://doi.org/10.1099/mic.0.29031-0

Agarwala, R., Barrett, T., Beck, J., Benson, D. A., Bollin, C., Bolton, E., Bourexis, D., Brister, J. R., Bryant, S. H., Canese, K., Charowhas, C., Clark, K., Dicuccio, M., Dondoshansky, I., Federhen, S., Feolo, M., Funk, K., Geer, L. Y., Gorelenkov, V., … Zbicz, K. (2016). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, *44*(D1), D7–D19. https://doi.org/10.1093/nar/gkv1290

Batani, G., Bayer, K., Böge, J., Hentschel, U., & Thomas, T. (2019). Fluorescence in situ hybridization (FISH) and cell sorting of living bacteria. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-55049-2

Bowman, J. (2006). The Methanotrophs — The Families *Methylococcaceae* and *Methylocystaceae*. In *The Prokaryotes* (pp. 266–289). Springer, New York. https://doi.org/10.1007/0-387-30745-1_15

Bussmann, I., Horn, F., Hoppert, M., Klings, K. W., Saborowski, A., Warnstedt, J., & Liebner, S. (2021). Methylomonas albis sp. nov. and Methylomonas fluvii sp. nov.: Two cold-adapted methanotrophs from the river Elbe and emended description of the species *Methylovulum psychrotolerans*. *Systematic and Applied Microbiology*, *44*(6). https://doi.org/10.1016/j.syapm.2021.126248

Dalton, H. (2005). The Leeuwenhoek Lecture 2000: The natural and unnatural history of methane-oxidizing bacteria. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1458), 1207–1222. https://doi.org/10.1098/rstb.2005.1657

Dedysh, S. N., Berestovskaya, Y. Y., Vasylieva, L. V., Belova, S. E., Khmelenina, V. N., Suzina, N. E., Trotsenko, Y. A., Liesack, W., & Zavarzin, G. A. (2004). *Methylocella tundrae sp. nov*., a novel methanotrophic bacterium from acidic tundra peatlands. *International Journal of Systematic and Evolutionary Microbiology*, *54*(1), 151–156. https://doi.org/10.1099/ijs.0.02805-0

Dedysh, S. N., Haupt, E. S., & Dunfield, P. F. (2016). Emended description of the family *Beijerinckiaceae* and transfer of the genera *Chelatococcus* and *Camelimonas* to the family *Chelatococcaceae fam. nov*. *International Journal of Systematic and Evolutionary Microbiology*, *66*(8), 3177–3182. https://doi.org/10.1099/ijsem.0.001167

Dedysh, S. N., & Knief, C. (2018). Diversity and phylogeny of described aerobic methanotrophs. In *Methane Biocatalysis: Paving the Way to Sustainability* (pp. 17–42). Springer International Publishing. https://doi.org/10.1007/978-3-319-74866-5_2

Dedysh, S. N., Liesack, W., Khmelenina, V. N., Suzina, N. E., Trotsenko, Y. A., Semrau, J. D., Bares, A. M., Panikov, N. S., & Tiedje, J. M. (2000). *Methylocella palustris gen. nov., sp. nov*., a new methane-oxidizing acidophilic bacterium from peat bogs, representing a novel subtype of serine- pathway methanotrophs. *International Journal of Systematic and Evolutionary Microbiology*, *50*(3), 955–969. https://doi.org/10.1099/00207713-50-3-955

Delghandi, M., Delghandi, M.P., Goddard, S. (2022). The Significance of PCR Primer Design in Genetic Diversity Studies: Exemplified by Recent Research into the Genetic Structure of Marine Species. In: Basu, C. (eds) PCR Primer Design. *Methods in Molecular Biology*, vol 2392. Humana, New York, NY. https://doi.org/10.1007/978-1-0716-1799-1_1

Deutzmann, J. S., Hoppert, M., & Schink, B. (2014). Characterization and phylogeny of a novel methanotroph, *Methyloglobulus morosus gen. nov., spec. nov*. *Systematic and Applied Microbiology*, *37*(3), 165–169. https://doi.org/10.1016/j.syapm.2014.02.001

Dougherty, K., Smith, B. A., Moore, A. F., Maitland, S., Fanger, C., Murillo, R., & Baltrus, D. A. (2014). Multiple phenotypic changes associated with large-scale horizontal gene transfer. *PLoS ONE*, 9(7). https://doi.org/10.1371/journal.pone.0102170

Eddy, S. R. (2011). Accelerated profile HMM searches. *PLoS Comput. Biol. 7*, e1002195. https://doi.org/10.1371/journal.pcbi.1002195

Gerts, E. M., Yu, Y. K., Agarwala, R., Schäffer, A. A., & Altschul, S. F. (2006). Composition-based statistics and translated nucleotide searches: Improving the TBLASTN module of BLAST. *BMC Biology*, *4*. https://doi.org/10.1186/1741-7007-4-41

Gęsicka, A., Oleskowicz-Popiel, P., & Łężyk, M. (2021). Recent trends in methane to bioproduct conversion by methanotrophs. In *Biotechnology Advances* (Vol. 53). Elsevier Inc. https://doi.org/10.1016/j.biotechadv.2021.107861

Haque, M. F. U., Xu, H. J., Colin Murrell, J., & Crombie, A. (2020). Facultative methanotrophs – diversity, genetics, molecular ecology and biotechnological potential: A mini-review. In *Microbiology (United Kingdom)* (Vol. 166, Issue 10, pp. 894–908). Microbiology Society. https://doi.org/10.1099/mic.0.000977

Hyatt, D., Chen, G.-L., Locascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). *Prodigal: prokaryotic gene recognition and translation initiation site identification*. http://www.biomedcentral.com/1471-2105/11/119

Iguchi, H., Yurimoto, H., & Sakai, Y. (2016). *Methylovulum*. In *Bergey's Manual of Systematics of Archaea and Bacteria* (pp. 1–4). Wiley. https://doi.org/10.1002/9781118960608.gbm01416

Kapli, P., Yang, Z., & Telford, M. J. (2020). Phylogenetic tree building in the genomic age. In *Nature Reviews Genetics* (Vol. 21, Issue 7, pp. 428–444). Nature Research. https://doi.org/10.1038/s41576-020-0233-0

Katoh, K., Kuma, K. I., Toh, H., & Miyata, T. (2005). MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic acids research*, *33*(2), 511-518.

Katoh, K., Rozewicki, J., & Yamada, K. D. (2018). MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*, 20(4), 1160–1166. https://doi.org/10.1093/bib/bbx108

Khadka, R., Clothier, L., Wang, L., Lim, C. K., Klotz, M. G., & Dunfield, P. F. (2018). Evolutionary History of Copper Membrane Monooxygenases. *Frontiers in Microbiology*, *9*. https://doi.org/10.3389/fmicb.2018.02493

Khalifa, A., Gyu Lee, C., Ogiso, T., Ueno, C., Dianou, D., Demachi, T., Katayama, A., & Asakawa, S. (2015). *Methylomagnum ishizawai gen. nov., sp. nov*., a mesophilic type i methanotroph isolated from rice rhizosphere. *International Journal of Systematic and Evolutionary Microbiology*, *65*(10), 3527–3534. https://doi.org/10.1099/ijsem.0.000451

Knief, C. (2015). Diversity and habitat preferences of cultivated and uncultivated aerobic methanotrophic bacteria evaluated based on pmoA as molecular marker. In *Frontiers in Microbiology* (Vol. 6, Issue DEC). Frontiers Media S.A. https://doi.org/10.3389/fmicb.2015.01346

Lee, E. Y. (2019). *Microbiology Monographs Series Editor: Alexander Steinbüchel*. http://www.springer.com/series/7171

Letunic, I., & Bork, P. (2021). Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*, *49*(W1), W293–W296. https://doi.org/10.1093/nar/gkab301

Machacova, K., Warlo, H., Svobodová, K., Agyei, T., Uchytilová, T., Horáček, P., & Lang, F. (2023). Methane emission from stems of European beech (Fagus sylvatica) offsets as much as half of methane oxidation in soil. *New Phytologist*, *238*(2), 584–597. https://doi.org/10.1111/nph.18726

Madden T. The BLAST Sequence Analysis Tool. 2002 Oct 9 [Updated 2003 Aug 13]. In: McEntyre J, Ostell J, editors. The NCBI Handbook [Internet]. Bethesda (MD): *National Center for Biotechnology Information* (US); 2002-. Chapter 16. Available from: http://www.ncbi.nlm.nih.gov/books/NBK21097/

Mahner, M., & Kary, M. (1997). What Exactly Are Genomes, Genotypes and Phenotypes? And What About Phenomes? In *J. theor. Biol* (Vol. 186).

Mattock, J., Watson, M. A (2023). comparison of single-coverage and multi-coverage metagenomic binning reveals extensive hidden contamination. *Nat Methods*. https://doi.org/10.1038/s41592-023-01934-8

Minh, B.Q., Schmidt, H. A., Chernomor, O. Schrempf, D. Woodhams, M. D., von Haeseler, A., Lanfear, R. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol*., 37:1530-1534. https://doi.org/10.1093/molbev/msaa015.

Murphy, C. L., Sheremet, A., Dunfield, P. F., Spear, J. R., Stepanauskas, R., Woyke, T., Elshahed, M. S., & Youssef, N. H. (2021). *Genomic Analysis of the Yet-Uncultured Binatota Reveals Broad Methylotrophic, Alkane-Degradation, and Pigment Production Capacities*. https://doi.org/10.1128/mBio

Nakamura, T., Yamada, K. D., Tomii, K., & Katoh, K. (2018). Parallelization of MAFFT for large-scale multiple sequence alignments. *Bioinformatics*, *34*(14), 2490-2492. Available from: https://mafft.cbrc.jp/alignment/server/

National Center for Biotechnology Information (NCBI)[Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988] – [18.07.2023]. Available from: https://www.ncbi.nlm.nih.gov/

Nayfach, S., Roux, S., Seshadri, R., Udwary, D., Varghese, N., Schulz, F., Wu, D., Paez-Espino, D., Chen, I. M., Huntemann, M., Palaniappan, K., Ladau, J., Mukherjee, S., Reddy, T. B. K., Nielsen, T., Kirton, E., Faria, J. P., Edirisinghe, J. N., Henry, C. S., … Eloe-Fadrosh, E. A.

(2021). A genomic catalog of Earth's microbiomes. *Nature Biotechnology*, *39*(4), 499–509. https://doi.org/10.1038/s41587-020-0718-6

Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274. https://doi.org/10.1093/molbev/msu300

Op den Camp, H. J. M., Islam, T., Stott, M. B., Harhangi, H. R., Hynes, A., Schouten, S., Jetten, M. S. M., Birkeland, N. K., Pol, A., & Dunfield, P. F. (2009). Environmental, genomic and taxonomic perspectives on methanotrophic *Verrucomicrobia*. In *Environmental Microbiology Reports* (Vol. 1, Issue 5, pp. 293–306). https://doi.org/10.1111/j.1758-2229.2009.00022.x

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*, *20*(2), 289–290. https://doi.org/10.1093/bioinformatics/btg412

Parks, D. H., Rinke, C., Chuvochina, M., Chaumeil, P. A., Woodcroft, B. J., Evans, P. N., Hugenholtz, P., & Tyson, G. W. (2017). Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nature Microbiology*, *2*(11), 1533–1542. https://doi.org/10.1038/s41564-017-0012-7

Peng, S., Lin, X., Thompson, R. L., Xi, Y., Liu, G., Hauglustaine, D., Lan, X., Poulter, B., Ramonet, M., Saunois, M., Yin, Y., Zhang, Z., Zheng, B., & Ciais, P. (2022). Wetland emission and atmospheric sink changes explain methane growth in 2020. *Nature*, *612*(7940), 477–482. https://doi.org/10.1038/s41586-022-05447-w

R Core Team. (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/

RStudio Team. (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA. URL http://www.rstudio.com/

Samanta, D., Govil, T., Saxena, P., Gadhamshetty, V., Krumholz, L. R., Salem, D. R., & Sani, R. K. (2022). Enhancement of Methane Catalysis Rates in *Methylosinus trichosporium OB3b*. *Biomolecules*, *12*(4). https://doi.org/10.3390/biom12040560

Setubal, J. C. (2021a). Metagenome-assembled genomes: concepts, analogies, and challenges. In *Biophysical Reviews* (Vol. 13, Issue 6, pp. 905–909). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/s12551-021-00865-y

Setubal, J. C. (2021b). Metagenome-assembled genomes: concepts, analogies, and challenges. In *Biophysical Reviews* (Vol. 13, Issue 6, pp. 905–909). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/s12551-021-00865-y

Singh, V. K., & Kumar, A. (2001). Primer Design 27. In *Molecular Biology Today* (Vol. 2, Issue 2). www.caister.com/bacteria-plant

Srinivas, V., Banerjee, R., Lebrette, H., Jones, J. C., Aurelius, O., Kim, I. S., Pham, C. C., Gul, S., Sutherlin, K. D., Bhowmick, A., John, J., Bozkurt, E., Fransson, T., Aller, P., Butryn, A., Bogacz, I., Simon, P., Keable, S., Britz, A., … Högbom, M. (2020). High-Resolution XFEL Structure of the Soluble Methane Monooxygenase Hydroxylase Complex with its Regulatory Component at Ambient Temperature in Two Oxidation States. *Journal of the American Chemical Society*, *142*(33), 14249–14266. https://doi.org/10.1021/jacs.0c05613

Stothard, P. (2000). *Internet On-Ramp The Sequence Manipulation Suite: JavaScript Programs for Analyzing and Formatting Protein and DNA Sequences* (Vol. 28, Issue 6). http://www.kazusa.or.jp/codon/

Tang, K. W., McGinnis, D. F., Ionescu, D., & Grossart, H. P. (2016). Methane production in oxic lake waters potentially increases aquatic methane flux to air. In *Environmental Science and Technology Letters* (Vol. 3, Issue 6, pp. 227–233). American Chemical Society. https://doi.org/10.1021/acs.estlett.6b00150

van Spanning, R. J. M., Guan, Q., Melkonian, C., Gallant, J., Polerecky, L., Flot, J. F., Brandt, B. W., Braster, M., Iturbe Espinoza, P., Aerts, J. W., Meima-Franke, M. M., Piersma, S. R., Bunduc, C. M., Ummels, R., Pain, A., Fleming, E. J., van der Wel, N. N., Gherman, V. D., Sarbu, S. M., … Bitter, W. (2022). Methanotrophy by a Mycobacterium species that dominates a cave microbial ecosystem. *Nature Microbiology*, *7*(12), 2089–2100. https://doi.org/10.1038/s41564-022-01252-3

Vorobev, A. V., Baani, M., Doronina, N. V., Brady, A. L., Liesack, W., Dunfield, P. F., & Dedysh, S. N. (2011). *Methyloferula stellata gen. nov., sp. nov.*, an acidophilic, obligately methanotrophic bacterium that possesses only a soluble methane monooxygenase. *International Journal of Systematic and Evolutionary Microbiology*, *61*(10), 2456–2463. https://doi.org/10.1099/ijs.0.028118-0

Whittenbury, R., Davies A N, S. L., & Davey, D. J. F. (1970). Exospores and Cysts Formed by Methane-utilizing Bacteria. In *Journal of General Microbiology (IgTO)* (Vol. 61). https://doi.org/10.1099/00221287-61-2-219

Wickham H (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4, http://ggplot2.org.

Wickham H, Bryan J (2023). readxl: Read Excel Files. https://readxl.tidyverse.org, https://github.com/tidyverse/readxl.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L. (2012). *Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction*. http://www.biomedcentral.com/1471-2105/13/134

Yu, B., Zhang, C. (2011). In Silico PCR Analysis. In: Yu, B., Hinchcliffe, M. (eds) In Silico Tools for Gene Discovery. *Methods in Molecular Biology*, vol 760. Humana Press. https://doi.org/10.1007/978-1-61779-176-5_6

Zhang, Z., Poulter, B., Feldman, A. F., Ying, Q., Ciais, P., Peng, S., & Li, X. (2023). Recent intensification of wetland methane feedback. *Nature Climate Change*. https://doi.org/10.1038/s41558-023-01629-0

Zuser, C. (2023). bachelor-thesis. Available at: https://github.com/christinazuser25/bachelor-thesis.